Ecological Applications, 31(3), 2021, e02283
© 2020 by the Ecological Society of America

Matching methods to quantify wildfire effects on forest carbon mass in the U.S. Pacific Northwest

Hyeyoung Woo,[1,3] Bianca N. I. Eskelson,[1] and Vicente J. Monleon[2]

[1]*Department of Forest Resources Management, Faculty of Forestry, The University of British Columbia, 2424 Main Mall, Vancouver, British Columbia V6T 1Z4 Canada*
[2]*USDA Forest Service, Pacific Northwest Research Station, 3200 SW Jefferson Way, Corvallis, Oregon 97331 USA*

Citation: Woo, H., B. N. I. Eskelson, and V. J. Monleon. 2021. Matching methods to quantify wildfire effects on forest carbon mass in the U.S. Pacific Northwest. Ecological Applications 31(3):e02283. 10.1002/eap.2283

Abstract. Forest wildfires consume and redistribute carbon within forest carbon pools. Because the incidence of wildfires is unpredictable, quantifying wildfire effects is challenging due to the lack of prefire data or controls from experiments over a large landscape. We explored a quasi-experimental method, propensity score matching, to estimate wildfire effects on aboveground forest woody carbon mass in Washington and Oregon, United States. Observational data, including national forest inventory plot measurements and satellite imagery metrics, were utilized to obtain a control set of unburned plots that are comparable to burned plots in terms of environmental conditions as well as spatial locations. Three matching methods were implemented: propensity score matching (PSM), spatial matching (SM), and distance-adjusted propensity score matching (DAPSM). We investigated if propensity score matching with and without spatial adjustment led to different outcomes in terms of (1) balance in covariate distributions between burned and control plots, (2) mean carbon mass obtained from the selected control plots compared to burned and all unburned plots, and (3) estimates of wildfire effects by burn severity. We found that PSM and SM, which use only the environmental covariate set or the spatial distance for estimating propensity scores, respectively, did not appear to produce a comparable set of control plots in terms of the estimated propensity scores and the outcomes of mean carbon mass. DAPSM was the preferred method both in balancing the observed covariates and in dealing with unobservable confounding variables through spatial adjustment. The average wildfire effects estimated by DAPSM showed clear evidence of redistribution of carbon among aboveground woody pools, from live to dead trees, but the consumption of total woody carbon by wildfire was not substantial. Only moderate burn severity led to significant reduction of total woody carbon mass across Washington and Oregon forests (64% of control plots remained on average). This study provides an applied example of a quasi-experimental approach to quantify the effects of a natural disturbance for which experimental settings are unavailable. The study results suggest that incorporating spatial information in addition to environmental covariates would yield a comparable set of control plots for wildfire effects quantification.

Key words: burn severity; cause-and-effect estimation; distance adjustment; forest inventory data; forest woody carbon; observational study; propensity score; quasi-experiment; unobservable confounding variable.

Introduction

Forest wildfires can lead to substantial carbon loss by the immediate removal of forest fuels (Dixon et al. 1994) as well as redistribution of carbon among forest pools (Keith et al. 2014). The change in forest carbon due to wildfire affects regional carbon dynamics that prompt ecological and climatic changes on a landscape (Flannigan et al. 2009). Thus, wildfire effects need to be identified at regional scales as a part of carbon management strategies over a region (Hudiburg et al. 2011). In the U.S. Pacific Northwest, where forests are abundant and exposed to high risk of wildfires (Hurteau and North 2010), forest managers and biologists have been trying to quantify wildfire effects on forest carbon to assess the regional carbon cycle (Mitchell et al. 2009, Hudiburg et al. 2011). Combined with increasing risk of wildfires due to drought and high fuel density in Pacific Northwest forests (Law and Waring 2015), wide variation in vegetation and burn severity on the landscape requires consistent monitoring and a refined quantification process for wildfire effects on carbon mass (Mitchell et al. 2009, Wimberly and Liu 2014).

Manuscript received 24 June 2020; revised 1 September 2020; accepted 5 October 2020. Corresponding Editor: Bradford P. Wilcox.
[3]E-mail: hywoo@mail.ubc.ca

Check for updates

Estimating wildfire effects on forest carbon, however, is challenging because areas that burn typically lack comparable controls or prefire data (Johnson et al. 2007). Wildfires are considered as stochastic events, and the probability that a forest burns depends on fuel bed composition, climate, and topography (Pew and Larsen 2001, Ager et al. 2007, Parisien et al. 2012) as well as the amount of carbon from forest biomass (Ottmar 2014). Therefore, simple comparisons of carbon mass between burned and unburned forests may result in biased estimates of wildfire effects on forest carbon owing to the pre-existing differences in the environmental conditions among burned and unburned forests (Meigs et al. 2009). When the wildfire effect quantification is conducted at a regional scale with variability in environmental conditions, it is impractical to obtain measurements of controls or prefire data over a large spatial and temporal coverage (Kremens et al. 2010). Instead of comparing burned and unburned forests, wildfire effects on forest carbon have been investigated through statistical models (e.g., Mitchell et al. 2009), simulations (e.g., Amiro et al. 2001), case studies of individual fires (e.g., Campbell et al. 2007, Johnson et al. 2007), and prescribed fires (Eskelson and Monleon 2018). The assessment of wildfire effects from case studies for high-severity fires on a particular area, however, may not be capture all the variability need for landscape-level estimations (Zheng et al. 2004), nor to the majority of wildfires with low to moderate severities (Eskelson and Monleon 2018).

The preexisting difference in forest conditions between burned and unburned forest stands that induces different wildfire probability is referred to as selection bias (Caliendo and Kopeinig 2008). In a randomized experiment, the selection bias is eliminated by the random assignment of a treatment (Heckman et al. 1996). Wildfire effects, however, can only be assessed by observational study data due to the unpredictability of wildfires and the spatiotemporal scales that do not allow the implementation of randomized experiments. Butsic et al. (2017) and Larsen et al. (2019) suggested quasi-experimental methods to analyze causal relationships from observational data for ecological problems. Propensity score matching (Rosenbaum and Rubin 1983) is one of the quasi-experimental methods that balances the probability of treatment between treated and control groups. While propensity score matching is widely used for observational studies in other fields of science such as health (e.g., D’Agostino 1998, Hernán et al. 2004), social policy program evaluation (e.g., Caliendo and Kopeinig 2008, Keele et al. 2015), and econometrics (e.g., Lechner 2002), the method is not commonly applied to quantify effects of natural disturbances on forests. Only a few examples of employing propensity scores are available in wildfire research (e.g., Butsic et al. 2017), and those are mostly related to evaluating management policies such as wildfire mitigation programs (Butry 2009) or wildfire prevention education (e.g., Prestemon et al. 2010).

In terms of wildfire effect analysis, the presence of wildfire is equivalent to the treatment in an experiment and the change in carbon mass is the outcome of interest (Ager et al. 2010, Butsic et al. 2017). Given that the treatment units are forest inventory plots, the propensity score is the probability of a plot being burned by wildfire. By selecting a set of unburned plots with a similar propensity score distribution to the set of burned plots, the selection bias due to the pre-existing difference between burned and unburned plots can be eliminated, conditional on the propensity scores (Rosenbaum and Rubin 1983). Various environmental covariates that account for fuels, climate, and topography can be obtained from observational data such as forest inventory plot measurements and satellite imagery metrics to estimate propensity scores (Butsic et al. 2017). Lately, there have been efforts to incorporate a measure of spatial distance in estimating propensity scores to account for covariates that are unobservable from available data (Keele et al. 2015, Papadogeorgou et al. 2019). Ecologically, forests that are spatially close to each other share similar environmental characteristics (Legendre 1993). Given that geographical proximity is often used as a factor to describe similarity among ecological structures or as a confounding variable that was not observed in the dataset (Dray et al. 2006), it may be beneficial to utilize spatial distance in estimating propensity scores for forest inventory plot data that are recorded with locations. Propensity score matching to solve ecological problems has been suggested only recently using a simulated dataset (e.g., Butsic et al. 2017), and mostly focused on the intuitional framework rather than the technical details (Larsen et al. 2019). Use of actual forest measurements to determine empirical effect estimates will elucidate the applicability of matching methods to estimate wildfire effects with observed environmental variables and distance measures.

In this study, we used matching methods in conjunction with national forest inventory data to quantify regional wildfire effects on aboveground woody forest carbon mass in Washington and Oregon, USA. The objectives were both methodological as well as applied. We implemented three matching methods: (1) propensity score matching, (2) spatial matching, and (3) distance-adjusted propensity score matching. The first objective was to diagnose the balance achieved between burned and control plots under the three matching methods, to address how environmental covariates and spatial distance can contribute to the removal of the selection bias between burned and unburned plots. The second objective was to compute the carbon masses of several forest pools, live trees, standing dead trees, and coarse woody materials, from each set of control plots to show how the three matching methods may provide different outcomes when answering an ecological question. The third objective was to estimate the difference in carbon mass by burn severity obtained through the three matching methods. The estimates of difference in the aboveground woody carbon mass between burned and control plots provided a regional assessment of average wildfire effect over the heterogeneous landscape of the U.S. Pacific Northwest.

Given the remote-sensing data and the extensive national forest inventory data with detailed environmental conditions and spatial locations, we discussed the applicability of propensity score matching as a quasi-experimental method to wildfire effect analysis with and without the inclusion of a spatial distance measure. This study highlights how the theoretical framework of matching methods can be applied to real-world data sets to solve ecological problems and how users can assess the performance of different matching approaches and choose the method that may provide the most reliable inferences with regards to the ecological question asked. Because matching methods have rarely been used for ecological applications, our comprehensive discussion of several approaches in a real-world application provides an important contribution to users of matching methods.

Study Area and Data

Study area

The study area encompasses the states of Washington and Oregon, USA (Figure 1). The landscape is largely distinguished by coastal temperate rainforests and drier forests west and east of the Cascade Range, respectively. Evergreen conifers such as Douglas-fir (*Pseudotsuga menziesii*), western hemlock (*Tsuga heterophylla*), Pacific silver fir (*Abies amabilis*), and ponderosa pine (*Pinus ponderosa*) are dominant over much of the landscape. The proportion of public forestland tends to be high: more than 50% of Washington and 60% of Oregon (Donnegan et al. 2008, Campbell et al. 2010). Fire regimes in the study area vary substantially, with fire return intervals ranging from 15 to 937 yr depending on the regional climate variation and forest types (Perry et al. 2011). The western part of the region tends to have less frequent, stand-replacing fires than the dry interior, which exhibits a variety of burn severities from low-severity fires in ponderosa pine forests to high-severity fires in subalpine, mixed conifer forests (Reilly et al. 2017).

Forest Inventory and Analysis (FIA) data

Since 1999, the Forest Inventory and Analysis (FIA) program has provided comprehensive information of nationwide forest plots based on an annual data collection protocol (Smith 2002). Following the annual inventory protocol that was applied to Washington and Oregon from 2001 and 2002, respectively, all forested lands are systematically sampled with one FIA plot for approximately every 2,500 ha (6,000 acres; Bechtold and Patterson 2005).

One FIA plot comprises measurements from four fixed circular 7.3 m radius subplots with plot centers 36 m apart (Appendix S1: Fig. S1). Year of measurement and topographical attributes (i.e., elevation, slope, and aspect), and plot location (i.e., latitude and longitude) are recorded. Measurements of standing trees and downed woody materials are collected on each plot with a 10-yr remeasurement cycle.

From 2001 to 2016, 16,239 plots with forested conditions (i.e., at least 10% potential cover by live trees; Bechtold and Patterson 2005) were measured by the FIA program in Washington and Oregon, USA. Of these, 8,184 plots were measured twice with a 10-yr remeasurement cycle in our data, producing separate measurements for the same plot location. Therefore, each plot can have multiple plot measurements. We considered only plots that had a minimum of 25% forested area, which resulted in a total of 23,150 plot measurements including remeasurements. The topography and land cover variables—physiographic class (PHYSCL), land ownership (OWNS), elevation (ELEV, m), slope (SLOPE, degrees), and cosine transformation of aspect (cosASP; Weiskittel et al. 2011)—were obtained from FIA data at the plot level and included as covariates to perform matching (Table 1). For ease of analysis, the categorical variables, physiographic class (PHYSCL) and land ownership (OWN), were classified into three (mesic, xeric, and hydric) and two (public and private) categories, respectively. Because a plot may straddle several values of the categorical variables, we used the value that covered the greatest proportion of the area. The proportion of forested area (PROPforest) was also calculated for each plot observation.

Fire data

Monitoring Trends in Burn Severity (MTBS) data from 1984 to 2016 (Fig. 1) were utilized to identify the presence of wildfires, burn severities, and date of wildfire for each plot (data *available online*).[4] National MTBS burn severity mosaics are provided as tagged image file format (TIFF), classifying burn severities into four major categories (e.g., low-low, low, moderate, and high) according to the differenced Normalized Burn Ratio (dNBR) from Landsat satellite images with 30 m of spatial resolution (Eidenshink et al. 2007). The MTBS data used in this study include records of over 1,000 fires in Washington and Oregon, USA, since 1984, which were approximately 40 ha in size or larger.

Burned plots were identified as FIA plot measurements that experienced any wildfire within 5 yr prior to the field measurement. In contrast, plot measurements that experienced no fire within at least 10 yr prior to the measurement were identified as unburned plots. Plots that had fires more than once within 5 yr were dropped to exclude the effect of reburn from the analysis. Any prefire measurements of burned plots found within the unburned group were also discarded to completely

[4] http://www.mtbs.gov/

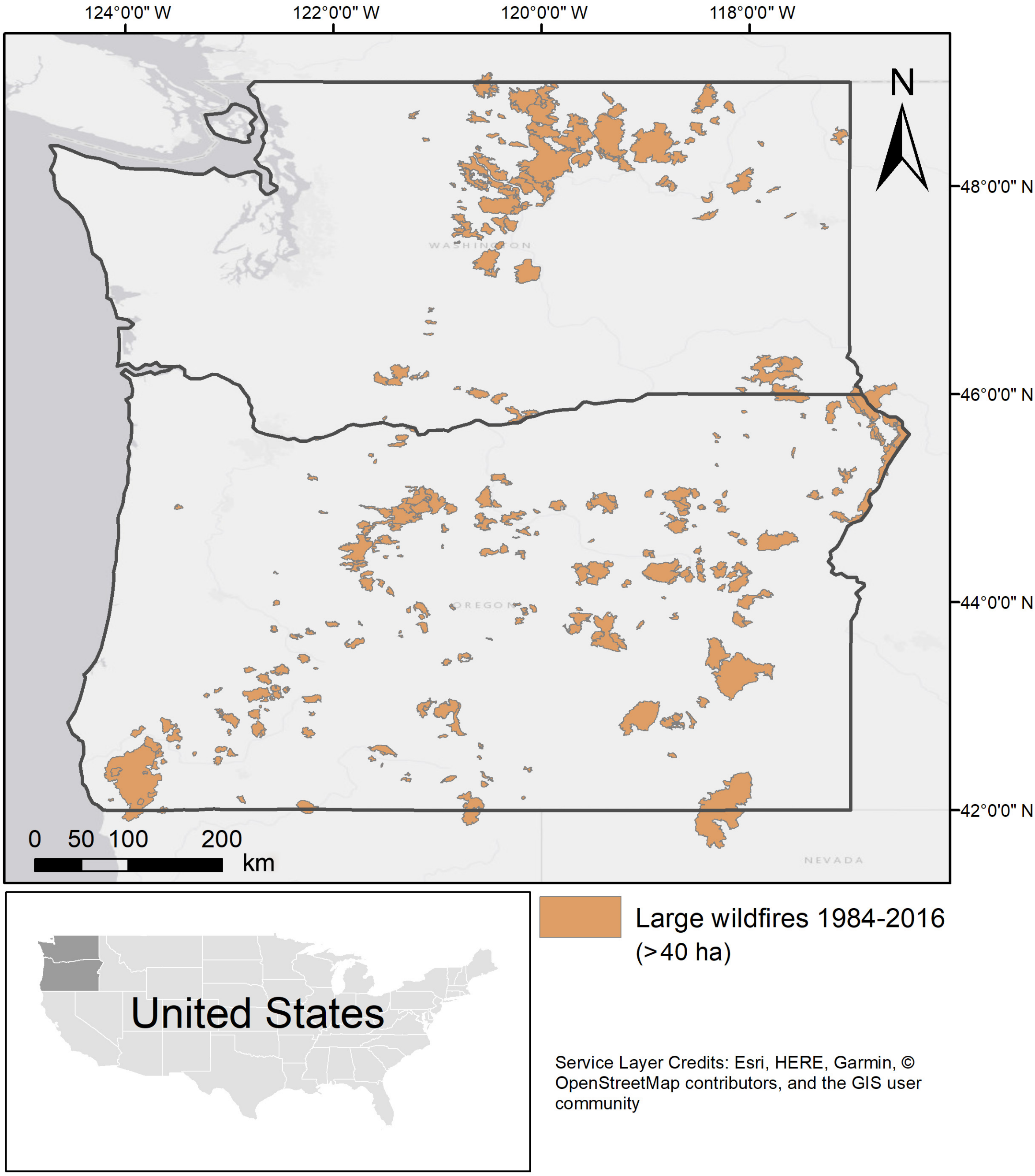

FIG. 1. Study area (Washington and Oregon, USA) and spatial distribution of large wildfires (e.g., >40 ha) from 1984 to 2016.

separate burned and unburned groups during the implementation of matching. A total of 611 burned plots and 22,539 unburned plot measurements were identified (Table 2). Approximately 70% of the burned plots were measured before 2011, thus those plot locations retained a single measurement in our data due to the 10-yr remeasurement cycle. For the 611 burned plot measurements, the burn severity was assigned as the most frequent severity around the selected plot location, using the values on the plot center and eight adjacent 30-m pixels from the MTBS burn-severity maps (Appendix S1: Fig. S1).

Climate and remotely sensed data

Thirty-year normals of PRISM (Daly et al. 2008) spatial climate data from 1981 to 2010 were obtained with 30-m resolution and spatially overlaid with each of the

Table 1. Summary statistics of topographic and land cover variables obtained from FIA plot data.

Variable and description	Burned (n = 611)	Unburned (n = 22,539)
Topography		
ELEV		
Elevation (m)	1,241.9 (439.63)	1,073.78 (534.08)
SLOPE		
Slope (degrees)	37.16 (23.88)	30.41 (24.28)
cosASP		
Aspect (degrees azimuth), cosine value	0.03 (0.72)	0.11 (0.74)
Land cover		
PROPforest		
Proportion of forested area	0.95 (0.14)	0.95 (0.14)
OWN		
Owner group code (categorical)†	Number of plots (%)	
1, Public lands	552 (90.34)	17,610 (77.76)
2, Private lands	59 (9.66)	5,037 (22.24)
PHYSCL		
Physiographic class code (categorical)‡		
1, Xeric sites	292 (47.79)	5,239 (23.13)
2, Mesic sites	319 (52.21)	17,279 (76.30)
3, Hydric sites	0 (0)	129 (0.57)

Notes: For continuous variables, the mean with the standard deviation in parentheses is provided. For categorical variables, the number of plots in each category is specified with the proportions in parentheses.

†Public lands in OWN are forests owned by the federal, state, or local governments.

‡Xeric sites are low or deficient in available moisture (e.g., dry tops, slopes, or deep sand), while mesic sites refer to sites with moderate but adequate available moisture (e.g., flatwoods, rolling uplands, or flood plains and bottomlands). Hydric sites exhibit abundant or overabundant moisture all year, including swamps, ponds, drains, or wet pocosins. See FIA user manual (https://www.fia.fs.fed.us/library/database-documentation/index.php) for a detailed description of the variables.

Table 2. Number of FIA plots in each burn-severity class.

Burn severity	Number of plots	Group
0 (unburned)	22,539	unburned
1 (low-low)	120	burned
2 (low)	201	burned
3 (moderate)	145	burned
4 (high)	145	burned

FIA plot locations. The data contain information on precipitation, temperature, vapor pressure, and moisture stress, which are closely linked to wildfire regimes over large areas (Parisien and Moritz 2009). Climate variables found to be highly correlated with each other (e.g., $\hat{\rho} > 0.8$ in inter-variable Pearson's correlation coefficients) were dropped to avoid multicollinearity among the covariates (Adelson 2013). As a result, six variables were selected as covariates to estimate propensity scores (Table 3): annual precipitation (annpre), mean annual temperature (anntmp), difference in precipitation between August and December (cvpre), difference in temperature between August and December (diftmp), mean summer vapor pressure deficit (smrnmvpd), and growing season moisture stress (smrtp).

Vegetation indices such as the normalized difference vegetation index (NDVI) and tasseled cap metrics (TC; Crist and Cicone 1984) are considered to be correlated with biophysical attributes and aboveground biomass (Duane et al. 2010, Pflugmacher et al. 2012). Using 30-m resolution, TC brightness (TC1), greenness (TC2), and wetness (TC3) indices as well as NDVI in the middle of the growing season were derived from Landsat spectral bands from 1984 to 2016 for the study area. For each of the plot locations, the remotely sensed metrics were averaged using nine adjacent pixel values (Appendix S1: Fig. S1). Burned plot measurements were associated with Landsat metrics from the year prior to the fire. Unburned plot measurements were associated with metrics of the measured year. For example, a plot measurement that experienced wildfire in 2012 and was measured in 2014 was classified as a burned plot and was associated with 2011 Landsat metrics. On the other hand, a plot measurement that burned in 2002 and was measured in 2014 was classified as unburned plot, and associated with 2014 Landsat metrics. All climate and remotely sensed metrics were obtained from the Landscape Ecology, Modeling, Mapping, and Analysis (LEMMA) team (data *available online*).[5]

Methods

Matching burned and unburned plots

Propensity score matching chooses a set of control units out of all the untreated units based on similarity of propensity scores. The propensity score is the probability that a subject is assigned the treatment, given a set of observed covariates (Rosenbaum and Rubin 1983). If the propensity score distributions are similar between the treated and the control groups, the selection bias due to the pre-existing differences between treated and untreated groups no longer exists conditionally on the covariates (Rosenbaum and Rubin 1983). The distribution of probabilities receiving the treatment becomes uniform between treated and untreated groups, which is similar to what is achieved in randomized experiments (McCaffrey et al. 2004).

With regards to wildfire effect analysis, wildfire is the treatment, and forest inventory plots are the experimental units either treated or untreated. The two groups of plots, treated or untreated, exhibit selection bias of

[5] https://lemma.forestry.oregonstate.edu/

TABLE 3. Mean of PRISM and Landsat data derived covariates, with the standard deviation in parentheses.

Variable	Description	Burned (n = 611)	Unburned (n = 22,539)
Climate			
Annpre	annual precipitation (measured in mm; multiplied by 100 and ln-transformed) from PRISM 30-yr normals	679.98 (65.25)	701.84 (66.64)
Anntmp	mean annual temperature (°C × 100)	737.22 (270.01)	755.11 (221.89)
cvpre	coefficient of variation of mean monthly precipitation of December and July (wettest and driest months)	6,928.74 (2,163.23)	7,266.69 (1,747.79)
diftmp	difference between mean August maximum temperature and mean December minimum temperature (°C × 100)	3,053.12 (293.17)	2,882.3 (455.35)
Smrmnvpd	mean summer vapor pressure deficit (hPa)	1,244.12 (228.68)	1,091.14 (276.31)
smrtp	growing season moisture stress (ratio of temperature to precipitation from May to September) (°C/ln mm)	277.78 (52.17)	262.93 (45.07)
Remotely sensed data			
NDVI	normalized difference in vegetation index	0.50 (0.14)	0.64 (0.14)
TC1	tasseled cap brightness	2,550.40 (644.25)	2,361.04 (696.60)
TC2	tasseled cap greenness	787.46 (444.06)	1,281.09 (620.91)
TC3	tasseled cap wetness	−1,251.07 (579.17)	−721.95 (593.07)

wildfires due to different environmental conditions (Cardille et al. 2001). Through matching, a set of unburned control plots that has a similar propensity score distribution to that of burned plots can be obtained. Hereafter, we refer to the untreated group including all unburned plot measurements (n = 22,359) as "unburned plots," and denote the subset of unburned plot measurements selected through matching as "control plots."

Burned plots were matched one-to-one with the nearest unburned plot in terms of the propensity scores using a greedy algorithm (e.g., Austin 2011), generating the same number of control plots as burned plots (n = 611). Three matching approaches were conducted using different sets of propensity scores estimated based on (1) a set of covariates representing environmental plot characteristics (propensity score matching [PSM]; e.g., Pufahl and Weiss 2009), (2) the spatial distance between burned and control plots (spatial matching [SM]; e.g., Samal et al. 2004), and (3) an equally weighted combination of both the set of covariates and the spatial distance (distance-adjusted propensity score matching [DAPSM]; e.g., Chagas et al. 2012). The overall process of data use and analysis are outlined in Fig. 2.

Propensity score matching (PSM).—The propensity score $p(X)$ is usually defined as a function of covariates X using binary logistic regression (Austin 2011) such that

$$p(X) = \Pr(T = 1|\boldsymbol{X}) = \mathrm{E}(T|\boldsymbol{X}) = \frac{1}{1 + e^{-\boldsymbol{X}\beta}} \quad (1)$$

where T is a binary treatment that identifies burned (T = 1) or unburned (T = 0) FIA inventory plots, $\boldsymbol{X}$+ $(X_1,\ldots,X_n)$ is the vector of covariates and $\beta = (\beta_1,\ldots,\beta_n)$ are the parameters. In this study, the propensity score presents the probability that a forest inventory plot experiences a wildfire, given environmental conditions such as climate, topography, and land cover. Sixteen covariates from FIA plot data, PRISM, and Landsat metrics (Tables 1, 3) were included in the covariate set to estimate the propensity score of each plot with logistic regression.

Spatial matching (SM).—In spatial data, the geographic distance between a pair of observations can be considered as a measure of similarity. Selection of control groups based on such spatial distance is reasonable provided that spatially close inventory plots usually share analogous environmental characteristics (Franco-Lopez et al. 2001). In practice, burned plots are often compared with adjoining plots just outside of the fire boundary (e.g., Ghebrehiwot et al. 2011). Spatial matching in this study denotes a matching method that takes into account the geographic distance to select the nearest unburned plot as control for a burned plot. For the FIA inventory plots, the standardized Euclidean distance (Dist_{ij}) between any pairs of burned and unburned forest inventory plots i,j was calculated as

$$\mathrm{Dist}_{ij} = \frac{d_{ij} - \min d_{\mathrm{all}}}{\max d_{\mathrm{all}} - \min d_{\mathrm{all}}} \quad (2)$$

where d_{ij} is the Euclidean distance between plot i, j and d_{all} denotes the distances of all burned and unburned plot pairs. The distance is standardized with the maximum and minimum value of all observed distances to have the same scale as the propensity scores [0,1] (Papadogeorgou et al. 2019). If a plot had multiple measurements, spatial matching resulted in the same propensity score; the algorithm randomly chose one of the measurements as match.

Distance-adjusted propensity score matching (DAPSM).—The propensity scores were derived based on both the environmental covariate set and the spatial distance. Papadogeorgou et al. (2019) developed the distance-adjusted propensity score (DAPS) that combines

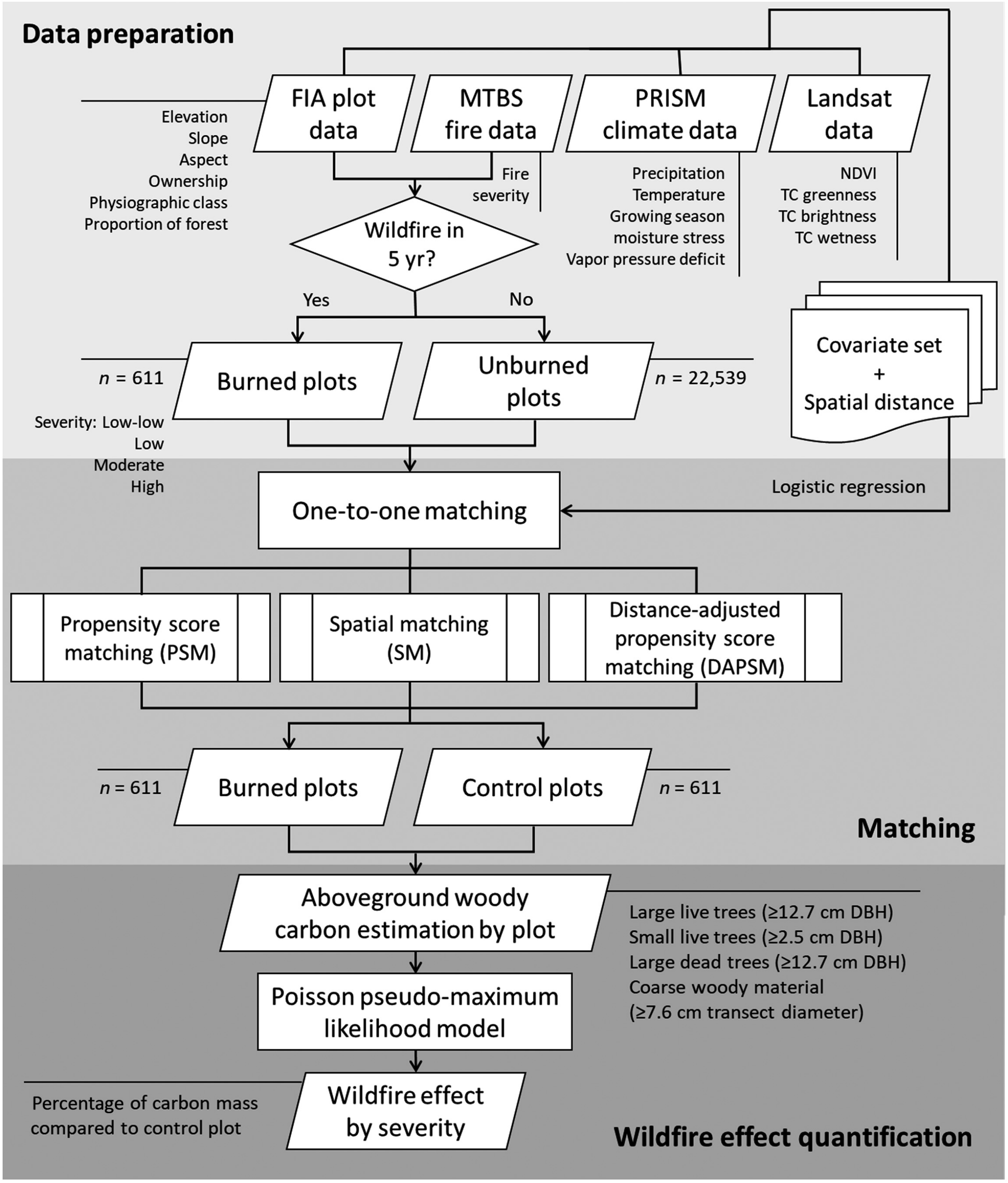

FIG. 2. Flow of the analysis in this study.

the propensity score estimate $p(X)$ with the standardized Euclidean distance

$$\text{DAPS}_{ij} = w \times \left|p(X)_i - p(X)_j\right| + (1-w) \times \text{Dist}_{ij} \quad (3)$$

where i and j are the unburned and burned plots to be paired, and w denotes the weight that puts emphasis either on covariates or distance in estimating DAPS. $p(X)$ and Dist_{ij} are the propensity scores and the relative distance, respectively, as defined previously. The choice of w depends on the balance of observed covariates or on the influence of unmeasured spatial confounding variables, and the value of w is often chosen by the researcher, based on the subject-matter knowledge regarding the potential confounding variables

(Papadogeorgou et al. 2019). We tried three values of w (e.g., $w = 0.25$, 0.5, 0.75). The means of the outcome variable from selected control plots were not significantly different among the values of w ($P = 0.999$; chi-square test). Therefore, we chose a weight of $w = 0.5$ to equally account for the observed environmental covariates and the spatial distance. Thus, DAPSM in this study represents the midpoint between PSM ($w = 1$) and SM ($w = 0$), where PSM is based only on the propensity scores estimated from the environmental covariates and where the environmental covariates are not included in the propensity score estimation for SM. All three matching methods—PSM, SM, and DAPSM—and propensity score estimation were performed using the DAPSm package (Papadogeorgou et al. 2019) in the R software (R Core Team 2013).

Diagnosis of balance between burned and control plots

Once the control plots were obtained through matching with burned plots, we diagnosed if the balance between burned and control plots had been achieved. The balance in propensity scores and in covariate distributions of burned and control plots was checked both numerically and graphically following Stuart (2010). The comparisons of the propensity score distribution before and after matching were visually presented as jitter plot (Stuart and Green 2008). Two-sample t and chi-square tests were used to test for differences in propensity scores as well as covariate distributions for continuous and categorical variables, respectively (Everitt 1992).

The balance in 16 individual covariates under each of the matching scenarios was further assessed using the absolute standardized mean differences (ASMD; Rosenbaum and Rubin 1985). ASMD denotes the mean difference of covariate values between burned and control plots relative to their standard deviations, and should generally be less than 0.2 when the balance has been achieved (Stuart and Rubin 2004). All analyses were conducted using the R software version 3.4.0 (R Core Team 2013).

Calculations of aboveground woody carbon mass

For live trees, the regional biomass oven-dry mass (Woodall et al. 2012) of individual trees was obtained from FIA plot data. The biomass includes stem, bark, and live branches of all large live trees (≥12.7 cm DBH) and small trees (≥2.54 cm DBH). All tree biomass from a forest inventory plot was summed as per hectare value, and the carbon mass (Mg/ha) was then computed for each plot by multiplying the carbon ratio (i.e., 0.5; Woodall et al. 2012).

The carbon mass from large standing dead trees (≥12.7 cm DBH) was calculated following Eskelson et al. (2016), considering the regional biomass oven-dry masses are estimated based on live trees thus often not applicable to dead trees with missing or broken top parts. We computed individual snag biomass based on the tree volume using DBH and height measurement, assuming a conic-paraboloid tree shape (Fraver et al. 2007). For trees with broken tops, original heights without the broken tops were estimated from height–diameter equations (Barrett 2006) to calculate the volume of frustums using measured heights and the estimated original heights. The tree volumes were multiplied with the wood bulk density and decay-class reduction factors (Harmon et al. 2008) to obtain dead woody biomass. Carbon mass was calculated as half of the biomass and summarized as per hectare value (Woodall et al. 2012).

Calculation of carbon mass in coarse woody materials (CWM) over 7.6 cm of transect diameter was based on the line-intersect sampling protocol of the FIA program (Woodall and Monleon 2008). Measurements of CWM loads on four 7.3 m long transects in every FIA plot were converted to carbon mass (Mg/ha) considering species-specific wood density and decay class (Harmon et al. 2008), similarly to the large standing dead trees.

The carbon masses in this study included four different aboveground woody carbon pools calculated as above: (1) large live trees, (2) small live trees, (3) large dead trees, and (4) CWM. Total carbon mass was computed as the sum of carbon masses of the four pools. Small standing dead trees under 12.7 cm DBH were not measured by FIA before 2016. Therefore, this carbon pool was not included in the analysis. Using post-2016 measurements, we performed exploratory analyses to confirm that the contribution of small dead trees to the aboveground woody carbon mass was minimal.

Average wildfire effects on aboveground woody carbon masses

When the balance in the propensity score distributions between burned and control plots is achieved, the potential outcome Y of the wildfire ($Y(T)$) is regarded to be independent of the propensity scores $p(X)$ given that the presence of T is correctly specified with the covariates (Rosenbaum and Rubin 1983). Here the potential outcome of interest is aboveground woody carbon mass contained in the plots that were either burned ($T = 1$) or unburned ($T = 0$). The average effect of wildfire is simply the difference in the mean carbon masses between burned and control plots.

The carbon mass in unburned plots and in plots burned with different severities is always positive and can often be zero. We fitted Poisson pseudo-maximum likelihood models with burn severity (Table 2) as covariate to estimate the average wildfire effect as the relative amount of carbon mass to control plots at a given burn severity. The aboveground woody carbon mass from four different pools, large live trees, small live trees, large dead trees, and CWM, and the total carbon mass (Mg/ha) was set as the response variable, and five burn-severity classes (Table 2) including unburned status from controls ($n = 611$) were used as the explanatory variable.

Models were constructed for each combination of carbon pools (i.e., large live trees, small live trees, large dead trees, CWM, and total) and matching methods (i.e., PSM, SM, and DAPSM) including a random effect of the matched pairs describing the correlation between the pairs (Austin 2008). The estimated coefficients of the fire severities can be interpreted as the proportion of carbon mass in burned plots, relative to that in unburned controls, under the corresponding burn severity (Long and Freese 2006). The 95% confidence intervals of the estimated coefficients were constructed using Tukey's pairwise comparison (Malison and Baxter 2010). Model fitting and computation of summary statistics were conducted using the SAS 9.4 PROC GLIMMIX procedure (SAS Institute, Cary, North Carolina, USA).

Results

Balance in propensity scores and in covariate distributions achieved under both PSM and DAPSM

Thee three matching scenarios, PSM, SM, and DAPSM, created one-to-one matches of control plots to all 611 burned plots, although the spatial distributions of control plots were different for each of the three scenarios (Fig. 3). PSM showed a scattered pattern of control plots over most of the forested lands in Washington and Oregon such as the Cascade Mountain Range and around Harney Basin, except for a few plots on the Olympic Peninsula and coastal Oregon where no large wildfire was reported during the study period (Fig. 1). DAPSM tended to attain control plots that were farther to the burned plots than the control plots selected under SM (Fig. 3). The mean distances between control plots and the closest burned plots were approximately 20.2, 3.2, and 4.7 km under PSM, SM, and DAPSM, respectively.

The propensity scores estimated from the logistic regression model ranged from approximately 0 to 0.44 across all burned and unburned plots (Fig. 4). Out of the sixteen environmental covariates, ten were found to be statistically significant in the linear logistic regression model to estimate the propensity scores (Appendix S1: Table S1). All six climate variables including mean annual precipitation (annpre) and temperature (anntmp), slope (SLOPE), land ownership (OWN), physiographic class (PHYSCL), and tasseled cap brightness (TC1) were strongly associated with the probability of wildfires ($P < 0.05$). The distribution of propensity scores was heavily right-skewed with a few plots having large propensity scores, meaning that those few plots had much higher chances of wildfire. The mean propensity score of all control plots was significantly smaller than the mean of the burned plots ($P < 0.0001$ in Welch's two-sample t test), with no statistical difference from those of the burned plots under PSM and DAPSM (Appendix S1: Table S2). The mean propensity score of SM-matched control plots was significantly smaller than those of burned plots ($P = 0.004$; Welch's two-sample t test). However, SM-matching greatly reduced the observed discrepancy in the mean propensity score between burned and unburned plots.

The comparisons of the individual covariate distributions based on two-sample t and chi-square tests also exhibited similar results of balance achievement under each method. When unmatched, the empirical means were significantly different ($P < 0.05$ in t test) between burned and control plots for 13 out of 16 covariates, excluding proportion of forest (PROPforest), annual temperature (anntmp), and tasseled cap brightness (TC1). When matched by PSM and DAPSM, all covariates achieved balance at the $\alpha = 0.05$ level. SM left slope (SLOPE) and physiographic class of the site (PHYSCL) imbalanced ($P = 0.010$ and $P = 0.0002$, respectively).

In terms of the absolute standardized difference in means (ASDM), 11 out of the 16 covariates were found to be imbalanced (ASDMs > 0.2) between burned and unburned plots (Fig. 5). Summer mean vapor pressure deficit (Smrmnvpd) exhibited the largest ASDM (0.607), followed by physiographic class (PHYSCL, 0.542), difference in mean precipitation (diftmp, 0.451), and tasseled cap greenness (TC2, 0.447). This means that those variables may account for a large portion of the selection bias. When matched based on PSM and DAPSM, ASDM of all covariates resulted in values below the cut-off value of 0.2 (Fig. 5). ASDM of PROPforest slightly increased after matching in the process of balancing the other covariates, although the increments were minimal (black lines in Fig. 5). Under SM, on the other hand, the ASDM of PHYSCL (0.236) was above the cut-off value of 0.2. SLOPE also showed relatively large ASDM (0.148) compared to the other covariates.

Discrepancy in carbon mass of standing and downed wood between unburned and control plots

The mean of total carbon mass (Mg/ha) of all unburned plots ($n = 22{,}539$) was notably larger than that of control plots ($n = 611$) or burned plots with low-low severity ($n = 102$), implying that there are considerable differences in stand level characteristics between burned and unburned forests (left four bars in Fig. 6). The PSM-selected control plots had the greatest amount of mean total carbon mass due to the large live tree carbon mass (77 Mg/ha) compared to that observed under SM and DAPSM (~70 Mg/ha). DAPSM produced control plots with the smallest total carbon mass among the three matching methods, and the composition ratio of four pools within the total carbon mass was similar to the ratio observed from the burned plots with low-low severity. Carbon masses from CWM and small live trees in DAPSM-selected controls were slightly larger than those in the low-low-severity plots. The total carbon mass in PSM-selected control plots was significantly different from the other two methods ($P = 0.044$) due to the large tree carbon mass, otherwise the carbon masses

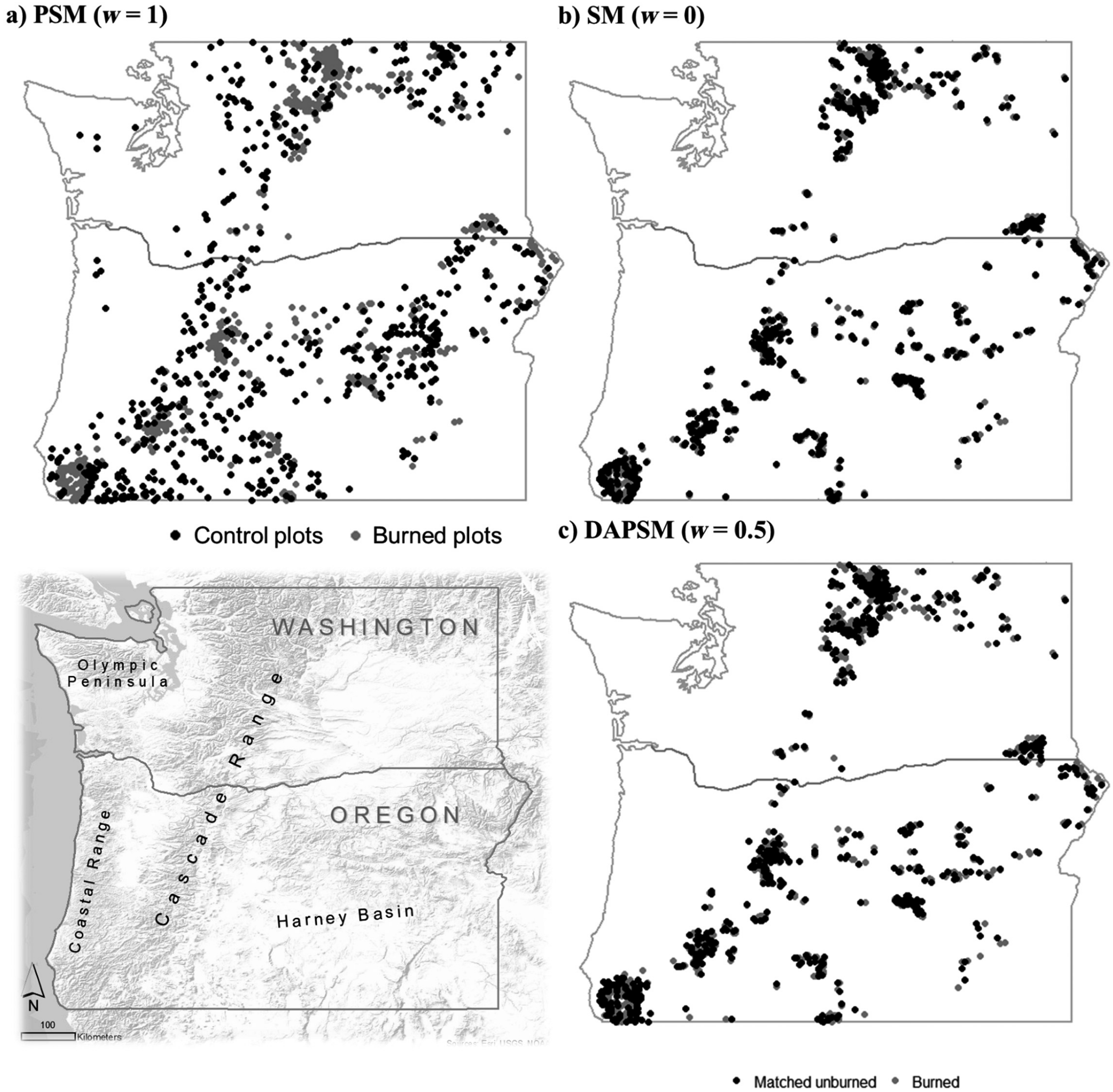

FIG. 3. Spatial distributions of burned plots (gray dots) and control plots (black dots) over Washington and Oregon under three matching strategies: (a) propensity score matching (PSM, $w = 1$), (b) spatial matching (SM, $w = 0$), and (c) distance-adjusted propensity score matching (DAPSM, $w = 0.5$).

from individual pools were statistically the same among the three matching methods ($P > 0.05$).

The total carbon mass in the burned plots was not statistically different between low-low and low burn severity ($P = 0.5$ in two-sample t test), despite the difference in the composition ratio of four pools (right four bars in Fig. 6). The plots with moderate wildfires had a significant reduction of the mean total carbon mass compared to the control plots, largely due to the decrease of carbon mass from large live trees (Fig. 6). The plots with high-severity fire contained similar total carbon mass to those with low-low and low-severity fire, but the mean of total carbon mass was significantly lower in high-severity plots than that observed in the control plots ($P < 0.006$ under all method). Looking at individual carbon pools, the mean carbon mass from large and small live trees significantly decreased with increasing burn severity (Fig. 6). Meanwhile, the opposite trend was observed for large dead trees (≥12.7 cm DBH), meaning that live carbon mass was consumed or converted into dead carbon experiencing wildfire. The mean of carbon mass in CWM did not vary substantially across severities (6.42–8.62 Mg/ha across all severities), and was the lowest under moderate severity (Fig. 6). Under high-severity fire, the mean carbon mass in CWM was significantly larger than under low-low severity.

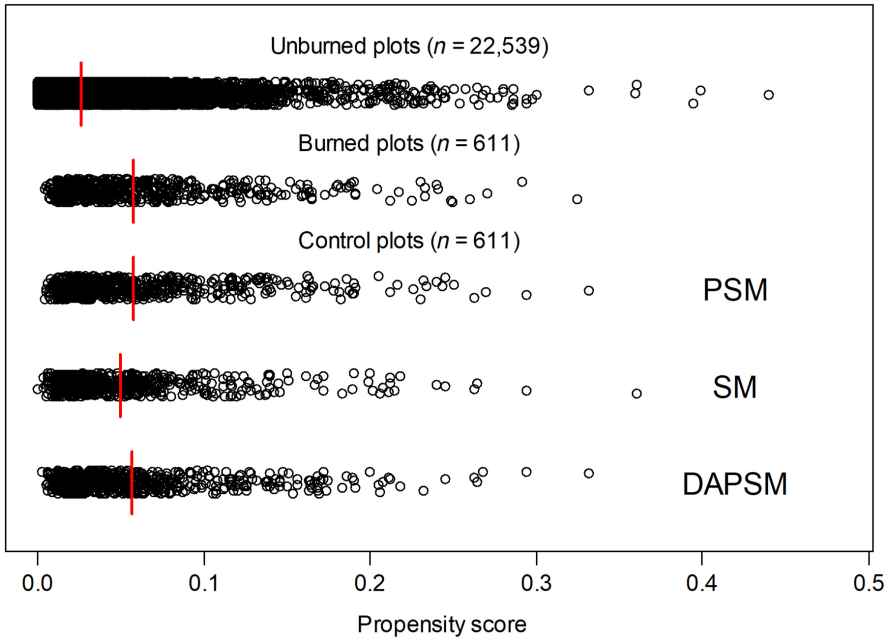

Fig. 4. Jitter plot of propensity scores from unburned plots (n = 22,539), burned plots (n = 611), and control plots (n = 611) under three matching scenarios: PSM, SM, and DAPSM. Red lines denote the mean value of propensity scores in each category: 0.025 for control, 0.057 for burned, 0.057, 0.049, and 0.056 for control plots matched under PSM, SM, and DAPSM, respectively.

Average wildfire effect on aboveground woody carbon mass under three matching methods

While SM and DAPSM indicated a significant reduction of total carbon mass under low-severity fire (95% CI 63–105% and 69–115%, respectively), PSM failed to detect a statistically significant difference under the same carbon pool and burn severity (95% CI 57–98%; Fig. 7 a). This tendency of contradicting results in detecting a statistically significant difference compared to controls between PSM and the other two matching methods also appears for large live trees under low-severity fire (Fig. 7 b) as well as for CWM under low and moderate fire (Fig. 7e). The only exception was in small live tree carbon mass under low-low-severity fire, where SM-matched control plots identified a difference for this carbon pool while DAPSM matching did not detect a difference from control plots (Fig. 7c).

In terms of the estimated proportion by severities, low-low-severity fires did not alter the total carbon mass or the carbon mass in any of the pools except for the small live trees under PSM and SM (Fig. 7c). Small live tree carbon mass significantly decreased by 22–28% of the control plots under all methods in low-severity fire, while the carbon mass from large dead trees increased approximately 350 % on average across all three matching methods. In total carbon mass, however, little difference under low burn severity was observed due to the conversion of carbon mass from live to dead trees. Plots that experienced moderate burn severity showed a significant decrease of total carbon mass under all three matching methods to the extent of 58–66% of control plots on average ($P < 0.05$ for all methods). There was a large reduction in live tree carbon mass, while the increase in dead tree carbon was similar to the level observed for low-severity fire (approximately 340–370% of control plots across all methods). Under high-severity fires, on the other hand, the observed total carbon mass was larger than under moderate-severity fires, but not significantly. Carbon mass from CWM did not show as clear of a monotonic increase or decrease along with increasing burn severity as the other pools.

Discussion

Balance in observed covariates may not guarantee unbiased estimation of average wildfire effects

The selection bias induced by the difference in covariates between burned and unburned plots was mostly removed under all three matching methods, except for SM showing imbalance in physiographic class and potentially slope. Slope and physiographic class are strongly related to the moisture condition of sites (Megahan 1983). Therefore, they are important variables in describing the probability of wildfire occurrence (Marques et al. 2012) as well as forest carbon mass (Sharma et al. 2011). The observed imbalance under SM in the covariate distributions suggests that spatially close unburned plots may not be entirely similar to burned plots in terms of their environmental conditions. Even

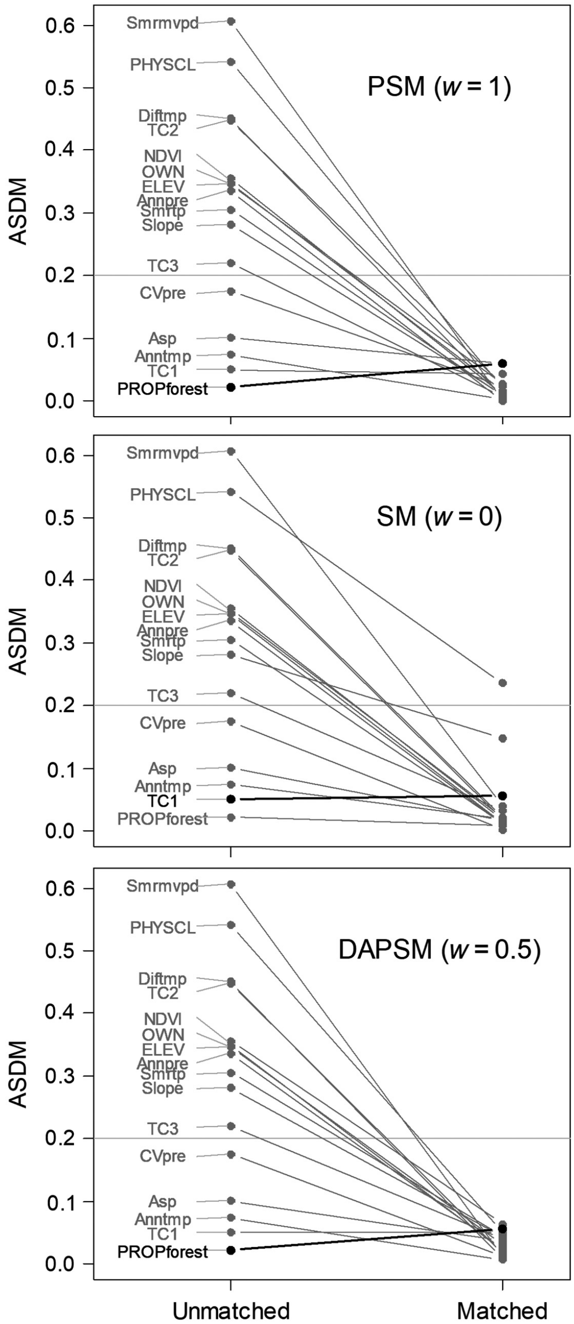

Fig. 5. Absolute standardized mean difference (ASMD) of 16 environmental covariates before and after matching under three strategies. The covariates are defined in Table 1 and Table 3. ASDM of the covariates with gray lines decreased after matching, while black lines denote that ASDM increased after matching. ASDM over the cut-off value (0.2, gray line) denotes the covariate means are different between burned and unburned plots.

though many studies use measurements of control plots adjacent to the fire perimeter to compare with the burned plots (e.g., Ghebrehiwot et al. 2011, Heckman et al. 2013, Oliveras et al. 2018), researchers should always check if the environmental conditions of the control plots are similar to those of the burned plots to assure that the estimation of average wildfire effects is unbiased (Certini et al. 2011). Nevertheless, the balance acquired under SM except for only two covariates is noteworthy, because SM matched burned and unburned plots solely based on distances, regardless of any environmental conditions.

Meanwhile, PSM considered only the available environmental covariates in matching. The balance achievement in covariate distributions under PSM was anticipated, because the method used the same covariate set for propensity score estimation. However, PSM resulted in choosing several inventory plots in coastal forests, where wildfires are rare (Gavin et al. 2013) and contain larger carbon mass compared to inland forests due to the climate (Gholz 1982). Estimating wildfire effects using the measurements from those plots may be misleading, because the amount of carbon mass in the two regions considerably differ (Palmer et al. 2018, 2019). Incorporating forest type to estimate the propensity scores may prevent selecting control plots from very dissimilar environments to the burned plots. For a small sample size of 54 plots, Campbell et al. (2007) were able to use a forest type variable in their analysis of carbon change in litter and duff by the 2002 Biscuit fire in southern Oregon, USA. However, for our larger sample size we did not have forest type information available for all plots. Specifically, for many of the burned plots that were only measured after they burned, we were lacking this information. Therefore, we could not include forest type in the matching process.

PSM controls had the largest mean carbon mass among the three sets of matched controls, and the amount was higher than that observed from low-low-severity plots. Low-low burn severity comprises very low tree mortality or unburned area within the fire perimeter following MTBS classification (Cocke et al. 2005, Eidenshink et al. 2007). Thus, low-low-severity fires hardly affect the woody carbon mass in forests (Campbell et al. 2007, Whittier and Gray 2016). Large live tree carbon mass in plots that burned with low-low severity is expected to show little difference from control plots. The carbon mass was significantly different between low-low-severity plots and PSM-selected controls in large live tree pool, which contradicts the general understanding of low-low fires. This implies that the selection bias in estimating wildfire effects has not been fully addressed under PSM, despite the balance achieved in propensity scores and covariate distributions. On the other hand, SM produced control plots with similar amount of carbon mass to that of low-low-severity plots, even though the balance was not achieved. This result suggests that spatial distance may have played a more important role in matching for wildfire effect quantification than the environmental covariates.

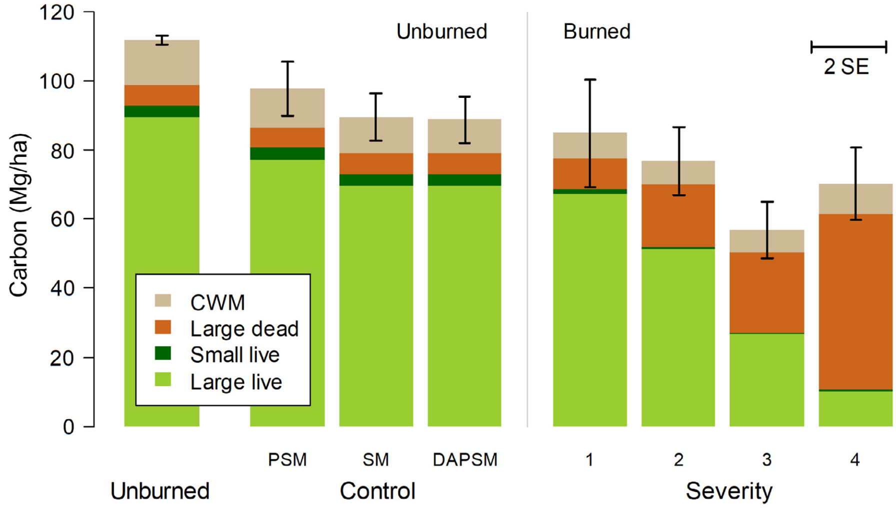

FIG. 6. Mean total carbon masses (Mg/ha) by severity classes, shown with 2 SE. All unburned plots (n = 22,539) and control plots (n = 611 for each method) denote the unburned plots before and after matching under three matching methods: PSM, SM, and DAPSM.

DAPSM is preferable for wildfire effects estimation

Unobserved covariates may result in bias in matching estimates due to residual imbalances between treated and control groups (Rosenbaum 1987). Propensity score matching warrants unbiased effect estimation conditionally on the covariates included, based on the assumption that there are no unobserved confounding variables (Rosenbaum and Rubin 1983). The assumption is impossible to test and it is difficult to identify the presence of unobserved confounding variables from the observed dataset, especially for ecological data for which the relationship between treatment and outcome is complex (Yuan 2010, Gross and Rosenheim 2011). Wildfires and carbon content in forests are associated with a large number of environmental covariates (Ager et al. 2007, Parisien et al. 2012). Therefore, it is practically impossible to address all the confounding variables in propensity score estimation.

Research in other fields of study has demonstrated that the problem of undetected confounding variables can be mitigated by incorporating spatial information, such as adding spatially autocorrelated random effects (Verbitsky-Savitz and Raudenbush 2012) or minimizing the spatial distance in matching (Keele et al. 2015). PSM in our study used propensity scores estimated from 16 covariates that cover a wide range of environmental conditions such as topography, climate, land cover from ground measurements and remotely sensed metrics. However, there still may be confounding variables related to the probability of wildfires and carbon mass that were not included in this set. Based on our results, the bias due to unobserved confounding variables can be reduced by introducing spatial information. DAPSM met the diagnostic criteria for checking the balance in propensity scores, and the selected control plots contained similar amount of carbon mass to low-low-severity plots. Therefore, we conclude that DAPSM is the preferred method for estimating wildfire effects, because it reduces the selection bias due to both observable and unobservable confounding variables.

DAPSM estimation can further improve when combined with other evaluation methods (Caliendo and Kopeinig 2008), especially when longitudinal data are available (Smith and Todd 2005). For example, future studies, for which measurements before and after wildfire are available, may incorporate a difference-in-differences analysis (Heckman et al. 1998) for the matched samples with a time relevant variable (e.g., years since fire). Difference-in-differences analysis in combination with propensity score matching may produce accurate and robust estimators (Ryan et al. 2015, Ho et al. 2007) by obtaining before- and after-measurements of burned plots and comparing each case to before- and after-measurements of control plots, respectively. Difference-in-differences matching analysis was found to be effective for removing selection bias due to unobserved confounding variables (Smith and Todd 2005), and has been applied in research quantifying the effects of economic policies (e.g., Girma and Görg 2007), social (e.g., Guo et al. 2006) and health programs (e.g., Whittaker et al. 2016), as well as forest management strategies (e.g., Brandt et al. 2015, Nolte et al. 2017), where repeated measurements over time were available. In our data, repeated measurements before and after wildfire were only available for a subset of plots. Further accumulation of FIA data will allow for more accurate and temporally invariant estimation of wildfire effects (Smith and Todd 2005).

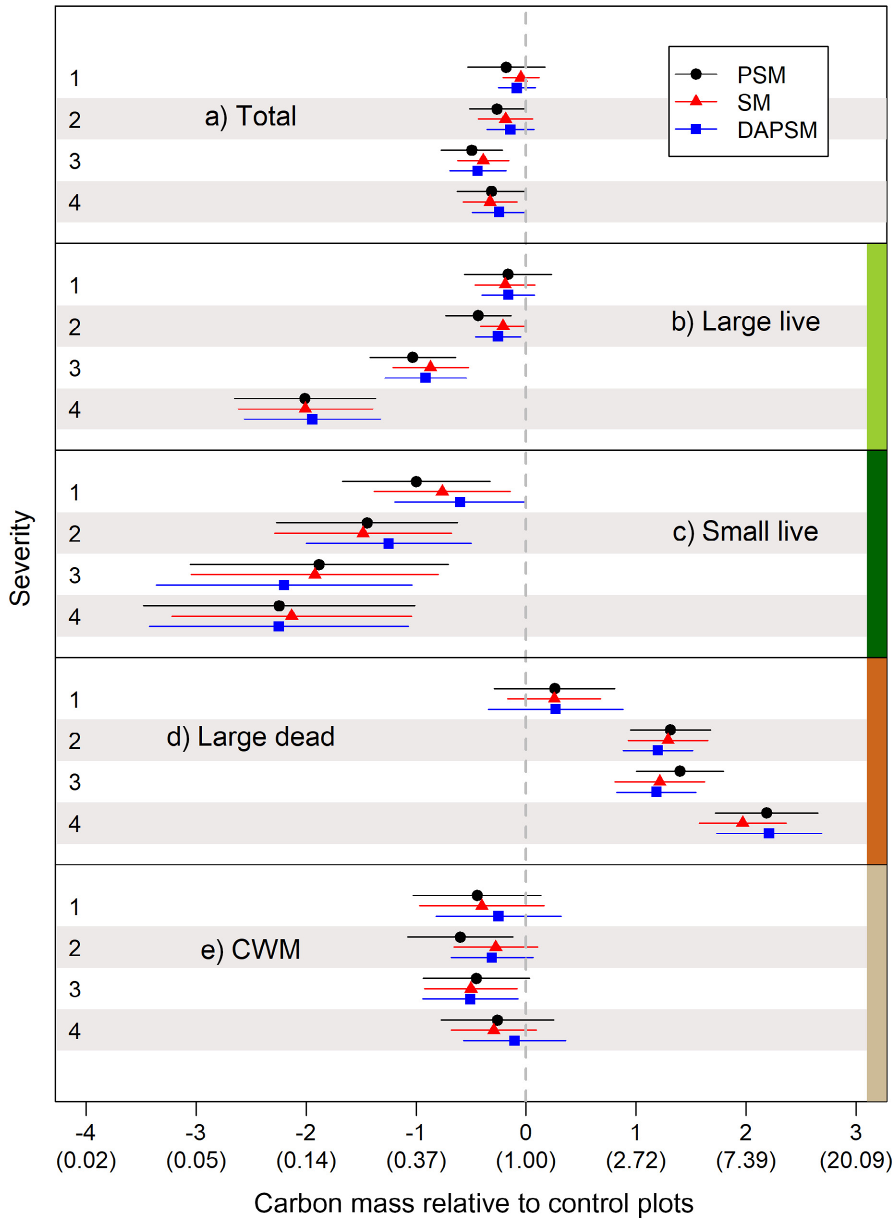

FIG. 7. Percentage of carbon mass relative to control plots (natural-log-transformed) estimated from the Poisson pseudo-maximum likelihood model were presented with 95% confidence intervals with Tukey's adjustment. The exponential of the natural log (given in parenthesis) denotes the percentage of remaining carbon mass compared to control plots under each burn severity (e.g., 1, low-low; 2, low; 3, moderate; or 4, high). For example, approximately 14% of carbon mass from large live trees remained under high-severity fire, while carbon mass from large dead trees increased over seven times under high-severity fire. The intervals including the gray vertical line showed no difference from control plots under 95% confidence level.

Estimated average wildfire effects on carbon mass by severity from DAPSM

Based on the DAPSM estimates, on average, the total carbon mass from aboveground woody pools in forests of the U.S. Pacific Northwest was barely consumed by wildfire. Instead, there was redistribution of carbon from live to dead trees. Meigs et al. (2009) reported that approximately 13–35% of carbon is emitted due to combustion from wildfires across all severities in the ponderosa pine and Douglas-fir forests in the East Cascades of Oregon, USA. The results were consistent with those of Campbell et al. (2007), and both studies were conducted for large, stand-replacing wildfires (e.g., >1,000 ha) using detailed prefire data or unburned plot measurements. Our estimation of average wildfire effects on carbon mass was focused on reduction of stem bole, bark, and large branches based on the allometric equations of tree biomass utilized in FIA program (Smith et al. 2006). Therefore, our approach excluded carbon from foliage, shrubs, and small downed wood that occupies a large portion of carbon consumption due to wildfires according to Meigs et al. (2009) and Campbell et al. (2007). Our result of little reduction in woody carbon mass by wildfire may also be related to the classification of burned plots. We specified burned plots that experienced any fire within 5 yr prior to field measurement. Hence, the estimated wildfire effects may include delayed mortality and short-term regeneration as well as the immediate consumption of trees (Carlson et al. 2012). There may be controversy on how to distinguish wildfire-affected forests from post-disturbance forests because carbon dynamics after a wildfire substantially vary by the nature of the disturbance (Raymond et al. 2015, Bartels et al. 2016). Limiting the criterion for classifying burned plots will allow us to quantify more immediate effects of wildfires on forest woody carbon, yet the decrease in sample size may hinder generalization of the average effect estimates over the region.

Among the different burn severities, only moderate-severity fires led to a statistically significant decrease in total carbon mass, while high-severity fire did not show statistically significant differences in the mean of total carbon mass compared to control plots. This result seemingly contradicts the general thought that more woody carbon is consumed under more severe fire. Peterson et al. (2019) found similar results for the Boulder Creek fire in British Columbia, Canada, where less carbon remained under moderate burn severity than under high severity in coarse woody materials. They attributed this to the preexisting difference between the moderate-severity fire plots and the other severity plots (Peterson et al. 2019). However, in our data, no statistically significant differences in stand structure or species composition were observed between moderate-severity fire plots and the other severities during the preliminary analyses. We speculated that the reason may be due to disagreement of burn severity classification between the ground measurements and satellite imagery metrics. Whittier and Gray (2016) found that MTBS burn severity classification tended to classify burn severities as less severe than the tree-mortality based classification from field data did, potentially due to the delayed mortality of large trees after a fire. Continued research on wildfire effects using ground measurements may help improve burn severity classification in relation with actual vegetation changes along with time since fire.

Conclusion

Propensity score matching allowed obtaining control plots for which environmental conditions were comparable to burned plots to quantify average wildfire effects on aboveground forest woody carbon across the U.S. Pacific Northwest. Among the three types of propensity scores computed for forest inventory plots, DAPSM matching was the most preferable for selection bias reduction in terms of balance achievement and dealing with unobservable confounding variables. The average wildfire effect was calculated as the proportion of forest woody carbon mass relative to control plots that showed similar distributions of environmental covariates to burned plots. This study serves as an applied example of matching based on empirical data such as forest inventory measurements and satellite imagery metrics. This study supports Papadogeorgou et al.'s (2019) that both spatial information and observed covariates should be included for matching to address natural disturbances for which cause-and-effect estimation are complicated. Our results provide regional estimates of average wildfire effects across all severities over Washington and Oregon, USA, without the need for generalization from local case studies. Matching methods will be a useful tool that can enable effect quantification for other natural disturbances over landscapes, such as insects and diseases, drought, and windstorms. Quasi-experimental methods supported by observational data allow unbiased estimation of effects caused by natural disturbances for which experimental settings are unavailable.

Acknowledgments

This study was supported by a 4-yr doctoral fellowship from the Faculty of Forestry provided by the University of British Columbia, Canada. We thank Matthew Gregory from the Landscape Ecology, Modeling, Mapping, and Analysis (LEMMA) team for his assistance with the PRISM and Landsat data retrieval and compilation, and Nguyet Anh Nguyen for editorial assistance. We also thank the many individuals involved in the design, field data collection, quality assurance, and processing of the U.S. Forest Service Forest Inventory and Analysis Program. We thank two anonymous reviewers for their valuable feedback and suggestions.

Literature Cited

Adelson, J. L. 2013. Educational research with real-world data: Reducing selection bias with propensity score analysis. Practical Assessment, Research, and Evaluation 18:1–11.

Ager, A. A., M. A. Finney, B. K. Kerns, and H. Maffei. 2007. Modeling wildfire risk to northern spotted owl (*Strix occidentalis caurina*) habitat in Central Oregon, USA. Forest Ecology and Management 246:45–56.

Ager, A. A., M. A. Finney, A. McMahan, and J. Carthcart. 2010. Measuring the effect of fuel treatments on forest carbon using landscape risk analysis. Natural Hazards and Earth System Sciences 10:1–12.

Amiro, B. D., J. B. Todd, B. M. Wotton, K. A. Logan, M. D. Flannigan, B. J. Stocks, J. A. Mason, D. L. Martell, and K. G. Hirsch. 2001. Direct carbon emissions from Canadian forest fires, 1959–1999. Canadian Journal of Forest Research 31:512–525.

Austin, P. C. 2008. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. Statistics in Medicine 27:2037–2049.

Austin, P. C. 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behavioral Research 46:399–424.

Barrett, T. M. 2006. Optimizing efficiency of height modeling for extensive forest inventories. Canadian Journal of Forest Research 36:2259–2269.

Bartels, S. F., H. Y. H. Chen, M. A. Wulder, and J. C. White. 2016. Trends in post-disturbance recovery rates of Canada's forests following wildfire and harvest. Forest Ecology and Management 361:194–207.

Bechtold, W. A., and P. L. Patterson. 2005. The enhanced forest inventory and analysis program-national sampling design and estimation procedures. General Technical Report SRS-80. U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, North Carolina, USA.

Brandt, J. S., V. Butsic, B. Schwab, T. Kuemmerle, and V. C. Radeloff. 2015. The relative effectiveness of protected areas, a logging ban, and sacred areas for old-growth forest protection in southwest China. Biological Conservation 181:1–8.

Butry, D. T. 2009. Fighting fire with fire: estimating the efficacy of wildfire mitigation programs using propensity scores. Environmental and Ecological Statistics 16:291–319.

Butsic, V., D. J. Lewis, V. C. Radeloff, M. Baumann, and T. Kuemmerle. 2017. Quasi-experimental methods enable stronger inferences from observational data in ecology. Basic and Applied Ecology 19:1–10.

Caliendo, M., and S. Kopeinig. 2008. Some practical guidance for the implementation of propensity score matching. Journal of Economic Surveys 22:31–72.

Campbell, J., D. Donato, D. Azuma, and B. Law. 2007. Pyrogenic carbon emission from a large wildfire in Oregon, United States. Journal of Geophysical Research: Biogeosciences 112. http://dx.doi.org/10.1029/2007jg000451

Campbell, S., K. Waddell, and A. Gray. 2010. Washington's forest resources, 2002–2006: Five-year forest inventory and analysis report. General Technical Report PNW-GTR-800. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, Oregon, USA.

Cardille, J. A., S. J. Ventura, and M. G. Turner. 2001. Environmental and social factors influencing wildfires in the Upper Midwest, United States. Ecological Applications 11:111–127.

Carlson, C. H., S. Z. Dobrowski, and H. D. Safford. 2012. Variation in tree mortality and regeneration affect forest carbon recovery following fuel treatments and wildfire in the Lake Tahoe Basin, California, USA. Carbon Balance and Management 7:1–17.

Certini, G., C. Nocentini, H. Knicker, P. Arfaioli, and C. Rumpel. 2011. Wildfire effects on soil organic matter quantity and quality in two fire-prone Mediterranean pine forests. Geoderma 167:148–155.

Chagas, A. L. S., R. Toneto, and C. R. Azzoni. 2012. A spatial propensity score matching evaluation of the social impacts of sugarcane growing on municipalities in Brazil. International Regional Science Review 35:48–69.

Cocke, A. E., P. Z. Fulé, and J. E. Crouse. 2005. Comparison of burn severity assessments using Differenced Normalized Burn Ratio and ground data. International Journal of Wildland Fire 14:189–198.

Crist, E. P., and R. C. Cicone. 1984. A physically-based transformation of Thematic Mapper data—The TM Tasseled Cap. IEEE Transactions on Geoscience and Remote Sensing GE-22:256–263.

D'Agostino, R. B. Jr. 1998. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Statistics in Medicine 17:2265–2281.

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology: a Journal of the Royal Meteorological Society 28:2031–2064.

Dixon, R. K., A. M. Solomon, S. Brown, R. A. Houghton, M. C. Trexier, and J. Wisniewski. 1994. Carbon pools and flux of global forest ecosystems. Science 263:185–190.

Donnegan, J. A., S. J. Campbell, and D. L. Azuma. 2008. Oregon's Forest Resources, 2001–2005: Five-year Forest Inventory and Analysis Report. US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 765 p. 186.

Dray, S., P. Legendre, and P. R. Peres-Neto. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecological Modelling 196:483–493.

Duane, M. V., W. B. Cohen, J. L. Campbell, T. Hudiburg, D. P. Turner, and D. L. Weyermann. 2010. Implications of alternative field-sampling designs on Landsat-based mapping of stand age and carbon stocks in Oregon forests. Forest Science 56:405–416.

Eidenshink, J., B. Schwind, K. Brewer, Z.-L. Zhu, B. Quayle, and S. Howard. 2007. A project for monitoring trends in burn severity. Fire Ecology 3:3–21.

Eskelson, B. N. I., and V. J. Monleon. 2018. Post-fire surface fuel dynamics in California forests across three burn severity classes. International Journal of Wildland Fire 27:114–124.

Eskelson, B. N. I., V. J. Monleon, and J. S. Fried. 2016. A 6 year longitudinal study of post-fire woody carbon dynamics in California's forests. Canadian Journal of Forest Research 46:610–620.

Everitt, B. S. 1992. The analysis of contingency tables. CRC Press, Boca Raton, Florida, USA.

Flannigan, M. D., M. A. Krawchuk, W. J. de Groot, B. M. Wotton, and L. M. Gowman. 2009. Implications of changing climate for global wildland fire. International Journal of Wildland Fire 18:483–507.

Franco-Lopez, H., A. R. Ek, and M. E. Bauer. 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. Remote Sensing of Environment 77:251–274.

Fraver, S., A. Ringvall, and B. G. Jonsson. 2007. Refining volume estimates of down woody debris. Canadian Journal of Forest Research 37:627–633.

Gavin, D. G., L. B. Brubaker, and D. N. Greenwald. 2013. Postglacial climate and fire-mediated vegetation change on the western Olympic Peninsula, Washington (USA). Ecological Monographs 83:471–489.

Ghebrehiwot, H. M., M. G. Kulkarni, M. E. Light, K. P. Kirkman, and J. Van Staden. 2011. Germination activity of smoke residues in soils following a fire. South African Journal of Botany 77:718–724.

Gholz, H. L. 1982. Environmental limits on aboveground net primary production, leaf area, and biomass in vegetation zones of the Pacific Northwest. Ecology 63:469–481.

Girma, S., and H. Görg. 2007. Evaluating the foreign ownership wage premium using a difference-in-differences matching approach. Journal of International Economics 72:97–112.

Gross, K., and J. A. Rosenheim. 2011. Quantifying secondary pest outbreaks in cotton and their monetary cost with causal-inference statistics. Ecological Applications 21:2770–2780.

Guo, S., R. P. Barth, and C. Gibbons. 2006. Propensity score matching strategies for evaluating substance abuse services for child welfare clients. Children and Youth Services Review 28:357–383.

Harmon, M. E., C. W. Woodall, B. Fasth, and J. Sexton. 2008. Woody detritus density and density reduction factors for tree species in the United States: a synthesis. General Technical

Report NRS-29. U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, Pennsylvania, USA.

Heckman, J. J., H. Ichimura, J. Smith, and P. Todd. 1996. Sources of selection bias in evaluating social programs: An interpretation of conventional measures and evidence on the effectiveness of matching as a program evaluation method. Proceedings of the National Academy of Sciences USA 93:13416–13420.

Heckman, J., H. Ichimura, J. Smith, and P. Todd. 1998. Characterizing selection bias using experimental data. Econometrica 66:1017–1098.

Heckman, K., J. Campbell, H. Powers, B. Law, and C. Swanston. 2013. The influence of fire on the radiocarbon signature and character of soil organic matter in the Siskiyou National Forest, Oregon, USA. Fire Ecology 9:40–56.

Hernán, M. A., S. Hernández-Díaz, and J. M. Robins. 2004. Structural approach to selection bias. Epidemiology 15:615–625.

Ho, D. E., K. Imai, G. King, and E. A. Stuart. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Political Analysis 15:199–236.

Hudiburg, T. W., B. E. Law, C. Wirth, and S. Luyssaert. 2011. Regional carbon dioxide implications of forest bioenergy production. Nature Climate Change 1:419–423.

Hurteau, M. D., and M. North. 2010. Carbon recovery rates following different wildfire risk mitigation treatments. Forest Ecology and Management 260:930–937.

Johnson, D., J. D. Murphy, R. F. Walker, D. W. Glass, and W. W. Miller. 2007. Wildfire effects on forest carbon and nutrient budgets. Ecological Engineering 31:183–192.

Keele, L., R. Titiunik, and J. R. Zubizarreta. 2015. Enhancing a geographic regression discontinuity design through matching to estimate the effect of ballot initiatives on voter turnout. Journal of the Royal Statistical Society: Series A (Statistics in Society) 178:223–239.

Keith, H., D. B. Lindenmayer, B. G. Mackey, D. Blair, L. Carter, L. McBurney, S. Okada, and T. Konishi-Nagano. 2014. Accounting for biomass carbon stock change due to wildfire in temperate forest landscapes in Australia. PLoS ONE 9: e107126.

Kremens, R. L., A. M. S. Smith, and M. B. Dickinson. 2010. Fire metrology: current and future directions in physics-based measurements. Fire Ecology 6:13–35.

Larsen, A. E., K. Meng, and B. E. Kendall. 2019. Causal analysis in control–Impact ecological studies with observational data. Methods in Ecology and Evolution 10:924–934.

Law, B. E., and R. H. Waring. 2015. Carbon implications of current and future effects of drought, fire and management on Pacific Northwest forests. Forest Ecology and Management 355:4–14.

Lechner, M. 2002. Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies. Review of Economics and Statistics 84:205–220.

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? Ecology 74:1659–1673.

Long, J. S., and J. Freese. 2006. Regression models for categorical dependent variables using Stata. Stata Press, College Station, Texas, USA.

Malison, R. L., and C. V. Baxter. 2010. Effects of wildfire of varying severity on benthic stream insect assemblages and emergence. Journal of the North American Benthological Society 29:1324–1338.

Marques, S., J. Garcia-Gonzalo, B. Botequim, A. Ricardo, J. G. Borges, M. Tomé, and M. M. Oliveira. 2012. Assessing wildfire occurrence probability in *Pinus pinaster* Ait. stands in Portugal. Forest Systems 21:111–120.

McCaffrey, D. F., G. Ridgeway, and A. R. Morral. 2004. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. Psychological Methods 9:403–425.

Megahan, W. F. 1983. Hydrologic effects of clearcutting and wildfire on steep granitic slopes in Idaho. Water Resources Research 19:811–819.

Meigs, G. W., D. C. Donato, J. L. Campbell, J. G. Martin, and B. E. Law. 2009. Forest fire impacts on carbon uptake, storage, and emission: the role of burn severity in the Eastern Cascades, Oregon. Ecosystems 12:1246–1267.

Mitchell, S. R., M. E. Harmon, and K. E. B. O'Connell. 2009. Forest fuel reduction alters fire severity and long-term carbon storage in three Pacific Northwest ecosystems. Ecological Applications 19:643–655.

Nolte, C., B. Gobbi, Y. le Polain de Waroux, M. Piquer-Rodríguez, V. Butsic, and E. F. Lambin. 2017. Decentralized land use zoning reduces large-scale deforestation in a major agricultural frontier. Ecological Economics 136:30–40.

Oliveras, I., R. M. Román-Cuesta, E. Urquiaga-Flores, J. A. Quintano Loayza, J. Kala, V. Huamán, N. Lizárraga, G. Sans, K. Quispe, and E. Lopez. 2018. Fire effects and ecological recovery pathways of tropical montane cloud forests along a time chronosequence. Global Change Biology 24:758–772.

Ottmar, R. D. 2014. Wildland fire emissions, carbon, and climate: modeling fuel consumption. Forest Ecology and Management 317:41–50.

Palmer, M., O. Kuegler, and G. Christensen. 2018. Oregon's forest resources, 2007–2016: Ten-year Forest Inventory and Analysis report. General Technical Report PNW-GTR-976. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, Oregon, USA.

Palmer, M., O. Kuegler, and G. Christensen. 2019. Washington's forest resources, 2006–2015: Ten-year Forest Inventory and Analysis report. General Technical Report PNW-GTR-971. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, Oregon, USA.

Papadogeorgou, G., C. Choirat, and C. M. Zigler. 2019. Adjusting for unmeasured spatial confounding with distance adjusted propensity score matching. Biostatistics 20:256–272.

Parisien, M.-A., and M. A. Moritz. 2009. Environmental controls on the distribution of wildfire at multiple spatial scales. Ecological Monographs 79:127–154.

Parisien, M.-A., S. Snetsinger, J. A. Greenberg, C. R. Nelson, T. Schoennagel, S. Z. Dobrowski, and M. A. Moritz. 2012. Spatial variability in wildfire probability across the western United States. International Journal of Wildland Fire 21:313–327.

Perry, D. A., P. F. Hessburg, C. N. Skinner, T. A. Spies, S. L. Stephens, A. H. Taylor, J. F. Franklin, B. McComb, and G. Riegel. 2011. The ecology of mixed severity fire regimes in Washington, Oregon, and Northern California. Forest Ecology and Management 262:703–717.

Peterson, K. F., B. N. I. Eskelson, V. J. Monleon, and L. D. Daniels. 2019. Surface fuel loads following a coastal–transitional fire of unprecedented severity: Boulder Creek fire case study. Canadian Journal of Forest Research 48:925–932.

Pew, K. L., and C. P. S. Larsen. 2001. GIS analysis of spatial and temporal patterns of human-caused wildfires in the temperate rain forest of Vancouver Island, Canada. Forest Ecology and Management 140:1–18.

Pflugmacher, D., W. B. Cohen, and R. E. Kennedy. 2012. Using Landsat-derived disturbance history (1972–2010) to predict current forest structure. Remote Sensing of Environment 122:146–165.

Prestemon, J. P., D. T. Butry, K. L. Abt, and R. Sutphen. 2010. Net benefits of wildfire prevention education efforts. Forest Science 56:181–192.

Pufahl, A., and C. R. Weiss. 2009. Evaluating the effects of farm programmes: results from propensity score matching. European Review of Agricultural Economics 36:79–101.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org

Raymond, C. L., S. Healey, A. Peduzzi, and P. Patterson. 2015. Representative regional models of post-disturbance forest carbon accumulation: integrating inventory data and a growth and yield model. Forest Ecology and Management 336:21–34.

Reilly, M. J., C. J. Dunn, G. W. Meigs, T. A. Spies, R. E. Kennedy, J. D. Bailey, and K. Briggs. 2017. Contemporary patterns of fire extent and severity in forests of the Pacific Northwest, USA (1985–2010). Ecosphere 8:e01695.

Rosenbaum, P. R. 1987. Sensitivity analysis for certain permutation inferences in matched observational studies. Biometrika 74:13–26.

Rosenbaum, P. R., and D. B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70:41–55.

Rosenbaum, P. R., and D. B. Rubin. 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. American Statistician 39:33–38.

Ryan, A. M., J. F. Jr Burgess, and J. B. Dimick. 2015. Why we should not be indifferent to specification choices for difference-in-differences. Health Services Research 50:1211–1235.

Samal, A., S. Seth, and K. Cueto 1. 2004. A feature-based approach to conflation of geospatial sources. International Journal of Geographical Information Science 18:459–489.

SAS Institute. 2017. SAS/STAT 9.4 User's Guide. SAS Institute, Cary, North Carolina, USA.

Sharma, C. M., S. Gairola, N. P. Baduni, S. K. Ghildiyal, and S. Suyal. 2011. Variation in carbon stocks on different slope aspects in seven major forest types of temperate region of Garhwal Himalaya, India. Journal of Biosciences 36:701–708.

Smith, J. A., and P. E. Todd. 2005. Does matching overcome LaLonde's critique of nonexperimental estimators? Journal of Econometrics 125:305–353.

Smith, J. E., L. S. Heath, K. E. Skog, and R. A. Birdsey. 2006. Methods for calculating forest ecosystem and harvested carbon with standard estimates for forest types of the United States. General Technical Report NE-343. U.S. Department of Agriculture, Forest Service, Northeastern Research Station, Newtown Square, Pennsylvania, USA.

Smith, W. B. 2002. Forest inventory and analysis: a national inventory and monitoring program. Environmental Pollution 116:S233–S242.

Stuart, E. A. 2010. Matching methods for causal inference: A review and a look forward. Statistical Science 25:1–21.

Stuart, E. A., and K. M. Green. 2008. Using full matching to estimate causal effects in nonexperimental studies: examining the relationship between adolescent marijuana use and adult outcomes. Developmental Psychology 44:395–406.

Stuart, E. A., and D. B. Rubin. 2004. Matching methods for causal inference: designing observational studies. Harvard University Department of Statistics Mimeo, Cambridge, Massachusetts, USA.

Verbitsky-Savitz, N., and S. W. Raudenbush. 2012. Causal inference under interference in spatial settings: a case study evaluating community policing program in Chicago. Epidemiologic Methods 1:107–130.

Weiskittel, A. R., J. A. Kershaw, J. K. Vanclay, and D. W. Hann. 2011. Forest growth and yield modeling. Wiley-Blackwell, Chichester, UK.

Whittaker, W., L. Anselmi, S. R. Kristensen, Y.-S. Lau, S. Bailey, P. Bower, K. Checkland, R. Elvey, K. Rothwell, and J. Stokes. 2016. Associations between extending access to primary care and emergency department visits: a difference-in-differences analysis. PLoS Medicine 13:e1002113.

Whittier, T. R., and A. N. Gray. 2016. Tree mortality based fire severity classification for forest inventories: a Pacific Northwest national forests example. Forest Ecology and Management 359:199–209.

Wimberly, M. C., and Z. Liu. 2014. Interactions of climate, fire, and management in future forests of the Pacific Northwest. Forest Ecology and Management 327:270–279.

Woodall, C. W., and V. J. Monleon. 2008. Sampling protocol, estimation, and analysis procedures for the down woody materials indicator of the FIA program. General Technical Report NRS-22. U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, Pennsylvania, USA.

Woodall, C. W., C. H. Perry, and J. A. Westfall. 2012. An empirical assessment of forest floor carbon stock components across the United States. Forest Ecology and Management 269:1–9.

Yuan, L. L. 2010. Estimating the effects of excess nutrients on stream invertebrates from observational data. Ecological Applications 20:110–125.

Zheng, D., J. Rademacher, J. Chen, T. Crow, M. Bresee, J. Le Moine, and S.-R. Ryu. 2004. Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. Remote Sensing of Environment 93:402–411.

Supporting Information

Additional supporting information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/eap.2283/full

Data Availability

Forest inventory and image data sources are provided in Appendix S2: Table S1. Data are available in an open repository, with the exception of plot locations due to confidentiality issues; users needing the exact plot coordinates may contact the FIA program directly as outlined in Appendix S2.