Environmental Modelling and Software 144 (2021) 105163

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft

Sensitivity analysis on distance-adjusted propensity score matching for wildfire effect quantification using national forest inventory data

Hyeyoung Woo[a,*], Bianca N.I. Eskelson[a], Vicente J. Monleon[b]

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

ARTICLE INFO

Keywords:
Data availability
Sample size
Environmental covariates
Natural disturbance
Quasi-experimental method
Monte Carlo simulation

ABSTRACT

Propensity score matching (PSM) and distance-adjusted PSM enable estimation of causal effects from observational data by selecting controls that are similar to treated observations in terms of environmental covariates and spatial locations. Quantifying effects of natural disturbances such as wildfires often encounters limited availability of observational data due to the scarcity of ecological events or cost of data collection over large areas. Using empirical data of national forest inventory plot measurements, we conducted a sensitivity analysis of distance-adjusted PSM on data availability for wildfire effect quantification. Using Monte Carlo simulations, we assessed the influence of sample size and covariates on the balance of propensity score distributions and the performance of effect estimates. The inclusion of the distance measure in matching compensated for the omission of key covariates. This study provides a practical guide of determining sample size and covariates for matching with spatial information to analyze ecological data.

1. Introduction

Natural disturbances such as wildfires, disease and insect outbreaks drive changes in the forest environment on a large scale. Wildfires, one of the major natural disturbance agents, cause the instant removal and redistribution of forest carbon (Dixon et al., 1994; Keith et al., 2014) as well as long-term effects such as delayed tree mortality (Carlson et al., 2012) and carbon losses through decomposition of dead trees (Eskelson et al., 2016). However, the effects of natural disturbances such as wildfire on forest dynamics are difficult to estimate due to the lack of an experimental setting for natural disturbances. Quantification of the impacts of natural disturbances and other ecological processes has been largely based on statistical modelling using observational data (Liu et al., 2011) instead of randomized experiments that control for confounding factors. Recent studies suggest quasi-experimental approaches (Greenstone and Gayer 2009) for impact quantification of wildfires (Butsic et al., 2017a; Larsen et al., 2019; Woo et al., 2021) to allow causal inference from observational studies. Quasi-experimental methods attribute causal relationships between a treatment (e.g., natural disturbance) and outcome (e.g., change in forest environment) from observational data by specifying a control group that is as similar as possible to the treatment group based on potential confounding factors (Campbell and Stanley 2015).

Propensity score matching (PSM) is one of the quasi-experimental methods that enables cause-and-effect analysis from observational data (Rosenbaum and Rubin 1983). A set of observations that were affected by a natural disturbance (i.e., treated group) is matched to observations that were not affected (i.e., untreated or raw control group) based on propensity scores. The propensity score denotes the probability of experiencing the natural disturbance conditional on the environmental covariates (Rosenbaum and Rubin 1983). Balancing the propensity score distributions between treated and control groups removes the possible selection bias introduced by the environmental covariates, leading to a similar setting as that of a randomized experiment, which assigns a known probability of treatment to the treated and control groups (Stuart and Rubin 2004). The matched observations (i.e., control group) with balanced propensity score distributions allow unbiased estimation of the outcome of interest (Stuart 2010; Austin 2009).

Ecological data are often associated with geographical locations, and the spatial information of the data can reflect some of the confounding variables that were unmeasured (Dray et al., 2006). Because PSM including geographical coordinates as covariates did not ensure the spatial proximity between treated and untreated groups, distance-adjusted propensity score matching (DAPSM) has been

* Corresponding author.
E-mail address: hyeyoung.woo@alumni.ubc.ca (H. Woo).

https://doi.org/10.1016/j.envsoft.2021.105163
Accepted 6 August 2021
Available online 10 August 2021

suggested by Papadogeorgou et al. (2019). DAPSM incorporates spatial distance measures in addition to the estimated propensity scores to account for unobserved covariates. The application of PSM and DAPSM to ecological data, however, was introduced only recently (e.g., Butsic et al., 2017a; Woo et al., 2021), so the performance of these matching methods in terms of balance between treated and control groups and unbiasedness of the estimates under different data availability scenarios is still unknown. Unlike large survey data from health science and econometrics (e.g., >30,000 treated and >60,000 untreated from Smeeth et al., 2009; >3000 treated and >100,000 untreated from Huber et al., 2013), ecological data sets associated with natural disturbances are often limited in size due to the scarcity of the natural disturbance events (e.g., wildfire, disease, or insect outbreak) or the cost of data collection over large spatial and temporal scales (Kremens et al., 2010). In matching, availability of data can be identified based on the number of treated observations, raw control observations, and environmental covariates to estimate the propensity scores. Limited data on any of these may affect the reliability of the causal effect quantification (Stuart 2010). The number of treated observations is directly related to the variance of the estimated treatment effect (Brookhart et al., 2006), and the number of raw control observations relative to the treated observations is associated with the bias in propensity scores (Heinrich et al., 2010). The availability of covariates influences the exclusion of important confounding variables when estimating the propensity scores (Heckman et al., 1997). Given that matching based on limited data may impose more severe bias on the effect estimation than regression adjustments (Stuart 2010), it is critical to understand the performance of matching methods under different data availability scenarios and identify a marginal size of data and covariate set required to implement matching methods.

This study provides a sensitivity analysis of matching on data availability in terms of the balance achieved in propensity score distributions and of the estimated effect using a case study of wildfire effect estimation. Woo et al. (2021) quantified the effect of wildfires on aboveground live woody forest carbon using DAPSM and PSM based on an empirical dataset of 23,150 forest inventory plot measurements from Washington and Oregon, United States of America (USA). Taking advantage of the large number of forest inventory plot measurements used in Woo et al. (2021), we performed a Monte Carlo simulation to generate subsets of the full dataset that contained 611 burned plots (i.e., treated) and 22,539 unburned plots (i.e., raw control). The propensity scores of the burned and unburned plots were estimated based on sixteen covariates that describe the different environment of plots, with and without distance adjustment. We examined three scenarios of data availability: 1) number of burned plots, 2) ratio of unburned plots relative to burned plots, and 3) availability of environmental covariates. For 3), we considered availability of the covariates based on data sources to define four groups: topography, climate, land cover, and remote-sensing variables. We first assessed how well the propensity score distributions were balanced between the burned and the control plots through matching. We then observed how the estimated wildfire effect changed under the different scenarios. The effects of wildfire were quantified as the difference in live woody carbon mass between the burned and the control plots. We presented the balance achieved and effect estimates under DAPSM and PSM to assess the bias and variability expected within each matching method and data availability scenario.

By providing a sensitivity analysis of DAPSM and PSM based on real-world data, the results of this study suggest practical guidelines of data availability to be considered for wildfire effect quantification and, by extension, estimation of other natural disturbance (e.g., Arovaara et al., 1984) and management effects (e.g., Butsic et al., 2017b). Utilization of national forest inventory data will broaden the opportunity to conduct effect analyses over large spatial and temporal scales with spatially balanced samples of forest conditions. The diagnostics of balance achievement by matching presented in this study are applicable for designing future quasi-experiments where the spatial distances as well as the environmental covariates play an important role in distinguishing treated and raw control observations. By implementing matching methods as quasi-experimental approach, causal effects can be estimated from observational data to inform ecological problems for which experimental settings are impractical.

2. Data

2.1. Forest inventory data

We used national forest inventory data of Washington and Oregon, USA, from the United States Department of Agriculture Forest Service Forest Inventory and Analysis (FIA) program (Smith 2002). Since 2000, the FIA program has collected field measurements of forested lands (i.e., at least 10 percent potential cover by live trees) by measuring permanent plots (4047-square meter in size) that represent a spatially balanced sample of one plot for every 2400 ha (Bechtold and Patterson 2005). A plot consists of four circular subplots, and the measurements of trees and environmental attributes are made at both plot and subplot level. Ten percent of the plot locations are visited annually, resulting on a 10-year remeasurement cycle (Bechtold and Patterson 2005).

Between 2001 and 2016, there were a total of 23,150 inventory plot measurements in Washington and Oregon forests including 7994 remeasurements. We classified the plot measurements into burned (i.e., treated, *t*) and unburned (i.e., raw control, *rc*) based on the Monitoring Trends in Burn Severity (MTBS, http://www.mtbs.gov/) map. Burned plots were identified by spatially overlaying the plot locations on the fire maps. Every FIA plot measurement that fell within the perimeter of a wildfire that occurred less than five years prior to the measurement was labeled as burned. Meanwhile, plots that had not burned at least ten years prior to measurement were labeled as unburned. Here we refer to the burned measurements as 'burned plots' and unburned measurements as 'unburned plots' for simplicity. Thus, unburned plots may have repeated measurements at the same location. To remove dependencies among burned and unburned plots, previous measurements of burned plots were excluded from the pool of raw control plots. Similarly, if a plot burned twice, the measurements after the first wildfire were also excluded to preclude the effect of reburn (Woo et al., 2021). A total of 611 burned plots and 22,539 unburned plot measurements were used for the analysis. More details on the compilation of the data are available in Woo et al. (2021).

2.2. Environmental covariates

For each forest inventory plot, we retrieved sixteen environmental covariates that are associated with the probability of wildfire occurrence. These environmental covariates were grouped into four categories: 1) topography (TOPO), 2) climate (CLIM), 3) land cover (LAND), and 4) remotely-sensed variables (REMO) (Table 1; additional details in Woo et al., 2021).

Topography and land cover variables of the forest inventory plots were measured in the field and obtained from the FIA database. Topography covariates included elevation, slope, and aspect. Elevation was measured at the center of the plot, and slope and aspect were measured at subplot level according to the land classification within the subplot (e.g., forested/non-forested land). We used the values of slope and aspect that accounted for the largest percentage of the forested land classification within the subplots. Land cover covariates including land ownership, physiographic condition, and proportion of forested area (Table 1) were measured at subplot level. For the class variables, we used two categories of land ownership (public and private) and three of physiographic class (mesic, xeric, and hydric) (Woo et al., 2021). We assigned the land cover categories that accounted for the largest percentage within the four subplots.

Climate covariates consisted of six PRISM-derived variables related to precipitation, temperature, vapor pressure, and moisture stress (Daly

Table 1
Descriptions and sources of sixteen environmental variables used for matching burned and unburned forest inventory plots.

Variable	Abbr.	Measurement unit/Description
Topography (TOPO)		
Elevation	ELEV	Metres
Slope	SLOPE	Degrees
Aspect	ASPECT	Cosine value, degrees azimuth
Land cover (LAND)		
Proportion of forested area	PROPforest	% of plot area with forested condition (at least 10 percent potential cover by live trees)
Owner group code	OWNGC	Categorical: Public, Private
Physiographic class	PHYSCL	Categorical: Xeric, Mesic, Hydric
Climate (CLIM)		
Annual precipitation	PRECP	ln mm * 100
Mean annual temperature	TEMP	degrees C * 100
Coefficient of variation in precipitation	CVPRECP	Variation between mean monthly precipitation of December and July (wettest and driest months)
Mean temperature difference	TEMPDIF	degrees C * 100, difference between mean August maximum temperature and mean December minimum temperature
Mean summer vapor pressure deficit	SVPD	hPa
Growing season moisture stress	GSMS	degrees C/ln mm, ratio of temperature to precipitation from May–September
Remotely-sensed data (REMO)		
Normalized difference in vegetation index	NDVI	
Tasseled cap brightness	TC1	
Tasseled cap greenness	TC2	
Tasseled cap wetness	TC3	

et al., 2008). The climate covariates were derived from thirty-year normal data (1981–2010) with 30-m resolution. The remote-sensing data included four spectral and vegetation indices including the normalized difference vegetation index (NDVI), and tasseled cap brightness, greenness, and wetness in the middle of growing season, all derived from annual Landsat satellite images with 30-m resolution. Unburned plots were associated with the remotely-sensed variables in the year that the plot was measured, and burned plots were associated with the variables measured in a year before the wildfire to account for stand characteristics that may be related to the risk of burning. To assign the climate and remotely-sensed covariates to each of the plot measurements the values of 9 pixels around the plot center location were averaged. The climate and remotely-sensed metrics were obtained from the Landscape Ecology, Modeling, Mapping, and Analysis (LEMMA) team (https://lemma.forestry.oregonstate.edu/).

2.3. Aboveground live woody carbon

We obtained aboveground live woody carbon mass per plot (Mg ha^{-1}) for both burned and unburned plots, based on all live trees over 2.54 cm diameter at breast height (DBH) measured within the forest inventory plot by the FIA protocol (Bechtold and Patterson 2005). Aboveground woody biomass including stem, bark, and live branches was estimated by the Pacific Northwest Research Station, USDA Forest Service (Woodall et al., 2012). We converted the aboveground live woody biomass into carbon mass using a carbon ratio multiplier of 0.5 (IPCC 2006). The obtained carbon mass was summarized as per ha values (Mg ha^{-1}) for each plot, and used as the outcome variable in propensity score matching to assess the effect of wildfires. The estimated wildfire effect is the difference in the outcome variable between burned and unburned plots.

3. Matching methods

3.1. PSM and DAPSM for assessing wildfire effects on aboveground live woody carbon

Propensity scores in PSM denote the probability of receiving the treatment conditional on the environmental covariates. The probability assigned to each observation – in this study, the forest inventory plot – is generally estimated using logistic regression. Distance-adjusted propensity scores in DAPSM consider spatial distances between burned and unburned plot locations in matching. They are a weighted average between the propensity scores estimated from environmental covariates and the normalized Euclidean distances between plot locations (Papadogeorgou et al., 2019). Because the spatial distance may take into account unobserved covariates that were not included in the covariate set to estimate the propensity scores, DAPSM should be preferred to PSM or spatial matching (SM) that considers only the distances for matching (Woo et al., 2021).

The quantification of wildfire effects on aboveground live woody carbon using the matching methods followed three steps: First, we estimated propensity scores of burned and unburned plots using logistic regression with the sixteen environmental covariates. Then, we estimated the distance-adjusted propensity scores as a linear combination of the propensity scores and spatial distances between pairs of burned and unburned plots. We used a weight of 0.5 to equally emphasize the observed environmental covariates and the spatial distance, given an earlier study that found other weights (i.e., 0.25, 0.5, and 0.75) did not result in any difference in the effect estimation (Woo et al., 2021). Finally, we matched the burned plots and unburned plots using PSM and DAPSM to obtain a set of control plots that share similar covariates and geographical locations. We employed matching without replacement with a greedy algorithm (Austin 2011) that sequentially finds the smallest difference in the distance-adjusted propensity scores between the burned and the unburned plots. This algorithm is the most commonly used in propensity score matching implementation (Austin and Cafri 2020). The matched unburned plots were identified as control plots. The estimates of the wildfire effects were then quantified as the difference in aboveground live woody carbon mass between the burned and control plots.

3.2. Sensitivity analysis of DAPSM under different data availability scenarios

We ran Monte Carlo simulations to perform a sensitivity analysis on the balance of propensity score distributions as well as effect estimates when reducing the number of treated and raw control observations and covariate sets. We created random subsets of forest inventory plot data from the full data consisting of 611 burned plots (N_t=611) and 22,539 unburned plots (N_{rc}=22,539). The full data was set as our reference population for the Monte Carlo simulations (i.e., simulation population), and the propensity scores estimated based on all available environmental covariates using the simulation population were considered to be the true propensity scores in our simulation. Because the number of possible combinations of all covariates is very large, we grouped the environmental covariates sharing similar attributes and sources as stated in Section 2.2 based on the idea that the attainability of variables is dependent on the accessibility to the data source. For example, TOPO including elevation, slope, and aspect are easily available for researchers through open sources such as digital elevation models (DEM) or geographical information systems (GIS) for most regions (Bungartz et al., 2018). The order of accessibility to the four groups of covariates in our study was identified as TOPO-CLIM-REMO-LAND from the least to the most costly to obtain. We considered four scenarios of environmental covariate sets for propensity score estimation: 1) full set of covariates with all four groups TOPO + CLIM + REMO + LAND, 2) TOPO + CLIM + REMO, 3) TOPO + CLIM, and 4) TOPO + REMO

Table 2
Simulated data availability scenarios and their values and conditions for DAPSM and PSM. SM simulated n_t and n_{rc}/n_t only.

Simulated scenarios	Values/Conditions
Number of burned plots (n_t)	500, 250, 150, 100, 75, 50, 25
Matching ratio (n_{rc}/ n_t)	30, 25, 20, 15, 10, 5, 4, 3, 2
Covariate set	1) TOPO + CLIM + REMO + LAND– full set of environmental covariates 2) TOPO + CLIM + REMO 3) TOPO + CLIM 4) TOPO + REMO

(Table 2).

We examined three simulation scenarios of data availability: number of burned plots (n_t), number of unburned plots relative to burned plots (n_{rc}/n_t), and environmental covariate set. We varied numbers of burned plots (n_t) and unburned plots relative to burned plots (n_{rc}/ n_t) for the random subsets of data from the simulation population (Table 2) under each of the four covariate set scenarios. From each combination of subsets and covariate sets, the plots were matched using different matching methods: DAPSM based on both the propensity scores and plot locations, and PSM based only on the propensity scores. Additionally, we implemented SM by matching the plots based only on the distances among the plot locations. The process was iterated 100 times for each combination of simulation values and conditions under different matching methods. We observed the propensity score distributions and the effect estimates under the data availability scenarios. We also computed mean distance to the control plots from burned plots under each matching method: DAPSM, PSM, and SM.

3.3. Evaluation of results

3.3.1. Descriptive statistics for evaluation of balance in propensity scores

The quality of the matching was assessed by the balance in propensity scores between treated and control groups (McCaffrey et al., 2013; Stuart 2010). Here, balance means that the distribution of propensity scores of the matched pairs were similar, thus the matching reduced the potential selection bias between treated and raw control groups caused by the environmental covariates included in the model. Rubin (2001) suggested three statistics to assess balance: 1) the difference in the means of the propensity scores between the treated and control groups, 2) the ratio of the variances of the propensity scores, and 3) the ratio of the variances of the residuals of the covariates after matching (Table 3). 1) and 2) represent the first and second moment of the overall propensity score distribution, respectively. The ratio of the variances of the residuals of the covariates is obtained by regressing each of the covariates on the estimated propensity scores. If the matching was successful and the control group is similar to the treated group, the means of difference and the variance ratios of the propensity scores should be close to zero and one, respectively. Ratio of variance between 0.8 and 1.2 is considered adequate, while values below 0.5 or over 2 are too extreme (Rubin 2001). For the ratio of the variances of the residuals of the covariates, we computed the percentage of covariates that are within each of the variance ratio ranges (i.e., $\leq$0.8 and $<$1.2, $<$0.5 or $\geq$ 2) for the 16 variables considered, even if the model only included a subset of covariates.

Table 3
Baseline statistics of bias and variance ratio in propensity scores computed from the simulation population (N_t=611, N_{rc}/N_t = 36.7) and controls obtained from DAPSM with full covariate set including TOPO, LAND, REMO, and CLIM.

	Raw control	Matched control
	N_t = 611, N_{rc} = 22,359	N_t = 611, N_c = 611
1) Bias in mean PS: E(PS_c) - E(PS_t)	0.031	0.001
2) Variance ratio in PS: Var(PS_c)/Var(PS_t)	3.108	1.037
3) Variance ratio of the residuals for individual covariates	Percent of covariates for which variance ratio of the residuals falls within specified range	
$<$0.5	0	0
$\geq$0.5 and $<$0.8	18.75	0
$\leq$0.8 and $<$1.2	50	100
$\geq$1.2 and $<$2	31.25	0
$\geq$2	0	0

E: expected value, PS: propensity score, c: control group, and t: treated group.

In addition to the statistics suggested by Rubin (2001), we compared the absolute standardized mean difference (ASMD) in the sixteen covariates between the unmatched and matched burned and unburned plots to examine the balance at the individual covariate level (Nolte et al., 2017; Austin 2009). ASMD was computed as the average difference standardized by the standard deviation of the variable for continuous covariates and using the average difference in proportions for categorical variables (Stuart 2010). ASMD values less than 0.2 indicate adequate balance in the covariates, while ASMD values between 0.2 and 0.4 indicate moderate imbalance and ASMD values over 0.4 indicate large imbalance (Cohen 2013).

Before matching, the bias and variance ratio of the estimated propensity scores using the full population of burned and unburned forest inventory plots (N_t=611 and N_{rc}/N_t = 36.7) and the full covariate set (16 variables; TOPO + CLIM + REMO + LAND) was very large (Table 3). After DAPSM, the bias decreased by approximately 97% (from 0.031 to 0.001). The variance ratio decreased from 3.11 to a value very close to 1. Thus, DAPSM with all 16 covariates resulted in a control group similar to the burned plots in terms of propensity scores (Rubin 2001). Regarding the ratio of the variance of the residuals of the covariates, 50% of the covariates (8 out of 16) showed relatively large variance ratios (e.g., $<$0.8 or $\geq$ 1.2) between treated and raw control, meaning that there was imbalance between burned and unburned plots for those covariates (Table 3). When DAPSM was applied, the variance ratios ranged within 0.8–1.2 for all covariates (100%, 16 out of 16) (Table 3). These three statistics obtained from DAPSM and the full data set were the baseline to compare bias and variance ratios from the subsets of data and covariate set using DAPSM and PSM. We graphically presented the comparisons of the descriptive statistics along with the combinations of the three simulation scenarios for effective visualization (Pianosi et al., 2016).

3.3.2. Performance of matching estimates

The bias reduction and variability of matching estimates was examined using the distributions of the simulated wildfire effect estimates computed under varying sample sizes and covariate sets. The wildfire effects were calculated as the difference in the aboveground live woody carbon mass between the burned and control plots. We obtained the baseline estimate of wildfire effects on aboveground live woody carbon mass based on the simulation population to benchmark the effect estimation. Based on the simulation population, the difference in the mean aboveground live woody carbon mass between the burned and the control plots was estimated as 30.95 Mg ha^{-1} in absolute value. The difference in the mean carbon mass between burned and raw control plots was 50.82 Mg ha^{-1}, so that the wildfire effect on aboveground live woody carbon mass was overestimated by 19.87 Mg ha^{-1} on average due to the selection bias introduced by the different environmental conditions between plots that burn and those that do not. We presented the simulated wildfire effect estimates from the subsets of data against the baseline estimate to assess the mean bias and the variability of the matching estimates. Also, we computed relative bias of the simulated wildfire effect estimates compared to the simulation population under different data availability scenarios. The simulated estimates and their relative bias were displayed as boxplots and matrix plots, respectively.

4. Results

4.1. Balance in propensity scores under different data availability scenarios

4.1.1. Smaller mean bias in propensity scores under larger n_t and covariate set with CLIM

The three scenarios of data availability in matching (i.e., n_t, n_{rc}/ n_t, and covariate set) jointly affected the mean bias in propensity scores under DAPSM. A notable change in the mean bias in propensity scores was found among different covariate sets. While there was little deviation from the baseline bias under DAPSM with the full covariate set including TOPO, CLIM, REMO, and LAND (<0.003, Fig. 1a), dropping LAND variables (Fig. 1b) and CLIM (Fig. 1c) resulted in larger bias on average: 0.005–0.007 and 0.007–0.019, respectively. Excluding REMO from the covariate set also increased the bias but not much (e.g., <0.002 on average; Fig. S1 in Supplementary materials). The mean bias in propensity scores was relatively constant across n_t, with a minimal increase in bias as n_{rc}/n_t decreased to 10 (Fig. 1a and b). When the covariate set excluded CLIM variables (Fig. 1c), however, the increase of bias was remarkable compared to the covariate sets including CLIM, especially when n_{rc}/n_t was less than 10. In contrast to the covariate sets including CLIM (Fig. 1a and b), the bias increased with decreasing n_t for the covariate set that excluded CLIM variables (Fig. 1c).

Comparing DAPSM to PSM shows that including distance in the matching process did not result in large differences in mean bias in propensity scores as long as CLIM variables were included in the covariate set (Fig. 1a vs. 1d and Fig. 1b vs. 1e). Without CLIM variables, PSM produced much larger bias than DAPSM regardless of n_{rc}/n_t (Fig. 1f). Without any environmental covariates included in the matching process (i.e., SM with distance only; Fig. 1e), the mean bias showed a similar pattern to DAPSM without CLIM (Fig. 1c): small increase in mean bias on average and a sharp increase in mean bias when $n_{rc}/n_t<10$.

4.1.2. Smaller variance ratio in propensity scores under large n_t and covariate set with CLIM

The variance ratio in propensity scores between the burned and the control plots generally followed similar patterns as the mean bias in propensity scores: matching based on covariate sets that excluded CLIM variables led to a large increase in the mean variance ratio with decreasing n_t and n_{rc}/n_t (Fig. 2). When using the full covariate set, the variance ratio fell within the range of 0.8–1.2 (i.e., close to 1) for most combinations of n_t and n_{rc}/n_t except for n_t less than 25 and n_{rc}/n_t less than 5 (Fig. 2a). Dropping the LAND variables resulted in variance ratios larger than 1.2 for the majority of combinations of n_t and n_{rc}/n_t, but not in extreme values exceeding 2 (Fig. 2b). Excluding CLIM variables from the covariate set introduced large variability in the variance ratio among different n_t and n_{rc}/n_t (Fig. 2c). Small n_t (i.e., <75) and n_{rc}/n_t (i.e., ≤5) exhibited extreme variance ratios over 2, indicating imbalance in the

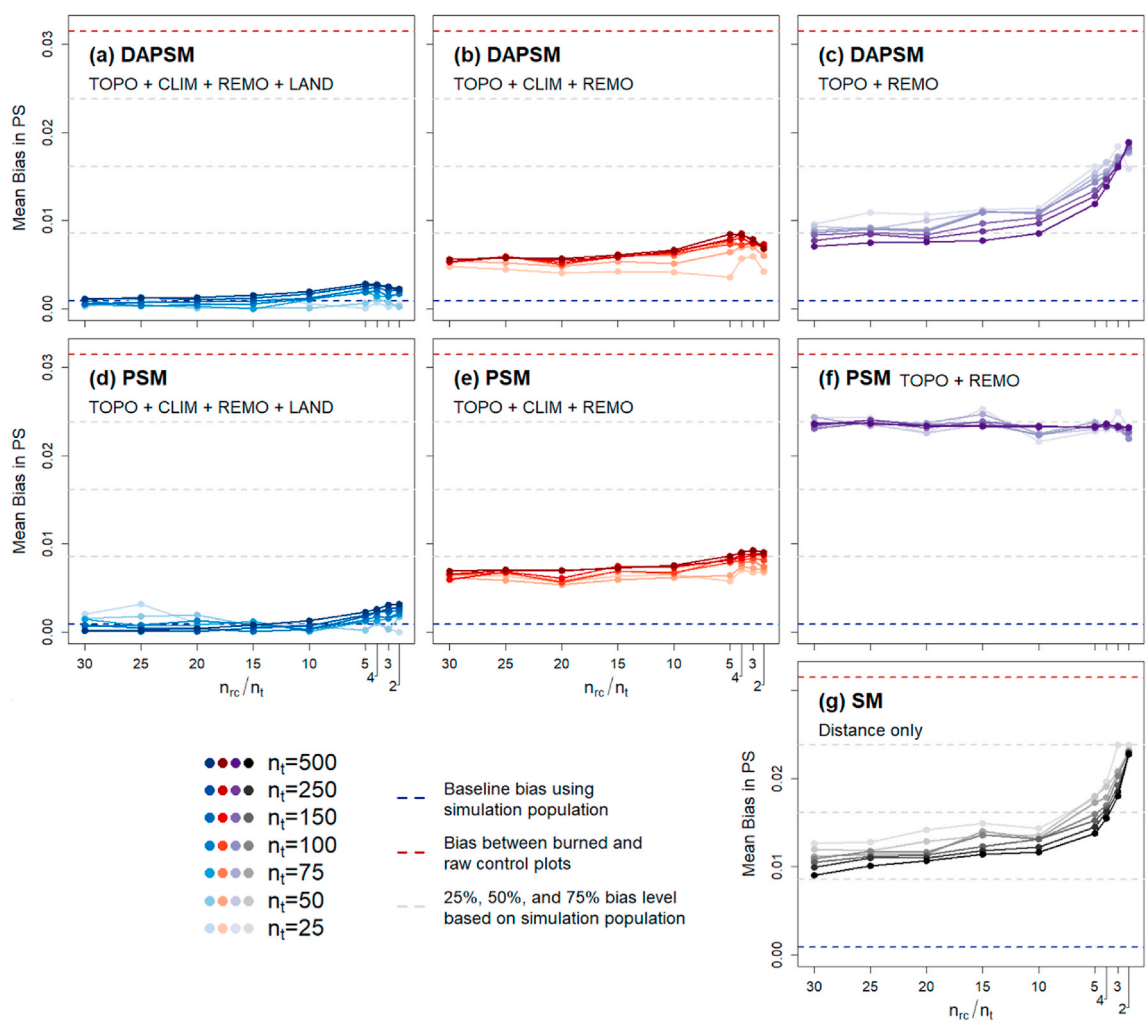

Fig. 1. Mean bias in propensity scores (PS) under different number of burned plots (n_t), unburned plots to burned plots (n_{rc}/n_t), and covariate set. The burned and unburned plots were matched based on each combination of simulation scenarios using (a)–(c) DAPSM, (d)–(f) PSM, and (g) SM.

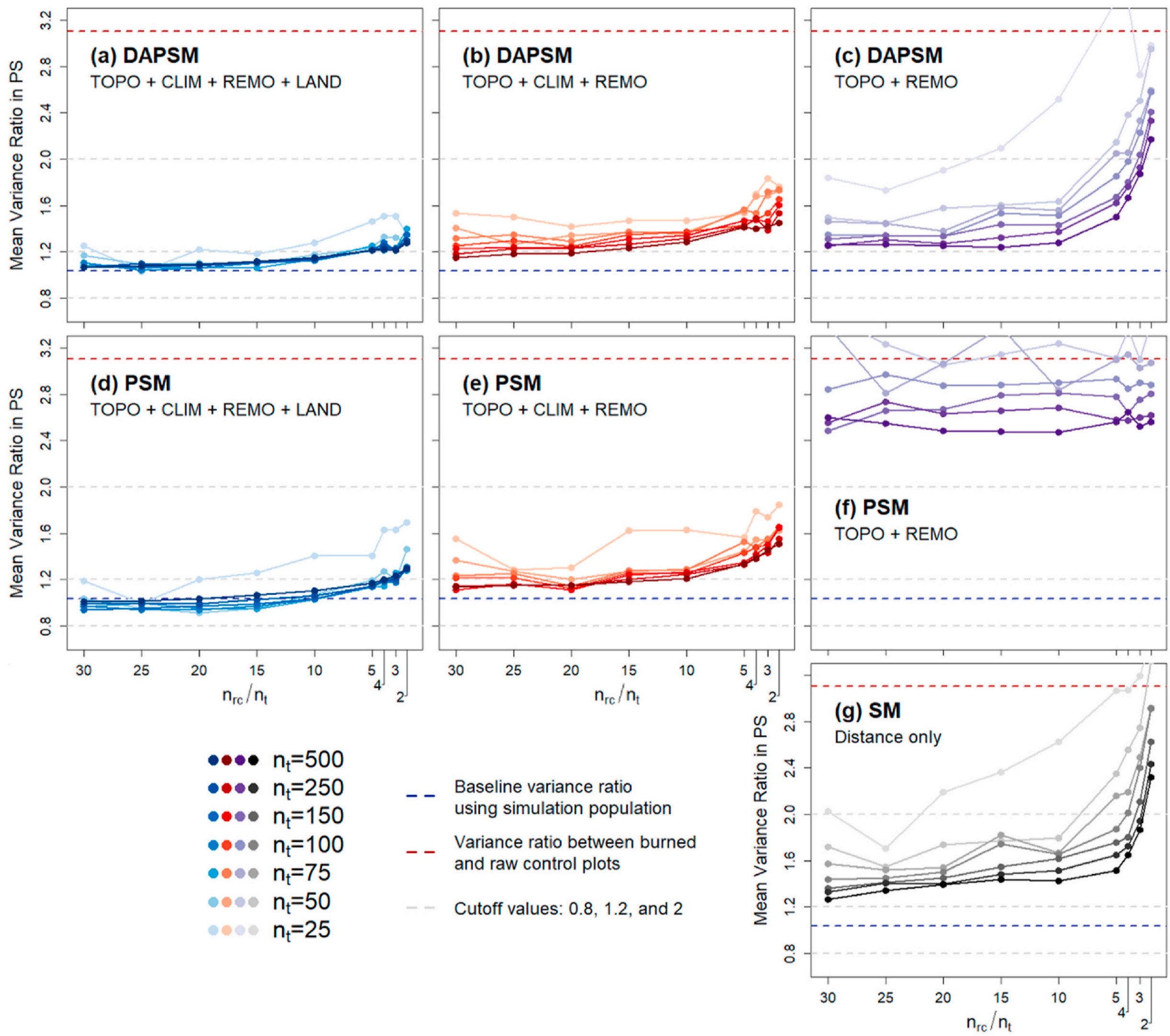

Fig. 2. Mean variance ratio in propensity scores (PS) between burned and control plots under different combinations of number of burned plots (n_t), unburned plots to burned plots (n_{rc}/n_t), and covariate set, matched based on (a)–(c) DAPSM, (d)–(f) PSM, or (g) SM. The cutoff values of variance ratio from Rubin (2001): values between 0.8 and 1.2 are considered to be adequate, while values > 2 are extreme. Values outside the range of the graphs are available in Fig. S2 in the supplementary materials.

distributions of propensity scores between burned and control plots.

PSM produced slightly smaller variance ratios than DAPSM did when CLIM was included in the covariate set (Fig. 2a vs. 2d, Fig. 2b vs. 2e). PSM without CLIM, however, did not reduce the variance in propensity scores between burned and unburned plots (red lines in Fig. 2f). Once distance was included in the matching process, however, the variance ratio substantially dropped, especially for large n_t (≥100) and n_{rc}/n_t (≥15) values even if no other covariates were included (Fig. 2g).

4.1.3. Smaller variance ratio of the residuals of the covariates under DAPSM

Comparing raw control to controls obtained through DAPSM and PSM, the improvement in the variance ratio of residuals of individual covariates through matching was obvious in all covariate sets, showing increase in the percentages of variance ratio between 0.8 and 1.2 (Fig. 3a–d). DAPSM exhibited larger percentages of variance ratio between 0.8 and 1.2 than PSM did, indicating the method was more effective achieving balance in terms of individual covariates. SM based on distance only was also more balance in individual covariates than PSM, especially when n_{rc}/n_t>10 (Fig. 3d).

As n_t decreased, increasingly extreme variance ratios of the residuals were observed for the matched plots, regardless of n_{rc}/n_t (Fig. 3a–d). n_t = 25 barely reduced the variance ratio for any n_{rc}/n_t through matching. When climate variables (CLIM) were included in the covariate set, the distribution of variance ratios obtained through both DAPSM and PSM did not differ much among the covariate sets (Fig. 3a–c). TOPO and REMO variables, however, showed little initial imbalance in the variance ratio of residuals between burned and raw control plots (Fig. 3d) compared to other covariate sets. In that case, PSM introduced more severe imbalance in variance ratio than the raw controls (Fig. 3d).

4.1.4. DAPSM balances some covariates that were not included in propensity score model

The average ASMD computed for each covariate showed major differences between DAPSM and PSM under the same covariate set (Table 4, Fig. S4). Comparing the raw controls of unburned plots to the burned plots (i.e., unmatched), 11 out of 16 covariates were imbalanced (i.e., ASMD > 0.2). Matching based on the full covariate set achieved balance in all covariates under both DAPSM and PSM. When the covariate sets were reduced, PSM resulted in imbalance in the omitted covariates (e.g., PHYSCL, TEMP, TEMPDF, SVPD, and GSMS under PSM with TOPO + REMO; Table 4). On the other hand, DAPSM accomplished

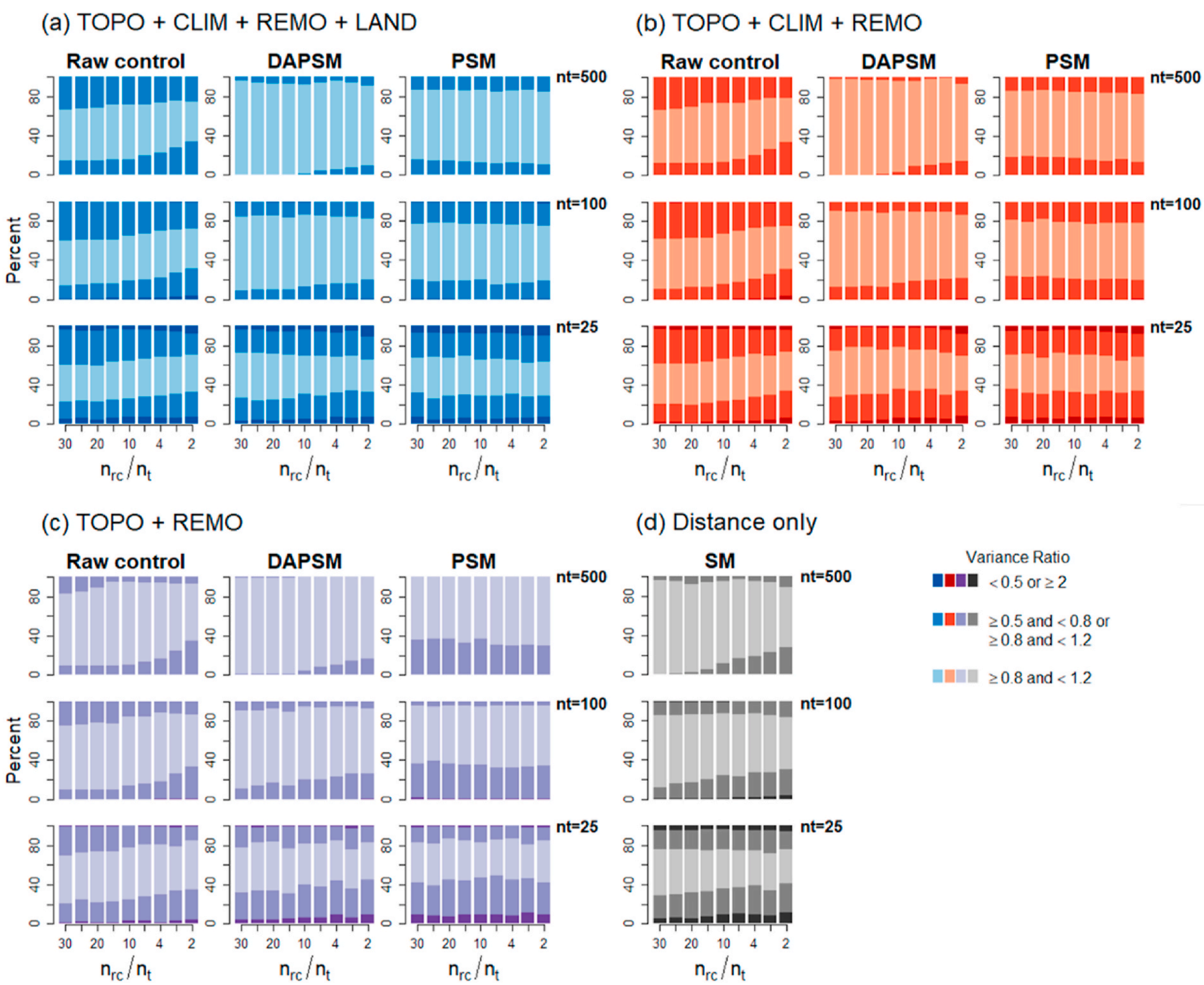

Fig. 3. Percentages of the number of covariates of which residual variance ratio falls within each range under the covariate set (a) TOPO + CLIM + REMO + LAND, (b) TOPO + CLIM + REMO, and (c) TOPO + REMO using DAPSM and PSM and (d) SM. Darker colours present more extreme variance ratios: < 0.5 or ≥2 (extreme), ≥0.5 and <0.8 or ≥ 0.8 and <1.2, and >0.8 and ≥1.2 (adequate).

Table 4
Average ASMD across all combinations of number of burned plots (n_t) and unburned plots to burned plots (n_{rc}/n_t) under different covariate sets and matching methods: unmatched, DAPSM, PSM, and SM. Values under different number of burned plots (n_t) and unburned plots to burned plots (n_{rc}/n_t) are available in Fig. S4 in the supplementary materials.

		TOPO + CLIM + REMO + LAND		TOPO + CLIM + REMO		TOPO + REMO		
	Unmatched	DAPSM	PSM	DAPSM	PSM	DAPSM	PSM	SM
ELEV	**0.347**	0.077	0.108	0.075	0.118	0.072	0.084	0.103
SLOPE	**0.292**	0.089	0.094	0.100	0.098	0.087	0.084	**0.226**
ASPECT	0.129	0.091	0.080	0.089	0.083	0.091	0.082	0.130
PHYSCL	**0.552**	0.117	0.072	**0.286**	**0.292**	**0.287**	**0.340**	**0.343**
OWNGC	**0.356**	0.078	0.083	0.178	**0.206**	0.136	0.192	0.160
PROPforest	0.097	0.094	0.090	0.127	0.135	0.125	0.137	0.121
PRECP	**0.349**	0.082	0.091	0.078	0.092	0.091	0.195	0.087
TEMP	0.119	0.057	0.092	0.055	0.104	0.072	**0.205**	0.063
CVPRECP	0.199	0.053	0.081	0.051	0.080	0.064	0.128	0.056
TEMPDF	**0.454**	0.073	0.101	0.071	0.107	0.138	**0.219**	0.126
SVPD	**0.610**	0.079	0.077	0.075	0.076	0.193	**0.436**	0.166
GSMS	**0.317**	0.096	0.079	0.089	0.087	0.133	**0.371**	0.113
NDVI	**0.367**	0.089	0.086	0.086	0.089	0.090	0.079	0.136
TC1	0.103	0.095	0.077	0.092	0.082	0.099	0.084	0.122
TC2	**0.460**	0.079	0.092	0.077	0.093	0.075	0.076	0.164
TC3	**0.232**	0.090	0.081	0.087	0.085	0.093	0.082	0.122

ASMD values over the cut-off of 0.2 were highlighted in bold. ASMD >0.2 and ≤0.4 indicates moderate imbalance, and ASMD >0.4 indicates severe imbalance.

balance in climate variables even though the CLIM covariates were excluded in the propensity score estimation. SM generally achieved balance in most of the covariates except for SLOPE and PHYSCL despite the withdrawal of all covariates in the matching process.

4.2. Performance of wildfire effect estimates under different data availability scenarios

Again, the three data availability scenarios n_t, n_{rc}/n_t, and covariate set affected the bias and variance of wildfire effect estimates collectively. Under DAPSM, the number of burned plots in the simulation subsets (n_t) was related to the bias of the estimates of average wildfire effect, whereas the ratio of unburned to burned plots (n_{rc}/n_t) influenced the variability of the estimates (Fig. 4). Dropping CLIM from the covariate set increased the bias and variability under small values of n_t and n_{rc}/n_t (Fig. 4c). When $n_t = 500$, the simulated mean effect estimate obtained from DAPSM was 30.84 Mg ha^{-1}, similar to the baseline estimate (30.95 Mg ha^{-1}). The estimates of wildfire effect across all n_t and n_{rc}/n_t values under DAPSM deviated on average by 25–30% from the baseline estimate when the covariate set included CLIM variables (Fig. 4a and b), whereas the average deviation was 43% for the covariate sets without CLIM variables (Fig. 4c). n_t less than 100 tended to show large variability in the wildfire effect estimates, resulting in positive effect estimates for a few subsets.

The bias of the effect estimates tended to increase with decreases in n_t and n_{rc}/n_t under DAPSM (Fig. 4a–c), while the bias of PSM-derived estimates was mostly dependent on covariate sets only, not on n_t or n_{rc}/n_t (Fig. 4d–f). The matrix plots (Fig. 5) suggested that the bias was susceptible to small n_{rc}/n_t than small n_t. When CLIM was included in the covariate set, controls from DAPSM with $n_t = 100$ and $n_{rc}/n_t = 20$ were found to show 20% bias compared to the simulation population (Fig. 5a and c). For small numbers of n_t and n_{rc}/n_t (e.g., $n_t<100$ and $n_{rc}/n_t<10$), PSM exhibited smaller bias than DAPSM once CLIM was included in the covariate set (Fig. 5).

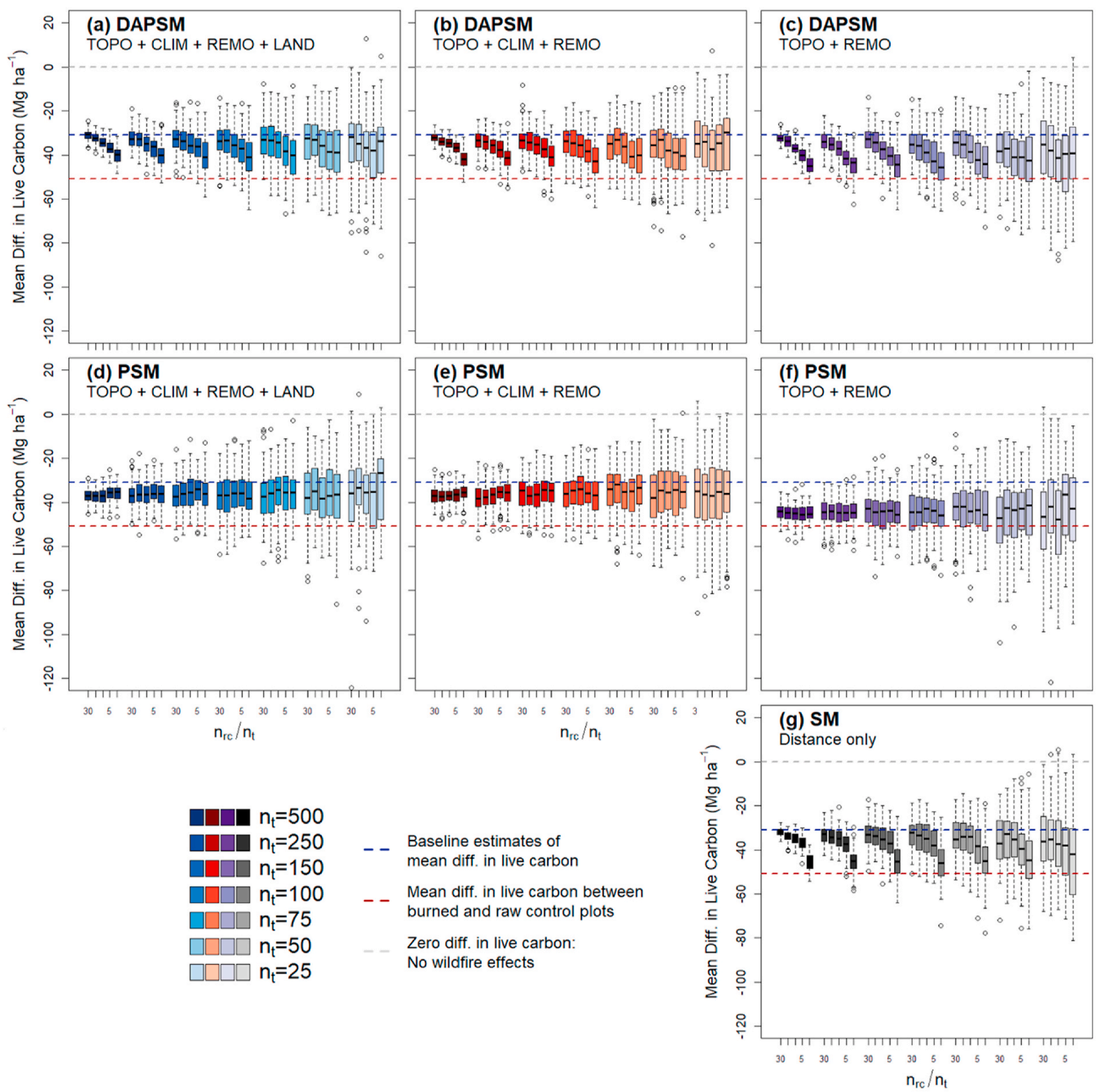

Fig. 4. Distribution of the estimates of wildfire effect on aboveground live woody carbon under different n_t, n_{rc}/n_t, and covariate sets obtained through (a)–(c) DAPSM, (d)–(f) PSM, and (g) SM. For n_{rc}/n_t, the ratio of 30, 20,10, 5, and 2 were plotted for simplicity.

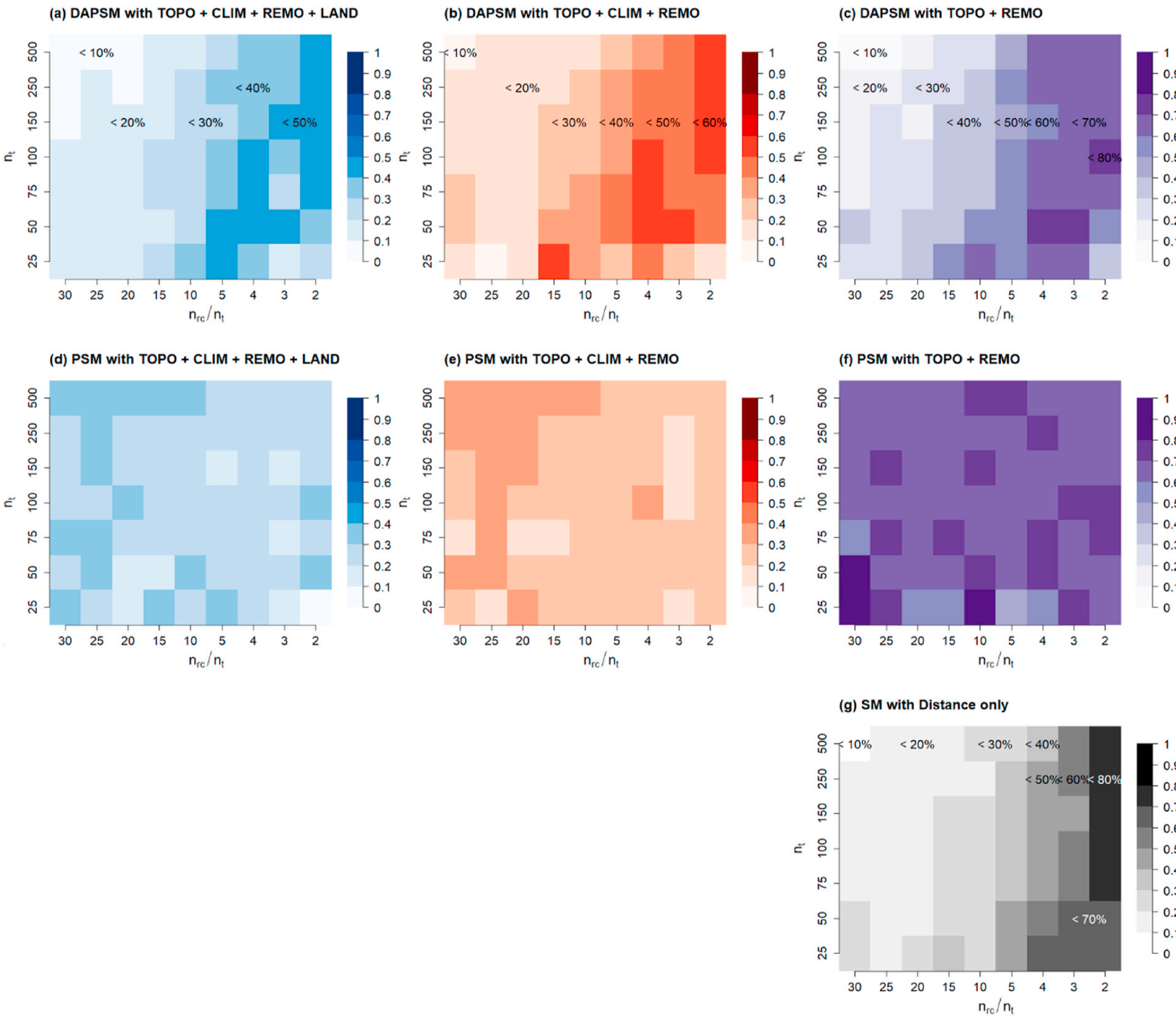

Fig. 5. Matrix plot of percent bias in the wildfire effect estimates under different combinations of n_t, n_{rc}/n_t, and covariate sets. For example, the combination of n_t = 150 and n_{rc}/n_t = 15 exhibited less than 20% of bias compared to the baseline estimate under (a) TOPO + CLIM + REMO + LAND, <30% under (b) TOPO + CLIM + REMO, and <40% under (c) TOPO + REMO when using DAPSM.

4.3. Distance between matched pairs under different matching methods

The mean distance between the burned plots and the matched control plots under PSM ranged between 340 km and 360 km regardless of the number of burned and unburned plots. It did not change substantially, regardless of n_t and n_{rc}/n_t. The mean distance under PSM was similar to the mean distance of all pairwise burned plots and raw control plots in the full data (363.1 km). The maximum pairwise distance between burned and raw control plots in the full data was 945.8 km.

Under DAPSM, the mean distance varied from approximately 7 km–80 km, increasing with decreasing number of observations (Fig. 6). The mean distance changed little among the covariate sets under the same matching method. SM showed the smallest mean distance, ranging from 4.4 to 65 km (Fig. 6).

5. Discussion

5.1. Covariates and distance affect the balance achievement in PS distributions

When matching on the propensity scores alone (PSM), the covariates included in the models were the key driver determining the balance of the propensity score distributions between treated and controls. Neither the sample size nor the ratio of the number of treated plots to controls significantly affected the performance of matching. CLIM variables were the key covariates to estimate the propensity scores for matching, so that covariate sets without CLIM failed to fully reduce the selection bias between the burned and the unburned plots. Our identification of CLIM as the key covariates was expected, as climate is one of the most critical factors determining wildfire occurrence (McKenzie et al., 2004; Weisberg and Swanson 2003) as well as carbon mass in forests (Stegen et al., 2011) at a regional scale. The results imply that to achieve balance in propensity scores, it is more effective to identify and include key covariates than to increase the sample size.

Once distance was included in the matching process (DAPSM), the importance of CLIM variables subsided except when the sample size was very small. Large bias and variance ratio in propensity scores under PSM without CLIM compared to DAPSM suggests that the spatial distance partially incorporated the effect of CLIM variables in estimating the probability of wildfire. The results of average ASMD also specifically confirm that the distance measure contributed to balancing the environmental covariates that were not included in the propensity score estimation model, while PSM obtained balance only on the covariates that were included in the model. Closer plots have higher chance to share similar climate conditions (Di Cecco and Gouhier 2018), thus the spatial proximity may have compensated for the exclusion of CLIM

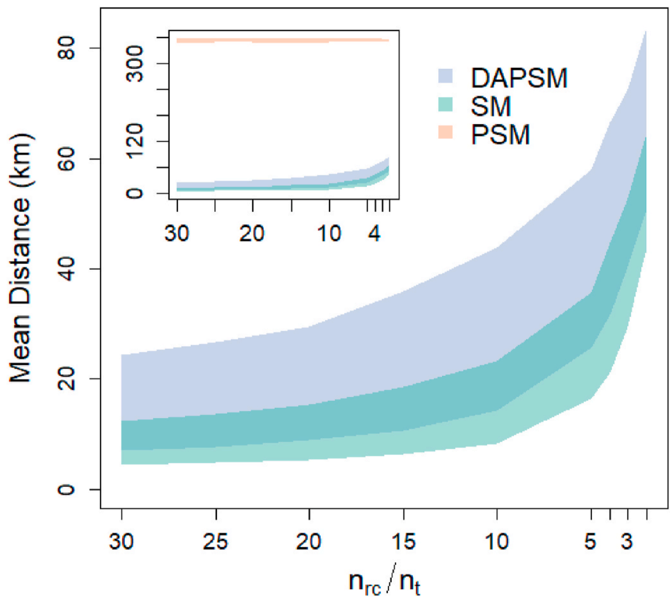

Fig. 6. Mean distance between matched pairs of burned and control plots under DAPSM, SM, and PSM across n_{rc}/n_t.

variables. Even SM, based solely on spatial distance, outperformed PSM without CLIM variables. The results emphasize the importance of incorporating spatial measures to account for both observed and unobserved confounding factors (Woo et al., 2021; Papadogeorgou et al., 2019).

However, incorporating spatial proximity was no longer effective as the observations became sparser over the study area. The bias and variance ratio in propensity scores were remarkably large under sample sizes n_t around 150 or less and $n_{rc}/n_t<10$. This finding indicates that the benefit of spatial distance operates at a particular scale where the geographic variation lines up with the heterogeneity of important covariates. The data of national forest inventory plot measurements utilized in our study were distributed systematically over the states of the Pacific Northwest at an approximate distance of 5 km between plots (Bechtold and Patterson 2005). Hence, the spatial distance among the plots captured relatively large-scale variation of climate rather than local geography such as aspect and slope. In our data with the full sample, the average distance between matched pairs under DAPSM was about 7 km. When the distance between the pairs of treated and control units became extreme due to the small sample size over the region, spatial distance did not account for the variation of climate. When the sample size decreased below $n_t<150$ and $n_{rc}/n_t<10$, the mean distance was approximately 20 km. Thus, the ability of spatial distance to account for the effect of covariates may be linked with the patterns of landscape variables depending on the spatial resolution (Wang et al., 2014; Turner et al., 1989).

The three data availability scenarios – the number of treated, number of raw control relative to treated, and covariate set – interplay with each other to achieve the balance in PS distribution between the treated and the controls. As the performance of the models estimating propensity scores degrades due to the omission of key covariates, spatial information will compensate the lack of data unless the observations are too sparse. Likewise, when less observations are available and the geographic locations do not provide additional information, covariates will become relatively more important. This is especially true when the heterogeneity of the key covariates does not align with the spatial scale of the data.

5.2. Sample size suggestion based on the effect estimates using DAPSM and forest inventory data

The effect of spatial distance in the performance of the matching was also evident from the estimated effect of wildfire, defined as the difference in live carbon between burned and control plots. DAPSM was superior to PSM showing less bias, most likely because the spatial distance accounted for additional information from unobserved covariates. These results agree with a recent clinical study that shows that spatial propensity score matching outperformed non-spatial matching (Davis et al., 2021). However, our study further showed that including the distance measure was less effective or counterproductive when the sample size was very small and the distance between possible matches too large. Then, the spatial distance provided no additional information, resulting in an even larger bias in the estimates. While the critical sample size may be scale dependent, based on the results of simulated estimates, we suggest a sample size of at least $n_t = 100$ and $n_{rc}/n_t = 10$ as the threshold to apply DAPSM over Washington and Oregon. This threshold of number of treated and raw control observations can serve as a practical guideline when implementing DAPSM for further research using forest inventory at regional scales. Further research may not be limited to wildfire effects on carbon mass but on various outcomes of interest that are closely related with climate variables, such as vegetation structure (Duguy et al., 2013), biodiversity (Gill et al., 2013), or land use (Duguy et al., 2012).

The bias in the estimates of wildfire effects observed from our study was generally larger than the results from Pirracchio et al. (2012) who obtained less than 10% of relative bias when decreasing the sample size from 1000 to 40. The relative bias from Pirracchio et al. (2012) was obtained from Monte Carlo simulations using an artificial dataset for which the data generation process was known. It is common to use an artificial dataset for sensitivity analyses of PSM (e.g., Coffman et al., 2020; Rubin and Thomas 1996) based on the true propensity score model or true estimates. Our simulation, on the other hand, was conducted with empirical data of forest inventory measurements that involves additional uncertainty from sampling and measurement errors (Rudolph and Stuart 2018; Bennett et al., 2013). The treatment in our study was a binary factor of either presence or absence of wildfire, because the effect of different severities of wildfires was beyond the research interests of this study. Large variability across fire severities in burned plots may have increased the bias of our estimated wildfire effects. Nevertheless, our study reflects actual relationships among the environmental variables under the presence of wildfires contrary to artificial data (Lechner and Wunsch 2013). Further application of various extended matching methods including propensity score estimation using generalized boosted models (McCaffrey et al., 2013) will help to assess multiple treatment effects from observational data.

5.3. Future work using forest inventory data

DAPSM can be utilized to examine any post-disturbance effects based on the current forest inventory data as long as the number of disturbed plot measurements is sufficient, i.e., ≥100. If the number of treated plots for a given disturbance is less than the desired threshold over a spatial region, one could possibly increase the spatial and temporal intensity of the inventory plot measurements in the database. Such temporal or spatial intensifications are already effective as supplemental studies under the FIA program: the fire effects and recovery study (FERS), used to assess fire effects and post-fire carbon dynamics (e.g., Eskelson et al., 2016), and the sudden oak death syndrome assessment (Jain and Fried 2010). Extension of matching methods such as matching with replacement (e.g., Nolte et al., 2013) can also resolve the small sample size problem (Austin and Cafri 2020), while maintaining high quality of matching (Abadie and Imbens 2006). Taking advantage of the well-defined population and unified protocol of measurements, the national forest inventory dataset combined with a variety of quasi-experimental methods will provide opportunities of continuous causal effect analysis for ecological data.

6. Conclusion

This study presented a sensitivity analysis of distance-adjusted propensity score matching (DAPSM) in relation to data availability. We examined three data availability scenarios in matching: number of available burned plots, number of unburned plots relative to burned plots, and environmental covariate set to estimate the propensity scores for the quantification of wildfire effects using national forest inventory plot measurements in Washington and Oregon. We conducted Monte Carlo simulations of random subsets of burned and unburned plots to quantify the wildfire effects under different covariate sets. For each simulation, performance of DAPSM was evaluated as the bias and variance ratio of propensity scores as well as the bias and variability of the estimates of wildfire effect. We studied the influence of distance measure in matching by comparing the performance of DAPSM to that of matching without distance adjustment.

We found that the three data availability scenarios interacted to affect the performance of matching. Among the three scenarios, whether the covariate set included climate variables or not dominated the degree of balance in propensity score distributions between burned and control plots. Climate variables turned out to be the key covariates to analyze the wildfire effect on forest carbon mass. The sample size, especially the number of unburned plots relative to the burned plots, was closely related with the distance between the matched pairs. Large distance among the plots due to sparseness of observations over the region resulted in severe bias and variability in the effect estimates because the key covariates were not accounted for by the distance measure.

Based on our results, the importance of key covariates and the inclusion of the spatial measure were emphasized for applying propensity score matching methods to account for both observed and unobserved confounding variables. We also suggest a marginal sample size with at least a hundred treated observations and a tenfold larger number of raw controls to implement DAPSM over Washington and Oregon to achieve balance in propensity score distributions and low bias in effect estimates. Future studies implementing matching approaches for effect analyses using forest inventory plot data will find this research useful to determine the sample size and covariate set in relation with spatial distances among the observations. National forest inventory data facilitate monitoring and the assessment of comprehensive forest resources over large spatial and temporal scales. Utilization of DAPSM with forest inventory data enables rigorous research on effect analysis not limited to wildfires but applicable to other disturbances and ecological processes where a randomized experiment is infeasible.

Declaration of competing interest

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

Acknowledgements

This study was supported partly by 4-year 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. 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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2021.105163.

References

Abadie, A., Imbens, G.W., 2006. Large sample properties of matching estimators for average treatment effects. Econometrica 74, 235–267.
Arovaara, H., Hari, P., Kuusela, K., 1984. Possible Effect of Changes in Atmospheric Composition and Acid Rain on Tree Growth. An Analysis Based on the Results of Finnish National Forest Inventories. Finnish Forest Research Institute.
Austin, P.C., 2009. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat. Med. 28, 3083–3107.
Austin, P.C., 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav. Res. 46, 399–424.
Austin, P.C., Cafri, G., 2020. Variance estimation when using propensity-score matching with replacement with survival or time-to-event outcomes. Stat. Med. 39, 1623–1640.
Bechtold, W.A., Patterson, P.L., 2005. The enhanced forest inventory and analysis program-national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC US Dep. Agric. For. Serv. South. Res. Station. 85, 80.
Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., 2013. Characterising performance of environmental models. Environ. Model. Softw. 40, 1–20.
Brookhart, M.A., Schneeweiss, S., Rothman, K.J., Glynn, R.J., Avorn, J., Stürmer, T., 2006. Variable selection for propensity score models. Am. J. Epidemiol. 163, 1149–1156.
Bungartz, H.-J., Kranzlmüller, D., Weinberg, V., Weismüller, J., Wohlgemuth, V., 2018. Advances and New Trends in Environmental Informatics. Springer.
Butsic, V., Lewis, D.J., Radeloff, V.C., Baumann, M., Kuemmerle, T., 2017a. Quasi-experimental methods enable stronger inferences from observational data in ecology. Basic Appl. Ecol. 19, 1–10.
Butsic, V., Munteanu, C., Griffiths, P., Knorn, J., Radeloff, V.C., Lieskovský, J., Mueller, D., Kuemmerle, T., 2017b. The effect of protected areas on forest disturbance in the Carpathian Mountains 1985–2010. Conserv. Biol. 31, 570–580.
Campbell, D.T., Stanley, J.C., 2015. Experimental and Quasi-Experimental Designs for Research. Ravenio Books, Boston, MA.
Carlson, C.H., Dobrowski, S.Z., Safford, H.D., 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 Manag 7, 1–17.
Coffman, D.L., Zhou, J., Cai, X., 2020. Comparison of methods for handling covariate missingness in propensity score estimation with a binary exposure. BMC Med. Res. Methodol. 20, 168. https://doi.org/10.1186/s12874-020-01053-4.
Cohen, J., 2013. Statistical Power Analysis for the Behavioral Sciences. Academic press, New York.
Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., Pasteris, P.P., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol. a J. R. Meteorol. Soc. 28, 2031–2064.
Davis, M.L., Neelon, B., Nietert, P.J., Burgette, L.F., Hunt, K.J., Lawson, A.B., Egede, L.E., 2021. Propensity score matching for multilevel spatial data: accounting for geographic confounding in health disparity studies. Int. J. Health Geogr. 20, 1–12. https://doi.org/10.1186/s12942-021-00265-1.
Di Cecco, G.J., Gouhier, T.C., 2018. Increased spatial and temporal autocorrelation of temperature under climate change. Sci. Rep. 8, 1–9.
Dixon, R.K., Solomon, A.M., Brown, S., Houghton, R.A., Trexier, M.C., Wisniewski, J., 1994. Carbon pools and flux of global forest ecosystems. Science 263, 185–190.
Dray, S., Legendre, P., Peres-Neto, P.R., 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol. Modell. 196, 483–493. https://doi.org/10.1016/j.ecolmodel.2006.02.015.
Duguy, B., Alloza, J.A., Baeza, M.J., De La Riva, J., Echeverría, M., Ibarra, P., Llovet, J., Cabello, F.P., Rovira, P., Vallejo, R.V., 2012. Modelling the ecological vulnerability to forest fires in mediterranean ecosystems using geographic information technologies. Environ. Manage. 50, 1012–1026. https://doi.org/10.1007/s00267-012-9933-3.
Duguy, B., Paula, S., Pausas, J.G., Alloza, J.A., Gimeno, T., Vallejo, R.V., 2013. In: Navarra, A., Tubiana, L. (Eds.), Effects of Climate and Extreme Events on Wildfire Regime and Their Ecological Impacts BT - Regional Assessment of Climate Change in the Mediterranean: Volume 2: Agriculture, Forests and Ecosystem Services and People. Springer Netherlands, Dordrecht, pp. 101–134. https://doi.org/10.1007/978-94-007-5772-1_6.
Eskelson, B.N.I., Monleon, V.J., Fried, J.S., 2016. A 6 year longitudinal study of post-fire woody carbon dynamics in California's forests. Can. J. For. Res. 46, 610–620. https://doi.org/10.1139/cjfr-2015-0375.
Gill, A.M., Stephens, S.L., Cary, G.J., 2013. The worldwide "wildfire" problem. Ecol. Appl. 23, 438–454.
Greenstone, M., Gayer, T., 2009. Quasi-experimental and experimental approaches to environmental economics. J. Environ. Econ. Manage. 57, 21–44.
Heckman, J.J., Ichimura, H., Todd, P.E., 1997. Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. Rev. Econ. Stud. 64, 605–654.

Heinrich, C., Maffioli, A., Vazquez, G., 2010. A Primer for Applying Propensity-Score Matching. Inter-American Dev. Bank.

Huber, M., Lechner, M., Wunsch, C., 2013. The performance of estimators based on the propensity score. J. Econom. 175, 1–21.

IPCC, 2006. IPCC guidelines for national greenhouse gas inventories. In: Agriculture, Forestry and Other Land Use, ume 4. Institute for Global Environmental Strategies (IGES).

Jain, T.B., Fried, J.S., 2010. Field instructions for the annual inventory of California, Oregon, and Washington 2010: supplement for: fire effects and recovery study. US Dep. Agric. For. Serv. Pacific Northwest Res. Station. For. Invent. Anal. Resour. Monit. Assess. Program. 30.

Keith, H., Lindenmayer, D.B., Mackey, B.G., Blair, D., Carter, L., McBurney, L., Okada, S., Konishi-Nagano, T., 2014. Accounting for biomass carbon stock change due to wildfire in temperate forest landscapes in Australia. PloS One 9, e107126. https://doi.org/10.1371/journal.pone.0107126.

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

Larsen, A.E., Meng, K., Kendall, B.E., 2019. Causal analysis in control–impact ecological studies with observational data. Methods Ecol. Evol. 10, 924–934. https://doi.org/10.1111/2041-210X.13190.

Lechner, M., Wunsch, C., 2013. Sensitivity of matching-based program evaluations to the availability of control variables. Labour Econ 21, 111–121.

Liu, S., Bond-Lamberty, B., Hicke, J.A., Vargas, R., Zhao, S., Chen, J., Edburg, S.L., Hu, Y., Liu, J., McGuire, A.D., Xiao, J., Keane, R., Yuan, W., Tang, J., Luo, Y., Potter, C., Oeding, J., 2011. Simulating the impacts of disturbances on forest carbon cycling in North America: processes, data, models, and challenges. J. Geophys. Res. 116, G00K08. https://doi.org/10.1029/2010JG001585.

McCaffrey, D.F., Griffin, B.A., Almirall, D., Slaughter, M.E., Ramchand, R., Burgette, L.F., 2013. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. Stat. Med. 32, 3388–3414.

McKenzie, D., Gedalof, Z., Peterson, D.L., Mote, P., 2004. Climatic change, wildfire, and conservation. Conserv. Biol. 18, 890–902.

Nolte, C., Agrawal, A., Silvius, K.M., Soares-Filho, B.S., 2013. Governance regime and location influence avoided deforestation success of protected areas in the Brazilian Amazon. Proc. Natl. Acad. Sci. Unit. States Am. 110, 4956–4961.

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

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

Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., Wagener, T., 2016. Sensitivity analysis of environmental models: a systematic review with practical workflow. Environ. Model. Softw. 79, 214–232.

Pirracchio, R., Resche-Rigon, M., Chevret, S., 2012. Evaluation of the Propensity score methods for estimating marginal odds ratios in case of small sample size. BMC Med. Res. Methodol. 12, 70. https://doi.org/10.1186/1471-2288-12-70.

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

Rubin, D.B., 2001. Using propensity scores to help design observational studies: application to the tobacco litigation. Health Serv. Outcome Res. Methodol. 2, 169–188.

Rubin, D.B., Thomas, N., 1996. Matching using estimated propensity scores: relating theory to practice. Biometrics 52, 249–264. https://doi.org/10.2307/2533160.

Rudolph, K.E., Stuart, E.A., 2018. Using sensitivity analyses for unobserved confounding to address covariate measurement error in propensity score methods. Am. J. Epidemiol. 187, 604–613. https://doi.org/10.1093/aje/kwx248.

Smeeth, L., Douglas, I., Hall, A.J., Hubbard, R., Evans, S., 2009. Effect of statins on a wide range of health outcomes: a cohort study validated by comparison with randomized trials. Br. J. Clin. Pharmacol. 67, 99–109.

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

Stegen, J.C., Swenson, N.G., Enquist, B.J., White, E.P., Phillips, O.L., Jørgensen, P.M., Weiser, M.D., Monteagudo Mendoza, A., Núñez Vargas, P., 2011. Variation in above-ground forest biomass across broad climatic gradients. Global Ecol. Biogeogr. 20, 744–754.

Stuart, E.A., 2010. Matching methods for causal inference: a review and a look forward. Stat. Sci. 25, 1–21. https://doi.org/10.1214/09-STS313.

Stuart, E.A., Rubin, D.B., 2004. Matching Methods for Causal Inference: Designing Observational Studies. Harvard Univ. Dep. Stat. mimeo.

Turner, M.G., O'Neill, R.V., Gardner, R.H., Milne, B.T., 1989. Effects of changing spatial scale on the analysis of landscape pattern. Landsc. Ecol. 3, 153–162.

Wang, W.J., He, H.S., Spetich, M.A., Shifley, S.R., Thompson, F.R., Dijak, W.D., Wang, Q., 2014. A framework for evaluating forest landscape model predictions using empirical data and knowledge. Environ. Model. Softw. 62, 230–239.

Weisberg, P.J., Swanson, F.J., 2003. Regional synchroneity in fire regimes of western Oregon and Washington, USA. For. Ecol. Manage. 172, 17–28.

Woo, H., Eskelson, B.N.I., Monleon, V.J., 2021. Matching methods to quantify wildfire effects on forest carbon mass in the U.S. Pacific Northwest. Ecol. Appl. 31, e02283.

Woodall, C.W., Perry, C.H., Westfall, J.A., 2012. An empirical assessment of forest floor carbon stock components across the United States. For. Ecol. Manage. 269, 1–9.