Mapping wildland fuels and forest structure for land management: a comparison of nearest neighbor imputation and other methods

Kenneth B. Pierce Jr., Janet L. Ohmann, Michael C. Wimberly, Matthew J. Gregory, and Jeremy S. Fried

Abstract: Land managers need consistent information about the geographic distribution of wildland fuels and forest structure over large areas to evaluate fire risk and plan fuel treatments. We compared spatial predictions for 12 fuel and forest structure variables across three regions in the western United States using gradient nearest neighbor (GNN) imputation, linear models (LMs), classification and regression trees (CART), and geostatistical methods (kriging and universal kriging (UK)). Local-scale map accuracy varied considerably across sites, variables, and methods. GNN performed best for forest structure variables in Oregon, but LMs and UK were better for canopy variables and for forest structure variables in Washington and California. Kriging performed poorly throughout, and kriging did not improve prediction accuracy when added to LMs as UK. GNN outperformed CART in predicting vegetation classes and fuel models, complex variables defined by multiple attributes. Regional distributions of vegetation classes and fuel models were accurately represented by GNN and very poorly by CART and LMs. Despite their often limited accuracy at the local scale, GNN maps are useful when information on a wide range of forest attributes is needed for analysis and forest management at intermediate, i.e., landscape to ecoregional, scales.

Résumé : Les aménagistes du territoire ont besoin d'informations fidèles au sujet de la distribution géographique des combustibles en milieu naturel et de la structure de la forêt sur de vastes régions pour évaluer le risque d'incendie et planifier les traitements des combustibles. Nous avons comparé les prédictions spatiales pour 12 variables reliées aux combustibles et à la structure de la forêt dans trois régions de l'ouest des États-Unis en utilisant l'imputation par l'analyse de gradient du plus proche voisin (AGPV), des modèles linéaires (ML), des arbres de régression et de classification (ARC) et des méthodes géostatistiques (krigeage et krigeage universel (KU)). La précision des cartes locales variait considérablement selon les stations, les variables et les méthodes. L'imputation par l'AGPV avait la meilleure performance pour les variables de la structure de la forêt en Oregon mais les ML et le KU étaient meilleurs pour les variables du couvert et de la structure de la forêt dans l'État de Washington et la Californie. Le krigeage a donné des résultats médiocres dans tous les cas et n'améliorait pas la précision lorsqu'il était ajouté aux ML sous forme de KU. L'imputation par l'AGPV était meilleure que les ARC pour prédire les classes de végétation et les modèles de combustibles, des variables complexes définies par de multiples attributs. La distribution régionale des classes de végétation et des modèles de combustibles était correctement représentée par l'imputation par l'AGPV et très inadéquatement par les ARC et les ML. Malgré une précision souvent limitée à l'échelle locale, les cartes produites à partir de l'imputation par l'AGPV sont utiles lorsque des informations sur une large gamme d'attributs de la forêt sont nécessaires pour l'analyse et l'aménagement forestier à des échelles intermédiaires allant du paysage à l'écorégion.

[Traduit par la Rédaction]

Introduction

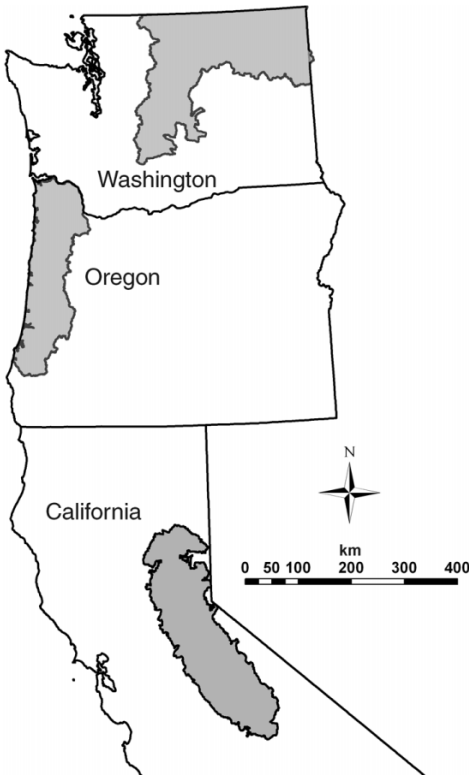
Fuels, weather, and topography dictate wildland fire behavior, but of these, only fuels can be directly managed (Rothermel 1972). Accordingly, forest managers need consistent information about the spatial patterns of wildland fuels over large areas to develop landscape-scale strategies for implementing fuel reduction treatments and reducing the risk of uncharacteristically large and severe wildfires (Keane et al. 2001). Such information allows managers to identify areas with risk of extreme fire behavior and to prioritize the locations of fuel treatments accordingly. In addition, spatial simulations of fire behavior require spatially explicit data on fuel loadings and forest structure at broad scales to predict the spatial patterns of fire behavior and effects (Finney 2004). In this paper, we investigate a new approach to map-

Received 10 September 2008. Accepted 17 June 2009. Published on the NRC Research Press Web site at cjfr.nrc.ca on 9 October 2009.

K.B. Pierce, Jr.[1] **and J.L. Ohmann.**[2] Pacific Northwest Research Station, USDA Forest Service, 3200 SW Jefferson Way, Corvallis, OR 97331, USA.
M.C. Wimberly. South Dakota State University, Wecota Hall, Box 506B, Brookings, SD 57007, USA.
M.J. Gregory. Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR 97331, USA.
J.S. Fried. Pacific Northwest Research Station, USDA Forest Service, P.O. Box 3890, Portland, OR 97208, USA.

[1]Present address: Washington Department of Fish and Wildlife, 600 Capitol Way North, Olympia, WA 98501, USA.
[2]Corresponding author (e-mail: johmann@fs.fed.us).

Fig. 1. Locations of the three study areas in the western United States.



ping the regional patterns of wildland fuels, through the imputation of forest inventory plots using the gradient nearest neighbor (GNN) method (Ohmann and Gregory 2002), and we compare results with three other commonly used methods.

Wildland fuels affect the ignition, behavior, and effects of fire. The task of mapping wildland fuels consists of describing the spatial patterns of vegetation attributes that influence fire dynamics, usually partitioned into surface and canopy fuels. Canopy fuel attributes, such as canopy bulk density and vertical crown continuity, influence the likelihood of crown fire and the resulting levels of tree mortality. Surface fuels and standing dead trees (snags) affect the probability of crown scorch or fire (Scott and Reinhardt 2001). Forest structure characteristics such as stand height and canopy cover also influence fire behavior and effects by affecting wind speeds and fuel moisture. Surface fuels, such as grass, leaf litter, deadwood, and shrubs, influence fire behavior by providing the substrate that can either carry or suppress fire, depending on fuelbed structure and moisture content (Agee et al. 2002). Surface fuels are typically grouped into sets of common conditions and referred to as "fuel models" (Anderson 1982; Ottmar et al. 2007), where each model is a labeled vector of numeric attributes used to parameterize a physics-based predictive model of fire behavior attributes such as intensity and spread rate. The most widely used fuel model classification is the system of 13 "stylized" Anderson fuel models (Anderson 1982), which describe surface fuels and expected fire behavior based on broad vegetation structure, management, and total fuel load. Customized models (Reinhardt and Crookston 2003) can be created, and there also is a set based on the National Fire Danger Rating System. Although fuel models were developed to aid communication about fire behavior under different conditions of weather and slope, they have been incorporated into forest management decision processes as an indicator of fire hazard and, when mapped, as a key input to fire-growth models.

Several studies have mapped wildland fuels at the regional scale, typically relying on remotely sensed imagery (Van Wagtendonk and Root 2003; Rollins et al. 2004). However, unlike vegetation classes that can be predicted based on spectral properties viewed from above the canopy (Cohen et al. 2001), surface fuels cannot be directly viewed by aerial and space-based spectral sensors and have remained difficult to map over large areas (Keane et al. 2001). Canopy fuel variables, such as height of the canopy base and canopy bulk density, also are challenging to assess using spectral remote sensing techniques (Scott and Reinhardt 2001; Cruz et al. 2003).

Researchers have attempted to map wildland fuels with classification and regression trees (CART) (Breiman 1984; Rollins et al. 2004) and image classification (Van Wagtendonk and Root 2003). With these approaches, fuel maps can be generated by mapping vegetation and associating a fuel variable to a reference vegetation type through a look-up table or classification rules (Keane et al. 2000, 2001; Mbow et al. 2004), or by relating field-measured fuel variables directly to spatial predictor variables such as spectral, topographic, and climate indices (Rollins et al. 2004). Because the relationships with these variables differ considerably among fuel variables, map accuracies tend to be low. For example, an extensive study on the Gila National Forest that mapped 10 fuel models achieved 36% accuracy (Keane et al. 2000). A more recent study of the Kootenai River basin predicted fuel loads with 12% explained variation, and three fuel models with 80% accuracy (Rollins et al. 2004).

One generally accepted approach to regional vegetation mapping involves using statistical models to fit the relationship between dependent variables measured on field plots and independent variables from GIS data sets of climate, topography, disturbance history, and the satellite-imagery-based spectral signatures of known vegetation conditions (Franklin 1995; Guisan and Zimmerman 2000). Typically, single attributes are mapped, and common techniques include generalized linear models (Nelder and Wedderburn 1972), generalized additive models (Yee and Mitchell 1991), CART (Breiman 1984), supervised classification, and geostatistical methods (Legendre and Legendre 1998).

Nearest neighbor imputation is a recently developed approach to mapping multiple forest attributes (Moeur and Stage 1995; Tomppo et al. 1999). Nearest neighbor methods use a set of spectral and environmental characteristics to determine which field plots are most similar to a target (map) location. Nearest neighbor methods have primarily been used to impute whole stands with tree lists using photointerpreted covariates (LeMay and Temesgen 2005), or to estimate specific quantities such as forest cover (McRoberts et al. 2002) or timber volume (Franco-Lopez et al. 2001) over large areas with satellite imagery. Unlike regression-based predictions, values assigned to each map unit using single-neighbor imputation are the original values from one imputed field plot.

Table 1. Comparison of study areas from plot and spatial data.

	Biome		
	Washington, temperate steppe	Oregon, coastal	California, mediterranean
Forested area (ha)	3 178 572	2 293 458	3 190 895
Nonforest area (ha)	1 822 578	572 985	968 943
Total area (ha)	5 001 150	2 866 444	4 159 841
Chaparral area (ha)*			153 168
Elevation and climate†			
Elevation (m)	1203	299	1784
Annual precipitation (mm)	864	2054	1081
Annual temperature (°C)	5	10	8
Annual frost days	204	50	161
Plot and imagery summary			
FIA plots on nonfederal lands	468	385	200
CVS/R6 plots on national forests	1808	279	
R5 plots on national forests			1288
Plots in national parks	49		347
BLM plots on BLM lands		99	
Total plots	2325	763	1835
Model-building plots	1825	613	1435
Validation plots	500	150	400
Landsat (TM 5 and 7) imagery dates	1992, 2000	1996	1992, 2000
Plot measurement dates	1991–2000	1993–1997	1988–2000
Summary of tree data on plots			
Mean basal area that is hardwood (%)	1	19	15
Mean canopy cover (SD) (%)	54 (26)	69 (21)	47 (27)
Mean basal area of trees ≥2.54 cm DBH (m^2/ha)	24	31	29
Mean QMD‡ of conifers ≥2.54 cm DBH (cm)	21	34	29

Note: Plot data are from USDA Forest Service (FS), Pacific Northwest Research Station, Forest Inventory and Analysis (FIA); USDA FS, Region 5 (R5); USDA FS, Region 6 (R6), Current Vegetation Survey (CVS/R6); Bureau of Land Management (BLM); and research plots in North Cascades and Yosemite National Parks.

*Included in nonforest area.

†Mean values at forested plots.

‡QMD, quadratic mean diameter.

Gradient nearest neighbor (GNN) imputation (Ohmann and Gregory 2002) employs direct gradient analysis using canonical correspondence analysis (CCA) (Ter Braak 1987) to assign weights to predictor variables. For delineating and interpreting plant communities, CCA has been shown to provide a better overall view of general ecological gradients and biodiversity than multiple individual generalized linear models (GLM), because CCA explicitly models the co-occurrence of species (Guisan et al. 1999). Single-neighbor imputation used by GNN maintains the covariance structure of multiple attributes of vegetation at a target map location. This property benefits many applications in forest ecosystem management and landscape planning, which often require information about many variables simultaneously (Keane et al. 2001; Miller and Landres 2004). For example, planning of fuel treatments for reducing fire risk typically requires considering potential effects on other forest values such as wildlife habitat, which may require information on a broader suite of vegetation attributes. Nearest neighbor maps are especially useful for regional simulation modeling efforts. Nearest neighbor methods are being increasingly used in the United States (McRoberts et al. 2002; Ohmann and Gregory 2002) and globally (e.g., Tomppo et al. 2008) for forest research, mapping, and assessment.

Although GNN has been successfully applied to map forest composition and structure in coastal Oregon (Ohmann and Gregory 2002; Ohmann et al. 2007), its efficacy in other forest ecosystems is unknown. Moreover, nearest neighbor techniques have not been evaluated specifically for mapping wildland fuels. The objective of this study was to evaluate the accuracy of GNN imputation and three other commonly used methods for mapping forest structure and wildland fuels. This paper presents a comparison of these methods for three forest biomes in the western United States, using a selection of fuel variables available from regional forest inventory plots and measures of forest structure that are commonly used in forest planning and management in the region. Our analyses address those fuel attributes that are currently available from regional forest inventories, rather than providing a complete accounting of all canopy and surface fuels.

Methods

Study areas

We applied GNN imputation and other methods to a wide range of forest ecosystems in three study areas in the western United States: northeastern Washington, coastal Oregon,

Table 2. Variables summarized from plot data and used to compare the performance of modeling methods.

Variable	Description
Continuous variables	
CANCOV	Canopy cover of live trees (%)
BAA	Basal area (m^2/ha) of live trees ≥2.54 cm DBH
QMD	Quadratic mean diameter (cm) of trees ≥2.54 cm DBH
LTPH	No. of live trees ≥25 cm DBH per hectare
STPH	No. of snags ≥25 cm DBH per hectare
STNDHGT	Mean stand height (m)
HCB	Height to canopy base (m), the mean height to the base of the live crown, averaged over all trees on plot
CANMASS	Canopy fuel mass (kg/ha)
CBD	Canopy bulk density (kg/m^3)
DVPH	Volume of downed wood 12–25 cm diameter at large end (Washington and California) or 13–20 cm diameter at large end (Oregon) (m^3/ha)
Discrete variables	
FUELMOD	Anderson or custom fuel model, with Anderson (1982) number in parentheses: ShortGrass, short grass (1); TimberG, timber (grass and understory) (2); Brush, brush (5); DorBrush, dormant brush (6); CTLitter, closed timber litter (8); HWLitter, hardwood litter (9); TimberL, timber (litter and understory) (10); LightSlash, light logging slash (11); MedSlash, medium logging slash (12); Chaparral, chaparral (California only) (13)
VEGCLASS	Vegetation class, modified from Johnson and O'Neil (2001), where BAHP is the percentage of total basal area that is hardwood. Sparse (CANCOV <10); open (CANCOV 10–39); hardwood (hdw) sapling–pole (sap–pole) (CANCOV ≥40, BAHP ≥65, QMD <25); hdw small–medium–large (sm–md–lg) (CANCOV ≥ 40, BAHP ≥65, QMD ≥25); mixed (sap–pole) (CANCOV ≥40, BAHP 20–64, QMD <25); mixed (sm–md) (CANCOV ≥40, BAHP 20–64, QMD ≤50); mixed large – very large (lg–vl) (CANCOV ≥40, BAHP 20–64, QMD ≥50); conifer (con) (sap–pole) (CANCOV ≥40, BAHP <20, QMD <25); con (sm–md) (CANCOV ≥40, BAHP <20, QMD 25–50; con (lg) (CANCOV ≥40, BAHPH <20, QMD 50–75); con (vl) (CANCOV ≥40, BAHP <20, QMD ≥75)

Note: See Appendix for detailed methods used to compute canopy fuel variables and assign fuel models.

and the central Sierra Nevada in California (Fig. 1, Table 1). Of the three sites, California has the highest mean elevation and spans a wide variety of broadleaf and coniferous forest types that typically are uneven aged. Forests in Oregon tend to be even aged, contain a mixture of broadleaf and coniferous species, and attain the greatest tree size and biomass of the three study areas. Forests in Washington are more conifer dominated and have less variation in mean tree size than forests in the other regions. Climate in Oregon is more strongly maritime than that in the other regions, with much higher annual precipitation, the warmest temperatures, and the most growing degree-days.

Forest plot data

We obtained vegetation data from nearly 5000 regional forest inventory plots installed on systematic grids by several federal agencies (Table 1). Most plots consisted of fixed- and variable-radius subplots and transects covering about 1 ha each. For use in modeling and mapping, we derived several summary variables from the live trees, standing dead trees (snags), and downed wood (coarse woody debris) tallied on the plots. To compare the performance of the modeling methods, we focussed on 10 continuous and 2 categorical variables (Table 2). These live- and dead-tree summary variables were selected to represent a cross-section of measures of vegetation structure and composition that are relevant to forest, wildlife, and fuels management. Because sampling protocols varied somewhat among data sources, particularly for downed wood and fine fuels, we limited our analyses to those attributes that were consistently measured across all plots.

We computed several canopy fuel variables using methods described in the Appendix. The height and live crown ratio of individual trees were used to derive a stand-level height to canopy base (HCB) for each plot, computed as the simple mean height to the lowest branch (Cruz et al. 2003). Canopy fuel mass (CANMASS) was computed as the total mass of all conifer tree crowns on the plot, including foliage and all live and dead branches. Canopy bulk density (CBD) was calculated using allometric equations for canopy leaf mass and the length of the canopy derived from tree heights and crown ratios (Reinhardt and Crookston 2003).

We assigned stylized fuel models to plots (see Appendix) based on relationships derived for each of the study areas from regional variants of the Fire and Fuels Extension to the Forest Vegetation Simulator (FVS-FFE) (Reinhardt and Crookston 2003), from the literature (Anderson 1982; Scott and Reinhardt 2001; Reinhardt and Crookston 2003), or from experts in the region. These approaches used various measures of total surface fuel loading, canopy cover, dominant tree species, stand height, and height to canopy base. We used these methods for consistency with one of the main uses of fuel models in simulations, namely FVS-FFE. The lack of an accepted process for translating specific field measurements into fuel models is an important problem that is beyond the scope of this paper.

Spatial data

Spatial data for modeling consisted of climate model predictions, topographic indices, geographic coordinates, and Landsat Thematic Mapper (TM) images (Table 3). Data on disturbance history were from a partial coverage of recent fires (Washington), maps of forest disturbance since 1972 derived from multitemporal Landsat imagery (Oregon)

Table 3. Spatial data obtained from remote sensing, 10 m resolution digital elevation models (DEM), climate models, and other digital GIS coverages.

Variable	Description
Climate	
ANNFROST	Mean number of days per year when daily minimum temperature is <0.0 °C
ANNGDD	Total annual growing degree-days
ANNPRE	Total annual precipitation (cm)
ANNSW	Annual sum of total daily incident shortwave radiative flux (accounts for cloudiness) ($MJ \cdot m^{-2} \cdot day^{-1}$)
ANNVP	Mean annual vapor pressure (Pa)
AUGMAXT	Mean maximum August temperature (°C)
CONTPRE	Percentage of annual precipitation falling in June–August
CVPRE	Coefficient of variation of December (wettest) and July (driest) mean monthly precipitation
SMRPRE	Total summer (May–September) precipitation (cm)
STRATUS	Percentage of July hours with a marine cloud layer <1524 m and visibility <8 km (unpublished data from C. Daly)
Remote sensing	
TMx	Landsat Thematic Mapper (TM) band, where x = 1–5 and 7
R43	Ratio of TM bands 4 and 3
R54	Ratio of TM bands 5 and 4
R57	Ratio of TM bands 5 and 7
ADTMx	Additive difference of TM bands, where x = 1–5 and 7. Differences in values between pairs of neighboring cells are calculated and then summed across the 13-pixel plot footprint (a measure of image texture).
ADR43	Additive difference of R43 within the plot footprint
ADR54	Additive difference of R54 within the plot footprint
ADR57	Additive difference of R57 within the plot footprint
Topography	
ASPTR	Cosine transformation of aspect (°) (Beers et al. 1966), 0.0 (southwest) to 2.0 (northeast), from DEM
DEM	Elevation (m), from DEM
PRR	Potential relative radiation (Pierce et al. 2005), from DEM
SLPPCT	Slope (%), from DEM
TPI150	Topographic position index, calculated as the difference between a cell's elevation and the mean elevation of cells within a 150 m radius window (A. Weiss, personal communication)
Disturbance	
YSDIST	Years since disturbance, derived from Landsat time-series data (Healey et al. 2005) or GIS coverages
YSFIRE	Years since fire, derived from Landsat time-series data (Healey et al. 2005) or GIS coverages
DISTF	Forest land not recently disturbed, derived from Landsat time-series data (Healey et al. 2005)
Location	
X	Longitudinal position (m) from Universal Transverse Mercator coordinate system
Y	Latitudinal position (m) from Universal Transverse Mercator coordinate system

Note: All coverages were resampled to 30 m resolution. Climate data are from Daymet (http://www.daymet.org), except where noted, and were interpolated from 1 km resolution data.

(Healey et al. 2005), and a map of disturbances since 1950 produced by the California Department of Forestry (California), all of which were converted to years since disturbance. Landsat TM imagery was selected to best match the year of plot sampling at each site (Table 1), using scenes from mid to late summer. We used two imagery years in Washington and California to accommodate a wide range of plot measurement dates. Individual TM scenes were radiometrically normalized before creating regional mosaics. Where two imagery years were used, the two mosaics also were normalized. All spatial data were resampled to 30 m resolution.

For use in model development, the values for spatial variables at each plot location were sampled as the mean of thirteen 30 m pixels in a diamond-shaped footprint centered on each field plot, designed to match the size and configuration of the plot on the ground. Over 90% of our field plots consisted of four or five subplots covering about 1 ha. The multipixel footprint was designed to capture spatial values across a generalized area, given that plot locations are known to have error, and to describe the same physical area as the ground data. We excluded from modeling those plots that straddled distinct boundaries in land cover or forest condition or that had been disturbed between plot measurement and imagery date. Plots in naturally heterogeneous areas were retained.

Model development with GNN imputation and other methods

We compared GNN with other common modeling approaches for 12 forest structure and fuel variables (Table 2). The 10 continuous variables were modeled using univariate linear multiple regression models (LMs) and with simple kriging and universal kriging (UK). The two categorical var-

Fig. 2. Ordination diagrams from canonical correspondence analysis (CCA) used in gradient nearest neighbor mapping, showing relationships of vegetation and fuels response variables to spatial variables and ordination axes (axes 1 and 2 for Washington and Oregon; axes 2 and 3 for California). Locations of the response variables are centroids of abundances weighted by plot scores. Arrow lengths for the predictor variables show correlation strength with the ordination axes. Response variables are basal area (m^2/ha) of live conifers by size class (BAC*X.Y*), basal area (m^2/ha) of live hardwoods by size class (BAH*X.Y*), snag density (trees/ha) by size class (STPH*X.Y*), and downed wood volume (m^3/ha) by size class (DVPH*X.Y*), where *X* and *Y* denote the minimum and maximum diameters of the size class. See Table 3 for definitions of spatial variables.

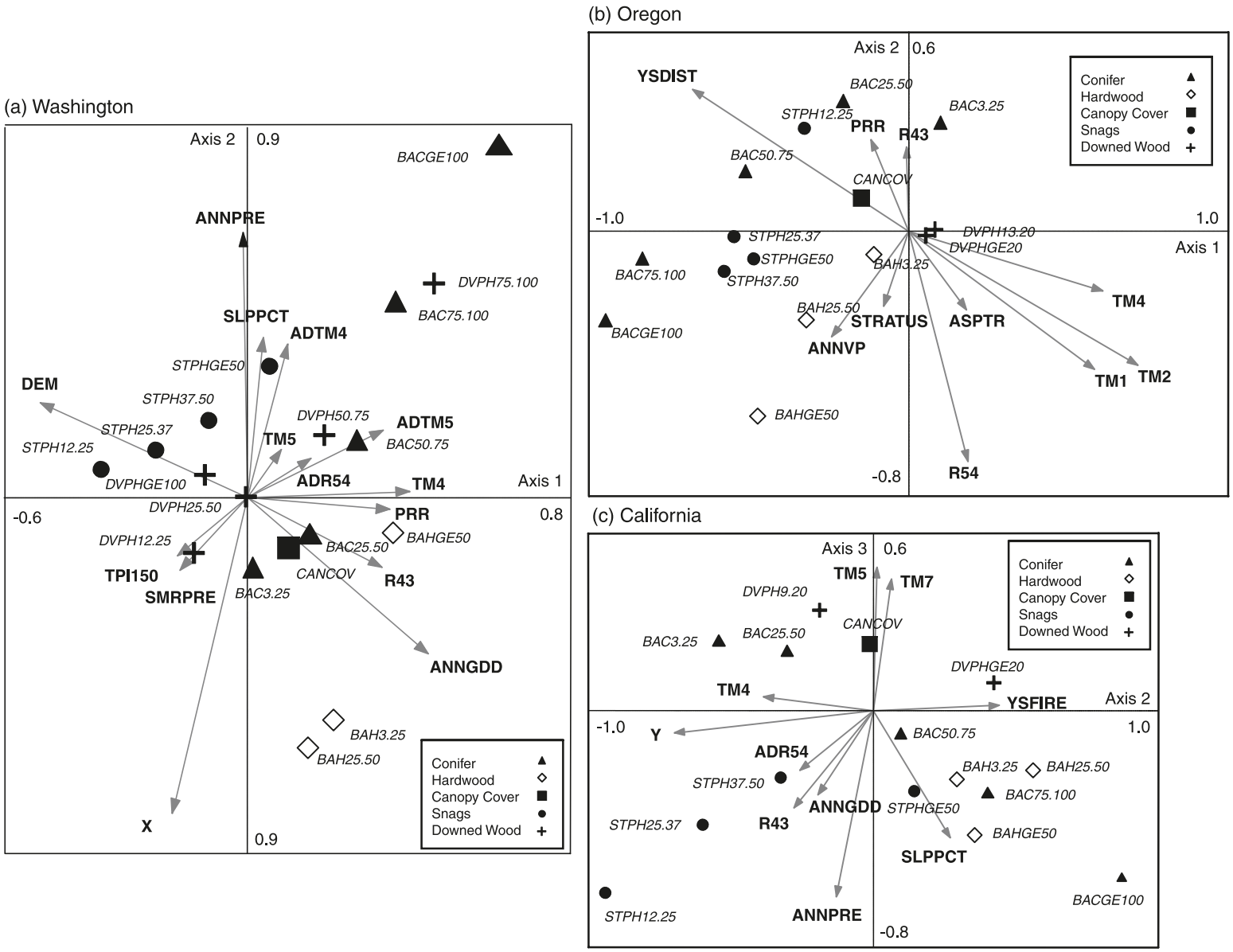


Table 4. Explanatory variables selected for the linear models in descending order of selection.

Response variable	Model
Washington	
CANCOV	–TM5 + ANNPRE + R57 + PRR + ASPTR + ADTM7 + *X*
BAA	–TM5 + ANNPRE + PRR + ASPTR – SMRPRE
QMD	–TM5 + PRR + ADTM5 + SLPPCT – *X* – DEM
LTPH	–TM5 + PRR – CONTPRE + ASPTR
STPH	–AUGMAXT – TM4 – CONTPRE – R54
STNDHGT	–TM7 + PRR + ADTM5 – DEM – TM4 + ASPTR + SLPPCT + R57
HCB	–TM7 + SMRTMP + PRR – TM4
CANMASS	–TM7 + PRR + ADTM5 – DEM – TM4 + ASPTR + SLPPCT + R57
CBD	–TM5 + ANNPRE + PRR + *X* + ASPTR + R57 + ADTM3 – ADTM4 – ANNVP – SMRPRE
DVPH	–TM3 + DEM – SLPPCT + ASPTR
Oregon	
CANCOV	–TM5 + TM4 + YSDIST
BAA	–TM5 + YSDIST + PRR – TM4 + ASPTR – ANNSW + ADTM4 + SLPPCT
QMD	–TM5 + ADR54 + YSDIST + ANNVP + R54 – TM3 + PRR + ASPTR – SMRTP
LTPH	YSDIST – TM2 + ANNVP + R54 + *Y*
STPH	–TM4 + ANNPRE – TM3 – ANNFROST
STNDHGT	–TM2 + STRATUS + DISTF – TM4 + ASPTR + PRR
HCB	–TM2 + YSDIST + ANNVP + *Y*
CANMASS	–TM5 + PRR + ADR54 – ANNSW + ASPTR
CBD	–TM5 + PRR + TM4 + TM1 + DEM + ASPTR
DVPH	TM3 + ANNFROST – ADTM2
California	
CANCOV	–TM5 + ANNPRE – DEM + ADTM5 + YSFIRE + ASPTR + *Y*
BAA	–TM5 – YSFIRE + ANNPRE – R57
QMD	–TM5 + ADTM5 + ANNPRE – TM4
LTPH	–TM5 + CVPRE + YSFIRE
STPH	–TM5 + ADR54 + SMRPRE – TM4
STNDHGT	–TM3 + CVPRE + ADTM5 + SMRPRE – TM4 – TM5
HCB	–TM3 + CVPRE + ADTM5 – TM4
CANMASS	–TM5 + SMRPRE – R57 – SLPPCT + CVPRE + ASPTR + PRR + ADTM5 + YSFIRE – TM1
CBD	–TM5 – SMRTP – SLPPCT + CVPRE + ASPTR – ADTM4 + ADTM5 + *Y* – R57 + PRR – TM1 + TM4
DVPH	–TM5 –ANNFROST – ADTM5 – DIFTMP

Note: All response variables were square-root transformed. See Table 2 for definitions of vegetation and fuel response variables and Table 3 for definitions of explanatory variables.

Table 5. Prediction accuracy for continuous vegetation and fuel variables from gradient nearest neighbor imputation (GNN), linear models (LM), kriging (K), and universal kriging (UK).

	Washington				Oregon				California			
	GNN	LM	K	UK	GNN	LM	K	UK	GNN	LM	K	UK
CANCOV	0.17	0.51	0.15	0.45	0.56	0.52	0.07	0.54	0.33	0.31	0.13	0.35
BAA	0.16	0.49	0.19	0.51	0.59	0.54	0.04	0.55	0.32	0.32	0.09	0.31
QMD	0.07	0.05	0.05	0.13	0.70	0.61	0.14	0.69	0.13	0.17	0.16	0.21
LTPH	0.29	0.37	0.12	0.34	0.57	0.34	0.02	0.34	0.29	0.24	0.07	0.28
STPH	0.05	0.15	0.02	0.16	0.31	0.40	0.22	0.40	0.09	0.11	0.10	0.14
STNDHGT	0.07	0.27	0.11	0.29	0.69	0.74	0.06	0.77	0.22	0.30	0.25	0.34
HCB	0.07	0.16	0.07	0.19	0.41	0.58	0.08	0.49	0.08	0.11	0.04	0.09
CANMASS	0.15	0.27	0.19	0.47	0.54	0.62	0.01	0.64	0.35	0.48	0.24	0.36
CBD	0.16	0.43	0.14	0.45	0.38	0.61	0.03	0.63	0.37	0.47	0.13	0.47
DVPH	0.07	0.13	0.13	0.20	0.16	0.10	0.05	0.07	0.01	0.04	0.02	0.04

Note: The values represent the squared correlation between the validation data sets and the model predictions. See Table 2 for variable definitions.

iables were modeled using CART analysis (Breiman 1984). We assessed accuracy for all methods by comparing model-predicted values with those observed on the ground for the 20% sample of validation plots (Table 1). A mask of nonforest areas was applied to all maps using 1992 National Land Cover Data. The models apply to forest land only because the forest inventories do not collect data on nonforest land.

GNN imputation (Ohmann and Gregory 2002) consists of

Table 6. Proportion of validation plots correctly classified and kappa statistics (in parentheses) for gradient nearest neighbor (GNN) and classification and regression tree (CART) models for categorical variables.

	Washington		Oregon		California	
	GNN	CART	GNN	CART	GNN	CART
Vegetation class	0.38 (0.23)	0.19 (0.10)	0.49 (0.59)	0.42 (0.51)	0.36 (0.39)	0.36 (0.40)
Fuel model	0.35 (0.10)	0.15 (0.13)	0.44 (0.22)	0.41 (0.20)	0.43 (0.23)	0.43 (0.24)

Note: See Table 2 for variable definitions.

direct gradient analysis (CCA) (Ter Braak 1987) followed by single-neighbor imputation to assign the closest sample plot in gradient space to each unsampled target (map) location. Because our goal was to map forest structure and fuels, for CCA we specified response variables to represent variation in the size, abundance, and composition of live trees, snags, and downed wood (Fig. 2). All response variables were square-root transformed to dampen the effects of dominant species. We used a stepwise procedure (CANOCO version 4.5) to select a parsimonious subset of spatial predictors (Table 3) and reject highly collinear predictors (variance inflation factors > 50) (Ter Braak 1987), to provide weights for those predictors that best discriminated differences among plots in terms of the response variables. CCA produces a set of orthogonal axes that are linear combinations of all predictor variables. We used the first eight ordination axes, weighted by their eigenvalues, to calculate multivariate distances in gradient space, which were used to rank the sample plots in terms of similarity to each target location (30 m square pixel).

For the 10 continuous variables modeled with LMs, we used multiple regressions with forward selection to examine the correlation between predictors and the values from model-building plots. Variables were added to the models until no predictor had a correlation greater than 0.10 with the model residuals and was considered significant after Bonferroni correction (Sokal and Rohlf 1995). Significance values were derived from the type III ANOVA t value and standard error. The subsequent model was then used to predict the values at the validation locations.

In the kriging analysis, we estimated values for the validation plots using the autocorrelation function of the model variable (Haining 1990). Variograms were fit on X and Y location for each of the 10 continuous variables for the modeling-building plots using the GSTAT package for R with the spherical autocorrelation function. UK was performed for each of the variables by adding the same terms derived from the LMs to the model statement for the variograms. This method effectively adds additional predictive power to the LMs by adding a term for the spatial autocorrelation in the residuals from the LM.

The two discrete variables, vegetation class (VEGCLASS) and stylized fuel models (FUELMOD), were modeled using GNN and CART. For GNN, the predicted map values for VEGCLASS and FUELMOD at validation-plot locations were those of the imputed nearest neighbor plots. CART models were developed using RPART in R version 2.4. CART is a variable selection technique that makes successive splits on the spatial variables that best partition the plots into groups (Venables and Ripley 2002). Values for the spatial variables assigned to model-building plots were used as predictors, and plot classifications of VEGCLASS and FUELMOD were the response variables. All spatial variables were included in the CART models. We also developed a map of VEGCLASS from the LMs, by combining spatial predictions for individual continuous variables used in the classification (CANCOV, QMD, and BAHP (Table 2)).

Accuracy assessment and model evaluation

To assess accuracy at the local (plot) scale, we compared predicted values from the univariate methods and from GNN with observed values for the 20% sample of validation plots. For the 10 continuous variables we calculated squared correlations between model-predicted and observed values. For the two discrete variables we compared the class assignments for VEGCLASS and FUELMOD with the actual assignments using confusion (error) matrices and calculated simple prediction percentages, producer's and user's accuracies (Congalton and Green 1999), and kappa statistics (Wilkie and Finn 1996). Producer's accuracy is the proportion of plots observed for a certain class that were predicted as that class. User's accuracy is the proportion of plots predicted as a certain class that were observed as that class.

To assess accuracy at the regional scale for discrete variables, we evaluated the GNN, LM, and CART models for VEGCLASS, and GNN and CART for FUELMOD. Because the classification rules used for developing FUELMOD maps used both continuous and discrete variables, no comparable map could be derived with LMs. For each variable, we compared the distribution of forest area among discrete classes from the mapped predictions to design-based estimates from the plot sample. The design-based estimates used plot area expansion factors provided by the inventory programs, which account for differences in inventory design.

Results

Ecological and spectral correlates in the models

Over the 30 linear models (10 variables in each of the three study regions), a Landsat TM variable was selected first by the forward stepwise algorithm in 28 cases (Table 4). Similarly, a TM variable was selected first in the three CART models for vegetation class for each study region (data not shown). Output from CART models is in the form of complex splitting trees, in which variables enter at different levels in the tree and at different splitting values, and we do not discuss these results further.

In Washington (Fig. 2*a*), dominant ecological gradients in forest structure and composition were most strongly associated with environmental factors and less so with disturbance history (Landsat and disturbance variables). Axis 1 was a gradient in elevation and temperature, with larger conifers and snags at lower elevations and warmer temperatures, and

Fig. 3. Confusion matrices represented as strip plots with user's accuracy depicted as open symbols and producer's accuracy depicted as solid symbols. The triangles represent the gradient nearest neighbor model and the squares represent linear models for (*a*) vegetation class and (*b*) fuel models. See Table 2 for definitions of vegetation classes and fuel models. Under each class label, numbers of validation plots are listed for Washington (WA), Oregon (OR), and California (CA), respectively.

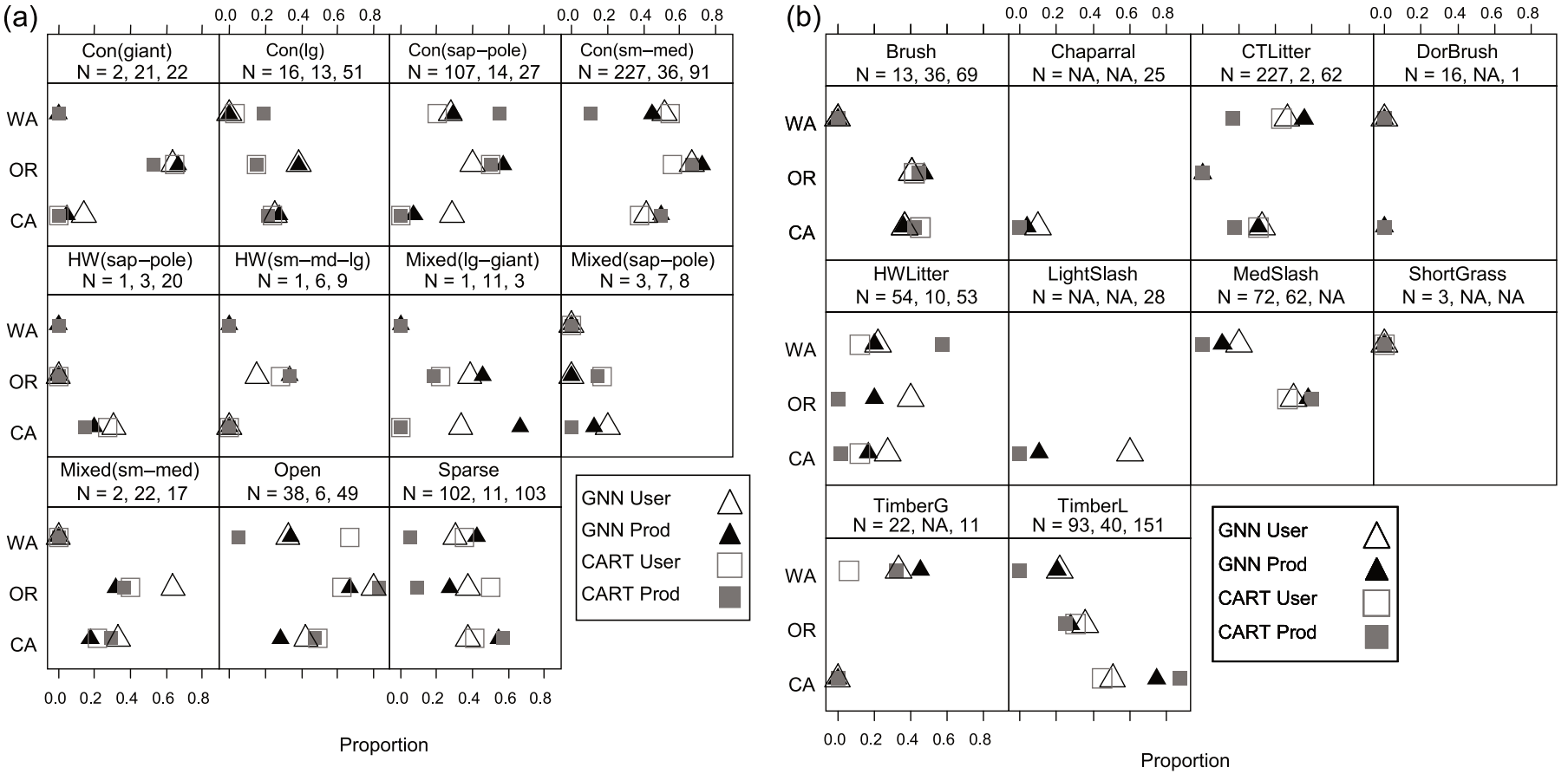


Fig. 4. Distribution of forest area among vegetation classes from plot-based estimates and from spatial predictions from gradient nearest neighbor (GNN), classification and regression trees (CART), and linear models (LMs). See Table 2 for definitions of vegetation classes.

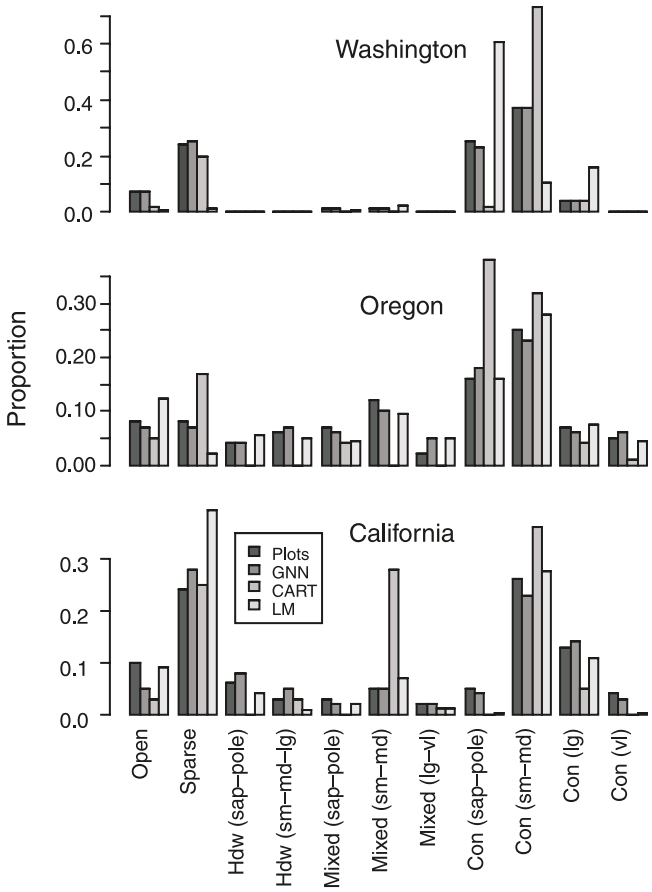
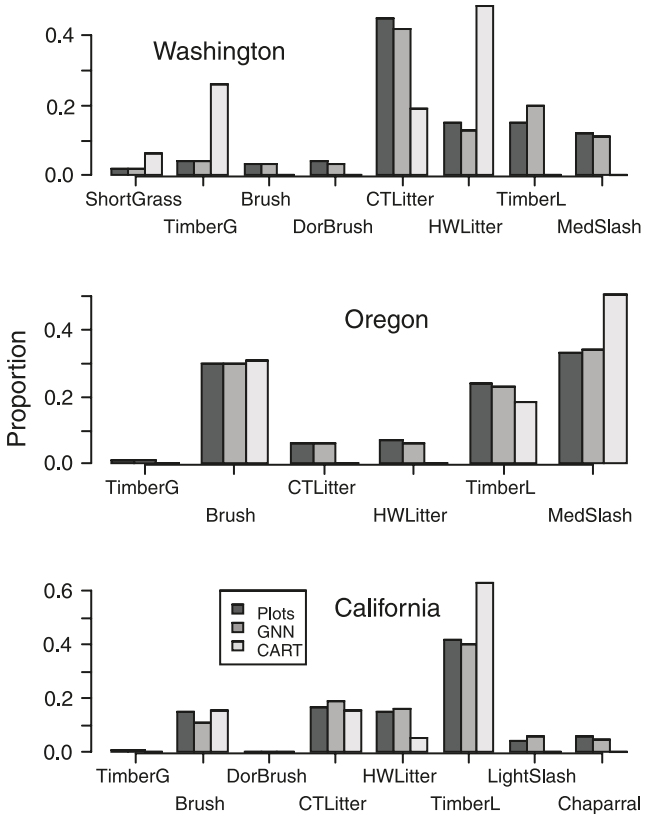


Fig. 5. Distribution of forest area among fuel models from plot-based estimates and from spatial predictions from gradient nearest neighbor (GNN) and classification and regression trees (CART). See Table 2 for definitions of fuel models.



smaller trees at cooler mountainous locations. Axis 2 was a longitudinal gradient in moisture, from areas of greater precipitation and steeper slopes in the west to the drier climate to the east. Gradients in conifer, snag, and downed wood size paralleled one another, responding similarly to environmental gradients, except for the largest size class of downed wood, which was associated with stands of smaller trees. Hardwoods were concentrated in warm low-elevation areas with longer growing seasons.

In Oregon (Fig. 2*b*), dominant gradients in forest vegetation were most strongly associated with disturbance history (years since disturbance and Landsat variables) and much less so with environmental gradients. Axes 1 and 2 captured a gradient from younger forests of smaller trees, which were more likely to contain hardwoods, to older conifer-dominated forests of larger trees. Sizes of conifers, hardwoods, and snags responded similarly to disturbance and environmental gradients. The size classes of downed wood were centrally located in the ordination, indicating a lack of correlation with our spatial variables and with live trees and snags.

The California model (Fig. 2*c*) was similar to Washington's model, sharing seven of the same or similar predictors. Axis 1 (not shown) was a gradient in composition, separating hardwood- and conifer-dominated forests. Hardwoods were predominant at lower elevations with more growing degree-days and on steeper slopes of canyons and riparian corridors. Axes 2 and 3, which were more relevant to fuels, are shown in Fig. 2*c*. These axes captured a gradient in tree size and forest development since disturbance. Smaller trees occurred in areas of recent disturbance, more often in the northern part of the study area, and larger trees were found in undisturbed areas to the south. Sizes of conifers, snags, and downed wood followed similar patterns. Snags were more abundant at northerly locations and in areas of recent fire, while hardwoods, downed wood, and conifers were more associated with southerly locations and less recent fire activity.

Comparisons of local-scale accuracy among modeling methods

Local-scale (plot) prediction accuracy for continuous variables by kriging was consistently poor (Table 5). Squared correlations obtained with kriging were worse than those obtained for all other methods, variables, and study areas, with the exception of a few variables in Washington and California where GNN accuracy was particularly poor.

The relative performance of the other three methods (GNN, LM, and UK) varied with region, variable, and evaluation method (Tables 5 and 6). In Washington, UK and LM performed better than GNN for almost all continuous variables (Table 5). In Oregon, GNN and UK were generally better than LM for forest structure variables, whereas LM and UK were better for canopy fuel variables. In California, accuracy among the three methods was more similar than it was in the other regions.

In general, UK only marginally changed the outcomes for LMs with the exception of canopy fuel mass, where UK was better than LM in Washington but worse than LM in California (Table 5). For some variables and some regions, none of the methods yielded good results. Correlations for downed wood never exceeded 0.20, and correlations for snag density were similarly low in Washington and California (Table 5). Correlations for HCB were very low in Washington and California.

Local-scale accuracy from GNN and CART for the discrete variables vegetation class and fuel model are shown in Table 6. In Washington and Oregon, GNN predictions were more accurate overall than CART predictions, with higher correct proportions and kappa statistics except for the kappa for fuel model in Washington, which was poorly predicted with both methods. In California, accuracy was nearly the same for both methods.

GNN predictions were more accurate than CART predictions for 13 of the 25 combinations of vegetation class and study region that contained plot observations (Fig. 3*a*). Producer's accuracies for GNN were better in 13 cases, and user's accuracies were better in 11 cases with one tie. For the 15 of 21 combinations of fuel model and study region that contained observations, GNN was more accurate in 11 cases and CART was better in four cases (Fig. 3*b*).

Regional-scale prediction accuracy for discrete variables from GNN, CART, and LMs

Overall, GNN outperformed both CART and LMs in terms of regional distributions of vegetation classes (Fig. 4) and fuel models (Fig. 5). The GNN predictions were closer to the plot-based estimates than were the CART predictions for 23 of 29 combinations of region and vegetation class, and closer than LM predictions for 25 of 29 combinations. The mean difference in area proportions between plot- and GNN-based estimates for vegetation classes across study sites was 0.01, compared with 0.05 for CART and 0.06 for LMs. The CART area distributions had major differences with plot-based estimates for size classes in all three study regions, tending to under- or over-predict the smallest and largest size classes. Furthermore, CART did not accurately predict distributions by cover classes in Washington nor by composition classes in Oregon and California. The LMs represented area distributions among vegetation classes reasonably well in Oregon, but they had major problems in Washington and California.

The GNN predictions more closely approximated the plot-based estimates than did the CART predictions for all but one of 22 combinations of region and fuel model (Fig. 5). (LMs could not be used to predict categorical fuel models, so are not compared.) Differences in area proportions between plot- and model-based estimates for fuel models averaged 0.01 for GNN and 0.09 for CART. As with vegetation class, differences between CART- and plot-based estimates of fuel models were most notable in Washington.

Discussion

Effects of forest ecological distributions and dynamics on model results

Among the three sites, prediction accuracies for canopy fuel and forest structure variables were best overall in Oregon, lowest in Washington, and intermediate in California (Table 5). All response variables except downed wood were best predicted in Oregon; this result is not surprising because the response variables have strong local correlations with overstory canopy characteristics that are directly measured by satellite imagery, and because they often vary regionally along climate gradients. The even-aged nature of forests in Oregon probably explains much of these regional differences. Many forest structural characteristics and associated fuels variables follow predictable changes with stand development, and previous research has demonstrated that Landsat TM imagery is effective at classifying forests along the successional sequence (Cohen et al. 2001; Ohmann and Gregory 2002). In contrast, more stands in Washington and California are uneven aged and more variable in tree diameter and height, and spatial variability in stand characteristics is difficult to distinguish at the 30 m resolution of the Landsat TM imagery. In addition, Oregon forests are largely dominated by a single conifer species, *Pseudotsuga menziesii* (Mirb.) Franco, and a single hardwood species, *Alnus rubra* Bong. Forest plots in Washington and California were more species rich, further complicating relationships of forest structure with environmental gradients and remotely sensed measurements of the canopy.

The deadwood variables were most weakly correlated with available spatial data and thus hardest to predict. The amount and characteristics of deadwood at a given site reflect complex interactions among numerous ecological processes, including growth and mortality, competition, disturbance, and fragmentation and decay (Harmon et al. 1986; Spies et al. 1988). Snags and downed wood are particularly difficult to predict at the stand level because they are hidden beneath the forest canopy, and because patterns are more influenced by disturbance history than by environmental variability. Downed wood was poorly predicted by all methods in all study areas. Prediction accuracy for snag density was highest in Oregon, where stands are predominantly even aged, and the distribution of deadwood often follows predictable patterns with time since fire or clear-cutting (Spies et al. 1988). In contrast, forests in Washington and California have experienced a greater amount of patchy low- to moderate-severity disturbances such as surface fires, insect outbreaks, and selective cutting. Spatial patterns of downed wood therefore are decoupled from overstory forest structure, and prediction of deadwood spatial patterns is nearly impossible in the absence of detailed spatial data on historical disturbances.

Imputation versus other methods: trade-offs between local and regional accuracy

For maps of continuous measures of forest structure and canopy fuels assessed at the local (plot) scale, the GNN method performed the best only for forest structure variables in the Oregon study area (Table 5). The LM and UK methods achieved better local accuracies for canopy fuels in all regions and for most forest structure variables in Washington and California, although accuracy in California was fairly similar across methods for all continuous variables. In terms of local map accuracy for discrete vegetation classes and fuel models, GNN generally outperformed CART in Or-

egon and Washington, but the two methods performed similarly in California (Table 6). These results suggest that GNN can be useful when the aim is to map an array of forest attributes, particularly when the local covariance of multiple attributes is important. However, the efficacy of GNN relative to that of other methods for mapping single vegetation or fuel attributes — when assessed at the local scale — will vary with attribute and location. GNN prediction accuracy for total basal area was comparable to results from another study that evaluated single-neighbor imputation (Franco-Lopez et al. 2001), in which the square root of the mean squared model residual (RMSE) was 60% of the mean basal area. Expressed in the same terms, GNN RMSEs for total basal area in Washington, Oregon, and California were 62%, 45%, and 77%, respectively.

At the regional scale, GNN strongly outperformed CART and LMs (Figs. 4 and 5). One reason for the superior performance of GNN at broad spatial extents is that GNN relies on the plot data values used to generate the design-based estimate against which the maps were evaluated. Differences between GNN and design-based distributions occur only because the weights applied to plots differ. In the area estimates for GNN, plot values for variables were weighted by the number of pixels in the landscape to which they were imputed. Plot weights for the design-based estimates were calculated from the ratio of the sampling stratum area to the number of plots within that stratum under standard inventory design-based estimation.

More importantly, the better performance of GNN over the other methods at the regional level can be explained by fundamental differences in modeling approaches. Single-plot imputation, which assigns a whole plot and its associated observed values to a map unit, differs greatly from using individual regression models to independently predict multiple attributes. Maps based on LMs tend to be more regionally homogenous because they rely on modeling departure from the mean. In the absence of strong relationships, these maps predict a range of values that is narrower than the range of the observed data because of the well-known phenomenon of "regression to the mean" (Gelman and Hill 2007). CART also tends to be regressive and therefore reduce variability in the predictions. In contrast, single-plot imputation maps preserve the range of variability observed in different environmental settings. For instance, if surveyed plots on northeast-facing slopes at 300 m elevation with moderate precipitation have high structural variability, imputed values for similar locations will maintain the same range of variability. Spatial patterns in the resulting imputation maps tend to exhibit much greater fine-scale variability, in contrast to those based on LMs or CART, which predict a more narrow range of mean values. This limited variability in predicted values from LMs may result in greater map accuracy at the local (plot) level but at the cost of limiting the range of variability across the broader region (Moeur and Stage 1995; Ohmann and Gregory 2002). The failure of LMs to represent regional distributions of vegetation classes and fuel models was striking (Figs. 4 and 5). Several classes were severely overpredicted or sometimes even entirely missing in the predictions from LMs and CART.

A common criticism of LMs for mapping spatial characteristics is their failure to address spatial autocorrelation. We included kriging and UK in our suite of analyses to see whether accounting for spatial autocorrelation would improve our estimates. However, kriging performed poorer than both GNN and LMs, and kriging did not substantially improve accuracy when added to LMs as UK. These results indicate relatively weak spatial autocorrelation once relationships in the linear predictions were taken into account, and that spatial autocorrelation at the scale at which regional sample-based inventories are conducted is minimal. Unmeasured local spatial structure is the product of spatial processes such as seed dispersal and disturbance, local phenomena that are not captured by grids of plots spaced >5 km apart. Regional spatial autocorrelation, at the scale of our plot spacing, is driven by climate and topography, which were already included as spatial predictors in the models.

Benefits and limitations of imputation fuel maps for forest management

Single-neighbor imputation offers several benefits for mapping fuels and vegetation. The full range of variation is represented across a region, and the covariance of numerous plot attributes is maintained within each map unit (Moeur and Stage 1995; Ohmann and Gregory 2002). Because the full suite of plot attributes is retained, including tree-level tallies, new classifications or descriptor variables can be calculated and mapped after-the-fact to support new management or research needs, without developing new models. This feature may prove useful as new stylized fuel models are developed to complement the Anderson models (e.g., Scott and Burgan 2005) or as part of other systems (Ottmar et al. 2007) or as new definitions of key fuel variables are developed. GNN and other nearest neighbors approaches are easily transferred to other parts of the United States or the world where regionally consistent inventory data are available.

However, single-neighbor imputation maps have several limitations. Local-scale map accuracy for individual attributes is often lower than that obtained with alternative methods. Predictions are subject to natural variability within an environmental or spectral envelope, i.e., within gradient space. In addition, imputation methods cannot predict outside of the range of the data set, since map values are constrained to values observed on the plots. Accordingly, the full range of natural variability across a region needs to be sampled for imputation to work effectively.

Despite the often poor accuracy at a local scale, the GNN method of distributing known plot values to ecologically similar locations captures the range of variability in a region. GNN maps provide information for planners and managers on a wide range of forest attributes that are relevant to timber management, wildlife habitat assessment, and fuel management. Feedback from land managers indicates that GNN maps are most useful at intermediate scales (watersheds to ecoregions), where they provide a general picture of landscape composition and heterogeneity. Unfortunately, we lack appropriate ground-truth data (i.e., intensively sampled watersheds or mapped stands with tree-level data) for assessing accuracy at these intermediate scales. In general, we think GNN maps are appropriately used in landscape analysis to support strategic and planning applications, but local-scale accuracy may be insufficient

for prescribing treatments for specific sites. Nevertheless, GNN maps may provide guidance about which areas could most benefit from more exhaustive sampling to plan management actions. As better spatial and field data sets become available, they can be incorporated into GNN models to improve map accuracy.

Limitations to regional wildland fuel mapping in general

There are several limitations to mapping wildland fuels and forest structure at regional scales, most of which have been well covered by Keane et al. (2001). Unfortunately, most of the pitfalls they highlighted in 2001 still remain. Most importantly, many vegetation attributes (e.g., coarse downed wood, height to canopy base, understory structure, snag density) are not directly sensed by remote sensing methods (although this may change with wider availability of lidar data) and are poorly related to climate variables. Patterns may be more closely related to past local disturbance or accidents of history that are challenging to capture in regional spatial data for predictive modeling. This limitation is particularly true for below-canopy disturbances such as fuel treatments and other forest management activities. Regionally consistent maps of past fire, harvest, insect and disease, and other disturbances, as well as within-patch variability in disturbance severity, are lacking in most regions, but their availability is increasing, which offers some promise for regional mapping efforts.

Another key limitation for regional fuel mapping is the quality and consistency of field measurements on regional inventory plots. The lack of field data for nonforested areas is problematic because the interspersion of land cover types can be critical to assessing fire hazard and predicting fire behavior. Our models applied to forest land only, and we relied on ancillary regional and national land cover data to map areas of nonforest. The occurrence of errors in the separation of major land cover classes has been the most common criticism of our maps by users, particularly the separation of recently disturbed forest from areas of shrub and herbaceous cover with no forest potential.

More research is needed on the implications of sampling and scaling issues for regional fuel mapping. Our models and accuracy assessments rest on the assumption that measured quantities of canopy and surface fuels and large deadwood on plots are representative of actual conditions and are measured without error. Yet for attributes such as downed wood, transect length on most plots is inadequate to provide precise estimates. Average transect length on the Oregon and Washington plots was about 75 m, whereas the current Fire Effects Monitoring and Inventory protocol suggests a minimum of three 75 m transects, and Brown (1974) suggested fifteen to twenty 35 to 50 ft (1 ft = 0.3048 m) transects. Fine fuels data were collected inconsistently across our plot data sets, but even if available, they would probably suffer similar limitations for regional fuel modeling. Other fuel variables, such as height to canopy base, suffer from scale and measurement challenges that may need to be addressed to improve their usefulness to other applications.

Some measures of surface and canopy fuels important to fire behavior may vary at a finer spatial resolution than the size of inventory plots such as those used in this study. Ultimately, it has not yet been demonstrated that fuel inputs needed for current models of fire behavior can be measured in the field at the appropriate scale and with acceptable consistency or accuracy. Surface fuel models are a representation of expected fire behavior and their selection requires expert knowledge and interpretation. Studies have shown that fire behavior experts often disagree on fuel model assignments in the field. Regional fuel mapping will not be fully successful until fire modelers can parameterize their models with variables that can be reliably assessed on the landscape. A set of fuel models more tightly coupled to field measurements would improve the state of wildland fuel inventories and management.

Acknowledgements

We thank the Joint Fire Science Program and the USDA Forest Service Western Wildlands Environmental Threat Assessment Center for supporting this research, and the Pacific Northwest Forest Inventory and Analysis Program, Regions 5 and 6, and the Bureau of Land Management for providing inventory plot data. Jan van Wagtendonk and Dave Peterson generously provided plot data for national parks. We thank Andrew Weiss for his geospatial contributions. We also thank Rebecca Kennedy, Jonathon Thompson, and two anonymous reviewers for their constructive reviews of earlier versions of this manuscript.

References

Agee, J.K., Wright, C.S., Williamson, N., and Huff, M.H. 2002. Foliar moisture content of Pacific Northwest vegetation and its relation to wildland fire behavior. For. Ecol. Manage. **167**(1–3): 57–66. doi:10.1016/S0378-1127(01)00690-9.

Anderson, H.E. 1982. Aids to determining fuel models for estimating fire behavior. USDA For. Serv. Gen. Tech. Rep. GTR-INT-122.

Beers, T.W., Dress, P.E., and Wensel, L.C. 1966. Aspect transformation in site productivity research. J. For. **64**: 691–692.

Breiman, L. 1984. Classification and regression trees. Wadsworth International Group, Belmont, Calif.

Brown, J.K. 1974. Handbook for inventorying downed woody material. USDA For. Serv. Gen. Tech. Rep. GTR-INT-16.

Cohen, W.B., Maiersperger, T.K., Spies, T.A., and Oetter, D.R. 2001. Modelling forest cover attributes as continuous variables in a regional context with Thematic Mapper data. Int. J. Remote Sens. **22**(12): 2279–2310. doi:10.1080/014311601300229827.

Congalton, R.G., and Green, K. 1999. Assessing the accuracy of remotely sensed data: principles and practices. Lewis Publications, Boca Raton, Fla.

Cruz, M.G., Alexander, M.E., and Wakimoto, R.H. 2003. Assessing canopy fuel stratum characteristics in crown fire prone fuel types of western North America. Int. J. Wildland Fire, **12**(1): 39–50. doi:10.1071/WF02024.

Finney, M.A. 2004. FARSITE: fire area simulation model development and evaluation. USDA For. Serv. Res. Pap. RMRS-4.

Franco-Lopez, H., Ek, A.R., and 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**(3): 251–274. doi:10.1016/S0034-4257(01)00209-7.

Franklin, J. 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. Prog. Phys. Geogr. **19**(4): 474–499. doi:10.1177/030913339501900403.

Gelman, A., and Hill, J. 2007. Data analysis using regression and

multilevel/hierarchical models. Cambridge University Press, New York.

Guisan, A., and Zimmerman, N.E. 2000. Predictive habitat distribution models in ecology. Ecol. Model. **135**(2–3): 147–186. doi:10.1016/S0304-3800(00)00354-9.

Guisan, A., Weiss, S.B., and Weiss, A.D. 1999. GLM versus CCA spatial modeling of plant species distribution. Plant Ecol. **143**(1): 107–122. doi:10.1023/A:1009841519580.

Haining, R.P. 1990. Spatial data analysis in the social and environmental sciences. Cambridge University Press, Cambridge, UK; New York.

Harmon, M.E., Franklin, J.F., Swanson, F.J., Sollins, P., Gregory, S.V., Lattin, J.D., Anderson, N.H., Cline, S.P., Aumen, N.G., Sedell, J.R., Lienkaemper, G.W., Cromack, K., Jr., and Cummins, K.W. 1986. Ecology of coarse woody debris in temperate ecosystems. Adv. Ecol. Res, **15**: 133–302. doi:10.1016/S0065-2504(08)60121-X.

Healey, S.P., Cohen, W.B., Zhiqiang, Y., and Krankina, O.N. 2005. Comparison of Tasseled Cap-based Landsat data structures for use in forest disturbance detection. Remote Sens. Environ. **97**(3): 301–310. doi:10.1016/j.rse.2005.05.009.

Johnson, D.H., and O'Neil, T.A. 2001. Wildlife-habitat relationships in Oregon and Washington. 1st ed. Oregon State University Press, Corvallis, Ore.

Keane, R.E., Mincemoyer, S.A., Schmidt, K.M., Long, D.G., and Garner, J.L. 2000. Mapping vegetation and fuels for fire management on the Gila National Forest Complex, New Mexico. USDA For. Serv. Gen. Tech. Rep. RMRS-GTR-46-CD.

Keane, R.E., Burgan, R., and Van Wagtendonk, J. 2001. Mapping wildland fuels for fire management across multiple scales: integrating remote sensing, GIS, and biophysical modeling. Int. J. Wildland Fire, **10**(4): 301–319. doi:10.1071/WF01028.

Legendre, P., and Legendre, L. 1998. Numerical ecology. Elsevier, New York.

LeMay, V., and Temesgen, H. 2005. Comparison of nearest neighbor methods for estimating basal area and stems per hectare using aerial auxiliary variables. For. Sci. **51**: 109–119.

Mbow, C., Goita, K., and Benie, G.B. 2004. Spectral indices and fire behavior simulation for fire risk assessment in savanna ecosystems. Remote Sens. Environ. **91**(1): 1–13. doi:10.1016/j.rse.2003.10.019.

McRoberts, R.E., Nelson, M.D., and Wendt, D.G. 2002. Stratified estimation of forest area using satellite imagery, inventory data, and the k-Nearest Neighbors technique. Remote Sens. Environ. **82**(2-3): 457–468. doi:10.1016/S0034-4257(02)00064-0.

Miller, C., and Landres, P. 2004. Exploring information needs for wildland fire and fuels management. USDA For. Serv. Gen. Tech. Rep. RMRS-GTR-127.

Moeur, M., and Stage, A.R. 1995. Most Similar Neighbor: an improved sampling inference procedure for natural resource planning. For. Sci. **41**: 337–359.

Nelder, J.A., and Wedderburn, R.W.M. 1972. Generalized linear models. J. R. Stat. Soc. [Ser A], **135**(3): 370–384. doi:10.2307/2344614.

Ohmann, J.L., and Gregory, M.J. 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest-neighbor imputation in coastal Oregon, USA. Can. J. For. Res. **32**(4): 725–741. doi:10.1139/x02-011.

Ohmann, J.L., Gregory, M.J., and Spies, T.A. 2007. Influence of environment, disturbance, and ownership on forest vegetation of coastal Oregon. Ecol. Appl. **17**(1): 18–33. doi:10.1890/1051-0761(2007)017[0018:IOEDAO]2.0.CO;2. PMID:17479832.

Ottmar, R.D., Sandberg, D.V., Riccardi, C.L., and Prichard, S.J. 2007. An overview of the fuel characteristic classification system — quantifying, classifying, and creating fuelbeds for resource planning. Can. J. For. Res. **37**(12): 2383–2393. doi:10.1139/X07-077.

Pierce, K.B., Jr., Lookingbill, T., and Urban, D. 2005. A simple method for estimating potential relative radiation (PRR) for landscape-scale vegetation analysis. Landsc. Ecol. **20**(2): 137–147. doi:10.1007/s10980-004-1296-6.

Reinhardt, E.D., and Crookston, N.L. 2003. The fire and fuels extension to the forest vegetation simulator. USDA For. Serv. Gen. Tech. Rep. RMRS-GTR-116.

Rollins, M.G., Keane, R.E., and Parsons, R.A. 2004. Mapping fuels and fire regimes using remote sensing, ecosystem simulation, and gradient modeling. Ecol. Appl. **14**(1): 75–95. doi:10.1890/02-5145.

Rothermel, R.C. 1972. A mathematical model for predicting fire spread in wildland fuels. USDA For. Serv. Res. Pap. INT-RP-438.

Scott, J.H., and Burgan, R.E. 2005. Standard fire behavior fuel models: a comprehensive set for use with Rothermel's surface fire spread model. USDA For. Serv. Gen. Tech. Rep. RMRS-GTR-153.

Scott, J.H., and Reinhardt, E.D. 2001. Assessing crown fire potential by linking models of surface and crown fire behavior. USDA For. Serv. Res. Pap. RMRS-RP-29.

Sokal, R.R., and Rohlf, F.J. 1995. Biometry: the principles and practice of statistics in biological research. W.H. Freeman, New York.

Spies, T.A., Franklin, J.F., and Thomas, T.B. 1988. Coarse woody debris in Douglas-fir forests of western Oregon and Washington. Ecology, **69**(6): 1689–1702. doi:10.2307/1941147.

Ter Braak, C.J. 1987. The analysis of vegetation–environment relationships by canonical correspondence analysis. Vegetatio, **69**(1–3): 69–77. doi:10.1007/BF00038688.

Tomppo, E., Goulding, C., and Katila, M. 1999. Adapting Finnish multi-source forest inventory techniques to the New Zealand preharvest inventory. Scand. J. For. Res. **14**: 182–192. doi:10.1080/02827589950152917.

Tomppo, E., Olsson, H., Stahl, G., Nilsson, M., Hagner, O., and Katila, M. 2008. Combining national forest inventory field plots and remote sensing data for forest databases. Remote Sens. Environ. **112**(5): 1982–1999. doi:10.1016/j.rse.2007.03.032.

Van Wagtendonk, J.W., and Root, R.R. 2003. The use of multitemporal Landsat normalized difference vegetation index (NDVI) data for mapping fuel models in Yosemite National Park, USA. Int. J. Remote Sens. **24**(8): 1639–1651. doi:10.1080/01431160210144679.

Venables, W.N., and Ripley, B.D. 2002. Modern applied statistics with S. 4th ed. Springer, New York.

Wilkie, D.S., and Finn, J.T. 1996. Remote sensing imagery for natural resources monitoring: a guide for first-time users. Columbia University Press, New York.

Yee, T.W., and Mitchell, N.D. 1991. Generalized additive models in plant ecology. J. Veg. Sci. **2**(5): 587–602. doi:10.2307/3236170.

Appendix A. Methods for calculating canopy variables and assigning stylized fuel models for field plots

This Appendix outlines procedures used to calculate canopy fuel variables, canopy structure variables, and fuel models for forest inventory plots. Canopy variables and fuel models are not measured directly on plots. Rather, these at-

tributes are modeled at the individual-tree level using allometric equations as a function of tree species, diameter at breast height (DBH), height, and live crown ratio, which are recorded in the field. The tree-level measures are then expanded to per-unit-area values at the plot level based on subplot size.

Canopy fuel and canopy structure variables

Crown variables calculated at the individual-tree level

Forest inventories do not directly measure canopy fuels. Instead, we modeled the masses of crown components (foliage, live branches, and dead branches) at the individual-tree level as a function of field-recorded attributes of tree species, DBH, total height, and live crown ratio, and then expanded the tree-level attributes to the plot level. We used published equations (Brown 1978; Snell and Brown 1980; Snell and Anholt 1981; Reinhardt and Crookston 2003) to predict the total live and dead crown masses of individual trees (Table A1). Crown fraction equations (Hasenauer and Monserud 1996; Temesgen et al. 2005) also were used to subdivide the live crown mass into foliage and branch components. Canopy volume per unit area and height to the base of the live canopy also were estimated from heights and live crown ratios recorded for individual trees. Missing tree heights were derived following accepted protocols (Hann 1998; Crookston and Stage 1999). We used equations developed by Monleon (Monleon et al. 2004) to "uncompact" the field-recorded, "compacted" live crown ratios for use in subsequent calculations. However, canopy variables derived from both compacted and uncompacted crowns did not differ appreciably, so we used the field-recorded, compacted crown ratio in all calculations. Additional sensitivity testing of the effects of compacted versus uncompacted crowns in various fire models is needed.

Calculation of canopy variables at the plot level

Canopy variables calculated for forest inventory plots are summarized in Table A2.

Canopy cover (CANCOV)

Canopy cover was calculated by estimating the mean crown width using equations based on DBH, height, stem density and species (Crookston and Stage 1999).

Stand height (STNDHGT)

Stand height was calculated as the mean height of all dominant and codominant trees on a plot, weighted by their densities.

Canopy mass (CANMASS) and canopy fuel mass (CANFWT)

CANMASS (kg/ha) was computed as the total mass of all conifer tree crowns on the plot, including foliage and all live and dead branches. CANFWT was computed as the total mass of available canopy fuels (kg/m^2), which included the mass of live conifer foliage plus half the mass of live and dead conifer branches in the 0–0.25 in. size class (Reinhardt and Crookston 2003).

Height to base of live canopy (HCB)

We used a weighted mean of height to crown base for all trees on the plot, with each tree weighted by its expansion factor (Cruz et al. 2003).

Canopy bulk density (CBD)

CBD characterizes the mass of canopy fuels per unit of canopy volume and is a key input to indices of crown fire risk (Scott and Reinhardt 2001) and spatial models of fire behavior (Finney 2004).

CBD was computed by vertically partitioning the live canopy into a series of 0.3 m thick layers (Sando and Wick 1972; Scott and Reinhardt 2001). The proportion of each tree's crown mass that fell within each vertical layer was computed by assuming a uniform distribution of crown mass between the crown base and the top of each tree and summed over all trees. Canopy bulk density was then computed for each layer as

$$\mathrm{CBD}_l = \mathrm{CANFWT}/\mathrm{CL}_l$$

where CANFWT_l was the total canopy fuel mass in layer l expressed in kilograms per square metre, and CL_l was the depth of layer l in metres. The vertical canopy bulk density profile was then smoothed using a 4.5 m running mean, and CBD was computed as the maximum value of this running mean.

Assigning field plots to fuel models

Fuel models for plots in Washington and Oregon

We assigned fuel models to plots in the Washington and Oregon study areas using classification rules derived from published documentation for the Fire and Fuels Extension to the Forest Vegetation Simulator (FVS-FFE) (Reinhardt and Crookston 2003, 2004). We modified the classification rules to reflect the fuels and forest structure variables available for plots in our database. Classification steps were as follows:

- Step 1: Categorize each plot as either "high fuel" or "low fuel" based on the loadings of small downed fuels (≤3 in. diameter) and large downed fuels (>3 in.). Small downed fuel information was only available on a subset of plots, so a logistic regression equation was developed from those plots to determine when plots were high fuel or low fuel based on overstory forest structure.
- Step 2: If the plot is high fuel, assign one of the high-fuel models (10 or 12) based on the loadings of small and large downed fuels. Otherwise, go to step 3 (Washington (WA) or Oregon (OR)).
- Step 3 (WA): If the plot is low fuel, determine whether the plot falls within the Northern Idaho or the East Cascades variant of FVS-FFE. Then choose a set of classification rules based on the species group that accounts for the majority of stand basal area. Go to step 4 (WA).
- Step 4 (WA): Using the classification rules chosen in step 3 (WA), assign one of the low-fuel models (1, 5, 6, 8, or 9) based on canopy cover, quadratic mean diameter, and whether the canopy is composed of a single stratum or is multistoried.
- Step 3 (OR): If the plot is categorized as having low fuels, choose a set of classification rules based on the species group that accounts for the majority of stand basal area. Go to step 4 (OR).
- Step 4 (OR): Using the classification rules chosen in step

Table A1. Crown mass variables calculated at the individual-tree level.

Variable	Description
LCMASS	Total mass of a tree's live foliage and branches (kg)
DCMASS	Total mass of a tree's dead branches (kg)
LCFRAC1	Proportion of a tree's live crown mass comprised of foliage
LCFRAC2	Proportion of a tree's live crown mass comprised of foliage plus branches <0.25 in.
DCFRAC1	Proportion of a tree's dead crown mass comprised of branches <0.25 in.

Note: 1 in. = 2.54 cm.

Table A2. Canopy fuel and canopy structure variables.

Variable	Description
CANCOV	Total canopy cover (%) of live trees
STNDHGT	Mean stand height (m)
CANMASS	Total canopy mass of all conifers (kg/ha)
CANFWT	Total mass of all available canopy fuels (kg/ha)
HCB	Weighted mean height to canopy base of all trees in the plot (m) (Cruz et al. 2003)
CBD	Canopy bulk density computed with vertical layering method (kg/m^3) (Sando and Wick 1972; Scott and Reinhardt 2001)

3 (OR), assign one of the low-fuel models (usually 5, 8, or 9) based on canopy cover and quadratic mean diameter.

Fuel models for plots in California

We adapted methods for assigning fuel models to each plot from the description of the Western Sierra variant of FVS-FFE, modifying them to reflect the fuel and forest structure variables available for our plots and to incorporate recommendations of a local expert from the California Department of Forestry (D. Sapsis, personal communication). Classification steps were as follows:

- Step 1: Classify each plot into a forest cover type based on the California Wildlife Habitat Relationships System (Mayer and Laudenslayer 1988).
- Step 2: Classify each plot into canopy cover and size classes based on the California Wildlife Habitat Relationships System (Mayer and Laudenslayer 1988).
- Step 3: Assign a fuel model to each plot based on its dominant species class, canopy cover class, and size class.

References

Brown, J.K. 1978. Weight and density of crowns of Rocky Mountain conifers. USDA For. Serv. Res. Pap. RP-INT-197.

Crookston, N.L., and Stage, A.R. 1999. Percent canopy cover and stand structure statistics from the Forest Vegetation Simulator. USDA For. Serv. Gen. Tech. Rep. RMRS-GTR-24.

Cruz, M.G., Alexander, M.E., and Wakimoto, R.H. 2003. Assessing canopy fuel stratum characteristics in crown fire prone fuel types of western North America. Int. J. Wildland Fire, **12**(1): 39–50. doi:10.1071/WF02024.

Finney, M.A. 2004. FARSITE: fire area simulation model development and evaluation. USDA For. Serv. Res. Pap. RMRS-4.

Hann, D.W. 1998. Equations for predicting the largest crown width of stand-grown trees in western Oregon. Research Contribution 17. Forest Research Laboratory, Oregon State University, Corvallis, Ore.

Hasenauer, H., and Monserud, R.A. 1996. A crown ratio model for Austrian forests. For. Ecol. Manage. **84**(1–3): 49–60. doi:10.1016/0378-1127(96)03768-1.

Mayer, K.E., and Laudenslayer, W.F. (*Editors*). 1988. A Guide to wildlife habitats of California. Department of Fish and Game, Sacramento, Calif.

Monleon, V.J., Azuma, D., and Gedney, D. 2004. Equations for predicting uncompacted crown ratio based on compacted crown ratio and tree attributes. West. J. Appl. For. **19**: 260–267.

Reinhardt, E.D., and Crookston, N.L. 2003. The fire and fuels extension to the forest vegetation simulator. USDA For. Serv. Gen. Tech. Rep. RMRS-GTR-116.

Reinhardt, E.D., and Crookston, N.L. 2004. The fire and fuels extension to the forest vegetation simulator (Addendum). USDA For. Serv. Gen. Tech. Rep. RMRS-GTR-116.

Sando, R.W., and Wick, C.H. 1972. A method of evaluating crown fuels in forest stands. USDA For. Serv. Res. Pap. RP-NC-84.

Scott, J.H., and Reinhardt, E.D. 2001. Assessing crown fire potential by linking models of surface and crown fire behavior. USDA For. Serv. Res. Pap. RMRS-RP-29.

Snell, J.A.K., and Anholt, B.F. 1981. Predicting crown weight of coast Douglas-fir and western hemlock. USDA For. Serv. Res. Pap. PNW-RP-81.

Snell, J.A.K., and Brown, J.K. 1980. Handbook for predicting residue weights of Pacific Northwest conifers. USDA For. Serv. Gen. Tech. Rep. PNW-GTR-103.

Temesgen, H., LeMay, V., and Mitchell, S.J. 2005. Tree crown ratio models for multi-species and multi-layered stands of southeastern British Columbia. For. Chron. **81**: 133–141.