IOP Publishing Environ. Res. Lett. 19 (2024) 124008 https://doi.org/10.1088/1748-9326/ad8766

ENVIRONMENTAL RESEARCH
LETTERS

RECEIVED
29 May 2024

REVISED
3 October 2024

ACCEPTED FOR PUBLICATION
16 October 2024

PUBLISHED
25 October 2024

LETTER

Lidar-derived estimates of boreal shrub biomass in Southcentral Alaska

Kaili A Martin[1], **Roman J Dial**[1,*], **Bruce D Cook**[2], **Douglas C Morton**[2], **Erin I Larson**[3] **and Hans-Erik Andersen**[4]

[1] Institute of Culture and Environment, Alaska Pacific University, Anchorage, AK, United States of America
[2] NASA Goddard Space Flight Center, Greenbelt, MD, United States of America
[3] Alaska Center for Conservation Science, University of Alaska Anchorage, Anchorage, AK, United States of America
[4] USDA Forest Service, Pacific Northwest Research Station, University of Washington, Seattle, WA, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: roman@alaskapacific.edu

Keywords: shrubification, shrubs, biomass, G-LiHT, lidar, boreal, Alaska

Supplementary material for this article is available online

Abstract

Despite widespread observations of shrub proliferation and expansion (shrubification), few studies quantify shrub biomass at the regional scale. Here we describe and implement a two-part modeling approach to estimate and map tall shrub (diameter at root collar > 2.5 cm) expected aboveground biomass, or E[SHB], across a 16.6 million ha boreal region where shrubification occurs in wetlands and subalpine ecosystems. Using $n = 384$ field plots nested within $m = 11$ study sites across southcentral Alaska, we constructed random forest models of the probability (pSH) that shrub wood volume surpasses tree wood volume, and generalized additive models of aboveground biomass of tall shrubs (SHB) using rasterized aerial lidar variables collected by NASA Goddard's Lidar, Hyperspectral, and Thermal (G-LiHT) Airborne Imager, together with gridded climate data from a Parameter-elevation Regressions on Independent Slopes Model 30-year normal climatology (PRISM). Applying those models to G-LiHT tiles, then averaging across tiles within $n = 843$ watersheds covering 9.2 million ha, we estimated that below 1000 m asl, the area-weighted mean value of E[SHB] = pSH x SHB = 6.6 Mg ha^{-1} with sd = 4.5 Mg ha^{-1}. This is the first study to estimate current shrub biomass density at the regional scale in southcentral Alaska and serves as a biomass baseline for measuring and modeling aboveground carbon fluxes where plant communities are undergoing climate-driven change.

1. Introduction

High latitude warming changes Arctic and boreal ecosystems with global implications (Chapin *et al* 2005, Box *et al* 2019, Overland *et al* 2019). Above 60°N, the proliferation and expansion of alder (*Alnus* spp.), willow (*Salix* spp.), and birch (*Betula* spp.) shrubs is among the most widespread terrestrial change (García Criado *et al* 2020, Mekonnen *et al* 2021). This shrubification affects snowpack dynamics, permafrost stability, soil conditions (structure, biology, hydrology and nutrient availability), and surface albedo by transforming aboveground vegetation structure, thereby altering the surface energy balance and vegetation carbon stocks, as well as modifying or replacing habitats for animals and shorter statured plants (Myers-Smith *et al* 2011, Tape *et al* 2016, García Criado *et al* 2020, Mekonnen *et al* 2021).

Early studies detected shrubification in Arctic (Sturm *et al* 2001) and boreal Alaska (Klein *et al* 2005, Dial *et al* 2007) using repeat optical imagery. Recent studies combine Landsat imagery with field data to map shrub expansion across Alaska (Macander *et al* 2022) and to estimate shrub biomass in tundra ecosystems of Arctic Alaska (Berner *et al* 2019). However, biomass estimates of shrubs in boreal ecosystems have not yet been quantified at regional scales. The development and demonstration of feasible, robust methods to calculate regional shrub biomass would improve monitoring protocols of carbon stocks and fluxes, such as the US Forest Service Forest Inventory Analysis (FIA) program.

FIA evaluates forest resources and assesses ecosystem management by sampling and quantifying timber, vegetation, and woody debris at thousands of sites across the USA, including Alaska, where boreal forests and woodlands contain large stores of carbon (Douglas *et al* 2014, Shoot *et al* 2021), much of it in tall shrubs. Currently, FIA surveys do not include tall shrubs, thereby underestimating aboveground carbon (Barrett and Gray 2011, PNW-FIA 2019, Alonzo *et al* 2020) and contributing to uncertainty associated with carbon estimates (Pan *et al* 2011, Bradshaw and Warkentin 2015). Due to the expensive logistical challenges of conducting repeat field inventories in remote regions, carbon monitoring at high latitudes will increasingly rely on remote sensing ground-truthed through regionally-specific fieldwork (Wulder *et al* 2012).

Here we describe a field-informed, airborne lidar-based modeling approach to estimate the aboveground dry biomass density of tall shrubs (SHB), defined as the biomass of shrubs with diameter at root collar (DRC) $\geqslant$2.5 cm (and generally taller than 1 m), across southcentral Alaska (figure 1(a)), where ongoing shrubification occurs in boreal wetlands (Klein *et al* 2005, Berg *et al* 2009) and subalpine communities (Dial *et al* 2007, 2016). Using lidar data from G-LiHT (Cook *et al* 2013, Babcock *et al* 2018, Alonzo *et al* 2020) and climate data from PRISM (Daly *et al* 1994, 2018, 2008), we constructed predictive models of shrub biomass across the region. We modeled expected SHB (E[SHB]) using a two-part modeling approach because satellite- and aerial-based remote sensing cannot always distinguish between tree and shrub plant functional types (PFTs) where the two are mixed in forests, woodlands, and forest-tundra ecotones. Using the two-part model to increase accuracy and reduce bias, we estimated and mapped the mean E[SHB] in over 800 United States Geological Survey (USGS) Hydrologic Unit Code 12 watersheds (HUC12s) overflown by G-LiHT (figure 2).

2. Methods

2.1. Overview

Field-data collected at $n = 384$ plots (168.1 m^2) distributed as 14–60 field plots per site at 11 sample sites (table 1) across a 500 000 ha region (figure 1(b)) informed modeling. Each plot included tree diameter at breast height (DBH) and measurements of shrubs (figure 3). Using published regional allometries (Alonzo *et al* 2020, Dial *et al* 2021b), we calculated for each plot the total biomass of trees (TRB) with DBH $\geqslant$2.5 cm and SHB. We classified plots by dominant PFT (shrub, tree, herbaceous). If SHB exceeded TRB, PFT was defined as 'shrub' ($n = 212$ plots); if TRB exceeded SHB, PFT was defined as 'tree' ($n = 157$); if both TRB and SHB were zero, the plot PFT was defined as 'herbaceous' ($n = 15$).

We evaluated 36 variables as predictors (table S1). Using PFT classifications from all plots, we constructed models giving pSH, the probability that SHB exceeded TRB. Our definition of pSH is that it identifies the probability that shrub wood volume surpasses tree wood volume, given climate, canopy structure, and topography, rather than as a percent cover on the landscape or the chance that a random sample contains a tall shrub.

To estimate pSH, we constructed random forest classification models (RFs) using package *randomForest* (Liaw and Wiener 2002) in R (v. 4.3.3, R Core Team 2024). RF classification models produced three probabilities: pSH = Pr(PFT = shrub), pTR = Pr(PFT = tree), and pHE = Pr(PFT = herbaceous). To avoid bias in estimates of SHB, we constructed models of SHB using only shrub plots in generalized additive models (GAMs) from R's *mgcv* package (Wood 2017). The product of pSH and SHB gives

$$\mathrm{E}[\mathrm{SHB}] = \mathrm{pSH} \times \mathrm{SHB}, \tag{1}$$

a simple formulation to reduce bias and increase accuracy given the limitations of rasterized aerial lidar in resolving vegetation structure.

2.2. Field-methods

Our approach scaled plot data of individual shrub measurements (figure 3, supplementary material, *Field data collection*), where SHB was calculated as kg plot^{-1}, to watersheds, where E[SHB] was estimated as Mg ha^{-1}. The area of interest (AOI) encompassed 16.6 million ha including the Copper and Matanuska-Susitna river basins; adjacent watersheds of western Cook Inlet; and portions of the Kenai Peninsula (figure 1(a), figure 2). Forest dominance included white spruce (*Picea glauca*), birch (*Betula* spp.), or aspen (*Populus tremuloides*) trees on well-drained soils; black spruce (*Picea mariana*) on wetter soils, particularly in permafrost regions of the Copper River basin; and Sitka spruce (*P. sitchensis*) and hemlock (*Tsuga* spp.) in coastal areas.

2.2.1. Experimental design

We chose sites (figure 1(b)) based on representativeness of vegetation, G-LiHT coverage (figure 1(a)), and ease of access. Sample plots (area = 168.1 m^2; table 1) were systematically located 26 m apart along 1–5 transects roughly parallel to elevation contours situated 40–1500 m apart. All *Alnus* and *Salix* stems with DRC $\geqslant$2.5 cm rooted within plots were identified to genus. We recorded several shrub measurements if DRC $\geqslant$10.2 cm (figure 3, supplementary material *Estimating plot shrub biomass*) to reduce bias and increase accuracy (Dial *et al* 2021b). All trees with DBH $\geqslant$2.5 cm rooted within plots were identified to species (*Betula, Picea, Populus, Tsuga*) or genus (*Salix*) and DBH recorded. Using a Trimble Geo 7x

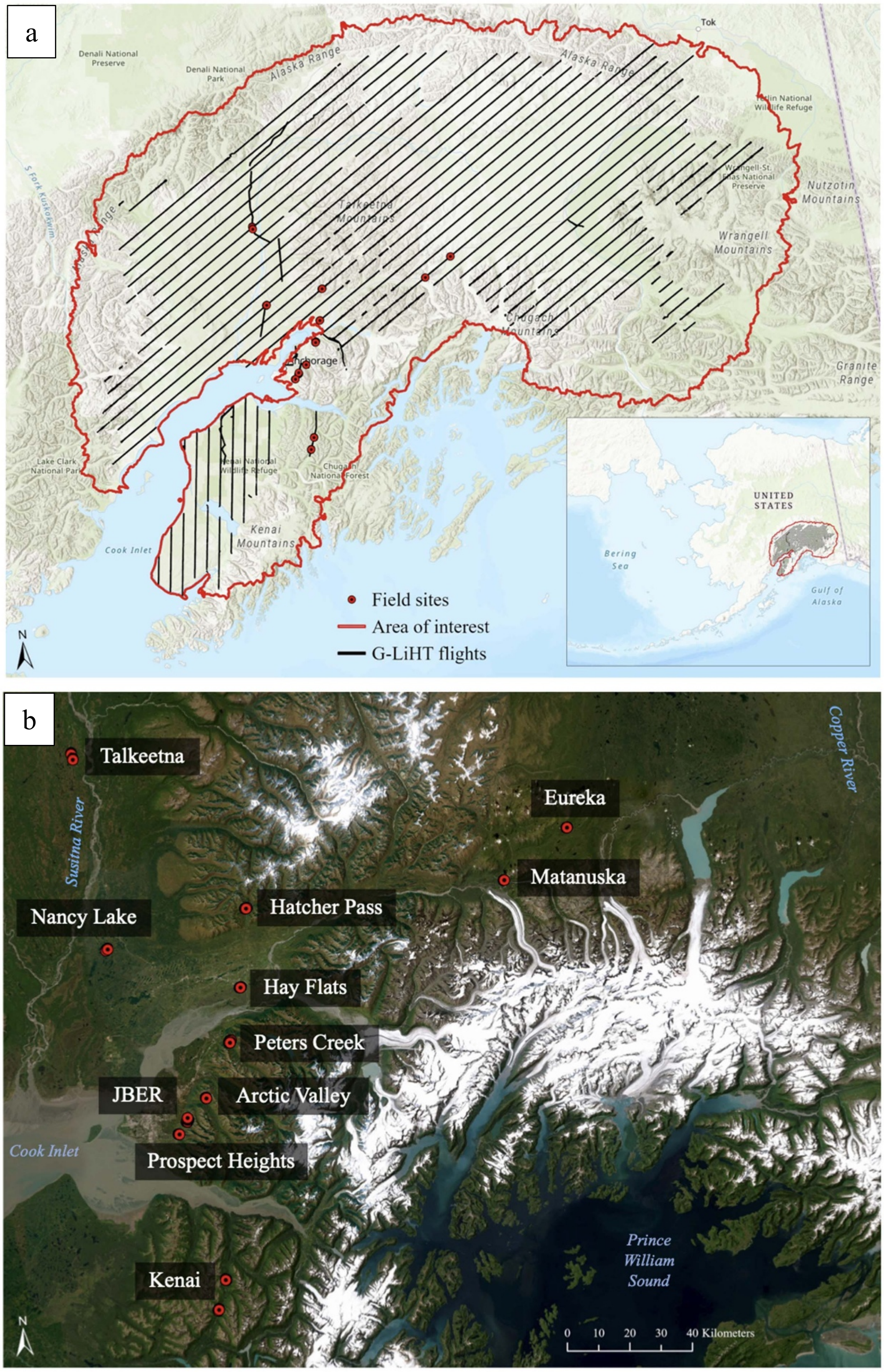

Figure 1. Area of interest (AOI) (A). G-LiHT flight lines (black) and boundary enclosing HUC12s. (B). Field sites in Susitna and Copper River basins. Red circles represent field sites.

Decimeter Handheld NMEA GNSS equipped with a Zephyr 3 Rover antenna mounted on a 2 m survey pole, we geolocated field plot centers then post-processed with GPS Pathfinder Office, differentially correcting with the nearest CORS base station, for precision error $\leqslant$0.1 m.

2.2.2. Classifying plots

The classification of all 384 plots by PFT (shrub, tree, herbaceous) was based on TRB calculated from regional tree allometry (Alonzo *et al* 2020) and on SHB calculated from regional shrub allometry (Dial *et al* 2021). Classification based on TRB and SHB

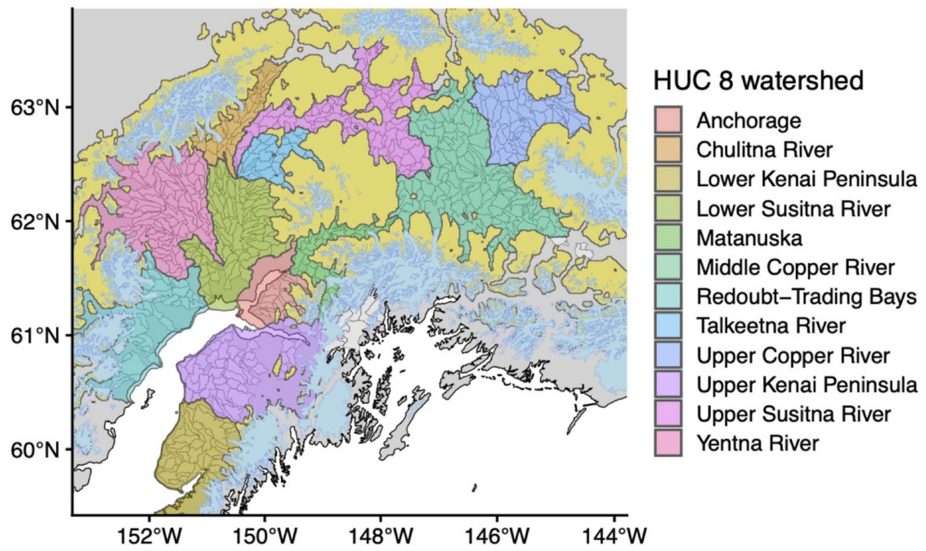

Figure 2. USGS Hydrologic Unit Code (HUC) polygons in southcentral Alaska. Smallest polygons show HUC12 boundaries as thin gray lines. Colored polygons show HUC8s with colors matched to watershed names in legend. White represents ocean waters, blue represents glaciers and perennial snowfields, tan represents elevations above 1000 m asl, and gray polygons are no-data regions.

(table 1) gave 212 shrub plots, 157 tree plots, and 15 herbaceous plots. Whereas the pSH estimates used the PFT classifications of all plots to inform RFs, the SHB estimates used as response values in GAMs were selected from shrub plots (figure 4).

2.3. Modeling

We used both topographic and canopy structural variables in modeling pSH and SHB (table S1). Predictors included G-LiHT canopy height models (CHMs, 1 m resolution) in m above ground level (m agl); G-LiHT digital terrain models (DTMs, 1 m resolution) in m above sea level (m asl); and the gridded 30 year (1980–2010) mean climatology of PRISM. To match G-LiHT data products to field plot size, we aggregated CHMs and DTMs as 13 m pixels, then applied the pSH and SHB models to calculate E[SHB] along G-LiHT flightlines. Given the flightline values of E[SHB], we attributed USGS HUC12 watershed polygons that were non-glacierized and below 1000 m asl with the means of pSH, SHB, E[SHB], and the standard error of E[SHB]. We omitted pixels above 1000 m asl because tall shrubs there are rare in southcentral Alaska (Dial *et al* 2007, 2016, Rinas *et al* 2017). These constraints left $n = 843$ HUC12 watersheds (mean area = 109 km^2, sd = 50.1 km^2) covering 9.2 million ha.

2.3.1. DTMs and CHMs

In July 2018, G-LiHT flew lidar along a systematic sample of 71 flight lines, ∼250 m wide and spaced ∼10 km apart (figure 1). The aircraft acquired airborne lidar scanning data with two 1550 nm Riegl VQ 480i instruments configured with a 300 kHz pulse repetition rate at a nominal flight altitude of 330 m. Each lidar recorded up to 8 returns per ∼10 cm diameter pulse, and a multifaceted mirror distributed the pulses evenly across the swath, producing ∼12 pulses m^{-2}. The DTMs and CHMs were generated following standard G-LiHT data processing protocols (Cook *et al* 2013) yielding a median absolute deviation that compared to ground-based GNSS validation points within 0.32 m.

From the G-LiHT Data Center (https://glihtdata.gsfc.nasa.gov, accessed 9 September 2024), we downloaded 8807 matching CHM and DTM lidar tiles as GeoTIFF rasters projected in UTM Zone 5 or 6 with WGS-84 horizontal and EGM96 vertical datums. We discarded tiles completely overlapping Cook Inlet, leaving 8756 tiles within the AOI. For model building, we used the **extract()** function in R's *raster* package (Hijmans 2020) to summarize structure (CHMs) and topography (DTMs) in field plots using the mean, maximum (max), and standard deviation (sd). For model application, we used **aggregate()** in *raster* to aggregate CHMs and DTMs as 13 m pixels (using mean, max, and sd), thereby nearly matching pixel area (169 m^2) to field plot size (168.1 m^2). The aggregated mean and max values of CHM were unable to distinguish among trees, shrubs, and herbaceous vegetation directly because G-LiHT CHM values were downloaded as rasterized, 1 m pixel averages. Instead, we used lidar and climate variables in RF classification modeling to estimate pSH, the probability that SHB exceeded TRB in each pixel equivalent to Pr(PFT) = shrub. The total area with real number pixel values (data pixels) covered 0.473 million ha, or about 3%, of the AOI.

Table 1. Study site (n = number of plots per site) characteristics.

				G-LiHT lidar		PRISM climate			
Site	n	Mean tall shrub biomass (kg m^{-2})	Range of tall shrub biomass (kg m^{-2})	Elevation range (m asl)	Max. canopy height range (m agl)	Mean growing season[b] temp. range (°C)	Mean winter[c] temp. range (°C)	Total growing season precip. range (mm)	Total winter precip. range (mm)
Kenai (KE)	60	2.1	0–12.3	391–520	1–26	9–10	−6, −8	215–245	307–394
Hatcher Pass (HP)	38	1.6	0–6.5	698–799	1–7	8–9	−7, −8	326–357	231–251
Prospect Heights (PH)	25	1.2	0.2–5.1	366–402	3–18	11	−6	165–174	138–151
Arctic Valley (AV)	28	1.1	0–7.9	615–722	1–16	8–9	−7, −8	168–178	148–184
JBER (JB)	50	1.1	0–7.6	533–755	0–15	9–10	−6, −7	176–191	155–183
Matanuska (MA)[a]	14	1.0	0–3.0	724–803	10–21	10	−13	168	67
Peters Creek (PC)	42	1.0	0–7.5	511–774	2–18	8–10	−7, −8	179–214	117–123
Eureka (EU)	35	1.0	0.1–3.1	958–980	2–14	8	−14, −15	190–200	86–87
Talkeetna (TK)	37	0.7	0–1.3	97–111	3–25	13	−9	279–280	190–192
Hay Flats (HF)	26	0.4	0.1–1.4	6–8	3–15	13	−8	167–170	95–96
Nancy Lake (NL)[a]	29	0.0	0	66–75	9–28	12	−11	211–214	147–148

[a] Site lacked shrub plots.
[b] May-August.
[c] November-February.

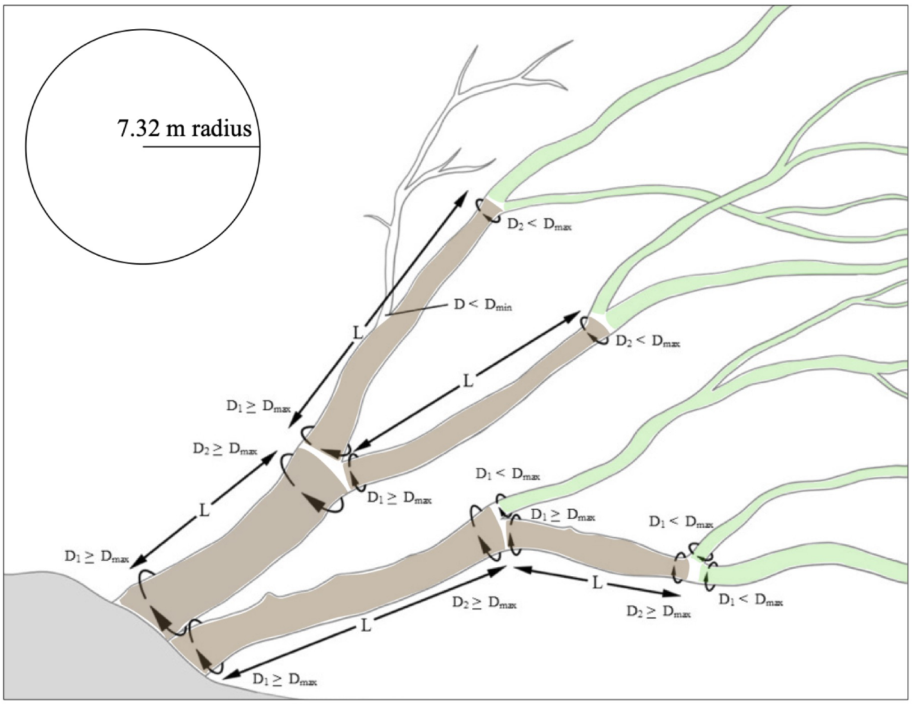

Figure 3. Field protocol for measuring tall shrubs that improves biomass estimate accuracy. Biomass allometry followed (Dial *et al* 2021b), who showed that multi-component allometry can reduce individual biomass uncertainty by 50% and plot level uncertainty by 40% over single component allometry. All shrubs with $2.5 \leqslant \mathrm{DRC} < 10.2$ cm $= D_{max}$ were considered as 'aerial tips'. Shrub aerial tip dry biomass B (kg) was calculated as $B_{tip} = 0.56 \times 0.07\, D^{2.437}$, where $D = \mathrm{DRC}$. Shrubs with $\mathrm{DRC} \geqslant 10.2$ cm $= D_{max}$ were considered two-component structures: aerial tips with node diameter $D \leqslant 10.2$ cm and stem frustra with bounding nodes of $D > 10.2$ cm. Each internode between the aerial tip and the ground with diameter $D > 10.2$ cm $= D_{max}$ was considered a frustrum, measured for its length L and two diameters D_1 and D_2. Dry biomass of an internode with node diameters D_1 and D_2, both $\geqslant 10.2$ cm was calculated as $B_{frustrum} = 0.44 \times \mathrm{L}\,({D_1}^2 + {D_2}^2 + D_1 D_2)/12$. Green colored segments use aerial tip allometry. Brown colored shrub segments use frustrum allometry. Inset shows plot design. Illustration courtesy of Julia Ditto.

2.3.2. Gridded climate data

In addition to topographic and structural variables, we used climatic variables (table S1) for modeling. Raster layers of PRISM gridded climatology were downloaded (means of 1981–2010, 800 m pixel resolution https://prism.oregonstate.edu/projects/public/alaska/grids/, accessed 9 September 2024) as monthly and annual mean temperature (°C) and total precipitation (mm). Using ArcGIS Pro (ESRI Version 3.0.3), rasters were reprojected to UTM Zone 5 or UTM Zone 6 and the 800 m PRISM pixels were disaggregated such that each 13 m pixel within a single 800 m pixel received the same climatic variable value. Correlated climatic variables (Pearson correlation $r \geqslant 0.7$) were combined as a single variable, with temperature variables averaged and precipitation variables summed over sequential months.

2.4. Modeling E[SHB]

PFT classification and SHB in field plots provided the response variables in pSH and SHB models (table S1) respectively. Models describing pSH and SHB were first explored using an array of transformations and variable combinations within a variety of modeling frameworks (GLMs, GAMs, RFs). We investigated models with interactions, but those sometimes led to improbable predictions outside the range of observations and so were abandoned. Through an iterative process, we evaluated model accuracy of RF classifiers as models of pSH, and of GAMs as models of SHB. We calculated E[SHB] (equation (1)) within each 13 m data pixel of each G-LiHT tile. Pixel values of SHB, pSH, and E[SHB] were then each averaged if overflown by at least $n = 30$ tiles. In addition, we calculated the uncertainty of E[SHB] for each watershed (see *2.5. Model application and uncertainty* below).

2.4.1. Modeling pSH

We constructed RF classification models from plot PFTs ($n = 394$) defined by the relative biomass of trees and shrubs (figure 4); in application, we used only pSH, the probability that SHB exceeded TRB. The plot-level PFT classifications informed the RF model response variable; lidar and climate metrics at each plot acted as predictors. Among RF candidates, the final model was selected based on its goodness of fit metrics out-of-bag error and multi-class area under the receiver operating characteristic curve, conditional as the model with the fewest number of

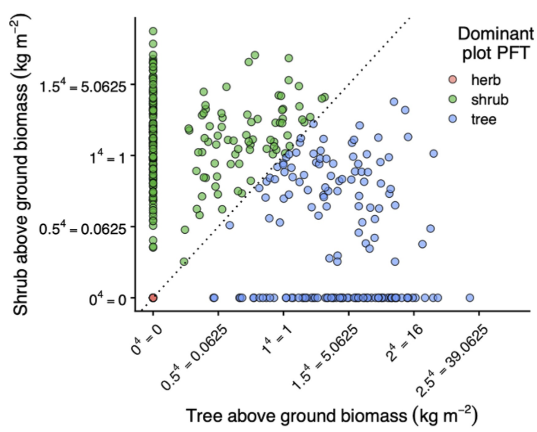

Figure 4. Total aboveground biomass in field plots for shrubs (SHB) vs. trees (TRB). Each point represents a field plot. The black line (slope = 1, intercept = 0) represents where plots with shrub biomass equals tree biomass. Points above the line were considered shrub plots ($n = 212$). Points below the line were considered forest plots ($n = 157$). Herbaceous plots were considered points where both shrub and tree basal areas were zero ($n = 15$).

variables with fit better than or equal to other models (parsimonious candidate). The parsimonious candidate provided pSH for each pixel.

In southcentral Alaska, shrubs are generally <6 m tall (Alonzo *et al* 2020). Thus, we set pSH = 0 where 13 m mean CHM pixels values exceeded 6.5 m agl. We lacked both sufficient field data to parameterize a continuous model of height thresholds and sufficient lidar data to model individual plant bodies. Thus, mean CHM could fall slightly below 6.5 m within a pixel dominated by young or otherwise short-statured trees, yet receive a high pSH value. Separating shrubs from trees is the central problem in boreal forest SHB estimates. Future studies should apply more data-rich lidar products to better discern shrubs from trees.

2.4.2. Modeling SHB

We used only shrub plots ($n = 212$) to construct models of SHB because including forest and herbaceous plots led to inaccurate, biased fits to observed plot data. SHB models used field-calculated biomass as response and lidar and climate metrics as predictor variables (table S1). The mean and maximum of CHMs, and of DTMs, were often correlated, but because our goal was prediction, collinear lidar variables were included to improve model fit. We selected a single model for application based on R^2, root-mean-squared-error (RMSE), Akaike information criterion (AIC), and parsimony. Models with interaction smooths performed well but produced unrealistic predictions when applied to data tiles. We abandoned those for additive-only smooths.

2.5. Model application and uncertainty

HUC12 estimates of pSH, SHB, E[SHB], and the uncertainty of E[SHB] were mapped across the AOI. Shapefiles for HUC12s in the AOI were downloaded (www.usgs.gov/national-hydrography/access-national-hydrography-products, accessed 9 September 2024) and reprojected to UTM Zone 5 and UTM Zone 6 coordinate reference systems in ArcGIS Pro. We first calculated SHB and pSH for each data pixel, then applied equation (1). Pixel values were averaged by tile for tiles with $n \geqslant 20$ data pixels ($n = 6424$ tiles). pSH, SHB, and E[SHB] were averaged for each HUC12 watershed with $m \geqslant 30$ tiles as the mean of tile values. Using a state-wide glacier shapefile in a GIS, we masked glacier-covered landscapes, often below 1000 m asl, because field plots did not include glacier surfaces.

Using the fact that the variance of a sum is the sum of the variances and applying the Central Limit Theorem, we calculated uncertainty (SE_{huc}) associated with E[SHB] within each HUC12 watershed. The uncertainty propagated neither errors of allometry, nor errors inherent in the RF estimate of pSH. Instead, calculation of SE_{huc} used the standard error (SE_i) of each pixel's GAM predicted ABG, the number of pixels in each tile (n) within the m tiles in a HUC12, and the variance of errors in the model E[SHB] as residuals^2:

$$SE_{huc} = \frac{\sqrt{\mathrm{var}\,(\mathrm{HUC12})}}{m}, \tag{2}$$

where

$$\mathrm{var}\,(\mathrm{HUC12}) = \sum_{j \in \mathrm{tiles}}^{m} \mathrm{var}\left(\mathrm{tile}_j\right),$$

and

$$\mathrm{var}\left(\mathrm{tile}_j\right) = \sum_{i \in \mathrm{pixels}}^{n} SE_i^2 + n\ \mathrm{residuals}^2.$$

2.6. Comparisons to deciduous shrub top cover

Macander *et al* (2022) mapped probabilities of seven PFTs as 'top cover' in 30 m pixels across Alaska by training two machine-learning tree-based models using a suite of environmental and spectral covariates. Of their seven PFTs, top cover of deciduous shrubs had the highest RMSE (19%). One machine-learning model estimated PFT probability; the second estimated abundance as top cover. Most Macander *et al* training data (https://akveg.uaa.alaska.edu/, accessed 9 September 2024) in southcentral Alaska relied on low-flying helicopter surveys (Nawrocki *et al* 2020). Assuming that top cover of deciduous shrubs and their aboveground biomass are related, we downloaded Macander *et al*

ABoVE_PFT_Top_Cover_DeciduousShrub_2020.tif (https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=2032, accessed 9 September 2024), a raster giving the probability that a 30 m pixel supported deciduous shrubs as top cover (pTC) in 2020. We extracted pTC values from pixels below 1000 m asl, found the mean pTC by G-LiHT tile for HUC12s with 5 tiles or more ($n = 647$ HUC12s), then correlated tile mean pTC to corresponding mean E[SHB].

3. Results

3.1. Field plots

Shrub plots, those with aboveground shrub biomass (SHB) exceeding TRB (figure 4), made up 55% of all field plots (tables 1 and 2). When averaged across all shrub plots ($n = 212$ plots), the mean ± sd of SHB was 2.2 ± 2.1 kg m^{-2} (22 ± 21 Mg ha^{-1}) and TRB was 0.18 ± 0.45 kg m^{-2} (1.8 ± 4.5 Mg ha^{-1}). In contrast, averaged across all tree plots, SHB (0.34 ± 0.58 kg m^{-2}, 3.4 ± 5.8 Mg ha^{-1}, $n = 157$ plots) was about 15% that of shrub plots and TRB (5.9 ± 5.5 kg m^{-2}, 59 ± 55 Mg ha^{-1}) was over 30 times that found in shrub plots.

3.2. pSH model

The final RF classifier of PFTs in field plots ($n = 384$ plots) included mean and max canopy height and winter temperature as predictors (table 3). Maximum canopy height was the most important variable (mean decrease Gini = 70.3), followed by mean canopy height (64.6), and winter temperature (52.9). The model predicted shrub PFT most accurately (class error = 0.08); followed by forest PFT (0.19); and herbaceous PFT (0.67), accuracies that were reflected in the multi-class AUROC scores (table 4).

We applied the RF classifier to all 13 m data pixels below 1000 m asl to calculate pSH, then averaged pSH as the mean within each tile with ⩾ 20 pixels (0.31 ± 0.20, $n = 6450$ tiles), and finally averaged the mean tile pSH to attribute HUC12 values (figure 5(a)). Considering HUC12-level pSH by HUC8s as subregions (figure 2) suggested that coastal areas surrounding Cook Inlet (Anchorage, Redoubt-Trading Bays, and the Upper and Lower Kenai Peninsula) supported above average pSH. Higher pSH than average also occurred in mountain valleys. Lower pSH occurred in low elevation drainages, with low elevation watersheds in the Copper River basin supporting a higher pSH than low elevation watersheds in the Susitna River basin. Across the 843 HUC12s, pSH was 0.33 ± 0.16.

3.3. *SHB* model

The top SHB models in terms of R^2, RMSE and AIC included CHM (structural) and PRISM (climatic) variables (table 3). Using observations from shrub PFT plots, the selected GAM with SHB as response variable included mean and maximum measures of structural predictors, as well as November–February (winter) and May-August (growing season) climate predictors. The second lowest AIC (ΔAIC = 5.4) also included both structural and elevation variables, but not climate variables. However, the model with the third lowest AIC (ΔAIC = 5.6) included two structural and a single climate predictor and displayed higher R^2 and lower RMSE than the second-lowest ΔAIC, supporting the approach that well-performing regional models include both structural and climatic predictors.

The selected GAM of SHB was applied to all 13 m data pixels below 1000 m asl and averaged by tile (17.7 ± 9.9 Mg ha^{-1}, $n = 6264$ tiles). To map SHB across the AOI, we averaged mean tile SHB across tiles within HUC12s (figure 5(b)). Defining subregions as HUC8s (figure 2), then mapping HUC12-level SHB, suggested that unglaciated coastal areas and most of Copper River and the upper Susitna River basins supported higher SHB–while the remainder of the Susitna basin had lower–SHB than average. Across the $n = 843$ HUC12s, overall mean SHB was 18.5 ± 8.0 Mg ha^{-1}, which are overestimates of actual tall shrub biomass because the SHB model cannot fully distinguish between trees and shrubs.

3.4. E[SHB] model

Mean E[SHB] per tile with ⩾ 20 pixels calculated using equation (1) for every data pixel below 1000 m asl was 5.9 ± 5.4 Mg ha^{-1}, $n = 6424$ tiles (figure 5(c)). We treated tile means as samples for HUC12s, finding the mean and sd of E[SHB] averaged across $m = 843$ HUC12 values as 6.6 ± 4.6 Mg ha^{-1}. Within subregions (figure 2), higher probabilities of tall shrubs (pSH) tended to match greater values of E[SHB]. However, whereas pSH (figure 5(a)) was high in the Upper Susitna River, Middle and Upper Copper River HUC8 watersheds, E[SHB] was low (figure 5(c)). In contrast, the Lower Kenai Peninsula and Redoubt-Trading Bays west of Cook Inlet each had high pSH and high E[SHB], while Yentna River had both low pSH and E[SHB]. The Lower Kenai Peninsula was notable for its high uncertainty in E[SHB] (figure 5(d)).

3.5. E[SHB] uncertainty

Preliminary analysis showed performance improved if using shrub plots (i.e. those with SHB > TRB) alone to construct SHB models. For example, using a 'one-model-approach' ($n = 384$ plots) to construct a GAM model of SHB using the same variables as the best performing model (table 3) gave RMSE = 206 kg plot^{-1} after setting all negative SHB predictions to zero kg plot^{-1}. There was a strong positive bias in the one-model-approach that over-predicted plot SHB (figure 6(a)). In particular, the median difference between modeled and observed plot SHB in the one-model-approach was 14.3 kg plot^{-1} and the absolute

Table 2. Number of plots per site per vegetation classification with measured species.

Site	Plots	Shrub	Forest	Herbaceous	Shrub genera	Tree species
KE	60	30	26	4	*Alnus, Salix*	*Betula spp., Picea glauca, Picea sitchensis, Populus spp., Salix spp., Tsuga spp.*
HP	38	32	0	6	*Alnus, Salix*	No forest plots
PH	25	23	2	0	*Alnus*	*Picea glauca, Betula spp.*
AV	28	18	6	4	*Alnus, Salix*	*Populus spp.*
JB	50	43	6	1	*Alnus, Salix*	*Betula spp., Picea glauca, Populus spp., Salix spp., Tsuga spp.*
MA	14	0	14	0	*Alnus, Salix*	*Betula spp., Picea glauca, Populus spp., Salix spp.*
PC	42	30	12	0	*Alnus, Salix*	*Betula spp., Picea glauca, Populus spp., Salix spp.*
EU	35	18	17	0	*Salix*	*Picea glauca, Picea mariana*
TK	37	1	36	0	*Alnus*	*Betula spp., Picea glauca, Picea mariana*
HF	26	17	9	0	*Alnus, Salix*	*Betula spp., Picea glauca*
NL	29	0	29	0	*No shrub plots*	*Betula spp., Picea glauca, Picea mariana*
Totals	384	212	157	15		

Table 3. Variables included and performance statistics for top five GAM SHB models.

Mean canopy height	Max. canopy height	Mean elevation	Max elevation	Winter precipitation	Growing season precipitation	Winter temperature	R^2	RMSE (kg m^{-2})	ΔAIC
x	x				x	x	0.663	1.18	0
x	x			x			0.652	1.21	5.6
x	x	x	x				0.648	1.22	5.4
x	x	x					0.644	1.23	9.4
x	x		x				0.641	1.24	9.3

x indicates $p(F) < 0.005$.

Table 4. Variables included and performance statistics for top five RF classifier models.

Mean canopy height	Max. canopy height	Mean elevation	Growing season precipitation	Winter temperature	Total Out-of-bag error (%) Total	Shrub	Forest	Herb.
x	x	x		x	14	8.0	17.8	53.3
x	x			x	14	8.0	16.5	60.0
x	x	x	x	x	14	9.4	16.5	53.3
x	x	x			16	9.9	19.1	53.3
	x	x			17	11.8	17.8	73.3

difference between median observed and median predicted was 66.1 kg $plot^{-1}$.

The two-part-model approach as E[SHB] informed by $n = 212$ plots for SHB and $n = 384$ plots for pSH then applied to the $n = 384$ plots improved overall accuracy over the one-model-approach by about 10 kg $plot^{-1}$ (RMSE = 196.8 kg $plot^{-1}$), reduced interquartile distance of errors by 30%, and removed bias (figure 6(a)). The median of the difference between E[SHB] and observed plot ABG was 0.0 kg $plot^{-1}$. The absolute difference between median observed plot SHB and median E[SHB] was 7.2 kg $plot^{-1}$, 58.9 kg $plot^{-1}$ less than the difference in median observed and modeled using the one-model-approach. The residual standard error was 195.3 kg $plot^{-1}$, an inflated value because shrub plots had pSH $\leqslant 1$ in the calculation of E[SHB]. The application of equation (2) provided the spatial uncertainty of the E[SHB] estimates (figure 5(d)).

E[SHB] tended to overpredict plot SHB (figure 6(b)) in herbaceous plots (median error = 54.1 kg $plot^{-1}$, RMSE = 86.5 kg $plot^{-1}$), lacked any bias in tree plots (median = 0.0 kg $plot^{-1}$, RMSE = 132 kg $plot^{-1}$), and underpredicted plot SHB in shrub plots (median error = −29.0 kg $plot^{-1}$, RMSE = 236.0 kg $plot^{-1}$). Setting pSH = 1 for shrub plots in calculation of E[SHB] reduced bias and RMSE (median error = 4.0 kg $plot^{-1}$, RMSE = 205.0 kg $plot^{-1}$) in shrub plots.

3.6. Comparisons to deciduous shrub top cover

Overall, half of correlations between tile mean pTC (figure 7) and tile mean E[SHB] by HUC12 were strongly positive (median $r = 0.59$; figure 8(a)) but 18% were negative. Watersheds north and west of Cook Inlet tended to have the highest correlations, whereas watersheds northeast of Cook Inlet (Copper

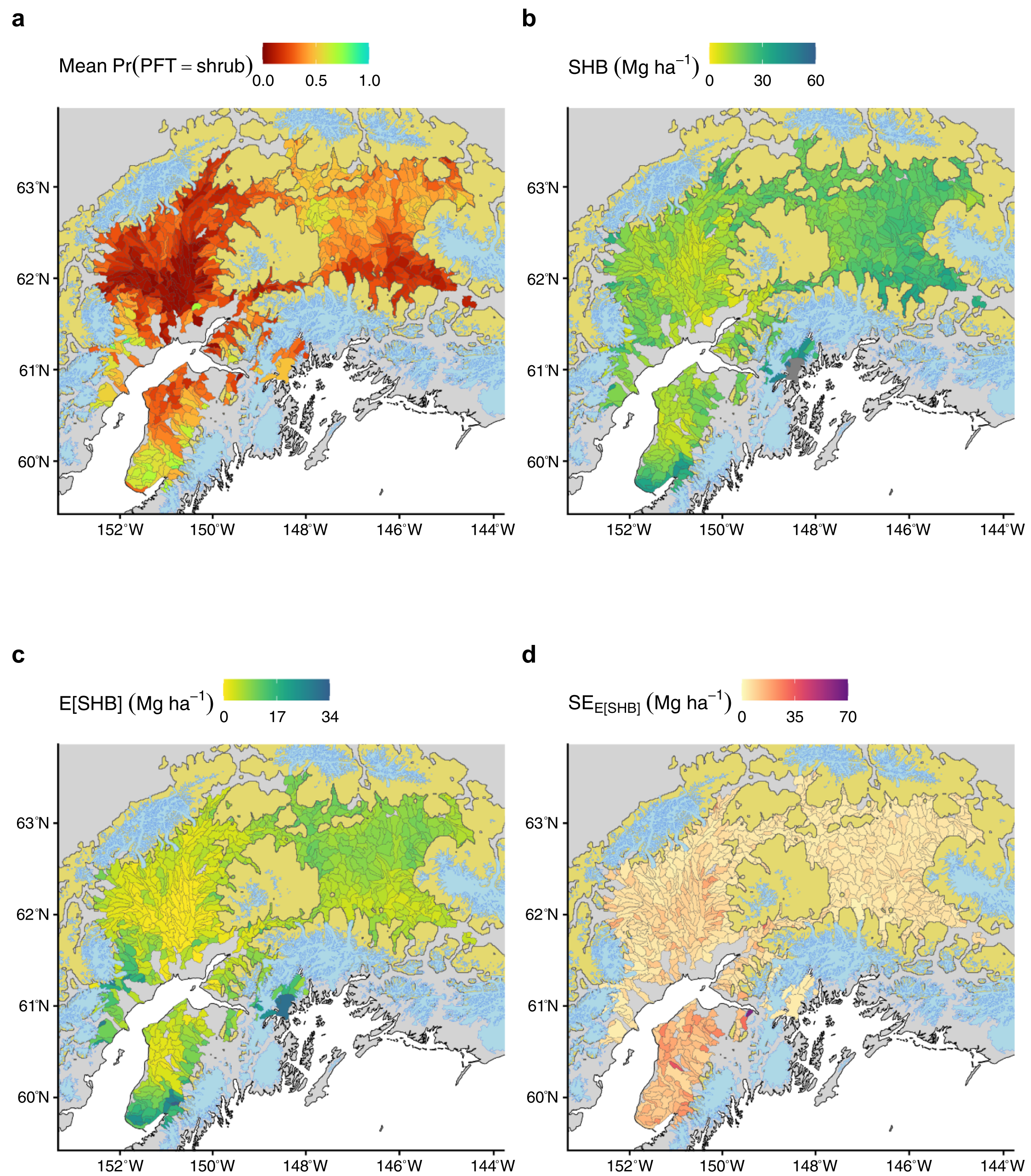

Figure 5. Tall shrub modeling across southcentral Alaska by hydrologic units (HUC12s). (a). Mean probability of tall shrub biomass (pSH). (b). Mean shrub above-ground biomass (SHB) assuming shrub biomass exceeds tree above-ground biomass (TRB) as SHB > TRB. (c). Mean expected tall shrub biomass (E[SHB]) calculated using equation (1) as E[SHB] = pSH × SHB. (d). Standard error of E[SHB] calculated using equation (2). White represents ocean waters, blue represents glaciers and perennial snowfields, and tan represents elevations above 1000 m asl. HUC12 boundaries are thin gray lines while gray polygons are no-data regions.

River, Upper Susitna) tended to have the lowest (figures 6(a) and (8b)).

4. Discussion

This study applied regional allometry to trees (Alonzo *et al* 2020) and shrubs (Dial *et al* 2021b) in 384 field plots to calculate aboveground biomass of trees and shrubs. These plots were integrated with strip samples of airborne lidar data and a gridded climate dataset to estimate aboveground shrub biomass at the scale of USGS HUC12 watersheds across a 16.6 million ha boreal region in southcentral Alaska. Selected predictor variables included canopy height, elevation, temperature, and precipitation to capture the regional determinants of shrub distribution and the local determinants of abundance, as used by other mapping studies of vegetation in Alaska (Alonzo *et al* 2020, Shoot *et al* 2021, Macander *et al* 2022, Orndahl *et al* 2022, Seider *et al* 2022). We found that, when averaged across 843 HUC12s, the expected aboveground biomass (E[SHB] in equation (1)) was 6.6 ± 4.6 Mg ha^{-1}(0.66 ± 0.46 kg m^{-2}).

Currently the national FIA monitoring program in the USA does not use shrubs in calculations of aboveground biomass or carbon (Barrett and Gray

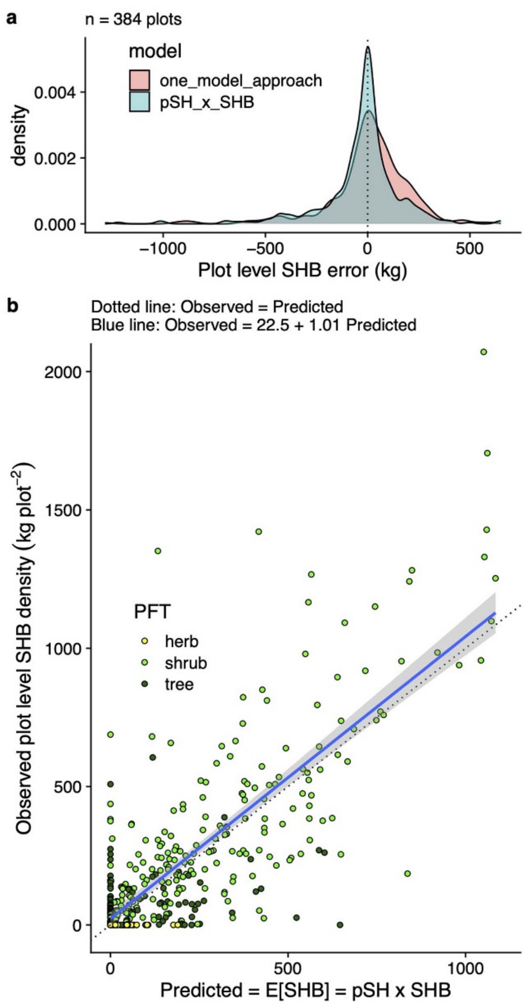

Figure 6. Plot-level biomass error of tall shrub biomass using E[SHB]. (a). Density distributions of model error defined as fitted minus observed tall shrub biomass in field plots. Blue density curve (Simple_SHB) gives error distribution of an 'all-plots-approach', where all $n = 384$ field plots inform a single-part model (GAM) predicting aboveground tall shrub biomass. The red density curve (pSH_SHB) gives the error distribution of the same $n = 384$ plots, but uses the two-part model E[SHB] = pSH × SHB, whereby tall shrub biomass is calculated as the product of a model giving the probability that the biomass of shrubs in a plot exceeds the biomass of trees in the plot (pSH model trained on all $n = 384$ plots) multiplied by a model giving tall shrub biomass in a plot where the tall shrub biomass exceeds the tree biomass in the plot (SHB model fit using $n = 212$ plots where tall shrub biomass > tree biomass). (b). Scatterplot of observed vs. predicted shrub biomass per plot shrub biomass per plot using the two-part model.

2011, Alonzo *et al* 2020), even though in some places, like southcentral Alaska, tall shrubs contribute substantially to aboveground biomass. Moreover, evidence of tall shrub expansion into wetlands (Klein *et al* 2005, Berg *et al* 2009) and sub-alpine communities (Dial *et al* 2007, 2016, Rinas *et al* 2017) in southcentral Alaska underscores the need for baseline information on shrub biomass to assess further changes in vegetation structure and composition in response to climate change. Finally, much of Alaska–like most of the planet's high-latitude ecosystems that are located distant from roads–are more economically sampled using remote sensing platforms like G-LiHT.

4.1. Two-part modeling

Two-part modeling approaches are usefully applied to taxa growing both as under- and overstory plants (Nawrocki *et al* 2020, Macander *et al* 2022, Travers-Smith *et al* 2024). Because trees confound and obscure remote sensing of tall shrubs in forest and woodland communities (Macander *et al* 2022), we applied a two-part modeling approach to predict shrub distribution as the probability (pSH) that shrub aboveground biomass surpassed tree biomass and–given biomass dominance by shrubs–predicted shrub abundance as aboveground biomass (SHB). The product of these quantities, E[SHB] = pSH × SHB, improved accuracy and reduced bias. Our definition of pSH is that it identifies the probability that shrub wood volume surpasses tree wood volume, given climate, canopy structure, and topography, rather than as a percent cover on the landscape or the chance that a random sample contains a tall shrub.

One outcome of the two-part modeling approach here was that shrub abundance was not overestimated, particularly in forests where field plots indicated low shrub abundance and where rasterized lidar easily confounds canopy structure of trees and shrubs. In field plots, the 95-percentile of TRB (122 kg m^{-2} = 1220 Mg ha^{-1}) was 20 times the 95-percentile of SHB (5.5 kg m^{-2} = 55 Mg ha^{-1}). Thus, a watershed-level estimate reaching 60 Mg ha^{-1} (SHB, figure 5(b)) is far less likely than one reaching only 30 Mg ha^{-1} (E[SHB], figure 5(c)); that is, there are likely no areas as large as a HUC12 watershed with a mean aboveground shrub biomass density surpassing the maximum observed plot-level SHB. Overall mean watershed E[SHB] (6.6 Mg ha^{-1}), was about one-third overall mean watershed SHB (18.5 Mg ha^{-1}), a fraction near the watershed mean pSH (0.33).

Based on our modeling of tall shrubs at the scale of HUC12 watersheds, the overall spatial patterns of shrubs as dominant PFT (figure 5(a)), and abundance as biomass (figure 5(b)) appear to differ across southcentral Alaska. SHB is relatively high in the Copper River basin but pSH is low, giving E[SHB] as moderate (figure 5(c)) with relatively low uncertainty (figure 5(d)). In contrast, both SHB and pSH are relatively high in the upper Susitna River basin, also with relatively low uncertainty. The Kenai Peninsula displays high E[SHB] but relatively high uncertainty (figure 6(c)).

4.2. Estimate comparisons

We compare our modeling efforts to those of a localized landscape estimate of shrub abundance (Alonzo *et al* 2020) and to a broad-scale gridded product of deciduous shrubs as top cover (Macander *et al* 2022). Alonzo *et al* (2020) modeled portions of two HUC12s within the Anchorage HUC8

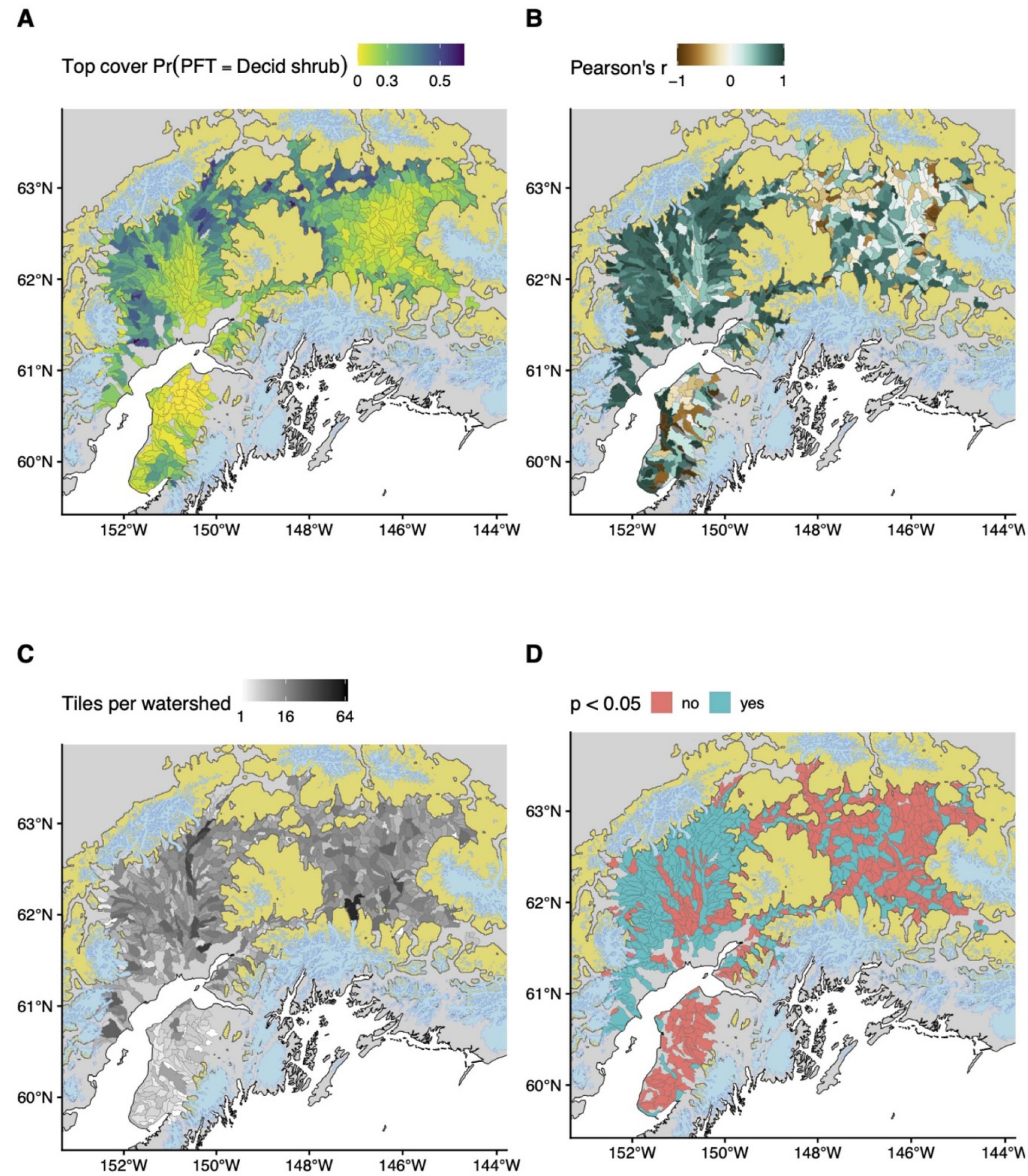

Figure 7. Mean probability of deciduous shrub top cover in southcentral Alaska by hydrologic units (HUC12s). (A). Mean probability per HUC12 watershed of deciduous tall shrubs as top cover in 2020 (pTC) extracted from Macander *et al* (2022). (B). Correlation coefficient of tile means of pTC and E[SHB] averaged by HUC12s. (C). Tile count per HUC12 with both pTC and E[SHB] tiles with $\geqslant 20$ pixels. (D). P-value corresponding to correlation in B given sample size in C. White represents ocean waters, blue glaciers and perennial snowfields, tan elevations above 1000 m asl. HUC12 boundaries are thin gray lines while gray polygons are no-data regions.

using 168.1 m^2 field plots and a subset of G-LiHT used here. They reported the best lidar-based model as a Gaussian GLM of SHB regressed against seven variables, attaining $R^2 = 0.77$ and an RMSE ($1.2\ kg\ m^{-2} = 201.7\ kg\ plot^{-1}$) similar to that found here. Their mapping produced an SHB estimate (mean $\pm$ 95%CI $= 2.56 \pm 1.8\ kg\ m^{-2}$) that is greater, but, due to high uncertainties, overlapping with E[SHB] estimates of the HUC12s sampled here: Peters Creek ($0.78 \pm 0.93\ kg\ m^{-2}$) and Ship Creek ($0.42 \pm 2.0\ kg\ m^{-2}$).

At the regional scale (Macander *et al* 2022), about half of HUC12 watersheds showed positive correlations between E[SHB] and deciduous shrubs as top cover (pTC), whereas almost 20% of watersheds showed negative correlations (figure 8(a)). While negative correlations may indicate shrub understories yielding high E[SHB] but low pTC, the spatial variability in correlation of E[SHB] with pTC (figures 7(b) and 8(b)) may also reflect different sampling efforts between studies. Macander *et al* under-sampled the Anchorage area, whereas we under-sampled

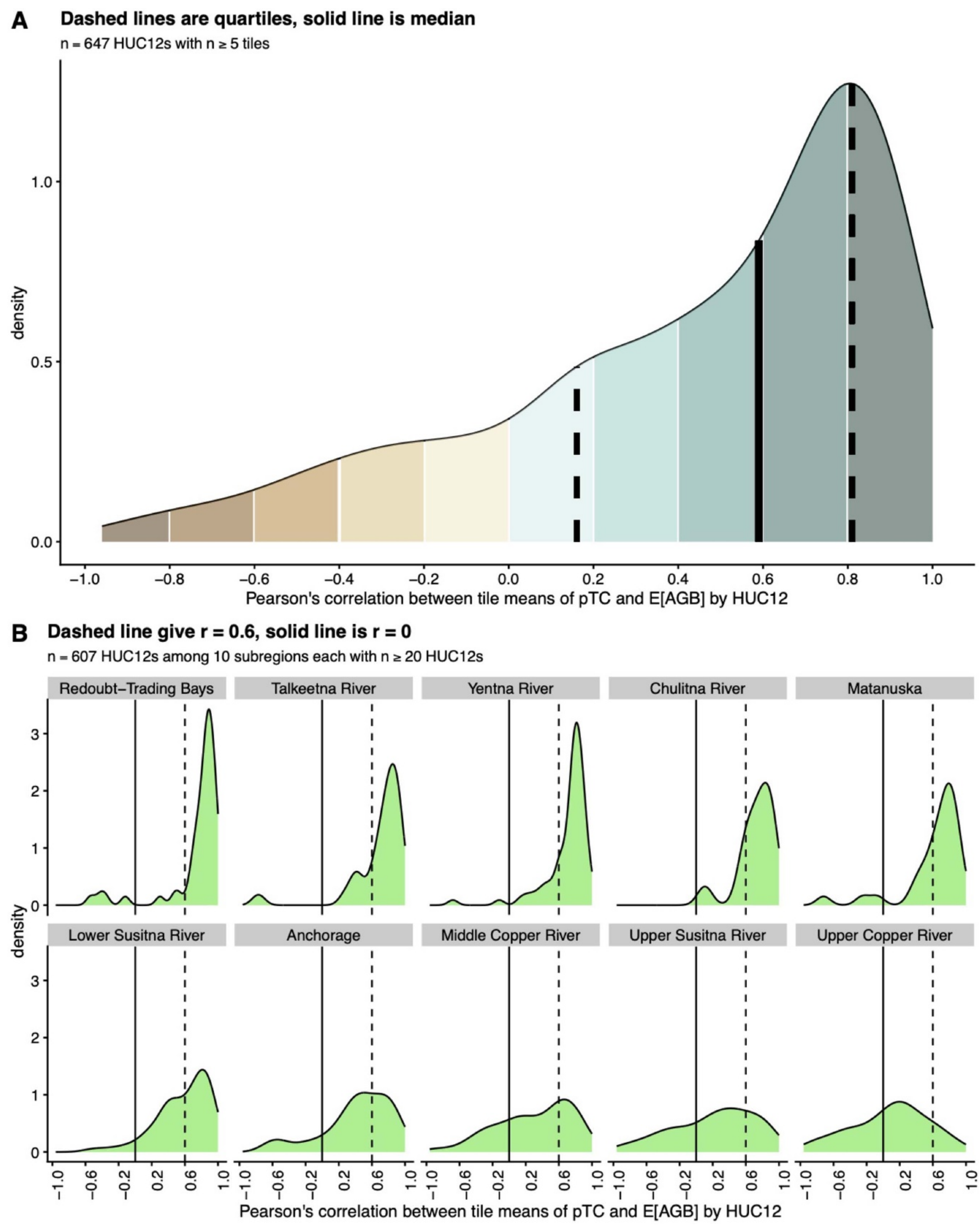

Figure 8. Correlations between E[SHB] and pTC were mostly strong ($r \geqslant 0.6$). (A). Distribution of correlations between tile means of E[SHB] and pTC within 647 HUC12 watersheds in southcentral Alaska, where each HUC12 contained $n \geqslant 5$ G-LiHT tiles. (B). Distribution of correlations between tile means of E[SHB] and pTC within 607 HUC12 watersheds by HUC8 subregion (figure 2), where each HUC8 contained $m \geqslant 20$ HUC12s and each HUC12 contained $n \geqslant 5$ G-LiHT tiles.

the Copper and Upper Susitna river basins. Further refinements of both modeling approaches seem warranted.

5. Conclusions

Shrub abundance in Alaska suggests that monitoring programs inventory tall shrubs if carbon sequestration is of interest. Because topography and climate largely determine the distribution and abundance of vegetation across boreal landscapes (Roland *et al* 2017, Macander *et al* 2022), these variables perform well in modeling tall shrub biomass. We find comparisons between this and more data-rich studies encourage the application of climate data to capture regional variation in growing conditions, vegetation structure to capture landscape differences, and elevation metrics to capture both.

Shrubification in southcentral Alaska is spatially variable and has occurred at different rates

(Klein *et al* 2005, Dial *et al* 2007, 2016, Berg *et al* 2009, Rinas *et al* 2017, Macander *et al* 2022), making it important to understand the relationship between shrubs and climate change for predicting shrub expansion over time. The positive response of pSH with increasing winter temperatures above −10 °C suggests that warming will increase distribution upward into subalpine ecosystems (Dial *et al* 2007, 2016, Rinas *et al* 2017) and outward into drying wetlands (Klein *et al* 2005, Berg *et al* 2009), increases with multiple ecosystem consequences (Mekonnen *et al* 2021).

To accurately capture shrub biomass trends, field and modeling methods will best inform when matched with lidar repeated across the G-LiHT flight paths. Because tall shrubs grow rapidly and respond quickly to disturbance, deconvolution methods that separate growth from disturbance (Dial *et al* 2021a) could characterize trends. Moreover, more precise estimates of TRB based on height and DBH collocated with shrub biomass would improve above-ground carbon estimates. In Alaska, previous studies have used hyperspectral and multispectral data to obtain site-specific spectral indices for estimating shrub biomass (Berner *et al* 2019, Alonzo *et al* 2020), classifying forest type (e.g. Shoot *et al* 2021), and assessing vegetation composition (e.g. Berner *et al* 2019, Macander *et al* 2022, Orndahl *et al* 2022). These data and more sophisticated structural lidar products than rasters would inform future tall shrub biomass modeling.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.5061/dryad.t4b8gtj95. Data will be available from 05 May 2025.

Acknowledgments

This study was supported by USFS joint venture agreements 19-JV-11261989-062 and 19-JV-11261989-112 to Roman Dial. We thank the summer field teams members Allen Dahl, Julia Ditto, Russell Wong, Scout Donahue, Toshio Matsuoka, Nikki Castro-Morales, Madeline Sadler, Kela Vicich, Amy Wockenfuss, and Madeline Zietlow for their efforts collecting data and Jason E Geck for GIS advice.

Author contributions

Conceptualization: RJD, HEA, BDC, DCM
Data curation: KAM, RJD, BDC
Formal analysis: KAM, RJD
Funding acquisition: HEA, BDC, DCM, RJD
Investigation: HEA, BDC, DCM, RJD, KAM
Methodology: HEA, BDC, DCM, RJD, KAM
Project administration: RJD, KAM, HEA
Resources: HEA, BDC, DCM, RJD, KAM
Supervision: RJD, KAM, HEA, EIL
Validation: RJD, KAM
Visualization: RJD, KAM
Writing—original draft: KAM, RJD
Writing—review & editing. KAM, RJD, HEA, EIL, BDC, DCM

ORCID iDs

Kaili A Martin https://orcid.org/0009-0003-5191-2253
Roman J Dial https://orcid.org/0000-0002-8189-7248
Hans-Erik Andersen https://orcid.org/0000-0002-5546-9522

References

Alonzo M, Dial R J, Schulz B K, Andersen H E, Lewis-Clark E, Cook B D and Morton D C 2020 Mapping tall shrub biomass in Alaska at landscape scale using structure-from-motion photogrammetry and lidar *Remote Sens. Environ.* **245** 111841

Babcock C *et al* 2018 Geostatistical estimation of forest biomass in interior Alaska combining Landsat-derived tree cover, sampled airborne lidar and field observations *Remote Sens. Environ.* **212** 212–30

Barrett T M and Gray A N 2011 Potential of a national monitoring program for forests to assess change in high-latitude ecosystems *Biol. Conserv.* **144** 1285–94

Berg E E, Hillman K M, Dial R and DeRuwe A 2009 Recent woody invasion of wetlands on the Kenai peninsula lowlands, south-central Alaska: a major regime shift after 18 000 years of wet sphagnum-sedge peat recruitment *Can. J. For. Res.* **39** 2033–46

Berner L T, Jantz P, Tape K D and Goetz S J 2019 Tundra plant above-ground biomass and shrub dominance mapped across the north slope of Alaska *Environ. Res. Lett.* **13** 35002

Box J E *et al* 2019 Key indicators of Arctic climate change: 1971–2017 *Environ. Res. Lett.* **14** 4

Bradshaw C J A and Warkentin I G 2015 Global estimates of boreal forest carbon stocks and flux *Glob. Planet. Change* **128** 24–30

Chapin F S, Sturm M, Serreze M C, McFadden J P, Key J R, Lloyd A H and Welker J M 2005 Role of land-surface changes in arctic summer warming *Science* **310** 657–60

Cook B D, Nelson R F, Middleton E M, Morton D C, McCorkel J T, Masek J G, Ranson K J, Ly V and Montesano P M 2013 NASA Goddard's lidar, hyperspectral and thermal (G-LiHT) airborne imager *Remote Sens.* **5** 4045–66

Daly C, Halbleib M, Smith J I, Gibson W P, Doggett M K, Taylor G H, Curtis J and Pasteris P P 2008 Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States *Int. J. Climatol.* **28** 2031–64

Daly C, Neilson R P and Phillips D L 1994 A statistical-topographic model for mapping climatological precipitation over mountainous terrain *J. Appl. Meteorol. Climatol.* **33** 140–58

Daly C, Smith J and Halbleib M 2018 1981–2010 High-resolution temperature and precipitation maps for Alaska Oregon State University (available at: https://prism.oregonstate.edu/projects/public/alaska/report/)

Dial R J, Berg E E, Timm K, McMahon A and Geck J 2007 Changes in the alpine forest-tundra ecotone commensurate with recent warming in southcentral Alaska: evidence from orthophotos and field plots *J. Geophys. Res.* **112** 1–15

Dial R J, Smeltz S T, Sullivan P F, Rinas C L, Timm K, Geck J E, Tobin C S, Golden T S and Berg E C 2016 Shrubline but not treeline advance matches climate velocity in montane ecosystems of south-central Alaska *Glob. Change Biol.* **22** 1841–56

Dial R, Chaussé P, Allgeier M, Smeltz T S, Golden T, Day T, Wong R and Andersen H E 2021a Estimating net primary productivity (NPP) and debris-Fall in forests using lidar time series *Remote Sens.* **13** 891

Dial R J, Schulz B, Lewis-Clark E, Martin K and Andersen H E 2021b Using fractal self-similarity to increase precision of shrub biomass estimates *Ecol. Evol.* **11** 4866–73

Douglas T A, Jones M C, Hiemstra C A and Arnold J R 2014 Sources and sinks of carbon in boreal ecosystems of interior Alaska: a review *Elem. Sci. Anth.* **2** 000032

García Criado M, Myers-Smith I H, Bjorkman A D, Lehmann C E R, Stevens N and Josée F M 2020 Woody plant encroachment intensifies under climate change across tundra and savanna biomes *Glob. Ecol. Biogeogr.* **29** 925–43

Hijmans R J 2020 Raster: geographic analysis and modeling with raster data R package version 0.4.1 (available at: https://CRAN.R-project.org/package=raster)

Klein E, Berg E E and Dial R 2005 Wetland drying and succession across the Kenai peninsula lowlands, southcentral Alaska *Can. J. For. Res.* **35** 1931–41

Liaw A and Wiener M 2002 Classification and regression by random Forest *R News* **2** 18–22

Macander M J, Nelson P R, Nawrocki T W, Frost G V, Orndahl K M, Palm E C, Wells A F and Goetz S J 2022 Time-series maps reveal widespread change in plant functional type cover across arctic and boreal Alaska and Yukon *Environ. Res. Lett.* **17** 54042

Mekonnen Z A *et al* 2021 Arctic tundra shrubification: a review of mechanisms and impacts on ecosystem carbon balance *Environ. Res. Lett.* **16** 53001

Myers-Smith I H *et al* 2011 Shrub expansion in tundra ecosystems: dynamics, impacts and research priorities *Environ. Res. Lett.* **6** 045509

Nawrocki T W, Carlson M L, Osnas J L, Trammell E J and Witmer F D 2020 Regional mapping of species-level continuous foliar cover: beyond categorical vegetation mapping *Ecol. Appl.* **30** 1–22

Orndahl K M, Macander M J, Berner L T and Goetz S J 2022 Plant functional type aboveground biomass change within Alaska and northwest Canada mapped using a 35-year satellite time series from 1985 to 2020 *Environ. Res. Lett.* **17** 115010

Overland J, Dunlea E, Box J E, Corell R, Forsius M, Kattsov V, Olsen M S, Pawlak J, Reiersen L O and Wang M 2019 The urgency of Arctic change *Polar Sci.* **21** 6–13

Pacific Northwest Research Station, Forest Inventory and Analysis (PNW-FIA), Forest Inventory and Analysis Resource Monitoring and Assessment Program 2019 *Field instructions for the annual inventory of Alaska* (U.S. Department of Agriculture, Forest Service (USDA FS)) (available at: https://research.fs.usda.gov/pnw/understory/2019-pnw-fia-alaska-field-manual)

Pan Y *et al* 2011 A large and persistent carbon sink in the world's forests *Science* **333** 988–93

R Core Team 2024 *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing)

Rinas C L, Dial R J, Sullivan P F, Smeltz T S, Tobin S C, Loso M and Geck J E 2017 Thermal segregation drives patterns of alder and willow expansion in a montane ecosystem subject to climate warming *J. Ecol.* **105** 935946

Roland C A, Stehn S E and Schmidt J H 2017 Species richness of multiple functional groups peaks in alpine tundra in subarctic Alaska *Ecosphere* **8** e01848

Seider J H, Lantz T C and Bone C 2022 Tundra shrub expansion in a warming climate and the influence of data type on models of habitat suitability *Arct. Antarct. Alp. Res.* **54** 488–506

Shoot C, Andersen H E, Moskal L M, Babcock C, Cook B D and Morton D C 2021 Classifying forest type in the national forest inventory context with airborne hyperspectral and lidar data *Remote Sens.* **13** 1863

Sturm M, Racine C and Tape K 2001 Increasing shrub abundance in the Arctic *Nature* **411** 546–7

Tape K D, Gustine D D, Ruess R W, Adams L G and Clark J A 2016 Range expansion of moose in arctic Alaska linked to warming and increased shrub habitat *PLoS One* **11** e0152636

Travers-Smith H, Coops N C, Mulverhill C, Wulder M A, Ignace D and Lantz T C 2024 Mapping vegetation height and identifying the northern forest limit across Canada using ICESat-2, Landsat time series and topographic data *Remote Sens. Environ.* **305** 114097

Wood S N 2017 *Generalized Additive Models: An Introduction with R* 2nd edn (CRC Press/Taylor & Francis Group)

Wulder M A, White J C, Nelson R F, Næsset E, Ørka H O, Coops N C, Hilker T, Bater C W and Gobakken T 2012 Lidar sampling for large-area forest characterization: a review *Remote Sens. Environ.* **121** 196–209