Title

Improving Methods to Validate Forest Growth and Quantify Landscape Tradeoffs in Fire Prone Ecosystems

Author

Herbert, Claudia

Publication Date

2022

Peer reviewed|Thesis/dissertation

Improving Methods to Validate Forest Growth and Quantify Landscape Tradeoffs in Fire Prone Ecosystems

By

Claudia Herbert

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Environmental Science, Policy, and Management

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Van Butsic, Chair
Professor Scott L. Stephens
Professor Matthew D. Potts

Spring 2022

Copyright page

Abstract

Improving Methods to Validate Forest Growth and Quantify Landscape Tradeoffs in Fire Prone Ecosystems

by

Claudia Herbert

Doctor of Philosophy in Environmental Science, Policy, and Management

University of California, Berkeley

Professor Van Butsic, Chair

Land system sciences offer techniques to study tradeoffs and solutions for ecosystems that are vulnerable to–and solutions for–climate change. This dissertation investigates tradeoffs for fire, forest, and carbon management in California. Throughout this work, I focus on methods to improve inference using remote sensing, spatial data analysis, and quasi-experimental methods. In the first chapter, I test assumptions about green landscapes as solutions to threats of fire. I use remote sensing and causal inference techniques to test whether green landscapes can alter fire severity and limit fire spread around communities, using golf courses as an example. I find that golf courses act as buffers, reducing fire severity by 49% relative to similar burned vegetation and that golf courses limit fire spread better than other vegetated landscape like regional parks. In the second chapter, I evaluate models used to make management decisions and discuss how errors from these models have landscape and policy implications. I conduct a model validation study for the widely used Forest Vegetation Simulator and find a significant overestimate in predicted growth rates relative to observed growth. I find that these errors in carbon growth rate are most pronounced on national forest lands, for stands aged 50-100 years old, forests managed as reserves, and for the least productive forest sites. Model assumptions of higher productivity than the observed growth will bias management away from more proactive silviculture techniques and can overestimate their impacts as a natural climate solution. In the third chapter, I document how land use policies for forests create tradeoffs that may undermine policymakers' objectives. I explore how operationalizing carbon management in forests creates tradeoffs between sustaining high levels of biomass, and carbon, and removing biomass to promote forest structures that are more resistant to disturbance. I find that forests managed for carbon credits are not changing fuels like forests managing for fire risk but are having changes in vegetation greenness as if they were. Throughout this work, I demonstrate how the application of rigorous methods of measurement and estimation can support improved decisions that deliver climate and ecosystem benefits for California.

Table of Contents

Acknowledgments

Thank you to my advisors, Van Butsic and Scott L. Stephens. I will forever be grateful to have learned from both of you. To Van, thank you for meeting with me weekly and giving me space to explore too many research ideas while miraculously keeping me on track. To Scott, I admire your integrity in both working with people and in science; you are a model for centering integrity in research that I hope to emulate. To Trevor Keenan, Siamak Khorram, Jonas Meckling, and Matthew D. Potts, thank you for serving as committee members on my qualifying examinations and dissertation committee. Your time and support have helped me become a better researcher and scientist.

My experience of graduate school was enriched by my delightfully different, and equally wonderful, lab groups: thank you to the former and current members of the Land Use Change and Stephens labs, especially Katherine Siegel, Carmen Tubbesing, and Jared Stapp. Thank you to my collaborator Barbara Haya for working with me on the complex landscape of carbon markets and pushing me to improve the clarity of my writing. Learning to teach has been one of the most rewarding experiences of graduate school. To the students in Spring 2019 ESPM 72 and Fall 2020 ESPM 172 classes, thank you for being a receptive bunch and sharing my enthusiasm for GIS and Remote Sensing. And to Ryann Madden, our wonderful Graduate Student Affairs Officer, thank you for making the bureaucratic parts of graduate school a little more bearable.

Validation studies are necessary but difficult; many thanks go to Jeremy S. Fried for his collaboration and for providing endless hours of guidance and support to accomplish this work. Thank you to the FIA field crews for collecting the field data, the FIA office team for analysis, and the FVS help desk for being incredibly helpful and responsive to emails—my work would not have been possible without all the people I never met. I am also grateful to William Stewart for providing feedback on this work.

Thank you to the professors that intensely informed how I think during my undergraduate years at Berkeley, especially Kate O'Neill, Ignacio Chapela, Laurel Larsen, and Michael O'Hare. You all make your own research so compelling, it's hard not to develop interdisciplinary interests. And to my dear Forestry Camp friends, spending a summer learning in the woods sparked my research interest in graduate school. I cannot imagine sharing that formative experience with anyone else.

For this dissertation, I was able to answer questions about California with rich resource input because of advances in cloud-based computing and open data science. Thank you to the contributors to the open data science movement from OpenStreetMap, R Core Team, the legion of individuals answering questions on developer forums, open access products like Google Earth Engine, and additional Python packages like 'geemap' from Professor Qiusheng Wu.

My academic journey would not have been possible without the early advocacy and support of my parents and sisters. To my parents, thank you for helping me navigate a world where the letters won't sit still. Reading is the best gift I've been given, and I hope to have made your efforts worthwhile. To my sisters Natalie, Gloria, and sister by choice, Ally, thank you for being my eternal copyeditors and role models. Thank you to my big bros Jake and Jordan and nephew Leo for the constant love. To my amazing partner Max, you have been the best part of the past four years. Thank you for being an incredible source of joy and support throughout graduate school.

Introduction

Managing natural resources often has distributed costs and benefits that creates winners and losers. How these distributions appear across different scales—spatially and temporally—is a part of land systems sciences (Meyfroidt et al. 2022). For example, land managers make practical decisions on the use of available resources to reduce near-term fire risk while weighing the cost of additional resource inputs. They may also use models to support their decision-making without clarity on how model assumptions may inaccurately predict vegetation dynamics. Resource managers may also consider establishing new streams of income like carbon credits to help subsidize management that may create new constraints or incentives for how they weigh risk of fire. In these scenarios and others, improving the measures and methods for researchers and managers may be a key intervention in supporting better land use outcomes. Monitoring the initial outcomes of resource decisions and considering unintended consequences can gear land management policies towards more win-wins or less lopsided distributions in costs and benefits.

In this dissertation, I address questions that involve tradeoffs at the intersection of fire, forest, and carbon management in California. Chapter 1 asks whether irrigated green landscapes protect communities from fire. By establishing whether irrigated landscapes can protect communities and quantifying their buffering effect, the resources required to maintain these irrigated landscapes can be weighed against the potential benefit. Next, open-access, spatially comprehensive forest vegetation models like the Forest Vegetation Simulator are important for landscape forest management, but these models may lack accuracy. Chapter 2 asks: how valid is this openly accessible model, and for which lands does the model perform better? By understanding whether the model produces valid growth predictions and where the errors are more pronounced, resources can be targeted for model calibration and improvement. Finally, Chapter 3 asks: how do policies incentivizing increasing aboveground carbon manage the associated fire risk? Understanding how forest carbon projects manage fire risk can help inform what tradeoffs they may create with fire and other disturbance risks. This dissertation combines existing methods and datasets in innovative analyses to answer these questions.

To collect data for this research, I use a combination of remote sensing and spatial data science to adapt existing datasets to be used in quasi-experimental designs. Remote sensing has been an important tool in land system sciences (Dong et al. 2019). Open access datasets and cloud-based image processing hosted on platforms like Google Earth Engine have increased the capacity of researchers generating more specialized remote sensing analysis (Gorelick et al. 2017). This boon in image processing lets researchers test assumptions and generate data at previously cost-prohibitive scales. I also rely on open-source data collected by groups like OpenStreetMap and publicly available datasets collected by public agencies like National Aeronautics and Space Administration (NSAS), United States Geological Survey (USGS), and the United States Forest Service. Open access and existing datasets are often available in more regions and have long archives that allow the methods demonstrated in this dissertation to be adapted to other contexts and periods.

To analyze data collected for this research, I use inferential statistics, including in quasi-experimental designs. Inferential statistics and multivariate analyses improve our ability to understand associations where many attributes can differ between observations over time. Using statistical methods, I estimate how biophysical factors explain historic outcomes. In Chapters 1 and 3, I design tests to improve comparison between groups of observed factors using quasi-experimental designs. These studies leverage propensity score matching as a method to create control groups, which improve our ability to estimate how different treatments compare. Many research questions that relate to land system sciences want to establish causal order, but randomized controlled trials are likely impractical for the problems I investigate. The quasi-experimental methods I apply to these research questions offer a rigorous, data-driven approach that can better establish internal and external validity than other observational methods.

In Chapter 1, I quantify how a given landscape has altered fire severity and limited fire spread using empirical observations of fires between 1986 through 2020. Prior to this work, green landscapes like golf courses or vineyards were commonly thought to offer a buffering effect between wildland fires and built infrastructure like housing developments (Chirouze et al. 2021; P. Gross 2008). This research pulls from satellite imagery, datasets of golf course locations, and several other biophysical variables to i) test whether this buffering effect can be observed in the data and ii) quantify the strength of the effect. Simply comparing fire severity outcomes on golf courses with non-golf course vegetation would lead to a finding of a buffering effect, because golf courses probably differ from non-golf courses in ways that would confound a comparison. I use propensity score matching to identify similar non-golf course controls and perform a multiple linear regression to estimate the treatment effect of a golf course relative to similar non-golf course vegetation. Another way a golf course could buffer a community is by not burning—i.e., stopping fire spread. Observing that golf courses frequently occur on the boundaries of fire perimeters is not an endorsement that they limit fire spread. I develop a way of comparing a given landscape's overlap with a fire perimeter to establish how that landscape limits fire spread relative to other types of landscapes.

In Chapter 2, I perform a model validation study and find that the errors produced by the Forest Vegetation Simulator (FVS) are greater for some forest conditions than others. These non-random errors are important for informing model calibration and for identifying areas of needed model improvement. In the largest model validation study of FVS spanning four model variants, I first compare model outputs with real world observations using equivalency tests. To further diagnose where these model predictions are inaccurate, I use linear regression and prediction to estimate the amount of error the model produces across different types of forest properties and growing conditions. This method lets me estimate the size and direction of error (e.g., under or overestimate). Forest vegetation models like FVS are one way people make long term decisions about managing forests. I find that at a decadal projection, the model predictions of growth are not like the observed growth rate. This presents a problem because the expected growth from this model will be both unrealistic and more pronounced systematically for the same types of forests.

In Chapter 3, I estimate the management and trends in vegetation productivity for forests that are generating carbon credits as improved forest management (IFM) projects under California's compliance market. Using propensity score matching in a quasi-experimental design, I compare these results to management in forests that have received fuel treatments to assess how carbon management manifests in forest practices. I find that between 2016 and 2020, the average forest carbon project in California was not altering fuel loads consistent with fuel management. Using a Kendall-Mann test for vegetation trends on a timeseries of satellite imagery, I find a declining trend in vegetation greenness for IFM projects in the North Coast, which is a proxy for vegetation productivity. Vegetation greenness on IFM projects in both the North Coast and Southern Cascade regions declined in greenness at a similar rate to forests that received fuel treatments. Because we observed that these projects are not reducing fuels, which could cause a decline in vegetation greenness, I conclude the management the is occurring within these forests is not improving forest health or reducing fire risk.

My research is designed to help close gaps between assumptions and system-level drivers using available data to inform decision making. California is a compelling research setting because it shares a physical climate with other places (e.g., Mediterranean, Australia) while having a singular climate for policy on climate action (Sneeuwjagt, Kline, and Stephens 2013; Stephens et al. 2009). In addition, California is a leader in environmental and climate action, making insights learned here more able to inform policy intervention (Vogel and Kagan 2002). The investment in public datasets and the state-lead commitment to mitigating climate change make it a promising setting to study these resource management questions.

References

Chirouze, M.; et al. 2021. "Quantifying Insurance Benefits of a Nature-Based Approach to Reducing Risk: Wildfire Risk Reduction Buffers." https://www.marshmclennan.com/content/dam/mmc-web/insights/publications/2021/december/MMC_TNC_Quantifying_Insurance_Benefits.pdf.

Dong, Jinwei et al. 2019. "Remote Sensing and Geospatial Technologies in Support of a Normative Land System Science: Status and Prospects." Current Opinion in Environmental Sustainability 38: 44–52. https://linkinghub.elsevier.com/retrieve/pii/S1877343519300041.

Gorelick, Noel et al. 2017. "Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone." Remote Sensing of Environment 202: 18–27. https://linkinghub.elsevier.com/retrieve/pii/S0034425717302900.

Gross, P. 2008. "Golf Courses on the Fire Line." https://gsrpdf.lib.msu.edu/?file=/2000s/2009/091113.pdf.

Meyfroidt, Patrick et al. 2022. "Ten Facts about Land Systems for Sustainability." 119(7): 1–12.

Sneeuwjagt, Rick J., Tim S. Kline, and Scott L. Stephens. 2013. "Opportunities for Improved Fire Use and Management in California: Lessons from Western Australia." *Fire Ecology* 9(2): 14–25.

Stephens, Scott L et al. 2009. "Urban–Wildland Fires: How California and Other Regions of the US Can Learn from Australia." *Environmental Research Letters* 4(1): 014010. https://iopscience.iop.org/article/10.1088/1748-9326/4/1/014010.

Vogel, D., and R. Kagan. 2002. "Dynamics of Regulatory Change: How Globalization Affects National Regulatory Policies An Introduction." National Regulations in a Global Economy: 1–28.

Chapter 1: Assessing the effectiveness of green landscape buffers to reduce fire severity and limit fire spread in California: Case study of golf courses

Claudia Herbert, Van Butsic

This article has been published previously and is reproduced here with permission from co-authors and the Graduate Division

Herbert C, Butsic V. Assessing the Effectiveness of Green Landscape Buffers to Reduce Fire Severity and Limit Fire Spread in California: Case Study of Golf Courses. *Fire*. 2022; 5(2):44. https://doi.org/10.3390/fire5020044

Abstract

Communities looking to improve fire protection may consider incorporating landscape features that 'buffer' the effects of a fire between developed and undeveloped lands. While landscapes such as golf courses, vineyards, or agriculture are already being considered part of this buffer zone, few empirical studies demonstrate causally how well these different landscape features operate as a fire buffer. This research selects golf courses as an example of a possible buffer landscape and proposes methods to test if this buffer alters fire severity and limits fire spread. Using propensity score matching and multiple linear regression, we demonstrate golf courses that burned in California between 1986 and 2020 had a predicted 49% reduction in fire severity relative to otherwise similar vegetated land. This reduction in fire severity is regionally dependent, with the effect of golf course buffering landscapes most pronounced in the North Bay region. For limiting fire spread, golf courses function similarly to hardscaped land uses such as airports, suggesting that irrigation and vegetation management can be effective in creating desired buffering qualities. These methods suggest that artificially created irrigated green zones act as effective buffers, providing de facto fuel breaks around communities, and can be reproduced for other potential buffering landscape features. This study does not advocate for the use of any specific anthropo-genic landscape feature, but rather highlights that community-based fire hazard reduction goals could be attained through considering landscape features beyond fuel reduction manipulations.

Keywords: WUI; vegetation fire; propensity score matching; fire severity; landscape ecology

Introduction

Increasingly destructive fire seasons in the Western United States underscore the need for people and communities to proactively plan for living with fire (Keeley and Syphard 2019; Kramer et al. 2019; Schoennagel et al. 2017). A suite of approaches, ranging from reducing fuels near homes to changing building materials or the spatial configuration of buildings, are considered fire risk mitigation strategies. These mitigation strategies are actions people or communities can take prior to a fire to reduce the harm from a fire event or reduce the chance of ignition (Calkin et al. 2014; Moritz et al. 2014). Fire risk mitigation strategies can occur across a range of spatial scales, from the individual building to the neighborhood or community level (Moritz and Butsic 2020). Regardless of the spatial scales over which these strategies are implemented, their core purpose is to slow or prevent fires from spreading, hopefully reducing fire intensity enough to provide opportunities for fire suppression agencies to safely control fire spread. Such strategies require maintenance to retain the same fire behavior-altering benefits over time (Agee et al. 2000).

At the community and neighborhood scales, strategic land use planning in the Wildland Urban Interface may be an effective wildfire risk mitigation tactic. Strategic planning includes considering how to arrange different land uses to benefit from changes in fuel availability or conditions. An example is to 'buffer' between vegetated lands and human developments (M. Moritz and Butsic 2020). Proponents of buffering communities point to landscape features such as orchards, vineyards, parks, and golf courses as landscapes that could interrupt fuel continuity and limit wildfire spread, acting as de facto fuel breaks (Chirouze et al. 2021). A benefit of this approach is that these are often economically generating land uses, so the upkeep necessary to deliver the fire risk benefit is tied to its productive activities; therefore, fire risk benefits happen without policy interventions or public funding, differing from more traditional fuel break maintenance. While the idea of incorporating buffering landscapes seems intuitive, outside of case studies, there is limited empirical evidence across broad spatial scales to quantify the impact of such buffer zones (Chirouze et al. 2021; Patrick Gross 2008).

One reason for this lack of evidence is that it is methodologically difficult to quantify buffering effects across a range of geographies. To formalize whether a landscape feature acts as a 'buffer,' we propose that effective buffers should both (i) reduce fire severity when burned and, (ii) more optimistically, not burn—limiting fire spread, either by not having enough fuel to carry fire or by providing access for suppression. By quantifying these two properties, we can measure a given landscape feature's buffering capacity, allowing it to be evaluated against competing land uses or fire risk mitigation strategies.

Fire severity is a measure of biomass or soil change used to gauge how intensely an ecosystem was impacted by fire (Lentile et al. 2006). It is directly related to the amount of energy a fire produced, also called fire intensity (Keeley 2009). When fire intensity is high, direct fire sup-pression is less effective and carries greater risk for fire suppression crews (National Wildfire Coordinating Group n.d.). Reduced fire severity can indicate areas of reduced fire intensity, possibly providing opportunity for effective fire suppression. However, trying to establish that a landscape reduces fire severity causally requires a control observation to compare outcomes. Using golf courses as an example, identifying a control group can be difficult outside of experimental settings because factors that influence where a golf course is located could also influence fire severity, such as slope, vegetation type, or vegetation moisture. Therefore, establishing that a reduction in severity occurred on a buffer requires identifying statistically similar controls that burned in the same fire or under similar conditions. Fortunately, pre-processing methods such as propensity score matching are an established method for identifying such quasi-experimental controls and have been used in numerous land use policy evaluations (Butry 2009; Butsic et al. 2017; Ramsey et al. 2019; Woo, Eskelson, and Monleon 2021). With a pre-processed dataset, analysis such as linear regression can be used to predict treatment effect of buffers on fire severity.

Second, an effective buffer should not burn, either because it is not flammable or because it provides access for suppression. If a landscape feature does not carry fire when a fire approaches, it could be considered to limit fire spread. However, limiting fire spread is another challenging concept to test using observational data. One approach could be comparing fire boundaries with buffering landscape features to examine the frequency of certain types of landcovers near a fire edge. However, observing a landscape feature at the edge does not necessarily indicate that the landscape feature had a special property that limited fire spread. Instead, a landscape feature could occur near a fire edge due to chance or due to an unrelated process such as ease of delineating a cartographic boundary near an established edge of a large landscape feature. For possible buffering landscapes that are observed near a fire edge, we could test if there is something special about the type of landscape by comparing it with otherwise similarly shaped and sized vegetated landscapes. This would allow for testing the hypothesis that certain types of landscape management confer advantages in limiting fire spread.

We selected golf courses as a proof of concept for testing the methods developed here to evaluate buffering capacity. In California, golf courses are found throughout the state and have a different vegetation and irrigation intensity than surrounding vegetation. Golf courses are potentially effective as fuel breaks because they are typically at least 150 feet wide in most places and consist of mowed grasses—consistent with fuel break design (CAL FIRE (California Department of Forestry and Fire Protection) n.d.). Importantly, golf courses are already considered part of 'Buffer Zones' by the National Wildfire Coordinating Group, because they are areas of reduced vegetation separating wildlands to vulnerable residential or business developments (National Wildfire Coordinating Group 2015). We propose two sets of quantitative methods for evaluating a given land-cover's observed buffering capacity when exposed to a fire, to answer:

1. Do golf courses alter fire severity relative to similar vegetation?
2. Do golf courses limit fire spread? How does this compare to other landscape features like parks or airports?

Materials and Methods

Study Period and Regions

The study region included fires in California that burned between 1986 and 2020 on or adjacent to golf courses. The perimeters of golf courses in California were gathered from OpenStreetMap Overpass Turbo API by querying for "golf course" under the "leisure" category (OpenStreetMap Contributors. 2015; Overpass Turbo Golf Course Query n.d.). These golf course polygon returns were limited to the state of California, determined by the California TigerLines Shapefile in ESRI's ArcGIS Pro (version 2.7.1) (ArcGIS Pro n.d.; TIGER/Line: Current State and Equivalent National 2019). Overlap between golf courses and fires were determined using CalFire California Fire and Resource Assessment Program (FRAP) fire perimeter data (CAL FIRE (California Department of Forestry and Fire Protection) n.d.). The fire boundaries reduced the 961-golf course dataset from OpenStreetMap in the state to 89 golf courses that had some spatial overlap with fire.

To answer the first question, if golf courses change fire severity outcomes, we examined 22 fires that burned at least a quarter of a golf course (Figure 1). We used the 25% area cutoff to ensure that there would be sufficient golf course area to sample for our statistical procedures. In total, we examined 29 different golf courses. Nine fires burned more than one golf course and three golf courses burned twice in the multi-decade study period, meaning golf courses burned 32 times over the 35-year study period.

The second analysis, on whether golf courses appear to limit fires spread, spanned 122 cases of golf courses that intersected a fire perimeter between 1986 and 2020. To interpret the effectiveness of golf courses relative to other landscape features, we also looked at parks and airports in California that burned or had some overlap with fire between 1986 and 2020. We limited the park data to be similar in size to golf courses by creating a cutoff of one standard deviation of golf course sizes, or 36 acres to 226 acres. Concerning vegetation management, we assumed that parks of this size would have less water-intensive management than golf courses but may have some fuel management (e.g., grazing, chemical or mechanical removal, prescribed fire). On the other hand, airports are likely managed for more limited vegetation with more hardscape than a golf course or regional park but, such as parks, are unlikely irrigated. This add-ed 121 parks and 49 airports to our study, also queried from OpenStreetMap Overpass Turbo API (Overpass Turbo Airport Query n.d.; Overpass Turbo Park Query n.d.).

Fire Severity Data

Fire severity is a measure of how fire intensity affected ecosystems (Keeley 2009). We approximate this initial impact from fire as changes in vegetation surface reflectance from before and after a fire. Data for fire severity was generated in Google Earth Engine (GEE), which provided access Landsat 5 Thematic Mapper (TM), Landsat 7 Enhanced Thematic Mapper (ETM+), and Landsat 8 Operational Land Imager and Thermal Infrared Sensor (OLI/TIRS) archived imagery (Gorelick et al. 2017). We used GEE to process vegetation fire severity maps for the 22 fires that burned 29 golf courses and generated differenced Normalized Burn Ratio with offset ($dNBR_{offset}$, eq. 1) maps to calculate severity. The $dNBR_{offset}$ is the average dNBR value from pixels in relatively homogenous, unchanged areas outside the fire perimeter, intended to account for differences in phenology or precipitation between pre- and post-fire images (Parks et al. 2018). This GEE script is based on the Parks et al. 2018 paper (corrected in 2021) and used the same cloud masking algorithm, three-sensor harmonizing method, and $dNBR_{offset}$ calculations described in that paper.

$$dNBR_{offset} = ((NBR_{prefire} - NBR_{postfire}) \times 1000) - dNBR_{offset}, \quad (1)$$

The pre-fire date was determined using imagery composited from one and a half months prior to the fire 'alarm date,' recorded in CalFire FRAP. This six-week window ensured that there was enough imagery to create a clear-sky composite of vegetation pre-fire. Similarly, the post-fire images were selected one and a half months post fire 'alarm date.' Some fires in the CalFire FRAP include containment dates that could have been used as the end date of fire, but these dates are not available for all relevant fires and may have more bureaucratic relevance than ecological importance, as these dates reflect when post-fire surveys determined that fires ended (Data Element Standard n.d.). Using the imagery immediately after fire, we calculate an initial burn severity to capture the immediate change from fire, closer approximating fire severity instead of an extended burn severity measure (Keeley 2009). We expected golf courses that did burn to receive swift management actions, meaning that using imagery too long after a fire would include the management response more strongly than adjacent vegetation. While we are interested in approximating fire severity using remote sensing data, oftentimes, this same technique is used to produce burn severity maps and the terms are used interchangeably (Keeley 2009). We refer to our produced data as burn severity values for consistency within the remote sensing literature but discuss our results in terms of fire severity.

After maps of continuous burn severity values are produced, burn severity data is often categorized using ground-truth measures of vegetation burn severity. Monitoring Trends in Burn Severity (MTBS) is a concerted multi-agency effort to produce such categorical burn severity maps (MTBS Project (USDA Forest Service/U.S. Geological Survey) n.d.). Because MTBS was produced for only 15 of the 22 fires relevant to this study, and ground-truth data was not available for all the study sites, we used the raw continuous approximations of burn severity for analysis. As a robustness check, we sampled the categorical MTBS maps for the available 15 fires at the same points as our continuous burn severity estimates. We used the MTBS categories to develop distributions of continuous burn severity values to interpret our predicted $dNBR_{offset}$ values.

Propensity Score Matching and Linear Regression

Statistically, comparing burned golf course vegetation to vegetation outside of a golf course can lead to biased results because the treated and control observations may be different in ways that are not observed by the researcher. To overcome this, we used propensity score matching (PSM) to analyze nonrandom, observational data by creating a valid set of control and treatment observations (Austin 2011a). The treatment and control points were first generated using a 100 m grid that considered all points within the burned portion of a golf course perimeter to be ‘treatment,’ and all non-golf course points within a fire perimeter to be ‘control.’ Any vegetated pixel within the burned portion of the golf course was eligible for sampling, regardless of whether it was on the fairway or more densely vegetated fairway adjacent areas. The PSM used biophysical and socioeconomic variables we thought would be relevant for identifying non-golf course lands that are otherwise like golf courses, such as the slope, aspect, vegetation moisture, total rainfall, latitude, vegetation type, and average household income (Table 1).

We implemented the PSM using matching software for causal inference, MatchIt (version 4.3.0) package in R Studio using the package ‘dplyr’ to organize data (Ho et al. 2011; R Core Team 2021; RStudio Team 2020; H. Wickham et al. 2021). Our model used nearest neighbor matching without replacement and matched points exactly on fire name and landcover to ensure that comparisons about fire severity would be contained by fire and landcover. We used a caliper, or cutoff of maximum difference, of 0.1 to remove any treatment observations that did not have a control observation sufficiently similar. After matching, we have a dataset with a control observation similar to each of the remaining treatment observations, which allows us to use regression analysis to estimate the treatment effect.

To identify the treatment effect on the $dNBR_{offset}$ burn severity outcomes, we then applied multiple linear regression to a dataset consisting only of matched observations. Along with the treatment group, slope, latitude, NDMI from three to six months pre-fire (NDMI 6), total precipitation from October–December (Precipitation 3), median household income, and landcover used in the matching process as explanatory variables, we added a categorical regional variable to indicate whether the fire was in the Bay Area (northwest), North Interior (northeast), or Southern California (south). These regional categories are consistent with Syphard et al. (2021) that found housing pattern and vegetation impact on housing structure loss varied by region of California (Syphard, Rustigian-Romsos, and Keeley 2021). This region variable is included in the linear regression as an independent variable and an interaction term with treatment. Adding region to the regression as an interaction with treatment allowed us to calculate the treatment effect on burn severity for each region separately. This allows us to investigate whether the treatment effect of golf courses vary by region.

We implemented the multiple linear regression in the base 'stats' package in R. To interpret the linear regression output, we predicted the average marginal effect using the 'margins' package in R (version 0.3.26) (Leeper 2021). To predict and plot the regional and treatment interaction term, we used the 'emmeans' package in R (version 1.6.3) (Lenth n.d.). The difference in the predicted treatment $dNBR_{offset}$ from the control group's prediction is the estimated treatment effect of vegetation being managed as a golf course instead of a similar non-golf course. To interpret the continuous $dNBR_{offset}$ predicted values for the regional treatment effects, we use an empirical cumulative distribution function on regionally subset distributions of MTBS categories (unburned, low, moderate, and high) to calculate the predicted values percentile.

Measuring How Golf Courses Limit Fire Spread

For a landscape feature to be an effective buffer, it should both reduce fire behavior and limit fire spread. It is difficult to measure how landscape features limit spread, because a golf course occurring on the edge of a fire perimeter could be the result of a random process or cartographic convenience (using green golf courses to delineate fire boundary) instead of a buffering characteristic of the golf course. To test how golf courses could limit fire spread, we looked at (*i*) how frequently fires stop at golf courses and (*ii*) if this is different than other landscape features, such as parks or airports.

To calculate the frequency of fires stopping at golf courses, we used the ArcGIS Pro (Version 2.7.1) Tabulate Intersection tool to determine the amount of overlap between the OpenStreetMap golf course perimeters and the FRAP fire perimeters (ArcGIS Pro n.d.). This process included all golf courses that had some overlap or touched the edge of a fire between 1986 and 2020. Golf courses that had close to 0.0% percent overlap with fire represent areas where the golf course may have limited fire spread. We then used 'ggplot2' package in R to plot the distribution of golf course burned proportions (H. Wickham 2016). To help interpret how golf courses may limit fire spread, we recreated the analysis for parks and airports, landscape features that are similar in shape and size to golf courses but have different vegetation management and may exhibit different abilities to limit fire spread. The differences in expected vegetation management between golf courses, parks, and airports is how we test whether an effect of limiting fire spread is present, despite these landcovers being similar in size and shape as landscape features.

To test the differences in the proportion of overlap between the three types of land use, we use a Shapiro–Wilk test to determine that the distributions of proportion burned for the golf courses, parks, or airports are not normally distributed. We then performed a Mann-Whitney-Wilcoxon test to test for similarities in distribution between golf courses and the two alternative landcovers, parks and airports. A Wilcoxon test does not assume a normal distribution of observations and can test for similarities of distributions where the samples are independent (McKnight and Najab 2010). The null hypothesis is that golf courses and parks or airports burn at a similar proportion. If we reject the null, we conclude that the groups are different in terms of the proportion burned when they are near or inside of a fire perimeter. We performed a two-sided Wilcoxon test because we did not want to assume the distributions would always be higher or lower on airports or parks, relative to golf course distributions.

Along with comparing distributions, it may be useful to compare the most frequent burn percentages between golf courses with airports and parks. To do this we implemented a two-sample bootstrap test to sample medians with 2500 replications (Peng 2019). This approach allows us to calculate the differences in medians observed between the different landcover samples. If the distribution of the 95% of differences contain zero, then we can say that the medians between the sample are similar.

Robustness Checks

We were concerned that this polygon-overlap approach may depend on the precision of the polygon perimeters; differences in precision between polygons used for golf courses, airports, and parks could lead to erroneous conclusions. As a robustness check, we compared the OpenStreetMap park data with a well-established park dataset, coming from the California Protected Areas Database (CPAD) Units (About CPAD n.d.). We assessed agreement using a two-sample bootstrap to compare the median values with 2500 replications and found that there are no statistical differences between medians of the two sources of park data using a Confidence Interval of 95%, which suggested that the OpenStreetMap data sufficiently captured the expected burn proportion from a reputable source such as CPAD.

Moreover, we considered that determining which golf courses, parks, or airports were near fire edge may be sensitive to the precision of a FRAP fire perimeter. We constructed 10 and 20 m buffers around fires to compare how our selection of golf courses, parks, and airports could change if the fire perimeter was slightly different. We found that the number of golf courses, parks, and airports that would have been included in our analysis if the fire perimeter were 10- or 20- meters wider was relatively consistent with our original selection. If the fire perimeters had been 20 m wider, this would have increased the number of golf courses included in the study by a 7 percent difference, added zero additional parks, and added 4.8 percent airports. This exaggerated buffer adding few airports and golf course gave us confidence that the selection of golf courses, parks, and airports were relatively unsensitive to imprecisions in the fire perimeter dataset.

Results

Quality of Burn Severity and Matched Data

This PSM reduced the mean differences between treated and control observations for the continuous variables included in the linear regression (Figure 2). The caliper and exact matching on landcover and fire name removed 202 treatment observations that did not have a sufficiently similar control observation. This process reduced our dataset from 363,843 observations with 1572 treatment samples to 2740 observations of matched treatment and control points. The mean differences between the control and treatment datasets improved the most for slope, precipitation between October and December, NDMI three to six months prior to fire, and latitude. Because the PSM resolved differences between the average control and average treatment for all continuous variables, we felt that this matched dataset was well-suited for linear regression and prediction.

Linear Regression and Predicted Treatment Effect

According to our regression and prediction analysis, the treatment effect of a golf course across all California regions reduced dNBR fire severity by an average of 65 units, compared to otherwise similar vegetation (Figure 3). A lower $dNBR_{offset}$ for golf courses indicates less change in surface reflectance post fire. This difference between the predicted control (163.11) and treatment (97.96) is a 49.91% difference. Because there is no overlap between the predicted 95% confidence intervals, we conclude that this treatment effect difference between golf courses and otherwise similar non-golf course vegetation is statistically significant. However, because $dNBR_{offset}$ values are most meaningful within the context of a single fire, and these predictions include fires across California over decades, interpreting fire severity outcomes is less straightforward.

Based on the regional and treatment interaction term predictions, we found that the treatment effect does vary by region in California (Figure 4). We found the largest treatment effect in the northwest region. Based on the northwest predicted treatment average of −41.49 to the predicted control at 99.90, we used the MTBS regional distributions to assess where in the burn severity categories these predictions fall (Table 2). We suspect that the treatment effect in the northwest is roughly a whole burn severity class. The northwest treatment prediction falls at the 10^{th} percentile of the 'unburned' category, and the northwest control prediction is above the 63^{rd} percentile for 'unburned' and around the 23^{rd} percentile for 'low.' Because there is no overlap between the predicted 95% CIs, this treatment effect is statistically significant.

The south region, similarly, has no overlapping 95% CI, suggesting these differences are statistically significant. Predicted $dNBR_{offset}$ values for the south treatment group, 99.44, and control group, 164.59, were the highest of the three regions. When we compare these predicted values to the distribution of MTBS categories, we see that these predicted values are still falling towards the middle of the unburned or low-severity fire distributions, despite the higher raw predicted $dNBR_{offset}$ values.

The northeast predicted and controls do have some overlap in their CIs, but using a two-sided student's T-test, we find that these distributions are still significantly different at a .05 p-value cutoff. The northeast control group is burning at the 52nd percentile for 'unburned' and the 31st percentile for 'low' suggesting that it is somewhere between those groups. The northeast treatment is burning at the 18th percentile for the unburned severity and 10th percentile for low severity, suggesting it is mostly unburned. This distribution comparison between treatment and controls is very similar to the south distributions for MTBS. We believe that both regions have a treatment effect where golf courses are burning at lower severities, but it is not clear that this effect necessarily translates to a burn severity class difference.

Golf Courses Limiting Fire Spread

When a fire burns near a golf course, golf courses most frequently burn less than 10% of their area (Figure 5). The observed distribution of how much of a golf course, park, or airport burn exhibits bimodal peaks; most burning completely or hardly at all. Because the total number of golf courses, parks, and airports differ, to facilitate comparisons between the features, we produced a smoothed density estimate (Figure 5, panel b). This smoothed density estimate makes the bimodal distribution more apparent and reveals a right skew for the airport and golf courses and a left skew for the park dataset. This means that golf courses and airports tend to burn closer to zero when near a fire, whereas parks found near a fire are more frequently entirely contained within the fire perimeter.

Using a 5% significance level, the Shapiro–Wilk test indicated that the proportion of burned areas for golf courses, parks, and airports were not normally distributed. This is consistent with the bimodal distribution observed in the histograms. The Wilcoxon test (Table 3) found golf courses and parks were significantly different, so we accept the alternative hypothesis, that the golf courses and parks are different. Conversely, the golf courses and airports test were not statistically significate at this threshold, indicating evidence for the null hypothesis, that the distributions are equal.

Using the two-sample bootstrap, we found that the differences in the median percentage of golf course burned was not distinguishable from the airport median (Figure 6). Parks, on the other hand, did not have a 95% confidence interval that contained zero, indicating that the median differences between the percentage of golf and parks burned are not similar. These bootstrapped median differences are consistent with the results of the Wilcoxon test, where golf courses are like airports in how they burn but parks are not.

Discussion

Choosing to promote buffer zones around communities is a resource decision. Land use planners and communities should be able to access empirical data to assess a land use's buffering capacity to weigh tradeoffs in making decisions about how to allocate limited resources such as water, fire suppression efforts, or undeveloped land. Here, we provide empirical evidence for the effectiveness of golf courses as fire buffers across California along with the methods for others to recreate this analysis for other landcovers.

Our findings suggest that when a golf course burns, it tends to burn at a reduced severity relative to similar vegetation. Our regression analysis found that this amounted to a 49% difference between golf courses and non-golf course controls. This demonstrated reduction in fire severity fulfills the first criteria we define as 'buffering capacity.' If the reduced fire severity is the result of reduced fire intensity, then these landscape features may provide buffers where suppression crews could safely access and more effectively protect adjacent communities. Because $dNBR_{offset}$ is a metric best used to interpret fire or burn severity within an individual fire, these predicted values are meant to illustrate the direction and significance of golf courses treatment effects. Importantly, there are regional differences in the size of the fire severity treatment effect. We found evidence that the northwest region had the largest treatment effect, possibly reducing burn severities an entire MTBS category, from low severity to unburned. The south and northeast regions still had statistically significant treatment effects, but it is less clear that these differences translate to a whole burn severity category reduction.

One limitation of this research is our reliance on using fire severity as an approximations of fire behavior. Ultimately, for fire suppression and community protection, fire behavior metrics such as fire intensity are more meaningful. Advances in producing fire intensity measurements from thermal heat signatures, captured by satellites such as MODIS, VIIRS, or GOES, show promise for estimating Byram's Fire Intensity (Giglio, Schroeder, and Justice 2016; Ruecker, Leimbach, and Tiemann 2021). Our proposed approach could be adapted to incorporate fire intensity estimates instead of fire severity. However, a challenge inherent in a thermal-based fire intensity approach is that passive thermal energy is recorded at a coarser spatial resolution than spectral data (e.g., 250 m^2 versus 30 m^2 pixel sizes), so it may not be sufficient for studying smaller landscape features such as golf courses. Moreover, because golf courses still burn infrequently, averaging less than one golf course over this 35-year study period, a longer archive of imagery was useful for creating a dataset with sufficient observations to compare across regions and a range of vegetation. Future research to improve this method could incorporate field measurements of fire intensity and connect the field measurements with the high spatial-coverage of satellite-derived fire severity. This field data is not available for our historic dataset and requires the ability to collect data during a fire. Alternatively, instead of looking at the fire severity impacts, researchers could also examine other adverse fire consequences such as structure damage from fires.

Despite these shortcomings, the robust treatment effects we observe across all of California and within the region-specific treatment effects suggest that our approximation of fire severity is sufficiently capturing the golf course treatment effect. While our use of initial fire severity is intended to limit the role of post-fire recovery from influencing our results, the effect of swift recovery on more intensely managed landscapes such as golf courses may be influencing our findings.

Through comparing the proportion of golf course area that burns or is adjacent to fires, we found that the low proportion of burning suggests evidence that they act as buffers, limiting fire spread. Golf courses exhibited this low proportion of area burned similar to other landcovers, such as airports, that we expected to be effective at limiting fire spread given the hard-cover and regular vegetation maintenance airports receive. Golf courses did not burn like similarly sized parks, consistent with our expectations of the role of management, and suggesting that our way of approximating this true absence of golf courses burning is valid.

Despite evidence that golf courses may be effective buffers, both reducing fire severity and limiting fire spread, there still may be limitations in their effectiveness that are not apparent in our analysis. Our work relies on observations from historic fires, yet many of the most recent destructive fires in California have been notably wind-driven and may not be well represented in our observational dataset (Keeley and Syphard 2019). During these wind-driven fires, golf courses may not provide a wide enough buffer to protect structures from wildland fires, failing to provide effective buffer capacity (Syphard, Keeley, and Brennan 2011). This is where considering what other types of ‘buffering’ landscapes could be combined to create wider buffer matrix could be useful.

Being able to quantify the relative effectiveness of a land use allows planners, communities, and emergency responders to identify key parts of their landscapes that should be maintained or augmented as part of fire preparedness. The effect we report from the linear regression and polygon analysis likely captures some effect of additional fire suppression resources used on these landscapes, rather than solely a physical advantage. For example, golf courses may have more extensive fire suppression activities around them than a similarly vegetated landscape that is not generating similar economic activity per acre. We believe that this factor is why it is important to not overstate the role that vegetation management pre-fire alone contributes to our findings. Still, even if we are measuring the effects of emergency management responses to protect certain landscapes, it is an important step forward in the planning and fire risk literature to quantify the role of placing different land uses around human developments as part of a community-wide fuel treatment and fire preparedness plan. Beyond California or golf courses, other regions and land uses may have anecdotal evidence supporting a fire buffering effect. The methods outlined in this paper could be used to study whether this buffering effect is observed in other regions for golf courses or other land uses.

Along with suppression during a fire, pre-fire management irrigation is likely different between an open-space park and a golf course. The reported benefit golf courses may have over similar non-golf course vegetation during a fire comes at the cost of the additional irrigation, herbicide, energy, and labor used to manage these land uses. These are fixed inputs where the benefit of reduced fire behavior is only realized during a fire. Along with the direct costs associated with these inputs and reduction in wildlife habitat, there are other social and resource costs when directing limited resources such as water to irrigating a golf course in an arid ecosystem. Finally, any of the observed benefits connected to the physical properties of the fuels on the golf courses are a byproduct of the management. If the management changes, the benefits observed may also change. One benefit of economically generating landscapes such as agriculture or golf courses for land use planning is that they are revenue streams to support their continued management outside of fire related public grants.

Communities or planners considering incorporating managed green landscapes as fire buffers might envision it as part of a resilient community design for living with fire. However, concepts like resilience are difficult to operationalize. Choosing to send resources towards maintaining a landscape like a golf course may provide specific resilience to the threat of fire while not being compatible with other sustainability goals. Resilience is not just the inverse of vulnerability and across different scales of resilience, some interventions might have tradeoffs (Chelleri et al. 2015). For example, maintaining golf courses as fire buffers may increase community resilience to fire and, at another timescale, if the resources needed for maintaining the golf course worsen the local effects of climate change or drought, it could contribute to increased wildfire threats. Tradeoffs between resources and resilience across scales are core to land use planning and management (Chelleri et al. 2015). While we were able to estimate the benefit green landscape buffers might provide, some of the tradeoffs or costs associated with their maintenance might be harder to quantify or lack common measures. Thus, a community resilience assessment for incorporating managed landscapes as fire buffers should consider the incommensurable tradeoffs associated with this land use and resource decision(Copeland et al. 2020). Understanding where these green landscape buffers should be maintained will largely be context and region specific, and we hope that our outlined method and software will be one tool for decision makers to use to make choices suitable for their circumstances.

Conclusion

Golf courses can act as a fire buffer around communities, potentially reducing community level risk of wildfire. While anecdotal evidence of golf courses acting as buffers already exists in different regions, using empirical methods we were able to study golf courses in or near fires in California to robustly estimate this buffering effect. We found two main results which indicate golf courses may be effective buffers. First, relative to similar vegetation outside of golf courses, golf course vegetation burns at lower severity. Second, based on the higher frequency with which golf courses appear near fire edges instead of entirely within a fire, we demonstrate that golf courses ap-pear to limit fire spread. Taken together, this evidence suggests that golf courses can be part of a community land use plan to limit wildfire risk. Golf courses in California are just one example of a type of landscape feature which could be incorporated as part of a community buffer zone. We hope that the methods and the software demonstrated in this research will be adapted to study fire buffering effects for other land uses or regions. Ultimately, we hope these tools will help communities plan for fire and weigh the resource tradeoffs associated with the maintenance of those potential buffers.

Acknowledgements

We acknowledge funding Lotusland Investment Holdings, Inc./Bohn Valley, Inc for a financial gift which supported this research. We also acknowledge Scott L. Stephens, the four anonymous reviewers, and Dr. Grant Williamson for feedback on this manuscript.

Tables

Table 1. The variables and their origin used for the PSM and regression analysis. PSM used all the listed variables except for the Burn Severity, which was only used in the regression analysis.

Type	Name	Description	Spatial and Temporal Resolution	Source
Burn Severity	NBR offset	Normalized Burn Ratio Offset using pre-fire dates from 1.5 months pre-fire to 1.5 months post fire	30 m; 16-day	Landsat 5 TM, Landsat 7 ETM+, Landsat 8 OLI/TIRS (U.S. Geological Survey. n.d.)
Geography	Eastness	Aspect-derived measure of 'east' facing, determined by sine function transformation	90 m; DEM from 2000	NASA SRTM (NASA JPL 2020)
	Northness	Aspect-derived measure of 'north' facing, determined by cosine function transformation	90 m; DEM from 2000	NASA SRTM
	Slope	STRM-derived DEM in GEE to calculate slope	90 m; DEM from 2000	NASA SRTM
	Latitude	Latitude of each pixel in degrees determined by GEE function	NA	Google Earth Engine
Vegetation and Vegetation Moisture	NDMI 6	Normalized Difference Moisture Index, NDMI = (NIR – SWIR) / (NIR + SWIR), taken from clear-sky composited image between 3 and 6 months pre-fire alarm date	30 m; 16-day	Landsat 5 TM, Landsat 7 ETM+, Landsat 8 OLI/TIRS
	Precipitation	Total precipitation from 1 January – 31 March (mm)	5566 m; daily	CHIRPS (Funk et al. 2015)
	Precipitation 3	Total precipitation from 1 October – 31 December (mm)	5566 m; daily	CHRIPS
	Landcover	Dominant vegetation determined by satellite and ground-truthed data. The landcover closest to the date prior to fire is used.	30 m; epochs produced for 2001, 2004, 2006, 2008, 2011, 2013, 2016	NLCD Land Cover (Yang et al. 2018)
Suppression Effort	Median Income	Median household income from the five-year 2018, 2013,	Variable; 5-year and 10-year	US Census 5-year American Community Survey

	2009 ACS and 2000 Decennial surveys	and Decennial (Walker and Herman 2021)

Table 2. Predicted values and where the predicted $dNBR_{offset}$ values fall on corresponding MTBS distributions.

Region	Group	n.	Predicted	SE	LCL	UCL	Unburned	Low	Moderate	High
northeast	Cont.	142	91.36	27.88	36.69	146.02	52.17%	31.51%	12.03%	8.00%
northeast	Treat.	142	11.70	27.68	−42.58	65.97	18.45%	10.64%	4.16%	3.05%
northwest	Cont.	278	99.90	23.00	54.81	144.99	63.91%	23.18%	1.722%	0.12%
northwest	Treat.	278	−41.49	23.00	−87.17	4.19	10.52%	0.57%	0%	0%
south	Cont.	950	164.59	15.93	133.36	195.81	54.39%	31.94%	12.42%	12.17%
south	Treat.	950	99.44	15.93	68.22	130.66	44.44%	19.75%	6.14%	7.81%

Table 3. The Wilcoxon test statistic and p-value interpretation for the landcover percentage burned comparisons.

Comparison	Test Statistic	Significant (0.05 cutoff)
Golf Course—Airport	3551	No
Golf Course—Park	4996	Yes

Figures

Figure 1. Map with the locations of the 29 golf courses in California that had a quarter of their area burned in a fire between 1986 and 2020.

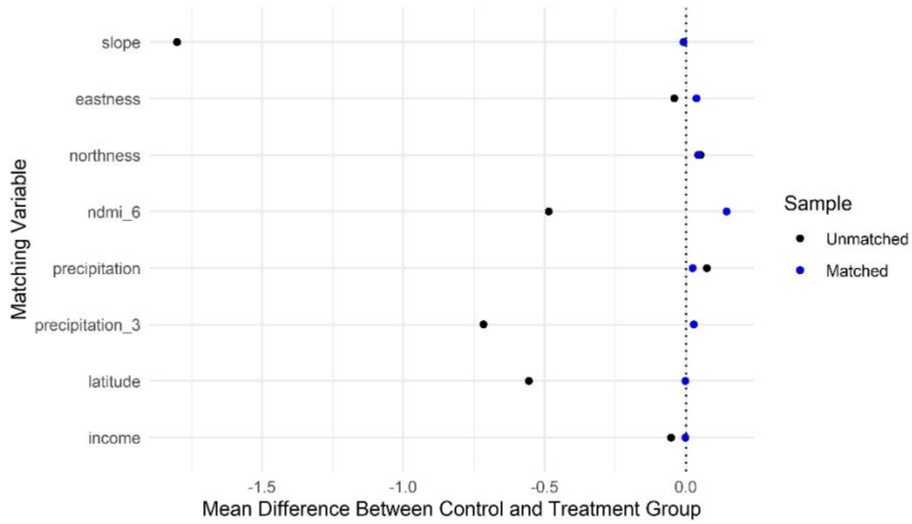


Figure 2. Love plot for the standardized mean differences for the continuous matching variable prior to and after the PSM generated by using the 'cobalt' package in R (Greifer 2021). Exact matches on categorical data such as landcover and fire name have been removed from this plot.

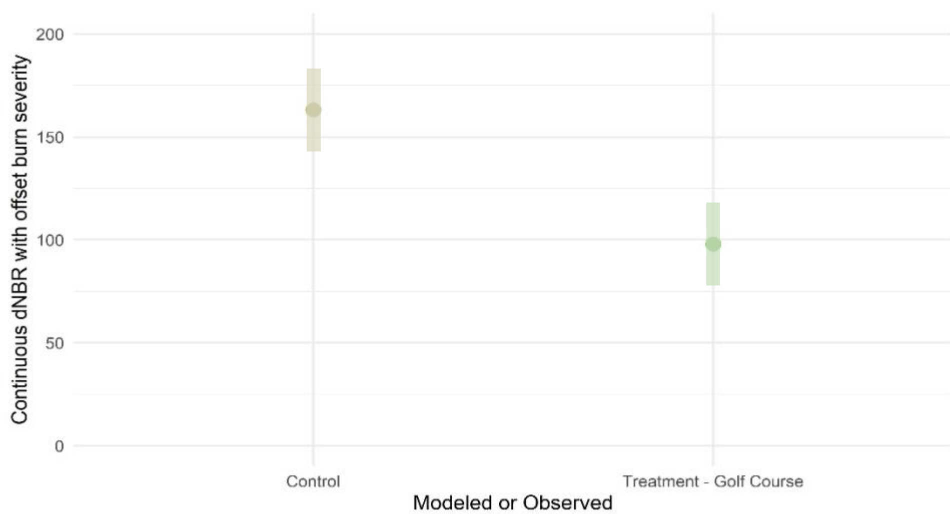


Figure 3. Predicted fire severity values for non-golf course vegetation (control, left) and golf courses.

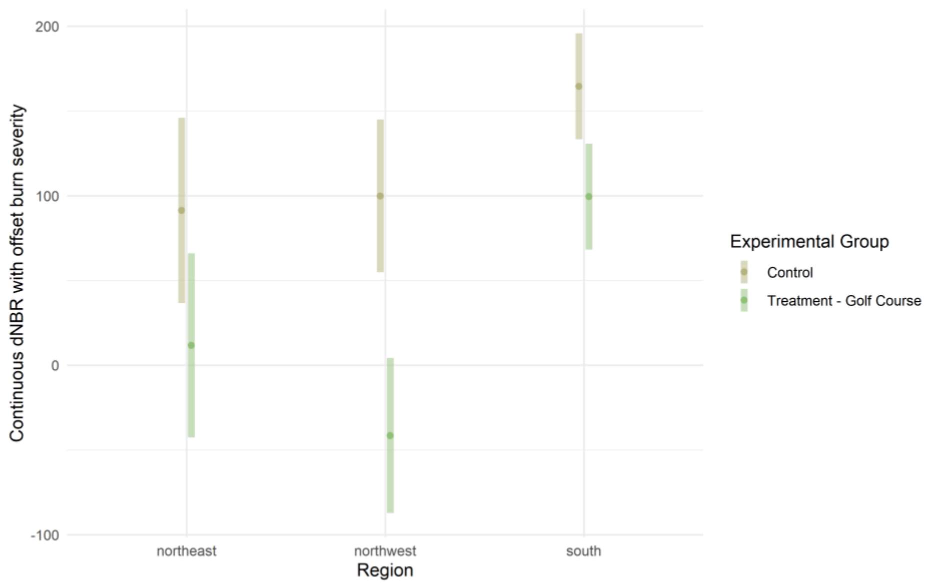


Figure 4. Predicted fire severity values for golf courses, golf course edges, and non-golf course vegetation by region.

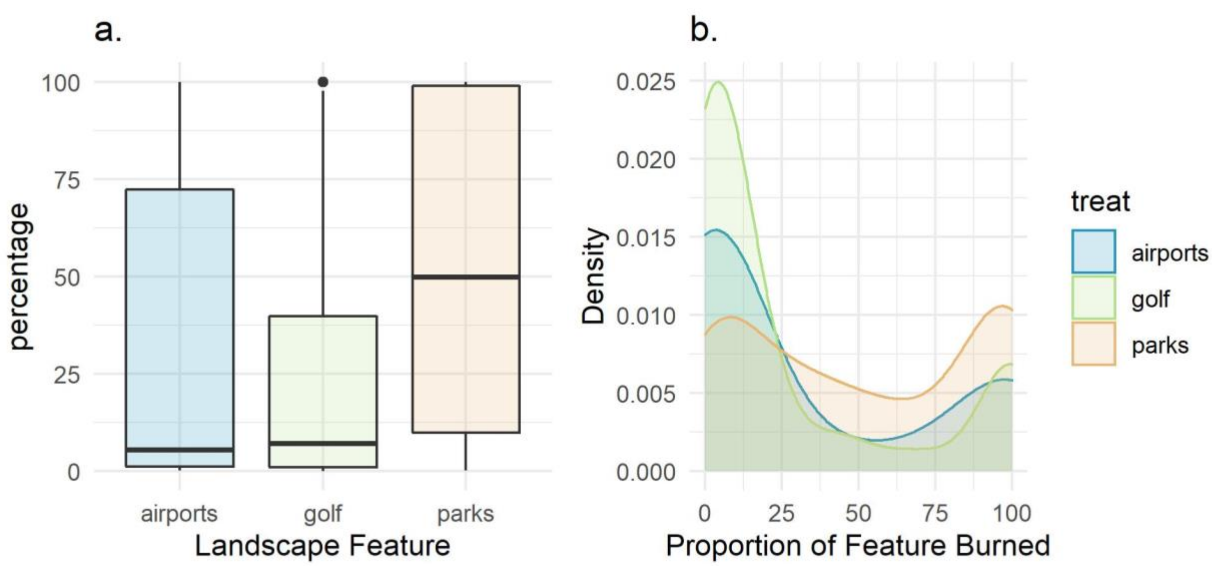


Figure 5. Boxplot (**a**) and smoothed density estimate (**b**) distribution of proportion of land burned of the golf, airport, or park polygons.

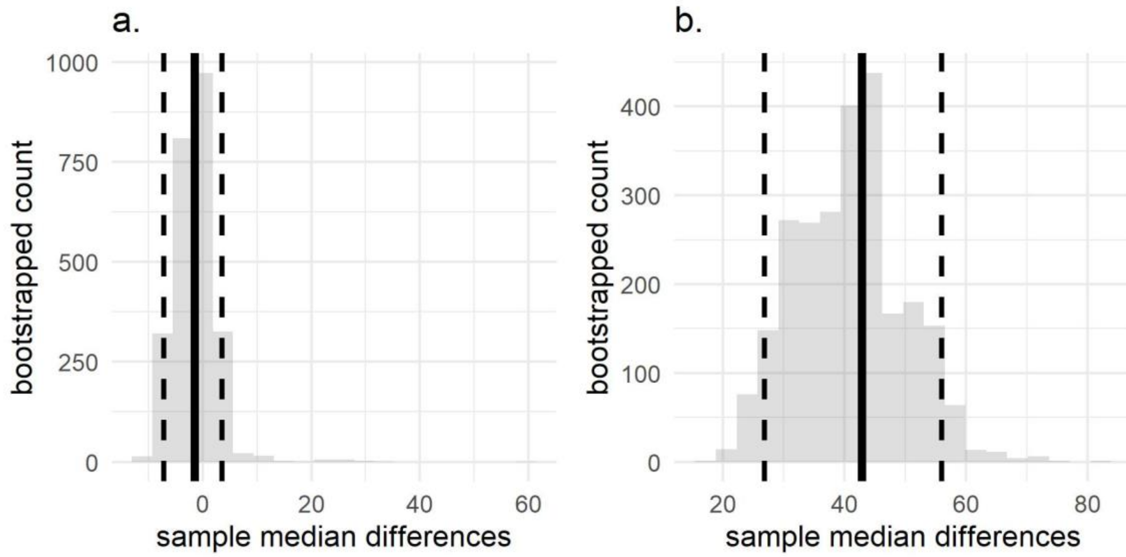


Figure 6. Results of the two-sample bootstrap to compare the median differences between (**a**) golf courses and airports and (**b**) golf courses and parks based on median percentage of land feature burned.

References

"About CPAD." https://www.calands.org/cpad/.

Agee, James K et al. 2000. "The Use of Shaded Fuelbreaks in Landscape Fire Management." Forest Ecology and Management 127(1–3): 55–66. https://linkinghub.elsevier.com/retrieve/pii/S0378112799001164.

"ArcGIS Pro." https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview.

Austin, Peter C. 2011. "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." Multivariate Behavioral Research 46(3): 399–424. http://www.tandfonline.com/doi/abs/10.1080/00273171.2011.568786.

Butry, David T. 2009. "Fighting Fire with Fire: Estimating the Efficacy of Wildfire Mitigation Programs Using Propensity Scores." Environmental and Ecological Statistics 16(2): 291–319. http://link.springer.com/10.1007/s10651-007-0083-3.

Butsic, Van et al. 2017. "Quasi-Experimental Methods Enable Stronger Inferences from Observational Data in Ecology." Basic and Applied Ecology 19: 1–10. http://dx.doi.org/10.1016/j.baae.2017.01.005.

CAL FIRE (California Department of Forestry and Fire Protection). "Cal Fire Fuel Breaks and Use During Fire Suppression 2019." https://www.fire.ca.gov/media/5585/fuel_break_case_studies_03212019.pdf.

———. "Fire Perimeters." https://frap.fire.ca.gov/frap-projects/fire-perimeters/.

Calkin, David E., Jack D. Cohen, Mark A. Finney, and Matthew P. Thompson. 2014. "How Risk Management Can Prevent Future Wildfire Disasters in the Wildland-Urban Interface." Proceedings of the National Academy of Sciences 111(2): 746–51. https://pnas.org/doi/full/10.1073/pnas.1315088111.

Chelleri, Lorenzo, James J Waters, Marta Olazabal, and Guido Minucci. 2015. "Resilience Trade-Offs: Addressing Multiple Scales and Temporal Aspects of Urban Resilience." Environment and Urbanization 27(1): 181–98. http://journals.sagepub.com/doi/10.1177/0956247814550780.

Chirouze, M.; et al. 2021. "Quantifying Insurance Benefits of a Nature-Based Approach to Reducing Risk: Wildfire Risk Reduction Buffers." https://www.marshmclennan.com/content/dam/mmc-web/insights/publications/2021/december/MMC_TNC_Quantifying_Insurance_Benefits.pdf.

Copeland, Samantha et al. 2020. "Measuring Social Resilience: Trade-Offs, Challenges and Opportunities for Indicator Models in Transforming Societies." International Journal of Disaster Risk Reduction 51: 101799. https://linkinghub.elsevier.com/retrieve/pii/S2212420920313017.

"Data Element Standard." Wildfire Coordinating Group (NWCG).

Funk, Chris et al. 2015. "The Climate Hazards Infrared Precipitation with Stations—a New Environmental Record for Monitoring Extremes." Scientific Data 2(1): 150066. http://www.nature.com/articles/sdata201566.

Giglio, Louis, Wilfrid Schroeder, and Christopher O. Justice. 2016. "The Collection 6 MODIS Active Fire Detection Algorithm and Fire Products." Remote Sensing of Environment 178: 31–41. https://linkinghub.elsevier.com/retrieve/pii/S0034425716300827.

Gorelick, Noel et al. 2017. "Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone." Remote Sensing of Environment 202: 18–27. https://linkinghub.elsevier.com/retrieve/pii/S0034425717302900.

Greifer, N. 2021. "Cobalt: Covariate Balance Tables and Plots." https://cran.r-project.org/package=cobalt.

Gross, Patrick. 2008. "Golf Courses on the Fire Line." : 13–16.
Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2011. "MatchIt : Nonparametric Preprocessing for Parametric Causal Inference." Journal of Statistical Software 42(8). http://www.jstatsoft.org/v42/i08/.
Keeley, Jon E. 2009. "Fire Intensity, Fire Severity and Burn Severity: A Brief Review and Suggested Usage." International Journal of Wildland Fire 18(1): 116. http://www.publish.csiro.au/?paper=WF07049.
Keeley, Jon E., and Alexandra D. Syphard. 2019. "Twenty-First Century California, USA, Wildfires: Fuel-Dominated vs. Wind-Dominated Fires." Fire Ecology 15(1): 24. https://fireecology.springeropen.com/articles/10.1186/s42408-019-0041-0.
Kramer, Heather Anu, Miranda H. Mockrin, Patricia M. Alexandre, and Volker C. Radeloff. 2019. "High Wildfire Damage in Interface Communities in California." International Journal of Wildland Fire 28(9): 641. http://www.publish.csiro.au/?paper=WF18108.
Leeper, T.J. 2021. "Margins: Marginal Effects for Model Objects."
Lenth, R.V. "Emmeans: Estimated Marginal Means, Aka Least-Squares Means."
Lentile, Leigh B. et al. 2006. "Remote Sensing Techniques to Assess Active Fire Characteristics and Post-Fire Effects." International Journal of Wildland Fire 15(3): 319. http://www.publish.csiro.au/?paper=WF05097.
McKnight, P.E., and J. Najab. 2010. "Mann-Whitney U Test." The Corsini encyclopedia of psychology.
Moritz, Max A. et al. 2014. "Learning to Coexist with Wildfire." Nature 515(7525): 58–66. http://www.nature.com/articles/nature13946.
Moritz, Max, and Van Butsic. 2020. Building to Coexist with Fire: Community Risk Reduction Measures for New Development in California. University of California, Agriculture and Natural Resources. https://anrcatalog.ucanr.edu/Details.aspx?itemNo=8680.
MTBS Project (USDA Forest Service/U.S. Geological Survey). "MTBS Data Access: Fire Level Geospatial Data." http://mtbs.gov/direct-download (February 2, 2020).
NASA JPL. 2020. "NASADEM Merged DEM Global 1 Arc Second V001 [Data Set]."
National Wildfire Coordinating Group. 2015. "Fire Terminology."
———. "Wildland Fire Suppression Tactics Reference Guide." https://www.coloradofirecamp.com/suppression-tactics/suppression-tactics-guide.pdf.
OpenStreetMap Contributors. 2015. "Planet Dump [Data File from 2021]."
"Overpass Turbo Airport Query." https://overpass-turbo.eu/s/1fM0
"Overpass Turbo Golf Course Query." https://overpass-turbo.eu/s/1fLY.
"Overpass Turbo Park Query." https://overpass-turbo.eu/s/1fLZ.
Parks, Sean et al. 2018. "Mean Composite Fire Severity Metrics Computed with Google Earth Engine Offer Improved Accuracy and Expanded Mapping Potential." Remote Sensing 10(6): 879. http://www.mdpi.com/2072-4292/10/6/879.
Peng, R.D. 2019. "Simpleboot: Simple Bootstrap Routines."
R Core Team. 2021. "R: A Language and Environment for Statistical Computing."
Ramsey, David S. L. et al. 2019. "Using Propensity Scores for Causal Inference in Ecology: Options, Considerations, and a Case Study" ed. Torbjørn Ergon. Methods in Ecology and Evolution 10(3): 320–31. https://onlinelibrary.wiley.com/doi/10.1111/2041-210X.13111.
RStudio Team. 2020. "RStudio: Integrated Development for R. RStudio." http://www.rstudio.com/.

Ruecker, Gernot, David Leimbach, and Joachim Tiemann. 2021. "Estimation of Byram's Fire Intensity and Rate of Spread from Spaceborne Remote Sensing Data in a Savanna Landscape." Fire 4(4): 65. https://www.mdpi.com/2571-6255/4/4/65.

Schoennagel, Tania et al. 2017. "Adapt to More Wildfire in Western North American Forests as Climate Changes." Proceedings of the National Academy of Sciences 114(18): 4582–90. https://pnas.org/doi/full/10.1073/pnas.1617464114.

Syphard, Alexandra D., Jon E. Keeley, and Teresa J. Brennan. 2011. "Factors Affecting Fuel Break Effectiveness in the Control of Large Fires on the Los Padres National Forest, California." International Journal of Wildland Fire 20(6): 764. http://www.publish.csiro.au/?paper=WF10065.

Syphard, Alexandra D., Heather Rustigian-Romsos, and Jon E. Keeley. 2021. "Multiple-Scale Relationships between Vegetation, the Wildland–Urban Interface, and Structure Loss to Wildfire in California." Fire 4(1): 12. https://www.mdpi.com/2571-6255/4/1/12.

"TIGER/Line: Current State and Equivalent National." 2019.

U.S. Geological Survey. "Landsat-5, Landsat-7, and Landsat-8 Imagery."

Walker, K., and M. Herman. 2021. "Tidycensus: Load US Census Boundary and Attribute Data as 'Tidyverse' and 'Sf'-Ready Data Frames." https://cran.r-project.org/package=tidycensus.

Wickham, H. 2016. "Ggplot2: Elegant Graphics for Data Analysis."

Wickham, H., R. François, L. Henry, and K Müller. 2021. "Dplyr: A Grammar of Data Manipulation." https://cran.r-project.org/package=dplyr.

Woo, Hyeyoung, Bianca N. I. Eskelson, and Vicente J. Monleon. 2021. "Matching Methods to Quantify Wildfire Effects on Forest Carbon Mass in the U.S. Pacific Northwest." Ecological Applications 31(3). https://onlinelibrary.wiley.com/doi/10.1002/eap.2283.

Yang, Limin et al. 2018. "A New Generation of the United States National Land Cover Database: Requirements, Research Priorities, Design, and Implementation Strategies." ISPRS Journal of Photogrammetry and Remote Sensing 146: 108–23. https://linkinghub.elsevier.com/retrieve/pii/S092427161830251X.

Transition between Chapters 1 and 2

In Chapter 1, I used observed data and causal inference methods to estimate the effect of green landscape features as fire buffers. For golf courses in California that burned or were near a fire perimeter between 1984 to 2020, I found that there is an observable buffering effect that both reduces fire severity on golf courses and limits fire spread. Estimating a buffering effect for green landscape features may be useful for community planners attempting to reduce the risk of fire to human developments and infrastructure.

Planning for communities or landscapes interventions can use historic observations for causal inference, as was done in Chapter 1, or it can use historic observations to develop models to anticipate how ecosystems will develop in the future. Both methods can be useful to weigh tradeoffs and evaluate competing needs for land use. These models developed to anticipate ecosystem development are process based models. Their usefulness in making decisions depend on the accuracy of how well their outputs mirror the observed future growth. While these models are regularly used to inform management decisions, there is less attention paid to how accurate their predictions are, in a process called model validation. This prompted the research questions I asked in Chapter 2.

In Chapter 2, I completed a model validation study for the Forest Vegetation Simulator (FVS) on anticipated carbon growth in California. I found that the 'out-of-the-box,' uncalibrated model runs of growth in FVS are not like the observed growth rates in California between 2001-2018. Errors produced by FVS that overestimate the rate of carbon accumulation in forests are more pronounced for lower productivity forest sites and National Forest Service forestlands. Overestimates in carbon growth rate can lead to forest management decisions that are based on unrealistic expectations of forest growth without management. Relative to forest carbon management, this may cause an overestimate in how fast forests sequester carbon without management.

Chapter 2. Validation of Forest Vegetation Simulator Model finds Overprediction of Carbon Growth in California

Claudia Herbert, Van Butsic, Jeremy Fried

Abstract

This research tests whether the Forest Vegetation Simulator (FVS) model produces carbon growth predictions in California similar to observed growth in a statewide validation study. Using regression-based, bootstrapped Equivalence Tests, we find that FVS as typically used out-of-the-box overpredicts carbon growth rates by 55%, on average, over the course of a short (ten-year) projection and that predicted growth is not equivalent to observed growth. Predictions of total carbon stock over such short projections of FVS growth may be equivalent to measured values, mainly because short-term growth is typically a small fraction of stocks. The sizeable deviations between predicted and actual growth rates are concerning for short model projections, and over longer periods, inaccurate predictions of growth will render predictions of carbon stock unreliable. To understand patterns in FVS errors that could reveal where the model is most and least reliable, we use linear regression, holding all other factors constant to determine how location and stand attributes such as FVS Variant, site productivity class, stand age, reserve status, and ownership affect prediction error. We find that the magnitude of errors produced by FVS is sensitive to the forest type and variant being modeled, with the largest overpredictions occurring in stands within the North Coast variant, growing on the lowest site class, having ages that are unknown or between 50 and 100 years, that are within governmentally designated reserved areas or on national forests. Direction and magnitude of errors are related to the stand attributes; these relationships may offer fruitful opportunities to improve the underlying growth models or calibrate the system to improve prediction accuracy. Our findings suggest that forest managers relying on out-of-the-box FVS growth models to forecast carbon sequestration implications of their management of California forests will obtain estimates that overstate the carbon that can be sequestered under light-touch or caretaker management, potentially leading to management decisions that fail to deliver the expected carbon sequestration benefits—a failure that could take a long time to recognize.

Keywords: Forest Vegetation Simulator (FVS); Model Validation; California; Equivalence Tests; Carbon

Introduction

In response to the threat of climate change, climate policies often highlight forest ecosystems as both a critical carbon sink and at risk from increased environmental stress accompanying climate change (Jiang et al. 2021; Keen et al. 2022). Policies for financing sustainable forest management across borders have evolved from programs aimed at reducing deforestation in the global south (e.g., Rainforest Alliance, REDD) to programs crediting forests anywhere for the carbon they store and sequester via carbon markets (e.g., REDD+, California Compliance Market Forest Offset Protocol). The scale of capital interested in reducing global emissions, and the potential for relatively cheap sequestration and storage in forests as offsets, have placed a spotlight on forests as Natural Climate Solutions (Griscom et al. 2017). In practice, revenue from carbon offsets can pay forest owners for keeping forests intact (avoided conversion), expanding or preventing loss of forests through re/afforestation, or increasing density of forest stocks through improved management (van der Gaast, Sikkema, and Vohrer 2018). The most common type of carbon offset by number of projects and tons credited in the United States for both the compliance and voluntary markets is this latter approach to offsets known as Improved Forest Management (IFM) (Haya et al. 2021). IFM offsets assign credit to changes in management (e.g., increasing rotation length, reducing harvest intensity, etc.) that increase carbon sequestered and provide an associated climate benefit. Outside of monetizing carbon for forest management, the largest forest owner by area and volume in the United States is the network of public forestlands managed by the Forest Service, which are not currently eligible for carbon credits. Despite the absence of carbon sequestration from the list of objectives guiding the agency, the Forest Service recognizes the importance of maintaining forests as a net carbon sink; over the past decade, the Forest Service has indicated that increasing carbon sequestration is part of their plan to mitigate climate change (Borys Tkacz et al. 2011). Arguably, the policy interest in forests and the need for accurately characterizing carbon dynamics is now greater than it has ever been.

Developing carbon-focused management plans and participating in a market for carbon credits requires consistent and accurate accounting of forest carbon stores and sequestration into woody biomass. For IFM offsets sold in the California Compliance Market, this is more than $1.7 billion worth of forest offsets predicated on the ability of projects to maintain the initial above average inventory (Badgley et al. 2022). Despite progress in developing comprehensive forest carbon stock and flux maps, data necessary for the more granular, project-level decision making needed for carbon markets is often limited in geographic and temporal extent (Harris et al. 2021). The most accurate practical approach to assessing carbon stocks relies on forest inventory measurements of tree height and diameter. Accurate flux measurement depends on either continuous measurement of canopy gas exchange or taking the difference of two or more stock measurements collected at different times, which is more temporally coarse but also more feasible over large areas. Retrospective analyses based on forest inventory assessments of carbon dynamics have only recently become possible with the availability in the U.S. of remeasured forest inventory data from a spatially balanced sample of all forested lands (Christensen et al. 2021). To credit current and future carbon sequestration, offset markets rely on periodic forest inventories and selected, peer-reviewed growth and yield models. For example, an IFM offset project can be inventoried as infrequently as once every twelve years. Over that interval, the project can rely on growth and yield models to estimate carbon stocks and flux to document management tradeoffs against anticipated carbon benefits (CARB Air Resources Board 2011). As forest carbon remains central to climate policy, the accurate performance of these growth and yield models should also be viewed as central to climate policy. Validating the accuracy of growth and yield models is an important step in connecting forest management with an honest characterization of climate benefits to support informed decisions about forests.

The Forest Vegetation Simulator (FVS; Dixon 2013) is one of the models approved by all carbon offset registries that accept Improved Forest Management projects (American Carbon Registry 2018, Climate Action Reserve 2017, Gibbon et al. 2013). FVS has been developed and supported by the USDA Forest Service for over three decades, developed to evaluate treatments to help managers make plans that address various forest objectives. This individual tree, distant-independent (not requiring spatial referencing of trees) empirical growth and yield model is widely used to predict stand dynamics, with and without management, in forests covering all ownerships across the United States (Shaw 2009). The base FVS model predicts growth in tree diameter and height and the probability of mortality over a user-selected period for each tally tree within a sampled stand. The aggregated projections of tree growth and mortality can be used to estimate future carbon stocks and change, sometimes referred to as carbon flux. Several extensions to FVS have been developed to account for changes in growth and carbon dynamics that might be expected under climate change, fire disturbance, and infestations by insects and pathogens (Crookston 2014; Crookston and Dixon 2005; Reinhardt and Crookston 2003). The full FVS system can be used to evaluate, for example, how forest carbon dynamics and other objectives for the forest would respond to different choices for rotation length, stand density management, and harvest operations to support strategies that maximize credit for forest carbon sequestration (Hoover and Rebain 2008).

Forest management strategies aimed at maximizing carbon credits may vary by location and context. This results in IFM carbon offset protocols avoiding prescriptive specifications about what management qualifies as 'improved.' Instead, they offer latitude for demonstrating, via rigorous analysis, both the current trajectory of carbon stocks and how this would change under silviculture designed to enhance climate benefits. Silviculture enhancing climate benefit could include increased retention of previously sequestered carbon or new sequestration such as storage of additional carbon both in the forest and in harvested wood products and solid waste disposal sites. Some researchers who use FVS to strategize which types of management can optimize future carbon have reported that limiting forest management through avoiding harvest, lengthening rotations, or maintaining near-maximum retention will result in greater climate benefits via reduced emissions and increased sequestration (Fischer, Cullen, and Ettl 2017; Kerchner and Keeton 2015). FVS was developed for evaluating different management scenarios, and yet carbon-focused management often suggests non-management, most notably on national forests. The validity of FVS-based analyses depends on both analyst assumptions and FVS generating accurate predictions of forest carbon trajectories. If FVS is increasingly used to simulate growth without management, then our expectations on continued carbon sequestration are based on model predictions.

As resource decisions with long-term consequences for the global climate are increasingly reliant on models like FVS, the need for robust model validation that exposes model bias and error has never been more important. A recent uptick in publications addressing model validation notwithstanding, fewer than one percent of published modeling studies focus explicitly on validation (Eker et al. 2018), perhaps owing to limited access to the consistently measured, longitudinal forest data that is essential to robust evaluation of model performance. The recently published remeasurement data from permanent Forest Inventory Analysis (FIA) plots is ideal for evaluating the validity of FVS growth predictions, as they provide a spatially balanced, statistically representative foundation for assessing FVS performance for multiple variants, forest types, and myriad sizes and species of trees (Crookston and Dixon 2005).

Previous evaluations of FVS reported both under- and overpredictions of tree growth, with sign and magnitude varying by model variant, species, and tree size (Bagdon et al. 2021; Canavan and Ramm 2000; Lacerte et al. 2004; Russell et al. 2015; Smith-Mateja and Ramm 2002). An important motivation for evaluating errors in FVS predictions is to identify conditions where it may be important to recalibrate and otherwise improve the model. In some cases, when deviations between predicted and actual growth exceed what is needed for FVS to be useful, re-engineering of model components may be advisable (Pokharel and Froese 2008). FVS is not one model; there are 22 geographically distinct variants, and each contains multiple component models, parameterizations, and assumptions believed to be appropriate for the species, forest types, and environmental attributes within the variant (Crookston and Dixon 2005). Only a few of these have been formally validated against observed tree growth. Notably, we were unable to find any studies evaluating FVS predictions of tree growth in California. This study appears to be the first formal validation of multiple FVS variants using a consistent, spatially balanced sample of longitudinal observations of forest growth, which presents the opportunity to compare prediction errors across variants, as well as species and stand characteristics.

This paper reports the performance of FVS when used to project forest carbon stocks and flux. This is an important area of research given the increasing use of FVS as a carbon calculator in support of carbon market offset projects. We evaluated the four FVS variants that account for 99% of California's forest land and for which no FVS validation exists. Specifically, we sought to address the following questions:

1. Are 10-year FVS projections of carbon stocks equivalent to field measured carbon stocks?
2. Are FVS predictions of growth in carbon equivalent to observed growth over 10 years?
3. If stocks and growth are not equivalent,
   a. Which types of forest lands generate predictions with the greatest departures from observed values?
   b. Which species and tree sizes lead to predictions with the greatest departures from observed values?

Materials and Methods

Study Area Description

We analyzed data from 3,340 FIA plots downloaded from the public Forest Inventory and Analysis Database (FIADB) repository (version 7.0, California file, part of the Pacific Northwest Work Unit, at https://apps.fs.usda.gov/fia/datamart/datamart.html) representing nearly all forested lands in California (Figure 1). These forests span a broad range of productivity classes and forest types, from redwood to California mixed conifer, and include dry forests dominated by oak and pinyon-juniper, where carbon stocks are typically much lower. Based on the FIA measurements at the initial visit (2001-2009), when these plots were installed under the annual inventory design, most forest carbon can be found on Federal lands, except in the Klamath Mountains (NC) variant, where private forests hold more carbon (Table 1). The largest carbon stocks tend to be on forests in the 50–100-year age class, perhaps because these account for the plurality of forest area, except for in the Western Sierra Nevada (WS) variant, where carbon stocks are greatest in 100–150-year-old forests. Including the entire state in this analysis offered the opportunity to evaluate the performance of multiple FVS variants across diverse forest landscapes.

FIA database description

To model forest growth in FVS, and the drivers of model validity, we relied on data from three tables in the FIA database (The Forest Inventory and Analysis Database: Database Description and User Guide for Phase 2 (version 9.0.1)): Plot (for location attributes such as elevation and applicable FVS variant), Condition (for site attributes like ownership, forest type, site class, and disturbance history), and Tree (species, diameter, and height) (Burrill et al. 2015). Most of this data is assessed in the field by specialized inventory data collection crews; computed attributes (e.g., tree volume and carbon) are calculated as part of the FIA compilation in the months following completion of data collection for a given year.

FIA's sampling intensity in California is a nominal one plot per 2387.6 hectares. Each plot consists of three component plot sizes, the area of which is distributed equally among 4 locations: four 2-meter radius microplots, totaling 1/75th acre, to sample trees 12.7 cm dbh; four 7.3-meter radius subplots, totaling 1/6th acre, to sample trees between 12.7 cm and 60.96 cm; and four 17.9-meter radius macroplots, totaling 1 acre, to sample trees larger than 60.96 cm in diameter. On plots containing more than one ownership class, forest type, reserve status, stand size class, or stand density class, these differences are mapped in the field as distinct "conditions", each with its own set of condition attributes and a "condition proportion" that reflects the mapped area of the condition on the full plot footprint. Plots without such condition "breaks" have a single condition.

We model each FIA condition in FVS as a "stand", and all the trees projected by FVS on that condition, or observed at the remeasurement, are assigned the site context values observed on the condition at the first field visit. Carbon stocks and carbon growth of all trees on a condition are summed, weighted by the inverse of plot size, adjusted for condition proportion, to arrive at carbon stock and flux due to growth for the condition. We refer to FIA conditions as stands, consistent with FVS terminology, for the remainder of this paper. To examine the role of tree size and species on FVS prediction performance, we also performed a tree-level validation of carbon stock and flux.

Under the enhanced annual inventory system (Bechtold and Patterson 2005) that began in 2001, FIA plots in the western US are visited and assessed at ten-year intervals, with ten percent of the plots (one panel) receiving a visit each year. This analysis includes all FIA plots in the annual inventory system that were first assessed in one of the nine panels assessed in 2001-2009 and remeasured approximately ten years later[1].

We sought to test FVS's capacity to predict forest growth in the absence of management and large-impact disturbance because, while those influences can be modeled in FVS and the performance of those model functions is of interest, the specific form of those disruptions to growth (e.g., fire severity, kind of thinning regime) could not be determined from the FIA data with sufficient precision to parametrize in FVS simulations. Stands for which crews implementing the remeasurement visit coded fire disturbance or any kind of human-initiated tree removal or surface disturbance (e.g., prescribed fire, chaining) during the remeasurement period were dropped, as were stands that changed status from forest to non-forest (conversion) or vice versa (reversion) between inventory visits, based on codes in FIADB's TREE_GRM_MIDPT table. These filtering rules removed 709 stands, mostly owing to fire disturbance or management activity. We coded a binary,

[1] remeasurement intervals for some (9 percent of) plots depart from the intended 10 years, typically by not more than 1-2 years, owing to delays presented by logistical challenges such as denied access and wildfire closures.

non-fire disturbance attribute based on whether stands were coded as having had >=25% of the trees or sample area affected by insects, disease, weather damage, or geologic disturbances damage between inventory visits. Stands that have any type of observable treatment during the remeasurement period were also removed to ensure that comparisons with our 'grow-only' FVS model runs would include only stands that did, in fact, grow-only. The ten percent of live trees that, for a variety of reasons (e.g., forking, swelling, an active wasp nest), had diameter measured at other than breast height at one or both visits, as indicated by the "Diameter Check" code in the Tree table, were removed because diameter is a critical input for FVS growth simulations and FVS is not designed to adjust diameters collected at non-standard heights (Figure 2). These trees were removed only after the FVS modeling so that their presence as competitors for resources would be reflected in modeled stand growth. The filtered dataset contains 69,480 trees on 3,024 conditions representing 42 different forest types and located as all or portions of 2,920 plots.

Modeling tree growth with FVS

FIA data for the 2001-2009 panels were prepared for modeling in FVS by loading them into the Bioregional Inventory Originated Simulation Under Management (BioSum) software (Fried et al. 2017), which assigned the correct FVS variant for each plot and generated FVS input files for each variant that could be used to perform a grow-only simulation for the length of the remeasurement period (Figure 1). There were 1,301 stands in the WS variant, 867 stands in the Inland California and Southern Cascades (CA) variant, 603 stands in the Klamath Mountains (NC) variant, and 347 in the South Central Oregon and Northeast California (SO) variant. We did not model the 16 stands in the Central Rockies (CR) variant because that CR variant contains <1% of eligible stands in California, and any conclusions reached based on such a small sample would not be generalizable to the variant, which covers many states. The four variants (WS, CA, NC and SO) were combined with the remeasurement periods (8 to 12 years) to create 18 variant/remeasurement period combinations that were simulated independently with specifications for the appropriate variant and growth period.

To evaluate FVS performance in estimating volume and carbon growth in live trees that remained alive, and volume and carbon stocks at the end of the growth period, we initiated an FVS simulation in the Suppose user interface (version 2.08) with no forest management activities—what we call a 'grow-only' run. We supplied all required and virtually all optional FVS stand and tree-level inputs (e.g., diameter, height, crown ratio, and, via computations with BioSum, FVS compatible site index) (Figure 2). We avoided any customization or adjustments Although some users have developed custom adjustments (e.g., overrides to the maximum allowable stand density index for a species via FVS's SDIMAX keyword) to achieve more realistic projections, we chose not to use these, consistent with our goal of evaluating the performance of FVS 'out-of-the-box.' We also did not provide the previous diameters or heights that FVS can use as a basis for model calibration, because these were not available in our dataset, and are not typically available for most users. The simulations projected tree growth for the length of the remeasurement period, then "cut" all trees so that when "cutlists" were subsequently loaded into BioSum, volume and biomass could be automatically calculated via FIA-curated equations embedded in BioSum. This ensured that carbon estimates for both projected and observed growth in tree diameter and height would be calculated via the same system of equations. The purpose of this validation was to compare the policy-relevant predictions (i.e., of volume and carbon) associated with modeled diameter and height growth, not the impact of volume and biomass allometry choices by the FIA and FVS programs.

Tree-level carbon and volume stocks and growth

To evaluate predictions of carbon stocks, we compared live tree, non-foliar, above-ground carbon, calculated as half of the DRYBIOT attribute found in FIADB's TREE_REGIONAL_BIOMASS table, where DRYBIOT (total above-ground tree biomass, oven-dry weight) was computed using PNW-RMA's standard equation system, for tree diameters and heights projected from time 1 (the installation visit) by FVS to the measured diameters and heights of those trees at time 2 (the FIA remeasurement visit). Volume growth predictions were based on PNW-RMA's VOLTSGRS attribute (gross volume total stem, which includes all bole wood from ground to tip). We expected that projected stocks would not differ greatly from observed stocks for such a short (~10 years) projection period because for stands free of disturbance and treatment, the predicted growth increment would typically be small relative to stocks, and most of the variation in time 2 stocks would track variation in time 1 stocks. Of greater interest is the performance of FVS in predicting carbon accumulation, in other words, the rate at which carbon becomes sequestered in live trees. We calculated this, for volume and carbon, as 'Annual Growth Increment' (equation 1).

$$Annual\ Growth\ Increment\ (Mg/ha/yr) = \frac{FIA\ Remeasurement\ or\ FVS\ Modeled\ (Mg/ha)\ -\ FIA\ Installation\ Measurement\ (Mg/ha)}{Remeasurement\ Period\ (yr)}$$

Regional calculations of biomass for each tree in the PNW RMA unit derive bole (stem) wood biomass from cubic foot volume (VOLTSGRS) and species-specific wood density (Miles and Smith 2009; equation 2); branch and bark biomass are calculated from equations developed for that tree's species or a related one (USDA Forest Service, PNW Research Station, Resource Monitoring and Assessment Program 2010; USDA Forest Service, PNW Research Station, Resource Monitoring and Assessment Program 2014). Carbon and volume are closely related because bole wood, the basis for volume calculation, accounts for most of the carbon in a tree. All non-Système International d'unités (SI) units, like cubic feet, are converted to SI units before analyzing results using the "measurements" package in R (Birk 2019).

$$bole\ wood\ carbon\ (Mg) = \big((cubic\ foot\ volume\ *\ wood\ density) \div\ 2000\big) \div .9071 \quad (2)$$

Stand-level carbon and volume

FVS projected and FIA Time 2 observations of tree volumes and carbon were aggregated to stand (FIA condition) level as weighted sums using the ratio of trees per acre unadjusted (TPA_CURR_UNADJ) and condition proportion (CONDPROP) from Time 1 as the weight (equation 3). TPA_CURR_UNADJ is the inverse of the plot size on which a tree was sampled; for example, a tree between 12.7 cm and 60.96 cm in diameter at Time 1 was sampled on a 1/6-acre subplot, so would have TPA_CURR_UNADJ=6. However, if that tree existed on a condition covering only half of the plot (CONDPROP=0.5), then for the purpose of carbon density calculation on a per-acre basis that tree sample represents 12 trees. Volume and carbon density were then converted and reported on a per hectare basis.

$$Stand\ stock\ or\ growth = \sum_{all\ trees} tree\ stock\ or\ growth \frac{TPA\_CURR\_UNADJ}{CONDPROP} \quad (3)$$

We removed extreme outliers (observations that were more than the 3rd quartile plus three times the interquartile range [IQR]) from Annual Growth Increment and Time 2 from the remainder of the analysis described in the proceeding sections as they do not appear to be linked to any aspect of FVS projection. Extreme outliers for Time 2 predictions were typically in stands with high carbon stocks. The NC Variant and stands over 200 years were over-represented, in terms of outliers, relative to their occurrence in the full dataset— almost a third of the outliers were redwood stands, mostly in the NC Variant region. FVS substantially overpredicted growth in some stands with very low volume, such as those dominated by hardwood saplings, accounting for additional outliers. Excluding these extreme outliers, which helped us better fit regressions to the non-outlier observations, removed less than 3% of observations, for a total of 3,118 trees that were outliers due to either the FIA observation or the FVS prediction.

Evaluation of bias and accuracy

The first validation diagnostics are numerical descriptions of bias and accuracy of our carbon and volume growth and stock values. Bias, or the systematic error, is the average deviation of repeated estimates from the true value (equation 4) (Cawrse et al. 2010). Bias shows the direction of systematic model deviation: in this context, negative signs for bias mean that FVS is overpredicting FIA growth. The RSME is also scale-dependent but will always be positive—it is a measure of accuracy, sensitive to outliers, where higher RMSE values indicate lower model accuracy (equation 5). Bias and RSME are the first steps in understanding FVS model accuracy relative to the FIA observations and, in the case of bias, the direction of model deviation. Bias and RSME are reported for the carbon and volume time 2 stocks, the annual growth increment, and a relative growth measure, defined here as percent of initial.

$$Bias = \sum_{i=1}^{n} (actual_i - predicted_i) / n \quad (4)$$

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(actual\ (FIA)_i - predicted\ (FVS)_i)^2}{n}} \quad (5)$$

Equivalence Tests

Equivalence testing has been demonstrated as a robust method for evaluating model outputs (Cawrse et al. 2010; Robinson, Duursma, and Marshall 2005). This approach reverses the conventional null hypothesis assumption that corresponding sets of modeled and observed data are similar and assumes, instead, that they are distinct, thereby shifting the burden of proof onto the modeled data (Robinson, Duursma, and Marshall 2005). To test this hypothesis of dissimilarity, a regression-based, two one-sided equivalence tests (TOST) compare the observed data with the mean-shifted FVS modeled data. This TOST is repeated twice: once to compare intercepts and a second time to compare the slope with a 1:1 slope. The intercept TOST tests for agreement between the mean observed and predicted value. The slope TOST tests for agreement between the plotted modeled and observed data and a 1:1 slope. If the modeled versus observed data has a slope near 1, this shows high model accuracy across the range of observations and suggests validation of model structure—or that the model is accurate for the 'right' reasons (Pokharel and Froese 2008; Andrew P. Robinson, Duursma, and Marshall 2005).

H_0: The observations are dissimilar; FVS is not predicting FIA observed values within the defined Smallest Effect Size of Interest (SESOI) (e.g., +/- 10%) of FIA observations for slope or intercept
H_1: The observations are similar, FVS is predicting FIA observed values within the SESOI to FIA observations

The TOST evaluates equivalence by treating model underprediction or overprediction as two independent tests: model predictions failing on just one side of the test means the model is not equivalent. The Smallest Effect Size of Interest (SESOI) sets the "cutoff" level of equivalence expressed as a percentage deviation from the mean observed value. This SESOI value is selected before conducting a TOST and can be interpreted as the maximum acceptable difference between observed and predicted values before the model would be considered not capable of offering reliable predictions for the purposes for which it is used. We conducted three TOSTs, with equivalence thresholds of 5%, 10%, and 25% that are consistent with previous validations of FVS predictions of tree growth (Bagdon et al. 2021; Pokharel and Froese 2008; Andrew P. Robinson, Duursma, and Marshall 2005). We perform a non-parametric bootstrap with 1000 replications to develop confidence intervals around our observation means to see if the predictions fall within our region of equivalence using the "equivalence" package in R Studio using an alpha level of 0.0125 (R Core Team 2021; A. P. Robinson 2016; RStudio Team 2020). If our SESOI-constructed Equivalence Levels do not fall within the bootstrap-constructed Confidence Levels, then the null hypothesis of dissimilarity is supported. A TOST rejection of the null hypothesis of dissimilarity for slope or intercept at either equivalence level can be interpreted as an equivalence between FVS predictions and FIA observations at that equivalence level. The null hypothesis of dissimilarity should be rejected for both slopes and intercept for the model to be considered equivalent.

Using Stand-Level Models to diagnose systemic FVS prediction errors
Because land management style and vegetation type may account for patterns of over-or underprediction, investigation of these patterns could highlight opportunities for future model refinement. We identified five condition-level variables that might drive FVS errors: (i) FVS variant, (ii) site class, (iii) stand age, (iv) reserve status, and (v) land ownership. We transformed stand age as a categorical variable by classification into 50-year bins.

To evaluate the influence of these five variables on model error, we fit a multiple linear regression model. This regression predicts carbon growth increment using potentially explanatory stand-level variables and interaction terms for whether the data point is observed or modeled. Adding interaction terms allow us to fit a linear model with different effects for the FVS and FIA values. This helps tease out the differences for how different values within categories, like stand age or site class, influence FVS errors in predicting carbon growth increment. When the FVS value exceeds the FIA value, overprediction is indicated; underprediction is indicated when the reverse is true.

The dataset for regression analysis contains two entries for each stand: one with FIA observed growth as the dependent variable and one with FVS-predicted growth, with all explanatory variables identical and a binary, 'run' variable indicating whether the record's dependent variable is FIA observed or FVS-predicted. The linear model controls for differences in explanatory variables such as stand density, site conditions, and initial stocks of forest carbon to predict the FVS or FIA growth. Along with the 'run' variable, FIA Time 1 value, and the five variables used as interaction terms, several other variables, drawn from both the FIA database and FVS computed values from the initial visit, were included: Structure Class, Aspect, Slope, Basal Area, Stand Density Index, Number of Strata, Total Cover, Quadratic Mean Diameter (QMD), Stratum 1 DBH, Forest Type, and Other Disturbance, a variable noting if there was a non-fire disturbance impacting over 25% of the trees or sample area during the remeasurement interval.

Using a Stepwise algorithm based on AIC from the "stats" base package in R, we dropped variables identified as reducing model fit if they were not also a component of an interaction term variable of interest (R Core Team 2021). Because the initial set of variables included some that may be largely redundant (e.g., highly correlated metrics of density which are often collinear), we also used variance inflation factors (VIF) to identify and further remove density and geophysical variables with VIF > 10. The final list of variables included the Run variable, FIA Time 1, the five interaction terms (e.g., Run by site class, variant, stands age, reserve status, and land ownership), Structure Class, Aspect, Slope, Total Cover, QMD, Stratum 1 DBH, and Forest Type. Site class was one variable that regularly emerged as having a high VIF, suggesting unsurprisingly high collinearity with other variables. For example, productivity is not independent of ownership and stocking. We did not drop site class as a predictor variable, despite such collinearity, because we believed it important to explore patterns of FVS error across an explicitly defined productivity gradient.

The regression is implemented using the 'lm' function in the R, also in the "stats" package, and model fit diagnostics are done using the R package "olsrr" (Hebbali 2020; R Core Team 2021). Data was organized for analysis using the "tidyverse" package and plots were made using "ggplot2" packages in R (H. Wickham 2016; Hadley Wickham et al. 2019). Because the interaction terms and the number of covariates make it difficult to interpret the coefficients in the linear model output, we use a prediction function to graph the mean predicted values for each category of interest. The output linear model is then used to predict and compare growth values for FVS versus FIA using the 'emmip' function in the "emmeans" package in R (Lenth 2022). The final linear model was significant at a .01 cutoff with an adjusted R-squared of 0.67.

Tree-Level Evaluation: how does error vary by tree diameter and species?
To further evaluate the trends of where bias was observed at the stand level, we used the tree-level measurements to test whether some species or diameter ranges might tend to have greater errors when predicting growth. This information could help prioritize opportunities to improve model predictions. For example, the growth model parameters for a particular species might need updating.

We expected that tree diameter class and species might be important predictors of FVS performance because FVS relies on tree size relative to the rest of the tress in a stand to calculate growth and allocate mortality, and because for some less common species, growth equations are "borrowed" from other species, potentially opening opportunities less accurate growth predictions for certain species (Crookston and Dixon 2005). We created five diameter classes to subset the trees: (i) less than 12.7 cm, (ii) 12.7 to 25.4 cm, (iii) 25.4 to 53.34 cm, (iv) 53.34 to 76.2 cm, and (v) greater than 76.2 cm. To analyze the role of species in tree-level errors, we selected the top six species by both sample tree frequency—*Pseudotsuga menziesii* (Douglas-fir, species code 202), *Abies concolor* (white fir, 15), *Pinus ponderosa* (ponderosa pine, 122), *Notholithocarpus densiflorus* (tanoak, 631), *Quercus chrysolepis* (canyon live oak, 805), *Calocedrus decurrens* (incense-cedar, 81), and by tree volume, which added *Sequoia sempervirens* (redwood, 211), *Abies magnifica* (California red fir, 20), and *Pinus jeffreyi* (Jeffery pine, 116) for a total of nine species. These subsets (for tree size and species) formed the basis of additional equivalence tests, replicating the analytic approach already described for stand level analysis. Subsets where the null hypothesis of dissimilarity cannot be rejected provide an indication of species and sizes for which out-of-the-box FVS simulation of carbon growth may not be valid.

Results

Bias and RMSE

Considering all forests in California, FVS predictions of carbon flux in live trees that remained live ~10 years later were 0.05 Mg/ha/year greater than observed (Table 2). In a state with 12.9 million hectares of forest, this implies a 6.45 Tg overestimation of carbon flux into live trees over the ten-year remeasurement period. The negative signs for bias for both volume and carbon measurements suggest that FVS is overpredicting relative to FIA observations. Because carbon and volume are recorded in different units and measure different entities (i.e., above ground live wood and bark versus only bole wood), we calculated relative growth to determine if Bias and RSME are similar between volume and carbon. The rest of this paper presents the findings only for carbon because the relative growth metric showed volume was within 1% and 4% of the relative growth bias and RMSE of carbon, respectively.

Equivalence Tests

The TOST intercept results show that the FVS predicted growth increments are not within 10% of the FIA observations at the mean and the slope result indicates a lack of agreement across stand-level predictions with measured values (Table 3) (Figure 3). FVS annual growth predictions are equivalent at the 25% level for mean values (intercept test) but not across the range of stand-level predictions (slope test). The predicted carbon stocks pass equivalence tests at the 5%, 10%, and 25% equivalence levels, so FVS projections of carbon stocks, one decade into the future, appear to be valid at the testing thresholds we assumed. For the growth TOSTs that fail to reject the null hypothesis of dissimilarity, we find the proportion of bootstrapped Confidence Levels falling below the Equivalence Level. The larger proportion of the bootstrapped Confidence Level being below the Equivalence Level is consistent with the negative bias results discussed previously, and it confirms that FVS overpredicts growth rate and stocks relative to observations recorded on FIA plots. As a robustness check, to investigate whether low productivity (<20 ft^3/ac/year growth) forests may account for at least some of the underprediction, we repeated the equivalence tests after removing these stands (those with site class=7). After this change, the Annual Growth Increment passed the slope test, but only at the 25% equivalence level.

Stand-Level Models

Consistent with the FVS overprediction established by the Bias and Equivalence Tests, the linear model predicts a mean FVS value, 0.259 that is 55% greater than the mean FIA value, 0.162, in bone dry Mg of carbon sequestered per hectare per year (Figure 4). For the explanatory variables tested with interaction, linear model-output prediction enables assessment of the FVS effect across those variables. Comparing the interaction term levels across the FVS variant, site productivity, stand age, reserve status and ownership, we find instances of FVS both over-and under-predicting growth. The differences between the mean FVS and FIA predicted increments on these interaction plots are the differences between FVS outputs and FIA observations, holding all other variables constant.

Variant: FVS underpredicts growth for the Southern Cascades (CA) and South Central Oregon and Northeast California (SO) variants by 10% of the FIA predicted values, and overpredicts by 22% of FIA values for the Western Sierra Nevada (WS) and 44% for the Klamath Mountains (NC) variants.

Most of California's forest carbon stocks are within the NC and WS variants (Table 1); the modest degree of underpredictions in CA and SO relative to the over predictions, particularly the NC variant, means that FVS is not accurately reflecting carbon accumulation rates for the regions with greater forest carbon stocks.

Site productivity: For Site Class values of 2 through 7, FVS overpredicts. FVS error increases as site productivity decreases, (and site class, an ordinal variable that increases as annual volume productivity decreases, increases), with the error greatest for site class 7 where the difference is statistically significant and FVS overpredicts by 60% of the predicted FIA value. For the most productive lands (site class = 1), FVS understates stocks and growth, though their overlapping standard error bars suggest this difference is not statistically significant. The progressive decrease in FVS prediction accuracy as site productivity decreases suggests an opportunity to build in a calibration or correction for FVS predictions.

Stand age: FVS overpredicts growth for all age cohorts. The under 50 years cohort overpredicts less than older cohorts, with less than a 4% overprediction of FVS relative to the predicted FIA value. FVS overestimation is present in all the older cohorts, with the overprediction greatest for the 50-100 and 100-150 cohorts. These older cohorts may be dominated by stands that have experienced little or no active management. FVS was developed to support decisions about management by projecting the results of alternative management activities that are more common in young growth stands. Older stands under "caretaker" 'grow-only' management may be less well represented in the data used to fit the FVS model. The stands that did not have an age recorded (categorized as age cohort 'NA') exhibit FVS over-prediction like the 50 to 100 and 100-150 age cohorts.

Reserve Status: FVS overstates growth for both unreserved and reserved forests, though overprediction is greater for reserved forests, at 15%, than for unreserved forests, where it is 9%. Our analysis holds all other factors like site class and stand age constant for these predictions, and we are only evaluating predictions for stands not managed over the growth interval; however, if less management drives greater discrepancy, it is certainly the case that there is less management on reserved forests.

Ownership: FVS overpredicts across ownership groups; on public forests, the overprediction is 18% above predicted FIA values. This overprediction is less for the 'other public' category, 15%, and overpredictions on private lands are comparatively modest, at 4%.

Tree-Level Models

Tree-level equivalence tests for all trees produced results consistent with what we observed at the stand level. FVS predictions of carbon stocks in stands projected forward ~10 years are equivalent to FIA observations of those stocks within 10% equivalence levels at the means (intercept) and across the range of observations (slope) (Table 4). Annual Growth predictions are equivalent to FIA values at 25% equivalence levels for the mean predictions, but not for the range of predictions agreement tested by the slope. For all the tree and diameter subsets, except for *Sequoia sempervirens* (211, SESE3) annual growth which fails the null hypothesis below the mean, the tests failed to reject the null hypothesis of dissimilarity and had the bootstraps falling below the mean. The bootstraps failing the null hypothesis below the mean suggests FVS over-prediction and is consistent with the stand level tests.

Exploring the diameter and species subsets, the species subsets are better at producing FIA observed growth than diameter class subsets. Douglas fir, ponderosa pine, and white fir in FVS are predicting FIA carbon stocks and growth at the means better than the all-species-combined tree dataset and better than any of the diameter class subsets. Within the diameter subsets, the smallest diameter class, representing trees less than 12.7 cm in diameter, is producing FVS stock and growth predictions that are the most dissimilar to the FIA observations.

Discussion

As society increasingly relies on forests to mitigate climate change, notwithstanding their own vulnerability to a changing climate, the need for accurate modeling of changes in forest carbon stocks with time and prospective management becomes ever-more pressing. Accuracy improvements begin with rigorous validation of existing models against quality-assured, longitudinal sequences of forest measurements. Ideally, such data consists of a field-collected sample that represents all forest conditions for which modeling projections are needed. We initiated such a validation effort for FVS covering nearly all the forests in California, a state with outsized forest policy ambitions, and one with a very large and spatially balanced, longitudinal sample of the forests that grow there.

We included only stands that did not receive management during the decade remeasurement period, dubbed grow-only, because we wanted to have many observations of FIA data that we could assess FVS against. Moreover, validating FVS predictions for management activities requires knowing exactly when and what kinds of management occured, data which is not readily available. While our validation does not address the accuracy of FVS projections involving management activity, validation of the base FVS growth model applies to management scenarios involving any growth projection. We find that, at least in California, uncalibrated FVS simulations do not accurately predict growth of carbon or wood volume within a range of error tolerance likely to be considered acceptable for characterizing carbon sequestration. The average growth predicted by FVS was only equivalent to the FIA measured value at the 25% equivalence level. Even at this level, the slope test failed to reject the hypothesis of dissimilarity, suggesting that FVS is not predicting growth equivalently to observed growth across the range of observed values. While it is hoped that model calibration might allow FVS to predict sufficiently precise growth increment for carbon project evaluation compared to using FVS out-of-the-box, the potential to apply such calibration is contingent on the available inventory data (e.g., either prior measurements or radial growth cores), which are not universally available.

FVS predictions of carbon stocks ten years hence, starting from current inventory data, were within 10% of FIA observations. We found that carbon stock predictions were equivalent to observations both at the mean and across stands with different beginning carbon stocks. Using a combination of current inventory and FVS projections may be a valid approach for short term predictions of carbon *stocks* (up to 10 years). For practical purposes, this suggests that FVS can appropriately be used, for example, to estimate initial carbon stocks for forests in California, consistent with the California Air Resources Board's (CARB) protocol for forest carbon offsets. Because forest inventory data may predate project initiation by a few years, forest growth models may be used to update inventory data-based carbon stocks data to the project start date. Our findings suggest that for such short periods, stock prediction errors may be within acceptable bounds.

While a 10% equivalence for short-term predictions of average carbon stocks might seem promising for FVS, the lack of equivalence for growth predictions will compromise the reliability of multi-decadal forest growth projections. Ten-year stocks are driven much more by initial stocks than by growth; the same is not true for stocks 50 years hence. This is a troubling result given that FVS is used to model growth in carbon stocks under the California Air Resources Board's (CARB) protocols, which require that management impacts on carbon sequestration be accounted over ten decades. Given the errors introduced by FVS growth prediction over even a modest ten-year projection, it may be technically infeasible to obtain valid characterizations of management tradeoffs at century timescales. Even assigning carbon credits for activities tracked and predicted over shorter periods may be too inaccurate for estimating carbon growth since the last inventory.

Our work illustrates how using multiple methods for validation can produce a better understanding of the nature of model performance. While an essential performance metric, bias is not the entire story. The linear regression-based equivalence tests offer two paths for understanding model performance: the intercept and slope tests. From these tests, we find FVS is better at predicting FIA at the means than it is for predicting accurately across the range of individual stands. This means FVS predictions may be performing better at understanding carbon for an average growing stand, but across the range of stand growth rates, it does not produce results that reflect FIA measurements.

The findings of the tree-level equivalence tests are both consistent with and depart from validation efforts undertaken for other FVS variants. In the Lake States variant, errors were also greatest for the smallest tree diameter classes (Lacerte et al. 2004). These errors predicting the growth of small-diameter trees should raise red flags for applying FVS where small trees are a substantial component of the forest (e.g., actively managed young growth). For species-specific prediction errors, based on the results of the parametric bootstrap in the equivalence tests, the growth of only one species, Sequoia sempervirens, was underpredicted by FVS. By contrast, Bagdon et al. (2021) found more variability, with some species subsets accurately predicted by FVS and others under- or overpredicted.

Based on our findings, reliable growth projections will almost certainly require calibration of the FVS model. FVS supports relatively automated calibration if there is previous diameter and height measurements or growth information in the data to be projected. However, users often have only a temporary plot to work with, making this calibration difficult unless radial growth cores have also been collected. Even with FIA data, there may be too few trees of a species and size class on the plot for calibration to be a viable option. That said, there is little doubt that FVS out-of-the-box will not generate reliable predictions of growth. We caution users against implementing FVS in this way and alert them to the need to calibrate. Our findings underscore the need for investment in calibration tools to support more reliable predictions.

FVS was designed to support management decisions and was largely fit using data from forests that were actively managed and not in late successional status. Our findings are consistent with the intended use of FVS, with results most accurate for stands aged less than 50 years and overprediction being greater for less intensively managed reserved forests. However, given that more than half of forests in California receive no management, and that many if not most carbon projects are linked to benefits assumed to accrue to drastically lengthened rotations and/or withdrawal from management, the dearth of forest growth models capable of accurately predicting prospective carbon growth rates is alarming. Inaccurate growth predictions may lead to management regimes, selected for their climate or carbon benefit, that fail to deliver anticipated benefits.

These errors in FVS predictions especially impact the Forest Service lands in California. Public land managers aim to deliver on multiple objectives, including promoting accelerated development of old growth forests while maintaining high rates of carbon sequestration. However, our results suggest that FVS overstates growth rates in stands that are older and/or experience less intensive management. Forest plans relying on the base FVS model without stand-specific calibration may misrepresent potential forest growth and associated benefits. Claims of increased carbon sequestration in California forests, based on out-of-the-box FVS simulations that don't feature active management, merit particularly tough scrutiny.

These results do not directly measure the potential for over-crediting of forest carbon offsets. However, our work does apply to landowners and advisors considering deployment of FVS to evaluate the potential of forest management options to add forest carbon as a revenue stream. Using a grow-only prescription as a baseline against which to compare alternative management scenarios in FVS will, on average, lead to an overprediction of baseline carbon stocks and growth. Significant overpredictions are consistent with work claiming FVS introduced over 70% of the modeling uncertainty when modeling increasing rotation length for Improved Forest Management offsets over 100 years (Fischer et al. 2016). Reliance on FVS as a basis for evaluating carbon management tradeoffs and expected payouts may lead to disappointment among landowners, carbon credit investors, and enterprises and institutions counting on forests carbon credits to meet climate goals. Notably, we found FVS accuracy is greater on private lands, which are more actively managed, than on public lands, which are less so. This is fortunate, in terms of impacts of FVS accuracy on California's carbon market, because only on private lands are currently eligible for participation, not the federal lands on which overstatement of carbon sequestration in the absence of management is most egregious.

Forest carbon protocols and management plans will continue to include options ranging from caretaker management to lengthened rotations as options for landowners to consider, while weighing putative climate mitigation benefits against other forest objectives. Overstated growth rates may enhance the appeal of lengthened rotations and caretaker management for those who seek to manage for carbon benefits—but such benefits may be illusory. Federal forest land contains the overwhelming majority of California's carbon stocks, much of it in stands with high carbon density. The lack of accurate carbon modeling capability for these forests poses a significant challenge for planning climate mitigation responses for these forests. The solution to this problem almost certainly involves better and more transparent processes for carrying out calibration. Ultimately, this will likely include updating the out-of-the-box model to reflect what can be learned through calibration. Until then, calibration will be essential for any growth projections intended to inform as to the outcome of management alternatives; even then, analysts are advised to consider what is now known as to how specific conditions in the forest being modeled may contribute to FVS growth prediction errors.

Acknowledgements
We gratefully acknowledge the US Forest Service for providing funding to support this research (Agreement 21-JV-11261979-047). We also acknowledge the helpful feedback from William Stewart and Scott L. Stephens on versions of this manuscript.

Tables

Table 1. Live tree carbon stocks observed at the remeasurement visit and flux calculated as stock change between the installation and remeasurement visits, for trees observed to be live at both visits and growing on land classified as forest at both visits, by stand age class and variant, and proportions of carbon stocks and flux by owner group.

Summary Carbon Stock and Observed Growth by Variant and Stand Age

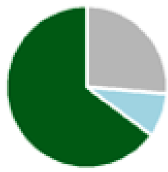
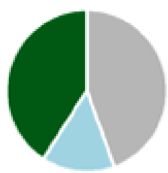
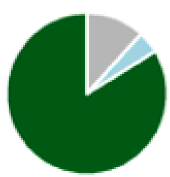
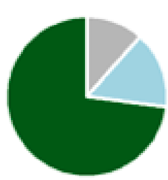
FVS Variant	*Ownership Proportion*	*Stand Age*	*n*	*'Grow Only' Aboveground Live Carbon (Mg)*
CA		<= 50	88	825,569
		50-100	315	6,265,330
		100-150	171	4,751,099
		150-200	52	1,541,091
		>200	58	2,485,565
		Unknown	185	1,551,650
NC		<= 50	154	3,573,573
		50-100	202	6,807,758
		100-150	88	4,291,585
		150-200	38	2,568,082
		>200	38	2,429,303
		Unknown	85	1385899.71
SO		<= 50	29	132,856
		50-100	187	1,886,297
		100-150	90	1,738,521
		150-200	18	337,833
		>200	23	566,941
		Unknown	4	7,547
WS		<= 50	108	850,302
		50-100	327	7,701,272
		100-150	275	8,853,684
		150-200	160	4,932,200
		>200	201	6,252,942
		Unknown	230	1,603,264

Table 2. Bias and RSME results for stand-level above-ground, non-foliar, tree carbon (wood and bark) and bole wood volume, comparing Time 2 (stocks predicted ~10 years forward), Relative Growth (with respect to initial stocks), and Annual Growth Increment

Bias statistics

Measure	*Metric*	*Time 2*	*Relative Growth (%)*	*Annual Growth Increment*
Bias	Carbon	-0.50 Mg/ha	-27.20	-0.05 Mg/ha
	Volume	-1.58 m^3/ha	-27.73	-5.63 m3/ha
RMSE	Carbon	1.49 Mg/ha	241.76	0.15 Mg/ha
	Volume	5.42 m^3/ha	245.32	19.25 m3/ha

Table 3. Equivalence test outcomes for intercept and slope for Time 2 (stand-level above-ground, non-foliar, carbon stocks) predicted ~10 years forward and Annual Growth at 5, 10, and 25% SESOI

Result of Carbon Intercept Test

SESOI	*Metric*	*Mean Observed*	*Mean Predicted*	*Hypothesis*	*CI Lower*	*CI Upper*	*EI Lower*	*EI Upper*	*Bootstrap Below*	*Bootstrap Within*	*Bootstrap Above*
5%	Time 2	12.26	12.76	Reject	12.2	12.31	12.12	13.4	0%	100%	0%
	Annual Growth Increment	0.21	0.26	Not Reject	0.2	0.21	0.25	0.27	100%	0%	0%
10%	Time 2	12.26	12.76	Reject	12.2	12.31	11.48	14.04	0%	100%	0%
	Annual Growth Increment	0.21	0.26	Not Reject	0.2	0.21	0.23	0.28	100%	0%	0%
25%	Time 2	12.26	12.76	Reject	12.2	12.31	9.57	15.95	0%	100%	0%
	Annual Growth Increment	0.21	0.26	Reject	0.2	0.21	0.19	0.32	0%	100%	0%

Result of Carbon Slope Test

SESOI	*Metric*	*Hypothesis*	*SESOI*	*CI Lower*	*CI Upper*	*EI Lower*	*EI Upper*	*Bootstrap Below*	*Bootstrap Within*	*Bootstrap Above*
5%	Time 2	Reject	5%	0.98	0.99	0.95	1.05	0%	100%	0%
	Annual Growth Increment	Not Reject	5%	0.71	0.77	0.95	1.05	100%	0%	0%
10%	Time 2	Reject	10%	0.98	0.99	0.9	1.1	0%	100%	0%
	Annual Growth Increment	Not Reject	10%	0.71	0.77	0.9	1.1	100%	0%	0%
25%	Time 2	Reject	25%	0.98	0.99	0.75	1.25	0%	100%	0%
	Annual Growth Increment	Not Reject	25%	0.71	0.77	0.75	1.25	76%	24%	0%

Table 4. Tree level equivalence test results for above-ground, non-foliar, tree carbon stocks predicted ~10 years forward and annual carbon growth for subsets of diameter groups and species. Blackened boxes indicate tests that failed to reject the null hypothesis of dissimilarity. The black box with the white asterisk indicates the test where the bootstrapped sample is above the equivalence region.

Intercept test	Full	DBH <12.7	DBH 12.7-25.4	DBH 25.4-53.34	DBH 53.34-76.2	DBH >76.2	202: PSME	122: PIPO	15: ABCO	631: LIDE3	805: QUCH2	81: CADE27	211: SESE3	20: ABMA	116: PIJE
Number of Observations	69,480	5,700	29,510	16,699	9,512	8,059	10,048	4,957	9,058	4,606	4,802	3,778	2,195	3,179	3,292
Result of Carbon test 10%															
Time 2		■													
Annual Growth	■	■	■	■	■	■				■	■	■	■ *	■	■
Result of Carbon test 25%															
Time 2															
Annual Growth		■	■							■		■	■		

Slope Test	Full	DBH <12.7	DBH 12.7-25.4	DBH 25.4-53.34	DBH 53.34-76.2	DBH >76.2	202: PSME	122: PIPO	15: ABCO	631: LIDE3	805: QUCH2	81: CADE27	211: SESE3	20: ABMA	116: PIJE
Result of Carbon test 10%															
Time 2		■	■										■		
Annual Growth	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Result of Carbon test 25%															
Time 2		■													
Annual Growth	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

□ Rejects Null Hypothesis of Dissimilarity

■ Does Not Reject Null Hypothesis of Dissimilarity

Figures

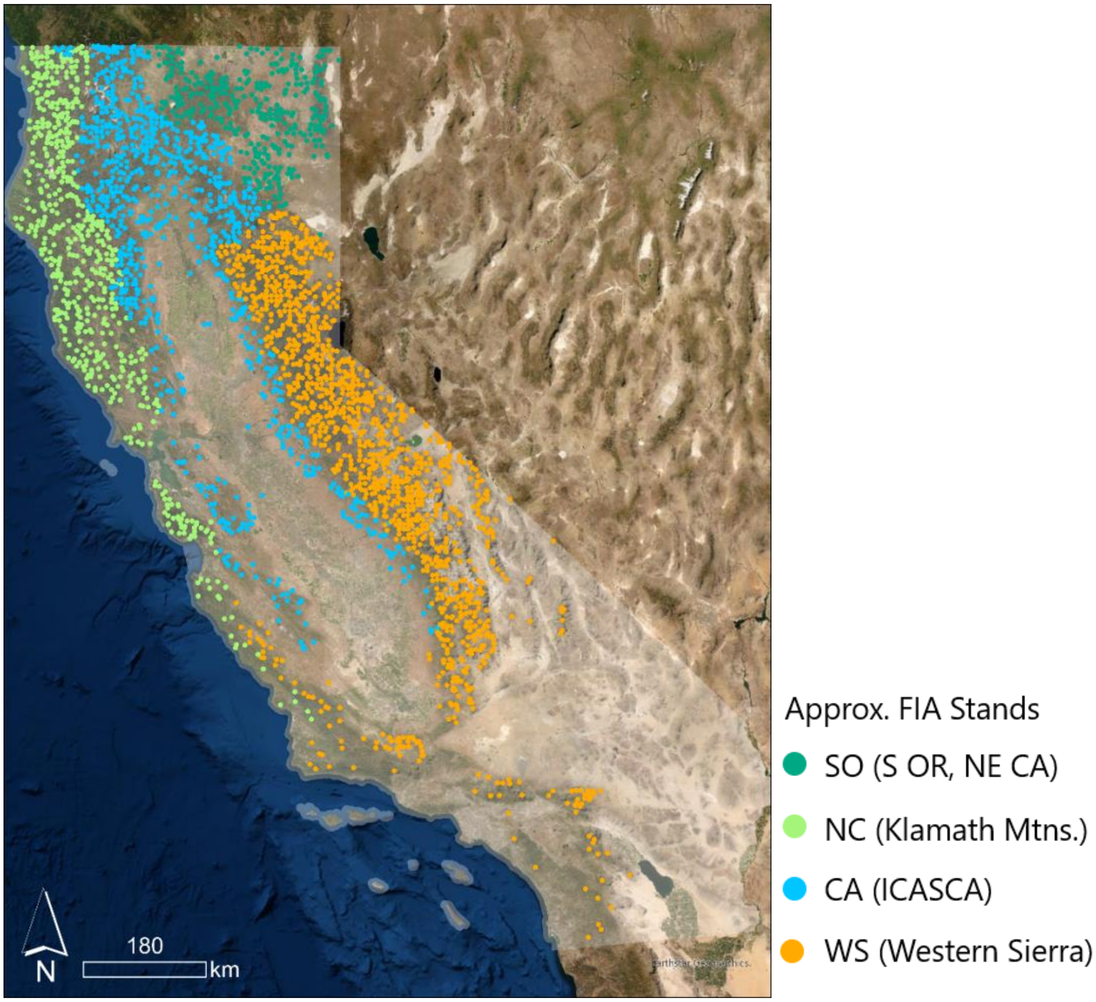


Figure 1. Approximate locations of FIA plots in California used for this analysis, shaded by FVS Variant.

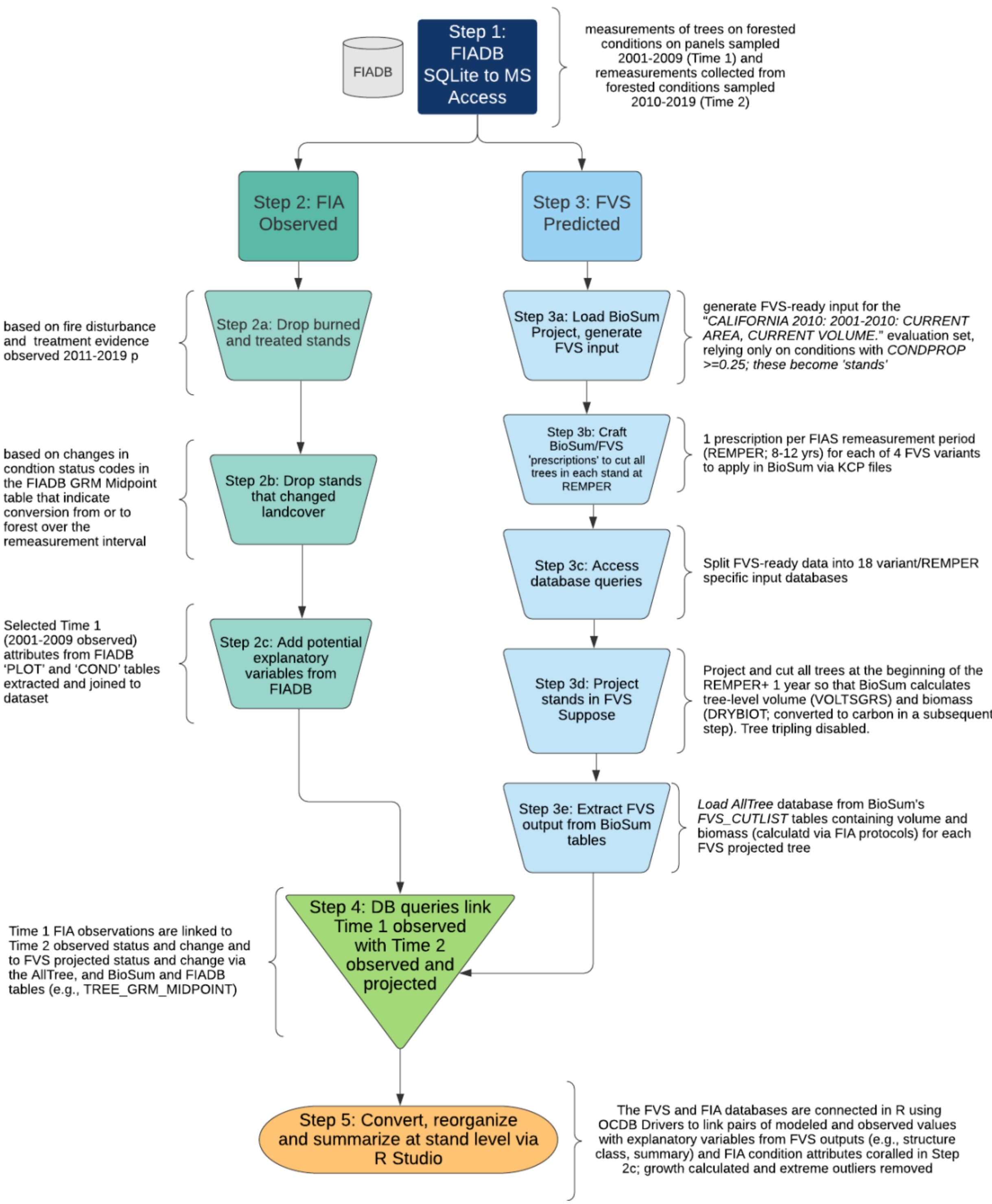


Figure 2. Data and workflow diagram for validation analysis.

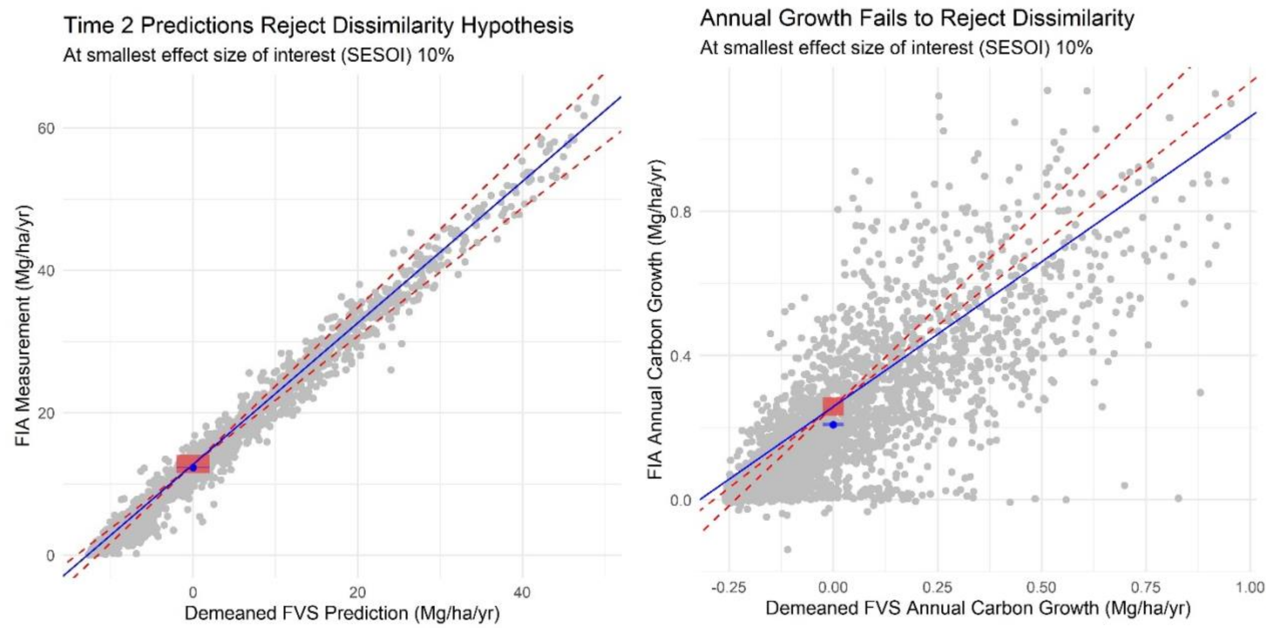


Figure 3. Scatter plots of above-ground, live tree, wood, and bark carbon at Time 2 (stocks ~10 years following initial measurement) and Annualized Growth for stands with equivalence confidence intervals with SESOI at 10%

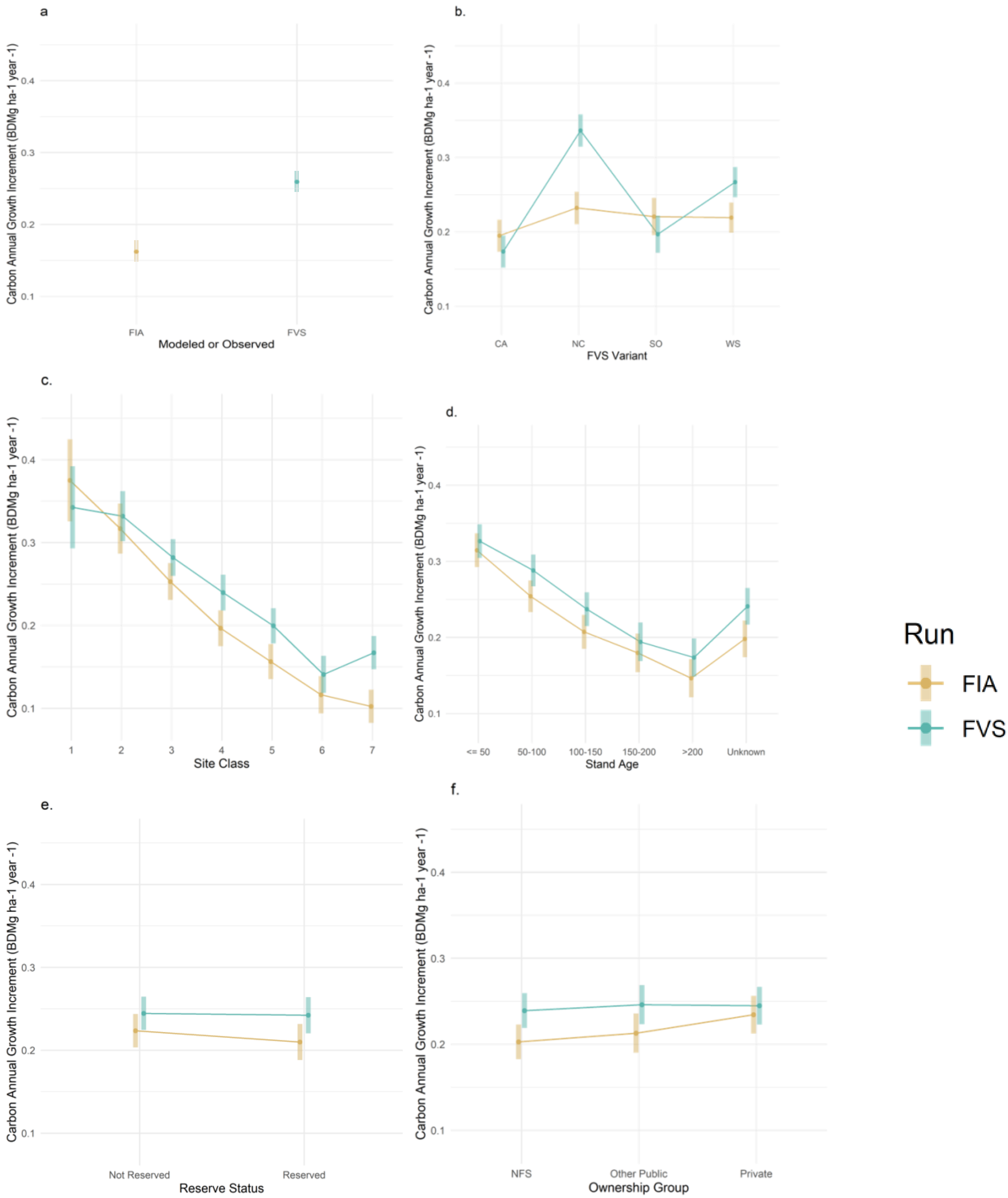


Figure 4. Average effect for the run (FIA vs FVS) variable (a) and interaction terms (b-f), showing the showing the relative effect of each level of each interaction term on the modeled FIA and FVS means for those levels, illustrating there FVS predictions over- and under-state modeled FIA values.

References

American Carbon Registry. 2018. “Improved Forest Management Methodology for Quantifying Removals and Emission Reductions through Increased Forest Carbon Sequestration on Non-Federal U.S. Forestlands.” Version 1.3, (Accessed 3 April 2022)Badgley, Grayson et al. 2022. “Systematic Over-Crediting in California’s Forest Carbon Offsets Program.” Global Change Biology 28(4): 1433–45.

Bagdon, Benjamin A. et al. 2021. “A Model Evaluation Framework Applied to the Forest Vegetation Simulator (FVS) in Colorado and Wyoming Lodgepole Pine Forests.” Forest Ecology and Management 480(October 2020): 118619. https://doi.org/10.1016/j.foreco.2020.118619.

Bechtold, William A, and Paul L Patterson. 2005. “The Enhanced Forest Inventory and Analysis Program — National Sampling Design and Estimation Procedures.” USDA General Technical Report SRS-80: 85.

Birk, M.A. 2019. “Measurements: Tools for Units of Measurement.” https://cran.r-project.org/web/packages/measurements/index.html.Borys Tkacz et al. 2011. National Roadmap for Responding to Climate Change. https://www.fs.fed.us/climatechange/pdf/Roadmapfinal.pdf

Burrill, E. A. et al. 2015. “The Forest Inventory and Analysis Database.” USDA Forest Service 2: 1026. https://www.fia.fs.fed.us/library/database-documentation/current/ver90/FIADB User Guide P2_9-0-1_final.pdf.

Canavan, Sean J., and Carl W. Ramm. 2000. “Accuracy and Precision of 10 Year Predictions for Forest Vegetation Simulator - Lake States.” Northern Journal of Applied Forestry 17(2): 62–70.

CARB Air Resources Board. 2011. Compliance Offset Protocol U.S. Forest Projects. http://www.arb.ca.gov/regact/2014/capandtrade14/ctusforestprojectsprotocol.pdf.

Cawrse, Dave et al. 2010. “Forest Vegetation Simulator Model Validation Protocols Validation Subcommittee FVS Steering Team.” Management (January).

Christensen, Glenn A. et al. 2021. “AB 1504 California Forest Ecosystem and Harvested Wood Product Carbon Inventory: 2019 Reporting Period DATA UPDATE.” US Forest Service and CA Dept of Forestry and Fire Protection Report (February).

Climate Action Reserve. 2017. “Forest Project Protocol Summary of Proposed Revisions.” Version 4.0. https://www.climateactionreserve.org/wp-content/uploads/2018/05/Forest-Project-Protocol-V4.0-package-05142018.pdf

Crookston, Nicholas L. 2014. “Climate-FVS Version 2: Content, Users Guide, Applications, and Behavior.” USDA Forest Service - General Technical Report RMRS-GTR (319 RMRS-GTR): 1–38.

Crookston, Nicholas L., and Gary E. Dixon. 2005. “The Forest Vegetation Simulator: A Review of Its Structure, Content, and Applications.” Computers and Electronics in Agriculture 49(1): 60–80.

Dixon, G. E. 2013. “Essential FVS: A User’s Guide to the Forest Vegetation Simulator”. January, 226. http://www.fs.fed.us/fmsc/ftp/fvs/docs/gtr/EssentialFVS.pdf

Eker, Sibel, Elena Rovenskaya, Michael Obersteiner, and Simon Langan. 2018. “Practice and Perspectives in the Validation of Resource Management Models.” Nature Communications 9(1): 1–10. http://dx.doi.org/10.1038/s41467-018-07811-9.

Fischer, Paul W., Alison C. Cullen, and Gregory J. Ettl. 2017. “The Effect of Forest Management Strategy on Carbon Storage and Revenue in Western Washington: A Probabilistic Simulation of Tradeoffs.” Risk Analysis 37(1): 173–92.

Fried, Jeremy S. et al. 2017. “Inventory-Based Landscape-Scale Simulation of Management Effectiveness and Economic Feasibility with BioSum.” Journal of Forestry 115(4): 249–57.

van der Gaast, Wytze, Richard Sikkema, and Moriz Vohrer. 2018. "The Contribution of Forest Carbon Credit Projects to Addressing the Climate Change Challenge." Climate Policy 18(1): 42–48.
Gibbon, A. et al. 2013. Methodology Element Assessment Report: VCS Version 3.
Griscom, Bronson W. et al. 2017. "Natural Climate Solutions." Proceedings of the National Academy of Sciences of the United States of America 114(44): 11645–50.
Harris, Nancy L. et al. 2021. "Global Maps of Twenty-First Century Forest Carbon Fluxes." Nature Climate Change 11(3): 234–40. http://www.nature.com/articles/s41558-020-00976-6.
Haya, B. et al. "Voluntary Registry Offsets Database." https://gspp.berkeley.edu/faculty-and-impact/centers/cepp/projects/berkeley-carbon-trading-project/offsets-database. (Accessed 3 April 2022)
Hebbali, A. 2020. "Olsrr: Tools for Building OLS Regression Models." https://cran.r-project.org/package=olsrr.
Hoover, C, and S Rebain. 2008. "The Kane Experimental Forest Carbon Inventory: Carbon Reporting with FVS." Carbon (March): 17–22.
Jiang, Lifen et al. 2021. "Country-Level Land Carbon Sink and Its Causing Components by the Middle of the Twenty-First Century." Ecological Processes 10(1).
Keen, Rachel M. et al. 2022. "Changes in Tree Drought Sensitivity Provided Early Warning Signals to the California Drought and Forest Mortality Event." Global Change Biology 28(3): 1119–32. https://onlinelibrary.wiley.com/doi/10.1111/gcb.15973.
Kerchner, Charles D., and William S. Keeton. 2015. "California's Regulatory Forest Carbon Market: Viability for Northeast Landowners." Forest Policy and Economics 50: 70–81. http://dx.doi.org/10.1016/j.forpol.2014.09.005.
Lacerte, V. et al. 2004. "Testing the Lake States Variant of FVS (Forest Vegetation Simulator) for the Main Forest Types of Northern Ontario." Forestry Chronicle 80(4): 495–506.
Lenth, R.V. 2022. "Emmeans: Estimated Marginal Means, Aka Least-Squares Means."
Miles, Patrick D, and W Brad Smith. 2009. "Specific Gravity and Other Properties of Wood and Bark for 156 Tree Species Found in North America." Res. Note. NRS-38: 35. https://www.nrs.fs.fed.us/pubs/rn/rn_nrs38.pdf.
Pokharel, Bharat, and Robert E Froese. 2008. "Evaluating Alternative Implementations of the Lake States FVS Diameter Increment Model." 255: 1759–71.
R Core Team. 2021. "R: A Language and Environment for Statistical Computing."
Reinhardt, Elizabeth D, and Nicholas L Crookston. 2003. "The Fire and Fuels Extension to the Forest Vegetation Simulator."
Robinson, A. P. 2016. "Equivalence: Provides Tests and Graphics for Assessing Tests of Equivalence." https://cran.r-project.org/package=equivalence.
Robinson, Andrew P., Remko A. Duursma, and John D. Marshall. 2005. "A Regression-Based Equivalence Test for Model Validation: Shifting the Burden of Proof." Tree Physiology 25(7): 903–13.
RStudio Team. 2020. "RStudio: Integrated Development for R. RStudio." http://www.rstudio.com/.
Russell, Matthew B et al. 2015. "Eastern Spruce Budworm in Northern Minnesota." (August).
Shaw, J D. 2009. "Using FIA Data in the Forest Vegetation Simulator." USDA Forest Service Proceedings: 1–16.
Smith-Mateja, Erin E., and Carl W. Ramm. 2002. "Validation of the Forest Vegetation Simulator Growth and Mortality Predictions on Red Pine in Michigan." Second Forest Vegetation Simulator Conference, U.S. Department of Agriculture, Forest Service, Rocky Mountain

Research Station. Fort Collins, CO, (September): 38–44. http://www.fs.fed.us/rm/pubs/rmrs_p025.pdf?origin=publication_detail.

USDA Forest Service, PNW Research Station, Resource Monitoring and Assessment Program. 2010. "Regional Biomass Equations Used by FIA to Estimate Bole, Bark, and Branches." 1–20. unpublished documentation on file with PNW-RMA in Portland, Oregon and available at: https://www.climateactionreserve.org/wp-content/uploads/2011/03/Biomass-Equations-Updated-3-25-2011.pdf

USDA Forest Service, PNW Research Station, Resource Monitoring and Assessment Program. 2014. "FIA Volume Equation Documentation." unpublished documentation on file with PNW-RMA in Portland, Oregon and available at: https://ww2.arb.ca.gov/sites/default/files/cap-and-trade/offsets/copupdatereferences/qm_volume_equations_pnw_updated_091914.pdf

Wickham, H. 2016. "Ggplot2: Elegant Graphics for Data Analysis."

Wickham, Hadley et al. 2019. "Welcome to the Tidyverse." Journal of Open Source Software 4(43): 1686. https://joss.theoj.org/papers/10.21105/joss.01686.

Transitions between Chapter 2 and 3

Forest land managers seeking to make evidence-based decisions on resource management need accurate models calibrated for their systems. In Chapter 2, I validated the Forest Vegetation Simulator on 99% of the trees in California and found the predicted growth rates are significantly dissimilar to observed growth rates over a short timeframe. Without further model calibration to correct for this overestimation of carbon growth rates in forests, forest managers may inappropriately select silvicultural treatments for forests that will under-sequester based on lower observed growth than the model projects.

Even with more accurate models, people managing forests are trying to achieve difficult objectives, like operationalizing 'forest resilience' or 'climate adapted forest.' These decisions will be context specific and present tradeoffs across different scales. In my third chapter, I explore how managers have interpreted objectives around carbon storage in California forests. Ecosystem carbon pools are dependent on the total carbon stock and the rate of carbon turnover. Managing forest carbon in ecosystems that are experiencing more severe fire disturbances creates tradeoffs for maintaining aboveground forest carbon because the rate of carbon turnover can increase. Land use sciences regularly investigates tradeoffs involved in land management. This field can be applied to evaluate the tradeoffs public agencies create when designing policy. In the third chapter, I explore how two state agencies that have created policies for private landowners to manage forests for climate goals.

Chapter 3. Managing nature-based solutions in fire-prone ecosystems: Competing management objectives in California forests evaluated at a landscape scale

Claudia Herbert, Barbara Haya, Scott L. Stephens, Van Butsic

Abstract

California's cap-and-trade compliance offset market incentivizes existing forest managers to maintain elevated carbon stocks. It provides these incentives without enforcing standardized fire mitigation practices despite many such additional aboveground carbon stocks being located in fire prone regions. Here, we evaluated the difference between management actions in California forests that participated in the carbon offset market versus those that engaged with other state programs to reduce wildfire risk via fuel reduction treatments. Using remotely sensed data from the California Forest Observatory and the Moderate Resolution Imaging Spectroradiometer, we compared the vertical forest structure and vegetation canopy trends on forest offsets with forests that are receiving fuel treatment. We found California forests managed for carbon under the Improved Forest Management (IFM) program by the California Air Resources Board had higher levels of biomass than forests managed for fire risk reduction as indicated by 2016 lidar-estimated fuel loads. Further evaluation of these IFM-participating forests between 2016 and 2020 found that such projects did not reduce their fuel loads, whereas lands receiving grants for fuel management did, indicating that on average, the IFM projects were not engaging in fuel reduction efforts. However, despite the differences in fuel management between IFM projects and active fuel treatments, we found that both types of management saw an average declining trend in vegetation greenness between 2015-2021. While declining greenness is expected of active fuel treatments associated with vegetation removal, such a trend in the case of IFM projects (i.e., forests without active vegetation removal) introduces additional wildfire risk. Managing forests for long-term carbon storage and sequestration should consider fire risk mitigation. If we are seeing little evidence of fuel reduction in the first decade of IFM projects implementation, assumptions around the century-long duration of such carbon stocks are potentially in need of recalibration. We recommend that policymakers reevaluate the incentives being directed at carbon stock preservation or expansion to better encompass the growing wildfire risk in California.

Keywords: forest carbon; wildfire; fuel management; remote sensing; Improved Forest Management; offsets

Introduction

Fire releases carbon stored in forest biomass into the atmosphere, making wildfires a threat to storing carbon in forests (Hurteau & Brooks, 2011; M. P. North & Hurteau, 2011). In California, wildfires are having a negative impact on aboveground live carbon stocks in forests: between 2001 to 2010 the estimated aboveground live carbon stocks in California declined by a rate of 0.8% per year. These declines were primarily driven by wildfire disturbances occurring on just 6% of the total forest area (Gonzalez et al., 2015). The emissions associated with burning forests drive most of the estimated greenhouse gas emissions from wildfires in California and these emissions may increase as climate change is expected to increase areas burned in California, in part from warming-driven fuel drying (California Air Resources Board, 2020; Westerling et al., 2011; Williams et al., 2019).

The growing impact of wildfire on carbon emissions is occurring within an already existing forest management crisis. Over a century of fire removal and suppression has altered species composition and function in the Western United States (Collins et al., 2017). Forests in California are, on average, supporting higher tree density, with an estimated six-to-seven-fold increase in tree density between 1911 to 2011 (M. P. North et al., 2022). Most forest stands in California support tree densities which place them in full competition, reducing tree growth efficiency and resilience to stressors like drought and fire, and putting them at high mortality risk (M. P. North et al., 2022). With expected increases in climatic stressors from climate change, these current forest densities are not sustainable. Using historic forest condition data to predict forest survival under future climate scenario RCP 8.5, research on the Sierra Nevada and Southern Cascade has found that forests may be unable to support aboveground biomass above 25% of its current average levels (Bernal et al., 2022).

From a carbon perspective, increases in tree density do not necessarily translate to increased forest carbon because the tree density has been primarily from small-diameter trees. The combination of removing frequent fires and legacy of logging the largest trees have caused a decline in large trees in California forests by 50% between the 1930s to the 2000s, being replaced with smaller trees, and species composition shifting away from pine dominance (McIntyre et al., 2015). This shift in dominance of smaller, fire-sensitive trees has reduced total live carbon stocks (M. North et al., 2009). The largest 1% of trees make up about 30% of live aboveground biomass in the United States, and generally, large trees dominate carbon stocks in temperate US forests (Lutz et al., 2018; Mildrexler et al., 2020). Additionally, because carbon accumulation increases continuously with tree size, large-diameter trees being replaced with small-diameter trees negatively impacts the rate of carbon sequestration, along with reducing total carbon stocks (Stephenson et al., 2014). These changes in forest structure and density are exacerbated by acute drought stressors, which create forest conditions where subsequent disturbances have the potential to be more severe than they otherwise would have been (Stephens et al., 2018). Another consequence of increasing disturbance severity is that these more severe disturbances can alter plant successions away from forests to grass or shrublands (Falk et al., 2022). Forest conversion to other types of vegetation cover can bring with it reduced aboveground carbon storage capacities and fire regimes that burn more severely (Coop et al., 2020). This interplay between climate change, fire exclusion, and antecedent disturbances changing forest composition creates the conditions that make mega-fires (Stephens et al., 2014).

Despite these challenges, for forests in the Western US, management aimed at vegetation removal from either forest restoration or fuels reduction could promote forests better able to withstand changing fire, drought, and beetle stressors (Stephens et al., 2021). Fuel treatments can address the imminent threat of extreme wildfires, while forest restoration efforts can reduce forest density by using tools like prescribed fire and restoration thinning treatments to reduce available fuels to prevent the worst effects of extreme wildfire (Stephens et al., 2020). Along with effectively reducing fire severity, meaning reduced tree mortality, fuel treatments can also reduce wildfire emissions (M. P. North & Hurteau, 2011). Widespread prescribed fire application could reduce fire related CO2 emissions in the Western US by 18-25% relative to a business-as-usual scenario (Wiedinmyer & Hurteau, 2010).

To increase the pace of implementing fuel treatments in California, the California Department of Forestry and Fire Protection (Cal Fire) provides landowners funding, planning, and implementation support to reduce costs of implementing fuel treatments. There are five Cal Fire programs that support fuel reductions: The California Fire Plan, Vegetation Management Program, Forest Health Grants, Fire Prevention Grants, and California Vegetation Treatment Program. Landowners supported by these grants for fuels reduction can achieve their goals using a diversity of methods, including prescribed fire, tree thinning, pruning, chipping, and roadway clearance (CAL FIRE, 2022). From a carbon accounting perspective, deciding to implement a fuel treatment means an immediate reduction in total amount of carbon, as fuel treatments remove or burn vegetation. Even if management causes some temporary reductions in carbon, over a century, large-scale restoration is projected to increase carbon stability (Liang et al., 2018). In fire prone ecosystems, stable carbon is the amount of carbon that is predicted to survive a wildfire. Managing stable carbon stocks in forests requires balancing total carbon pools against disturbance risks that emit carbon. Based on research from California forests, prescribed fire, mechanical treatments, or a combination thereof have been shown to increase stable live carbon (Foster et al., 2020; M. North et al., 2009). Programs like California Fire Plan explicitly reference fuel reductions as part of maintaining California's carbon sinks and have overarching goals of improving forest health through fire risk reduction (State Board of Forestry and Fire Protection & CAL FIRE, 2018).

Stable forest carbon management, which includes fuel management, is needed to maintain carbon durably in California forests. Yet, a large portion of funding in response to climate change has been to pay landowners for increasing aboveground forest carbon to offset greenhouse gas emissions elsewhere. One example of such a program is California's cap-and-trade Compliance market. This program is administered by California's Air Resources Board (CARB) which facilitates a carbon market where forest landowners can enroll in an Improved Forest Management (IFM) offset program and be paid for managing existing forests for carbon stock maintenance and expansion (Air Resources Board, 2011; Projects, n.d.). IFM projects in this market are intended to store carbon for 100-years above a regionally determined carbon baseline. Before the project is awarded carbon credits, the IFM project owner submits initial documents to registries disclosing how they intend on increasing aboveground live carbon for the next century. This can include strategies like increasing rotation length, decreasing harvest intensity, or switching to uneven age management. IFM projects, once validated and issued credits, cannot significantly reduce on-site carbon stocks without a penalty.

CARB and Cal Fire currently administer programs that pay private landowners for forest management, with both agencies interested in stabilizing and storing carbon in forests. Forest landowners in CARB's Improved Forest Management (IFM) offset program are incentivized to increase aboveground carbon, while the grantees of the Cal Fire program are supported to temporarily remove aboveground carbon. Forest carbon offset programs that require sustaining already elevated forest biomass create the possibility for conflicting incentives with stable carbon management. And these IFM programs are relatively popular; credits generated from IFM projects represent 82.8% of the credits that have been retired during the compliance periods in California's compliance market (CARB Offset Credit Issuance Table, n.d.). Focusing on increasing aboveground carbon in fire prone forests could lead to management decisions counter to stable carbon storage (Bernal et al., 2022; Hurteau et al., 2008; M. P. North & Hurteau, 2011). It is unknown if these IFM programs effectively target forest where biomass increases or decreases will lead to long term stable carbon on the landscape.

IFM projects are minimally incentivized to do fuel treatments under the current standard design. The buffer pool is a reserve of credits that every IFM project contributes a set amount of their issued credits to as an insurance policy. Typically, a project contributes 9-19% of issued credits, with an average around 15%, to this pool (CARB Air Resources Board, 2011). Projects that complete a 'qualified' fuel treatment can receive up to a 2% reduction on their buffer pool contribution, meaning the project owner receives a greater percentage of the credits generated. This 2% incentive for fuel treatments is the only way a project can reduce its contribution to the buffer pool for natural disturbance risks and requires a third-party auditor confirm the fuel treatment. Yet, based on the observed contributions projects are giving to the buffer pool, most of the IFM projects in California are not taking advantage of this buffer pool contribution reduction.

The problem for assessing the effectiveness of IFM project management strategies is that it is difficult to determine exactly which types of management takes place within IFM projects. While the projects submit documentation to public registries, in practice, these documents vary in the level of detail and supplemental information provided describing management plans. Information like whether a project owner plans on funding fuel management in the forms of mechanical thinning or prescribed fire between harvests is not required in the documentation, even if a IFM project owner intends on employing one of these techniques. The other limitation on relying on registry-submitted documentation is that it details a project owner's intent, but they are not bound to execute the detailed management, they only need to maintain carbon levels above a regionally determined baseline and any crediting increases above that baseline. It would be reasonable to expect some of these projects are using fuel management given the legacy of fire suppression increasing tree density and the growing understanding that climate change will exacerbate threats to storing carbon in trees, but use of fuel treatment in IFM projects has never been quantified.

Here we use novel remotely sensed data and a host of statistical techniques to answer two main questions. First, how common are fuel treatments on IFM lands and how do these treatments compare to treatments carried out under Cal Fire fuel reduction programs? Second, we ask if forest health, as proxied by vegetation greenness, has changed on IFM forest relative to forest that have received Cal Fire grants. By answering these questions, we hope to better understand if California's large carbon offset credit program is effectively addressing the tradeoffs between carbon, fuel management, and vegetation health outcomes.

Materials and Methods

Study Area

IFM projects under California's compliance offsets market can be developed on any private forest in the United States. We limit our analysis to the IFM projects in California, where projects are in the North Coast and Southern Cascade regions (LANDFIRE, 2001). This research spans forests in California that are managed under the IFM offsets or Cal Fire fuel treatments from the same region (Figure 1). We excluded fuel treatments from outside of these two regions because there may be regional variability in fuel loads and vegetation trends that would make them too different to compare to the IFM projects.

Data
Cal Fire's Management Activity Project Planning and Event Reporter (CalMAPPER) database includes location and project information for forest and fuel management projects across Cal Fire programs (CAL FIRE, 2022). We used GIS treatment polygons produced by Cal Fire which record areas of grant funded fuel treatments to establish how fuel treatments alter vegetation in forests within the same regions as the IFM projects.

CARB-approved IFM projects submit GIS shapefiles or maps of project boundaries to the three registries approved for CARB offset use, American Carbon Registry, Climate Action Reserve, and Verra. The dataset we use to determine where IFM projects occur in California are based on these submitted project boundaries to the registries (Stapp et al. 2022). This data represents 31 of the 40 CARB-approved IFM projects in California. The remaining nine projects listed on the public registries lacked spatial data. We have no reason to believe projects with missing geospatial data are absent for a systemic reason, so we proceed with the analysis using the available data, considering this a representative dataset of the CARB-approved IFM projects in California.

We limited the CalMAPPER treatment polygons to projects that had 'fuel reduction' recorded as the 'Treatment Objective' and were in the North Coast or the Southern Cascade region (CAL FIRE, 2022). This limited our dataset to projects that were in the California Fire Plan, Vegetation Management Program, and Forest Health grants. This dataset included 448 Cal Fire-approved fuel reduction projects. Cal Fire grants span across California and the data is provided both as a larger project boundary and the polygons where management activity occurs. This analysis uses the more specific polygons where management occurs for sampling, called the Treatment Polygons. We used the fuel treatments start and end dates to determine the status of fuel reduction activities and gain a better sense of how fuel treatments change vertical structures of fuel over time. Projects that had end dates prior to 2016 are recoded as 'Completed Fuel Treatments,' projects that were completed by 2020 were recorded as 'Active Fuel Treatments,' and projects that did not begin until after 2020 were recorded as 'Planned Fuel Treatments.'

The California Forest Observatory produced annual maps of fuels for forests in California at 10-meter spatial resolution from the summers of 2016 and 2020. The fuel structure data are modeled products based on satellite radar, satellite spectral imagery, and airborne lidar flights (California Forest Observatory, 2020). Airborne lidar is a useful tool for estimating fuel loads because it can penetrate canopy covers to estimate vertical structures of vegetation based on the timing and density of points that return to the sensor (Jarron et al., 2020; Mutlu et al., 2008; Su et al., 2016). The California Forest Observatory lidar-derived fuel maps represent the best available fuels dataset that has both high spatial resolution and is produced for multiple years. This spatial and temporal coverage was necessary for research studying how fuels change over time at a landscape scale. Fuel data includes ladder fuel density (%), surface fuel (unitless), canopy cover (%), canopy height (meters), canopy layer count (#), and canopy base height (meters), relevant to fuel and carbon discussion. Ladder fuel density ($r2 = 0.78$) considers the lidar point returns from 1 to 4 meters above ground returns which represent the proportion of surface fuels in the understory. Surface fuels are based on outputs from Scott & Burgan Fire Behavior Fuel Models where higher values show higher levels of surface fuels. Canopy cover ($r2 = 0.91$) reports the horizontal cover fraction from tree canopies. Canopy height ($r2 = 0.86$) is the distance between ground and top of canopy returns. Canopy layer count ($r2 = 0.765$) is the number of distinct vertical canopy layers, which could be a proxy for leaf area index or canopy complexity. And canopy base height ($r2 = 0.70$) is the distance between the ground and the lowest branches in the canopy, predictive of whether a surface fire will transition to a canopy fire (California Forest Observatory, 2020). For all but surface fuel estimates, the California Forest Observatory produces accuracy estimates (S. Figure 1). Using the California Forest Observatory produced API in Jupyter notebook, the six types of fuel data from 2016 and 2020 were queried for the study area.

Using the 'geemap' package for Google Earth Engine access in Python, this fuel data was sampled within IFM project and fuel treatment areas at a 300-meter sample grid. We choose this grid size as a good trade-off between increase sample observations and limiting spatial autocorrelation. Sampled observations of fuel and fuel change were exported as a CSV file for further analysis in R Studio.

Objective 1. Lidar-based fuels comparison

We compare changes in IFM projects with different stages of implemented fuel treatments (completed, active, and planned) to test if IFM projects alter fuels in ways that resembles fuel treatments. For example, if the IFM projects are managing fuels during the study period, then they might appear to change fuels in similar ways to active fuel treatments. We expected active fuel treatments would have a reduction in the fuel, while the completed and planned fuel treatments might have a positive change in fuel strata estimates, as vegetation recovers after previous fuel treatments or grows prior to future fuel treatments. Depending on how the IFM projects changed during this remeasurement relative to the three stages of fuel treatments, we can determine if fuels have been managed. The input of the fuel analysis is the CAFO lidar-based dataset which produces estimates of different vertical levels of carbon.

Because we were interested in observing which strata are most impacted by fuel treatments or IFM project management, we analyzed the six fuel treatments as separate regressions to understand how these management approaches altered forest and fuel structure. For example, using a popular 'thin from below' silvicultural approach will maintain high levels of canopy cover and canopy height and remove the non-dominant trees that could carry surface flames from surface to canopy. The result of such a treatment would be a reduction in surface and ladder fuel, an increase in canopy base height, a decrease canopy layer count, and little or no-change in canopy cover or canopy height. We report all tests for the following section for the six separate fuel strata because the direction of changes over time depends on the strata of fuel evaluated.

Fuels in IFM and fuel treatments

To understand if fuels differ between IFM projects and fuel treatments, we first compared fuel estimates for IFM projects and Cal Fire fuel treatments at the beginning of our study period. We compared, offsets to the three categories of fuel treatments and we visualized our results by creating a boxplot of the starting 2016 fuel estimates by fuel type and region (Figure 2).

Propensity score matching

Forests managed as offsets may differ from forests eligible for grant-funded fuel treatments, creating a form of selection bias. Therefore, we use propensity score matching to prepare a dataset that limits the differences between our controls, or fuel treatments, relative to our treatments, or IFM offsets (Austin, 2011). We use factors associated with vegetation growth because we thought factors that might contribute to whether a forest is an offset or receives a grant-funded fuel treatment might be associated with how productively vegetation grows on that site. Along with sampling how fuel changed at each sample, we sampled climatic variables like total winter precipitation and max summer temperature as annual composites from 2016-2020, physical variables like elevation, slope, and aspect used in parametric matching (Table 1). The climatic variable selected were based on variables considered important for predicting forest survival (Bernal et al., 2022). We implemented this parametric matching to identify fuel treatments that were otherwise like IFM offsets using all the variables noted in Table 1, the fuel estimate from 2016, and a variable signifying which region the IFM project or fuel treatment is located.

Linear Regression

This matched dataset was used in a linear regression to predict the change in fuel loading controlling for the variables used for matching, the starting fuel estimate from 2016, the region, a treatment term, indicating whether the sample was taken from an offset, or a fuel treatment completed by 2016, and an interaction term between region and treatment. The interaction term allowed us to independently predict the treatment effects for the two regions studied. We used the output of the linear regression in an estimated mean marginal effects model to predict the change in fuel load based on treatment and region using the interaction term. To understand the effect of fuel load changes on IFM projects relative to fuel treatments, we used the linear regression outputs to predict changes in fuel for a four-year period for the IFM projects and fuel treatments. This allowed us to report the effect in same units as the original fuel estimate, easing interpretation. Predicted changes in fuel estimates that are greater than zero indicate that fuel is predicted to increase for that group, whereas a change in fuel that is less than zero indicate that there is a predicted loss of fuel, consistent with vegetation removal. The pre-regression matching, linear regression, and prediction was repeated for the six types of fuels estimated from the California Forest Observatory dataset.

Object 2 Spectral Based comparison

Beyond changes in total fuel amount, changing vegetation greenness also has implications for forest carbon and fire risk. A forest may show signs of vegetation stress or declines in productivity through decrease in canopy vegetation greenness (Sims et al., 2006). Forests with reductions in greenness may also indicate areas of increased fuel dryness, making it more available to burn if there is a fire (Roberts et al., 2006). We used satellite-based spectral estimates of vegetation greenness using the Enhanced Vegetation Index (EVI). EVI was designed for use in high density vegetation areas and has been demonstrated to be efficient indicators of forest productivity and die back (Ogaya et al., 2015; Sims et al., 2006).

This estimate of vegetation greenness, was calculated in Google Earth Engine using the Moderate Resolution Imaging Spectroradiometer (MODIS) imagery, based on the Non-Parametric Trend Analysis tutorial (Nicholas Clinton, n.d.). The MODIS EVI data product was the basis of a constructed time series of vegetation greenness between 2015-2021 based on clear-sky, cloud-corrected imagery collected between August and September. A Kendall-Man non-parametric trend analysis was used to reduce the EVI time series imagery to identify areas of increasing or decreasing vegetation greenness. These analyzed timeseries data produced a single number of the observed trend in greenness for Objective 2. A positive EVI Kendall Mann value indicates that the vegetation greenness is increasing over time. The MODIS EVI data was sampled for the offset and project areas using the same sample grid at 300-meter density that was used for the fuel analysis in Objective 1.

Bootstrap for determining differences in the median greenness trend

To further analyze how the trends of vegetation greenness differ between the two types of forest management, we used a bootstrap to estimate the true differences between the sample median values. The strength of this approach is that it can take non-normally distributed observations, with differing variances, and develop a test that compares whether the two samples might be similar (Johnston & Faulkner, 2021). We used a two-sample bootstrap to test whether the median changes in vegetation greenness differed by offset and fuel treatment groups. This two-sample bootstrap test samples the different group with 5000 replications to estimate the true median differences. For observations like trends in vegetation greenness, where the observations may not be normally distributed, this is a non-parametric statistical test that can determine if there is a statistically significant difference between median values of the offset and fuel break groups and can estimate the magnitude of difference between groups. When the median differences for two groups contains zero, the two groups can be considered similar. To determine statistical significance, confidence intervals were constructed around the bootstrapped estimates and a 0.05 alpha level. Along with determining whether there is a statistically significant difference, this test also established the direction of differences between the two samples. A negative bootstrap differences indicate that the median IFM projects have vegetation greenness is lower than the median fuel treatments. A positive bootstrapped median difference indicates that the fuel treatments have improved vegetation greenness more than the IFM project.

Validation

Previous research has found that public datasets produced for tracking timber and non-timber management from the Forest Service and Cal Fire overrepresented areas of treatments (Knight et al., 2022). Since this work relies on Cal Fire's CalMapper data to create the sampling boundaries for the fuel treatment areas, we assessed if there was a logical change in carbon levels based on our classification of 'completed,' 'active,' and 'planned' fuel treatments. We expected that planned, projects that had not started by 2020, would have higher ladder and surface fuel estimates than active projects; the completed fuel treatments would have the lowest surface and ladder fuel estimates; and that the active fuel treatments would show the greatest reduction relative to the other two types of fuel treatments. Our validation check suggested that our sampling scheme using CalMAPPER was accurately capturing changes in fuel loads consistent with active, complete, and planned fuel treatments. Active fuel treatments tend to have more negative changes in fuels than the completed and planned fuel treatments (S. Figure 2).

Results

Objective 1. Lidar-based fuels comparison

Fuels in IFM and fuel treatments

IFM projects have higher starting levels of fuel estimates than the forest where fuel treatments were completed or were active during the observation period. Based on the 2016 CAFO fuel estimates, IFM projects have the higher average ladder, canopy cover, canopy height, and canopy layer count than the fuel treatments that were active after 2016 or completed by 2016 in both the Southern Cascade and North Coast region (Figure 2a, c, d, e).

For a fuel metric like canopy cover or canopy height, it makes sense that forests identified for carbon management would have high forest coverage, as tree height is correlated with tree size and dense forests would have higher canopy cover. The low canopy cover for fuel treatments in the

Southern Cascade region, as low as 25% for Active fuel treatments, could be attributed to the higher number of fuel treatments that are linear, possibly adjacent to roads. There are differences in the 2016 estimates of the vertical structure and canopy conditions estimates for forests where IFM projects are located relative to fuel treatments that will require addressing for further comparison.

Propensity score matching
The pre-regression matching improved the covariate balance between the control and treatment groups (S. Figure 3). There remained some differences for the temperature and precipitation variables that out method for matching could not resolve. Because the matching did improve the mean differences in the relevant 2016 fuel estimate (e.g., cc_2016 for canopy cover in 2016), which our boxplots indicated were different between the IFM projects and fuel treatment groups, we felt that this matched dataset effectively reduced some bias between treatment and control observations. The pre-regression matching reduced the dataset from 24,051 observations to 11,378 observations with equal number of treatment and control observations. There were too few observations from the planned fuel treatments in the Southern Cascade region to make inference, therefore this group is not reported.

Linear Regression
IFM projects have greater predicted changes in vertical structure and canopy composition than the fuel treatments (Figure 3). This indicates that even controlling for differences in biophysical factors that impact forest growth and the starting fuel loads, the observed changes in forest structure on IFM projects is increasing faster than the forests that are receiving or will receive fuel treatments. IFM projects have changes greater than zero for ladder and surface fuels that indicate increasing fuel amounts. However, the IFM projects are also the only group that had a predicted positive increase for forest height and canopy cover, which both would be consistent with increasing carbon stored in live tree carbon. Because there are not overlapping standard errors between the IFM projects and the active fuel treatments for any of the metrics, we conclude IFM projects are not reducing fuels like forests that are receiving fuel treatments.

In the North Coast region, where there were enough observations also evaluate planned fuel treatments, predictions of IFM projects changes were the closest to planned fuel treatments for changes in canopy base height and changes in canopy layer count. Of the predicted values from the fuel treatments, IFM projects did have predicted changes closer to the observations on planned fuel treatments. In the Southern Cascade region, IFM projects have predicted values closer to completed fuel projects.

Object 2 Spectral Based comparison
The average trend for the Kendall-Mann test of vegetation trends in greenness are reported for the IFM projects and fuel treatments for the two regions (Table 2). IFM projects in the North Coast had a negative trend in vegetation, expressed in a spectral EVI trend change of -29.10 (unitless). The negative value indicates a loss in vegetation greenness or productivity. The two IFM projects in the Southern Cascade region were the only group in our sample that had a positive trend in vegetation greenness, of 9.94 (unitless).

In the North Coast region, the completed fuel treatment in the North Coast had the most negative mean vegetation greenness trend. This could be from delayed treatment effects from prescribed fire and fuel management after management. Completed fuel treatments were omitted from the

Southern Cascade due to too few observations. In the Southern Cascade region, it is the planned fuel treatments that had the most negative value relative to the active and the IFM projects. For both regions, active fuel treatments cause a decline in vegetation greenness, indicated in the average negative vegetation trend (Table 2). This could be due to the vegetation removal or rearrangement associated with fuel treatments.

Bootstrap for determining differences in the median greenness trend

To further contextualize these observed trends in IFM projects relative to other forests, the bootstrap tests compare the vegetation trends on IFM projects with forests receiving fuel treatments. Active fuel treatments have similar median trends in vegetation to the IFM projects in both regions of California. The bootstrapped median difference for the IFM and active fuel treatments had a 95% confidence interval that contain zero, meaning the median differences in greenness trends between IFM projects and active fuel treatments could be zero (Figure 4c-d).

The largest difference in vegetation trends is in the Southern Cascade region between planned fuel treatments to IFM projects at 73 (unitless). This is consistent with the mean values, because the two IFM projects in this region had the highest overall positive greenness trend and the planned fuel treatments had the most negative fuel treatment (Table 2). Both the sample mean trend for IFM and planned fuel treatments in the North Coast are negative. Our estimated median difference being positive 20 (unitless) indicates IFM projects have a less negative trend in vegetation greenness relative to the planned fuel treatments.

Completed fuel treatments in the North Coast had median vegetation trend that was -55 relative to the IFM projects mean vegetation trend. Suggesting that there is at least a temporary decline in vegetation greenness following fuel reductions. This could be the byproduct unique to our observation window or could be capturing some delayed ecological stress following management. Our results can only suggest a difference in the distributions and the relative direction of difference.

Discussion

Previous research has used tools like modeling and field experiments to demonstrate that continuing to increase carbon in aboveground biomass will come with tradeoffs for long-term carbon management (Bernal et al., 2022; Foster et al., 2020; Liang et al., 2018). Using our remote sensing techniques, we found additional evidence that tradeoffs are already present in IFM carbon projects in fire prone regions between mitigating short-term fire risk and maintaining elevated carbon stocks for carbon crediting. We found IFM projects had higher lidar-estimated fuel loads in 2016 than the fuel treatments, indicating that these are forests with high starting levels of fuels. Between 2016-2020, these IFM projects did not reduce fuel loads in ways consistent with active or completed fuel treatments. From our vegetation greenness trend analysis, we found the management between IFM projects and active fuel treatments are having similar outcomes on vegetation (Figure 4). IFM projects are on average not reducing fuel loads, yet they are having reductions in vegetation greenness as if they were. Another reason why vegetation can lose greenness is a decline in fuel moisture, which also results in increased flammability (McEvoy et al., 2020). If these loses in vegetation greenness are from declining vegetation moisture, then the management occurring on IFM projects could be increasing fire risk through both not reducing fuel loads and increasing vegetation flammability.

Based on the first decade of the compliance market, our results suggest forests managed as offsets promote less beneficial outcomes for fire hazard reduction and vegetation greenness. There is an abundance of research that supports the idea management should be scaled in these types of forests—both for lofty concepts of forest resilience and for long term carbon storage, relevant to greenhouse gas accounting (Liang et al., 2018; M. P. North et al., 2022). And yet, when this climate-focused policy has been applied to forests, we do not observe carbon payments incentivizing the average project to implement more management. Only six of the 40 projects in California had qualified fuel treatments based on IFM project documentation submitted to carbon credit registries. Forest management for climate change should be aligned with other policies promoting more stable carbon management.

The timing of how CARB pays landowners could be creating perverse incentives with landowners maximizing carbon today rather than prioritizing managing for stable carbon. In California's compliance market IFM offset crediting program, the first credits issued reward the demonstrated carbon stock above a theoretical baseline amount of carbon determined regionally. Subsequent payments, represented as different vintages, are based on the additional growth that sustains carbon above the baseline. The California compliance market currently facilitated offset credits trading through 2030 while these IFM projects are intended to store carbon for 100 years (CARB Air Resources Board, 2011; State of California, 2017). This difference between the credit generation period and time carbon is to be stored could create a temporal tradeoff. Project managers are encouraged maximizing aboveground live carbon for the first decades of generating carbon credits, at the expense of further fuel management and fire risk reduction, which will be critical for storing carbon in forests through the end of century. Carbon credit generation is just one of multiple objectives most forest managers consider. There may be opportunities to realign incentives and support forest managers managing for carbon in fire prone ecosystems to better incorporate needs for mitigating fire risk and supporting carbon stability.

Thus, we turn to the incentives that this IFM offset program has created for managing forests. What does payments for forest carbon really incorporate? Creating incentives for maximizing aboveground carbon stored in forest biomass allows forest managers to shift economics towards delaying harvests. Based on registry submitted data, eight projects in California demonstrate increasing aboveground carbon through increasing rotation length. While delaying timber harvests is not mutually exclusive with no management occurring in these forests, some of these projects have no other management planned at the time of project documentation submission. Within the existing California IFM program, we have identified five potential ways policymakers could increase fuel management within IFM projects in fire prone regions. Each of these outlined approaches is accompanied by research questions to address whether our proposed idea would be able to deliver the desired outcomes of increasing the fire risk reduction on IFM projects in fire prone ecosystems.

The first approach is creating further incentives in buffer pool reductions for fuel break maintenance. As previously discussed, fuel treatments are currently incentivized via a reduction in the percentage of credits a project contributes to the buffer pool. But we observed that IFM projects have high starting levels of fuels in their vertical fuel structure and most of the projects do not show signs of fuel reduction. It could be possible that an incentive greater than 2% on the project's buffer pool contribution could drive more fire risk mitigation. Because the most carbon a project receives tends to be in the first year of accounting, a qualified fuel treatment approved at the start of the project would reduce the number of credits a project contributes to account for a century of natural disturbance insurance. If the 2% reduction in buffer pool contribution is not enough to offset this lose of credits generated, then this may be an area of policy improvement. Further research could study whether there are differences by landowners in their approach to management. If only some types of landowners completing qualified fuel treatments, it could be useful for identifying other barriers to management by other landowner groups.

Along with the previously discussed temporal misalignment, there is a spatial misalignment as this insurance pool encompasses IFM projects across the United States. Only in regions where there is moderate to high fire risk, would it make sense for projects to proactively be treating fuels. Yet these other regions contribute credits to compensate for fires in other regions. The buffer pool creates adverse selection like any insurance pool. Each landowner benefits from selling carbon credits, but the risk of reversal is spread out among all landowners.

This brings us to our second recommendation: qualified fuel treatments could be a standard to be eligible as an IFM project in moderate or high fire risk regions. At a minimum, IFM projects in fire prone regions could be required to include a statement about why they are not managing fuel explicitly as part of their carbon management strategy.

If changes directly to the carbon crediting standards are not feasible, policy could also target fuel treatments more directly. Our third and fourth recommendations include expansion of fuel treatments on or surrounding IFM projects. Instead of using the buffer pool to incentive fuel reductions after they occur, our third proposed approach could use grants to fund landowners more directly in the IFM program to implement fuel reductions. The proceeds generated from the Cap-and-Trade Auctions are deposited to the Greenhouse Gas Reduction Fund which supports California Climate Investments (CCI). Cal Fire's Fire Prevention Grants and the Forest Health Program are supported by the CCI, meaning IFM projects are indirectly already supporting fuel treatments in the state (California Air Resources Board, 2022). One possible approach would be proactively supporting landowners managing pools instead of an insurance reduction after the cost has been incurred.

The fourth recommendation could encourage the exiting Cal Fire fuel treatments to surround IFM projects. If maintaining elevated carbon stocks on IFM projects is still the direction market administrators and forest managers wish to pursue, strategically placed landscape fuel treatments surrounding IFM projects could decrease fire severity and promote forest recovery, helping carbon sinks (Tubbesing et al., 2019). We already observe IFM projects and Cal Fire fuel treatments in proximity, sometimes sharing even sharing project boundary boarders. The efficacy of strategically placed adjacent fuel treatments in protecting carbon within IFM projects could be further investigated for the regions IFM projects occur. If found to be effective, these co-occurring IFM projects and fuel treatments could be scaled up.

Finally, we would be remiss to not mention that more dramatic changes to carbon accounting could also shift the management occurring on IFM projects towards more proactive fire risk mitigation. For example, carbon credit-generating projects in fire prone regions might be evaluated not on total aboveground live carbon, but on the amount of carbon predicted to survive a fire disturbance. Such "avoided wildfire emissions" methodologies are already being developed by standard setting groups like Climate Action Reserve (Climate Action Reserve, 2022). Another accounting change could reevaluate how the carbon sequestered in wood products is incorporated in carbon crediting. Better incorporating woody products could better align forest and carbon management outcomes (Cabiyo et al., 2021). Further research could explore the implications of different changes in carbon accounting and whether such changes would be compatible with current carbon crediting standards.

IFM offsets are part of a climate change mitigation strategy. Climate change, fire exclusion, and antecedent disturbance contributes to the increasing presence of mega-fires as part of wildland fire regimes (Stephens et al., 2014). To disrupt this mega-fire triangle, fuel treatments or forest restoration attempt to disrupt the forest conditions created from fire exclusion and antecedent disturbance. While policymakers can support funding and reducing frictions to implementing these vegetation management projects, reducing greenhouse gas emissions could also help break this mega-fire triangle. It is striking those forests incorporated in cap-and-trade carbon markets enable others to emit greenhouse gases that will worsen the effect of climate change, while simultaneously expecting the forests that are being used to offset emissions are somehow immune to increasing threat from climate change. Worsening forest disturbances and carbon sink management are connected through processes like climate change. Policy decisions to further connect these systems can exacerbate climate change and forest management issues if tradeoffs are not addressed.

Acknowledgements
Thank you to Jared Stapp for sharing the spatial boundaries of IFM projects in California.

Tables

Table 1. List of biophysical variables used for propensity score matching, indicated with asterisk, and full list of variables included in linear regression.

Type	*Variable*	*Source*
Climate	Total winter precipitation 2016* (December 2015 – February 2016)	GRIDMET (4638.3 meters) Daily Total Precipitation (mm / day)
	Total winter precipitation 2017 (December 2016 – February 2017)	
	Total winter precipitation 2018 (December 2017 – February 2018)	
	Total winter precipitation 2019 (December 2018 – February 2019)	
	Total winter precipitation 2020 (December 2019 – February 2020)	
	Max summer temperature 2016* (June 2016 – August 2016)	GRIDMET (4638.3 meters) Max Temperature (K) (Abatzoglou, 2013)
	Max summer temperature 2017 (June 2017 – August 2017)	
	Max summer temperature 2018 (June 2018 – August 2018)	
	Max summer temperature 2019 (June 2019 – August 2019)	
	Max summer temperature 2020 (June 2020 – August 2020)	
Physical	Elevation*	SRTM (30 meters) Digital Elevation Model (meters)
	Northness* Aspect function on SRTM and then a cosine conversion to get degree facing north	SRTM-derived (degrees)
	Eastness* Aspect function on SRTM and then a sine conversion to get degree facing east	SRTM-derived (degrees)
	Slope* Slope function on SRTM	SRTM-derived (degrees) (Farr et al., 2007)
	Latitude*	Google Earth Engine (degrees)

Table 2. Average Kendall-Mann Trend in vegetation and the number of samples within each of the categories. Southern Cascade completed fuel treatments is omitted for having too few observations.

Kendall-Mann Trend	North Coast	S. Cascade
IFM	-29.10, *n = 16,577*	9.94, *n = 1,781*
Planned Fuel Treat.	15.30, *n = 409*	-59.37, *n = 104*
Active Fuel Treat.	-33.17, *n = 2,975*	-7.16, *n = 32*
Completed Fuel Treat.	-52.86, *n = 69*	*NA, n = 6*

Figures

Figure 1. Study Area Map showing the North Coast and southern Cascade region in northern California. This map includes fires between 2016-2020 and locations of IFM projects and fuel treatments.

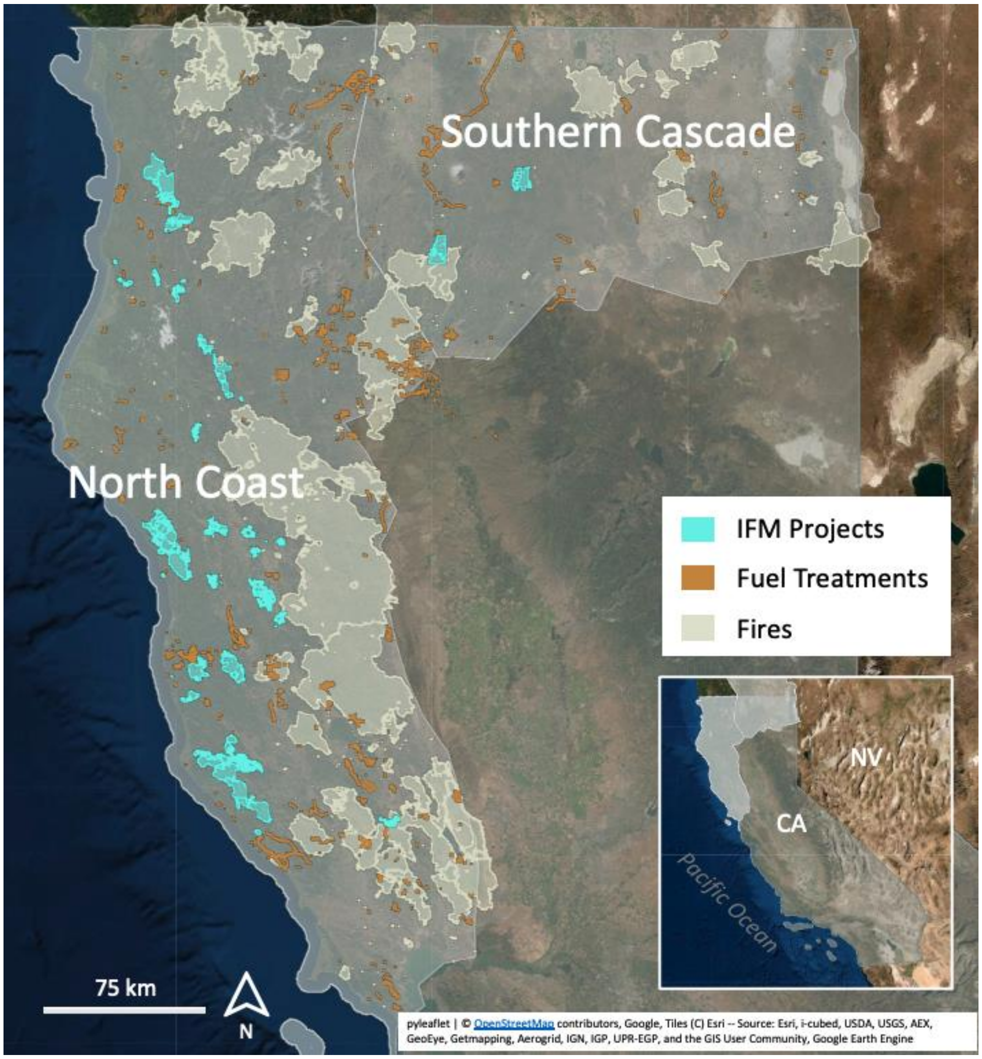


Figure 2. Boxplots of the 2016 fuel estimates for the IFM projects and fuel treatments (planned, active, and completed) across the Cascade and North Coast regions.

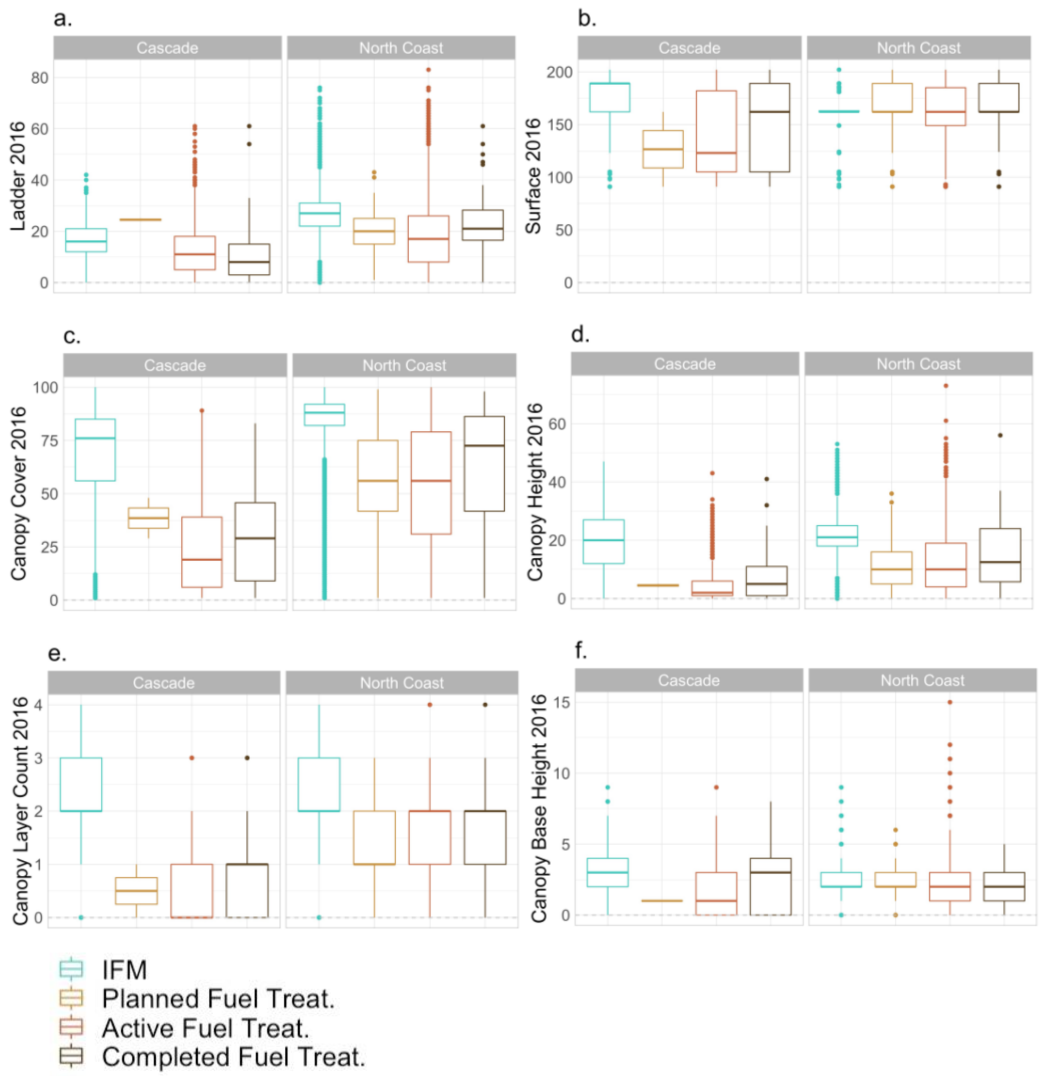


Figure 4. Estimated change between 2016 to 2020 for (a.) ladder fuel, (b.) surface fuel, (c.) canopy cover, (d.) canopy height, (e.) canopy layer count, (f.) canopy base height between offsets and fuel treatments completed prior to 2016. These graphs should the predicted change from the matched regression using the 2016 variable a-f, respectively, for each graph.

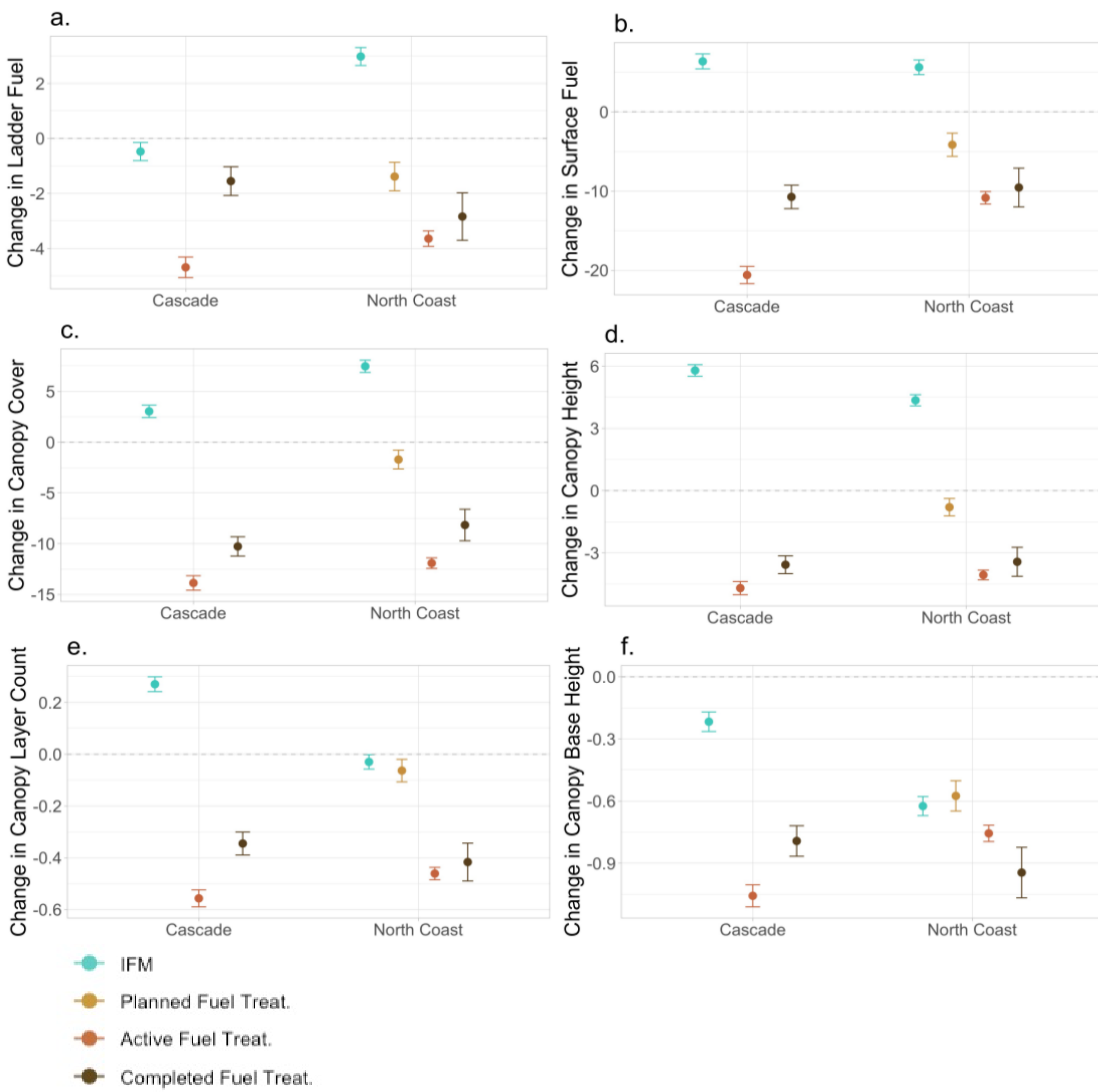
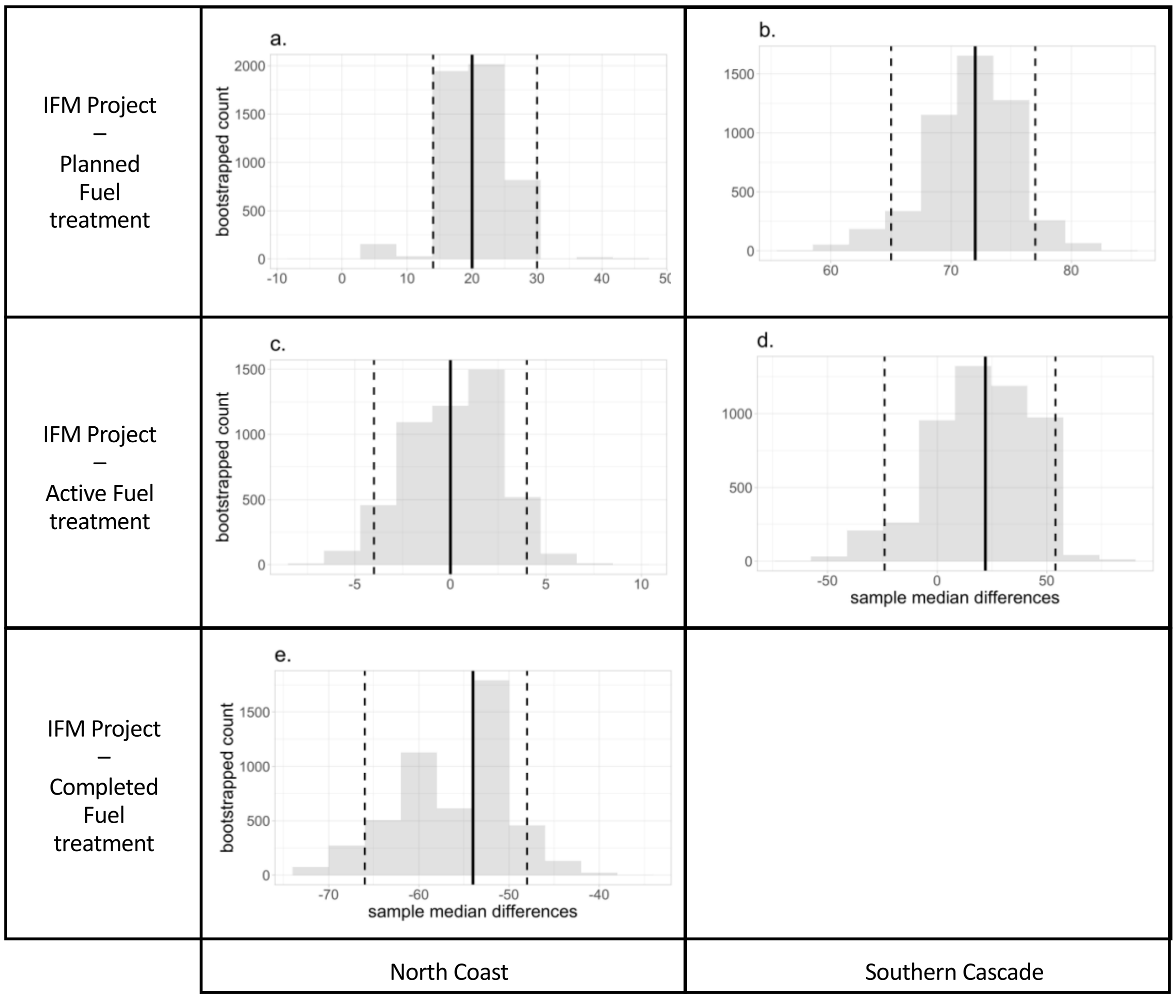


Figure 5. Bootstrapped median differences in the trend of EVI between samples from offsets and fuel treatments between the North Coast (a, c, e) and Southern Cascade (b, d) regions.



Reference

Abatzoglou, John T. 2013. "Development of Gridded Surface Meteorological Data for Ecological Applications and Modelling." International Journal of Climatology 33(1): 121–31. http://doi.wiley.com/10.1002/joc.3413 (January 30, 2020).

"ARB Offset Credit Issuance Table." California Air Resources Board. https://ww2.arb.ca.gov/our-work/programs/compliance-offset-program (April 12, 2022).

Austin, Peter C. 2011. "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." Multivariate Behavioral Research 46(3): 399–424. http://www.tandfonline.com/doi/abs/10.1080/00273171.2011.568786.

Bernal, Alexis A, Scott L Stephens, Brandon M Collins, and John J Battles. 2022. "Biomass Stocks in California's Fire-Prone Forests: Mismatch in Ecology and Policy." Environmental Research Letters 17(4): 044047.

Cabiyo, Bodie et al. 2021. "Innovative Wood Use Can Enable Carbon-Beneficial Forest Management in California." Proceedings of the National Academy of Sciences 118(49). https://pnas.org/doi/full/10.1073/pnas.2019073118.

CAL FIRE. 2022. "CalMAPPER Forest and Fuel Management Projects." https://www.arcgis.com/home/item.html?id=4a73accda8314ccd9442d365bf32f51d (April 12, 2022).

CAL FIRE (California Department of Forestry and Fire Protection). 2022. "Fuel Reduction." https://www.fire.ca.gov/fuels-reduction/ (May 1, 2022).

California Air Resources Board. 2020. Greenhouse Gas Emissions of Contemporary Wildfire, Prescribed Fire, and Forest Management Activities. https://ww3.arb.ca.gov/cc/inventory/pubs/ca_ghg_wildfire_forestmanagement.pdf.

———. 2022. "California Climate Investments 2021 Annual Report: Cap-and-Trade Auction Proceeds." (April). https://ww2.arb.ca.gov/sites/default/files/auction-proceeds/cci_annual_report_2022.pdf.

California Forest Observatory. 2020. "Data Description: Vegetation Structure and Fuels." Salo Sciences, Inc (April). https://salo.ai/assets/pdf/Forest-Observatory-Data-Description.pdf (April 12, 2022).

CARB Air Resources Board. 2011. Compliance Offset Protocol U.S. Forest Projects. http://www.arb.ca.gov/regact/2014/capandtrade14/ctusforestprojectsprotocol.pdf.

Climate Action Reserve. 2022. Avoided Wildfire Emissions Forecast Methodology in Development. https://climateforward.org/program/methodologies/avoided-wildfire-emissions/.

Collins, Brandon M. et al. 2017. "Impacts of Different Land Management Histories on Forest Change." Ecological Applications 27(8): 2475–86.

Coop, Jonathan D. et al. 2020. "Wildfire-Driven Forest Conversion in Western North American Landscapes." BioScience 70(8): 659–73.

Falk, Donald A et al. 2022. "Mechanisms of Forest Resilience." Forest Ecology and Management 512(November 2021): 120129. https://doi.org/10.1016/j.foreco.2022.120129.

Farr, Tom G. et al. 2007. "The Shuttle Radar Topography Mission." Reviews of Geophysics 45(2): RG2004. http://doi.wiley.com/10.1029/2005RG000183.

Foster, Daniel E. et al. 2020. "Potential Wildfire and Carbon Stability in Frequent-Fire Forests in the Sierra Nevada: Trade-Offs from a Long-Term Study." Ecosphere 11(8).

Gonzalez, Patrick et al. 2015. "Aboveground Live Carbon Stock Changes of California Wildland Ecosystems, 2001-2010." Forest Ecology and Management 348: 68–77. http://dx.doi.org/10.1016/j.foreco.2015.03.040.

Hurteau, Matthew D., and Matthew L. Brooks. 2011. "Short- and Long-Term Effects of Fire on Carbon in US Dry Temperate Forest Systems." BioScience 61(2): 139–46.
Hurteau, Matthew D., George W. Koch, and Bruce A. Hungate. 2008. "Carbon Protection and Fire Risk Reduction: Toward a Full Accounting of Forest Carbon Offsets." Frontiers in Ecology and the Environment 6(9): 493–98.
Jarron, Lukas R. et al. 2020. "Detection of Sub-Canopy Forest Structure Using Airborne LiDAR." Remote Sensing of Environment 244: 111770. https://linkinghub.elsevier.com/retrieve/pii/S0034425720301401.
Johnston, Matthew G., and Christine Faulkner. 2021. "A Bootstrap Approach Is a Superior Statistical Method for the Comparison of Non-Normal Data with Differing Variances." New Phytologist 230(1): 23–26.
Knight, Clarke A. et al. 2022. "Accurate Tracking of Forest Activity Key to Multi-Jurisdictional Management Goals: A Case Study in California." Journal of Environmental Management 302(PB): 114083. https://doi.org/10.1016/j.jenvman.2021.114083.
LANDFIRE. 2001. "Existing Vegetation Type Layer." http://www.landfire/viewer.
Liang, Shuang, Matthew D. Hurteau, and Anthony L. Westerling. 2018. "Large-Scale Restoration Increases Carbon Stability under Projected Climate and Wildfire Regimes." Frontiers in Ecology and the Environment 16(4): 207–12.
Lutz, James A. et al. 2018. "Global Importance of Large-diameter Trees." Global Ecology and Biogeography 27(7): 849–64. https://onlinelibrary.wiley.com/doi/10.1111/geb.12747.
McEvoy, Daniel J. et al. 2020. "Projected Changes in Reference Evapotranspiration in California and Nevada: Implications for Drought and Wildland Fire Danger." Earth's Future 8(11). https://onlinelibrary.wiley.com/doi/10.1029/2020EF001736.
McIntyre, Patrick J. et al. 2015. "Twentieth-Century Shifts in Forest Structure in California: Denser Forests, Smaller Trees, and Increased Dominance of Oaks." Proceedings of the National Academy of Sciences of the United States of America 112(5): 1458–63.
Mildrexler, David J. et al. 2020. "Large Trees Dominate Carbon Storage in Forests East of the Cascade Crest in the United States Pacific Northwest." Frontiers in Forests and Global Change 3. https://www.frontiersin.org/articles/10.3389/ffgc.2020.594274/full.
Mutlu, M., Popescu, S., Stripling, C., and Spencer, T. 2008. "Mapping Surface Fuel Models Using Lidar and Multispectral Data Fusion for Fire Behavior." Remote Sensing of Environment 112(1): 274–85. https://linkinghub.elsevier.com/retrieve/pii/S0034425707001988.
Nicholas Clinton. "Non-Parametric Trend Analysis." https://developers.google.com/earth-engine/tutorials/community/nonparametric-trends (April 12, 2022).
North, Malcolm, Matthew Hurteau, and James Innes. 2009. "Fire Suppression and Fuels Treatment Effects on Mixed-Conifer Carbon Stocks and Emissions." Ecological Applications 19(6): 1385–96. http://doi.wiley.com/10.1890/08-1173.1.
North, Malcolm P. et al. 2022. "Operational Resilience in Western US Frequent-Fire Forests." Forest Ecology and Management 507(December 2021).
North, Malcolm P., and Matthew D. Hurteau. 2011. "High-Severity Wildfire Effects on Carbon Stocks and Emissions in Fuels Treated and Untreated Forest." Forest Ecology and Management 261(6): 1115–20. http://dx.doi.org/10.1016/j.foreco.2010.12.039.
Ogaya, Romà, Adrià Barbeta, Corina Başnou, and Josep Peñuelas. 2015. "Satellite Data as Indicators of Tree Biomass Growth and Forest Dieback in a Mediterranean Holm Oak Forest." Annals of Forest Science 72(1): 135–44. http://link.springer.com/10.1007/s13595-014-0408-y.

Projects, U S Forest. Compliance Offset Protocol AIR RESOURCES BOARD. https://www.arb.ca.gov/cc/capandtrade/protocols/usforest/forestprotocol2015.pdf (May 7, 2019).
Roberts, D. A. et al. 2006. "Evaluation of Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and Moderate Resolution Imaging Spectrometer (MODIS) Measures of Live Fuel Moisture and Fuel Condition in a Shrubland Ecosystem in Southern California." Journal of Geophysical Research: Biogeosciences 111(G4). http://doi.wiley.com/10.1029/2005JG000113.
Sims, Daniel A. et al. 2006. "On the Use of MODIS EVI to Assess Gross Primary Productivity of North American Ecosystems." Journal of Geophysical Research: Biogeosciences 111(G4). http://doi.wiley.com/10.1029/2006JG000162.
State Board of Forestry and Fire Protection, and CAL FIRE (California Department of Forestry and Fire Protection). 2018. "2018 Strategic Fire Plan for California." : 36. https://osfm.fire.ca.gov/media/5590/2018-strategic-fire-plan-approved-08_22_18.pdf.
State of California. 2017. "Assembly Bill No. 398." (398): 1–21. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201720180AB398.
Stapp, J., Nolte, C., Potts, M., Baumann, M., Haya, B., Butsic, V. 2022. Early indications of effectiveness in California's forest offset program. In Press.
Stephens, Scott L. et al. 2014. "Temperate and Boreal Forest Mega-Fires: Characteristics and Challenges." Frontiers in Ecology and the Environment 12(2): 115–22.
———. 2018. "Drought, Tree Mortality, and Wildfire in Forests Adapted to Frequent Fire." BioScience 68(2): 77–88.
———. 2020. "Fire and Climate Change: Conserving Seasonally Dry Forests Is Still Possible." Frontiers in Ecology and the Environment 18(6): 354–60.
———. 2021. "Forest Restoration and Fuels Reduction: Convergent or Divergent?" BioScience 71(1): 85–101.
Stephenson, N. L. et al. 2014. "Rate of Tree Carbon Accumulation Increases Continuously with Tree Size." Nature 507(7490): 90–93. http://www.nature.com/articles/nature12914.
Su, Yanjun et al. 2016. "Forest Fuel Treatment Detection Using Multi-Temporal Airborne Lidar Data and High-Resolution Aerial Imagery: A Case Study in the Sierra Nevada Mountains, California." International Journal of Remote Sensing 37(14): 3322–45. http://www.tandfonline.com/doi/full/10.1080/01431161.2016.1196842.
Tubbesing, Carmen L. et al. 2019. "Strategically Placed Landscape Fuel Treatments Decrease Fire Severity and Promote Recovery in the Northern Sierra Nevada." Forest Ecology and Management 436: 45–55. https://linkinghub.elsevier.com/retrieve/pii/S0378112718320462.
Westerling, A. L. et al. 2011. "Climate Change and Growth Scenarios for California Wildfire." Climatic Change 109(SUPPL. 1): 445–63.
Wiedinmyer, Christine, and Matthew D. Hurteau. 2010. "Prescribed Fire As a Means of Reducing Forest Carbon Emissions in the Western United States." Environmental Science & Technology 44(6): 1926–32. https://pubs.acs.org/doi/10.1021/es902455e.
Williams, A. Park et al. 2019. "Observed Impacts of Anthropogenic Climate Change on Wildfire in California." Earth's Future 7(8): 892–910.

Conclusion

This work is motivated by and in response to the increasing stressors on human and ecological systems from climate change. I used analytical frameworks from land use sciences, data processing techniques from remote sensing, and data analysis techniques from econometrics to explore topics relevant to policy and management decisions in California. Addressing the mounting management challenges arising from climate change and the legacies of past management will require incorporating existing tools in innovative ways and closely reviewing new policy and management programs to measure effectiveness in results.

This research focused on improving the methods by which we test assumptions in technical management questions. The three chapters presented are unified by the types of analytical approaches used to set up hypothesis tests in natural experiments.

In the first chapter, I use inferential statistical analysis to understand how irrigated landscapes in California may provide benefits to adjacent communities during a fire. I found that golf courses, an example of irrigated landscapes, reduce fire severity by 49% relative to otherwise similar landscapes. This fire severity reduction effect was variable by region, with the effect most pronounced in the North Bay region. Using the perimeters of golf courses, large regional parks, and airports, I also find that golf courses are as effective in limiting fire spread as more hardscaped landcovers like airports. Increasingly devastating fire seasons are an example of how interactions between histories of ecosystem management, low-density urban expansion, and climate change create complex management decisions. Fire management near housing and built infrastructure in the Wildland Urban Interface (WUI) will continue to be an economically and socially charged topic as the threat of fire increases. Frequently, discussion surrounding vegetation management around homes focuses on altering exposure to risk by removal of vegetation. Along with vegetation removal, managing vegetation flammability can also be part of the toolkit for communities. Finally, irrigating landscapes in California is a resource decision that comes at the expense of other sustainability or water management goals and may not be appropriate in all contexts. To help inform decision makers on when additional resources could be used to maintain a green landscape, methods like those demonstrated in this dissertation will be useful for weighing tradeoffs and making context appropriate resource decisions.

In the second chapter, I produced predictions of vegetation growth using a vegetation growth simulator and used these outputs to perform a model validation of predicted carbon growth. Using equivalency tests, we demonstrate that FVS modeled outputs are not capturing observed carbon growth rates. FVS is an imperfect, yet widely relied on, model for understanding forest growth dynamics, especially for public forests. The overpredictions we observed from FVS could lead to management decisions that assume carbon will accumulate faster in forest stands than is realistic. One outcome of this overestimation is that it could possibly delay more active management interventions that could remove some within stand competition, increasing the carbon growth rate. We find that the overprediction is not random: we can systematically account for model error by including variables about the forest type (e.g., public or private, younger or older forests). We also found trees growing on the lowest site class, or areas of the lowest forest productivity, had greater errors in FVS outputs. Historically, forestry has been dominated by timber management that would have focused on the management of the highest site classes. Using the equations developed by foresters for forest ecology research will not always capture the ecosystem we now endeavor to manage. Identifying these issues is a start, but those of us in the field should work to improve what we measure and how we measure it to improve our ability to achieve the forest ecology outcomes we care about.

In the third chapter, I explore how Improved Forest Management (IFM) projects in California consider fuel reduction as part of their management between 2016-2020. The California Air Resources Board compliance market incorporates forests in the cap-and-trade program as offsets. I use remote sensing techniques to gauge how these projects are managing during the first decade of implementation. This assessment is crucial, because the projects are not required to submit information about smaller fuel management projects that maintain carbon above baseline levels. Furthermore, IFM projects frequently achieve increased carbon levels by deferring harvest by increasing rotation lengths, switching from even to uneven age management, or reducing harvest intensity. However, the demands of managing California forests will differ from those in other regions. For example, switching from uneven age management and reducing harvest intensity may be appropriate for a pine plantation in the southeast to increase aboveground forest carbon, but forests in California are dealing with another management crisis and have different forest practice rules. For forests in California, the aboveground tree density has increased dramatically over the past century due to fire suppression and the legacy of forest practices removing the largest, carbon dense trees. One standard that is flexible enough to be appropriate for the different management histories and needs of these geographies is unlikely to work for diverse forest needs.

Throughout my work, I focus on mitigating fire risk because the implications of historic management are clear. Research shows that not changing these forest structures will have negative carbon and forest ecology consequences in California. Increased fuel reductions and forest restoration are viable management solutions that can promote forest resilience. In the final chapter, I combine multiple remotely sensed datasets to understand how IFM projects in California are managing their fuels. Results demonstrate that projects have high levels of biomass that are not being reduced. Between 2017-2021, IFM projects in and outside of California are burning. This problem will likely continue as landscape fire risks are expected to increase for most regions in California. Projects claiming to improve forest management in these regions may consider mitigating fire risk or incorporating management that will promote forest resilience.

This work is an attempt to answer some of these questions regarding forests, fire, and risk. Undoubtedly, future work will be able to refine this research by incorporating better data as it becomes available. I look forward to more researchers applying these analytical tools to other ecosystems and to the same ecosystems over different timespans. Each of these papers was analyzed using primarily open data sources and analysis techniques that lend themselves to be reproduced. Further work could look at landscape features in other regions, inspect how errors in FVS have changed over time, or compare IFM projects and fuel treatments to other types of forest management techniques. These works are threads woven together by the way they apply methods to better quantify, qualify, and evaluate land management practices. The results of this work are then rife for application by land managers, fire system modelers, city planners, and policy makers.

Chapter 3 Appendix

S. Figure 1. Errors for California Forest Observatory fuel dataset. No estimates are produced for Surface fuel because it is a unitless estimate.

Fuels Error

Fuel Estimate	*unit*	*r^2*	*MAE*	*RSME*
ladder fuel	%	0.78	2.8	4.79
surface fuel	unitless	NA	NA	NA
canopy cover	%	0.91	7	12.2
canopy height	meters	0.86	1.97	3.61
canopy layer count	count	0.765	0.33	0.49
canopy base height	meters	0.70	1.47	2.53

S. Figure 2. Mean Changes in Fuels on Fuel treatments as part of data validation.

Mean Change in Fuels on Fuel Treatments

Fuel Estimate	*Planned*	*Active*	*Completed*
ladder fuel	0.05	-1.42	0.79
surface fuel	1.51	-5.56	-0.23
canopy cover	2.68	-5.56	-1.55
canopy height	2.19	-0.61	0.51
canopy layer count	0.25	-0.16	0.10
canopy base height	-0.38	-0.55	-0.71

S. Figure 3. Plots of how matching affected the mean differences between the matching variables.