Quasi-experimental methods for wildfire impact quantification: applications of distance-adjusted propensity score matching to forest inventory data

by

Hyeyoung Woo

B.Sc., Seoul National University, 2007

M.S., Forestry, University of Montana, 2015

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(FORESTRY)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

December 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Quasi-experimental methods for wildfire impact quantification: applications of distance-adjusted propensity score matching to forest inventory data

---

submitted by Hyeyoung Woo in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Forestry

Examining Committee:

Dr. Bianca N.I. Eskelson, Forest Resources Management, UBC

Supervisor

Dr. Lori D. Daniels, Forest and Conservation Sciences, UBC

Supervisory Committee Member

Dr. Vicente J. Monleon, U.S. Forest Service

Supervisory Committee Member

Dr. Peter L. Marshall, Forest Resources Management, UBC

University Examiner

Dr. R. Daniel Moore, Geography, UBC

University Examiner

Additional Supervisory Committee Members:

Dr. Céline Boisvenue, Canadian Forest Service

Supervisory Committee Member

Abstract

Quantifying wildfire impacts on forest ecosystems is challenging due to the lack of pre-fire data or controls from experiments over a large landscape. Quasi-experimental methods have been popular in various fields of science where experiments are difficult to implement. However, the application of quasi-experimental methods to ecological data have not yet been fully explored. In this dissertation, I applied quasi-experimental methods to quantify wildfire impacts on aboveground forest woody carbon mass using national forest inventory data from the United States of America (USA) and British Columbia (BC), Canada.

First, I compared distance-adjusted propensity score matching (DAPSM) with propensity score matching (PSM) and spatial matching (SM) to quantify the changes in forest woody carbon mass due to wildfires in Washington and Oregon, USA. Incorporating spatial information in addition to environmental covariates was essential to account for both observed and unobserved environmental covariates in matching. Thus, DAPSM was favored over PSM and SM.

Second, I conducted a sensitivity analysis on the performance of DAPSM with different data availability to provide a practical guide of sample size and environmental covariates required to quantify wildfire impacts. I found that the inclusion of the spatial distance compensated for the omission of key covariates, but this compensation was not effective for small sample sizes.

Third, I applied DAPSM with and without replacement to three datasets with small sample sizes collected for case-studies of wildfire impacts in south-central BC. DAPSM with replacement using BC forest inventory plot data enabled balancing the environmental covariates between burned and control plots under certain circumstances. The controls produced by DAPSM captured the trends in the amount of woody carbon masses under different fire severities,

implying that they may replace the pre-burn data once the propensity scores are adequately addressed.

Overall, the implementation of DAPSM allowed the assessment of wildfire impacts on forest carbon by building a causal relationship from observational forest inventory data. Based on applied examples, this dissertation provides guidelines for employing propensity score matching to quantify the impacts of natural disturbances. This research contributes to future studies considering quasi-experimental approaches for analyzing ecological data where controlled experiments are impossible.

Lay Summary

The impact of wildfires on forest carbon is difficult to quantify because pre-fire data are mostly lacking and randomized experiments are impossible. Quasi-experimental methods are widely used in health, econometrics, and social sciences to assess causal impacts from non-experimental research. Yet, these methods have been rarely used for impact analyses of natural disturbances. In this dissertation, I established guidelines for implementing quasi-experimental methods for ecological data, especially for spatially located data. I examined distance-adjusted propensity score matching (DAPSM) as one of the quasi-experimental methods to quantify wildfire impacts on aboveground forest woody carbon mass, based on national forest inventory data from the Pacific Northwest of the United States of America and south-central British Columbia, Canada. This dissertation provides innovative approaches using quasi-experimental methods with considerations of spatial components in ecological data with examples and suggestions for future research on impact analyses research for which controlled experiments are impossible.

Preface

The contributions of committee members to each research chapter are summarized in the tables below.

Chapter 2

Research Contribution	Hyeyoung Woo	Bianca Eskelson	Vicente Monleon	Lori Daniels	Celine Boisvenue	Total
Identify research problem	80	8	8	2	2	100
Designing research	80	9	9	1	1	100
Analyzing data	85	8	7	0	0	100
Manuscript writing	90	5	5	0	0	100

Chapter 3

Research Contribution	Hyeyoung Woo	Bianca Eskelson	Vicente Monleon	Lori Daniels	Celine Boisvenue	Total
Identify research problem	80	9	9	1	1	100
Designing research	90	5	5	0	0	100
Analyzing data	95	2	3	0	0	100
Manuscript writing	94	3	3	0	0	100

Chapter 4

Research Contribution	Hyeyoung Woo	Bianca Eskelson	Vicente Monleon	Lori Daniels	Celine Boisvenue	Total
Identify research problem	85	5	0	5	5	100
Designing research	85	5	0	5	5	100
Analyzing data	100	0	0	0	0	100
Manuscript writing	88	3	3	3	3	100

A version of Chapter 2 was published as:

Woo, Hyeyoung, et al. “Matching Methods to Quantify Wildfire Effects on Forest Carbon Mass in the U.S. Pacific Northwest.” Ecological Applications, vol. 31, no. 3, John Wiley & Sons, Ltd, Apr. 2021, p. e02283, doi:https://doi.org/10.1002/eap.2283. Journal Impact Factor in 2020: 4.657 (Clarivate Analytics).

A version of Chapter 3 was published as:

Woo, Hyeyoung, et al. “Sensitivity Analysis on Distance-Adjusted Propensity Score Matching for Wildfire Effect Quantification Using National Forest Inventory Data.” Environmental Modelling & Software, vol. 144, Elsevier, 2021, p. 105163, doi: https://doi.org/10.1016/j.envsoft.2021.105163. Journal Impact Factor in 2020: 5.288 (Clarivate Analytics).

A version of Chapter 4 is being prepared for peer-reviewed publication.

Table of Contents

Abstract iii

Lay Summary v

Preface vi

Table of Contents viii

List of Tables xiv

List of Figures xv

List of Abbreviations xvii

Acknowledgements xx

Dedication xxi

Chapter 1: Introduction 1

1.1 Background 1

1.2 Research goal and dissertation overview 4

Chapter 2: Matching methods to quantify wildfire effects on forest carbon mass in the United States Pacific Northwest 9

2.1 Introduction 9

2.2 Study area and data 13

2.2.1 Study area 13

2.2.2 Forest Inventory and Analysis (FIA) data 15

2.2.3 Fire data 17
2.2.4 Climate and remotely sensed data 18
2.3 Methods 20
2.3.1 Matching burned and unburned plots 20
2.3.1.1 Propensity score matching (PSM) 23
2.3.1.2 Spatial matching (SM) 23
2.3.1.3 Distance-adjusted propensity score matching (DAPSM) 24
2.3.2 Diagnosis of balance between burned and control plots 25
2.3.3 Calculations of aboveground woody carbon mass 26
2.3.4 Average wildfire effects on aboveground woody carbon masses 27
2.4 Results 28
2.4.1 Balance in propensity scores and in covariate distributions achieved under both PSM and DAPSM 28
2.4.2 Discrepancy in carbon mass of standing and downed wood between unburned and control plots 34
2.4.3 Average wildfire effect on aboveground woody carbon mass under three matching methods 35
2.5 Discussion 38
2.5.1 Balance in observed covariates may not guarantee unbiased estimation of average wildfire effects 38
2.5.2 DAPSM is preferable for wildfire effects estimation 40

2.5.3 Estimated average wildfire effects on carbon mass by severity from DAPSM .. 41

Chapter 3: Sensitivity analysis on distance-adjusted propensity score matching for wildfire effect quantification using national forest inventory data 44

3.1 Introduction 44

3.2 Data 47

3.2.1 Forest inventory data 47

3.2.2 Environmental covariates 49

3.2.3 Aboveground live woody carbon 52

3.3 Matching methods 52

3.3.1 PSM and DAPSM for assessing wildfire effects on aboveground live woody carbon 52

3.3.2 Sensitivity analysis of DAPSM under different data availability scenarios 53

3.3.3 Evaluation of results 55

3.3.3.1 Descriptive statistics for evaluation of balance in propensity scores 55

3.3.3.2 Performance of matching estimates 57

3.4 Results 58

3.4.1 Balance in propensity scores under different data availability scenarios 58

3.4.1.1 Smaller mean bias in propensity scores under larger $\boldsymbol{nt}$ and covariate set with CLIM 58

3.4.1.2 Smaller variance ratio in propensity scores under large $\boldsymbol{nt}$ and covariate set with CLIM 60

3.4.1.3 Smaller variance ratio of the residuals of the covariates under DAPSM 63

3.4.1.4 DAPSM balances some covariates that were not included in propensity score model 64

3.4.2 Performance of wildfire effect estimates under different data availability scenarios 67

3.4.3 Distance between matched pairs under different matching methods 70

3.5 Discussion 71

3.5.1 Covariates and distance affect the balance achievement in PS distributions 71

3.5.2 Sample size suggestion based on the effect estimates using DAPSM and forest inventory data 73

3.5.3 Future work using forest inventory data 74

Chapter 4: Quasi-experimental methods to support case studies of wildfire impacts 76

4.1 Introduction 76

4.2 Study area 80

4.3 Data 82

4.3.1 Fire data 82

4.3.1.1 Elephant Hill Fire data 82

4.3.1.2 Prouton Lakes Fire data 83

4.3.1.3 South-central BC fires data 83

4.3.2 Forest Analysis and Inventory Branch (FAIB) data 85

4.3.3 Environmental covariates 87
4.3.3.1 Topography 87
4.3.3.2 Climate 88
4.3.3.3 Remote-sensing biomass data 88
4.4 Methods 90
4.4.1 Evaluation of similarity in environmental covariates between burned and control plots 90
4.4.2 Distance-adjusted propensity score matching (DAPSM) with and without replacement 91
4.4.3 Computation of standing tree carbon 92
4.4.4 Process of analysis 92
4.5 Results 94
4.5.1 Similarity between burned and sampled unburned plots in the three datasets 97
4.5.2 Performance of DAPSM with replacement compared to DAPSM without replacement 99
4.5.3 Influence of spatial intensity of burned and unburned plots on DAPSM with and without replacement 102
4.5.4 Wildfire impact on woody carbon mass from standing trees estimated for the south-central BC fires 103
4.6 Discussion 105

4.6.1 Imbalance found in environmental covariates between sampled unburned plots and burned plots 105

4.6.2 Relative performance of DAPSM with replacement to without replacement was dependent on the spatial distribution of data 107

4.6.3 Impact of south-central BC fires on standing tree woody carbon mass: can DAPSM-selected controls replace pre-burn data? 109

Chapter 5: Conclusion 111

5.1 Overall importance and contributions 111

5.2 Cautions about implementing propensity score matching 111

5.3 Future research directions 116

5.3.1 Methodology may be applicable to quantifying effects of other disturbances . 117

5.3.2 Exploration of extensions of propensity score adjustment methods 117

5.3.3 Improvement of propensity score estimation 118

5.3.4 Downscaling of the temporal resolution of climate metrics 119

5.3.5 Implement modifications to existing inventory design 119

Bibliography 121

Appendices 136

List of Tables

Table 2.1 Summary statistics of topographic and land cover variables obtained from FIA plot data. 16

Table 2.2 Number of FIA plots in each burn severity class. 18

Table 2.3 Mean of PRISM and Landsat data derived covariates, with the standard deviation in parentheses. 20

Table 3.1 Descriptions and sources of sixteen environmental variables used for matching burned and unburned forest inventory plots. 51

Table 3.2 Simulated data availability scenarios and their values and conditions for DAPSM and PSM. SM simulated $\boldsymbol{nt}$ and $\boldsymbol{nrc/nt}$ only. 55

Table 3.3 Baseline statistics of bias and variance ratio in propensity scores (PS) computed from the simulation population ($\boldsymbol{Nt}$=611, $\boldsymbol{Nrc/Nt}$=36.7) and controls obtained from DAPSM with full covariate set including TOPO, LAND, REMO, and CLIM. 57

Table 3.4 Average ASMD across all combinations of number of burned plots ($\boldsymbol{nt}$) and unburned plots to burned plots ($\boldsymbol{nrc/nt}$) under different covariate sets and matching methods. ... 66

Table 4.1 Descriptions of the three datasets used for the analysis. 84

Table 4.2 Descriptions of the environmental covariates. 89

Table 4.3 Results of PCA for eight climate variables. 90

Table 4.4 Numbers of sampled unburned plots, burned plots, and raw control plots used in the analysis. 94

Table 4.5 The mean and standard deviation (s.d.) of propensity scores (PS) of burned plots, sampled unburned plots, raw controls, and controls matched with and without replacement for the three datasets. 101

Table 4.6 Mean distances (in km) among all burned plots, among all sampled unburned plots, and between matched pairs of burned and control plots selected through DAPSM with and without replacement for the three fire datasets. 103

List of Figures

Figure 2.1. Study area (Washington and Oregon, USA) and spatial distribution of large wildfires (e.g., >40 ha) from 1984-2016. 14

Figure 2.2 Flow of the analysis in this chapter. 22

Figure 2.3 Spatial distributions of burned plots (grey 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). 31

Figure 2.4 Histograms and 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. 32

Figure 2.5 Absolute standardized mean difference (ASMD) of 16 environmental covariates before and after matching under three strategies. 33

Figure 2.6 Mean total carbon masses (Mg ha^{-1}) by severity classes with their standard errors (SE). 36

Figure 2.7 Proportion of carbon mass relative to control plots (natural log) estimated from the Poisson pseudo-maximum likelihood model were presented with 95% confidence intervals with Tukey's adjustment. 37

Figure 3.1 Mean bias in propensity scores (PS) under different number of burned plots ($\boldsymbol{nt}$), unburned plots to burned plots ($\boldsymbol{nrc/nt}$), 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. 60

Figure 3.2 Mean variance ratio in propensity scores (PS) between burned and control plots under different combinations of number of burned plots ($\boldsymbol{nt}$), unburned plots to burned plots ($\boldsymbol{nrc/nt}$), and covariate set, matched based on (a) – (c) DAPSM, (d) – (f) PSM, or (g) SM. 62

Figure 3.3 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. 64

Figure 3.4 Distribution of the estimates of wildfire effect on aboveground live woody carbon under different $\boldsymbol{nt}$, $\boldsymbol{nrc/nt}$, and covariate sets obtained through (a)-(c) DAPSM, (d)-(f) PSM, and (g) SM. 68

Figure 3.5 Matrix plot of percent bias in the wildfire effect estimates under different combinations of $\boldsymbol{nt}$, $\boldsymbol{nrc/nt}$, and covariate sets. 69

Figure 3.6 Mean distance between matched pairs of burned and control plots under DAPSM, SM, and PSM across $\boldsymbol{nrc/nt}$. 70

Figure 4.1 Fire perimeters and sampling locations of the three fire datasets: South-central BC fires, Elephant Hill Fire, and Prouton Lakes Fire. 85

Figure 4.2 Locations of 632 Forest Analysis and Inventory Branch (FAIB) plots in the study area. 87

Figure 4.3 Spatial distributions of burned plots and control plots obtained from DAPSM with and without replacement for the three datasets: a) and b), Elephant Hill Fire data, c) and d), Prouton Lakes Fire data, and e) and f), south-central BC fires data. 96

Figure 4.4 Absolute standardized mean difference (ASMD) of the fifteen environmental covariates for a) Elephant Hill Fire data, b) Prouton Lakes Fire data, and c) south-central BC fires data. 98

Figure 4.5 Estimated woody carbon mass (Mg ha^{-1}) from standing trees (live and dead) in plots measured post-burn (Post), in plots measured pre-burn (Pre), and controls selected by DAPSM with replacement (MWR) and DAPSM without replacement (MWOR) by different fire severity (low, moderate, and high) in the south-central BC fires data. 105

List of Abbreviations

ρ – Pearson's correlation coefficients

annpre – annual precipitation

anntmp – mean annual temperature

ASMD – absolute standardized mean difference

asp – aspect

BC – British Columbia

CI – confidence interval

CLIM – climate

CMI – Change Monitoring Inventory

cosASP – cosine transformation of aspect in azimuth

cvpre – coefficient of variation in difference in precipitation between August and December

CWM – coarse woody material

DAPS – distance-adjusted propensity score

DAPSM – distance-adjusted propensity score matching

DBH – diameter at breast height

DEM – digital elevation models (DEM)

diftmp – difference in mean temperature between August and December

Dist – relative Euclidean distance

dNBR – difference Normalized Burn Ratio

E – expectations

ELEV – elevation in metres

FAIB – Forest Analysis and Inventory Branch

FIA – Forest Inventory and Analysis

GIS – geographical information systems (GIS)

LAND – land cover

Mg ha^{-1} – mega grams per hectare

MTBS – Monitoring Trends in Burn Severity

MWOR – matching without replacement

MWR – matching with replacement

N – population size

n – sample size

NDVI – Normalized Difference Vegetation Index

NFI – National Forest Inventory

OWNS – land ownership

p – p-value

PCA – principal component analysis

PHYSCL – physiographic class

PRISM – Parameter-elevation Regressions on Independent Slopes Model

PROPforest – proportion of forested area

PS – propensity score

PSM – propensity score matching

rc – raw control

REMO – remotely sensed variables

s.d. – standard deviation

SE – standard errors

SLOPE – slope in degrees

SM – spatial matching

smrnmvpd – mean summer vapor pressure deficit

smrtp – growing season moisture stress

SUP – Supplementary sample plots

t – treated

TC – tasseled cap metrics

TC1 – TC brightness index

TC2 – TC greenness index

TC3 – TC wetness index

TIFF – tagged image file format

TOPO – topography

US – United States

USA – United States of America

Var – variance

w – weight

YSM – Young Stand Monitoring

β – parameters

T – binary treatment

$Y(T)$ – outcome in treated unit

$p(X)$ – propensity score defined as a function of covariates X

$\boldsymbol{X}$ – vector of covariates

Acknowledgements

It was great luck for me to learn and live in the beautiful British Columbia. Without the contributions of lots of people, I doubt that my graduate program would have been a successful and rewarding experience.

First, I would like to express my sincere gratitude to my research supervisor, Prof. Bianca Eskelson, for her guidance, commitment, encouragement, and patience that made this dissertation possible. I also thank Dr. Vicente Monleon, Prof. Lori Daniels, and Dr. Céline Boisvenue for serving on my committee. I have learned much from their help and insights. I thank Prof. John Kershaw for acting as the external examiner for this dissertation.

There are tons of people who helped to finish my graduate course and dissertation project. I would like to thank Dr. Matthew Gregory from LEMMA, Dr. Rene De Jong from BC Ministry of Forests, and Haitao Li from UBC Geospatial team for providing the data needed for the research project. Various scholarships and fellowship from the University of British Columbia and the Faculty of Forestry were essential to support the project.

Lastly, I thank my wonderful family in Canada and Korea for their endless love and support.

Dedication

This dissertation is dedicated to the people who have supported me throughout my journey.

Thanks for their endless support and encouragement.

Chapter 1: Introduction

1.1 Background

Forests hold up to 90% of the aboveground terrestrial carbon stock (Houghton et al. 2009). Carbon sequestration by forests through photosynthesis is vital to offset the carbon emitted by fossil fuel emissions, and further to moderate the effect of climate change (Bonan 2008). The contribution of temperate forests in the United States (US) and Canada to the global forest carbon sink has been highly variable due to natural and anthropogenic disturbances such as fire, insects, and harvest (Kurz et al. 2008; Stinson et al. 2011; Van Mantgem et al. 2009). Therefore, quantifying the impact of such disturbances on forest carbon dynamics is important to strategically react to changes in climate.

Wildfire is one of the primary disturbance agents altering the carbon cycles in forests (Goetz et al. 2012; Goulden et al. 2011). Not only does the consumption of live and dead fuels cause instant and huge carbon emission from forests, but the post-fire tree mortality and changes in forest stand structure due to stand-replacing wildfires result in fluctuations in the carbon storage in the long term (Kashian et al. 2006, Regelbrugge and Conard 1993). Moreover, the atmospheric carbon released from forest fires accelerates the potential threat of more frequent and severe wildfire occurrences through positive feedbacks in the carbon cycle (Spracklen et al. 2009). The close linkage between wildfires and the changing role of forests as carbon sink-to-source calls for accurate measures of wildfire impact on the forest carbon budget at a regional scale (Brown 2002). Carbon consumption by wildfires, unlike the carbon removal intended by forest management, cannot be planned or adjusted, thus a more precise quantification process for this subject is required (French et al. 2002; Kasischke and Bruhwiler 2002).

The fact that wildfires cannot be fully controlled in experimental studies poses methodological challenges for deriving statistical inferences to quantify wildfire impacts. For example, wildfire impacts on forest carbon can be quantified as the amount of carbon mass altered when fires occur. In this context, the occurrence of wildfire can be considered a natural treatment imposed to forests (Butsic et al. 2017). Yet, the treatment is not randomly assigned to forest sites or plots as in experimental studies, because wildfire is a stochastic and unpredictable event. Therefore, randomized experimental approaches that examine the causal effect of wildfires from a design-based perspective are impossible (Taylor et al. 2013; Barrett et al. 2010). Instead, the connection between wildfires and forest carbon mass is often accounted for through non-experimental research based on survey data using post-fire measurements and/or satellite images (e.g., Stinson et al. 2011, Isaev et al. 2001). Statistical inferences in non-experimental research, however, are inherently confined by systematic differences between burned and unburned units (Johnson et al. 2007). The systematic difference between treated (i.e., burned plots) and untreated units (i.e., unburned plots) is referred to as selection bias (Caliendo and Kopeinig 2008), which leads to biased inferences of the quantified impact (Winship and Mare 1992).

The constraint of making causal inference from observational studies is not limited to forestry data. This topic has been thoroughly discussed in many fields of science where randomized experiments are not feasible. Research in the social sciences (Berk 1983), econometrics (Heckman 1977), and medicine (Hernán et al. 2004) includes development of quasi-experimental methods that approximate randomized experiments and provide viable alternatives to the random assignment of treatments (Cook and Campbell 1986). Quasi-experimental methods enable stronger, causal inferences from observational studies (Butsic et al. 2017) and have become widely utilized in disciplines that rely on non-experimental approaches.

Propensity score matching, developed by Rosenbaum and Rubin (1983), is one of the quasi-experimental methods that reduce selection bias by regulating the covariate distributions between treated and untreated groups (Stuart 2010). Propensity score matching has been considered as standard practice for inferring causality from observational data for three decades (Ali et al. 2019, Shipman et al. 2017, Lee and Little 2017), because the method allows direct comparison between treated and untreated groups (Morgan and Winship 2015). For wildfire or other natural disturbance analyses, however, few studies have applied this method, even though the data structure and questions are often similar to those of other disciplines that implement propensity score matching methods (Butsic et al. 2017, Larsen et al. 2019).

Ecological processes, including natural disturbances such as wildfires, are affected by spatio-temporal heterogeneity of the environment (Legendre and Fortin 1989). Because the environmental variables display spatial patterns, it is essential to consider the spatial structure of data for quantifying the impacts of the ecological processes (Legendre et al. 2002). The incorporation of a spatial component to propensity score matching methods has been shown to reduce potential bias due to spatial dependence of observations in political science (e.g., Keele et al. 2015), social studies (e.g., Chagas et al. 2012), economics (e.g., Gonzales et al. 2017), and health studies (e.g., Davis et al. 2021). Recently, distance-adjusted propensity score matching (DAPSM; Papadogeorgou et al. 2019) has been proposed to take into account spatial proximity between treated and untreated units for matching. Given the importance of spatial attributes in ecological data, the applicability of quasi-experimental methods on analyses of wildfire impacts needs to be evaluated. One opportunity is to assess the performance of distance-adjusted propensity score matching in comparison with that of conventional propensity score matching using empirical forest inventory data.

Forest inventories provide empirical data and monitoring information derived from field surveys with repeated measurements. Forest inventory data are managed at national or regional levels in numerous countries and are based on sample plots from a pre-defined spatial sampling protocol over a larger region (Tomppo et al. 2008). National forest inventory data are utilized for estimating forest-related variables such as forest biomass (Blackard et al. 2008) or for developing growth models (Soares et al. 1995) or models to assess wind and snow damage (Jalkanen and Mattila 2000). In the United States of America (USA) and Canada, countries where the forest industry is a large part of the economy (Gronowska et al. 2009), observational data in forest inventories have been collected using a systematic spatio-temporal sampling scheme. Combined with remotely sensed wildfire occurrence and burn severity data, the national forest inventory data enable the implementation of quasi-experimental methods for making causal inference of wildfire impacts on forest attributes such as aboveground woody carbon mass.

1.2 Research goal and dissertation overview

The main research goal of this dissertation is to build a framework implementing quasi-experimental methods to obtain causal inferences of wildfire impacts on forest carbon mass. Specifically, I have made innovative contributions to distance-adjusted propensity score matching methods using the national forest inventory data from the Pacific Northwest of the USA and south-central British Columbia (BC), Canada. To achieve my main research goal, specific research questions and objectives were posed in each chapter as described in the following overview and flow chart (Figure 1.1).

In Chapter 2, I present novel methods to quantify wildfire effects on aboveground forest woody carbon mass in Washington and Oregon, United States, using the Forest Inventory and Analysis (FIA) data. I compared three different matching methods for propensity score estimation, each

uniquely considering spatial measures: 1) propensity score matching (PSM), 2) spatial matching (SM), and 3) distance-adjusted propensity score matching (DAPSM). Remotely sensed data including the Normalized Difference Vegetation Index (NDVI) and burn severity classification, and PRISM climate variables (Daly et al. 2008) were also used to account for linkages between the presence of wildfire and environmental characteristics of the forest inventory plots when estimating the propensity scores. The specific research objectives addressed in this chapter are:

1. To diagnose the balance achieved between burned and control plots under the three matching methods (DAPSM, PSM, and SM) and to address how environmental covariates and spatial distance can contribute to the removal of the selection bias between burned and unburned plots.

2. 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 (DAPSM, PSM, and SM) may provide different outcomes when answering an ecological question.

3. To estimate the difference in carbon mass by burn severity obtained through the three matching methods (DAPSM, PSM, and SM).

In Chapter 3, I performed sensitivity analyses on the performance of distance-adjusted propensity score matching (DAPSM) under different data availability. Relying on the large number of burned and unburned plot measurements from the FIA data utilized in Chapter 2, I conducted Monte Carlo simulations of DAPSM with random subsets of the empirical data. I assessed the influence of sample size and covariate information on the balance of the propensity score distribution and the efficiency of wildfire effect estimates. The subsets varied according to

the number of burned plots, ratio of unburned plots to burned plots, and availability of environmental covariates. The specific research questions addressed in this chapter are:

1. How do the balance in the propensity score distribution and the estimated wildfire effects change under decreasing sample size?

2. How does the performance of DAPSM change under the limited availability of covariate information?

3. What is the impact of spatial distance in the performance of matching in the context of reduced sample size?

4. What are the marginal sample size and the key covariates to implement DAPSM for wildfire effect analyses using the forest inventory data?

In Chapter 4, I investigated matching approaches with and without -replacement for three case studies of wildfire impacts in south-central BC, Canada. Based on the marginal sample size and the key covariates found from Chapter 3, I examined the performance of DAPSM with and without replacement using three fire datasets collected following the extreme fire year of 2017. For each matching approach, the balance in the propensity score and the distribution of the environmental covariates was evaluated for the three datasets to interrogate the utility of DAPSM with replacement in relation to the spatial distributions of data from case studies with small sample size. The specific research questions addressed in this chapter are:

1. How much similarity with regards to environmental covariates is there between burned plots and unburned plots that were sampled after the wildfire? How similar are the

control plots selected by DAPSM to the burned plots in terms of their environmental covariates?

2. Does DAPSM with replacement improve the balance in environmental covariates for fire datasets with small sample sizes in comparison with DAPSM without replacement? Do the relative spatial locations of burned plots against unburned controls affect the performance of DAPSM in achieving balance in environmental covariates?

3. Do the controls selected by DAPSM with and without replacement produce similar estimation of wildfire impacts on forest woody carbon as the pre-burn data?

Lastly, Chapter 5 provides a summary of the results with an overall conclusion. I highlight the contributions of this dissertation to forest research on disturbance effects, and discuss limitations and future research directions.

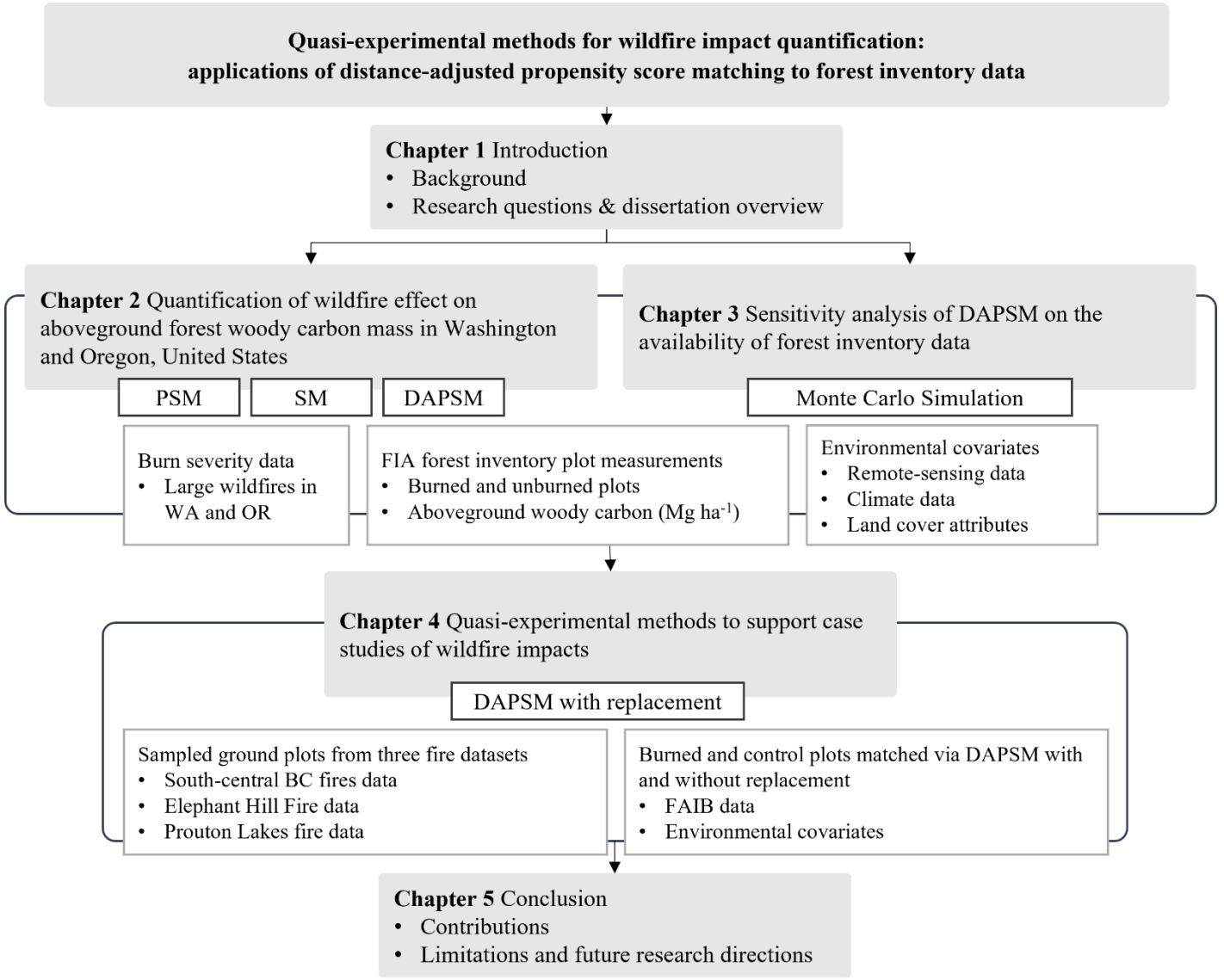


Figure 1.1. A flow chart of the dissertation structure with methods and data.

Chapter 2: Matching methods to quantify wildfire effects on forest carbon mass in the United States Pacific Northwest

2.1 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 long-term 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 United States (US) 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 (Hudiburg et al. 2011, Mitchell et al. 2009). 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 (Wimberly and Liu 2014, Mitchell et al. 2009).

Estimating wildfire effects on forest carbon, however, is challenging because areas that burn typically lack comparable controls or pre-fire 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 (Parisien et al. 2012, Ager et al. 2007, Pew and Larsen 2001) 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 pre-fire 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 capture all the variability need for landscape-level estimations (Zheng et al., 2004), nor the majority of wildfires with low to moderate severities (Eskelson and Monleon 2018).

The pre-existing 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 spatio-temporal 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., Hernán et al. 2004, D'Agostino 1998), social policy program evaluation (e.g., Keele et al. 2015, Caliendo and

Kopeinig 2008), 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 managemental 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 (Butsic et al. 2017, Ager et al. 2010). 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, I 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. I 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 US Pacific Northwest.

Given the remote-sensing data and the extensive national forest inventory data with detailed environmental conditions and spatial locations, I 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 1) how the theoretical framework

of matching methods can be applied to real-world data sets to solve ecological problems, and 2) how users can assess the performance of different matching approaches to choose the best method providing the most reliable inferences with regards to the ecological question asked. Because matching methods have rarely been used for ecological applications, the comprehensive discussion of several approaches in a real-world application provides an important contribution to users of matching methods.

2.2 Study area and data

2.2.1 Study area

The study area encompasses the states of Washington and Oregon, USA (Figure 2.1). 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 percent of Washington and 60 percent of Oregon (Campbell et al. 2010, Donnegan et al. 2008). Fire regimes in the study area vary substantially, with fire return intervals ranging from 15 to 937 years 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).

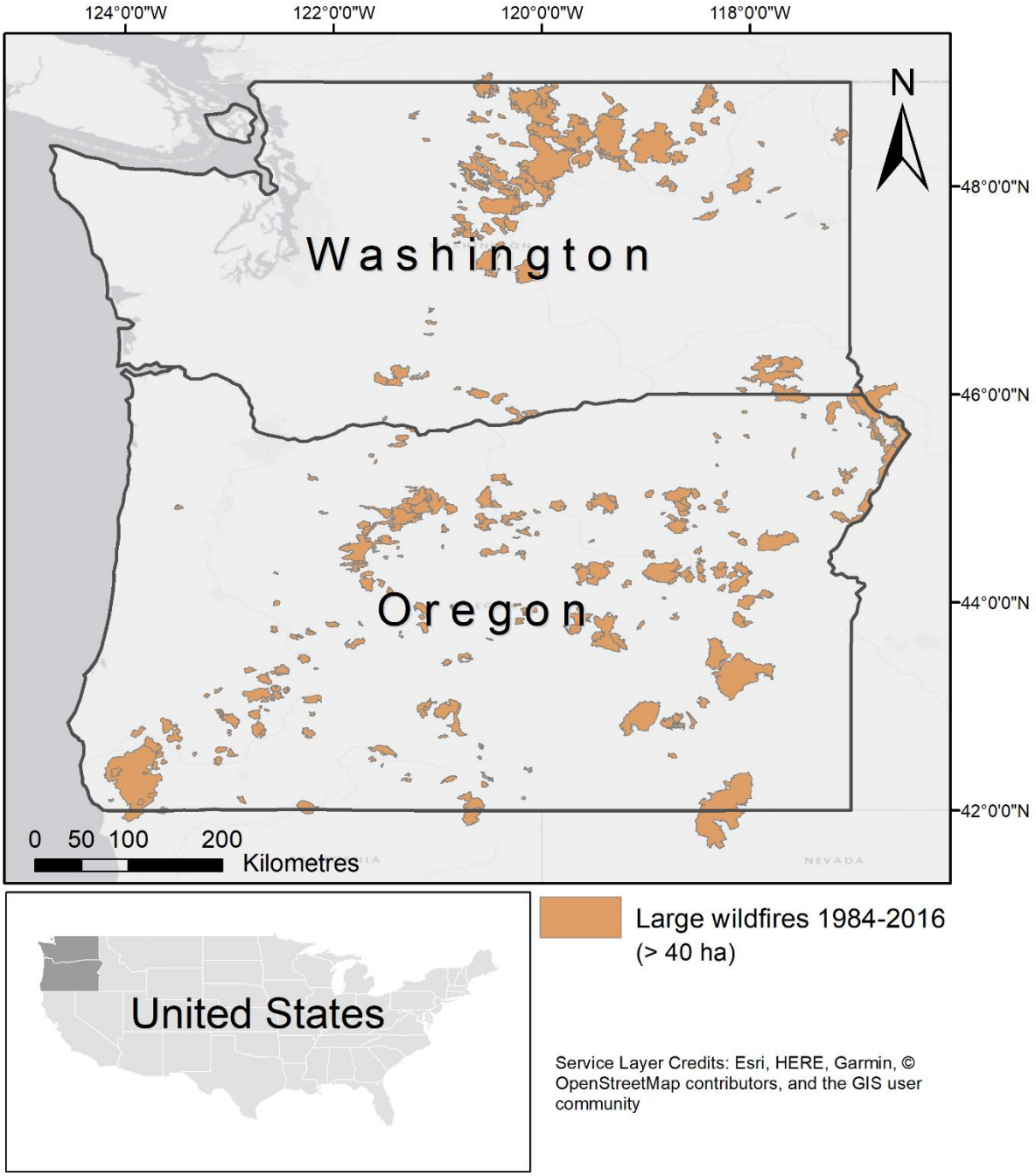


Figure 2.1. Study area (Washington and Oregon, USA) and spatial distribution of large wildfires (e.g., >40 ha) from 1984-2016.

2.2.2 Forest Inventory and Analysis (FIA) data

Since 1999, the Forest Inventory and Analysis (FIA) program has provided comprehensive information on 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 hectares (6,000 acres, Bechtold and Patterson 2005).

One FIA plot comprises measurements from four fixed circular 7.3 metre-radius subplots with plot centres 36 metres apart (Appendix A: Figure A.1.1). 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-year re-measurement cycle.

From 2001 to 2016, 16,239 plots with forested conditions (i.e., at least 10 percent potential cover by live trees) (Bechtold and Patterson 2005) were measured by the FIA program in Washington and Oregon, USA. Among them, 8,184 of the plots were measured twice with a 10-year re-measurement cycle in the data, producing separate measurements for the same plot location. Therefore, each plot can have multiple plot measurements. I considered only plots that had a minimum of 25% forested area, which resulted in a total of 23,150 plot measurements including re-measurements. The topography and land cover variables—physiographic class (PHYSCL), land ownership (OWNS), elevation (ELEV, metres), slope (SLOPE, degrees), and cosine transformation of aspect (cosASP, Weiskittel et al. 2007)—were obtained from FIA data at the plot level and included as covariates to perform matching (Table 2.1). 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, I used the value that covered the greatest proportion of the area. The proportion of forested area (PROPforest) was also calculated for each plot observation.

Table 2.1 Summary statistics of topographic and land cover variables obtained from FIA plot data. For continuous variables, the mean with the standard deviation in parentheses is provided. For categorical variables, the number of plots in each category are specified with the proportions in parentheses.

Variable	Description	Burned (n = 611)	Unburned (n = 22,539)
Topography			
ELEV	Elevation (in metres)	1,241.9 (439.63)	1,073.78 (534.08)
SLOPE	Slope (in degrees)	37.16 (23.88)	30.41 (24.28)
cosASP	Aspect (in 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)

*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.

2.2.3 Fire data

Monitoring Trends in Burn Severity (MTBS, http://www.mtbs.gov/) data from 1984 to 2016 (Figure 2.1) were utilized to identify the presence of wildfires, burn severities, and date of wildfire for each plot. National MTBS burn severity mosaics are provided as tagged image file format (TIFF), classifying burn severities into 4 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 hectares in size or larger.

Burned plots were identified as FIA plot measurements that experienced any wildfire within 5 years prior to the field measurement. In contrast, plot measurements that experienced no fire within at least 10 years prior to the measurement were identified as unburned plots. 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 within 5 years, the measurements after the first wildfire were also excluded to preclude the effect of reburn. A total of 611 burned plots and 22,539 unburned plot measurements were identified (Table 2.2). Approximately 70% of the burned plots were measured before 2011, those plot locations thus retained a single measurement in the data due to the 10-year re-measurement 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 centre and 8 adjacent 30-m pixels from the MTBS burn severity maps (Appendix A: Figure A.1.1).

Table 2.2 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	
3 (Moderate)	145	
4 (High)	145	

2.2.4 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 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 2.3): 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 9 adjacent pixel values (Appendix A: Figure A.1.1). 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 (https://lemma.forestry.oregonstate.edu/).

Table 2.3 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 (ln mm × 100) from PRISM 30-year 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 (degrees 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-September) (degrees 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 (644)	2,361 (697)
TC2	Tasseled cap greenness	787 (444)	1,281 (621)
TC3	Tasseled cap wetness	-1,251 (579)	-722 (593)

2.3 Methods

2.3.1 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 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, I 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 Figure 2.2.

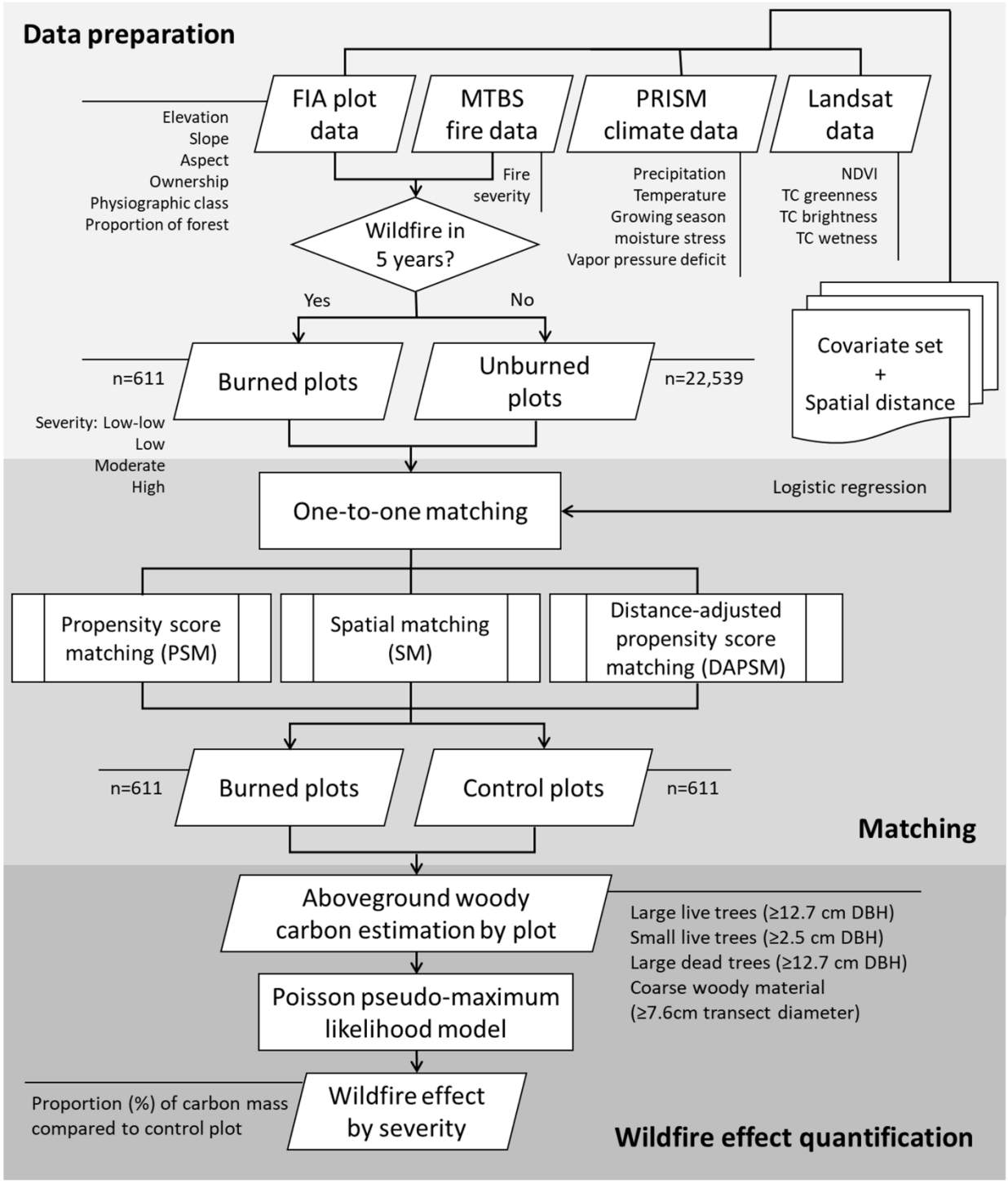


Figure 2.2 Flow of the analysis in this chapter.

2.3.1.1 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

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

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 represents 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 (Table 2.1 and 2.3) were included in the covariate set to estimate the propensity score of each plot with logistic regression.

2.3.1.2 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:

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

where d_{ij} is the Euclidean distance between plot i, j and d_{all} denotes 2-dimensional 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 the match.

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

$$\mathrm{DAPS_{ij}} = w * \left|p(X)_i - p(X)_j\right| + (1 - w) * Dist_{ij} \quad (3)$$

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). I 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, I 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).

2.3.2 Diagnosis of balance between burned and control plots

Once the control plots were obtained through matching with burned plots, I 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 were checked both numerically and graphically following Stuart (2010). The comparisons of the propensity score distribution before and after matching were visually presented as a 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 sixteen 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, and assesses the bias in the observed covariates across groups as a percentage of the pooled standardized deviation (Cohen 1988). ASMD is computed as

$$ASMD_k = \frac{\left|\bar{x}_{k,b} - \bar{x}_{k,c}\right|}{\hat{\sigma}_{k,b}}$$

where $\bar{x}_k$ and $\hat{\sigma}_k$ denotes the mean and the estimated standard deviation of covariate k. b and c represent burned and control group, respectively. Usually, ASMD values below 0.25 indicate that the distributions are balanced between the two groups (Stuart 2010, Stuart and Rubin 2004).

2.3.3 Calculations of aboveground woody carbon mass

For live trees, the regional biomass oven-dry weight (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 diameter at breast height, DBH) and small trees (≥ 2.54 and < 12.7 cm DBH). All tree biomass from a forest inventory plot was summed as per hectare value, and the carbon mass (Mg ha^{-1}) was then computed for each plot by multiplying by 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). The regional biomass oven-dry weights are estimated based on live trees, thus often not applicable to dead trees with missing or broken top parts. I 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^{-1}) 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, I performed exploratory analyses to confirm that the contribution of small dead trees to the aboveground woody carbon mass was minimal.

2.3.4 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 non-negative and can often be zero. I fitted Poisson pseudo-maximum likelihood models with burn severity (Table 2.2, Appendix A.3) as the 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^{-1}) were set as the response variable, and five burn severity classes (Table 2.2) including unburned status from controls (n=611) were used as the explanatory variables. 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). A 95% confidence interval of each estimated coefficient was 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 Inc., Cary, North Carolina, USA).

2.4 Results

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

The 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 (Figure 2.3). 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 (Figure 2.1). DAPSM tended to attain control plots that were farther from the burned plots than the control plots selected under SM

(Figure 2.3). The mean distances between control plots and the closest burned plots were approximately 20.2 km, 3.2 km, 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 (Figure 2.4). 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 A: Table A.1.1). 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 A: Table A.1.2). 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 α=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 (Figure 2.5). 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 (Figure 2.5). ASDM of PROPforest slightly increased after matching in the process of balancing the other covariates, although the increments were minimal (black lines in Figure 2.5). 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.

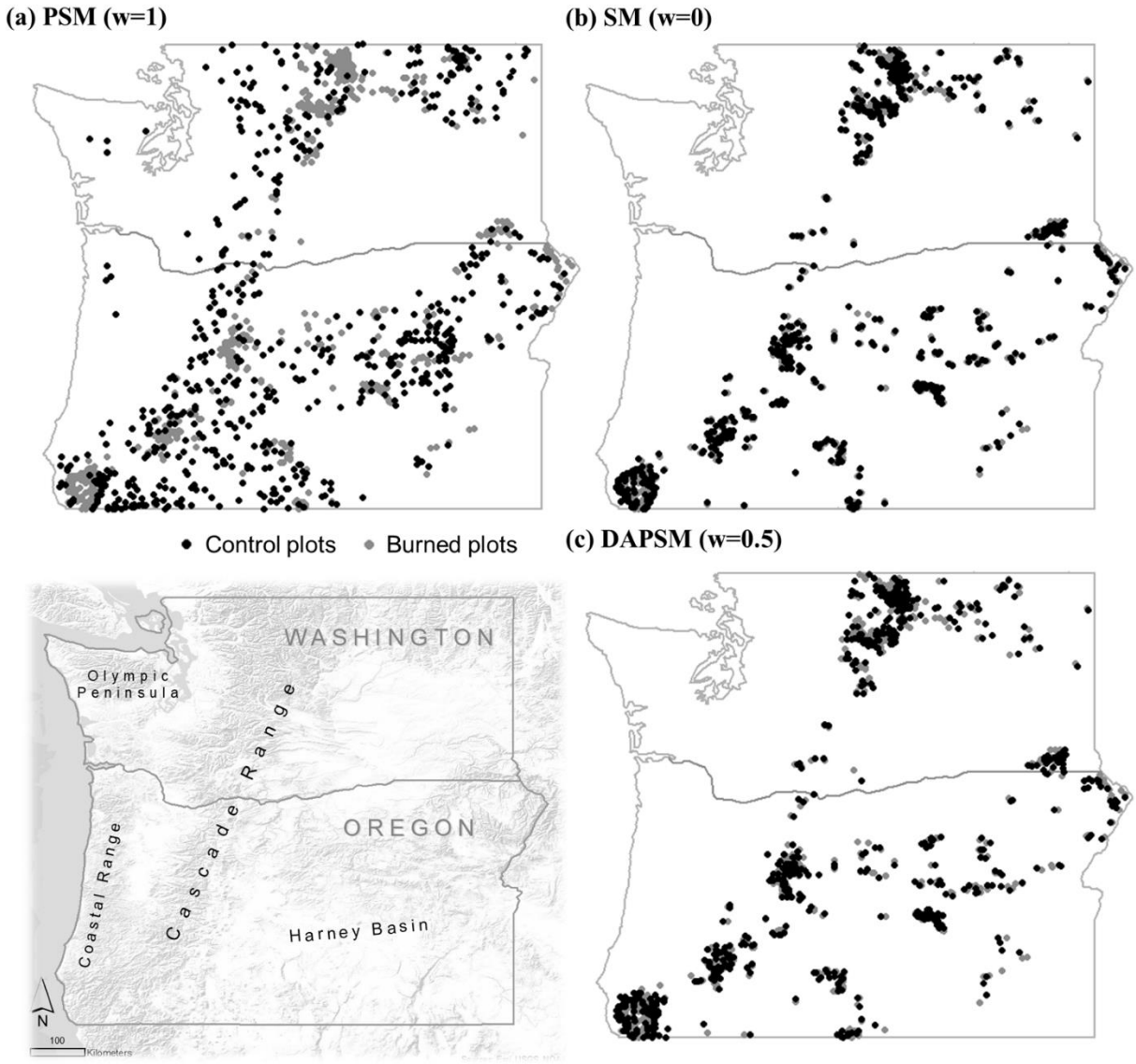


Figure 2.3 Spatial distributions of burned plots (grey dots) and control plots (black dots) over Washington and Oregon under three matching strategies: a) propensity score matching (PSM, w=1), b) spatial matching (SM, w=0), and c) distance-adjusted propensity score matching (DAPSM, w=0.5).

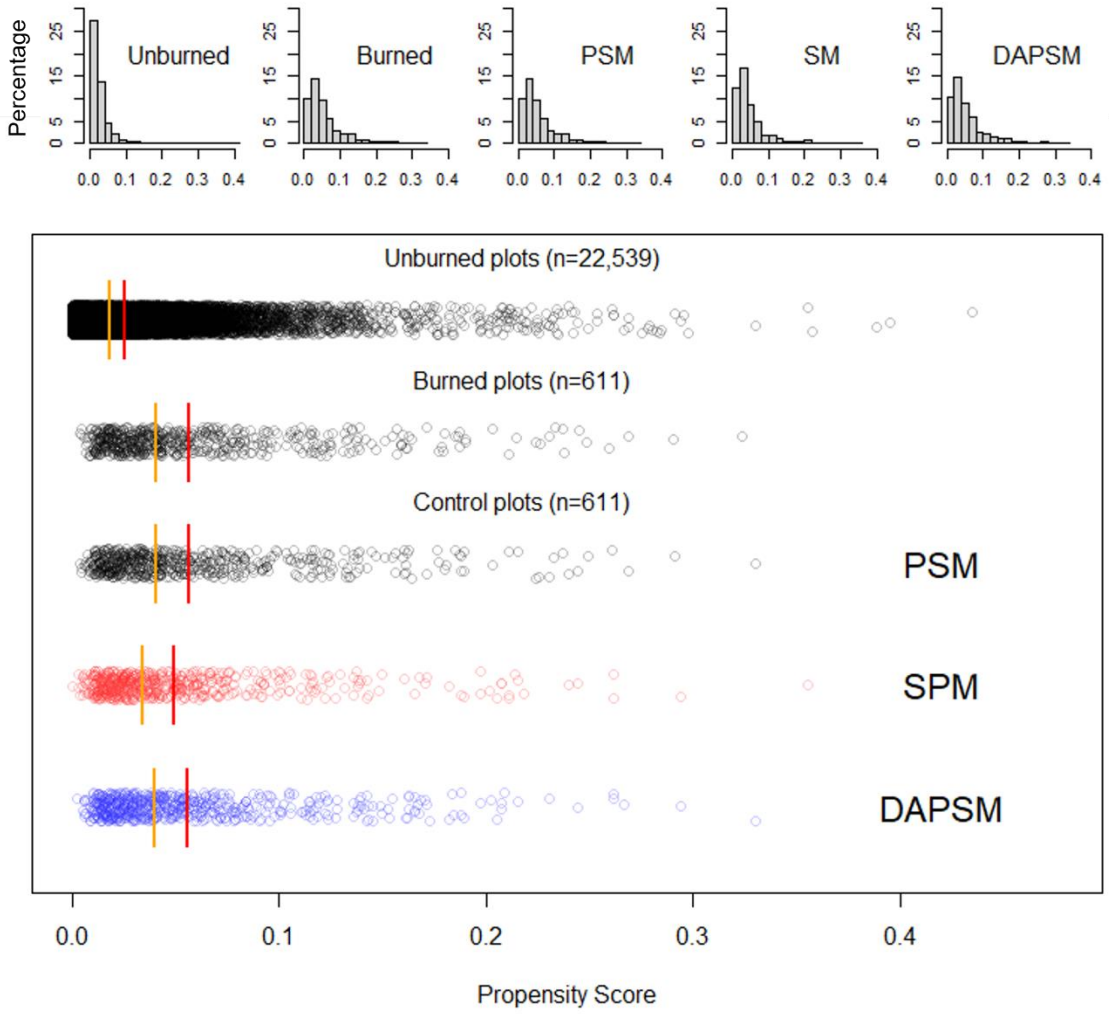


Figure 2.4 Histograms and jitter plot of propensity scores from unburned plots (n=22,539), burned plots (n=611), and control plots (n=611) under three matching scenarios: PSM, SM, and DAPSM. Red and orange lines on the jitter plot denote the mean and the median values of propensity scores in each category, respectively.

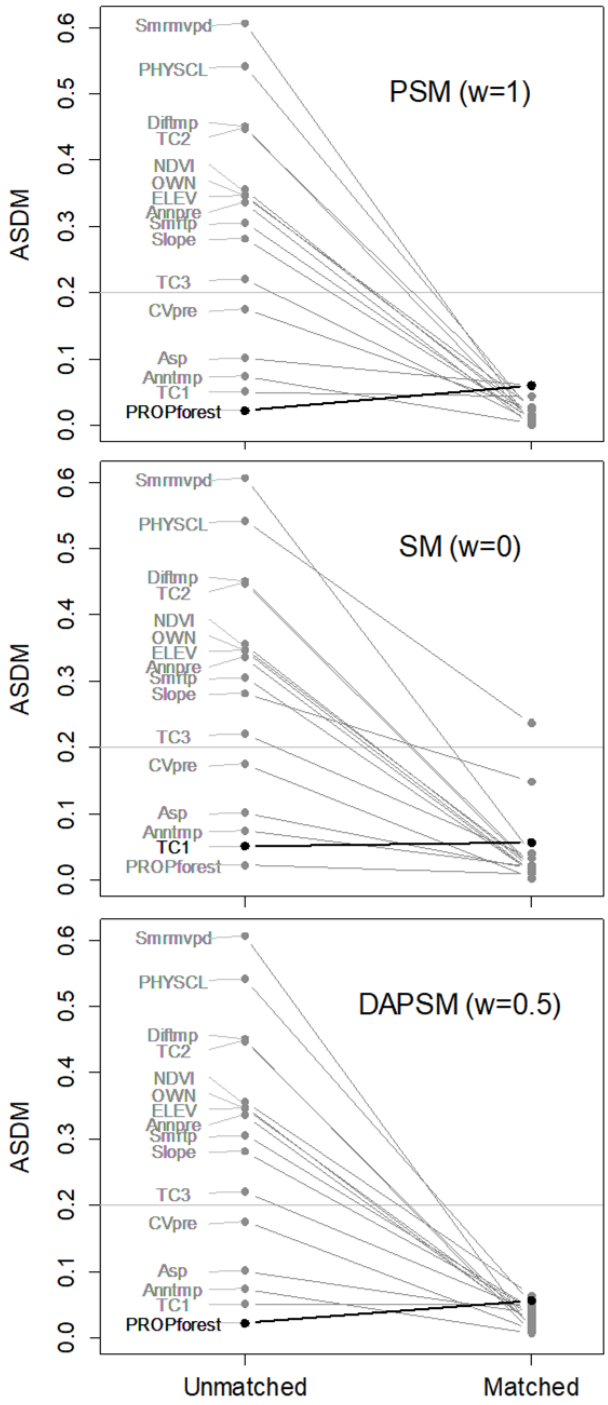


Figure 2.5 Absolute standardized mean difference (ASMD) of 16 environmental covariates before and after matching under three strategies. ASDM of the covariates with grey lines decreased after matching, while black lines denote that ASDM increased after matching. ASDM over the cut-off value (0.2, gray line) denotes the covariate means are different between burned and unburned plots.

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

The mean of total carbon mass (Mg ha^{-1}) 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 Figure 2.6). 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^{-1}) compared to that observed under SM and DAPSM (approximately 70 Mg ha^{-1}). 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 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 Figure 2.6). 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 (Figure 2.6). 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 methods). Examining individual carbon pools, the mean carbon mass

from large and small live trees significantly decreased with increasing burn severity (Figure 2.6). 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 after experiencing wildfire. The mean of carbon mass in CWM did not vary substantially across severities (6.42-8.62 Mg ha^{-1} across all severities), and was the lowest under moderate severity (Figure 2.6). Under high severity fire, the mean carbon mass in CWM was significantly larger than under low-low severity.

2.4.3 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% confidence interval (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%) (Figure 2.7a). 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 (Figure 2.7b) as well as for CWM under low and moderate fire (Figure 2.7e). 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 (Figure 2.7c).

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 (Figure 2.7c). 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.

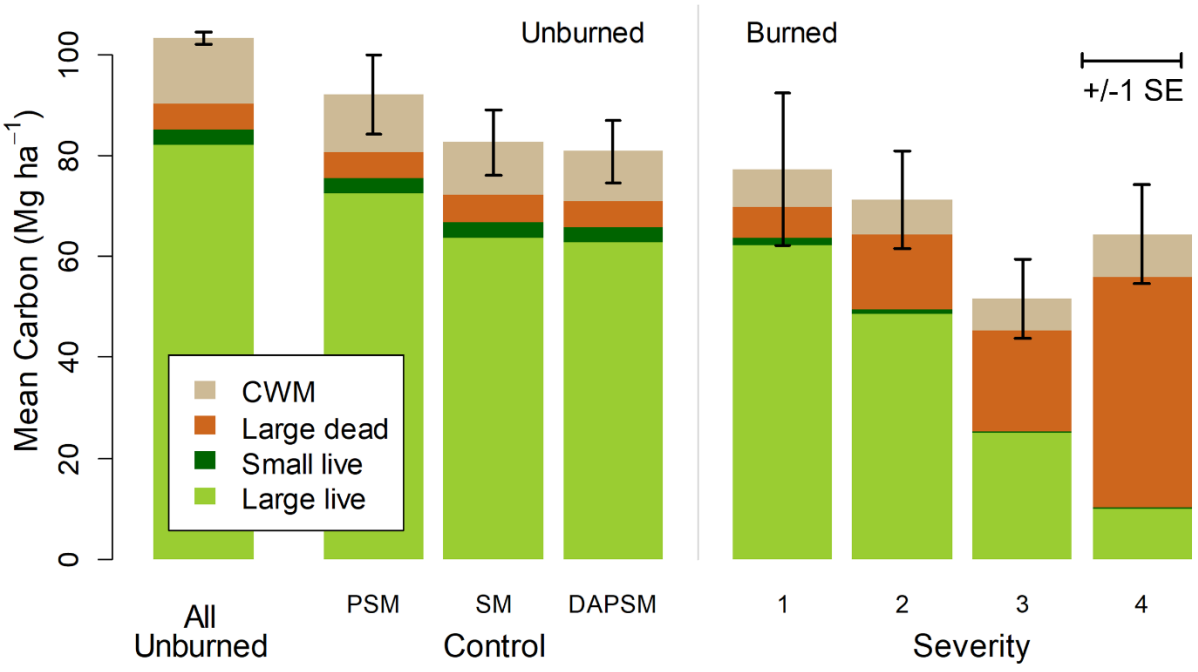


Figure 2.6 Mean total carbon masses (Mg ha^{-1}) by severity classes with their standard errors (SE). All unburned plots (n=22,539) and control plots (n=611 for each method) denote the unburned plots before and after matching under three matching methods: PSM, SM, and DAPSM.

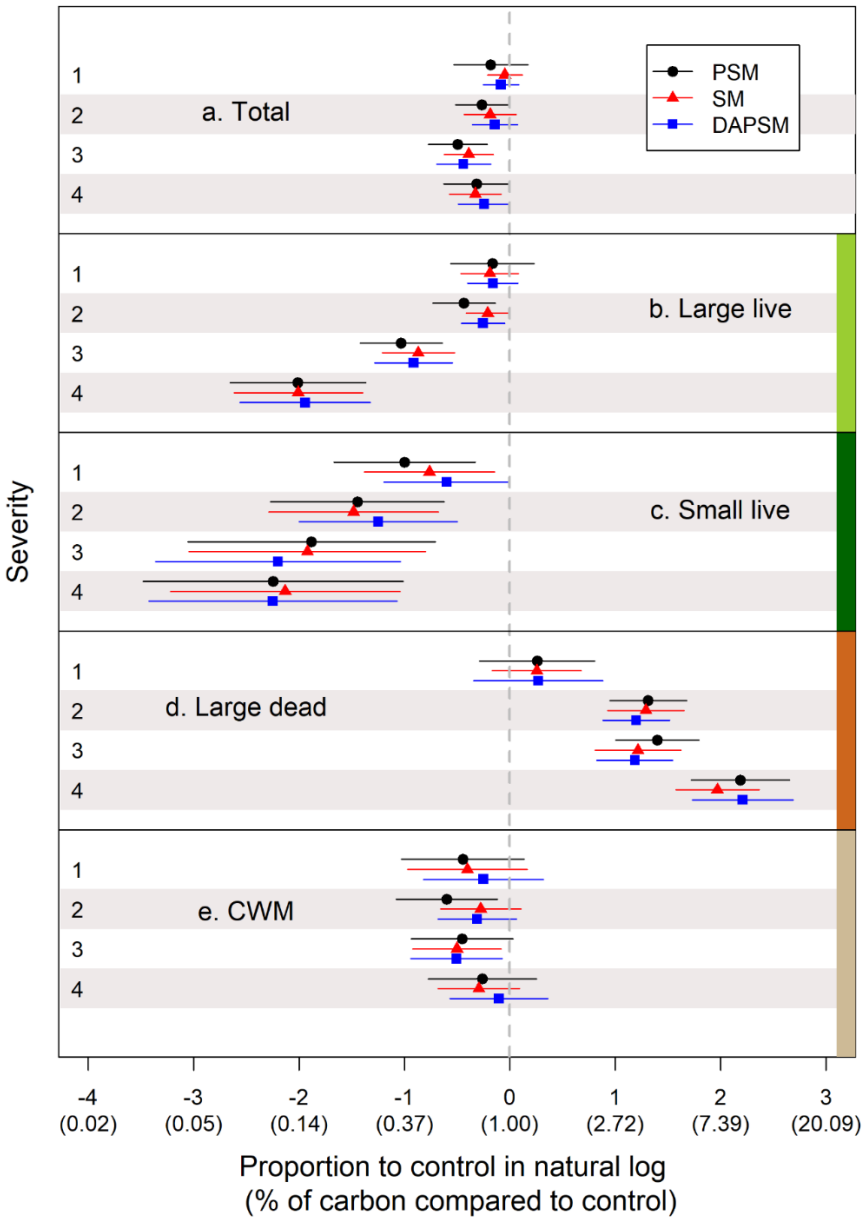


Figure 2.7 Proportion of carbon mass relative to control plots (natural log) estimated from the Poisson pseudo-maximum likelihood model were presented with 95% confidence intervals with Tukey's adjustment. The exponential of the natural log (parenthesis) denotes the fraction of remaining carbon mass 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 7 times under high severity fire. The intervals including the gray vertical line showed no difference from control plots under 95% confidence level.

2.5 Discussion

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

The difference in covariates between burned and unburned plots was mostly eliminated 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 though many studies use measurements of control plots adjacent to the fire perimeter to compare with the burned plots (e.g., Oliveras et al. 2018, Heckman et al. 2013, Ghebrehiwot et al. 2011), 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 2-dimensional distances, regardless of any environmental conditions including elevation.

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. 2019, 2018). 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 the larger sample size I did not have forest type information available for all plots. Specifically, for many of the burned plots that were only measured after they burned, I was lacking this information. Therefore, I 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 (Whittier and Gray 2016, Campbell et al. 2007). 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.

2.5.2 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 (Gross and Rosenheim 2011, Yuan 2010). Wildfires and carbon content in forests are associated with a large number of environmental covariates (Parisien et al. 2012, Ager et al. 2007). 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 this 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 the 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, I 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 be further improved 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. 2006) 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., Nolte et al. 2017, Brandt et al. 2015), where repeated measurements over time were available. In the 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 2006).

2.5.3 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 US 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 pre-fire data or unburned plot measurements. The 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, my 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). The result of little reduction in woody carbon mass by wildfire may also be related to the classification of burned plots. I specified burned plots that experienced any fire within 5 years 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 (Bartels et al. 2016, Raymond et al. 2015). Limiting the criterion for classifying burned plots will allow quantification of 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 pre-existing difference between the moderate-severity fire plots and the other severity plots (Peterson et al. 2019). However, in my 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. I 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.

Chapter 3: Sensitivity analysis on distance-adjusted propensity score matching for wildfire effect quantification using national forest inventory data

3.1 Introduction

Natural disturbances such as wildfires, disease and insect outbreaks drive changes in the forest environment on large scales. Wildfires, one of the major natural disturbance agents, cause the 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. 2021a) to allow for 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 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; > 3,000 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. (2021a, see Chapter 2) 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 a, see Chapter 2), I 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 environments of plots, with and without distance adjustment. I 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), I considered availability of the covariates based on data sources to define four

groups: topography, climate, land cover, and remote-sensing variables. I first assessed how well the propensity score distributions were balanced between the burned and the control plots through matching. I 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. I 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.

3.2 Data

3.2.1 Forest inventory data

I 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 (4,047-square meter in size) that represent a spatially balanced sample of one plot for every 2,400 hectares (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 to 2016, there were a total of 23,150 inventory plot measurements in Washington and Oregon forests including 7,994 remeasurements. I 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 I 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. Burned pots that had fires more than once within 5 years were dropped to exclude the effect of reburn from the analysis. Any pre-fire measurements of burned plots found within the unburned group were also discarded to completely separate burned and unburned groups during the implementation of matching (Woo et al. 2021a). 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. (2021a).

3.2.2 Environmental covariates

For each forest inventory plot, I 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 3.1; additional details in Woo et al. 2021a).

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). I 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 (public or private), physiographic condition, and proportion of forested area (Table 3.1) were measured at subplot level. For the class variables, I used two categories of land ownership (public and private) and three of physiographic class (mesic, xeric, and hydric) (Woo et al. 2021a). I 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 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/).

Table 3.1 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	cosASP	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	OWNS	Categorical: Public, Private
Physiographic class	PHYSCL	Categorical: Xeric, Mesic, Hydric
Climate (CLIM)		
Annual precipitation	annpre	ln mm * 100
Mean annual temperature	anntmp	degrees C * 100
Coefficient of variation in precipitation	cvpre	Variation between mean monthly precipitation of December and July (wettest and driest months)
Mean temperature difference	diftmp	degrees C * 100, difference between mean August maximum temperature and mean December minimum temperature
Mean summer vapor pressure deficit	smrnmvpd	hPa
Growing season moisture stress	smrtp	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	

3.2.3 Aboveground live woody carbon

I 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). I 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.

3.3 Matching methods

3.3.1 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 2-dimensional distances for matching (Woo et al. 2021a).

The quantification of wildfire effects on aboveground live woody carbon using the matching methods followed three steps: First, I estimated propensity scores of burned and unburned plots using logistic regression with the sixteen environmental covariates. Then, I estimated the distance-adjusted propensity scores as a linear combination of the propensity scores and spatial distances between pairs of burned and unburned plots. I used a weight of 0.5 to equally emphasize the observed environmental covariates and the spatial distance, given an exploratory analysis that found other weights (i.e., 0.25, 0.5, and 0.75) did not result in any difference in the effect estimation using the data (Woo et al. 2021a). Finally, I 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. I 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.

3.3.2 Sensitivity analysis of DAPSM under different data availability scenarios

I 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. I 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 the 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 the simulation. Because the number of possible combinations of all covariates is very large, I 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 this study was identified as TOPO-CLIM-REMO-LAND from the least to the most costly to obtain. I 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 3.2).

I 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. I 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 3.2) 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, I 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. I

observed the propensity score distributions and the effect estimates under the data availability scenarios. I also computed mean distance to the control plots from burned plots under each matching method: DAPSM, PSM, and SM.

Table 3.2 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

3.3.3 Evaluation of results

3.3.3.1 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 3.3). 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, I computed the percentage of covariates that are within each of the variance ratio ranges (i.e., < 0.5, ≥ 0.5 and < 0.8, ≤ 0.8 and < 1.2, ≥ 1.2 and < 2, ≥ 2) for the 16 variables considered, even if the model only included a subset of covariates.

In addition to the statistics suggested by Rubin (2001), I 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 3.3). 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 ≥ 1.2) between treated and raw control,

meaning that there was imbalance between burned and unburned plots for those covariates (Table 3.3). When DAPSM was applied, the variance ratios ranged within 0.8 to 1.2 for all covariates (100%, 16 out of 16) (Table 3.3). 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. I graphically presented the comparisons of the descriptive statistics along with the combinations of the three simulation scenarios for effective visualization (Pianosi et al. 2016).

Table 3.3 Baseline statistics of bias and variance ratio in propensity scores (PS) 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 N_t=611, N_{rc}=22,359	Matched control 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
≥ 0.5 and < 0.8	18.75	0
≤ 0.8 and < 1.2	50	100
≥ 1.2 and < 2	31.25	0
≥ 2	0	0

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

3.3.3.2 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. I 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. I 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, I 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.

3.4 Results

3.4.1 Balance in propensity scores under different data availability scenarios

3.4.1.1 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, Figure 3.1a), dropping LAND variables (Figure 3.1b) and CLIM (Figure 3.1c)

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; Appendix B: Figure B.1.1). 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 (Figure 3.1a and 3.1b). When the covariate set excluded CLIM variables (Figure 3.1c), however, the increase in 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 (Figures 3.1a and 3.1b), the bias increased with decreasing n_t for the covariate set that excluded CLIM variables (Figure 3.1c).

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 (Figure 3.1a vs. 3.1d and Figure 3.1b vs. 3.1e). Without CLIM variables, PSM produced much larger bias than DAPSM regardless of n_{rc}/n_t (Figure 3.1f). Without any environmental covariates included in the matching process (i.e., SM with distance only; Figure 3.1e), the mean bias showed a similar pattern to DAPSM without CLIM (Figure 3.1c): small increase in mean bias on average and a sharp increase in mean bias when n_{rc}/n_t<10.

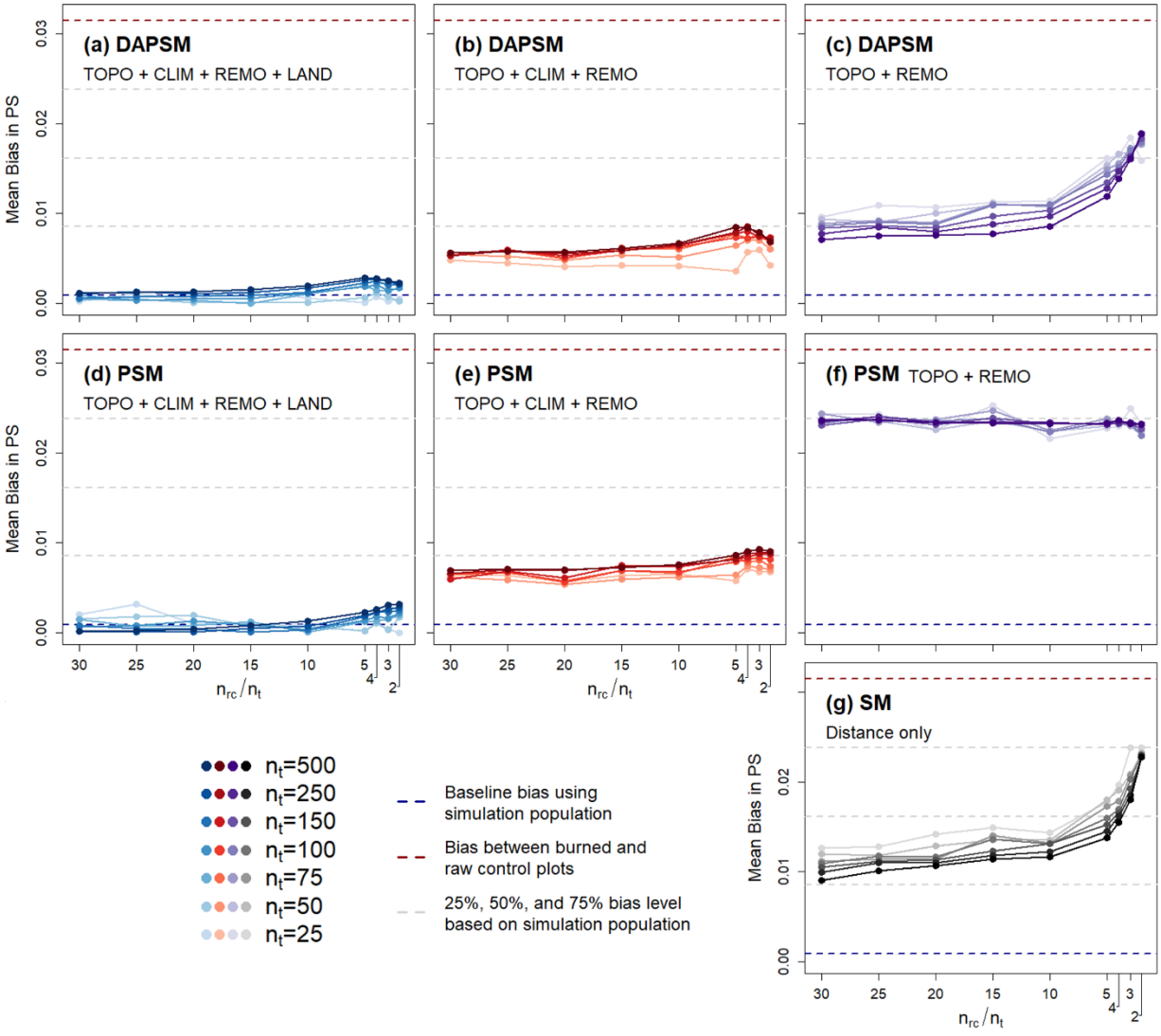


Figure 3.1 Mean bias in propensity scores (PS) under different number of burned plots (n_t), unburned plots to burned plots (n_{rc}/n_t), and covariate set. The burned and unburned plots were matched based on each combination of simulation scenarios using (a) – (c) DAPSM, (d) – (f) PSM, and (g) SM.

3.4.1.2 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 (Figure 3.2). When using the full covariate set, the variance ratio fell within the range of 0.8 to 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 (Figure 3.2a). 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 (Figure 3.2b). Excluding CLIM variables from the covariate set introduced large variability in the variance ratio among different n_t and n_{rc}/n_t (Figure 3.2c). Small n_t (i.e., <75) and n_{rc}/n_t (i.e., ≤5) exhibited extreme variance ratios over 2, indicating imbalance in the 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 (Figure 3.2a vs. 3.2d, Figure 3.2b vs. 3.2e). PSM without CLIM, however, did not reduce the variance in propensity scores between burned and unburned plots (red lines in Figure 3.2f). 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 (Figure 3.2g).

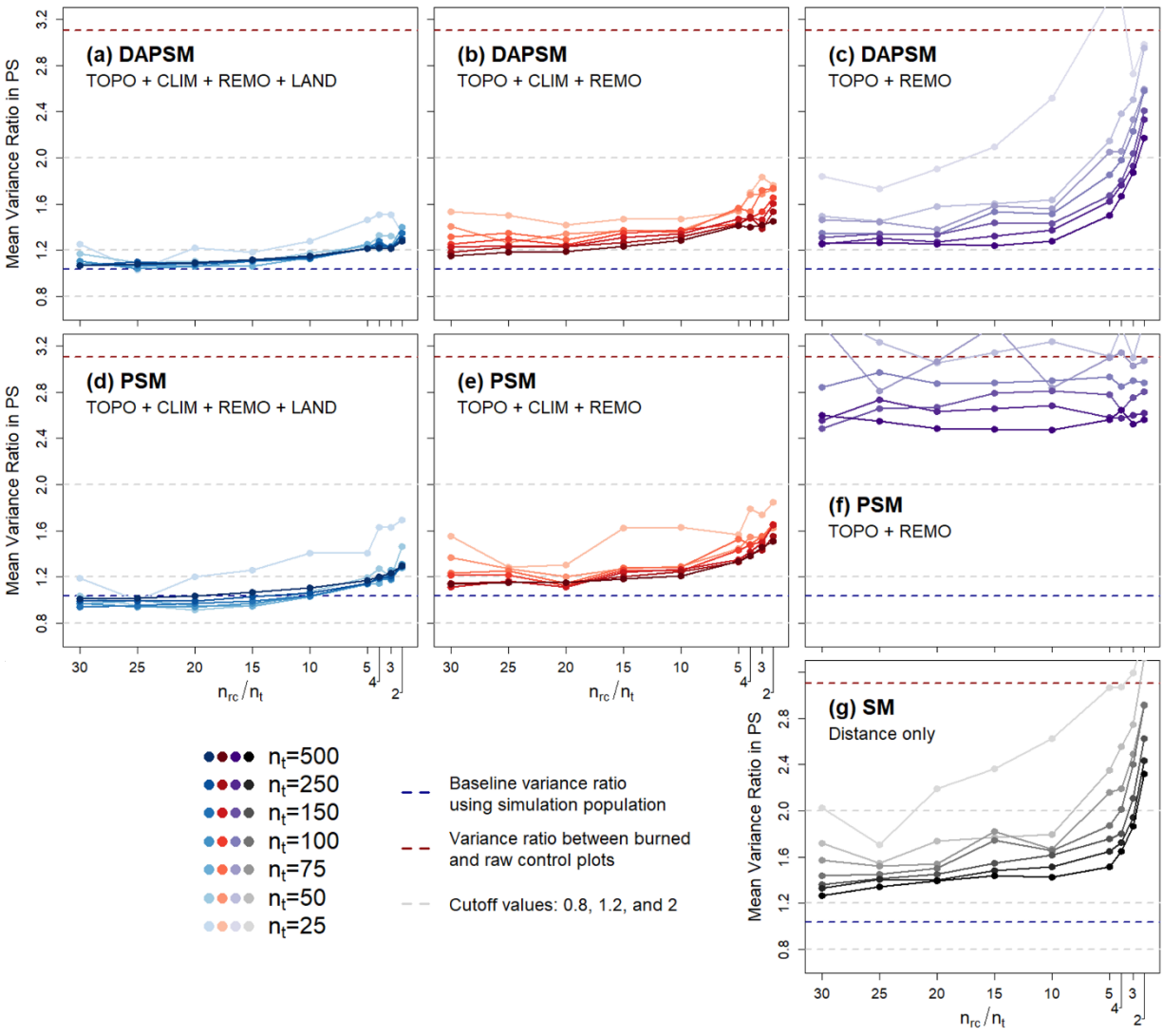


Figure 3.2 Mean variance ratio in propensity scores (PS) between burned and control plots under different combinations of number of burned plots (n_t), unburned plots to burned plots (n_{rc}/n_t), and covariate set, matched based on (a) – (c) DAPSM, (d) – (f) PSM, or (g) SM. The cutoff values of variance 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 Figure B.2.1 in Appendix B.

3.4.1.3 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 (Figure 3.3a-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 also showed more balance in individual covariates than PSM, especially when n_{rc}/n_t>10 (Figure 3.3d)

As n_t decreased, increasingly extreme variance ratios of the residuals were observed for the matched plots, regardless of n_{rc}/n_t (Figure 3.3a-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 (Figure 3.3a-3.3c). TOPO and REMO variables, however, showed little initial imbalance in the variance ratio of residuals between burned and raw control plots (Figure 3.3d) compared to other covariate sets. In that case, PSM introduced more severe imbalance in variance ratio than the raw controls (Figure 3.3d).

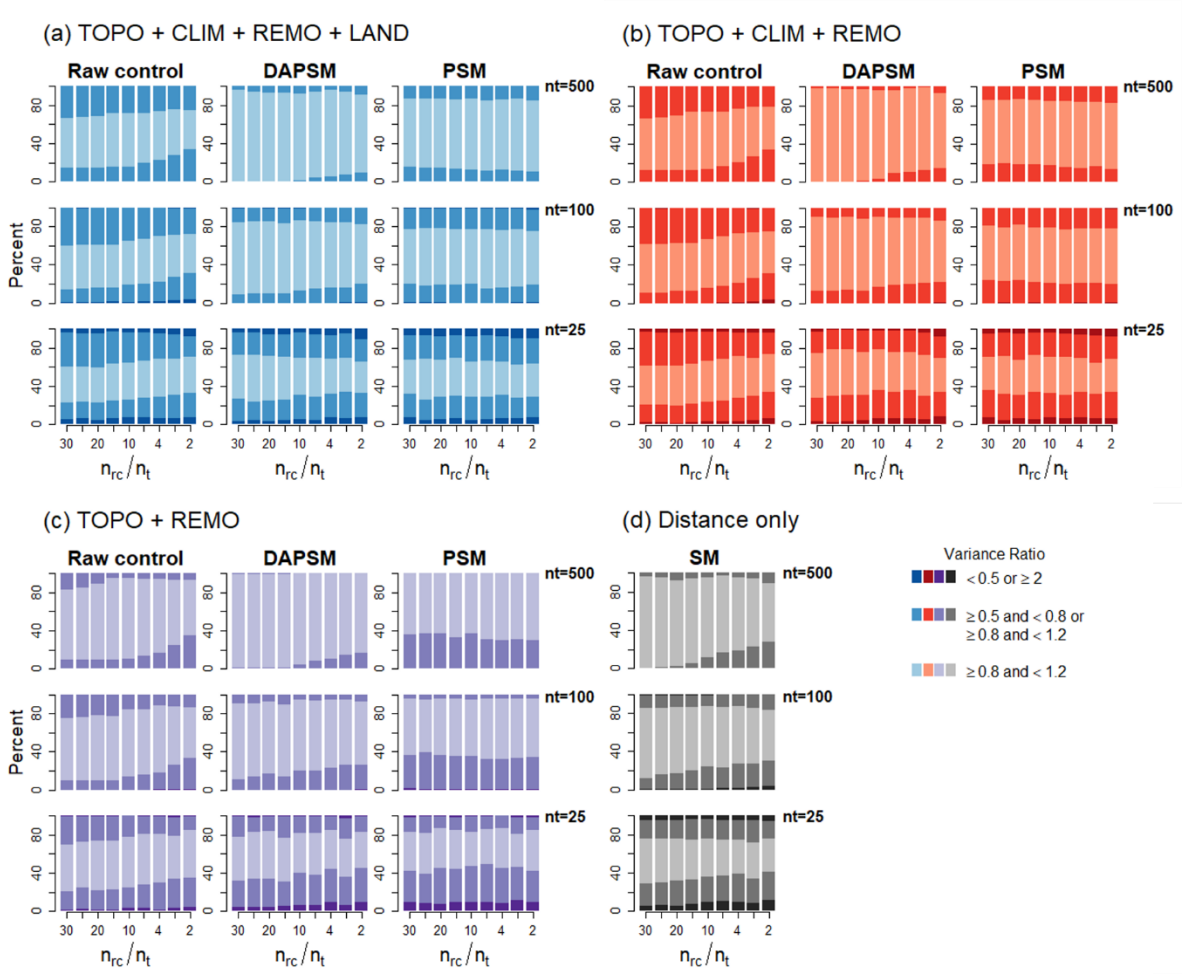


Figure 3.3 Percentages of the number of covariates of which residual variance ratio falls within each range under the covariate set (a) TOPO + CLIM + REMO + LAND, (b) TOPO + CLIM + REMO, and (c) TOPO + REMO using DAPSM and PSM and (d) SM. Darker colours present more extreme variance ratios: < 0.5 or ≥ 2 (extreme), ≥ 0.5 and < 0.8 or ≥ 0.8 and < 1.2, and > 0.8 and ≥ 1.2 (adequate).

3.4.1.4 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 3.4, Appendix B: Figure B.4). 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, anntmp, diftmp, smrnmvpd, and smrtp under PSM with TOPO + REMO; Table 3.4). On the other hand, DAPSM accomplished 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.

Table 3.4 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 Figure B.4 in Appendix B.

		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**
cosASP	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**
OWNS	**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
annpre	**0.349**	0.082	0.091	0.078	0.092	0.091	0.195	0.087
anntmp	0.119	0.057	0.092	0.055	0.104	0.072	**0.205**	0.063
cvpre	0.199	0.053	0.081	0.051	0.080	0.064	0.128	0.056
diftmp	**0.454**	0.073	0.101	0.071	0.107	0.138	**0.219**	0.126
smrnmvpd	**0.610**	0.079	0.077	0.075	0.076	0.193	**0.436**	0.166
smrtp	**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.

3.4.2 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 (Figure 3.4). Dropping CLIM from the covariate set increased the bias and variability under small values of n_t and n_{rc}/n_t (Figure 3.4c). 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 (Figure 3.4a and 3.4b), whereas the average deviation was 43% for the covariate sets without CLIM variables (Figure 3.4c). 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 (Figure 3.4a-c), while the bias of PSM-derived estimates was mostly dependent on covariate sets only, not on n_t or n_{rc}/n_t (Figure 3.4d-f). The matrix plots (Figure 3.5) suggested that the bias was more 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 (Figure 3.5a and 3.5c). 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 (Figure 3.5).

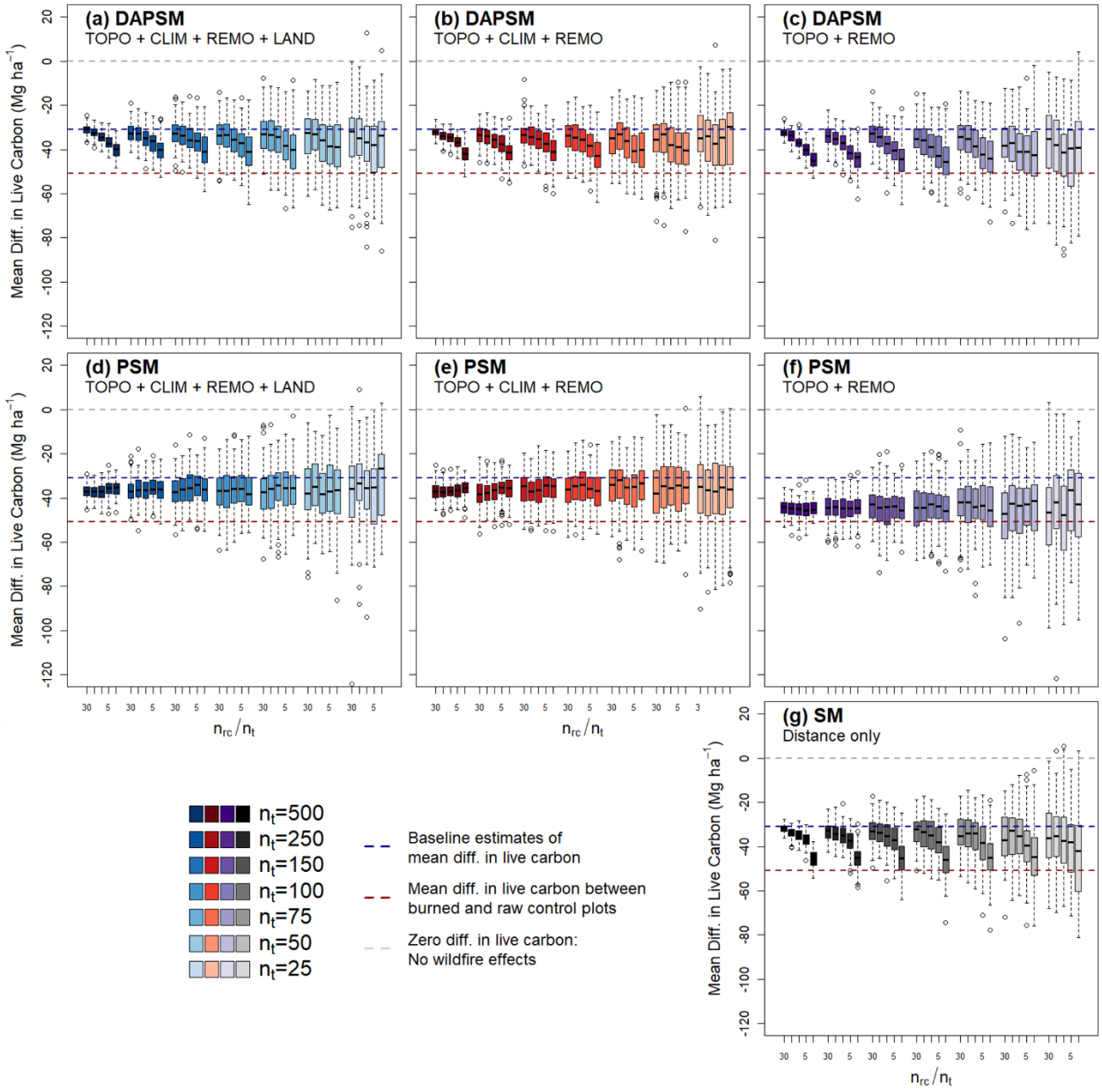


Figure 3.4 Distribution of the estimates of wildfire effect on aboveground live woody carbon under different n_t, n_{rc}/n_t, and covariate sets obtained through (a)-(c) DAPSM, (d)-(f) PSM, and (g) SM. For n_{rc}/n_t, the ratio of 30, 20 ,10, 5, and 2 were plotted for simplicity.

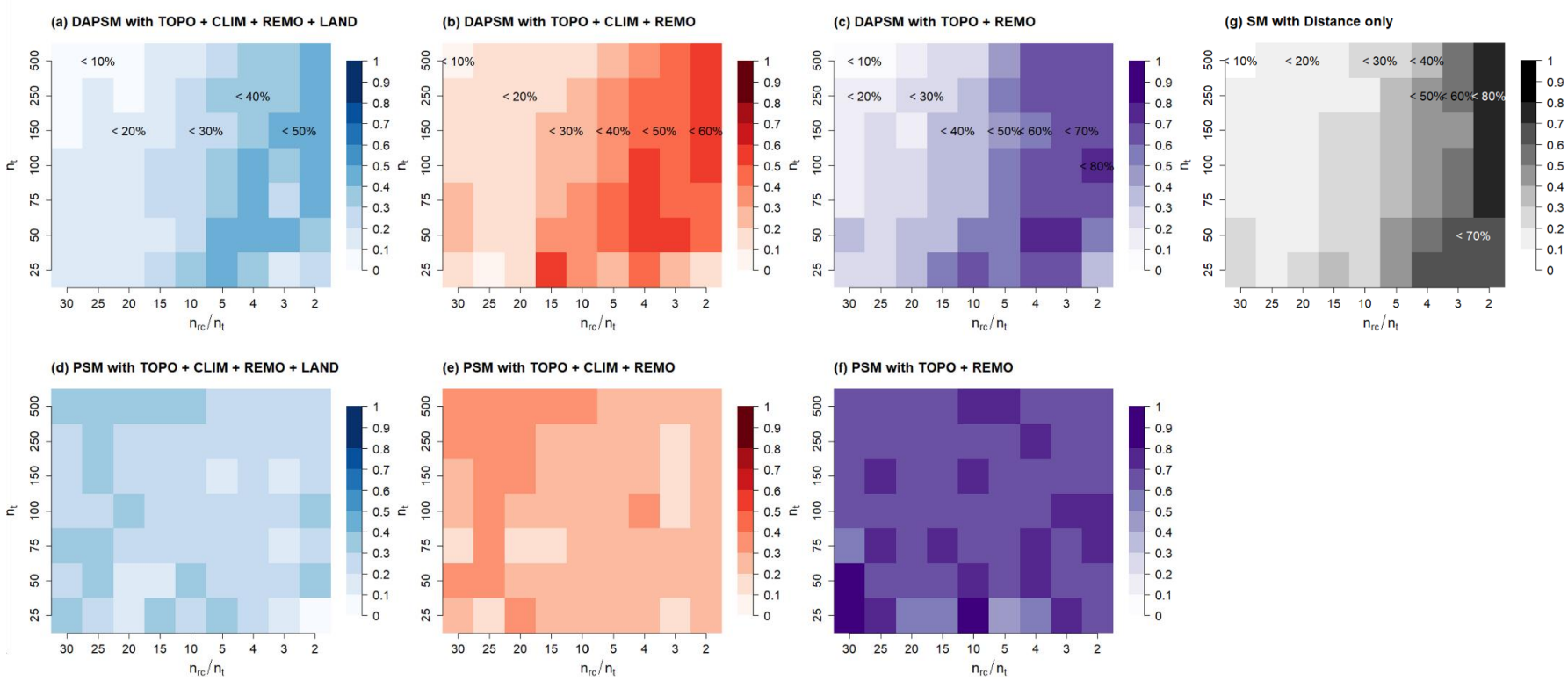


Figure 3.5 Matrix plot of percent bias in the wildfire effect estimates under different combinations of n_t, n_{rc}/n_t, and covariate sets. For example, the combination of n_t=150 and n_{rc}/n_t=15 exhibited less than 20% of bias compared to the baseline estimate under (a) TOPO + CLIM + REMO + LAND, <30% under (b) TOPO + CLIM + REMO, and <40% under (c) TOPO + REMO when using DAPSM.

3.4.3 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 to 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 to 80 km, increasing with decreasing number of observations (Figure 3.6). 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 (Figure 3.6).

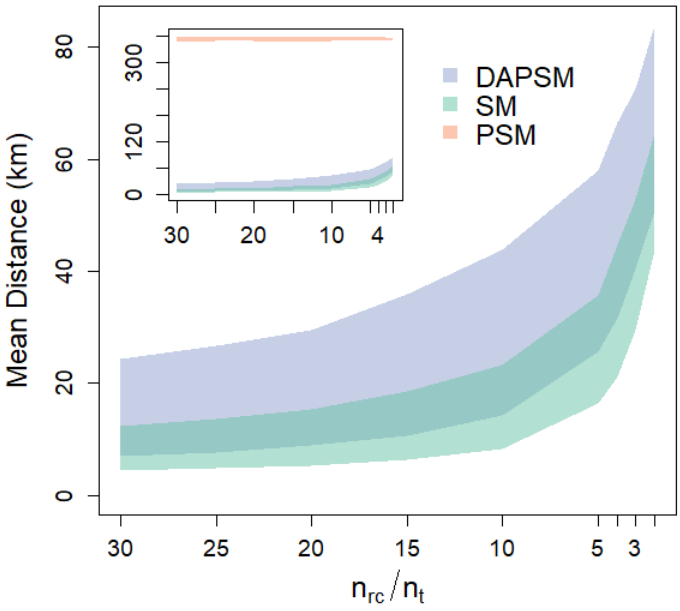


Figure 3.6 Mean distance between matched pairs of burned and control plots under DAPSM, SM, and PSM across n_{rc}/n_t.

3.5 Discussion

3.5.1 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. My 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 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. 2021a; see Chapter 2, 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 this 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 the 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.

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

The effect of spatial distance on the performance of the matching was 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, this 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, I 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 to 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 this 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 1,000 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. The simulation in this research, 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 this 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 the estimated wildfire effects. Nevertheless, my 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.

3.5.3 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 for continuous causal effect analysis of ecological data.

Chapter 4: Quasi-experimental methods to support case studies of wildfire impacts

4.1 Introduction

Wildfire is a global issue altering forest structure and composition, accelerated by climate changes that prolong dry seasons (Halofsky et al. 2020, Haughian et al. 2012). In south-central British Columbia (BC), Canada, small and large wildfires with different severities have continuously reshaped forest conditions for decades (Robinne et al. 2020), including stand structure (Marcoux et al. 2015), fuel loadings (Chen et al. 2015), and nutrient cycles (Smithwick et al. 2005). The year 2017 was one of the most destructive fire seasons in BC wildfire history, with over 1,353 wildfires burning more than 1.2 million hectares of forests (Crowley et al. 2019). These wildfires caused substantial changes in forest structure and fuel characteristics (Coogan et al. 2019), as well as forest carbon mass (Johnson et al. 2007). The influence of wildfires on forest conditions is projected to escalate due to the dry conditions of south-central BC (Hanes et al. 2019, Haughian et al. 2012). Assessing the impact of the wildfires on forest carbon and various environmental changes is crucial for fuel management to preserve ecological and social communities (Hessburg et al. 2021), and to strategically prepare for future disasters (Yong and Lemyre 2019).

Research efforts to examine the impact of wildfires on forest ecosystems often focus on conducting case studies across regions, due to practical difficulties for real-world experiments and stochastic characteristics of wildfires. For example, case studies have been used to investigate the long-term effect on soil chemical properties in a Mediterranean forest (Francos et al. 2018), the nitrogen concentration and production from burned watershed streams in Alberta

(Bladon et al. 2008), and the surface fuel carbon storage in coastal-transitional forests in BC (Peterson et al., 2019). These case studies provide detailed understanding of individual and local fire impact on forest attributes including soil nutrients, timber volume, and carbon, which can be applied to other forest ecosystems where the forest environments are similar (Alexander and Thomas 2003). To allow broader interpretations of the inferences from case studies and to increase the potential for further application of the inferences to other forest ecosystems in a different location, it is necessary to establish cause-and-effect relationships between the ecological event – in this case wildfires – and the outcome variables affected by wildfires (Grace and Irvine 2020).

Case studies of wildfire impacts have some common limitations when scaling up from an individual fire. First, most case studies do not allow the establishment of causal relationships, because they lack pre-fire data or controls that are available in experimental settings (Larsen et al. 2019, Butsic et al. 2017). Case studies with field sampling often utilize measurements from unburned sample plots installed within or adjacent to the fire perimeters as comparison to the measurements from burned plots (e.g., Francos et al. 2018, Rochester et al. 2010, Peterson et al., 2019). However, identifying comparable control plots during field campaigns is challenging. Few studies present explicit analyses showing that the environmental characteristics of unburned sample plots are comparable to those of the burned sample plots. Instead, studies include a short description that some of the topographic attributes such as aspect, elevation, and slope were similar between burned and control plots (e.g., Francos et al. 2018, Rochester et al. 2010). In addition, case studies are often based on limited sample size due to small spatial scales or practical issues (Bladon et al. 2008) such as large fire size in space, remote access, and cost of field work. The sample sizes from the case studies listed above (i.e., Francos et al. 2018,

Peterson et al. 2019, and Rochester et al. 2010) ranged from 37 to 55 plots across different classes of burn severity (e.g., unburned, low, moderate, high). The lack of comparable controls and limited sample sizes may make unbiased and precise quantification of wildfire impacts based on case study data difficult.

Quasi-experimental methods can mitigate bias that arises due to the lack of comparable controls (Campbell and Stanley 2015). Quasi-experimental methods mimic randomized experiments to examine impacts of treatments that are practically infeasible to randomize. In the context of wildfire research, survey data from inventory programs is often available to act as unburned control plots (e.g., Forest Inventory and Analysis program; Bechtold and Patterson 2005, National Forest Inventory plots in Canada; Gillis et al. 2005). Therefore, quasi-experimental methods are a powerful way to identify unburned plots that are similar to burned plots to enhance wildfire research (Woo et al. 2021a, see Chapter 2). Propensity score matching (PSM) is a quasi-experimental method used to select comparable controls based on the probability of wildfire occurrence at different plots estimated from environmental covariates (Rosenbaum and Rubin 1983). Similar propensity scores between burned plots and matched control plots indicate equivalent probabilities of wildfire occurrence conditional on environmental covariates (Stuart 2010).

Distance-adjusted propensity score matching (DAPSM) improves the matching procedure for selecting control plots; however, some knowledge gaps remain. While the standard propensity score accounts for the selected environmental covariates, DAPSM considers the influence of unobserved covariates by accounting for spatial proximity among the plots (Papadogeorgou et al. 2019). DAPSM outperformed PSM for quantifying regional wildfire effects on forest carbon using the national forest inventory data from the United States (Woo et al. 2021a, see Chapter 2).

Yet, DAPSM might be only reliable when the sample size is sufficiently large (e.g., ≥100 burned plots) (Woo et al. 2021b, see Chapter 3). If the sample size is small, e.g., as occurs in most case studies, matching with replacement has been proposed to decrease the bias in the propensity scores between the treated and the control groups (Stuart 2010). Moreover, the spatial arrangement of burned and unburned plots may influence the performance of DAPSM with and without replacement, because relative locations of sampling units affect the spatial analysis procedure (Borcard and Legendre 2002). The effects of the spatial distribution of data on the performance of DAPSM to assess wildfire impacts have yet to be assessed empirically.

This study evaluates the performance of DAPSM with and without replacement for case studies analyzing wildfire impacts on forest carbon mass with small sample size. For the case studies of wildfire impacts, three different post-fire datasets were collected from forests in south-central BC following the 2017 fire year. One of the datasets had pre- and post-wildfire measurements for the sampled plots, while the other two contained unburned plots sampled after the wildfire events. Aside from the sampled unburned plots in the datasets, I used forest inventory data from the Forest Analysis and Inventory Branch (FAIB) of the Province of BC to match unburned controls to the burned plots. The performance of DAPSM with and without replacement was evaluated in terms of the balance in environmental covariates achieved between the burned and control plots for each dataset. Because the FAIB data uses a grid-based systematic sampling while the three fire datasets had different spatial distributions of the sampled plots, I also examined the impact of different spatial scales of the burned and control plots on the matching performance. The wildfire impacts on forest woody carbon mass were estimated as the difference between burned and control plots obtained from DAPSM with and without replacement, in comparison with the pre-fire data.

I posed the following questions and hypotheses: 1. How similar were environmental covariates between burned plots and unburned plots sampled after the wildfire? How similar were the control plots selected by DAPSM to the burned plots in terms of their environmental covariates? 2. Does either DAPSM with or without replacement improve the similarity in environmental covariates for fire datasets with small sample sizes? If so, do the spatial distributions of burned and unburned controls affect the performance of DAPSM with replacement relative to without replacement? 3. Do the controls selected by DAPSM with and without replacement produce similar estimates of wildfire impacts on forest woody carbon compared to pre-burn data?

I hypothesize that similarity between burned and unburned control plots will improve by selecting control plots using DAPSM with replacement. I anticipate that the use of the quasi-experimental method will produce estimation of wildfire impacts on forest woody carbon comparable to the estimation using pre-burn data. This research, as far as I know, is the first study that implemented DAPSM with replacement for ecological data. Based on the empirical forest inventory plot data and post-wildfire data, this research will extend understanding of the performance of quasi-experimental methods for data with small sample size, especially in relation with the spatial distributions of the data. I anticipate that this research will be helpful for future wildfire studies with limited sample size where spatial adjustment is critical in designing impact analyses.

4.2 Study area

The study area is in the south-central part of BC, Canada, an area of approximately 500 square kilometers (km) from 119–126° west and 50–54° north, encompassing the three case studies (Figure 4.1). Located in the rainshadow of the Coast Mountains, but west of the Cariboo Mountains, the study area has a continental and dry climate with <500 mm annual precipitation

(Canadian Climate Normals 1971-2000 Station Data, https://climate.weather.gc.ca/climate_normals/). The study area includes eight biogeoclimatic zones that vary along elevational and latitudinal gradients (Meidinger and Pojar 1991). The warmest and driest zones are the Bunchgrass (BG; elevation = 150–600 meters above sea level), Ponderosa Pine (PP; 250–900 m), and Interior Douglas-fir (IDF; 350–1450 m) biogeoclimatic zones, comprising open- and closed-canopy forests dominated by of ponderosa pine (*Pinus ponderosa* Dougl. Ex Laws.) and interior Douglas-fir (*Pseudotsuga menziesii var glauca* (Beissn.) Franco), with minor components of trembling aspen (*Populus tremuloides* Michx) and paper birch (*Betula papyifera* Marsh.). With increasing elevation, the Sub-boreal Pine – Spruce (SBPS; 800-1300 m, Montane Spruce (MS; 1250–1700 m), and Engelmann Spruce–Subalpine Fir (ESSF; 1500–2300 m) zones include increasing components of lodgepole pine (*P. contorta* var. *latifolia* Douglas), hybrid spruce (*Picea engelmannii* Parry ex Engelm x *Picea glauca* (Moench) Voss), and subalpine fir (*Abies lasiocarpa* Hook (Nutt.)). In the northern part of the study area, the Interior Cedar Hemlock (ICH; 400–1500 m) and Sub-boreal Spruce (SBS; 750-1250 m) zones also include western red cedar (*Thuja plicata* Donn ex D.Don).

Historical wildfire regimes in the study area included mixed-severity regimes in the drier zones, with stand-maintaining surface fires burning at intervals of 10-40 years (Brookes et al. 2021). Across the study area infrequent stand-initiating fires burn at intervals of 100-250 years (Hessburg et al. 2019, Brookes et al. 2021). Fire intervals have been predicted to decrease while mean and maximum fire sizes are predicted to increase in accordance with climate change (Nitschke and Innes 2013, Hanes et al. 2019).

4.3 Data

4.3.1 Fire data

I obtained three datasets of post-fire ground plots from wildfires in the south-central part of BC: the Elephant Hill Fire data, Prouton Lakes Fire data, and plots from a group of six fires that burned across the Kamloops and Cariboo Fire Centres (hereafter south-central BC fires data; Figure 4.1). The three datasets included different sampling approaches and sampling intensity (Table 4.1). The Elephant Hill Fire data and the Prouton Lakes Fire data used geographically stratified random sampling within and adjacent to the fire perimeter. In contrast, the south-central BC data are from permanent sample plots collected by the provincial Forest Analysis and Inventory Branch (FAIB). The south-central BC data included the subset of the FAIB plots within the perimeters of six fires, and remeasurements of the plots conducted in the September 2017 after the fires. The south-central BC data included FAIB plots within the perimeter of the Elephant Hill Fire (Figure 4.1), but the locations and the number of the plots differ from the plots in the stratified-random dataset from the Elephant Hill Fire.

4.3.1.1 Elephant Hill Fire data

The Elephant Hill Fire (K20637) ignited on July 6th, 2017, and burned 191,865 hectares with mixed severity. After the fire, 105 sample plots were installed by the University of British Columbia (UBC) Tree Ring Lab (https://treering.forestry.ubc.ca/) with geographically stratified random sampling based on elevation and burn severity (Figure 4.1). The stratification included three elevation groups (BG and PP zones; IDF zone; MS and ESSF zones) and burn severity (unburned, low, moderate, and high based on remote sensing; Pickell et al. 2020) and plots were randomly located in patches that were within 1 km of a road. Some unburned plots were sampled

from forests adjacent to the fire perimeter. At each plot location, a circular plot with 11.28 m fixed radius was established to measure all live trees and standing dead trees ≥5cm diameter at breast height (DBH). Plot information such as geographical coordinates of the plot centers, burn severity (i.e., low, moderate, and high), and topography including elevation, slope, and aspect were recorded.

4.3.1.2 Prouton Lakes Fire data

The Prouton Lakes Fire (C30870) was a relatively small fire in UBC's Alex Fraser Research Forest that ignited on July 8th, 2017, and burned 859 hectares (Table 4.1). In 2018, 46 sample plots were installed in the northern part of the fire area (Figure 4.1) as part of a post-fire recovery experiment comparing forests that were unburned with those burned at moderate or high severity. Plots were located in patches of forest that burned at moderate (<80% tree mortality) and high (80-100% tree mortality) severity, and in adjacent unburned forests with similar structure, composition, and topographic attributes. Plots were measured following the same protocol as the Elephant Hill Fire plots (Table 4.1).

4.3.1.3 South-central BC fires data

The south-central BC fires data consisted of FAIB plots from six different fire perimeters (Figure 4.1). The fires ignited between July 6th and 7th, 2017, and the post-fire data from 35 FAIB plots were collected in September 2017. The FAIB plots were measured between 2009 and 2016 before the fire. Thus, the data contain both pre- and post-fire measurements. The south-central BC fires data consisted of different types of the FAIB forest inventory monitoring plots: 1 National Forest Inventory (NFI) plot, 13 Change Monitoring Inventory (CMI) plots, 7 Young Stand Monitoring (YSM) plots, and 14 Supplementary sample plots (SUP). Details of the plot

types that are part of the FAIB data are explained in the next section. Only one sampled unburned plot that fell within the fire perimeters was included in the south-central BC fires data (Table 4.1).

Table 4.1 Descriptions of the three datasets used for the analysis.

	Elephant Hill Fire	**Prouton Lakes Fire**	**South-central BC**
Sampling method	Stratified random sampling		FAIB plot locations
# of plots	105	46	35
High	33	20	9
Moderate	26	19	15
Low	30	-	10
Unburned	16	7	1
Measurement year	Jun-Jul 2018	Jul-Aug 2018	Sep 2017
Fire ID	K20637	C30870	C40621, C10970, C50744, K20637, C50647, C10784
Fire area (km^2)	1,920	8.6	57, 219, 250, 1920, 2393, 5209
Elevation range (m)	475-1,700	1,028-1,251	841-1,580
BEC* zone classification	BG, PP, IDF, MS, ESSF	ICH, SBS	IDF, MS, SBPS

*BEC: Biogeoclimatic Ecosystem Classification (Pojar et al. 1987)

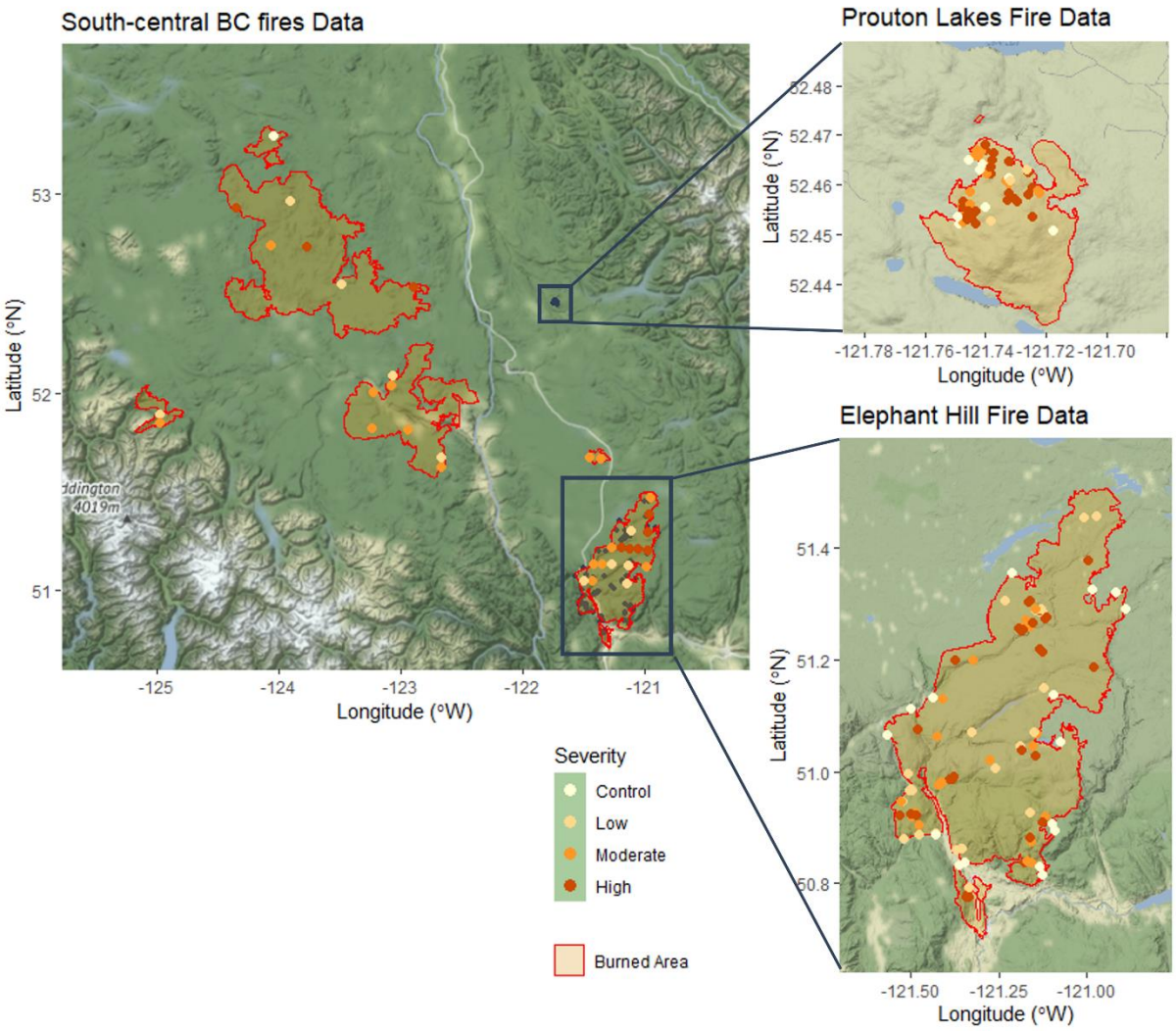


Figure 4.1 Fire perimeters and sampling locations of the three fire datasets: South-central BC fires, Elephant Hill Fire, and Prouton Lakes Fire.

4.3.2 Forest Analysis and Inventory Branch (FAIB) data

The FAIB data in this study consisted of 632 forest inventory plots in the study area, measured between 2000 and 2017 (pre-burn) within and outside the fire perimeters of the three fire datasets (Figure 4.2). The FAIB data included four different types monitoring plots: 56 National Forest Inventory (NFI) ground plots, 248 Change Monitoring Inventory (CMI) plots, 219 Young Stand Monitoring (YSM) plots, and 109 Supplementary sample plots (SUP). The population of NFI

plots covers the entire forests in Canada, sampled from a 20 km x 20 km national systematic grid (Gillis et al. 2005). Other types of plots including CMI, YSM, and SUP were developed by BC provincial programs for specific purposes such as monitoring average change in forest attributes over the province (i.e., CMI), timber management in 15 to 50-year-old stands (i.e., YSM), and increasing sample size through intensified subsamples (i.e., SUP). CMI follows the sampling method from the NFI system, sampled from a standard 20 km grid with a regular measurement interval of 5-10 years (Fenger and Bradford 2013). Meanwhile, the grid size of YSM and SUP varies across 4 x 4 km, 5 x 10 km, 10 x 10 km, or 10 x 20 km. All plots contain site information (e.g., geographical coordinates, elevation, and BEC zone classification), and lists of all live and dead trees measurements (e.g., species, DBH, height, and crown class) on a circular plot with 11.28m fixed radius.

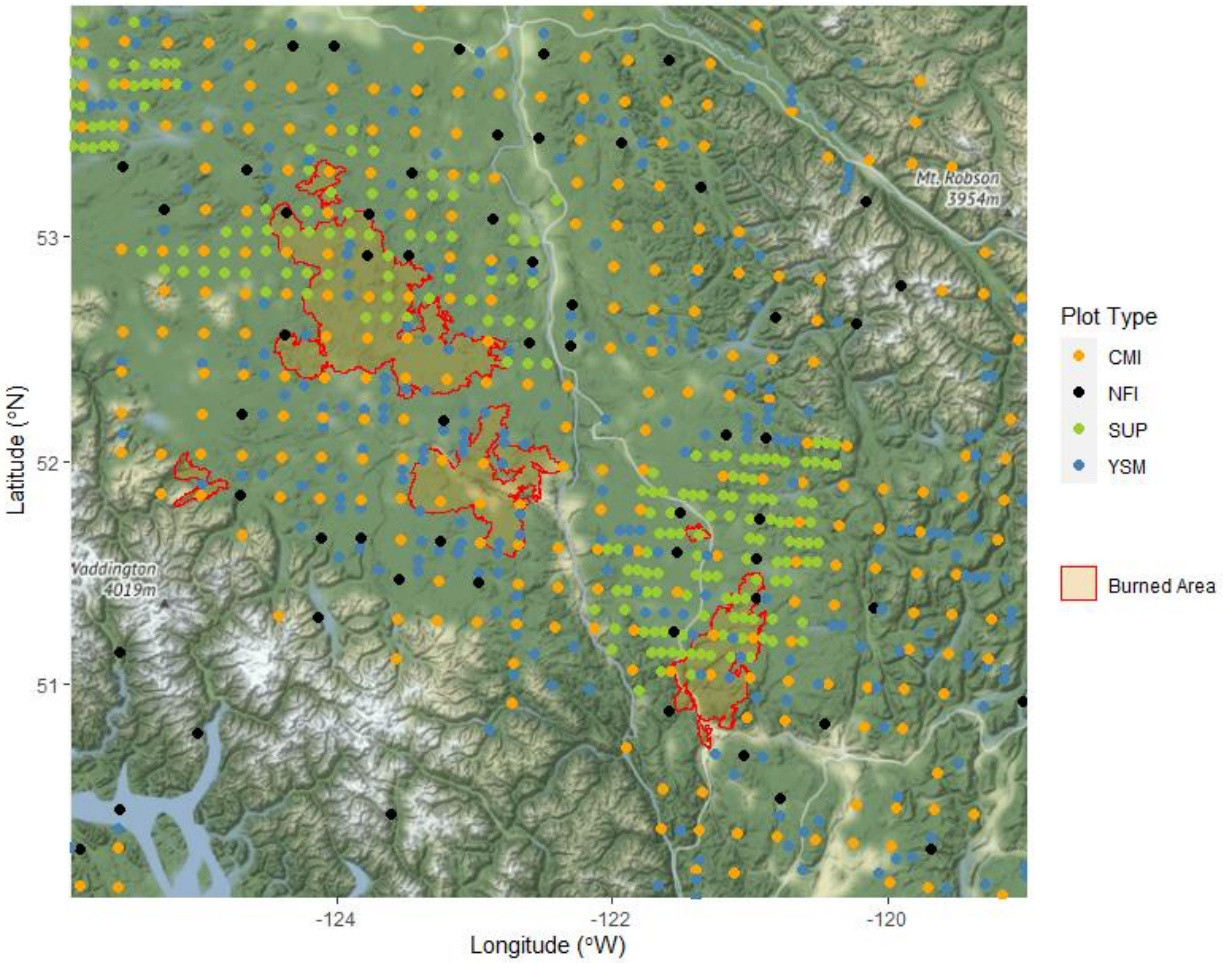


Figure 4.2 Locations of 632 Forest Analysis and Inventory Branch (FAIB) plots in the study area. The data contain measurements of 56 National Forest Inventory plots (NFI), 248 Change Monitoring Inventory plots (CMI), 109 Supplementary sample plots (SUP), and 219 Young Stand Monitoring plots (YSM) from 2000-2016.

4.3.3 Environmental covariates

4.3.3.1 Topography

A 30-m resolution Digital Elevation Model (DEM) over BC, Canada, was obtained from DataBC (Table 4.2, https://data.gov.bc.ca/) and processed by the UBC Geospatial Resources team.

Derivatives of the DEM including elevation (m), slope angle (%), and aspect (°) were calculated via ESRI ArcMap 10.8.1. Using the plot center coordinates, I extracted values for each topography variable for all the fire plots and unburned FAIB plots in the study area. Slope aspects in degrees were converted to cosine values.

4.3.3.2 Climate

I extracted eight variables of 30-year normal (1990-2020) climate data (Table 4.2) at each plot location from ClimateBC v7.00 (Wang et al. 2016). I conducted principal component analysis (PCA) to reduce the dimensionality of the climate variables (Gaveau et al. 2017). I obtained three component factors related to 1) precipitation, 2) temperature, and 3) difference in temperature of the warmest and coldest months (Table 4.3). The three principal components explained over 96% of the variance of the eight normal climate variables.

4.3.3.3 Remote-sensing biomass data

Remotely assessed biomass is closely associated with woody carbon and forest fuel loads, and the carbon and fuel loads are linked to the probability of wildfire occurrence (Fares et al. 2017) as well as the woody carbon mass (Houghton 2005). I utilized the global map of aboveground live woody biomass (Mg ha^{-1}) in the year of 2000 (Harris et al. 2021) and the Canadian forest total aboveground biomass in 2015 (Wulder et al. 2020) to estimate the pre-fire biomass at each burned-plot location. Both maps assess forest biomass based on species- and regional-specific height-biomass equations using remote-sensing data. I extracted the biomass values (Biomass2000 and Biomass2015; Table 4.2) at the plot center coordinates via ESRI ArcMap 10.1, similar to the process used for the topography variables.

Table 4.2 Descriptions of the environmental covariates.

	Variables and description	Source	Resolution
Topography	Elevation (m), slope (degree), cosine of aspect	DataBC (data.gov.bc.ca)	30m
Climate	30-year normal (1990-2020) variables	ClimateBC	800m
	MAT: mean annual temperature (°C)		
	MWMT: mean warmest month temperature (°C)		
	MCMT: mean coldest month temperature (°C)		
	TD: temperature difference between MWMT and MCMT (°C)		
	MAP: mean annual precipitation (mm)		
	MSP: mean annual summer (May to Sept.) precipitation (mm)		
	AHM: annual heat-moisture index (MAT+10)/(MAP/1000))		
	SHM: summer heat-moisture index ((MWMT)/(MSP/1000))		
Remote-sensing metric	Aboveground live woody biomass density in 2000 (Biomass2000)	Harris et al. 2021 Wulder et al. 2020	30m
	Forest total aboveground biomass in 2015 (Biomass2015)		

Table 4.3 Results of PCA for eight climate variables.

Principal components	Variables with high correlation	Contribution rate (%)	Cumulative contribution rate (%)
1. Precipitation	MAP	52.08	52.08
	MSP		
	AHM		
	SHM		
2. Temperature	MAT	35.60	87.67
	MWMT		
	MCMT		
3. Difference in temperature	TD	8.67	96.35

4.4 Methods

4.4.1 Evaluation of similarity in environmental covariates between burned and control plots

The similarity in the environmental covariates between burned and unburned plots was assessed based on both parametric and non-parametric statistics: Welch's two-sample t-test and the absolute standardized mean difference (ASMD, Rosenbaum and Rubin 1985). While the test of significance is frequently used to compare the distribution of burned and control groups, ASMD is particularly recommended to assess the goodness-of-fit for the propensity score model (Austin 2008, 2009), because of its scale independence (Ali et al. 2015). 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).

4.4.2 Distance-adjusted propensity score matching (DAPSM) with and without replacement

The distance-adjusted propensity scores (Papadogeorgou et al. 2019) were computed as a linear combination of the difference in propensity scores (PS) and the relative Euclidean distance between the pairs of treated and untreated observations, in this study burned and unburned plots, respectively. The propensity scores of each plot were estimated from a logistic regression of wildfire occurrence (i.e., 1-burned, 0-unburned) as a function of eight environmental covariates, including three PCA components of the climate variables (i.e., precipitation, temperature, and difference in temperature), three topographic variables (i.e., elevation, slope, and cosine value of aspect), and two remotely assessed biomass variables (i.e., Biomass 2000 and Biomass 2015). The relative Euclidean distance was computed for the pairs of burned and unburned plots as the standardized distance between the minimum and maximum distance of all pairs. The weight of the distance-adjusted propensity score was determined empirically by trying different weight values w from 0.9 to 0.1. Larger values of w put more emphasis on the covariates (Woo et al. 2021a, see Chapter 2; Papadogeorgou et al. 2019), which may result in failure to find close matches when the difference in propensity score distribution between burned and unburned control groups is substantial. Generally, a difference in propensity scores larger than 0.2 standard deviation of the propensity scores is considered substantial (Austin 2011), and the burned observations without close matches remain unmatched and are discarded (Rubin 2001). I selected the maximum w for which the matching was complete without any unmatched burned plots. As a result, w=0.2 was chosen, which put more emphasis on spatial proximity than on the environmental covariates.

Among the potential pool of unburned plots (i.e., raw controls), plots with the minimum distance-adjusted propensity score were selected as controls. I used one-to-one matching with a greedy algorithm (Austin 2008). Two different matching methods were employed: DAPSM with replacement and DAPSM without replacement. For DAPSM with replacement the same control can be matched to multiple burned plots, while DAPSM without replacement allows only one burned plot per control.

4.4.3 Computation of standing tree carbon

I calculated stem biomass in kilograms from wood and bark compartments of individual standing live and dead trees based on Canadian national biomass allometric equations (Lambert et al. 2005). The nonlinear equations were parameterized by Ung et al. (2008) for different species using DBH and height of trees. The computed woody biomass was converted as per hectare values for each plot through multiplying by per hectare factor (Avery and Burkhart 2015) and summarized by plots. Finally, the woody carbon mass (Mg ha^{-1}) from standing trees was obtained as half of the biomass. Because the information on decay class for dead trees were not collected in the field, I estimated the mean density reduction factor of standing dead trees by averaging the density reduction factors reported in Harmon et al. (2011) across all decay classes. As a result, a uniform multiplier of 0.95 was applied to live carbon mass to compute the dead tree carbon.

4.4.4 Process of analysis

I computed ASMD between burned and sampled unburned plots of the three fire datasets for the thirteen environmental variables including the eight ClimateBC 30 year normal climate variables (Table 4.1), three topographic variables and two remotely assessed biomass variables. I evaluated the balance for the original eight climate variables as well as the three PCA

components used for propensity score estimation, based on the idea that the goodness-of-fit diagnostics should be performed on the measured baseline characteristics of the sample (Austin 2008).

For DAPSM with and without replacement using the FAIB data, I set the selection pools for raw controls for each of the three fire datasets differently. For the Elephant Hill Fire and the Prouton Lakes Fire data, the raw controls included all 632 FAIB plots measured in 2000-2016 as well as the sampled unburned plots (Table 4.4). For the south-central BC fires data, the raw controls excluded the pre-measurements of the 34 burned plots (Table 4.4). I reported mean and variance of the estimated propensity scores for burned, sampled unburned plots, raw controls, and controls selected through matching to examine the similarity in the estimated propensity scores (i.e., probability of having wildfire) based on the environmental covariates (Baser 2006). The number of times each control was matched in DAPSM with replacement was recorded to examine potential decrease in sample size of control plots compared to DAPSM without replacement (Stuart 2010). I mapped the burned and selected control plot locations for each fire dataset to compare the spatial distribution of burned and raw controls. I also compared the ASMD values from DAPSM with and without replacement to investigate if the controls selected by matching were similar to the burned plots in terms of the observed environmental covariates. I additionally examined matching using only the FAIB plots as the raw controls for each of the three datasets (Appendix C).

The wildfire impacts on forest woody carbon were quantified for the south-central BC fires data as the mean difference in standing woody carbon mass between burned and control plots by fire severities. As the dataset contained pre- and post-fire measurements of the sampled plots, difference in the mean carbon mass between pre- and post-burn data was set as the benchmark of

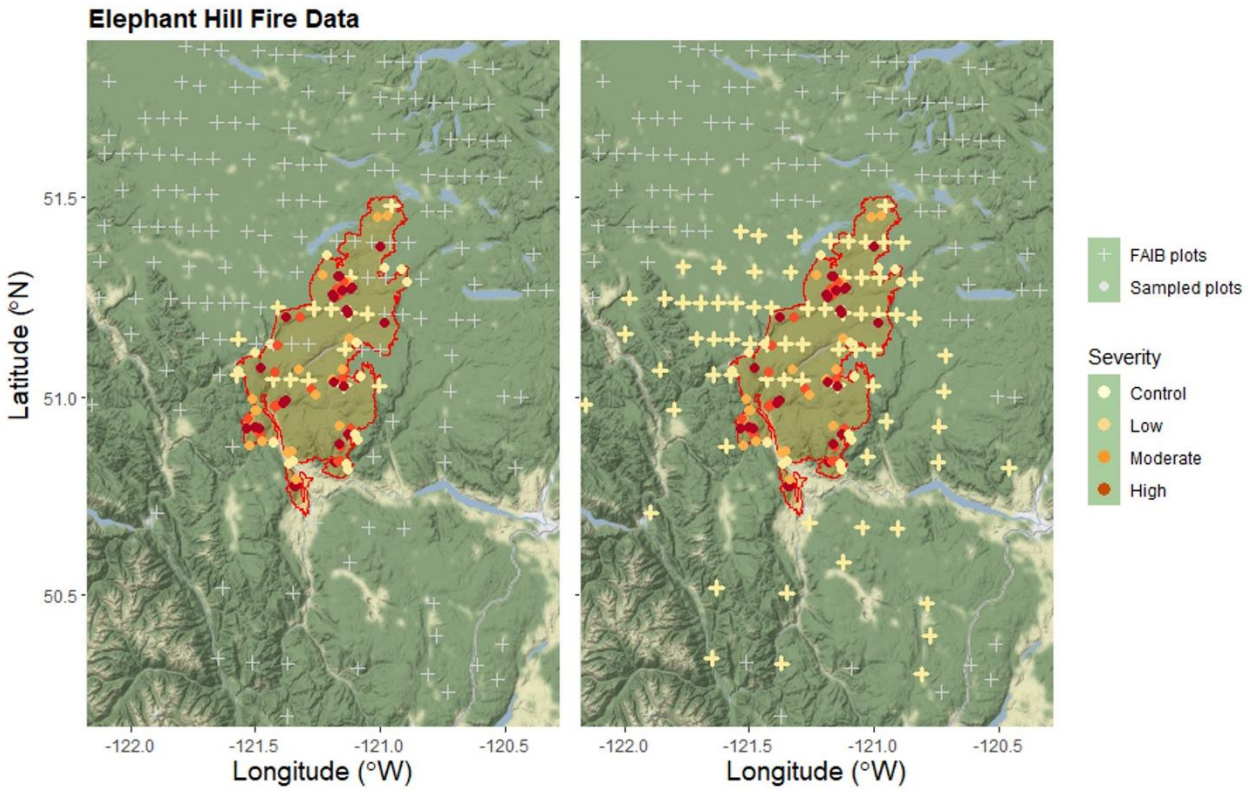
wildfire impact estimates. Therefore, the amount of mean woody carbon mass from standing trees (live and dead) was calculated for post-burn data (Post), pre-burn data (Pre), controls selected by DAPSM with replacement (MWR), and controls selected by DAPSM without replacement (MWOR). The wildfire impacts on standing tree carbon as estimated as the mean difference in standing woody carbon mass between Post and Pre, between Post and MWR, and between Post and MWOR.

Table 4.4 Numbers of sampled unburned plots, burned plots, and raw control plots used in the analysis.

Dataset	# Sampled unburned plots	# Sampled burned plots	# Raw controls
South-central BC fires	1	34	599
Elephant Hill Fire	16	89	648
Prouton Lakes Fire	7	39	639

4.5 Results

For each of the three fire datasets (i.e., Elephant Hill, Prouton Lakes, and south-central BC fires data), I obtained three sets of possible controls: 1) sampled unburned plots, 2) controls selected via DAPSM with replacement, and 3) controls selected via DAPSM without replacement. The spatial distributions of the sample plots and the selected controls were presented in Figure 4.3. The controls obtained via DAPSM were selected from the pool of raw controls which included both the sampled unburned plots and the FAIB plots (2000-2016). The results of DAPSM using different pools of raw control (i.e., the FAIB data only) were presented in Appendix C.2.


Elephant Hill Fire Data
Latitude (°N)
51.5
51.0
50.5
-122.0
-121.5
-121.0
-120.5
Longitude (°W)
FAIB plots
Sampled plots
Severity
Control
Low
Moderate
High

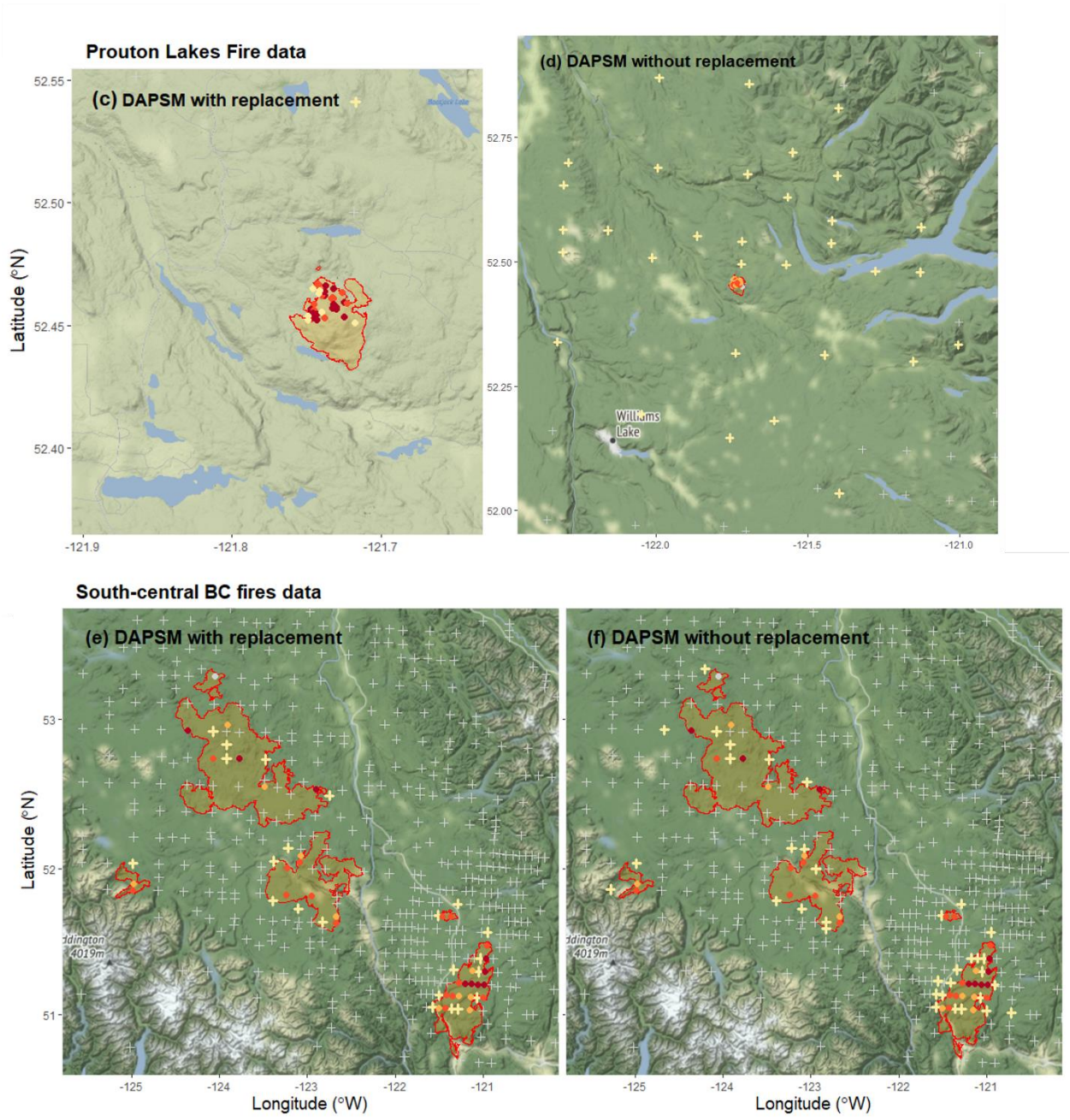


Figure 4.3 Spatial distributions of burned plots and control plots obtained from DAPSM with and without replacement for the three datasets: a) and b), Elephant Hill Fire data, c) and d), Prouton Lakes Fire data, and e) and f), south-central BC fires data. Crosses present locations of the FAIB plots, and circles present the sampled plots for each dataset. The plots selected as controls were marked as yellow.

4.5.1 Similarity between burned and sampled unburned plots in the three datasets

Based on the ASMD values of the individual environmental covariates, the climate variables of the Elephant Hill Fire data were fairly balanced between the burned plots and the sampled unburned plots with less than or close to the cut-off value of 0.2 (Figure 4.4). Among the topography covariates, the elevation of the sampled unburned plots tended to be lower than the elevation of the burned plots (Appendix C1: Table C.2.2), showing larger ASMD values (0.334) than the cut-off. However, the remote-sensing variables including Biomass2000 and Biomass2015 exhibited large imbalance for the sampled unburned plots (ASMD with 0.450 and 0.643, respectively; Figure 4.4a). On the other hand, the sampled unburned plots in the Prouton Lakes Fire data showed large imbalance for thirteen out of the fifteen covariates. The ASMD value was substantially larger for elevation (0.627) and Biomass2015 (0.815), implying that the sampled unburned plots may not be comparable to burned plots in the dataset (Figure 4.4b). The ASMD values for the south-central BC fires data were unavailable, because the data had only one sampled unburned plot (Table 4.1).

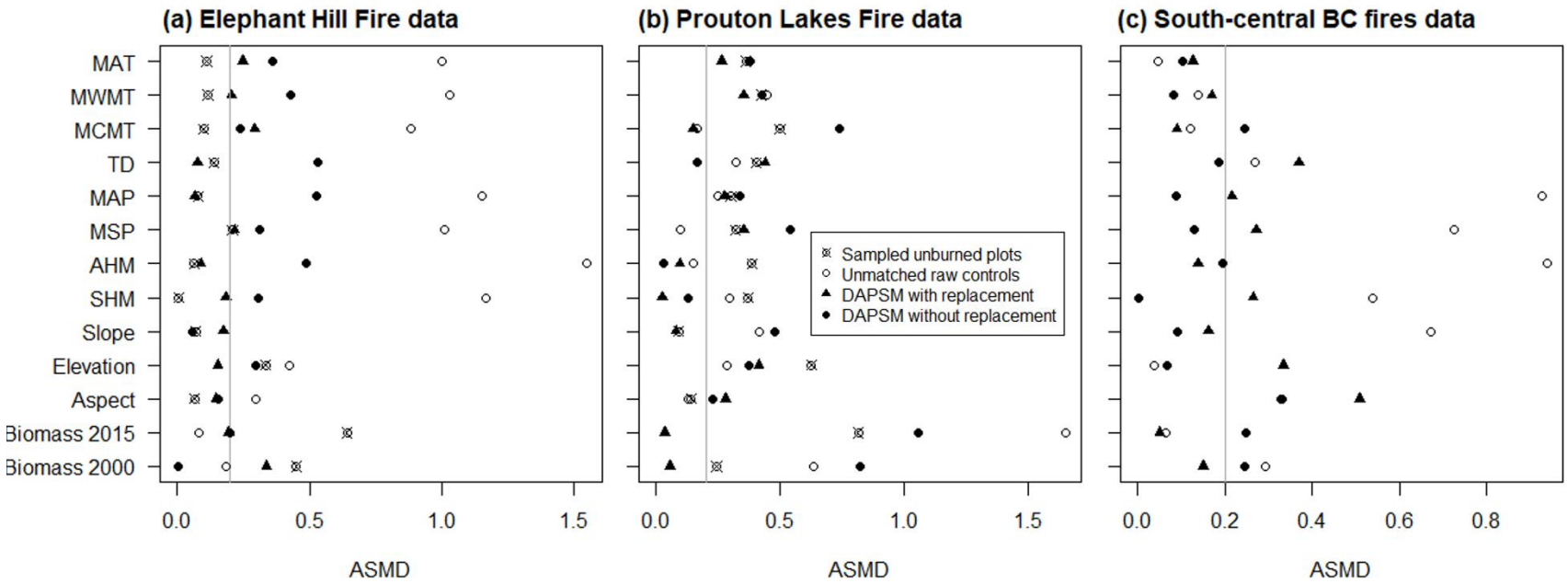


Figure 4.4 Absolute standardized mean difference (ASMD) of the fifteen environmental covariates for a) Elephant Hill Fire data, b) Prouton Lakes Fire data, and c) south-central BC fires data. Acronyms of the covariates were defined in Table 4.2. The ASMD values were computed between burned plots and the sampled unburned plots, unmatched raw controls, and control plots matched through DAPSM with and without replacement. The gray vertical line presents the cut-off value (0.2) of ASMD between balanced and imbalanced covariate.

4.5.2 Performance of DAPSM with replacement compared to DAPSM without replacement

For all three datasets, controls selected from DAPSM with replacement were more similar to the burned plots than those from DAPSM without replacement in terms of the mean PS (Table 4.5). DAPSM with replacement especially outperformed DAPSM without replacement for the Elephant Hill Fire data, resulting in more balance in the mean PS (Table 4.5). Meanwhile, there was little difference in the mean PS for the south-central BC fires data between DAPSM with and without replacement (Table 4.5). The same tendencies appeared in the matching using only the FAIB plots (Appendix C: Table C.2.1).

In terms of balancing the environmental covariates, the relative performance of DAPSM with replacement compared to without replacement differed by dataset. For the Elephant Hill Fire data, DAPSM with replacement was more effective than without replacement in reducing the ASMD below the cut-off value of 0.2 for most of the covariates except for MCMT and Biomass2000 (Figure 4.4). The mean ASMD values of the fifteen covariates were smaller under matching with replacement (0.177) than under matching without replacement (0.314) (Figure 4.4). The Prouton Lakes Fire data also showed smaller mean ASMD value from DAPSM with replacement (0.262) than from DAPSM without replacement (0.433) (Figure 4.4). For the south-central BC fires data, on the other hand, DAPSM with replacement performed worse than DAPSM without replacement, showing larger ASMD for TD, MAP, MSP, SHM, elevation, and especially aspect (Figure 4.4).

The number of the controls selected by DAPSM with replacement substantially decreased compared to DAPSM without replacement due to multiple matchings to one control. For the Elephant Hill Fire data, DAPSM with replacement selected 24 controls for 89 burned plots. Two

controls in the Elephant Hill Fire data were matched to over 10 burned plots (i.e., 12 and 13 burned plots), and five control plots were matched 5-9 times. Similarly, DAPSM with replacement for the Prouton Lakes Fire data resulted in eight controls including the seven unburned plot samples and one additional FAIB plot measurement as controls. One sampled unburned plot was matched to 10 different burned plots and three of the controls were matched to five burned plots each. Among the 24 controls in the south-central BC fires data, 16 were matched to more than one burned plot. The number of burned plots matched to a control plot ranged from one to five.

Compared with the sampled unburned plots, the controls selected from DAPSM with replacement were more similar to the burned plots in terms of the mean PS than the sampled unburned plots in both the Elephant Hill Fire and Prouton Lakes Fire data (Table 4.5). However, the difference between burned plots and controls selected from DAPSM with replacement was still statistically significant in the Prouton Lakes Fire data ($p=0.02$). For the south-central BC fires data, both DAPSM with and without replacement produced similar controls to the burned plots in terms of the mean PS (Table 4.5).

For the south-central BC fires data, DAPSM with and without replacement produced different sets of controls in their composition of plot types. Controls selected by DAPSM with replacement consisted of six SUP, five CMI, one NFI, and ten YSM plots, while DAPSM with replacement selected nineteen SUP, fourteen CMI, and two NFI. One YSM plot that has only one tree measurement was repeatedly matched to five different burned plots in DAPSM with replacement. The composition of plot types in controls selected by DAPSM without replacement was similar to that of the burned plots (i.e., 14 SUP, 13 CMI, 1 NFI and 7 YSM).

Table 4.5 The mean and standard deviation (s.d.) of propensity scores (PS) of burned plots, sampled unburned plots, raw controls, and controls matched with and without replacement for the three datasets: Elephant Hill Fire, Prouton Lakes Fire, and south-central BC fires data. n denotes the numbers of plots in each group. Statistically significant differences from the PS of burned plots were marked as bold.

	Elephant Hill Fire			Prouton Lakes Fire			South-central BC fires		
	Mean	s.d.	(n)	Mean	s.d.	(n)	Mean	s.d.	(n)
Burned plots	0.506	0.299	89	0.219	0.147	39	0.119	0.061	34
Sampled unburned plots	0.419	0.330	16	**0.115**	0.081	7	-	-	1
Raw controls	**0.079**	0.140	648	**0.058**	0.093	639	**0.050**	0.060	599
Controls – DAPSM with replacement	0.493	0.302	24	**0.144**	0.070	8	0.096	0.047	22
Controls – DAPSM without replacement	**0.278**	0.255	89	**0.088**	0.018	39	0.093	0.049	34

4.5.3 Influence of spatial intensity of burned and unburned plots on DAPSM with and without replacement

The performance of DAPSM with and without replacement to achieve balance between burned and controls differed among the three datasets with different spatial distances among the sampled plots (Figure 4.3, Table 4.6). For the Elephant Hill Fire data, for which the sampled burned and unburned plots were apart about 30 km on average, DAPSM with replacement clearly performed better than without replacement in balancing the mean PS and ASMD of the individual covariates. DAPSM with replacement selected plots that were located within or adjacent to the fire perimeter of the Elephant Hill Fire data (Figure 4.3a), producing similar mean distance between matched pairs (Table 4.6). Meanwhile, under DAPSM without replacement, some controls from FAIB plot measurements were located on the south of the fire perimeter to be matched to the burned plots on higher elevation (below the latitude 50.7 in Figure 4.3b). Because the FAIB plots were relatively sparse on the south of the fire area (Figure 4.3b), FAIB plots farther away from the burned plots were matched to burned plots (>50 km, Table 4.6), leading to larger discrepancies in the mean PS and ASMD values.

The spatial distribution of the burned plots in the Prouton Lakes Fire data was highly clustered on a fire that was small in size (Table 4.6), and there were few FAIB plots available near the fire perimeter (Figure 4.3c, 3d). DAPSM with replacement for this data set selected controls mostly from the sampled unburned plots repeatedly, with only one control from the FAIB plots (Figure 4.3c). DAPSM without replacement was bound to select controls from the FAIB plots that are far from the fire perimeter of the Prouton Lakes Fire (Table 4.6). Neither DAPSM with nor DAPSM without replacement were effective in reducing the imbalance in PS and the environmental covariates for the Prouton Lakes Fire data.

The plot locations of the south-central BC fires data were dispersed over the largest spatial area across several fire perimeters (Figure 4.3e, 4.3f). The spatial distribution of the sample plots in the south-central BC fires data was identical to the distribution of the FAIB data (approximately 130 km, Table 4.6) as they share the same data source. In this case, DAPSM with replacement performed similar to DAPSM without replacement in terms of balancing the mean PS and the environmental covariates.

Table 4.6 Mean distances (in km) among all burned plots, among all sampled unburned plots, and between matched pairs of burned and control plots selected through DAPSM with and without replacement for the three fire datasets.

	Burned plots	Sampled unburned plots	Matched pairs	
			DAPSM with replacement	DAPSM without replacement
Elephant Hill Fire	27.60	33.49	27.22	48.49
Prouton Lakes Fire	0.91	1.23	1.44	39.67
South-central BC fires	128.22	-	131.29	132.53

4.5.4 Wildfire impact on woody carbon mass from standing trees estimated for the south-central BC fires

The mean total woody carbon mass from standing trees was the largest in the plots burned with high-severity fire in the south-central BC fires data (49 Mg ha^{-1}, Figure 4.5), followed by moderate severity (32 Mg ha^{-1}) and low severity (25 Mg ha^{-1}) plots. Compared to the pre-burn data, there was large redistribution of carbon from live to dead trees after fire especially under the high severity fire. The amount of carbon reduced by the wildfires (i.e., difference in carbon mass between pre- and post-burn) was the highest under high severity (3.64 Mg ha^{-1}), followed by low severity (2.24 Mg ha^{-1}) and moderate severity (1.80 Mg ha^{-1}).

The controls selected by DAPSM without replacement for moderate and high severity groups (Figure 4.5) had a larger amount of live tree carbon mass than the pre-burn data, therefore the total carbon mass was also larger (Figure 4.5). For the low severity plots, the total woody carbon mass was less in the controls from DAPSM without replacement than in the pre-burn data. DAPSM with replacement, on the other hand, showed less carbon than the pre-burn data for the moderate and high severity groups (Figure 4.5).

Comparing the controls with the burned plots, DAPSM without replacement tended to overestimate the reduction of live carbon due to moderate and high severity wildfires (14 and 10 Mg ha^{-1} for moderate and high severity, respectively). DAPSM with replacement produced a counterintuitive result that the mean woody carbon mass increased compared to the burned plots (8 and 19 Mg ha^{-1} for moderate and high severity, respectively). The relative amount of total woody carbon mass by different fire severities showed similar trend between controls and burned plots: the largest total woody carbon in high severity plots, followed by moderate and low severity plots.

Figure 4.5 Estimated woody carbon mass (Mg ha^{-1}) from standing trees (live and dead) in plots measured post-burn (Post), in plots measured pre-burn (Pre), and controls selected by DAPSM with replacement (MWR) and DAPSM without replacement (MWOR) by different fire severity (low, moderate, and high) in the south-central BC fires data.

4.6 Discussion

4.6.1 Imbalance found in environmental covariates between sampled unburned plots and burned plots

Controls that are similar to burned plots in terms of environmental conditions allow accurate estimation of wildfire impact by reducing the influence of confounding factors that causes

inherently different probability of wildfires between burned and unburned areas (Stuart 2010, Woo et al. 2021a). Elevation and climate gradients are related to forest structure and fuel characteristics (Lentile et al. 2006, Halofsky et al. 2013), thus they influence the probability of wildfire occurrences (Nadeem et al. 2019, Westerling et al. 2006). The imbalance in elevation and climate covariates among the sampled unburned and burned plots may impose bias to any impact analysis. Both the Elephant Hill Fire data and the Prouton Lakes Fire data showed lower mean elevation of the unburned plot sample locations than that of the burned plot locations. This may imply that the fire burned at higher elevations, or the unburned plots were sampled at lower elevation, potentially for the ease of access to the site for the field crews. In the Prouton Lakes Fire data, moreover, all climate variables exhibited large discrepancies between the burned and the sampled unburned plots. The results show that without balancing the environmental covariates that influence the wildfire occurrence probability and the carbon mass of the plots, using the sampled unburned plots in the Elephant Hill Fire and the Prouton Lakes Fire data as comparison may induce bias in impact estimation.

Aside from topography variables, the burned plots were found to have larger amount of remotely assessed biomass in 2015 than the sampled unburned plots. Assuming that the remotely assessed biomass did not induce systematic bias, this may imply that the pre-burn conditions may not be similar between burned and sampled unburned plots. Impact quantification based on the sampled unburned plots may be misleading, because carbon mass is strongly associated with biomass estimation. Researchers should check the balance between burned plots and unburned plots sampled for comparison at the stage of research design to ensure unbiased impact quantification (Stuart 2010), especially when the outcome of interest for wildfire impact analysis is closely connected with the confounding factors (Cox et al. 2009).

4.6.2 Relative performance of DAPSM with replacement to without replacement was dependent on the spatial distribution of data

Propensity score matching with replacement, though less commonly implemented than matching without replacement, is generally considered to reduce bias in the mean PS when the pool of raw controls is limited (Austin and Cafri 2020). I found that DAPSM with replacement also resulted in less difference in the mean PS for the Elephant Hill Fire data. Because the mean PS presented in Table 4.5 was exclusively based on the environmental covariates, the results support that the general utility of matching with replacement is also applicable to DAPSM. However, the degree of improvement in the performance of DAPSM with replacement compared to the without-replacement approach was dependent on datasets with different spatial patterns. DAPSM with replacement was found to be effective when there were sufficient raw controls (i.e., the FAIB plots) within and around the fire perimeter (i.e., Elephant Hill Fire data). If the burned plots were distributed over a large region sharing the same spatial sampling method with the raw controls (i.e., south-central BC fires data), on the other hand, there was little difference in the performance between DAPSM with and without replacement for balancing both the PS and the environmental covariates. For spatially clustered data with few raw controls nearby (i.e., Prouton Lake Fire data), both DAPSM with and without replacement failed to produce balanced controls that were comparable to burned plots. Using forest inventory plots for DAPSM is not adequate when the locations of burned plots are clustered, as shown in the Prouton Lakes Fire data.

Implementations of analysis methods are affected by spatial scale of data (Heuvelink 1998), and changes in spatial scales often produce different results under the same method (Turner et al. 1989). The results of this study pointed out the relationship between performance of matching methods and spatial scales of wildfire data. I suggest to examine the spatial scales and patterns of

burned and raw control observations to select an appropriate matching method, because the general statistical strategy (e.g., matching with replacement to reduce bias) may not be applicable to spatial data. If there is sufficient overlap in spatial distribution of burned plots and the FAIB plot locations, as seen in the Elephant Hill Fire data, DAPSM with replacement may be useful to find comparable controls to the burned plots for the case studies of wildfires with small number of plots.

Even though matching with replacement was found to be preferable for small datasets to achieve balance between burned and control plots in environmental covariates, implementation of the method for impact quantification using spatial data needs extra attention. The main concern for the with replacement approach is the trade-off between bias and variance (Austin and Cafri 2020, Baser 2006). Because a control plot can be matched to different burned plots multiple times, the number of controls decreases, which may lead to large variance compared to the one-to-one matching without replacement (Austin 2014). The reason of repeated matchings in this study was partially due to the small weight (w) of environmental covariates for the distance-adjusted propensity score, which was necessary to prevent discarding any burned plots. Because I put more weight on the spatial proximity, controls that are closer to burned plots have been repeatedly selected. The results indicated that DAPSM with replacement introduced large dependency on several controls that were matched to over 10 different burned plots. Using those controls, the variance of the difference in the mean carbon between burned and unburned plots will be extremely blown up (Hill and Reiter 2006). Inferences on impact quantification based on matching with replacement are complicated and require approximation (Stuart 2010, Hill and Reiter 2006), which should be further investigated.

4.6.3 Impact of south-central BC fires on standing tree woody carbon mass: can DAPSM-selected controls replace pre-burn data?

Comparison of pre- and post-fire measurements in south-central BC fires data allowed direct assessment of wildfire impacts on woody carbon mass from standing trees. Such pre-burn data are not always available due to the cost of monitoring and the large spatio-temporal scale of disturbances (Greenbaum et al. 2021). Implementation of DAPSM methods may be an alternative in case the pre-burn data are unavailable, if the methods can produce similar results with the pre-burn data. The results of woody carbon mass from the controls selected by DAPSM with and without replacement did not completely align with the pre-burn data. I speculate the reason was because the distance-adjusted propensity score could not address all the potential environmental factors that influence the wildfire occurrence probability. However, the controls from DAPSM especially for DAPSM without replacement reflected the pattern in the pre-fire data in terms of the relative amount of carbon mass under different fire severities: largest carbon mass in the high severity plots, followed by carbon mass in moderate and low severity plots. The difference in the mean carbon mass among different fire severity plots suggests that fuel loads differ by the plots, potentially due to different forest types or forest development stages (Eskelson et al. 2016, Peterson et al. 2019). Assuming that the pre-fire data were not available, as in many wildfire impact analyses, carbon mass computed from sampled unburned plots would replace the pre-burn data across all severities, which cannot reflect the trend in carbon mass under different severities. On the other hand, DAPSM-selected plots were one-to-one matched to the burned plots, which allow accounting for the variation of carbon mass by fire severities. Once the potential environmental covariates are adequately considered, the DAPSM-selected controls may potentially replace pre-burn data.

In DAPSM with replacement, a control plot with little carbon mass was repeatedly matched to five different burned plots, which resulted in substantially less carbon mass compared to the pre-burn data or DAPSM without replacement. The results suggest a special precaution in matching with replacement: if the controls matched to burned plots multiple times contain extreme outcome values of interest (e.g., carbon mass), the impact estimation can be severely biased especially for the data with small sample size. In the controls selected for the south-central BC fires data, the different compositions of plot types in the controls and burned plots also likely affected the bias, because the controls from DAPSM with replacement contained many of YSM plots. YSM program was mainly developed for providing information on young stand growth (Fenger and Bradford 2013), thus the plot may contain young trees mostly. Inclusion of different plot types in propensity score estimation model may reduce the difference in carbon mass from the pre-burn data for matching. However, it was infeasible in this study due to the small number of burned plots.

Overall, this chapter presented empirical information to support future analyses for case studies of wildfire impact with small datasets: 1) post-fire sample plots may not be comparable between burned and unburned plots in their environmental conditions, thus it is important to ensure the covariate balance at the stage of research design; 2) matching with replacement can help to enhance the covariate balance for small datasets under some circumstances, for example when the spatial pattern of the raw control plots is similar to that of the burned plots; and 3) controls obtained through matching allowed pairwise comparison among different wildfire severities as the pre- and post-fire data do, provided that the potential environmental covariates are adequately considered in the process of matching.

Chapter 5: Conclusion

5.1 Overall importance and contributions

Due to the stochastic nature of wildfire occurrences as well as the large spatio-temporal scale of wildfire disturbances, it has been challenging to estimate the changes in forest woody carbon mass that can be directly attributed to wildfire with experimental settings. In this dissertation, I assessed the applicability of state-of-the-art quasi-experimental methods, with a focus on distance-adjusted propensity score matching (DAPSM), to forest data for quantifying wildfire impacts on forest woody carbon mass. I examined the performance of DAPSM to obtain a set of controls comparable to a randomized experiment using observational data from forest inventory plots. Specifically, I used Forest Inventory Analysis (FIA) data from the Pacific Northwest, United States of America, and Forest Analysis and Inventory Branch (FAIB) data from south-central British Columbia, Canada. In addition, I used remotely sensed wildfire data including burn severity and fire perimeter maps (http://www.mtbs.gov/). For regional fire effects estimation, I evaluated the similarity between burned plots and the controls selected by DAPSM in terms of their environmental covariates such as topography, climate, and vegetation indices. A sensitivity analysis on the performance of DAPSM in terms of the balance in environmental covariates between burned and control plots was conducted under varying data availability (i.e., number of burned plots, number of unburned plots, and environmental covariates). Using three case studies, I evaluated the similarity between burned plots and the controls selected by DAPSM with replacement in terms of their environmental characteristics and the spatial distributions of the data (i.e., clustered or grid-sampled). Overall, this research showcases the application of quasi-experimental methods to ecological data for making causal inferences from observational data. The dissertation results show evidence that incorporating a spatial component

in matching is important to address the influence of unobserved environmental covariates. The dissertation further provides a framework for identifying minimum sample size and covariate needs to implement DAPSM methods using forest inventory data. The framework established in this dissertation can be expanded to analyze other natural disturbances and ecological responses in general. The specific insights for implementing DAPSM using forest inventory data are listed below, followed by their contributions to current knowledge:

- DAPSM was superior to PSM in balancing the observed environmental covariates and in dealing with potential confounding covariates through spatial adjustment. The traditional PSM and spatial matching, which use only the covariates or the spatial distances in matching, respectively, did not produce a set of controls that were comparable to burned plots. DAPSM allowed better quantification of the causal effects of wildfires on forest carbon in the forests of Washington and Oregon, USA (Chapter 2).

- Climate variables were the key covariates to estimate the propensity scores for matching burned and control plots. However, inclusion of spatial distance in the matching process compensated for the omission of key covariates if the number of available sample plots was sufficiently large (e.g., ≥ 100 burned plots) (Chapter 3).

Both of these results show that the spatial components of ecological data are important to find a set of controls that are comparable to treated units in terms of environmental covariates to establish a quasi-experimental setting. Thus, the spatial distances among the observations need to be accounted for when employing propensity score matching methods. The importance of spatial components in propensity score matching has been gaining attention from disciplines where quasi-experimental methods are already prevalent (e.g., Davis et al. 2019, Papadogeorgou et al.

2019, Chagas et al. 2012), but has not been thoroughly discussed through applied examples using ecological data. The findings in my dissertation provide the empirical evidence for incorporating a spatial adjustment in matching to ecologists who analyze spatially indexed data and want to implement quasi-experimental methods for ecological problems for which experimental settings are infeasible.

- When the available forest inventory plots were spatially sparse, inclusion of spatial distance in propensity score estimation was not effective for increasing balance in environmental covariates. The suggested threshold of sample size for implementing DAPSM on the national forest inventory data in Washington and Oregon is ≥ 100 disturbance-affected plots and ≥ 1,000 disturbance-unaffected plots (Chapter 3).

The data requirements established in this dissertation contribute to the design of future wildfire impact analyses using forest inventory data. Utilization of forest inventory data will provide accurate estimates of forest resources especially for forest carbon from tree biomass (Liski et al. 2006).

- DAPSM with replacement using the forest inventory data from south-central British Columbia, Canada, could improve the performance of DAPSM reducing the imbalance in environmental covariates between burned and control plots for a particular case-study of wildfire impacts with small number of post-fire plot data. The spatial distributions of forest inventory plots and post-fire sampled burned plots influenced the performance of DAPSM with replacement relative to that of DAPSM without replacement. When the locations of burned plots were clustered within a small range of spatial extent, DAPSM

with replacement based on grid-based forest inventory plots resulted in large dependency on a few numbers of controls near the fire perimeter (Chapter 4).

- The controls matched one-to-one to burned plots through DAPSM could reflect the trend in woody carbon mass from standing trees by different fire severities. The result implies that DAPSM-selected controls accounted for the variability of carbon mass between different burn severities. With an adequate specification of the environmental covariates, matching can produce better controls than the sampled unburned plots where pre-burn data are unavailable (Chapter 4).

The results provide empirical supports to the claim that the choice of with- and without-replacement approach in spatial data should be based on realistic ecological conditions (Guillera-Arroita 2011). As the first research examining the performance of DAPSM with replacement, the above insights contribute to enhance current understanding on the use of quasi-experimental methods and their extension for wildfire impact analyses. In addition, the woody carbon mass estimation using the matched data and pre-burn data help to identify practical problems and suggestions for applying with-replacement approaches to ecological data, which are spatially distributed and often contain small number of samples.

5.2 Cautions about implementing propensity score matching

Quasi-experimental methods, including propensity score matching, are procedures for designing non-experimental studies to mitigate the selection bias that leads to different probability of treatment assignment (Stuart 2010). However, the mitigation is not guaranteed if the assumptions for the propensity scores are not met. The most important and strongest assumption is ‘no unobserved confounding variables’ (Rosenbaum and Rubin 1983), which was discussed in

Chapter 2 (2.5.2) of this dissertation. I suggested that the inclusion of a spatial component in the propensity score estimation model may ease the violation of the assumption to some degree. Undeniably, the spatial component does not address all unobserved confounding factors. It is recommended to include as many potential environmental covariates as possible in the propensity score estimation model (McCaffrey et al. 2004), given that the number of observations is large enough.

The other important assumption in propensity score methods is called 'common supports' assumption (Stuart 2011, Heckman et al. 1999), which means that the ranges of estimated propensity scores should overlap between the treated and control group. Finding close matches is impossible if the difference in the propensity scores between treated and untreated groups is substantial, resulting in bias in the impact quantification even with a well-defined propensity score model. Research in other fields has shown that large amounts of data can resolve the lack of common supports problem (Lechner 2001). Other remedies dealing with the violation of the assumption have been examined, for example trimming the observations with extreme values of propensity scores on a systematic basis (Crump et al. 2009). In Chapter 3, I proposed a marginal number of wildfire data to meet the common supports assumption. It is unclear, however, if other potential remedies (e.g., trimming) will be applicable to ecological data, given the analysis of Chapter 4 warranted that general strategies of reducing bias through matching with replacement did not work well for the spatially located wildfire data. Further studies should be conducted based on empirical and ecological data to understand the impact of assumption violation in implementing propensity score methods.

Above all, it should be noted that quasi-experimental methods cannot completely substitute well-established randomized experiments. Experiments with random assignment of treatment do not

require the above assumptions, thus always are preferred whenever available (Yue 2007). Studies that inevitably rely on nonrandomized conditions due to practical or ethical reasons can consider propensity score methods to analyze their data, as they consider traditional covariate adjustment methods. It is difficult to say which method produces more reliable inferences—quasi-experimental methods or traditional covariate adjustment methods, especially non-parametric analyses such as random forest and artificial intelligence imputation based on big data—because the scope of inferences differs between the two (Greenstone and Gayer 2009). A clear advantage of quasi-experimental methods over traditional covariate adjustment methods is that they can provide stronger inferences of causal effects from observational data, as stated in the Introduction of this dissertation.

5.3 Future research directions

Ecological research, including analyses on wildfire effects, aims to quantify causal impacts from observational data (Butsic et al. 2017, Larsen et al. 2019). Quasi-experimental methods have been developed and widely implemented in medical sciences over the past decades across disciplines that have similar objectives (Stuart 2010), yet the application of these methods to empirical ecological data have not been fully examined. Here I demonstrated spatially-adjusted propensity score matching methods using forest inventory data and plot-based fire data. There are more opportunities for future research to come with regards to propensity score matching and other quasi-experimental methods for impact quantification of many ecological processes. As one of the first studies applying quasi-experimental methods to empirical forest inventory data, challenges and opportunities in application have become apparent. The potential research directions and suggestions that I have identified are described in the following.

5.3.1 Methodology may be applicable to quantifying effects of other disturbances

The methodology of spatially adjusted propensity score matching and the framework for data requirements and covariate sets can be expanded to different natural disturbances in forests other than wildfires (e.g., disease, insects, wind, and drought), as well as various responses to the disturbances (e.g., effects on vegetation structure, wildlife, and impacts on other disturbances). The possible examples are also listed in Butsic et al. (2017).

5.3.2 Exploration of extensions of propensity score adjustment methods

A variety of extensions of propensity score adjustment methods besides matching are still to be explored for wildfires and other ecological processes with different data structures and distributions, especially with spatial components. Examples may include:

- Sub-classification for non-binary treatments (Imai and Van Dyk 2004), e.g., different severities of insects or disease infestation;

- Propensity score weighting for multilevel or clustered data (Li et al. 2013), e.g., impact of slash burn on a specific wildfire-affected area; and

- Covariate adjustment with propensity scores for estimating average treatment effect (ATE, Grilli and Rampichini 2011), e.g., impact of wildfires on smoke aerosols over both burned and unburned area.

Each of these methods needs to be validated by sensitivity analyses to gauge the associated bias under different circumstances of data and propensity score model availability, as shown in Chapter 3 of this dissertation.

5.3.3 Improvement of propensity score estimation

Regarding the suggested research directions outlined in 5.2.1 and 5.2.2 above, environmental covariates as well as site-specific factors that are related to both ecological processes and the outcome of the processes should be identified for estimation of propensity scores (Patorno et al. 2014), as done in Chapter 3 of this dissertation. The identification process includes types of the environmental covariates, orders of the covariates to be included in the propensity score estimation models, and interactions among the covariates (McCaffrey et al. 2004). In this dissertation, selection of covariates for the propensity score estimation models mostly relied on literature investigating the environmental covariates related to wildfire and forest aboveground woody carbon (e.g., Cardille et al. 2001, Ager et al. 2012, Parisien and Moritz 2009). I included topography, climate, and remotely sensed pre-fire biomass metrics that are generally considered to be strongly associated with both wildfire occurrence probability and woody carbon mass in the forests, which was necessary to analyze the multiple wildfires occurred over the Pacific Northwest region. However, for the small and local wildfires, for instance the Elephant Hill and Prouton Lake fires analyzed in Chapter 4, inclusion of BEC subzone, site series classification, or species composition that are available in the forest inventory data may increase the validity of the propensity score estimation model. Addition of such multi-categorical environmental covariates to the propensity score model is currently difficult due to the lack of measurements and small number of samples, hence a fire-specific propensity score model exploring various covariates and model types is required.

Throughout this research, the propensity score models were derived using logistic regression models to estimate the probability of wildfire occurrences in burned and unburned plots. Even though the logistic regression models are one of the most commonly used models for propensity

score matching methods (Baser 2006), a variety of parametric or non-parametric models can be employed for propensity score estimation. For example, ordinal logistic regression can be used when there are multiple fires in plots (Zanutto et al. 2005), and classification and regression trees are robust when the data are incomplete with missing values (Lee et al. 2010).

5.3.4 Downscaling of the temporal resolution of climate metrics

I derived 30-year normal climate metrics from PRISM and ClimateBC interpolation maps to represent the climate covariates in the USA and Canada, respectively. PRISM and ClimateBC models are frequently used to depict spatial climate patterns over a large area with reasonable accuracy. However, the interpolated metrics and long-term 30-year averages may not reflect the local variation of climate covariates at each plot location (Wang et al. 2016), especially for the case-studies for which plot locations are clustered within a small burned area. Downscaling the temporal resolution of climate metrics would likely help capture the local variability of fire climate (Abatzoglou and Brown 2012).

5.3.5 Implement modifications to existing inventory design

Advances in national forest inventory survey design in combination with remote sensing data are essential to support forest-related research by providing reliable and persistent observational data (Tomppo et al. 2008). Lindenmayer and Likens (2009) proposed an adaptive monitoring paradigm, which allows inventory and monitoring programs to evolve with new questions, new methods, and new technology. Thus, research for monitoring and measurement techniques need to be continuously carried-out, which in turn will increase the utility of forest inventory data for assessment of disturbance impacts using quasi-experimental methods.

Increment in spatial and temporal sampling intensity of the national forest inventory data will support further development of impact analyses for ecological processes using forest monitoring data. The examples of such intensifications include the fire effects and recovery study (FERS) of the Forest Inventory Analysis (FIA) program in USA (Chapter 3), and supplementary sample plots (SUP) of the Forest Analysis and Inventory Branch (FAIB) in Canada (Chapter 4). Improvement of forest inventory data can help in understanding complicated carbon dynamics in forests (Hararuk et al. 2017). The wildfire impact analyses using forest inventory data can be enhanced by detailed measurements across fuel and vegetation attributes in terms of various forest carbon pools (e.g., crowns, fine woody debris, soil, litter and duff). Modifications to existing inventory and monitoring networks in terms of spatial or temporal intensifications or additional measurements align with the adaptive monitoring paradigm described by Lindenmayer and Likens (2009). These kinds of modifications may be needed to answer new research questions and to address new research needs. New research questions may evolve around linkages between forest carbon and natural disturbances, which are projected to increase in frequency and severity from future climate changes (Spracklen et al. 2009).

Bibliography

Abadie, A., Imbens, G.W., 2006. Large sample properties of matching estimators for average treatment effects. Econometrica 74, 235–267.

Abatzoglou, J.T., Brown, T.J., 2012. A comparison of statistical downscaling methods suited for wildfire applications. Int. J. Climatol. 32, 772–780.

Adelson, J.L., 2013. Educational research with real-world data: Reducing selection bias with propensity score analysis. Pract. Assessment, Res. Eval. 18, 15.

Ager, A.A., Finney, M.A., McMahan, A., Carthcart, J., 2010. Measuring the effect of fuel treatments on forest carbon using landscape risk analysis. Nat. Hazards Earth Syst. Sci. 10, 1–12.

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

Ager, A.A., Vaillant, N.M., Finney, M.A., Preisler, H.K., 2012. Analyzing wildfire exposure and source–sink relationships on a fire prone forest landscape. For. Ecol. Manage. 267, 271–283.

Alexander, M.E., Thomas, D.A., 2003. Wildland fire behavior case studies and analyses: Other examples, methods, reporting standards, and some practical advice. Fire Manag. Today 63, 4–12.

Ali, M.S., Groenwold, R.H.H., Belitser, S. V, Pestman, W.R., Hoes, A.W., Roes, K.C.B., de Boer, A., Klungel, O.H., 2015. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. J. Clin. Epidemiol. 68, 122–131.

Ali, M.S., Prieto-Alhambra, D., Lopes, L.C., Ramos, D., Bispo, N., Ichihara, M.Y., Pescarini, J.M., Williamson, E., Fiaccone, R.L., Barreto, M.L., 2019. Propensity score methods in health technology assessment: principles, extended applications, and recent advances. Front. Pharmacol. 10, 973.

Amiro, B.D., Todd, J.B., Wotton, B.M., Logan, K.A., Flannigan, M.D., Stocks, B.J., Mason, J.A., Martell, D.L., Hirsch, K.G., 2001. Direct carbon emissions from Canadian forest fires, 1959-1999. Can. J. For. Res. 31, 512–525.

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., 2008. Goodness-of-fit diagnostics for the propensity score model when estimating treatment effects using covariate adjustment with the propensity score. Pharmacoepidemiol. Drug Saf. 17, 1202–1217.

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., 2014. A comparison of 12 algorithms for matching on the propensity score. Stat. Med. 33, 1057–1069.

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

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.

Barrett, K., Kasischke, E.S., McGuire, A.D., Turetsky, M.R., Kane, E.S., 2010. Modeling fire severity in black spruce stands in the Alaskan boreal forest using spectral and non-spectral geospatial data. Remote Sens. Environ. 114, 1494–1503.

Barrett, T.M., 2006. Optimizing efficiency of height modeling for extensive forest inventories. Can. J. For. Res. 36, 2259–2269.

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

Baser, O., 2006. Too much ado about propensity score models? Comparing methods of propensity score matching. Value Heal. 9, 377–385.

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 p. 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.

Berk, R.A., 1983. An introduction to sample selection bias in sociological data. Am. Sociol. Rev. 386–398.

Blackard, J.A., Finco, M. V, Helmer, E.H., Holden, G.R., Hoppus, M.L., Jacobs, D.M., Lister, A.J., Moisen, G.G., Nelson, M.D., Riemann, R., 2008. Mapping US forest biomass using nationwide forest inventory data and moderate resolution information. Remote Sens. Environ. 112, 1658–1677.

Bladon, K.D., Silins, U., Wagner, M.J., Stone, M., Emelko, M.B., Mendoza, C.A., Devito, K.J., Boon, S., 2008. Wildfire impacts on nitrogen concentration and production from headwater streams in southern Alberta's Rocky Mountains. Can. J. For. Res. 38, 2359–2371.

Bonan, G.B., 2008. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. Science 320, 1444–1449.

Borcard, D., Legendre, P., Avois-Jacquet, C., Tuomisto, H., 2004. Dissecting the spatial structure of ecological data at multiple scales. Ecology 85, 1826–1832.

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

Brookes, W., Daniels, L.D., Copes-Gerbitz, K., Baron, J.N., Carroll, A.L., 2021. A disrupted historical fire regime in central British Columbia. Front. Ecol. Evol. 9, 420.

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.

Brown, S., 2002. Measuring carbon in forests: current status and future challenges. Environ. Pollut. 116, 363–372.

Bungartz, H.-J., Kranzlmüller, D., Weinberg, V., Weismüller, J., Wohlgemuth, V., 2018. Advances and New Trends in Environmental Informatics. Springer.

Butry, D.T., 2009. Fighting fire with fire: estimating the efficacy of wildfire mitigation programs using propensity scores. Environ. Ecol. Stat. 16, 291–319.

Butsic, V., Lewis, D.J., Radeloff, V.C., Baumann, M., Kuemmerle, T., 2017. 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., 2017. The effect of protected areas on forest disturbance in the Carpathian Mountains 1985–2010. Conserv. Biol. 31, 570–580.

Caliendo, M., Kopeinig, S., 2008. Some practical guidance for the implementation of propensity score matching. J. Econ. Surv. 22, 31–72.

Campbell, D.T., Stanley, J.C., 2015. Experimental and quasi-experimental designs for research. Ravenio Books.

Campbell, J., Donato, D., Azuma, D., Law, B., 2007. Pyrogenic carbon emission from a large wildfire in Oregon, United States. J. Geophys. Res. Biogeosciences 112, 1–11.

Campbell, S., Waddell, K., Gray, A., 2010. Washington's forest resources, 2002–2006: Five-year forest inventory and analysis report. Gen. Tech. Rep. PNW-GTR-800. Portland, OR US Dep. Agric. For. Serv. Pacific Northwest Res. Station. 189 p. 800.

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

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, 7.

Certini, G., Nocentini, C., Knicker, H., Arfaioli, P., Rumpel, C., 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., Toneto, R., Azzoni, C.R., 2012. A spatial propensity score matching evaluation of the social impacts of sugarcane growing on municipalities in Brazil. Int. Reg. Sci. Rev. 35, 48–69.

Chen, X., Wei, X., Scherer, R., 2005. Influence of wildfire and harvest on biomass, carbon pool, and decomposition of large woody debris in forested streams of southern interior British Columbia. For. Ecol. Manage. 208, 101–114.

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

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.

Cohen, J., 2013. Statistical power analysis for the behavioral sciences. Academic press.

Coogan, S.C.P., Robinne, F.-N., Jain, P., Flannigan, M.D., 2019. Scientists' warning on wildfire—a Canadian perspective. Can. J. For. Res. 49, 1015–1023.

Cook, T.D., Campbell, D.T., 1986. The causal assumptions of quasi-experimental practice: The origins of quasi-experimental practice. Synthese 141–180.

Cox, E., Martin, B.C., Van Staa, T., Garbe, E., Siebert, U., Johnson, M.L., 2009. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report—Part II. Value Heal. 12, 1053–1061.

Crist, E.P., Cicone, R.C., 1984. A physically-based transformation of Thematic Mapper data---The TM Tasseled Cap. IEEE Trans. Geosci. Remote Sens. 256–263.

Crowley, M.A., Cardille, J.A., White, J.C., Wulder, M.A., 2019. Generating intra-year metrics of wildfire progression using multiple open-access satellite data streams. Remote Sens. Environ. 232, 111295.

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

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.

Davis, M.L., Neelon, B., Nietert, P.J., Burgette, L.F., Hunt, K.J., Lawson, A.B., Egede, L.E., 2019. Analysis of racial differences in hospital stays in the presence of geographic confounding. Spat. Spatio-temporal Epidemiol. 30, 100284.

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.

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

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.

Duguy, B., Paula, S., Pausas, J.G., Alloza, J.A., Gimeno, T., Vallejo, R. V, 2013. Effects of climate and extreme events on wildfire regime and their ecological impacts, in: Regional Assessment of Climate Change in the Mediterranean. Springer, pp. 101–134.

Eggleston, H.S., Buendia, L., Miwa, K., Ngara, T., Tanabe, K., 2006. 2006 IPCC guidelines for national greenhouse gas inventories.

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

Eskelson, B.N.I., Monleon, V.J., 2018. Post-fire surface fuel dynamics in California forests across three burn severity classes. Int. J. Wildl. Fire 27, 114–124.

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.

Everitt, B.S., 2019. The analysis of contingency tables. Chapman and Hall/CRC.

Fares, S., Bajocco, S., Salvati, L., Camarretta, N., Dupuy, J.-L., Xanthopoulos, G., Guijarro, M., Madrigal, J., Hernando, C., Corona, P., 2017. Characterizing potential wildland fire fuel in live vegetation in the Mediterranean region. Ann. For. Sci. 74, 1.

Fenger, M., Bradford, P., 2013. Natural Resource Monitoring in British Columbia: A Compilation of Provincial Government Initiatives. Ministry of Forests, Lands and Natural Resource Operations, Resource Practices Br., Victoria BC FREP Report 33.

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

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

Francos, M., Úbeda, X., Pereira, P., Alcañiz, M., 2018. Long-term impact of wildfire on soils exposed to different fire severities. A case study in Cadiretes Massif (NE Iberian Peninsula). Sci. Total Environ. 615, 664–671.

Fraver, S., Ringvall, A., Jonsson, B.G., 2007. Refining volume estimates of down woody debris. Can. J. For. Res. 37, 627–633.

French, N.H.F., de Groot, W.J., Jenkins, L.K., Rogers, B.M., Alvarado, E., Amiro, B., De Jong, B., Goetz, S., Hoy, E., Hyer, E., 2011. Model comparisons for estimating carbon emissions from North American wildland fire. J. Geophys. Res. Biogeosciences 116.

Gaveau, D.L.A., Epting, J., Lyne, O., Linkie, M., Kumara, I., Kanninen, M., Leader-Williams, N., 2009. Evaluating whether protected areas reduce tropical deforestation in Sumatra. J. Biogeogr. 36, 2165–2175.

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

Ghebrehiwot, H.M., Kulkarni, M.G., Light, M.E., Kirkman, K.P., Van Staden, J., 2011. Germination activity of smoke residues in soils following a fire. South African J. Bot. 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.

Gill, A.M., Stephens, S.L., Cary, G.J., 2013. The worldwide "wildfire" problem. Ecol. Appl. 23, 438–454.

Gillis, M.D., Omule, A.Y., Brierley, T., 2005. Monitoring Canada's forests: the national forest inventory. For. Chron. 81, 214–221.

Girma, S., Görg, H., 2007. Evaluating the foreign ownership wage premium using a difference-in-differences matching approach. J. Int. Econ. 72, 97–112.

Goetz, S.J., Bond-Lamberty, B., Law, B.E., Hicke, J.A., Huang, C., Houghton, R.A., McNulty, S., O'Halloran, T., Harmon, M., Meddens, A.J.H., 2012. Observations and assessment of forest carbon dynamics following disturbance in North America. J. Geophys. Res. Biogeosciences 117.

Gonzales, R., Aranda, P., Mendizabal, J., 2017. A Bayesian spatial propensity score matching evaluation of the regional impact of micro-finance. Rev. Econ. Anal. 9, 127-153.

Goulden, M.L., Winston, G.C., McMILLAN, A.M.S., Litvak, M.E., Read, E.L., Rocha, A. V, Rob Elliot, J., 2006. An eddy covariance mesonet to measure the effect of forest age on land–atmosphere exchange. Glob. Chang. Biol. 12, 2146–2162.

Grace, J.B., Irvine, K.M., 2020. Scientist's guide to developing explanatory statistical models using causal analysis principles. Ecology 101, e02962.

Greenbaum, N., Wittenberg, L., Malkinson, D., Inbar, M., 2021. Hydrological and sedimentological changes following the 2010-forest fire in the Nahal Oren Basin, Mt. Carmel Israel–a comparison to pre-fire natural rates. Catena 196, 104891.

Greenstone, M., Gayer, T., 2009. Quasi-experimental and experimental approaches to environmental economics. J. Environ. Econ. Manage. 57, 21–44.

Grilli, L., Rampichini, C., 2011. Propensity scores for the estimation of average treatment effects in observational studies. Train. Sess. Causal Inference, Bristol 28–29.

Gronowska, M., Joshi, S., MacLean, H.L., 2009. A review of US and Canadian biomass supply studies. BioResources 4, 341–369.

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

Guillera-Arroita, G., 2011. Impact of sampling with replacement in occupancy studies with spatial replication. Methods Ecol. Evol. 2, 401–406.

Guo, S., Barth, R.P., Gibbons, C., 2006. Propensity score matching strategies for evaluating substance abuse services for child welfare clients. Child. Youth Serv. Rev. 28, 357–383.

Halofsky, J.E., Hemstrom, M.A., Conklin, D.R., Halofsky, J.S., Kerns, B.K., Bachelet, D., 2013. Assessing potential climate change effects on vegetation using a linked model approach. Ecol. Modell. 266, 131–143.

Halofsky, J.E., Peterson, D.L., Harvey, B.J., 2020. Changing wildfire, changing forests: the effects of climate change on fire regimes and vegetation in the Pacific Northwest, USA. Fire Ecol. 16, 1–26.

Hanes, C.C., Wang, X., Jain, P., Parisien, M.-A., Little, J.M., Flannigan, M.D., 2019. Fire-regime changes in Canada over the last half century. Can. J. For. Res. 49, 256–269.

Hararuk, O., Shaw, C., Kurz, W.A., 2017. Constraining the organic matter decay parameters in the CBM-CFS3 using Canadian National Forest Inventory data and a Bayesian inversion technique. Ecol. Modell. 364, 1–12.

Harmon, M.E., Woodall, C.W., Fasth, B., Sexton, J., 2008. Woody detritus density and density reduction factors for tree species in the United States: a synthesis. Gen. Tech. Rep. NRS-29. Newt. Square, PA US Dep. Agric. For. Serv. North. Res. Station. 84 p. 29.

Harmon, M.E., Woodall, C.W., Fasth, B., Sexton, J., Yatkov, M., 2011. Differences between standing and downed dead tree wood density reduction factors: a comparison across decay classes and tree species. Res. Pap. NRS-15. Newt. Square, PA US Dep. Agric. For. Serv. North. Res. Station. 40 p. 15, 1–40.

Harris, N.L., Gibbs, D.A., Baccini, A., Birdsey, R.A., De Bruin, S., Farina, M., Fatoyinbo, L., Hansen, M.C., Herold, M., Houghton, R.A., 2021. Global maps of twenty-first century forest carbon fluxes. Nat. Clim. Chang. 11, 234–240.

Haughian, S.R., Burton, P.J., Taylor, S.W., Curry, C., 2012. Expected effects of climate change on forest disturbance regimes in British Columbia. J. Ecosyst. Manag. Vol 13, No 1 13.

Heckman, J.J., 2014. Sample selection bias as a specification error with an application to the estimation of labor supply functions. Princeton University Press.

Heckman, J.J., Ichimura, H., Smith, J., Todd, P., 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. Proc. Natl. Acad. Sci. 93, 13416–13420.

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.

Heckman, J., Ichimura, H., Smith, J., Todd, P., 1998. Characterizing selection bias using experimental data. National bureau of economic research.

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

Heinrich, C., Maffioli, A., Vazquez, G., 2010. A primer for applying propensity-score matching. Inter-American Dev. Bank.

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

Hessburg, P.F., Miller, C.L., Povak, N.A., Taylor, A.H., Higuera, P.E., Prichard, S.J., North, M.P., Collins, B.M., Hurteau, M.D., Larson, A.J., 2019. Climate, environment, and disturbance history govern resilience of western North American forests. Front. Ecol. Evol. 7, 239.

Hessburg, P.F., Prichard, S.J., Hagmann, R.K., Povak, N.A., Lake, F.K., 2021. Wildfire and climate change adaptation of western North American forests: a case for intentional management. Ecol. Appl. e02432.

Heuvelink, G.B.M., 1998. Uncertainty analysis in environmental modelling under a change of spatial scale, in: Soil and Water Quality at Different Scales. Springer, pp. 255–264.

Hill, J., Reiter, J.P., 2006. Interval estimation for treatment effects using propensity score matching. Stat. Med. 25, 2230–2256.

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

Houghton, R.A., 2005. Aboveground forest biomass and the global carbon balance. Glob. Chang. Biol. 11, 945–958.

Houghton, R. A., Hall, F., Goetz, S.J., 2009. Importance of biomass in the global carbon cycle. J. Geophys. Res. Biogeosci. 114, G00E03.

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

Hudiburg, T.W., Law, B.E., Wirth, C., Luyssaert, S., 2011. Regional carbon dioxide implications of forest bioenergy production. Nat. Clim. Chang. 1, 419–423.

Hurteau, M.D., North, M., 2010. Carbon recovery rates following different wildfire risk mitigation treatments. For. Ecol. Manage. 260, 930–937.

Imai, K., Van Dyk, D.A., 2004. Causal inference with general treatment regimes: Generalizing the propensity score. J. Am. Stat. Assoc. 99, 854–866.

Isaev, A.S., Korovin, G.N., Bartalev, S.A., Ershov, D. V, Janetos, A., Kasischke, E.S., Shugart, H.H., French, N.H.F., Orlick, B.E., Murphy, T.L., 2002. Using remote sensing to assess Russian forest fire carbon emissions. Clim. Change 55, 235–249.

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 p.

Jalkanen, A., Mattila, U., 2000. Logistic regression models for wind and snow damage in northern Finland based on the National Forest Inventory data. For. Ecol. Manage. 135, 315–330.

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

Kashian, D.M., Romme, W.H., Tinker, D.B., Turner, M.G., Ryan, M.G., 2006. Carbon storage on landscapes with stand-replacing fires. Bioscience 56, 598.

Kasischke, E.S., Bruhwiler, L.P., 2002. Emissions of carbon dioxide, carbon monoxide, and methane from boreal forest fires in 1998. J. Geophys. Res. Atmos. 107, FFR-2.

Keele, L., Titiunik, R., Zubizarreta, J.R., 2015. Enhancing a geographic regression discontinuity design through matching to estimate the effect of ballot initiatives on voter turnout. J. R. Stat. Soc. Ser. A 178, 223–239.

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.

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.

Kurz, W.A., Stinson, G., Rampley, G.J., Dymond, C.C., Neilson, E.T., 2008. Risk of natural disturbances makes future contribution of Canada's forests to the global carbon cycle highly uncertain. Proc. Natl. Acad. Sci. 105, 1551–1555.

Lambert, M.C., Ung, C.H., Raulier, F., 2005. Canadian national tree aboveground biomass equations. Can. J. For. Res. 35, 1996–2018.

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.

Law, B.E., Waring, R.H., 2015. Carbon implications of current and future effects of drought, fire and management on Pacific Northwest forests. For. Ecol. Manage. 355, 4–14.

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

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

Lee, B.K., Lessler, J., Stuart, E.A., 2010. Improving propensity score weighting using machine learning. Stat. Med. 29, 337–346.

Lee, J., Little, T.D., 2017. A practical guide to propensity score analysis for applied clinical research. Behav. Res. Ther. 98, 76–90.

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

Legendre, P., Dale, M.R.T., Fortin, M., Gurevitch, J., Hohn, M., Myers, D., 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. Ecography 25, 601–615.

Legendre, P., Fortin, M.J., 1989. Spatial pattern and ecological analysis. Vegetatio 80, 107–138.

Lentile, L.B., Smith, F.W., Shepperd, W.D., 2006. Influence of topography and forest structure on patterns of mixed severity fire in ponderosa pine forests of the South Dakota Black Hills, USA. Int. J. Wildl. Fire 15, 557–566.

Li, F., Zaslavsky, A.M., Landrum, M.B., 2013. Propensity score weighting with multilevel data. Stat. Med. 32, 3373–3387.

Lindenmayer, D.B., Likens, G.E., 2009. Adaptive monitoring: a new paradigm for long-term research and monitoring. Trends Ecol. Evol. 24, 482–486.

Liski, J., Lehtonen, A., Palosuo, T., Peltoniemi, M., Eggers, T., Muukkonen, P., Mäkipää, R., 2006. Carbon accumulation in Finland's forests 1922–2004–an estimate obtained by combination of forest inventory data with modelling of biomass, litter and soil. Ann. For. Sci. 63, 687–697.

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.

Long, J.S., Freese, J., 2006. Regression models for categorical dependent variables using Stata. Stata press.

Malison, R.L., Baxter, C. V, 2010. Effects of wildfire of varying severity on benthic stream insect assemblages and emergence. J. North Am. Benthol. Soc. 29, 1324–1338.

Marcoux, H.M., Daniels, L.D., Gergel, S.E., Da Silva, E., Gedalof, Z., Hessburg, P.F., 2015. Differentiating mixed-and high-severity fire regimes in mixed-conifer forests of the Canadian Cordillera. For. Ecol. Manage. 341, 45–58.

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

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.

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

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

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

Meidinger, D., Pojar, J., 1991. Ecosystems of British Columbia. Spec. Rep. Ser. For. Br. Columbia.

Meigs, G.W., Donato, D.C., Campbell, J.L., Martin, J.G., Law, B.E., 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., Harmon, M.E., O'Connell, K.E.B., 2009. Forest fuel reduction alters fire severity and long-term carbon storage in three Pacific Northwest ecosystems. Ecol. Appl. 19, 643–655.

Morgan, S.L., Winship, C., 2015. Counterfactuals and causal inference. Cambridge University Press.

Nadeem, K., Taylor, S.W., Woolford, D.G., Dean, C.B., 2019. Mesoscale spatiotemporal predictive models of daily human-and lightning-caused wildland fire occurrence in British Columbia. Int. J. Wildl. fire 29, 11–27.

Nitschke, C.R., Innes, J.L., 2013. Potential effect of climate change on observed fire regimes in the Cordilleran forests of South-Central Interior, British Columbia. Clim. Change 116, 579–591.

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. 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.

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

Ottmar, R.D., 2014. Wildland fire emissions, carbon, and climate: modeling fuel consumption. For. Ecol. Manage. 317, 41–50.

Palmer, M., Kuegler, O., Christensen, G., 2018. Oregon's forest resources, 2006–2015: Ten-year Forest Inventory and Analysis report. Gen. Tech. Rep. PNW-GTR-971. Portland, OR US Dep. Agric. For. Serv. Pacific Northwest Res. Station. 54 p. 971.

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

Parisien, M.-A., Moritz, M.A., 2009. Environmental controls on the distribution of wildfire at multiple spatial scales. Ecol. Monogr. 79, 127–154.

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

Patorno, E., Glynn, R.J., Hernández-Díaz, S., Liu, J., Schneeweiss, S., 2014. Studies with many covariates and few outcomes: selecting covariates and implementing propensity-score–based confounding adjustments. Epidemiology 25, 268–278.

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

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

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

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

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.

Pickell, P.D., Chavardès, R.D., Li, S., Daniels, L.D., 2020. FuelNet: An Artificial Neural Network for Learning and Updating Fuel Types for Fire Research. IEEE Trans. Geosci. Remote Sens.

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.

Pojar, J., Klinka, K., Meidinger, D. V, 1987. Biogeoclimatic ecosystem classification in British Columbia. For. Ecol. Manage. 22, 119–154.

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

Pufahl, A., Weiss, C.R., 2009. Evaluating the effects of farm programmes: results from propensity score matching. Eur. Rev. Agric. Econ. 36, 79–101.

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

Regelbrugge, J.C., Conard, S.G., 1993. Modeling tree mortality following wildfire in Pinus ponderosa forests in the central Sierra-Nevada of California. Int. J. Wildl. Fire 3, 139–148.

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

Robinne, F.-N., Hallema, D.W., Bladon, K.D., Buttle, J.M., 2020. Wildfire impacts on hydrologic ecosystem services in North American high-latitude forests: A scoping review. J. Hydrol. 581, 124360.

Rochester, C.J., Brehme, C.S., Clark, D.R., Stokes, D.C., Hathaway, S.A., Fisher, R.N., 2010. Reptile and amphibian responses to large-scale wildfires in southern California. J. Herpetol. 44, 333–351.

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

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

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

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

Rubin, D.B., Thomas, N., 1996. Matching Using Estimated Propensity Scores: Relating Theory to Practice. Biometrics 52, 249–264.

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.

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

Samal, A., Seth, S., Cueto 1, K., 2004. A feature-based approach to conflation of geospatial sources. Int. J. Geogr. Inf. Sci. 18, 459–489.

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

Shipman, J.E., Swanquist, Q.T., Whited, R.L., 2017. Propensity score matching in accounting research. Account. Rev. 92, 213–244.

Silva, J.S., Tenreyro, S., 2006. The log of gravity. Rev. Econ. Stat. 88, 641-658.

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, J.E., Heath, L.S., Skog, K.E., Birdsey, R.A., 2006. Methods for calculating forest ecosystem and harvested carbon with standard estimates for forest types of the United States. Gen. Tech. Rep. NE-343. Newt. Square, PA US Dep. Agric. For. Serv. Northeast. Res. Station. 216 p. 343.

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

Smithwick, E.A.H., Turner, M.G., Mack, M.C., Chapin, F.S., 2005. Postfire soil N cycling in northern conifer forests affected by severe, stand-replacing wildfires. Ecosystems 8, 163–181.

Soares, P., Tomé, M., Skovsgaard, J.P., Vanclay, J.K., 1995. Evaluating a growth model for forest management using continuous forest inventory data. For. Ecol. Manage. 71, 251–265.

Spracklen, D. V, Mickley, L.J., Logan, J.A., Hudman, R.C., Yevich, R., Flannigan, M.D., Westerling, A.L., 2009. Impacts of climate change from 2000 to 2050 on wildfire activity and carbonaceous aerosol concentrations in the western United States. J. Geophys. Res. Atmos. 114.

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. Glob. Ecol. Biogeogr. 20, 744–754.

Stinson, G., Kurz, W.A., Smyth, C.E., Neilson, E.T., Dymond, C.C., Metsaranta, J.M., Boisvenue, C., Rampley, G.J., Li, Q., White, T.M., Blain, D., 2011. An inventory-based analysis of Canada's managed forest carbon dynamics, 1990 to 2008. Glob. Chang. Biol. 17, 2227–2244.

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

Stuart, E.A., Green, K.M., 2008. Using full matching to estimate causal effects in nonexperimental studies: Examining the relationship between adolescent marijuana use and adult outcomes. Dev. Psychol. 44, 395.

Stuart, E.A., Rubin, D.B., 2004. Matching methods for causal inference: Designing observational studies. Harvard Univ. Dep. Stat. mimeo.

Taylor, S.W., Woolford, D.G., Dean, C.B., Martell, D.L., 2013. Wildfire prediction to inform fire management: Statistical science challenges. Stat. Sci. 28, 586–615.

Tomppo, E., Olsson, H., Ståhl, G., Nilsson, M., Hagner, O., Katila, M., 2008. Combining national forest inventory field plots and remote sensing data for forest databases. Remote Sens. Environ. 112, 1982–1999.

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.

Ung, C.-H., Bernier, P., Guo, X.-J., 2008. Canadian national biomass equations: new parameter estimates that include British Columbia data. Can. J. For. Res. 38, 1123–1132.

Van Mantgem, P.J., Stephenson, N.L., Byrne, J.C., Daniels, L.D., Franklin, J.F., Fulé, P.Z., Harmon, M.E., Larson, A.J., Smith, J.M., Taylor, A.H., 2009. Widespread increase of tree mortality rates in the western United States. Science 323, 521–524.

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

Wang, T., Hamann, A., Spittlehouse, D., Carroll, C., 2016. Locally downscaled and spatially customizable climate data for historical and future periods for North America. PLoS One 11, e0156720.

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.

Westerling, A.L., Hidalgo, H.G., Cayan, D.R., Swetnam, T.W., 2006. Warming and earlier spring increase western US forest wildfire activity. Science 313, 940–943.

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

Wimberly, M.C., Liu, Z., 2014. Interactions of climate, fire, and management in future forests of the Pacific Northwest. For. Ecol. Manage. 327, 270–279.

Winship, C., Mare, R.D., 1992. Models for sample selection bias. Annu. Rev. Sociol. 18, 327–350.

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

Woo, H., Eskelson, B.N.I., Monleon, V.J., 2021b. Sensitivity analysis on distance-adjusted propensity score matching for wildfire effect quantification using national forest inventory data. Environ. Model. Softw. 144, 105163.

Woodall, C.W., Monleon, V.J., 2008. Sampling protocol, estimation, and analysis procedures for the down woody materials indicator of the FIA program. Gen. Tech. Rep. NRS-22. Newt. Square, PA US Dep. Agric. For. Serv. North. Res. Station. 68 p. 22.

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.

Wulder, M.A., Hermosilla, T., White, J.C., Coops, N.C., 2020. Biomass status and dynamics over Canada's forests: Disentangling disturbed area from associated aboveground biomass consequences. Environ. Res. Lett. 15, 94093.

Yong, A.G., Lemyre, L., 2019. Getting Canadians prepared for natural disasters: a multi-method analysis of risk perception, behaviors, and the social environment. Nat. Hazards 98, 319–341.

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

Zanutto, E., Lu, B., Hornik, R., 2005. Using propensity score subclassification for multiple treatment doses to evaluate a national antidrug media campaign. J. Educ. Behav. Stat. 30, 59–73.

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

Appendices

Appendix A

A.1

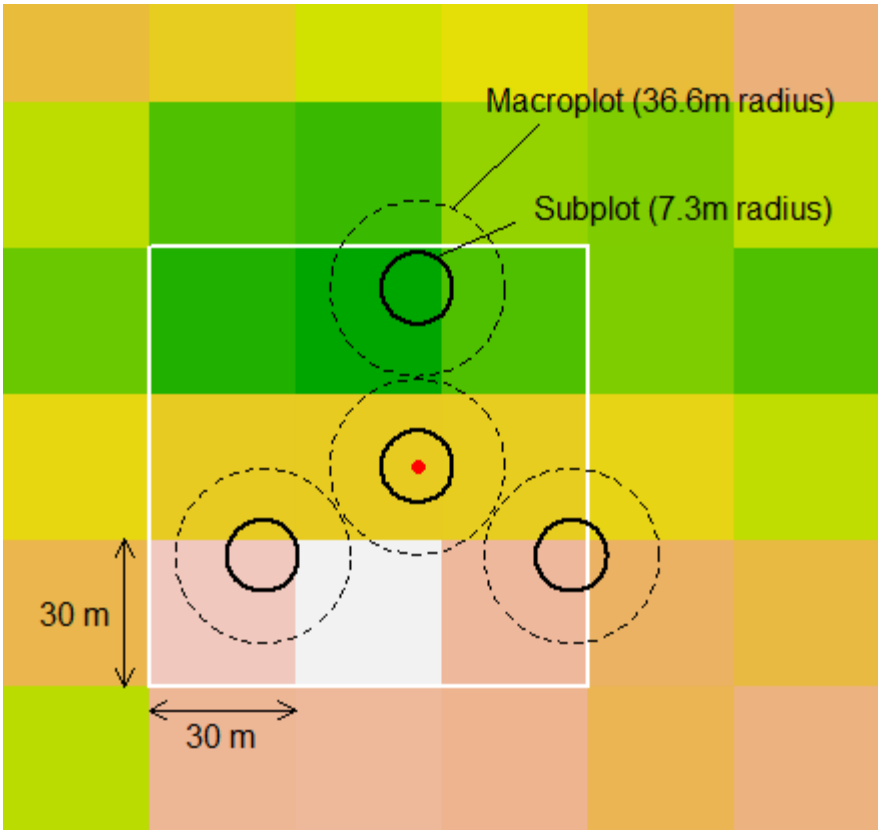


Figure A.1.1 Example of an FIA plot laid over a satellite imagery file with 30-m resolution. FIA plot consists of four concentric subplots and macroplots with approximately 7.3m and 36.6m radius respectively. All live trees over 12.7 cm DBH are measured within subplots, and trees over 61 cm and 76.2 cm are measured within macroplots in eastern and western part of WA and OR, respectively. When extracting satellite imagery (e.g., PRISM and Landsat) metrics, the centres of FIA plots (red dot) are spatially overlaid on the image and the pixel containing the centre and 8 adjacent pixel values (white square) are used to calculate the mean (continuous variables) or mode (categorical variables) of the metrics.

Table A.1.1 Coefficients and standard errors (SE) of sixteen covariates from the multiple logistic regression model estimating the propensity scores for burned and unburned plots. The estimated propensity scores are the same for PSM and DAPSM.

Variable	Description	Coefficient (SE)	Signif.
Topography			
ELEV	Elevation (in metres)	0.0001 (0.0000)	
SLOPE	Slope (in degrees)	0.0109 (0.0020)	***
cosASP	Aspect (in degrees azimuth), cosine value	0.2111 (0.2535)	
Land cover			
PROPforest	Proportion of forested area	0.3639 (0.3223)	
OWN	Owner group code (categorical)	-0.6659 (0.1516)	***
PHYSCL	Physiographic class code (categorical)	-0.7334 (0.0991)	***
Climate			
annpre	Annual precipitation (ln mm * 100) from PRISM 30-year normals	0.0182 (0.0024)	***
anntmp	Mean annual temperature (degrees C * 100)	-0.0058 (0.0009)	***
cvpre	Coefficient of variation of mean monthly precipitation of December and July (wettest and driest months)	-0.0003 (0.0000)	***
diftmp	Difference between mean August maximum temperature and mean December minimum temperature (degrees C * 100)	-0.0023 (0.0004)	***
smrmnvpd	Mean summer vapor pressure deficit (hPa)	0.0049 (0.0007)	***
smrtp	Growing season moisture stress (ratio of temperature to precipitation from May-September) (degrees C/ln mm)	0.0264 (0.0037)	***
Remotely sensed data			
NDVI	Normalized difference in vegetation index	0.6032 (1.1169)	
TC1	Tasseled cap brightness	-0.0004 (0.0002)	**
TC2	Tasseled cap greenness	-0.0002 (0.0003)	
TC3	Tasseled cap wetness	-0.0004 (0.0003)	

Significance (Signif): *** $p \leq 0.001$, ** $0.001 < p \leq 0.05$, * $0.05 < p \leq 0.1$

Table A.1.2 Means and standard deviations (SD) of propensity scores (PS) of burned, unburned, and matched unburned plots under three matching scenarios: propensity score matching (PSM), spatial matching (SM), and distance-adjusted propensity score matching (DAPSM).

	Number plots	Mean PS	SD
Unmatched unburned	22,539	0.025	0.028
Burned	611	0.057	0.050
Matched unburned			
PSM	611	0.057	0.050
SPM	611	0.049	0.046
DAPSM	611	0.056	0.049

A.2

Table A.2.1 Forest inventory and image data sources for topography, weather, and vegetation indices maps. All data contain the extent of Washington and Oregon, USA. Note that the Forest Inventory and Analysis (FIA) plot database is available at an open repository, with an exception of plot locations due to the confidentiality issue. Users needing the exact plot coordinates may contact the FIA program directly (www.fia.fs.fed.us/tools-data/customer-service/) with an approval from the FIA program and a Confidentiality Agreement, following the Food Security Act of 1985 that protects the privacy of the landowners and the integrity of the sample.

Map/Image Data Source	Resolution and Periods	Source
Forest Inventory and Analysis (FIA) plot data	Vector data; 2001-2016	U.S. Department of Agriculture Forest Service, Pacific Northwest Research Station, www.fs.fed.us/pnw/rma/fia-topics/inventory-data/
Wildfire burn severity maps	30m; 1984-2016	Monitoring Trends in Burn Severity (MTBS), http://www.mtbs.gov/
PRISM 30-year average climate maps	30m; 1981-2010	Landscape Ecology, Modeling, Mapping, and Analysis (LEMMA) team, https://lemma.forestry.oregonstate.edu/
Landsat 8 Satellite imagery metrics	30m; 1984-2016	Landscape Ecology, Modeling, Mapping, and Analysis (LEMMA) team, https://lemma.forestry.oregonstate.edu/

Table A.2.2 Summary of covariates of 1) burned plots, 2) all unburned plots, 3) control plots matched by a) propensity score matching (PSM, w=1), b) spatial matching (SM, w=0), and c) distance-adjusted propensity score matching (DAPSM, w=0.5). Two-sample t-tests for continuous variables and chi-square tests for categorical variables were conducted to see if the selected unburned plots are different from burned plots. Covariates that were significantly different from burned plots (α=0.05) are highlighted in bold.

	Burned n=611	Unburned n=22,539	PSM Control n=611	SM Control n=611	DAPSM Control n=611
Covariates (continuous)			Mean (SD)		
ELEV	1,242 (440)	**1,073 (535)**	1,254 (429)	1,248 (454)	1,243 (457)
SLOPE	376 (24)	**30 (24)**	37 (28)	**34 (24)**	36 (25)
ASP	0.03 (0.72)	**0.11 (0.74)**	0.08 (0.72)	0.02 (0.75)	0.01 (0.73)
PROPforest	0.95 (0.14)	0.95 (0.14)	0.96 (0.12)	0.95 (0.13)	0.94 (0.14)
annpre	680 (65)	**702 (67)**	680 (64)	679 (62)	678 (61)
anntmp	737 (270)	756 (222)	738 (213)	732 (277)	739 (272)
cvpre	6,929 (2,163)	**7,272 (1,745)**	6,979 (1,998)	6,925 (2,127)	6,957 (2,103)
diftmp	3,053 (293)	**2,881 (455)**	3,051 (360)	3,047 (303)	3,056 (311)
smrmnvpd	1,244 (229)	**1,091 (276)**	1,244 (247)	1,235 (231)	1,247 (225)
smrtp	278 (52)	**263 (45)**	278 (52)	277 (54)	280 (55)
NDVI	0.59 (0.12)	**0.64 (0.14)**	0.59 (0.14)	0.59 (0.13)	0.58 (0.13)
TC1	2,281 (654)	2,313 (621)	2,254 (587)	2,318 (609)	2,313 (628)
TC2	1,040 (445)	**1,275 (593)**	1,029 (445)	1,048 (441)	1,033 (458)
TC3	-831 (570)	**-706 (562)**	-827 (568)	-849 (560)	-854 (567)
Covariates (category)			Number of plots (%)		
OWN					
1: Public lands	552 (90)	**17,521 (78)**	549 (90)	548 (90)	556 (91)
2: Private lands	59 (10)	**5,018 (22)**	62 (10)	62 (10)	55 (9)
PHYSCL					
1: Xeric sites	292 (48)	**5,197 (23)**	285 (47)	**222 (36)**	276 (45)
2: Mesic sites	319 (52)	**17,213 (76)**	326 (53)	**3887 (64)**	335 (55)
3: Hydric sites	0 (0)	**129 (1)**	0 (0)	**1 (0.16)**	0 (0)

Table A.2.3 Summary statistics for carbon mass (Mg ha^{-1}) for each carbon pool. The numbers of plots included in each severity group (n) were shown with the number of plots with no carbon mass from the corresponding pool (# of 0s).

	Severity	Method	n (# of 0s)	Median	Mean	SD
Large live		PSM	611 (13)	47.21	77.23	84.29
	0	SPM	611 (26)	43.07	69.66	75.93
		DAPSM	611 (28)	43.90	69.74	74.15
	1		120 (12)	43.97	67.40	75.45
	2		201 (27)	34.49	51.14	55.87
	3		145 (43)	11.04	26.61	36.82
	4		145 (102)	0	10.14	27.68
Small live		PSM	611 (250)	0.79	3.39	6.76
	0	SPM	611 (251)	1.06	3.20	5.54
		DAPSM	611 (261)	0.63	3.19	6.20
	1		120 (82)	0	1.40	3.07
	2		201 (157)	0	0.80	3.10
	3		145 (125)	0	0.42	2.09
	4		145 (132)	0	0.39	1.89
Large dead		PSM	611 (166)	1.62	5.63	10.02
	0	SPM	611 (166)	1.88	6.28	10.76
		DAPSM	611 (177)	1.64	6.11	10.99
	1		120 (32)	3.03	8.69	12.62
	2		201 (18)	9.45	17.90	22.92
	3		145 (12)	15.40	23.35	23.71
	4		145 (4)	32.45	51.06	53.86
CWD		PSM	611 (94)	5.50	11.49	19.09
	0	SPM	611 (82)	5.94	10.45	13.36
		DAPSM	611 (90)	5.41	9.79	12.06
	1		120 (32)	2.00	7.39	14.88
	2		201 (45)	3.11	6.98	11.30
	3		145 (26)	3.00	6.42	9.59
	4		145 (18)	465	8.62	16.13

A.3

To calculate the average wildfire effects on aboveground woody carbon masses, Poisson pseudo-maximum likelihood models (Silva and Tenreyro 2006) were fitted for the carbon mass using the wildfire severities as the explanatory variable. The model was defined as following:

$$\log(\mu) = \beta_0 + \beta_1(severity = 1) + \beta_2(severity = 2) + \beta_3(severity = 3) + \beta_4(severity = 4)$$

where μ denotes the mean carbon mass (Mg ha^{-1}). $severity$ is a factor variable with five categories; control (severity=0, reference level), low-low (severity=1), low (severity=2), moderate (severity=3), and high (severity=4). The models were fitted for four different aboveground woody carbon pools; large live trees, small live trees, large dead trees, and coarse woody materials (CWD), as well as the total carbon mass including all the pools. The fixed effect of $severity$ was statistically significant in all models ($F < 0.0001$, Type III test).

Table A.3.1 Percent of aboveground woody carbon contents (Mg ha^{-1}) of large live trees (≥ 12.7 cm DBH), small live trees (≥ 2.54 cm DBH), large dead trees (≥ 12.7 cm DBH), and coarse woody debris (CWM, ≥ 7.6 cm) compared to the control plots selected from the three matching methods. The coefficients and the 95% CIs (in parentheses) were estimated from Poisson pseudo-maximum likelihood models with Tukey's adjustment for different fire severities.

		% of aboveground woody carbon compared to control plots					
Carbon pool	Severity	PSM		SM		DAPSM	
Large live	1	86	(56, 131)	85	(67, 110)	90	(68, 119)
	2	65	(47, 89)	83	(63, 106)	82	(65, 102)
	3	35	(23, 53)	42	(29, 60)	41	(28, 61)
	4	14	(7, 26)	14	(8, 26)	15	(8, 28)
Small live	1	39	(19, 80)	48	(25, 90)	58	(31, 102)
	2	23	(10, 53)	22	(10, 50)	28	(13, 60)
	3	15	(4, 50)	14	(4, 46)	11	(3, 36)
	4	10	(3, 36)	12	(4, 36)	11	(3, 35)
Large dead	1	103	(56, 188)	105	(65, 169)	111	(57, 216)
	2	354	(242, 519)	345	(233, 512)	316	(226, 443)
	3	365	(237, 562)	342	(220, 531)	346	(234, 512)
	4	880	(536, 1446)	708	(467, 1075)	937	(568, 1546)
CWM	1	64	(38, 115)	67	(38, 118)	78	(44, 137)
	2	55	(34, 89)	76	(52, 111)	73	(51, 106)
	3	64	(39, 103)	61	(40, 92)	60	(39, 93)
	4	77	(46, 129)	75	(51, 110)	90	(57, 112)
Total	1	83	(56, 120)	96	(80, 116)	93	(78, 112)
	2	75	(57, 98)	82	(63, 105)	89	(69, 115)
	3	58	(43, 78)	66	(51, 85)	64	(49, 84)
	4	71	(51, 99)	73	(57, 93)	78	(61, 100)

Appendix B

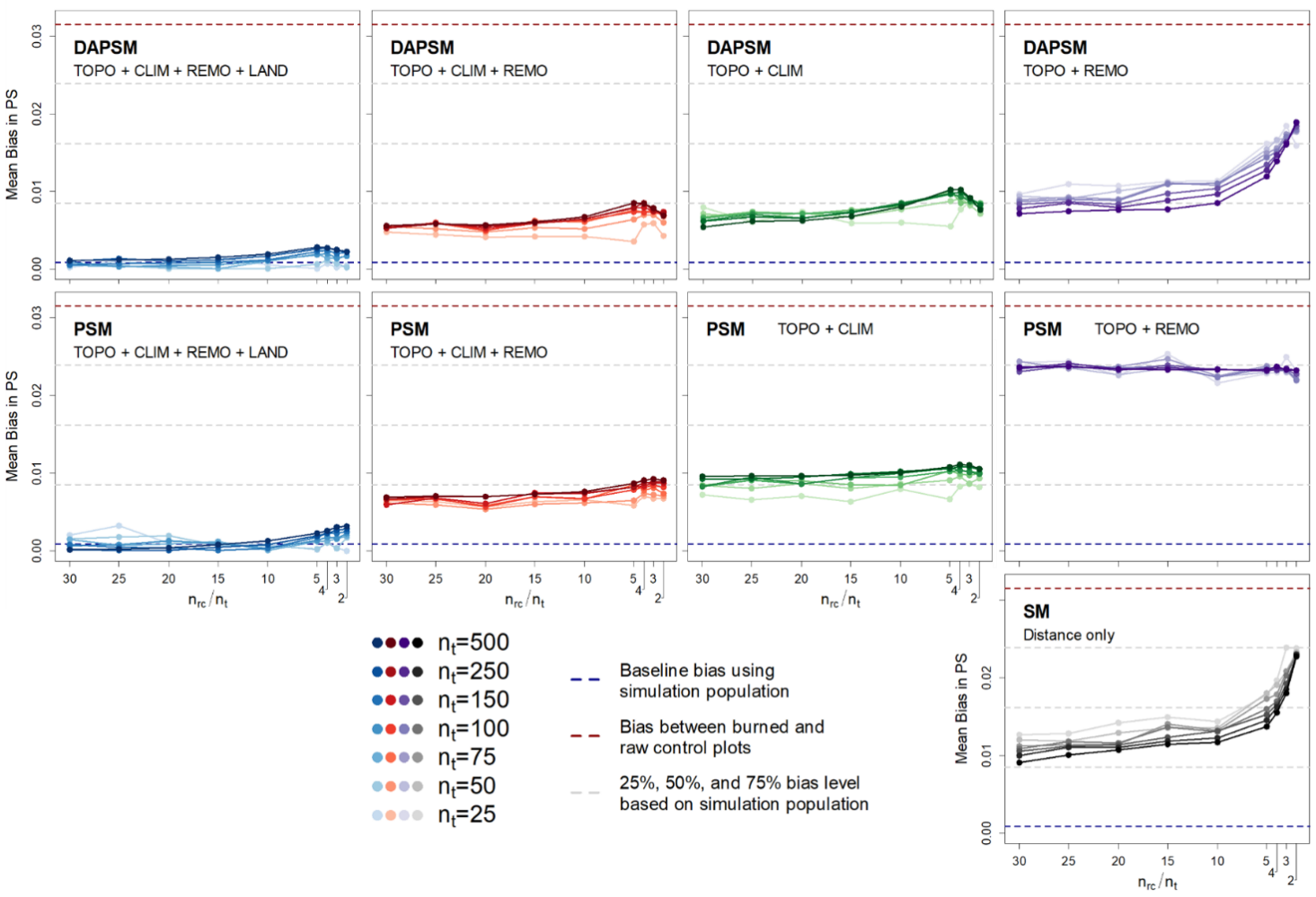


Figure B.1. Mean bias in propensity scores (PS) under different number of burned plots (n_t), unburned plots to burned plots (n_{rc}/n_t), and covariate set. The burned and unburned plots were matched based on each combination of simulation scenarios using DAPSM, PSM, and SM.

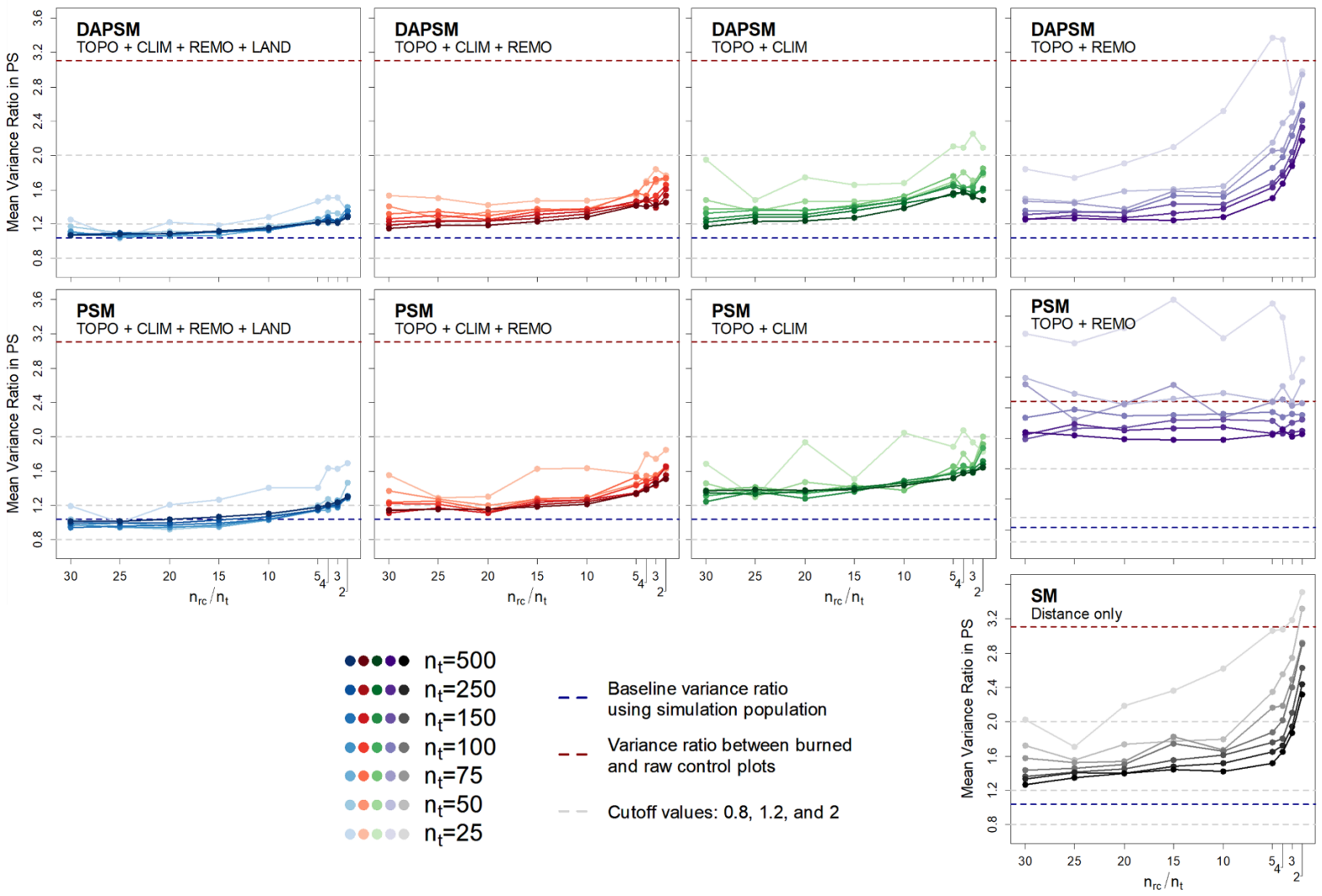


Figure B.2 Mean variance ratio in propensity scores (PS) between burned and control plots under different combinations of number of burned plots (n_t), unburned plots to burned plots (n_{rc}/n_t), and covariate set, matched based on DAPSM, PSM, or SM.

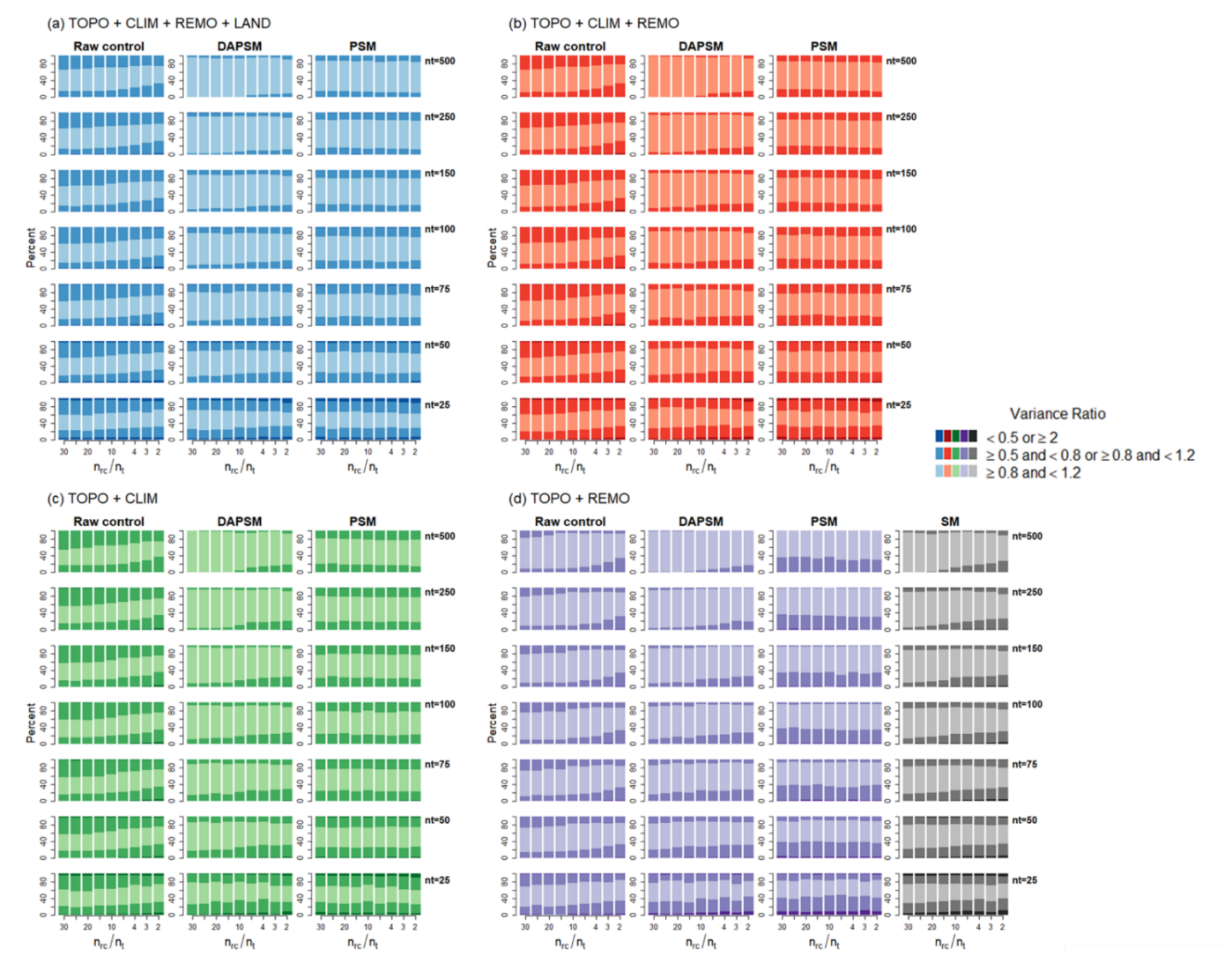


Figure B.3 Percentages of the number of covariates of which residual variance ratio falls within each range: < 0.5 or ≥ 2 (extreme), ≥ 0.5 and < 0.8 or ≥ 0.8 and < 1.2, and > 0.8 and ≥ 1.2 (adequate) under different covariate sets.

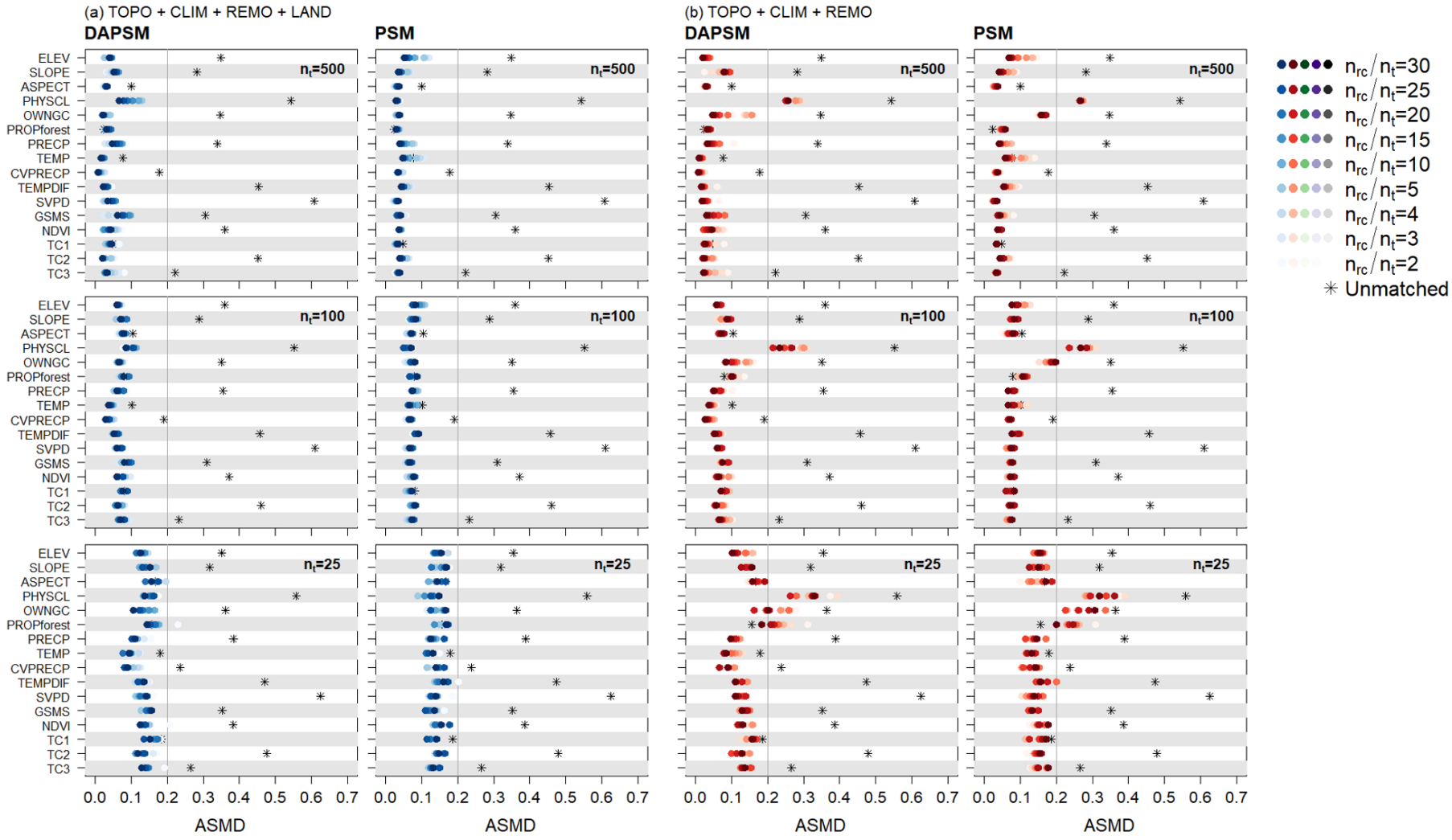


Figure B.4 Absolute standardized mean difference (ASMD) of sixteen covariates unmatched and matched by DAPSM and PSM with varying numbers of unburned plots to burned plots (n_{rc}/n_t) using different covariate sets (a)-(e) for n_t=500, 100, and 25.

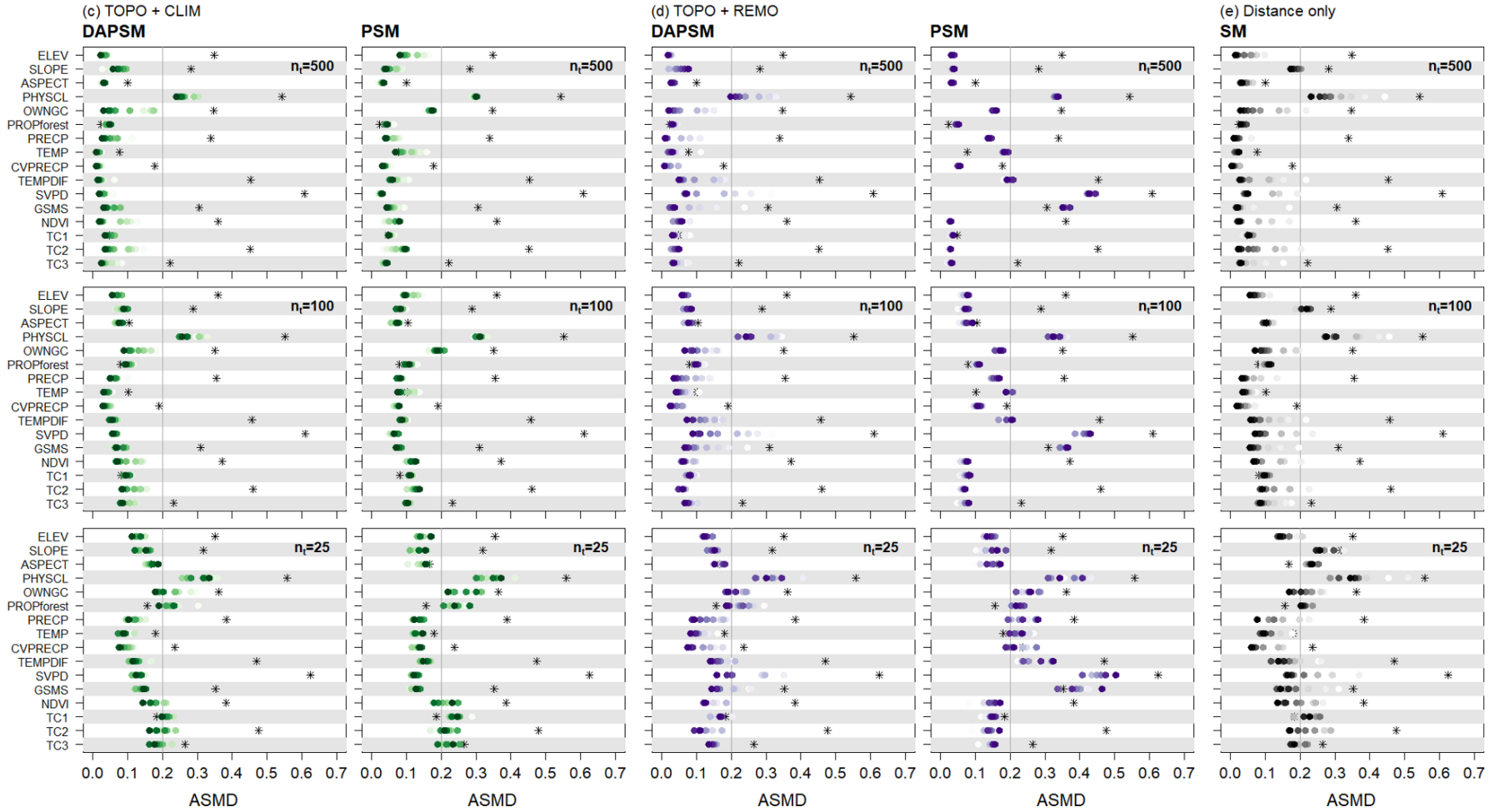


Figure B.4 (cont.) Absolute standardized mean difference (ASMD) of sixteen covariates unmatched and matched by DAPSM and PSM with varying numbers of unburned plots to burned plots (n_{rc}/n_t) using different covariate sets (a)-(e) for n_t=500, 100, and 25.

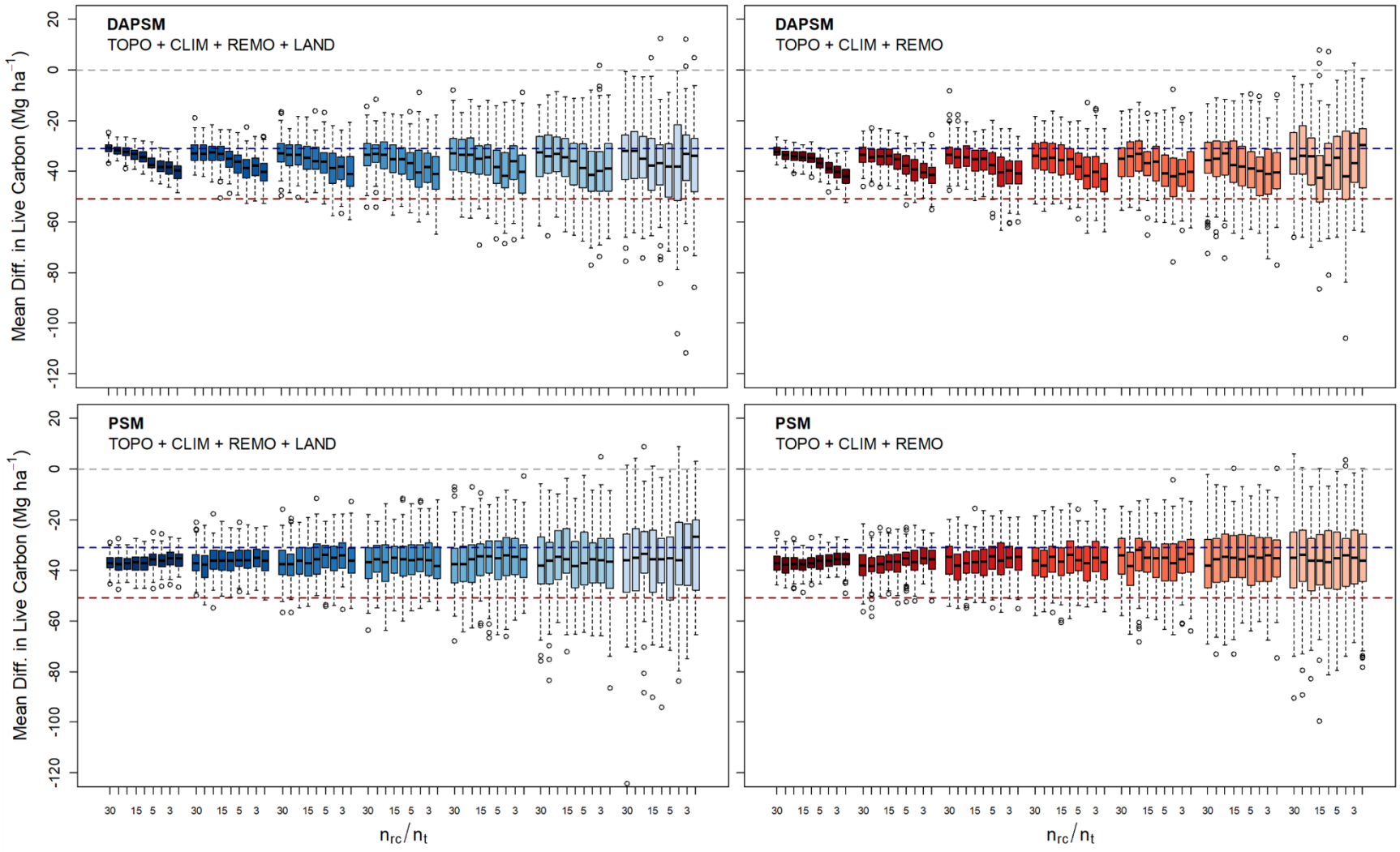


Figure B.5 Distribution of the estimates of wildfire effect on aboveground live woody carbon under different n_t, n_{rc}/n_t, and covariate sets (TOPO + CLIM + REMO + LAND and TOPO + CLIM + REMO) obtained through DAPSM and PSM.

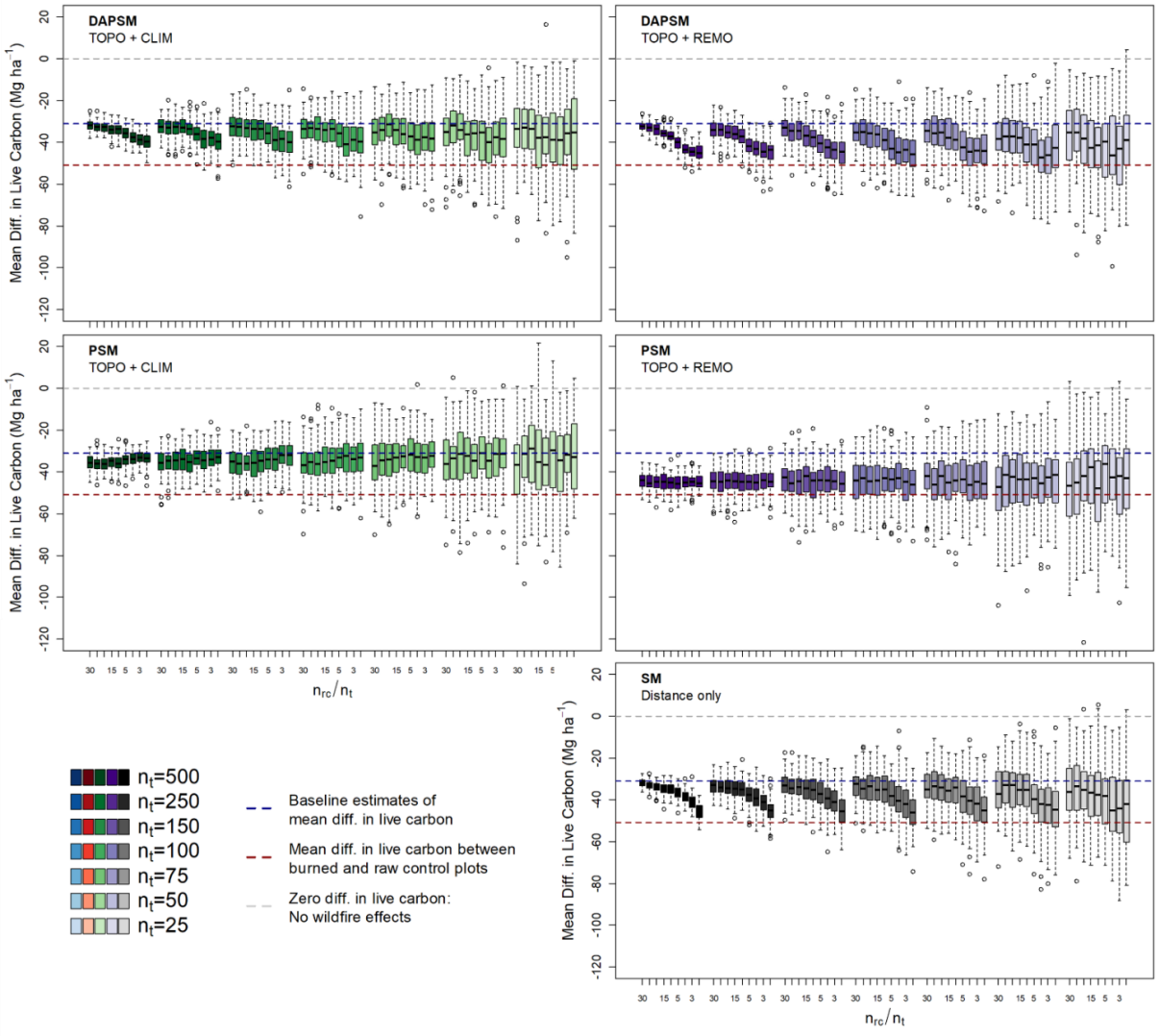


Figure B.5 (cont.) Distribution of the estimates of wildfire effect on aboveground live woody carbon under different n_t, n_{rc}/n_t, and covariate sets (TOPO + CLIM and TOPO + REMO) obtained through DAPSM, PSM, and SM.

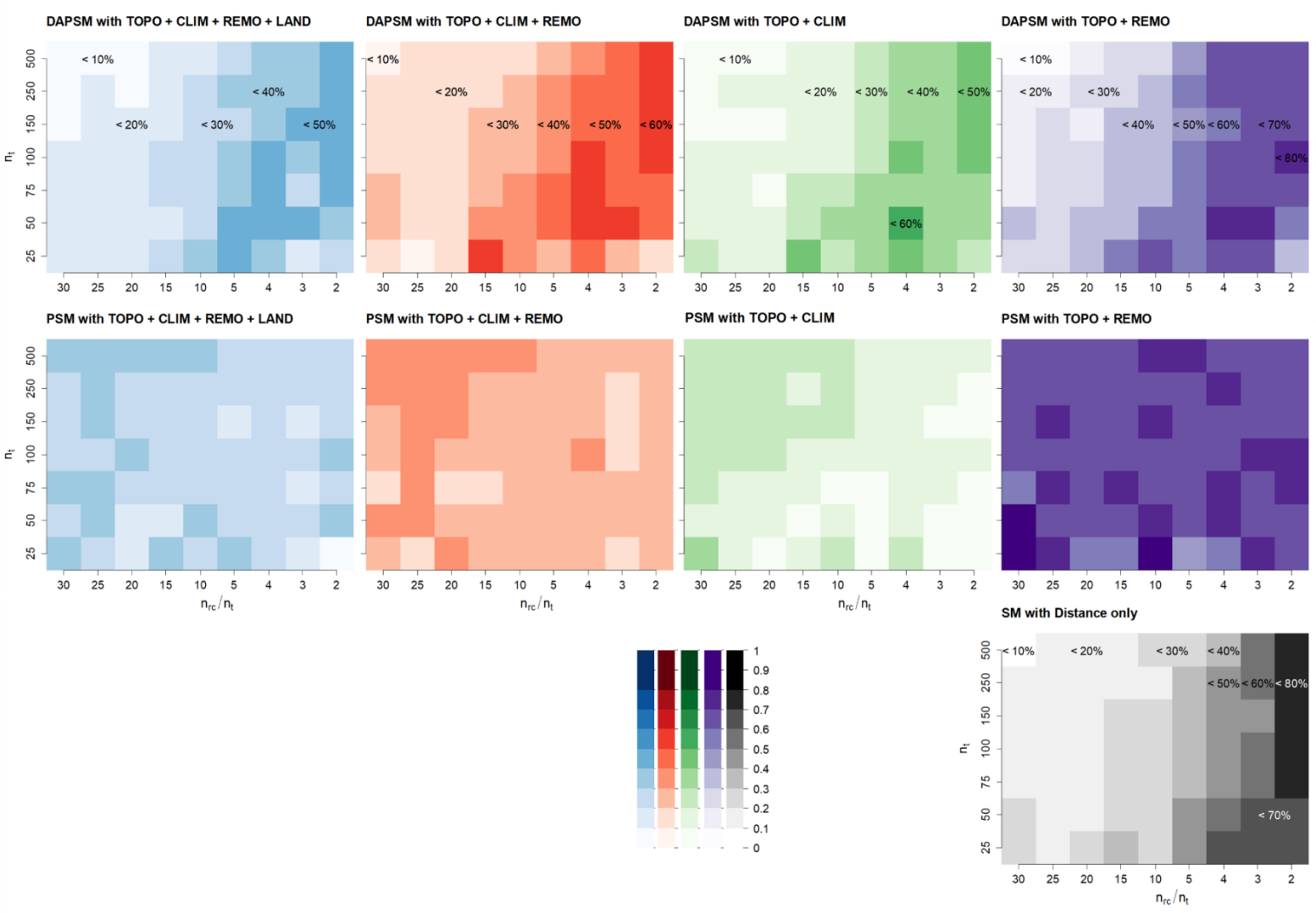


Figure B.6 Matrix plot of percent bias in the wildfire effect estimates under different combinations of n_t, n_{rc}/n_t, and covariate sets. The estimates were obtained through DAPSM, PSM, and SM.

Appendix C

C.1 Description of the additional analysis using the FAIB-only datasets

I examined an additional scenario of raw control pool for the three fire datasets: Elephant Hill Fire, Prouton Lakes Fire, and south-central BC fires data. I conducted matching with and without replacement using only the forest inventory plots from Forest Analysis and Inventory Branch (FAIB) data as the raw control plots, excluding the sampled unburned plots from the fire datasets. The raw control plots now contain 632 FAIB plots to match to each of the burned plots in the fire datasets (Table C.1.1). For the south-central BC fire data only, I excluded the pre-measurements of the 34 burned plots from the pool of raw controls (Table C.1.1). All the processes of distance-adjusted propensity score matching (DAPSM) and evaluation of the results were repeated based on this additional raw control pool.

Table C.1.1 Numbers of sampled unburned plots, burned plots, and raw control plots used in the additional analysis.

Dataset	# Sampled unburned plots	# Sampled burned plots	# Raw controls (FAIB plots only)
South-central BC fires	1	34	598
Elephant Hill Fire	16	89	632
Prouton Lakes Fire	7	39	632

C.2 Results of DAPSM with and without replacement from the additional analysis

The spatial distributions of burned and control plots obtained from DAPSM with and without replacement using the FAIB plots only are displayed in Figure C.2.1. I obtained different numbers of the controls as well as the mean propensity scores (PS) across the plot groups from those of the controls using both the sampled unburned plots and the FAIB plots as raw controls in matching (Table C.2.1). For the Elephant Hill Fire data, DAPSM with replacement selected 24 controls for 89 burned plots. One control plot in the Elephant Hill Fire data was matched to 20 different burned plots, and three control plots were matched 10-13 times. Furthermore, DAPSM with replacement for the Gavin Lake Fire data resulted in only three controls: One sampled unburned plot was matched to 32 burned plots. For the Williams Lake data, there was no difference in the controls from the other scenario of raw controls, which included FAIB plots and sampled unburned plots.

The absolute standardized mean difference (ASMD) and mean values of the thirteen environmental covariates among burned and control plots are reported in Figure C.2.2 and Table C.2.2. We also computed the mean distances among the sampled plots and matched pairs of burned and control plots (Table C.2.3).

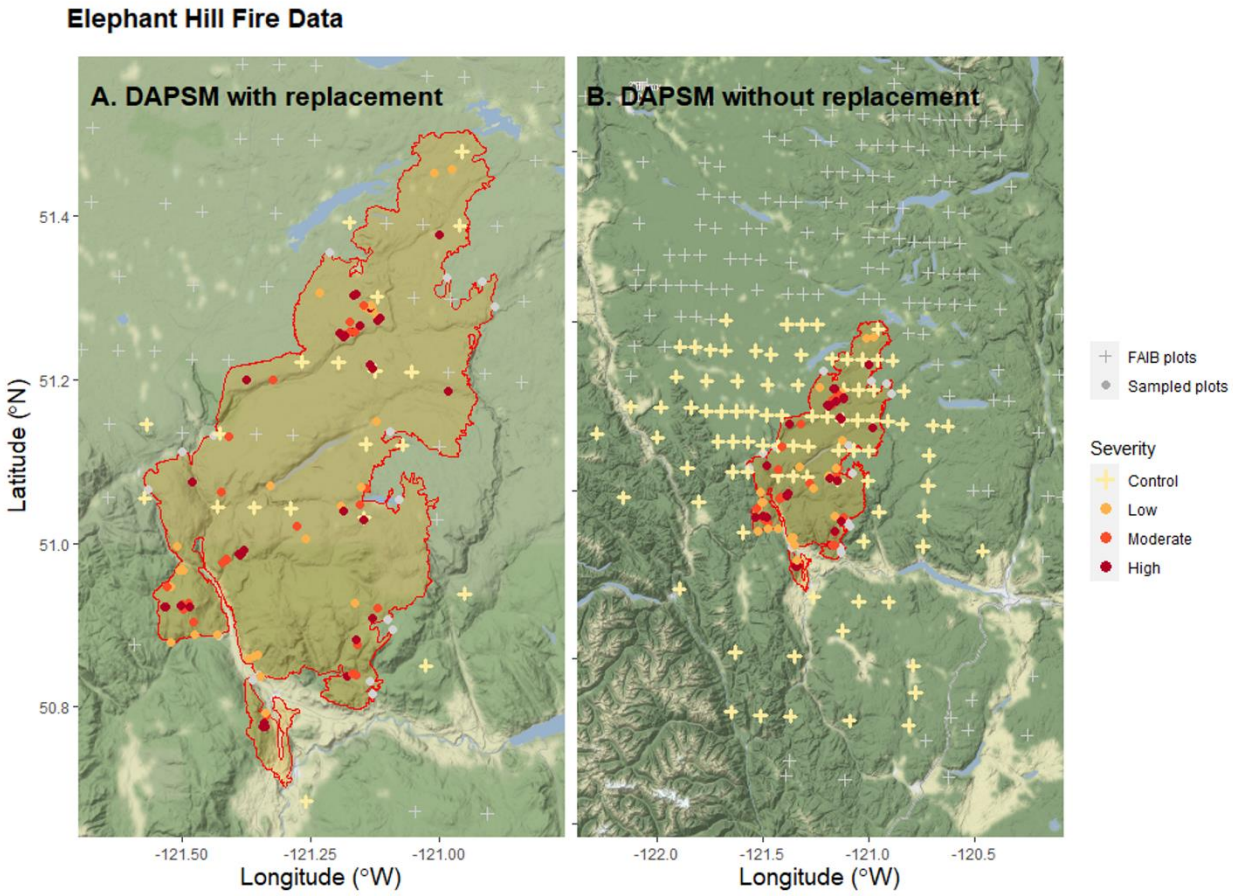


Figure C.2.1 Spatial distributions of burned plots and control plots for the Elephant Hill Fire data. The controls were obtained through a) DAPSM with replacement and b) DAPSM without replacement. Crosses present locations of the FAIB plots, and circles present the burned plots for each dataset. The FAIB plots selected as controls were marked as bold cross in yellow.

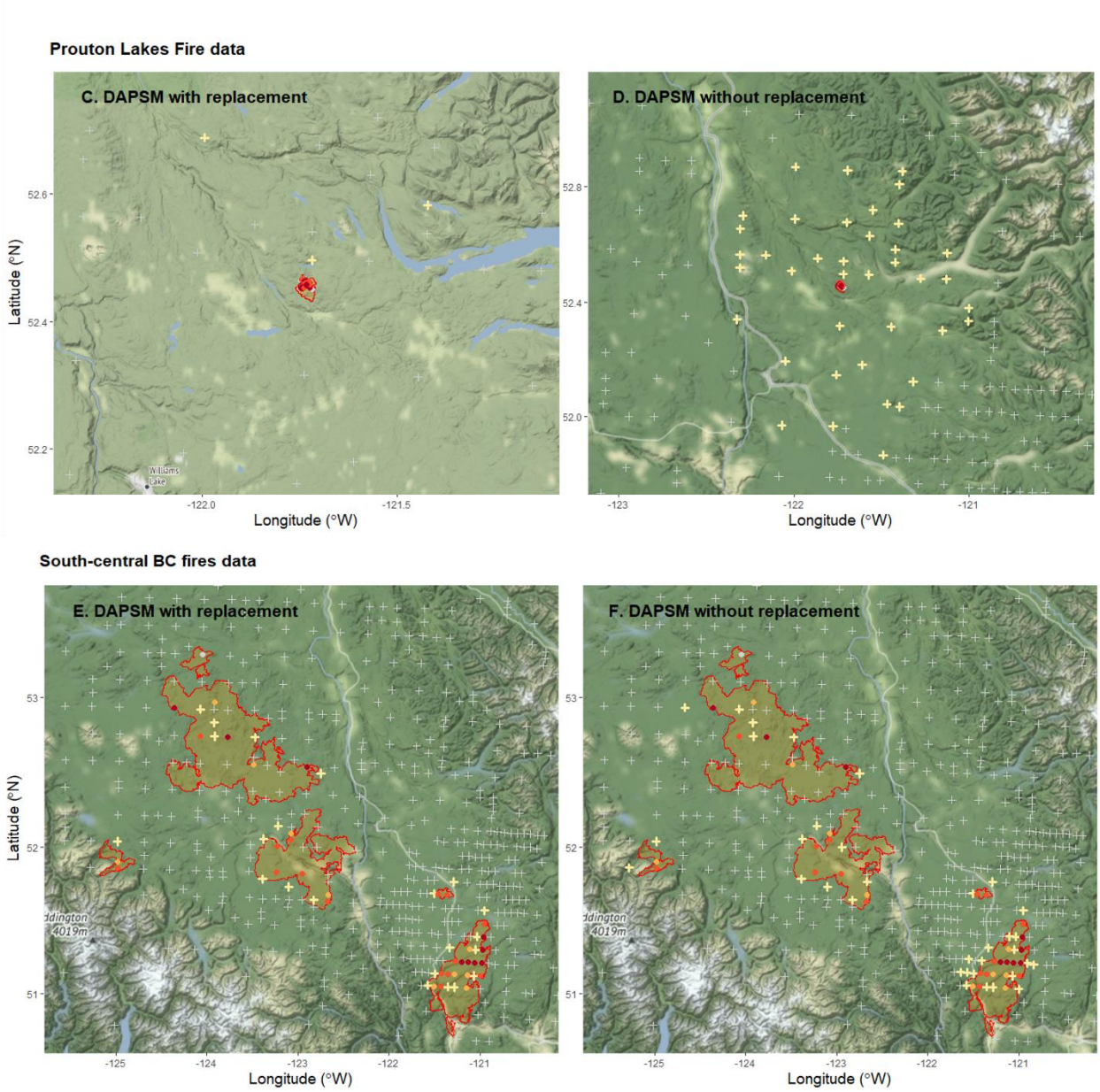


Figure C.2.1 (cont.) Spatial distributions of burned plots and control plots obtained from DAPSM with and without replacement for the Prouton Lakes Fire—c) and d), and the south-central BC fires—e) and f). Crosses present locations of the FAIB plots, and circles present the burned plots for each dataset. The FAIB plots selected as controls were marked as bold cross in yellow.

Table C.2.1 The mean and standard deviation (s.d.) of propensity scores (PS) of burned plots, raw controls, and controls matched with and without replacement for the three datasets: Elephant Hill Fire, and Prouton Lakes Fire, and south-central BC fires data. The raw controls included the FAIB plots only, excluding the sampled unburned plots in the three datasets. n denotes the numbers of plots in each group. Statistically significant differences from the PS of burned plots were marked as * (p<0.0001), ** (p<0.001), and * (p<0.05).**

		Burned plots	Raw controls		Controls			
					DAPSM without replacement		DAPSM with replacement	
Elephant Hill Fire	n	89	632		89		20	
	Mean PS	0.481	0.073	***	0.238	***	0.473	
	s.d.	0.302	0.136		0.250		0.304	
Prouton Lakes Fire	n	39	632		39		3	
	Mean PS	0.199	0.049	***	0.096	**	0.113	*
	s.d.	0.147	0.087		0.150		0.177	
South-central BC fires	n	34	599		34		22	
	Mean PS	0.119	0.050	***	0.093		0.096	
	s.d.	0.061	0.060		0.049		0.047	

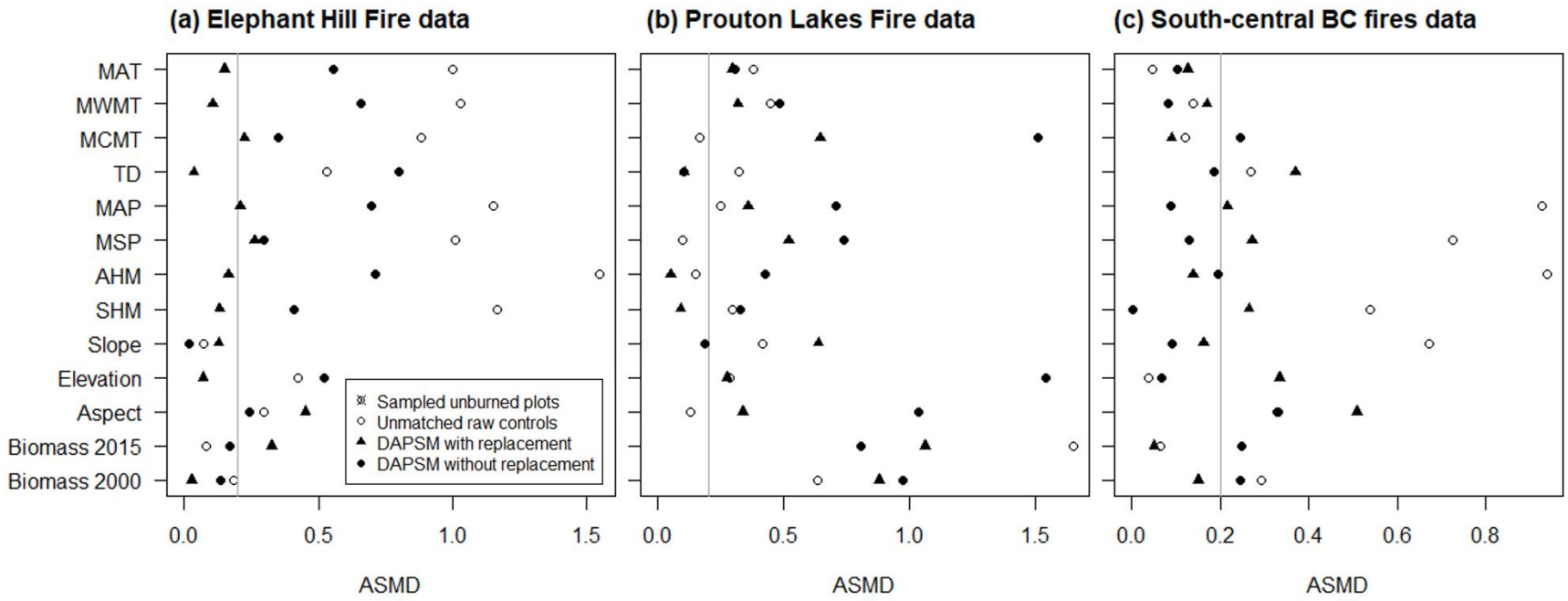


Figure C.2.2 ASMD of the fifteen environmental covariates for a) Elephant Hill Fire data, b) Prouton Lakes Fire data, and c) south-central BC fires data. The ASMD values were computed between burned plots and unmatched raw controls, control plots matched through DAPSM with and without replacement. The gray vertical line presents the cut-off value (0.2) of ASMD between balanced and imbalanced covariate.

Table C.2.2 Mean values of the 13 environmental covariates of burned, sampled unburned, controls selected through DAPSM with and without replacement of Elephant Hill Fire data.

	Burned plots	Raw controls	Controls	
			DAPSM with replacement	DAPSM without replacement
MAP	403.978	663.503	418.528	449.831
MSP	208.461	274.597	217.079	217.854
AHM	38.136	38.906	36.453	32.100
SHM	79.133	23.618	76.062	70.592
MAT	4.754	3.395	4.548	4.065
MWMT	15.874	14.073	15.689	14.809
MCMT	-5.807	-6.896	-6.022	-6.126
TD	21.678	20.968	21.708	20.933
Elevation	1069.742	1179.856	1086.000	1184.438
Slope	8.688	9.250	9.502	8.550
Aspect	-0.177	0.027	-0.440	-0.009
Biomass2000	94.764	104.785	96.236	101.775
Biomass2015	87.729	95.730	121.896	73.428

Table C.2.3 Mean distances (in kilometers) among all burned plots and between matched pairs of burned and control plots selected through DAPSM with and without replacement for the three datasets.

	Burned plots	Matched pairs	
		DAPSM with replacement	DAPSM without replacement
Elephant Hill Fire	27.60	31.58	55.34
Prouton Lakes Fire	0.91	8.58	50.29
South-central BC fires	128.22	131.29	132.53