More than climate? Predictors of tree canopy height vary with scale in complex terrain, Sierra Nevada, CA (USA)

Geoffrey A. Fricker*a,b,c,⁎, Nicholas W. Synesc, Josep M. Serra-Diazd,e,f, Malcolm P. Northg, Frank W. Davis h, Janet Franklina,c

a Department of Botany and Plant Sciences, University of California, Riverside, Riverside, CA 92507, USA
b Social Sciences Department, California Polytechnic University, San Luis Obispo, CA 93407, USA
c School of Geographical Sciences & Urban Planning, Arizona State University, P.O. Box 875302, Tempe, AZ 85287-5302, USA
d UMR Silva, AgroParisTech, Université de Lorraine, INRA, 54000 Nancy, France
e Section for Ecoinformatics and Biodiversity, Department of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus, Denmark
f Center for Biodiversity Dynamics in a Changing World (BIOCHANGE), Department of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus, Denmark
g USFS Pacific Southwest Research Station, Davis, CA 95618, USA
h Bren School of Environmental Science & Management, University of California, Santa Barbara, CA 93106, USA

ARTICLE INFO

Keywords:
Tree height
LiDAR
Mixed-conifer forest
Foothill oak woodland
Water-energy limitation
Climate
Soils
Topography

ABSTRACT

Tall trees and vertical forest structure are associated with increased productivity, biomass and wildlife habitat quality. While climate has been widely hypothesized to control forest structure at broad scales, other variables could be key at fine scales, and are associated with forest management. In this study we identify the environmental conditions (climate, topography, soils) associated with increased tree height across spatial scales using airborne Light Detection and Ranging (LiDAR) data to measure canopy height. The study was conducted over a large elevational gradient from 200 to 3000 m in the Sierra Nevada Mountains (CA, USA) spanning sparse oak woodlands to closed canopy conifer forests. We developed Generalized Boosted Models (GBMs) of forest height, ranking predictor variable importance against Maximum Canopy Height (CHmax) at six spatial scales (25, 50, 100, 250, 500, 1000 m). In our study area, climate variables such as the climatic water deficit and mean annual precipitation were more strongly correlated with CHmax (18–52% relative importance) than soil and topographic variables, and models at intermediate (50–500 m) scales explained the most variance in CHMax (R² 0.77–0.83). Certain soil variables such as soil bulk density and pH, as well as topographic variables such as the topographic wetness index, slope curvature and potential solar radiation, showed consistent, strong associations with canopy structure across the gradient, but these relationships were scale dependent. Topography played a greater role in predicting forest structure at fine spatial scales, while climate variables dominated our models, particularly at coarse scales. Our results indicate that multiple abiotic factors are associated with increased maximum tree height; climatic water balance is most strongly associated with this component of forest structure but varies across all spatial scales examined (6.9–54.8% relative importance), while variables related to topography also explain variance in tree height across the elevational gradient, particularly at finer spatial scales (37.15%, 20.26% relative importance at 25, 50 m scales respectively).

1. Introduction

Forest canopy height is strongly related to forest productivity and carbon sequestration (Keith et al., 2009). Tall and varied vertical forest structure provides habitat for wildlife, and increased canopy height and stem diameter is positively correlated with terrestrial plant diversity at multiple spatial scales (Cazzolla Gatti et al., 2017; Lindenmayer et al., 2012; Lutz et al., 2018; Marks et al., 2016; Slik et al., 2013; Terborgh, 1985). Overstory vegetation is also an important driver of near-surface micro-climate conditions important for plant growth, recruitment and regeneration (Chen et al., 1999). In spite of its importance to ecosystem processes and biodiversity conservation, environmental predictors of forest canopy height have been largely assessed at coarse spatial resolution over continental-to-global scales, despite significant regional and local variation (Tao et al., 2016; Zhang et al., 2016). A better understanding of abiotic drivers of forest height across scales, especially at...
scales relevant to forest dynamics and management, will help connect ecological theory to ecosystem management in an era of global change.

Water-energy dynamics have long been hypothesized to control growth and attainable tree height, and climatic factors affecting maximum tree height have been explored over large latitudinal and altitudinal gradients. Tree height may be constrained due to increased probability of hydraulic failure, as well as limited carbon assimilation in the upper canopy (Ishii et al., 2014; Koch et al., 2004; Ryan and Yoder, 1997), and limited water transport (Jensen and Zwieniecki, 2013). There is evidence for hydraulic resistance and stomatal conductance limiting both tree height and the leaf area to sapwood area ratio, particularly in older, larger individuals, a pattern that increases with tree age and appears to be consistent globally (McDowell et al., 2002; Ryan and Yoder, 1997; Schäfer et al., 2000). For eucalyptus forests in Australia, Givnish et al. (2014) found a strong relationship between precipitation and maximum tree height along a rainfall gradient, suggesting both allocational allometry and hydraulic limitation were determining maximum tree height. They proposed that hotter, drier conditions lead to negative feedbacks related to decreased vertical structure, potentially denser wood and lower hydraulic conductivity (Givnish et al., 2014).

Global-scale studies have shown that climatic factors related to water and energy balance are strong predictors of canopy height, although factor importance varies across biogeographical regions and latitudinal gradients (Cong et al., 2016; Moles et al., 2009; Zhang et al., 2016). Tall trees (> 25 m) are found in both temperate and tropical climates above a rainfall threshold of roughly 1500 mm and where rainfall and temperature variability are low (Scheffer et al., 2018). Globally, canopy height has a bimodal distribution, correlated with the distribution of tree cover; in regions with low precipitation, trees are short and sparse (savanna) whereas in regions with high precipitation, trees are tall and dense (forest). Landscape (kilometers) and local-scale variation (25–500 m) in energy and water balance associated with topography and soils may mediate coarse-scale climate regimes. For instance, topography mediates solar radiation and thus evapotranspiration and water deficit (Dubayah and Rich, 1995). Steeper topography enhances tree biomechanical damage by gravitational forces (King et al., 2009) and influences wind disturbance that could limit tree height (Larjavaara, 2010). Furthermore topography is also key in soil development and erosion, which in turn affects soil water retention (McNab, 1989; Moore et al., 1991), playing a key role in patterns of forest mortality (Anderegg and HilleRisLambers, 2016; Anderegg et al., 2016; Young et al., 2017). Additionally, soil properties influence tree height via nutrient availability (e.g. P, Mg and N) and water dynamics (Cramer, 2012; Huston, 1980). A survey of soil along an elevational transect adjacent to our study area found that soil pH decreases and soil carbon increases with elevation, with large breakpoints in nutrients and weathering coinciding with the transition from oak woodland to mixed-conifer forest, as well as the average effective winter snow line (Dahlgren et al., 1997).

Given the potential for multiple mediating factors at landscape-to-local scales, the goal of this study is to characterize the association of climate, topography and soil factors with forest height across spatial resolutions from 25 to 1000 m within temperate, mid-latitude woodlands and forests found at the same latitude. We use airborne Light Detection and Ranging (LiDAR) data over a 200–3000 m elevational gradient in the Sierra Nevada (California, USA) to determine (1) What is the distribution of tree height across this elevation gradient? and (2) Which climate, topography and soil variables have the greatest influence on maximum tree height and how do these relationships vary with scale? We expected water availability to limit maximum tree height in this region dominated by water-limited forest and woodland, and that factors related to climatic water balance would explain tree height variation at broad scales while topographic factors influenced water balance. We also expected maximum tree height to be greater where soil factors indicate greater availability of plant nutrients.

2. Methods

2.1. Study area

Our study area consists of four non-contiguous sites; three of these form part of the USA National Ecological Observatory Network (NEON; www.neonscience.org) D17 (Pacific Southwest, California, USA; Fig. 1).

Fig. 1. Study Area. Landsat satellite imagery map (true color) and NEON D17 Pacific Southwest sites. San Joaquin Experimental Range (SJER-yellow), Soaproot Saddle (SOAP-blue), Providence Creek (PROV-magenta), Teakettle Forest Watershed (TEAK-red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
This study area was selected because of the availability of prototype NEON airborne remotely sensed data acquired in 2013 using the Airborne Observation Platform (AOP). We used the maximum available data footprint around each research site. From low to high elevation and west to east, the four sites are the San Joaquin Experimental Range (SJER), Soaproot Saddle (SOAP), Providence Creek (PROV), and the Teakettle Watershed (TEAK) (Fig. 1). These sites span a 2800-m elevation gradient of decreasing average temperature and increasing precipitation (Goulden et al., 2012). Sites range from open oak woodland-savanna at 150–520 m at SJER, to conifer-dominated forests from 1390 to 3030 m at Teakettle (Barbour et al., 2007; Mooney and Zavaleta, 2016). Providence Creek and Soaproot Saddle are mid-elevation sites that capture the transition zone from open savanna to dense forest (Mooney and Zavaleta, 2016), and the upper elevation range of the Providence Creek watershed overlaps with the lower range of the Teakettle watershed around 1500 m (Fig. 1). The region has a typical Mediterranean-type climate with warm to hot (17–27° C) dry summers and cool to cold (10–0°C) wet winters (Ma et al., 2010). We were motivated to evaluate the use of publicly available NEON data that are intended for ecological monitoring and because the NEON D17 site was specifically designed to span multiple sites across the valley-montane transition.

The lowest elevation site SJER comprises about 6700 ha of oak woodland and savanna in the Sierra Nevada foothills (36° 58′ N, 119° 2′ W) in California’s Central Valley north-east of Fresno, CA (Radloff et al., 1991). The sparse canopy (< 25%) is dominated by two species of oak (Quercus wislizeni and Quercus douglasii) and foothill pine (Pinus sabiniana), and the understory is composed of scattered shrubs and a nearly continuous cover of herbaceous plants (mostly non-native annual grasses), and gently undulating terrain. This site is currently a functioning research rangeland laboratory associated with California State University, Fresno.

The two middle elevation transition sites Soaproot Saddle and Providence Creek are nearby and ecologically similar. Soaproot Saddle (3300 ha) lies in an intermediate location along the elevation gradient (37° 1′ N, 119° 15′ W), from 920 to 1590 m elevation in California’s southern Sierra Nevada Mountains. The forest is mixed deciduous/conifer forest dominated by ponderosa pine (Pinus ponderosa) and incense cedar (Calocedrus decurrens) with an open, structurally complex canopy and a dense understory and ground layer of shrubs and grasses. Topography is complex with broad hills and valleys. This site receives approximately 20% of annual precipitation as snow and 80% as rain and captures the snow-rain transition. The Providence Creek site (37° 3′ N, 119° 11′ W), a 1000 ha catchment, is the primary research area in the Southern Sierra Critical Zone Observatory (http://criticalzone.org/sierra/) and ranges in elevation from 1580 to 2190 m. Forest vegetation at Providence Creek is similar to Soaproot Saddle, composed of mid-elevation, mixed-conifer forest. The Providence Creek Watershed is part of the larger Kings River Experimental Watersheds research project managed by the USDA Forest Service, Pacific Southwest Research Station, and although included in the initial NEON Airborne Observation Platform data collection in 2013, it will not be collected in future NEON missions. The hydrology and setting of Providence Creek was described in detail by Hunsaker et al. (2012).

Mixed conifer/deciduous forest transitions to red-fir dominated conifer forest at the upper elevations of the 1500–3038 m Wishon watershed. The watershed extends uphill and north of the Wishon Reservoir and downhill to the south of the Wishon Dam where the 1250 ha Teakettle Experimental Forest is located (Kampe et al., 2013). The Teakettle Experimental Forest is located within this 18,500 ha watershed area at 36°58′N, 119°2′W, and at elevations 1900–2500 m. The forest is dominated by white fir (Abies concolor) in terms of basal area and tree density, but sugar pine (Pinus lambertiana) and Jeffrey pine (Pinus jeffreyi) are among the largest diameter and tallest trees. Incense cedar (Calocedrus decurrens), western white pine (Pinus monticola), and lodgepole pine (Pinus contorta) are also prevalent and scattered black oak (Quercus kelloggii) can be found in rocky areas, primarily at the lower elevations. Shrub cover typically consists of whitethorn ceanothus (Ceanothus cordulatus), and green leaf manzanita (Arctostaphylos patula) (North et al., 2002).

Past management activities can influence tree height distributions due to logging practices and forest clearing. Past management activities have influenced the current distribution and abundance of tall trees in the three study areas dominated by conifers (i.e., Soaproot Saddle, Providence Creek and Teakettle) where some logging has occurred beginning in the 1880s, which could blur the relationship between canopy height and abiotic factors. All of these three sites, however, have substantial areas where little to no tree removal occurred due to limited access and mill activity (McKelvey and Johnston, 1992). With the exception of the Teakettle Site’s highest elevations, most of these forests have been selectively harvested at least once over the last century, often removing the largest, commercially-valuable trees (i.e., ‘high grading’ (Rose, 1994)). As a result, residual old-growth stands containing some of the tallest trees could be associated with less mechanically accessible sites such as steeper, mid-slope positions. The Sierra Nevada Forest, however, has not been as heavily logged as many other National Forests particularly those in the northern Sierra Nevada (North et al., 2015, 2009). All three sites have substantial areas where little to no tree removal occurred due to limited access and mill activity (McKelvey and Johnston, 1992; Rose, 1994) and large, old trees are well-distributed across the landscape. Furthermore, previous studies in the Sierra Nevada based on models (Urban et al., 2000), historical data (Collins et al., 2015; Stephens et al., 2015), and LiDAR (Kane et al., 2015) as well as field sampling (Jylens and North, 2012) have found tall trees in mesic locations associated with large-scale climatic water balance and local topography (i.e., valley bottom and lower slope positions), in spite of the history of logging (see Table 1).

### 2.2. Airborne LiDAR data and vertical forest structure

Airborne LiDAR imagery across all sites was collected by the NEON Airborne Observation Platform during multiple flights in June 2013. NEON used an Optech Gemini small-footprint LiDAR sensor that records both discrete range and full waveform returns (Kampe et al., 2013). We used maximum canopy height ($CH_{\text{max}}$) as our response variable to explain the site’s potential for tree growth and as an effort to mitigate the effects of past disturbance from human or natural causes which might disproportionately affect mean canopy height. To control the LiDAR point classification we completely reclassified the point cloud and ran numerous smoothing and outlier point removal filters in addition to a manual classification accuracy check in Microstation’s Terrascan and QCoherent’s LP360 software. The canopy surface/digital

### Table 1

Site code, name, elevation range, climate and description of topography for each of the four study sub-sites.

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
<th>Elevation (m)</th>
<th>Vegetation</th>
<th>Topography</th>
</tr>
</thead>
<tbody>
<tr>
<td>SJER</td>
<td>San Joaquin Experimental Range</td>
<td>148–518</td>
<td>Open oak woodland savanna</td>
<td>Gentle, Rolling Hills</td>
</tr>
<tr>
<td>SOAP</td>
<td>Soaproot Saddle</td>
<td>921–1589</td>
<td>Transition zone from open savanna to dense forest</td>
<td>Complex Topography, broad hills and valleys</td>
</tr>
<tr>
<td>PROV</td>
<td>Providence Creek</td>
<td>1582–2192</td>
<td>Transition zone from open savanna to dense forest</td>
<td>Complex Topography, broad hills and valleys</td>
</tr>
<tr>
<td>TEAK</td>
<td>Teakettle Experimental Forest</td>
<td>1391–3038</td>
<td>Conifer-dominated forests</td>
<td>Steep, complex terrain</td>
</tr>
</tbody>
</table>
elevation model and canopy height model were all derived from this reclassified point cloud. To calculate vertical forest structure from LiDAR we first created a canopy height model (CHM) which is the first-return canopy surface model (CSM) minus the bare-earth digital elevation model (DEM). The 1-m resolution canopy surface model is created by taking the highest return from any ground- or canopy-classified point within each pixel (not including points that strike objects like birds, clouds, smoke, etc.). The digital elevation model is an interpolated, last-return “bare earth” surface which is then rasterized to 1 m to match the resolution of the canopy surface model. After subtracting the digital elevation model from the canopy surface model, the resulting canopy height model is a measure of vertical tree height with differences in topography removed (Næsset, 1997; Patenaude et al., 2004). CH_{max} is the highest value of the canopy height model pixel in the gridded cell at each spatial resolution (25, 50, 100, 250, 500, 1000 m).

The study area has numerous features that are not forested and were identified visually and manually removed from our analysis. These included highways, irrigation ponds, large lakes, private residences and a large utility ‘right-of-way’ corridor in which all tall vegetation has been removed. Grid cells which contained these features were manually digitized and removed. Because most of these structures or clearings were relatively small (< 100 m across), we only removed them from the analyses conducted at the finest spatial scales (25, 50, 100 m). Removing these features focuses the analysis on vegetation that has not undergone obvious human manipulation or clearing. Grid cells with maximum canopy values less than 3 m were also removed to avoid analyzing cells with no trees.

2.3. Predictor variables

2.3.1. Climate

We used annual precipitation, annual temperature seasonality, growing degree days (above 5 °C), maximum annual temperature, minimum annual temperature, and climatic water deficit (CWD) as the climate predictor variables (see abbreviations in Table 2). Annual temperature seasonality is the annual range in temperature, and growing degree days is the annual sum of mean daily temperatures minus 5 for all days with a mean daily temperature > 5 °C. Maximum and minimum temperature is the mean high and low temperature of the warmest and coldest months respectively. Climatic water deficit is quantified as the amount of water by which potential evapotranspiration exceeds actual evapotranspiration (Stephenson, 1998). The climate data used in our study were developed using the Basin Characterization Model (BCM) based on 270 m resolution digital elevation data (Flint et al., 2013). Historical Parameter-elevation Relationship on Independent Slopes Model (PRISM) precipitation and temperature data (Daly et al., 2008, 1994) were spatially downscaled from 800 m to 270 m using Gradient Inverse Distance Squared (GIDS) downscaling (Nalder and Wein, 1998), an approach which applies weighting to monthly point data, developing multiple regressions for every fine-resolution grid cell for every month. Using the PRISM climate variables and a 270 m digital elevation model, parameters weighting is based on the location and elevation of the coarse-resolution cells around each fine resolution cell to predict the climate variable in the fine resolution cell (Flint and Flint, 2012; Nalder and Wein, 1998). The BCM provided gridded estimates of 14 different variables including precipitation, climatic water deficit, temperature and seasonality. From the past 30-years of climate data, we calculated the mean and standard deviation of each of the climate predictor variables at each resolution as potential predictors of CH_{max}. We used these statistics to capture the average, and spatial variability of each of our predictor variables. At coarse scales, individual grids cells can contain large variations in individual variables and at fine spatial scales, climate variables contained no variability so only the mean value was used.

2.3.2. Topography

We focused on terrain variables that are considered proxies for ‘microclimates’ or topo-climates, where topographically-determined variability in radiation, and hydrologic environments might promote tree growth, or modify the regional climate at fine scales (Frey et al., 2016). We varied the spatial resolution of the digital elevation model from 1 to 20 m to identify effects of spatial scale on estimation of variables such as curvature which has been shown to be scale sensitive (Detto et al., 2013), and based on this we chose 1-m resolution for the final analysis. Standard deviation of elevation was calculated at each scale as a measure of terrain roughness (Wilson and Gallant, 2000). We processed the LiDAR digital elevation model to derive primary topographic attributes (Gallant and Wilson, 2000) including mean elevation, terrain slope and curvature at each scale (Moore et al., 1991), and also computed secondary attributes including potential solar radiation on a sloping surface (using the Areal Solar Radiation Model) (Fu and Rich, 2002), and soil wetness estimated using the Topographic Wetness

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Variable description</th>
<th>Units</th>
<th>Native resolution</th>
<th>Variable type</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAP</td>
<td>Mean annual precipitation</td>
<td>mm</td>
<td>270 m</td>
<td>Climate</td>
</tr>
<tr>
<td>ANR</td>
<td>Annual temperature range</td>
<td>Degrees Celsius</td>
<td>270 m</td>
<td>Climate</td>
</tr>
<tr>
<td>GDD</td>
<td>Growing degree days with base of 5 °C</td>
<td>Degree days</td>
<td>270 m</td>
<td>Climate</td>
</tr>
<tr>
<td>Temp_{max}</td>
<td>Maximum temperature of the warmest month</td>
<td>Degrees Celsius</td>
<td>270 m</td>
<td>Climate</td>
</tr>
<tr>
<td>Temp_{min}</td>
<td>Minimum temperature of the coldest month</td>
<td>Degrees Celsius</td>
<td>270 m</td>
<td>Climate</td>
</tr>
<tr>
<td>CWD</td>
<td>Climatic water deficit</td>
<td>mm</td>
<td>270 m</td>
<td>Climate</td>
</tr>
<tr>
<td>CURV</td>
<td>Slope curvature</td>
<td>(unitless) + convex, 0 flat, − concave</td>
<td>1 m</td>
<td>Topography</td>
</tr>
<tr>
<td>TWI</td>
<td>Topographic wetness index (up slope contributing area scaled by slope)</td>
<td>(unitless)</td>
<td>1 m</td>
<td>Topography</td>
</tr>
<tr>
<td>DEM Solar 3 m</td>
<td>Potential solar radiation on a sloping surface</td>
<td>Watts/m²</td>
<td>3 m</td>
<td>Topography</td>
</tr>
<tr>
<td>DEM_{std}</td>
<td>Standard deviation of elevation</td>
<td>m</td>
<td>1 m</td>
<td>Topography</td>
</tr>
<tr>
<td>AWC_{mean}</td>
<td>Available water content</td>
<td>cm water/cm soil</td>
<td>Vector</td>
<td>Soil</td>
</tr>
<tr>
<td>OM_{mean}</td>
<td>Organic matter</td>
<td>mg</td>
<td>Vector</td>
<td>Soil</td>
</tr>
<tr>
<td>pH_{mean}</td>
<td>Potential of Hydrogen</td>
<td>– 10 log H+</td>
<td>Vector</td>
<td>Soil</td>
</tr>
<tr>
<td>PARMATNM D</td>
<td>Geologic parent material</td>
<td>Rock type from Basalt, Till, Granite, etc.</td>
<td>Vector</td>
<td>Soil</td>
</tr>
</tbody>
</table>

Subscripts

Max: Maximum (ex. Temp_{max})
Min: Minimum (ex. Temp_{min})
Mean: Mean (ex. OM_{mean})
Sd: Standard Deviation (ex. DEM_{std})

Topography Data Source: http://data.neonscience.org/home.
Index, a physically-based basin contribution model (Beven and Kirkby, 1979). Equation below:

\[ \text{Topographic Wetness Index} = \ln \frac{\alpha}{\tan \beta + c} \]

where \(\alpha\) is the upslope contributing basin area, \(\beta\) is the slope at that cell as defined by Moore et al. (1991) and we modified the equation slightly by adding \(c\), a small constant (\(c = 0.01\)), to avoid division by zero in flat terrain cells. We calculated the topographic predictor variables using Python scripts in ArcGIS 10.3.

2.3.3. Soil

We selected soil variables that reflect the physical and chemical properties of soils that influence vertical vegetation structure. These included available water content, organic matter, pH and geologic parent material (Table 2). Soil data were obtained from the National Resource Conservation Service’s SSURGO and STATSGO national soil databases using the ArcGIS SSURGO Soil Data Development Toolbox (Soil Survey Staff United States Department of Agriculture, 2017). We gridded continuous and categorical soil variables using the Map Soil Properties and Interpretations tool in the gSSURGO Mapping Toolset in ArcGIS 10.3. We calculated the mean and standard deviations of Available Soil Water Content, OM and pH at each scale. We also included three categorical variables related to geologic substrate, rock type and geologic parent material.

2.4. Statistical analysis

Our statistical methods used generalized boosted models to predict \(CH_{\text{max}}\) as a response variable from environmental variables which characterized climate, topography and soil characteristics. The predictor variables were calculated from source data ranging in spatial resolution from 1 to 270 m (Table 2) and then gridded at six different spatial resolutions, resulting in a range of sample sizes (number of grid cells) available for each scale of analysis: 1000 m (\(n = 195\)), 500 m (841), 250 m (3826), 100 m (24,895), 50 m (102,001), and 25 m (400,460). Our study was designed to span a range of resolutions in order detect patterns in these scale-dependent correlations. The up-scaling of finer resolution to coarser resolutions was done by nearest neighbor averaging for continuous variables, and for the soil categorical variables, the category with most of the area in each grid cell was used to represent the entire grid cell.

We used generalized boosted (regression tree) models in \(R\) (R Core Team, 2017), Version 1.0.136, package ‘caret’ and ‘gbm’ (Kuhn, 2008; Ridgeway, 2007) to predict maximum canopy height variables from the environmental predictors. We chose generalized boosted models because they combine the strengths of two algorithms, regression trees (models that relate a response to their predictors by recursive binary splits) and boosting (an adaptive method for combining many simple models to give improved predictive performance). Boosted regression trees have been used extensively in ecological modelling (Elith et al., 2008). Generalized boosted models (GBMs) are a powerful ensemble statistical learning approach capable of achieving bias reduction through forward stagewise fitting, suitable for handling different types of predictor variables and their interactions, and able to characterize complex data-generating processes (Elith et al., 2008; Hastie et al., 2009). The final model can be understood as an additive regression model in which individual terms are simple trees, fitted in a forward, stage-wise fashion. Generalized boosted models provide an estimate of variance explained by the model and the relative importance of the predictor variables.

We initially explored many potential predictors within each group (climate, topography, and soil) and computed a preliminary set of generalized boosted models to screen variables. The results of the preliminary generalized boosted models were sorted by spatial resolution and variable importance was ranked to remove the lowest contributing third of all variables from subsequent modeling. The top predictor variables in each group are listed in Table 2 (see Table S1 for a full list of all variables initially considered). GBM models of maximum canopy height were then developed using the top two thirds of the candidate predictors from each group. Model parameters were calibrated with 10-fold cross-validation and a full factorial design with interaction depth varied over the integers from 1 to 5. The number of regression trees was varied from 2,000 to 10,000 in increments of 2,000 and the shrinkage rate was varied from 0.1 to 0.01, at intervals of 0.01.

The gbm package in \(R\), originally developed by Friedman (2001), estimates the relative influence of predictor variables. This measure of variable importance is defined as the number of times a variable is selected for splitting, weighted by the squared increase in explained deviance to the model as a result of each split, and averaged over all trees (Friedman and Meulman, 2003). Thus, each variable’s relative contribution (or importance) represents its percentage of the total contribution of all variables. Although variable importance is determined by splitting thousands of models in different trees, generalized boosted models should not be considered a statistical ‘black box’ since individual variable responses can be summarized, evaluated and interpreted similarly to a conventional regression model using partial dependence plots (Elith et al., 2008). In our study, variable importance is tracked relative to the other variables for models at each spatial scale.

We expect \(CH_{\text{max}}\) (our response variable) to be correlated with environmental predictors that we know are spatially structured (Lennon, 2000). We would expect environmental conditions to show positive spatial autocorrelation (SA), at spatial lags of tens to thousands of m for topography, and tens to hundreds of km in the case of climate. Boosted regression tree models (GBM) are more robust to the effects of SA on model fit, variable importance and estimated response curves than generalized linear models (Crase et al., 2012). Model residuals were tested for SA at each spatial scale (one-cell lag for 250, 500 and 1000 m scales, lags 1–4 for 100-m, lags 1–5 to 50-m and lags 1–6 for 25-m) to aid interpretation of the models. Analysis of SA in model residuals can suggest that there may either be missing (spatially structured) environmental predictors or that there are spatially structured data generating processes for the response variable, but cannot distinguish between these exogenous or endogenous causes (Dormann et al., 2007; Legendre et al., 2002).

3. Results

3.1. Canopy height on an elevation gradient

Estimated maximum tree height ranged from 3 to 70 m, measured at elevations ranging from 200 to 3000 m. The distribution of maximum tree height with elevation was non-linear, with a peak at about 2300 m and a secondary peak at about 1200 m. Maximum tree height is smallest at the lowest elevation in the transect but declines at both ends of the elevation gradient. We lacked observations between 500 and 950 m elevation – the elevation gap between the open oak woodland (San Joaquin Experimental Range) and the transition zone to mixed conifer (Soaproot Saddle) (Fig. 2). However, this gap is less than 14% of the total elevation range and our data do include the rain-snow transition or the water- to energy-limited forest transition at 2400 m.

3.2. Predictor variables associated with canopy height across scales

Overall variance in maximum height explained by generalized boosted models was roughly the same across the intermediate scales examined (50–500 m) ranging from 72 to 83%, and greatest at the 100–250 m scales. At both the coarsest (1000 m) and finest (25 m) spatial scales, the amount of variance explained was considerably lower than at middle scales, particularly at the finest spatial scale at which only 21% of total variance was explained. The relative influence of all aggregated climatic, topographic and soil predictors was similar across
scales; soil and topography converged in their importance at 500- to 1000-m scale, but still both were much less important than climate (Fig. 3). The relative influence of soil and topography variables decreased, and influence of climate increased, for coarser-scale models, and at 25 1000-m scale four of the five top-ranked predictor variables were climate predictors (Fig. 4).

We show the five top-ranked predictors for CH$_\text{max}$ at each scale (Fig. 4; variable importance ranking for all predictors is shown in Table S2). CH$_\text{max}$ is most strongly correlated with climate variables including climatic water deficit, growing degree days, and temperature. Annual Precipitation, climatic water deficit, standard deviation of climatic water deficit, minimum temperature, maximum temperature, growing degree days, standard deviation of growing degree days, and annual temperature seasonality all were included among the top five predictors for at least one of the spatial scales. Temperature variables related to growing season length (minimum temperature and growing degree days) and heat stress (maximum temperature) rank among the top predictors only at the coarser 250- and 500-m scale. Topoclimatic variables including solar insolation and topographic wetness index are important predictors at the finest (25–50 m) scale. The topographic variables are more strongly associated with canopy height compared to soil variables across scales, with a strong divergence at the 25-m scale (Fig. 3). The only soil attributes included in the top five predictors at any scale was average pH (Fig. 4), although other soil variables were included in the full models (Table S2).

Maximum canopy height declined with increasing CWD and had an approximately unimodal response to annual precipitation – height was greatest at middle levels of precipitation and declined at the very highest values of precipitation. Maximum height also was greatest at intermediate values of maximum temperature (Fig. S1a).

Models residuals were not significantly spatially autocorrelated (P > 0.05 based on Moran’s I) for the 1000-m, 500-m or 250-m coarser-scale models (Table S3). Residuals were spatially autocorrelated (P < 0.05) for 25-m, 50-m and 100-m finer-scale grids at all lags tested, suggesting that there are either additional spatially-structured environmental drivers not included in our model that may be important at those scales, or that there are endogenous factors (biological processes) causing tall trees to be near other tall trees and vice versa at those scales. These Moran’s I values were small, however, ranging from 0.02 to 0.33 on a scale of 1 to −1, where 0 indicated complete spatial randomness (Table S3). This suggests that SA was not strong; the Moran’s I values were nonetheless significantly different from zero because of the extremely large sample size – the statistic was calculated based on every cell in the study area grid.

3.3. More than climate I: terrain curvature and solar radiation

Although terrain curvature only explains 1.3–6.6% of total variance across scales, there is a consistent cross-scale association between terrain curvature and canopy height, with taller trees occurring in valley bottoms or on concave slopes (negative curvature). At fine spatial scales (25–100 m) the negative association of terrain curvature with height is the strongest, weakens at coarser spatial scales, and is weakest across scales for the oak woodland (SJER) site (Fig. 5).
3.4. More than climate II: soil parent material

Although soil variables were the least important factors associated with CHmax across all scales in our comparisons, there are instances where canopy height is stunted on specific soil types (Fig. 6). Maximum canopy height was greatest on residuum weathered from basalt, residuum weathered from andesite, and residuum/colluvium/till weathered from granite parent materials. Lower CHmax was found on residuum weathered from metasedimentary rock, alluvium/colluvium derived from granodiorite and residuum weathered from granite. The majority (∼85%) of the study area is underlain by granite parent material, but basalt is present, and weathering patterns and soil texture change along the elevation gradient and with topography.

4. Discussion

The results of this study highlight strong, scale-dependent associations between maximum canopy height and water availability as measured by the climatic water deficit, mean annual precipitation, and topographic factors across a ∼2800 m regional elevational gradient. Remarkably, despite the extensive disturbance history of the region, these environmental factors explain 70% of the variance in maximum canopy height within these mid-latitude temperate woodlands and forests. Generalized boosted models explain most of the variance in CHmax at spatial scales of 50–500 m. As predicted, coarse-scale patterns of canopy height (250–500 m) are associated primarily with climatic variables related to water balance. While climate variables still dominate at finer scales (50–100 m), topographic variables affecting moisture availability (terrain curvature, topographic wetness index, solar radiation model) become relatively more influential (Fig. 4). Although most of the area is underlain by granitic parent material, CHmax is also associated with parent material and associated soil properties, notably soil pH. We acknowledge that there is a roughly 450 m elevation gap in our data however this gap does not cover the rain-snow transition zone or elevations that coincide with critical zones of species turnover or water-energy limitation transition.

4.1. Climatic variables associated with maximum tree height

Temperate forest structure along the elevation gradient is limited by the availability of water and energy (Boisvenue and Running, 2006). At the dry low-elevation end of the moisture availability gradient, tree growth may be moisture limited, while at the moist end, light competition may drive forest height (Liénard et al., 2016). At higher elevations and latitudes with freezing winter temperatures and a short growing season, we would expect canopy height to be limited by low temperatures (Reich et al., 2015), as illustrated by the short, sparse nature of boreal forest canopies near arctic tree line, and shorter trees as alpine tree line is approached (Paulsen et al., 2000). In the tropics, however, global studies indicate that temperature is not a limiting factor for tree height (Way and Oren, 2010). Additionally, there is evidence that the world’s tallest trees are found in temperate latitudes and grow in similar (mild and stable) thermal climates (Larjavaara, 2014).

The overwhelming importance of climate variables describing water limitation found in this study is consistent with coarse-resolution, global-scale studies showing that water availability limits maximum canopy height in tropical and temperate regions (Scheffer et al., 2018; Zhang et al., 2016). Our results are also consistent with the characterization of forests below ∼2400 m in our study region as water limited (Das et al., 2013; Tague et al., 2009). Along this same gradient, annual evaporation and gross primary production have been found to be greatest at 1160 and 2015 m; both were lower at 405 m, coincident with less precipitation, and at 2700 m coincident with colder temperatures (Goulden et al., 2012). We found that climate variables reflecting energy limitation (minimum temperature, growing degree
Fig. 5. Maximum canopy height plotted as a function of terrain curvature at six scales. Negative curvature is concave up (valleys) and positive curvature is concave down (ridges).

Fig. 6. “Soil Type”: Boxplots showing Maximum Canopy height cross tabulated by Soil Geologic Parent Material. Line is median, box encompasses 25–75th percentile, whiskers encompass 5–95th percentile, dots are observations beyond that. The sample size is shown in each box (number) for the 1 ha (100 × 100 m grid cell) scale.
days) were also correlated with canopy height along our gradient that extended into energy-limited forests above 2400 m elevation with increased snow cover and shorter growing seasons. Lower CHmax values were found at low values of minimum and maximum temperature, high values of temperature seasonality, and low values of growing degree days.

While we did attempt to quantify both geological substrate and water availability, variables like geologic substrate type do not capture deep, subsurface porosity or water holding capacity (Meyer et al., 2007), and the climatic water deficit measure used only accounts for available moisture in the top layer of soil (Flint et al., 2013). A study of subsurface water in the Southern Sierra Critical Zone Observatory found that large trees are deeply rooted in highly porous saprolite (weathered subsurface rock at the base of the soil profile) with roots reaching 10–20 m below the surface. This porous rock layer contains large volumes of subsurface water and is vital to supporting the ecosystem through the summer dry season and extended droughts (Klos et al., 2018). Having spatially-explicit maps of subsurface porous rock containing water that can be tapped by large trees would improve our ability to model maximum tree height and predict future forest distribution. In spite of this limitation, CWD explained 18–52% of the variation in maximum tree height and was the most important predictor at every scale. The 25-m resolution models explained substantially less variance than those for coarser resolutions, and also showed the greatest spatial autocorrelation in residuals. This suggests that the mapped predictors used in this study do not describe patterns of maximum tree height at that scale, and that there are other exogenous or endogenous factors affecting CHmax at the local scale. Possibly, at that higher resolution there is a qualitative biological gap that could explain such differences. At 30 m, it is likely that we are switching from describing canopy to describing individual trees. At that level of organization (individuals vs. tree communities) it is likely that our ability to capture individual histories through climate decreases. Indeed, cross-scaling across levels of ecological organization still remains a challenge. We are uncertain why the explanatory power of the model declined from 500-m to 100-m resolution, but we note that the amount of variance explained by our 1000-m resolution models is about the same as was explained in a global model based on 55-km grids (Zhang et al., 2016).

4.2. Topography

Topography affects vertical forest structure by controlling environmental factors such as water drainage, solar radiation regime, soil depth, cold air pooling and wind exposure. As predicted, topographic effects were detected at the finest spatial scales in the generalized boosted models for CHmax, but show less importance at the coarse landscape scale where effects of climate dominate.

Terrain curvature, topographic wetness index and the solar radiation model all affect soil water balance and were important relative to the other topography variables. At fine scales (25–100 m), solar radiation was more important and at coarse scales (250–1000 m) terrain curvature was more important. This indicates that specific levels of solar exposure and topographic concavity can both promote taller tree growth, independent of meso-scale climate or soil characteristics. Tree-ring data from an Appalachian watershed showed differences in growth rates on different topographic aspects with nearly all species exhibiting faster growth rates on (cooler, shaded) northeast facing slopes compared to (warmer, drier) southwest facing slopes, presumably due to differences in solar radiation driving evaporative demand (Fekedulegn et al., 2003).

Taller trees generally occur in valleys as opposed to ridgetops (Fig. 5), and are found at the lowest levels of solar radiation; high levels of topographic radiation are associated with shorter tree heights at the finest spatial scales, suggesting the dominance of water-limitations (resulting from the positive relationship between insolation and water stress) on much of the gradient (Fig. S1). Tall trees found at intermediate levels of potential radiation may reflect the ameliorative effects of topography on climatic temperature limitations to tree height at higher elevations in the transect where the tallest trees are found. While other studies of canopy height in the Sierra Nevada Mountains have found a positive correlation between change in tree height and the topographic wetness index (Ma et al., 2018), our results showed that climatic variables are more strongly associated with canopy height over regional scales while topographic wetness is correlated with maximum height at local scales.

4.3. Soil

Among the soil variables considered, pH had the strongest association with CHmax, but this is likely because in our study region tall, coniferous trees are found on granitic-derived, shallow, poorly-developed acidic soils, while low elevation oak woodland trees are found on more basic soils that have developed on colluvium and alluvium. Low pH soils are probably not driving tall tree growth but pH is correlated with the elevational gradient in water availability and phylogenetically-determined limits to maximum tree height among the taxa that dominate different parts of the gradient. Soil pH is related to the amount of precipitation, with soils at higher elevations experiencing heavier leaching and consequently lower pH values. The lower pH values result in lower cation exchange capacity and nutrient poor soils at the highest elevations. Giant Sequoia trees (Sequoiadendron giganteum) are conifers found along our study gradient adjacent to our study sites and are among the tallest trees in the world. This suggests that soil nutrients or pH are not generally limiting to conifer growth compared to other predictor variables considered. Some soil geologic parent materials were associated with taller or stunted maximum canopy heights, but parent material was not highly ranked among soil variables across scales as a predictor of maximum height. Differences in forest structure are related to erosion rates, soil depth and nutrient deficiencies (Cramer, 2012), all of which are influenced by parent material. Our ranking of variable importance suggests that at low elevations water availability is limiting tree heights rather than nutrient limitation, but the effects of soil parent material are still present. In the San Joaquin Experimental Range, a relatively small area of forest (181 ha) is found on ‘Residuum derived from Metasedimentary Rock’ and contains the shortest trees of any geologic parent material type in our study. There is a distinct break in canopy height between this area and other adjacent areas in the open oak woodland savanna (SJER) which experience similar climatic and topographic conditions suggesting this soil parent material type is poorly suited for supporting large trees (Fig. 6).

Our ability to characterize an effect of soils properties on tree height was compromised by both the characteristics of the study area and the accuracy and precision of available large-scale mapped data. Geology and soil were not randomly distributed on our elevation transect, preventing us from disentangling the effects of substrate versus other factors. The soil types in the NRCS soil survey are based on relatively few field samples, and spatial interpolation to map units is based on aerial photographs and historic data (Peters and United States. Forest Service. Northern Research Station, 2013). So, while our models show that mapped soil characteristics and the geologic parent material are marginally important, comprehensive, spatially-explicit field soil surveys and maps would be needed to better understand the effects of soil nutrients and geology on forest height, particularly at fine spatial scales (Grunwald et al., 2011; Rossiter, 2006).

4.4. Management implications

Most Sierra Nevada forests lack resilience to wildfire and drought because historic logging practices and fire suppression have reduced large tree abundance and significantly increased fuel loads, stand density and water stress (Stephens et al., 2018). Current management
practices emphasize realigning forest conditions with topographic differences in water availability and local fire regime (North et al., 2009). A particular focus is on identifying and developing large, tall trees associated with sensitive vertebrate species such as the California spotted owl (Strix occidentalis occidentalis) and the fisher (Martes pennanti) in more mesic, productive sites buffered from high-severity wildfire and drought stress (North et al., 2017; Stephens et al., 2015). Our results suggest forest managers could identify such locations using both large-scale (i.e., > 500 m) differences in CWD from readily available mapped data (i.e., BASIN (Flint et al., 2013)) and fine-scale (i.e., 25–100 m) topographic indicators associated with higher soil moisture (i.e., GIS-generated topographic wetness index). This could help focus budget-constrained management practices in these key areas on reducing fuel loads and water competition, creating stand structures to protect and foster large, tall tree development.

In the context of global climate change, our findings suggest that as broad scale changes in climate lead to shifts in moisture and temperature regimes, large trees will only persist in their current range where topographic and soil conditions allow. Currently, coarse scale models of climate and ecosystem response lack the capacity to incorporate microclimatic variability critical to biodiversity refugia (Ashcroft et al., 2012; Dobrowski, 2011; Frey et al., 2016). Higher elevations that are currently snow covered for much of the winter and spring, will be less energy limited under a warmer climate and habitat loss at lower elevations could be offset by habitat gain at upper elevations. This warmer transition could also increase water stress as there is effectively less moisture available for plant growth at all elevations. This future scenario is supported by evidence of shifts in California’s forest towards smaller, denser forests with an increase in oak species (McIntyre et al., 2015).

The Southern Sierra Nevada Mountains lie at a particularly sensitive geographic junction where drier, warmer conditions will persist into the next century and already this area has experienced high canopy water loss and tree mortality, particularly during the most recent drought from 2012 to 2015 (Asner et al., 2016). As climate changes, species and consequently forest structure will also shift geographically. There is evidence of these shifts in progress along a nearby elevational gradient where Pinyus ponderosa and P. lambertiana experienced increased mortality compared to the other dominant tree species (Paz-Kagan et al., 2017). The Southern Sierra Nevada mountains are also home to the largest trees in the world (Great Sequoias Sequoia giganteum); although these trees did not occur within the footprint of the available LiDAR imagery, the climate is very similar to the mid-elevation transition sites (Soaproot Saddle/Providence Creek) and these isolated pockets of Sequoias will also experience Southern Sierran climatic changes in the next century. Extensive human management and fire in these forests has affected species composition and structure, highlighting the importance of anthropogenic influences on the forests of the Southern Sierra Nevada (Roy and Vankat, 1999). The elevation gradient spanned in this study allows us to make predictions about forest structure as climate changes in the next century, and we expect broad scale changes to be driven by water availability while fine-scale refugia will provide microclimatic buffering against hotter and drier conditions.

Acknowledgements

Funding was provided by the U.S. National Science Foundation (EF-1065864, -1550653, -1568262 and -1550640). The authors would like to thank D. Tazik, T. Goulden and N. Leisso from NEON with their support and patience providing guidance in using NEON data and derived products. We thank I. McCullough for his input.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foreco.2018.12.006.

References

Fu, P., Rich, P.M., 2002. A geometric solar radiation model with applications in
current landscape-scale ponderosa pine and mixed conifer forest structure in the Southern Sierra Nevada. Ecosphere 6, 1–63.