Unravelling Forest Complexity: Resource Use Efficiency, Disturbance, and the Structure-Function Relationship

Bailey A. Murphy[1], Jacob A. May[2], Brian J. Butterworth[3], Christian G. Andresen[4], and Ankur Rashmikant Desai[5]

[1]Department of Atmospheric and Oceanic Sciences
[2]Department of Geography
[3]3Cooperative Institute for Research in Environmental Sciences at the University of Colorado Boulder
[4]University of Wisconsin Madison
[5]University of Wisconsin-Madison

November 22, 2022

Abstract

Structurally complex forests optimize light and water resources to assimilate carbon more effectively, leading to higher productivity. Information obtained from Light Detection and Ranging (LiDAR)-derived structural complexity (SC) metrics across spatial scales serves as a powerful indicator of ecosystem-scale functions such as gross primary productivity (GPP). However, our understanding of mechanistic links between forest structure and function, and the impact of disturbance on the relationship, is limited. Here, we paired eddy covariance measurements of carbon and water fluxes in temperate forests collected in the CHEESEHEAD19 field campaign with drone LiDAR measurements of SC to establish which SC metrics were strong drivers of GPP, and tested potential mediators of the relationship. Mechanistic relationships were inspected at four metric calculation resolutions to determine whether relationships persisted with scale. Vertical heterogeneity metrics were the most influential in predicting productivity for forests with a significant degree of heterogeneity in management, forest type, and species composition. SC metrics included in the structure-function relationship as well as the strength of drivers was dependent on metric calculation resolution. The relationship was mediated by light use efficiency (LUE) and water use efficiency (WUE), with WUE being a stronger mediator and driver of GPP. These findings allow us to improve representation in ecosystem models of how SC impacts light and water-sensitive processes, and ultimately GPP. Improved models enhance our ability to simulate true ecosystem responses to management, resulting in a more accurate assessment of forest responses to management regimes and furthering our ability to assess climate mitigation and strategies.

Unravelling Forest Complexity: Resource Use Efficiency, Disturbance, and the Structure-Function Relationship

Bailey A. Murphy[1], Jacob A. May[2], Brian J. Butterworth[3], Christian G. Andresen[2], Ankur R. Desai[1]

[1]Department of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, Madison, Wisconsin, USA

[2]Department of Geography, University of Wisconsin-Madison, Madison, Wisconsin, USA

[3]Cooperative Institute for Research in Environmental Sciences at the University of Colorado Boulder, Boulder, Colorado, USA

Corresponding author: Bailey A. Murphy (bamurphy5@wisc.edu)

Key points:

- Vertical heterogeneity metrics are the most influential productivity drivers for heterogenous temperate forests
- The structure-function relationship is mediated by resource use efficiency, and water use efficiency is a strong driver of productivity
- The mechanistic relationship between forest structure and function is dependent upon structural metric calculation resolution

Abstract

Structurally complex forests optimize light and water resources to assimilate carbon more effectively, leading to higher productivity. Information obtained from Light Detection and Ranging (LiDAR)-derived structural complexity (SC) metrics across spatial scales serves as a powerful indicator of ecosystem-scale functions such as gross primary productivity (GPP). However, our understanding of mechanistic links between forest structure and function, and the impact of disturbance on the relationship, is limited. Here, we paired eddy covariance measurements of carbon and water fluxes in temperate forests collected in the CHEESEHEAD19 field campaign with drone LiDAR measurements of SC to establish which SC metrics were strong drivers of GPP, and tested potential mediators of the relationship. Mechanistic relationships were inspected at four metric calculation resolutions to determine whether relationships persisted with scale. Vertical heterogeneity metrics were the most influential in predicting productivity for forests with a significant degree of heterogeneity in management, forest type, and species composition. SC metrics included in the structure-function relationship as well as the strength of drivers was dependent on metric calculation resolution. The relationship was mediated by light use efficiency (LUE) and water use efficiency (WUE), with WUE being a stronger mediator and driver of GPP. These findings allow us to improve representation in ecosystem models of how SC impacts light and water-sensitive processes, and ultimately GPP. Improved models enhance our ability to simulate true ecosystem responses to management, resulting in a more accurate assessment of forest responses to management regimes and furthering our ability to assess climate mitigation and strategies.

Plain Language Summary

The way that trees are arranged within a forest impacts the forest's ability to use light and water resources for photosynthesis. Forests that are arranged in more complex ways do a better job of using available resources, and have higher rates of photosynthesis, or productivity. By combining data that describes the complexity of the forest with data that describes how much photosynthesis is taking place, we can better understand which factors impact that relationship, and which types of forest complexity are the most important. We used data from a temperate forest with a long history of management and found that vertical complexity was the most influential, and that the intensity of management had a large impact on the relationship between complexity and productivity. We also found that the relationship was controlled by how efficiently the forest used the available resources, and that the spatial resolution at which the data were examined changed the relationship. These findings will allow us to improve the mathematical models we use to test the impacts of forest management on forest productivity, which will enhance our ability to manage our resources in the face of climate change.

1 Introduction

Advancing our understanding of the relationship between forest structural complexity (SC) and key ecosystem functions such as carbon and water cycling requires quantification of the mechanistic links between structure and function. Mapping these links is a fundamental aspect of improving our ability to scale measurements from the leaf to stand to landscape level, a preeminent challenge in the field of ecosystem ecology (Bonan 2008, Fahey et al., 2019). Forest SC characterizes the three-dimensional arrangement of vegetation in a forest and includes variables such as rugosity, vertical complexity index, and mean canopy height (McElhinny et al., 2005, Atkins et al., 2018). Taken together, these variables constrain the ability of the forest to assimilate available resources, and thus the capacity for photosynthesis (Ehbrecht et al., 2021). The prevailing theory is that structurally complex forests are better able to optimize incoming light and water resources to assimilate carbon more effectively (Anten, 2016, Hardiman et al., 2011, Atkins et al., 2018, Gough et al., 2016). It has been suggested that heterogeneous mixed forests with higher levels of SC are tied to a heightened ability to capitalize on available resources, in part due to functional trait variability and niche differentiation (Zhang et al., 2012, Williams et al., 2017, Hillebrand et al., 2008, Danescu et al., 2016).

Studies have shown that integrating information obtained from SC metrics across spatial scales to describe overall SC can serve as a powerful indicator of ecosystem-scale functions such as gross primary productivity (GPP), augmenting other commonly measured characteristics including species composition and diversity (Atkins et al., 2018, Gough et al., 2019, Hardiman et al., 2011, Silva Pedro et al., 2017, Eitel et al., 2016, Fahey et al., 2019). Identifying not only which SC variables have the greatest potential to predict GPP, but what potential controls or influential factors of the structure-function relationship might exist is a vital aspect of this effort. Additionally, relationships between productivity and SC could provide mechanistic evidence for using these SC metrics as predictors of forest carbon storage capacity and functionality.

The application of Unoccupied Aerial System (UAS) Light Detection and Ranging (LiDAR) to derive physically based parameters such as SC variables helps to address critical knowledge gaps regarding our mechanistic understanding of how structure determines function (Atkins et al., 2018, Camarretta et al., 2020). As opposed to passive optical remote sensing approaches, active remote sensing tools such as LiDAR have demonstrated superior performance in capturing three-dimensional vertical profiles of stand structure (Lefsky et al., 1999, Eitel et al., 2016). In addition to allowing for the quantification of complexity, LiDAR-derived SC metrics can facilitate a deeper understanding of the relationship between complexity and successional processes in heterogeneous mixed temperate forests (van Ewijk et al., 2011). This is an important endeavor as much of the work done to develop related ecological theory has been conducted in more mature coniferous forests with comparatively less species diversity (Pregitzer and Euskirchen, 2004, Ryan et al., 1997, Kane et al., 2010), which likely experience different control mechanisms on productivity than temperate deciduous or mixed forests (Gough et al., 2016).

This approach is particularly useful in the temperate forests of the upper Midwest USA, where once even-aged forests are undergoing a transition to more complex systems as they approach advanced stages of successional development following a long history of intensive disturbance (Hardiman et al. 2011, Frelich, 1995, Bogdanovich et al., 2021). Variability in disturbance legacies combined with a primarily mixed deciduous-conifer forest composition and general landscape heterogeneity result in large variations in stand complexity at the ecosystem scale. As SC has been shown to be positively correlated with stand production, characterizing the mechanistic relationship between complexity and productivity will enable better representation of the potential impacts of these transitions on carbon sequestration in Midwestern forests (Forrester et al., 2013). The study design of the 2019 Chequamegon Heterogenous Ecosystem Energy-balance Study Enabled by a High-density Extensive Array of Detectors (CHEESEHEAD19) field experiment provided a unique opportunity to partially control for the influence of variability in climate, edaphic factors, and forest functional types on productivity, allowing for a more representative physiological understanding of the structure-function relationship than has been previously demonstrated.

The objective of this study was to identify mechanistic relationships between forest structure and function, explore potential controls or mediating factors on that relationship, and determine whether or not the structure-function relationship persisted when structural metrics were calculated at a variety of resolutions.

In pursuit of this objective, this project addressed four primary research questions:

1. Which SC metrics are most influential for the prediction of stand primary productivity in mixed temperate forests with a high degree of heterogeneity and a long history of management?

2. How do management legacies impact these influential SC metrics, and ultimately stand productivity?

3. Is the mechanistic relationship between forest structure and function direct, or is it mediated by other factors such as RUE?

4. Is the mechanistic relationship between forest structure and function dependent upon the scale of structural metric calculation?

2 Methods

2.1 Experimental design

During the CHEESEHEAD19 intensive field campaign spanning from June to October of 2019, 17 eddy covariance (EC) flux towers from the NSF Lower Atmosphere Observing Facility (LAOF) were deployed across the 10 x 10 km study domain. These 17 towers were in addition to the preexisting AmeriFlux tall tower US-PFa, and two towers supported by Dr. Paul Stoy, bringing the total number of EC towers to 20. The primary research interests of CHEESEHEAD19 were to explore potential drivers behind the enduring lack of energy balance closure frequently observed over heterogeneous landscapes, and to address persistent challenges associated with upscaling surface energy fluxes (Butterworth et al., 2021). The study period reflects both the summer season land-atmosphere exchange as well as exchanges during the transition of vegetation into senescence. This observational period was chosen to support the energy balance related research interests of CHEESEHEAD19, as it captures the shift in energy balance from a latent heat flux dominant landscape to a sensible heat flux dominant landscape (Butterworth et al., 2021). Of the 20 total towers, nine towers located in forested sites were selected to measure forest composition using UAS mounted LiDAR (Figure 1). These nine sites were selected given their 1) forested composition (several of the original 20 sites were located in wetland areas) and representative forest type, 2) overlap with flux tower footprints, and 3) ease of access for UAS flying. While climatic conditions and topography are shared across the nine sites, the selected sites span a range of successional stages, dominant vegetation types, management histories, and degrees of heterogeneity. Pairing EC surface-atmosphere carbon and water fluxes with LiDAR-derived forest SC metrics, mechanistic relationships between forest structure and function were established.

Mechanistic relationships were explored using best-subsets selection and structural equation modeling (SEM), specifically path analysis. The application of SEM allows for the establishment not only of which SC metrics are influential in predicting GPP, but the specific strengths, significance, and variability of their predictive power. Additionally, SEM allows for the testing of variables that potentially serve as mediators of the relationship between SC and GPP, through the comparison of reduced and saturated model designs (Fan et al., 2016). This study explored the viability of resource use efficiency (RUE) as a mediator of the structure-function relationship, as previous studies have demonstrated it to be a strong predictor of forest productivity (Atkins et al., 2018, Gough et al., 2019). Both water use efficiency (WUE) and light use efficiency (LUE) were used to represent overall stand RUE.

RUE describes how well a forest stand captures and utilizes its available resources to fix carbon dioxide, with greater efficiency typically resulting in higher levels of biomass production (Binkley et al., 2004, Anderson-Teixeira et al., 2021). This paper focuses specifically on light and water as the primary limiting resources controlling the rate of photosynthesis, although other factors including the supply of CO_2, concentration of photosynthetic enzymes such as Rubisco, and availability of catalysts including nitrogen and phosphorous have been explored at length in other studies (Tang et al., 2018, Hardiman et al., 2013, Ainsworth and Long 2004). Additionally, these mechanistic relationships were inspected at four structural metric calculation resolutions to determine whether the relationships persisted with scaling, or if they were simply artifacts of the resolution at which they were calculated. Structural metrics were calculated from LiDAR returns collected at spatial resolutions of 0.25 m, 2 m, 10 m, and 25 m.

2.2 Site description

The study area is a 10 x 10 km domain located in the Chequamegon-Nicolet National Forest in Northern Wisconsin. Most of the region is heavily forested and trees are primarily deciduous but a significant conifer presence exists as well. There is a high degree of heterogeneity representative of a typical mid-latitude forest, displaying a diverse array of wetlands, meadows, streams, and lakes in addition to forest cover (USDA Forest Service, 2011). Typical homogenous patches of land cover are generally around 20 hectares or less (Desai et al., 2015). Heterogeneity is further accentuated by a long history of non-uniform forest management practices including thinning and clear-cuts, resulting in increased variability in stand age and structure. Forests in Northern Wisconsin typically have an age distribution centered around 'middle age', or 40 – 90 years (Birdsey et al., 2014, Wisconsin Department of Natural Resources, 2019). This age pattern is reflective of the fact that the majority of the forested land was logged in the mid-19th to early 20th century to clear land for agricultural purposes (Desai et al., 2007, Gough et al., 2007, Rhemtulla et al., 2009), which was followed by subsequent periods of agricultural land abandonment, reforestation, fire suppression, and intensive timber harvest (Birdsey et al., 2006). In addition to human management, the region experiences natural disturbance due to windstorms, insect invasion, and occasionally fire (Gough et al., 2007). Fires were historically influential during times of land clearing and Euro-American settlement (Rhemtulla et al., 2009), but wind damage has had more consistent impacts on stand structure and carbon storage over time (Mladenoff et al., 2008).

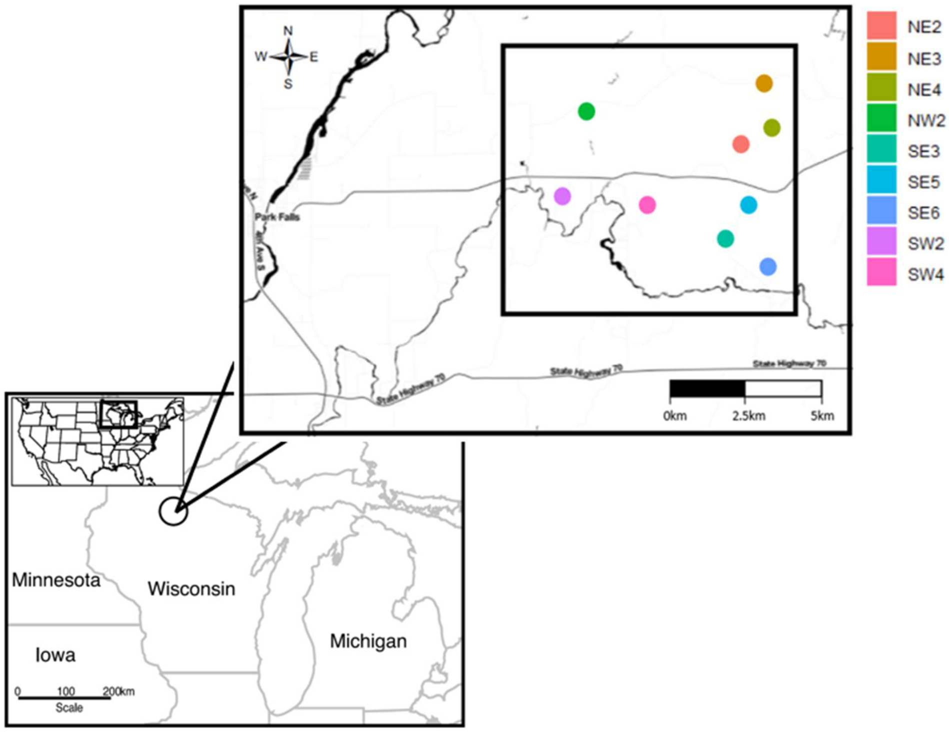


Figure 1. Map depicting the location of the study site within a regional and state context. The black circle on the state map depicts a 60km radius around the location of the Park Falls, Wisconsin WLEF tall tower. Colored dots represent the nine sites within the 10 x 10 km CHEESEHEAD19 study domain (represented by the black square) selected for measurement of forest composition.

The study domain is of relatively consistent low-grade elevation and human population is minimal. Slight variations in terrain elevation in combination with significant precipitation in all seasons results in a mix of saturated (wetland) and unsaturated (upland) sandy loam soils (Davis et al., 2003). Upland forests comprise roughly 65% of the landscape (Wisconsin Department of Natural Resources, 2019) and deciduous tree types include quaking aspen (*Populus tremuloides* Michx.), sugar maple (*Acer saccharum* Marsh), red maple (*Acer rubrum* L.), basswood (*Tilia americana*), beech (*Fagus grandifolia*), and several varieties of oak and birch. Coniferous tree varieties include balsam fir (*Abies balsamea*), red, white, and jack pine (*Pinus resinosa, Pinus strobus, Pinus banksiana*), and white spruce (*Picea glauca*). Wetlands are both forested and unforested and account for approximately 35% of the land cover (Wisconsin Department of

Natural Resources, 2019). Wetland tree species include alder (*Alnus incana*), cedar (*Thuja occidentalis*), tamarack (*Larix larcina*), and black spruce (*Picea mariana*) (Davis et al., 2003). The area has a Köppen climate classification of Dfb, and experiences a humid continental climate characterized by warm humid summers and cold snowy winters, with no significant difference in precipitation amount between seasons (Arnfield, 2020).

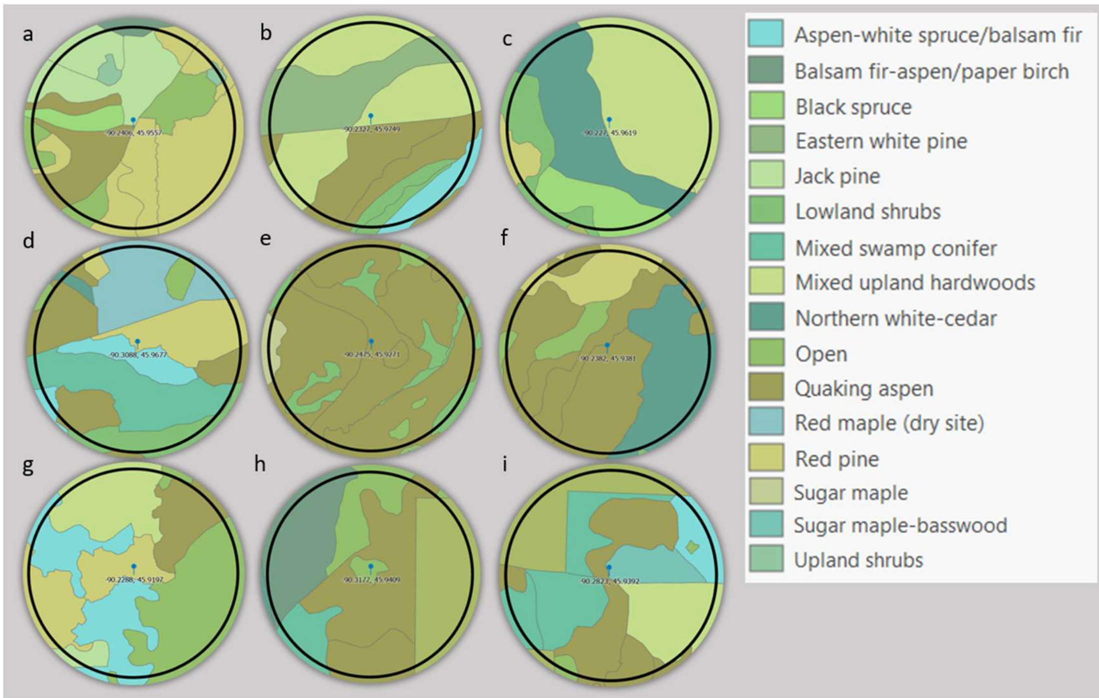


Figure 2. Vegetation coverage at each of the nine forested sites: **a)** NE2 **b)** NE3 **c)** NE4 **d)** NW2 **e)** SE3 **f)** SE5 **g)** SE6 **h)** SW2 and **i)** SW4. Coverage is segmented by both vegetation type and stand age.

2.3 Measurements

2.3.1 Flux Towers

Exchanges of carbon, water, and energy between the atmosphere and the land surface were collected at a frequency of 20 Hz using an open-path infrared H_2O and CO_2 gas analyzer (Campbell Scientific EC150) and sonic anemometer to measure three-dimensional wind speed (Campbell

Scientific CSAT3AW). In addition to flux-specific instrumentation, the nine selected sites were similarly outfitted with meteorological instruments including slow-response air temperature and humidity sensors (NCAR SHT), barometers (Vaisala PTB210), and 4-component radiometers (Hukseflux NR01). Gas analyzers, sonic anemometers, barometers, and radiometers were all mounted at the top of the EC towers above the local forest canopy, mounting heights are presented in Table 1. Additional instrumentation included tower-mounted air temperature sensors at two levels within the canopy (2 m and mid-canopy, which varied by site), and soil sensors (NCAR 4-level Tsoil, Meter EC-5 Qsoil, REBS HFT Gsoil, and Hukseflux TP01 Csoil) buried near the base of each tower in the upper soil profile (0 – 5 cm). Instrument power was supplied via exchangeable batteries, which occasionally resulted in minimal data loss due to limited recharging capacity at the field operations base. NR01 radiometer deployment was delayed for sites NW2, NE3, SW2, and SE5, therefore no data exists for approximately the first 25 days of the study period. Radiometer data was filtered for sensor wetness and cleaning periods. Gas analyzers were cleaned 2 – 3 times during the study, and data was filtered out for periods of significant nighttime dew formation, which resulted in sensor biases.

Table 1. LiDAR footprint size, instrument installation heights, and age and tree height metrics for each of the nine selected forest plots

Site	LiDAR Footprint (km^2)	Instrument Height (m)	Avg. Tree Height (m)	Avg. Age (years)
NE2	0.48	32	14.20	56.77
NE3	0.24	32	18.10	71.29
NE4	0.18	32	18.70	108.5
NW2	0.23	12	8.80	44.08
SE3	0.82	32	8.10	42.00
SE5	0.22	13	12.40	55.67
SE6	0.23	32	10.30	49.50
SW2	0.22	30	10.90	63.50
SW4	0.82	32	13.50	76.27

Turbulent fluxes of carbon, water, and energy were calculated every thirty minutes from high frequency (20 Hz) eddy covariance measurements. Prior to gap filling, a friction velocity (u*) threshold calculation was performed using the approach outlined in Wutzler et al. (2018), where the u* threshold is estimated with the moving point test. u* is a reference wind velocity that represents the shear stress arising through movement across the land surface. Below the u* threshold, turbulent mixing is weak enough that flux measurements are considered non-representative of the actual flux state, and thus net ecosystem exchange (NEE) flux data is filtered out during those time periods. Gap filling and filtering of flux data was performed using the software REddyProc (Wutzler et al., 2018). Prior to gap filling, an average of 37% of NEE values were missing across all nine sites, with individual site missing values ranging from 26% (SW2) to 61% (SE5). Missing data occurred to some degree at every site, although the reasons for missing data (equipment malfunction or cleaning, temporary power loss, moisture interference, etc.) varied. GPP was approximated from NEE using the flux partitioning method described in Reichstein et al. 2005 and was calculated using both the nighttime and the light response curve methods for respiration (Reichstein et al., 2012).

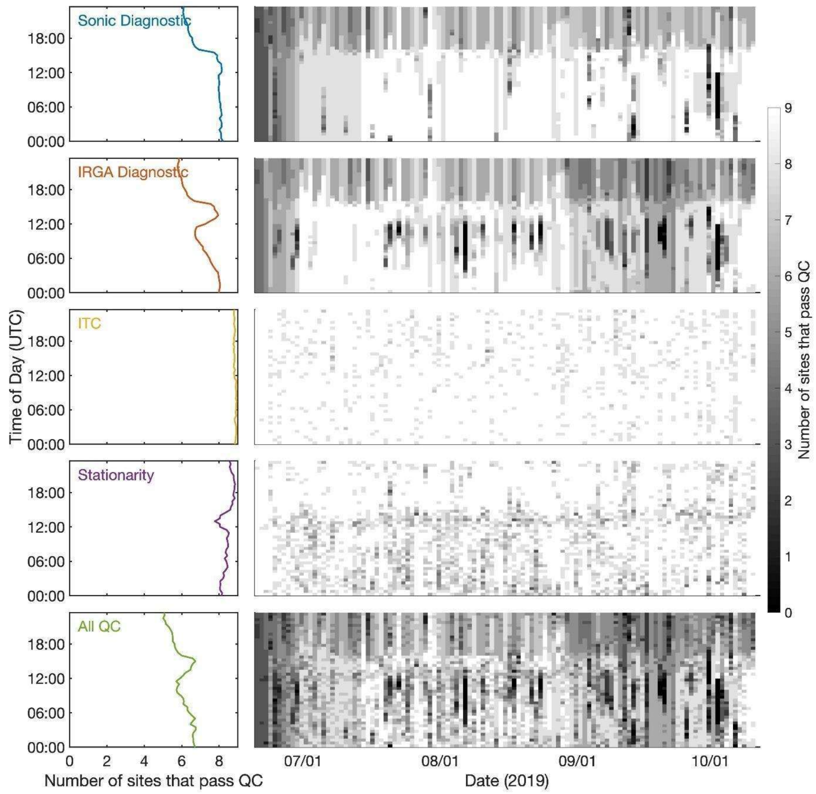


Figure 3. Results of the four quality control (QC) checks that were assessed for EC data, as well as the combined QC assessment for each site. The gray scale represents the number of sites that passed or failed the QC assessment for each date and time during the measurement period.

2.3.2 Drone-based LiDAR

To characterize three-dimensional forest structure, we employed a Routescene © LiDAR onboard a UAS hexacopter DJI M600 Pro to collect high-density 3D scans (~600 points m^{-2}). Over the span of June 25 – 29, 2019, we surveyed the footprints of the nine selected flux tower sites and areas ranging between 0.25 to 1 km^2 per site (Table 1) with a flight footprint of approximately 500 m x 500 m. Autonomous flights (with a duration of ~20 minutes each) were programmed using Universal ground Control Software (UgCS) v3.2.113. Flights were performed at a speed of 6 m s^{-1}, 60 m above ground level and 60 m side distance between parallel flight lines. Raw data was boresight calibrated, filtered and *.laz exported using Routescene proprietary software LidarViewer ©. Points within 1 mm radius were filtered and a box range filter centered on the sensor for each scan (scan rate 10 Hz) of 120 m width, 180 m height and

120 m length was applied, ensuring each flight line would have complete overlap with other flight lines. Random noise was addressed using a statistical outlier removal filter and combined (only for multiple flights per site) in CloudCompare v2.10.

2.3.3 Stand age and disturbance

Stand age and disturbance history data was obtained from the publicly available United States Department of Agriculture Forest Service Geodata Clearinghouse. All sites had multiple distinct age classes present, representing a range of successional statuses (Figure 4). The majority of the sites were dominated by stands in the young to middle age classes, although regeneration saplings younger than five years were not specifically accounted for. The young age class corresponds to the stand initiation and stem exclusion successional stages (Odum, 1969), and the middle age class, defined by Pan et al. (2011) as roughly 40 – 100 years, corresponds to the understory reinitiation stage. Two sites (NE4 and SW2) contain stands that fall within the old growth successional stage, characterized in the temperate Lake States (Minnesota, Wisconsin, and Michigan) by the presence of long-lived tree species that are at or greater than 120 years of age and exist in an advanced stage of structural development (Frelich, 1995). Forest Inventory Analysis data shows that the oldest forests sampled in the temperate Lake States region are between 200 – 210 years old (Birdsey et al., 2014).

Several sites have experienced significant disturbance in the form of clearcutting and harvest (Figure 5), with the most recent harvest taking place in 2016 (SE6), and the most recent clear cut occurring in 2013 at stands in sites SE5 and NW2. Harvest is broadly defined here to include selective and shelterwood cuts as well as any harvest that is not stand replacing, whereas a clear cut specifies a stand replacing harvest occurring within the last fifty years. In addition to anthropogenic disturbance, sites SE6 and SE3 experienced substantial hail damage in the year 2000, and large-scale defoliation resulting from Forest Tent Caterpillar infestation occurred across the domain in 2001 (Wisconsin Department of Natural Resources). Blowdown due to wind stress has also been noted at sites SW4, SE6, NW2, and NE2, with the damage being most substantial at site SE6. Neither wildfire nor prescribed burning management activities have been a significant disturbance factor at any of the study plots. Species specific planting has occurred at sites SW2, SE5, NW2, and NE2.

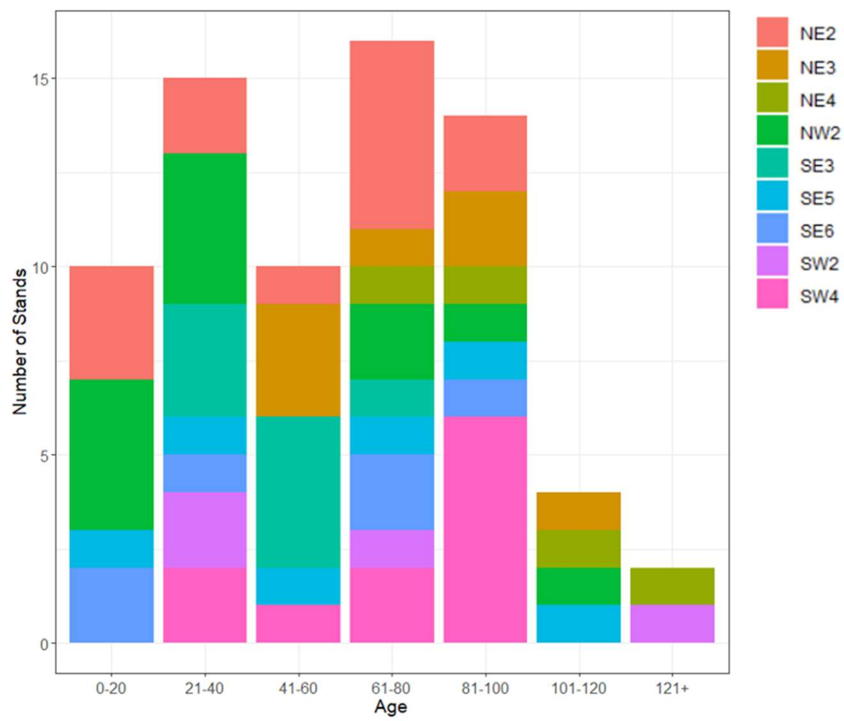


Figure 4. Age distribution at the nine selected forest sites, where colors represent different sites to highlight the presence of multiple forest age classes within a single site.

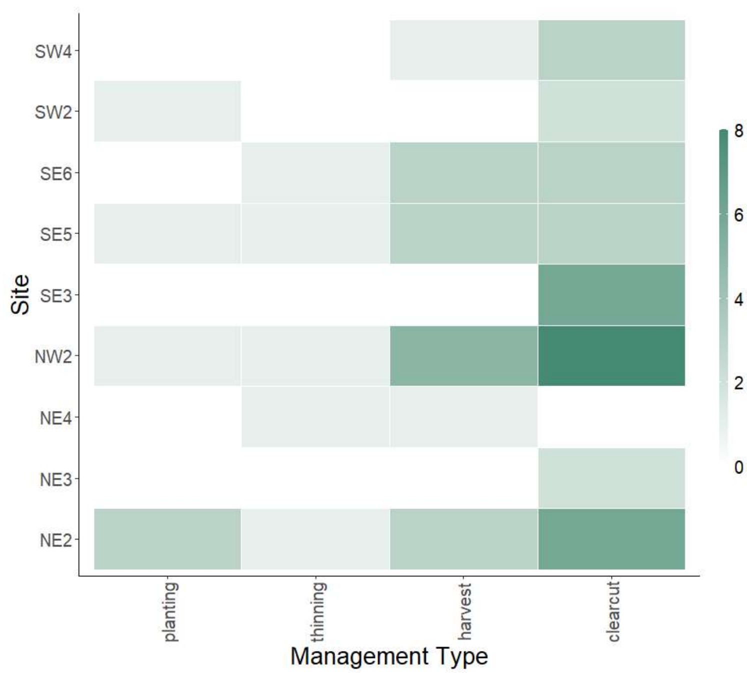


Figure 5. Management practices and frequency of occurrence recorded at each site.

2.4 Statistical Analysis

2.4.1 Metric extraction

LiDAR generated datasets were analyzed using the R programming language (R Core Team 2021, Version 4.0.4) package *lidR* (Roussel et al., 2020). The cloth simulation filter was used to identify ground points (Zhang et al., 2016) and triangulation was used to construct a digital terrain model from the ground points, which was then height-normalized. For each plot, 20 LiDAR metrics were calculated to describe tree height, arrangement, and stand complexity using the R programming language package *forestr* (Atkins et al., 2018). *forestr* gives a comprehensive formulation of metrics for characterizing forest canopy SC and arrangement using either portable canopy LiDAR or terrestrial laser scanning ground-based LiDAR platforms. Several metrics described in the *lidR* R library were adapted for an area-based approach with a UAS platform (Table 2). With the exception of “Rumple” and “VerticalDistMax”, each of the metrics were calculated by creating a raster of the site with a value for each pixel, then finding the average or standard deviation for all pixels within the site. For example, to find the average tree height, a raster of each site was first created where each pixel in the raster was assigned the average height of all the LiDAR returns within the pixel. For this metric LiDAR returns under 0.5 m were removed to exclude most ground points from the calculation. To summarize the data as a single number, the mean of all the pixels in the raster was used. Each raster-based metric was calculated at a resolution of 0.25 m, 2 m, 10 m, and 25 m per pixel to check for resolution dependencies.

Some metrics require additional explanation. Rumple was computed by creating a canopy height model for each site and dividing its area by the projected ground area. VerticalDistMax was computed by finding the vertical distribution of all the points in a site and determining which height bin contained the most points. Vertical bins of 0.5 m and a lower cutoff of 5 m were used to prevent the ground cover and understory from influencing the result. Both of these metrics were calculated on a per site basis instead of a per pixel basis. Leaf area index (LAI) was also calculated using the formulation provided in the *forestr* library (Atkins et al., 2018) and compared to LAI field measurements for verification, which showed a high correlation of $R = 0.78$ ($p < 0.05$).

339 **Table 2.** Description of LiDAR-derived forest complexity metrics

Metric	Description	Source
Rumple	Ratio of the top surface area of the canopy to the projected ground area	*Kane et al., 2010*
VerticalDistMax	Height with the most points, using 0.5m bins above a cutoff height of 5m	*Atkins et al., 2018*
maxZ_mean	Mean of max height of points in each pixel	*Atkins et al., 2018*
maxZ_sd	Standard deviation of max height of points in each pixel	*Atkins et al., 2018*
sdZ_mean	Mean of standard deviation of height of points in each pixel	*Atkins et al., 2018*
sdZ_sd	Standard deviation of standard deviation of height of points in each pixel	*Atkins et al., 2018*
meanZ_mean	Mean of mean of height of points in each pixel	*Atkins et al., 2018*
meanZ_sd	Standard deviation of mean of height of points in each pixel	*Atkins et al., 2018*
density_mean	Mean of density of points in each pixel	*Roussel et al., 2020*
density_sd	Standard deviation of density of points in each pixel	*Roussel et al., 2020*
gap_fraction	Fraction of pixels with returns below a cutoff height	*Atkins et al., 2018*
VCI_mean	Mean of vertical complexity index	*van Ewijk et al., 2011*
VCI_sd	Standard deviation of vertical complexity index	*van Ewijk et al., 2011*
LAI_mean	Mean of leaf area index	*Atkins et al., 2018*

LAI_sd	Standard deviation of leaf area index	*Atkins et al., 2018*
RH25	Mean of 25th quantile of point heights	*Schneider et al., 2017*
RH50	Mean of 50th quantile of point heights	*Schneider et al., 2017*
RH75	Mean of 75th quantile of point heights	*Schneider et al., 2017*
RH95	Mean of 95th quantile of point heights	*Schneider et al., 2017*
canopy_ratio	Mean of 95th quantile of heights minus the 25th quantile of heights divided by the 95th quantile of heights	*Schneider et al., 2017*

LUE was calculated as the ratio of total daily GPP to total daily incoming photosynthetic photon flux density (PPFD), where PPFD is the incident flux density of photosynthetically active radiation (PAR), or the number of photons incident per unit time on a unit surface (Olson et al., 2004). PPFD is considered a synonym for incident PAR (IPAR) (Olson et al., 2004). The exchange of carbon between the forest plots and the atmosphere was measured by the EC towers directly and partitioned into GPP and ecosystem respiration, R_{eco} (Reichstein et al., 2012). The site EC towers were only equipped to measure incoming and outgoing shortwave and longwave radiation as well as net radiation, as opposed to direct measurement of PPFD. Incoming shortwave radiation was converted to PPFD using a fraction of incoming solar irradiance in the photosynthetically active region of 0.50 (Knauer et al., 2018).

WUE describes the amount of carbon fixed per unit of water transpired (De Kauwe et al., 2013), and was calculated here as grams of carbon produced as biomass for every kilogram of water released through evapotranspiration (ET). ET is the sum of evaporation from the land surface and transpiration from vegetation, and is both the key process determining water use in forests (Fisher et al., 2017, Mathias and Thomas, 2021), and the primary process through which the carbon cycle is connected to and maintains the water cycle (Raupach et al., 2005). Since ET was not directly measured by this eddy covariance system, latent heat flux was used as its equivalent.

2.4.2 Model determination

A suite of linear regression models were tested to evaluate the relationships between SC metrics, RUE, and stand productivity. Non-linear models were not tested, as previous studies exploring multiple non-linear model representations have shown that although the relationships may in reality be non-linear, non-linear representations repeatedly failed to achieve statistical significance (Gough et al., 2019). The combination of SC and RUE metrics that best predicted stand GPP was assessed using best-subsets model selection. Model fit was evaluated using the Schwarz Bayesian Criterion (SBC), mean square error prediction (MSEP), and adjusted R^2 ($R^2{}_{adj}$), where the model with the lowest significant SBC ($p < 0.05$), lowest MSEP, and highest $R^2{}_{adj}$ was selected as optimal. SBC was used as opposed to Akaike information criterion to account for the presence of multiple predictive variables and a relatively small sample size.

High multicollinearity was a significant problem in determining which SC metrics were the most robust drivers of GPP. Several SC metrics had intercorrelation values that exceeded 0.98 and thus were not included in SEM. This included metrics related to the height at which a given quantile of returned energy was reached relative to the ground, metrics RH25, RH50, RH75, and RH95. Variance inflation factors (VIF) were calculated for the best-fit models, and models were classified as having severe multicollinearity if the average VIF was greater than 10. Pearson's correlation coefficients were used to determine the strength of pairwise interactions between variables for models where severe multicollinearity was a concern to determine which SC metric was likely driving the observed multicollinearity, and that variable was subsequently removed and the resulting model was reevaluated.

SEM was used to ascertain the mechanistic relationship between stand productivity and the influential SC metrics determined through best-subsets selection, as well as whether or not the relationship was direct or was mediated by RUE. Path analysis, a subset of SEM where models are created as a series of regressions to specify causal relationships between variables (Fan et al., 2016), was used to determine possible mediation effects of RUE through the comparison of reduced and saturated models. The reduced model allowed SC metrics to predict WUE and LUE, and WUE and LUE to then predict GPP. The saturated model allowed for the same prediction pipeline, but SC metrics could also bypass RUE and directly impact GPP (Figure 6).

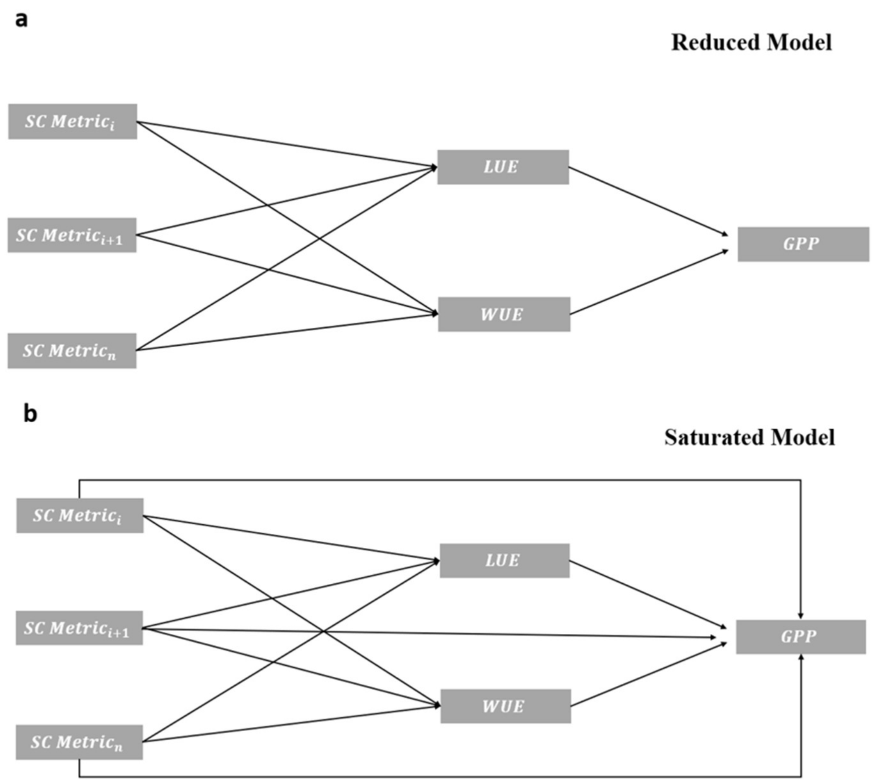


Figure 6. Conceptual figure outlining the **(a)** reduced and **(b)** saturated SEM model designs. The reduced model **(a)** restricts SC metrics to influencing the dependent variable, GPP, indirectly through their effect on LUE and WUE, whereas the saturated model **(b)** allows SC metrics to affect GPP both directly and indirectly through LUE and WUE. Arrows indicate the direction of influence from one variable to the next.

SEM was performed at each of the four LiDAR metric calculation resolutions to assess whether or not the mediation effect persisted with resolution changes. Reduced and saturated model fit was assessed using comparative fit index (CFI), standardized root mean square residual (SRMR), and SBC. CFI values closer to one indicate better model fit, so a threshold value of $\geq$ 0.80 was applied (Hu and Bentler, 1991). SRMR represents the difference between observed and expected variable correlations, and a threshold value of $\leq$ 0.90 was applied, with a lower value indicating a better model fit. Maximum likelihood estimation was used to determine model fit, and parameter estimates were standardized across all observed variables. Bootstrapping was used to test the significance of indirect effects between SC variables and productivity through LUE and WUE as well as for estimation of standard errors and bootstrap-based confidence intervals. 1,000 draws were performed for each indirect effect evaluated. Significance testing of mediation was performed using the R programming language (R Core Team 2021, Version 4.0.4) package *lavaan* (Rosseel, 2012).

3 Results

3.1 Stand productivity and resource use efficiency

Of the nine CHEESEHEAD19 sites examined here, eight were classified as net carbon sinks, where a negative flux value indicates a net flux of carbon into the ecosystem from the atmosphere. A single site (NE2) was classified as a net carbon source, albeit a minor one, with a net flux of 35 gC m^{-2} released to the atmosphere over the entire measurement period. Additionally, at eight out of the nine sites greater variability in daily fluxes was observed for GPP than NEE, with an average variance of 28 gC m^{-2} for GPP compared to 7.8 gC m^{-2} for NEE. Across all sites average daily GPP ranged from 2.6 gC m^{-2} to 14 gC m^{-2}, and average daily fluxes of NEE ranged between -3.5 gC m^{-2} and 0.30 gC m^{-2}. Substantial variability was observed in daily total ecosystem respiration (R_{eco}) as well, defined as the sum of both heterotrophic and autotrophic respiration, with an average variance of 20 gC m^{-2}. The highest productivity (represented as GPP) was observed at sites NE2, SW2, and SW4, with average GPP ranging from 10 – 14 gC m^{-2} day^{-1}. Although NE2 has the highest productivity of the nine sites, it also has the highest average daily R_{eco} (14 gC m^{-2} day^{-1}), resulting in its ultimate classification as a slight net carbon source to the atmosphere, as NEE = R_{eco} – GPP. The three sites with the lowest productivity are NW2, SE5, and NE4. NW2 has a higher number of clear cuts than all other sites, several stands described as wet conifer bogs, and includes stands ranging in age from 7 – 111 years. SE5 includes a mix of aspen, pine, and upland hardwoods ranging in age from 19 – 92 years. NE4 is a considerably older site, with stand age ranging from 76 – 15 years, and consisting of mixed upland hardwoods, pine, and northern white cedar. Over the course of the June-October observational period, productivity peaked in June to mid-July and decreased into fall as leaves began to senesce, with an average change in GPP across all nine sites of 19 gC m^{-2}. Of the sites, NW2 exhibited the least seasonal change in productivity, with a total difference of only 5.9 gC m^{-2} between the start and end of the study period.

Both LUE and WUE varied between sites, with the across-site average LUE equaling 0.70 gC MJ^{-1} and WUE equaling 4.1 gC kg H_2O^{-1}. Average LUE variance was 0.19 gC MJ^{-1} and average WUE variance was 1.4 gC kg H_2O^{-1}. Site NE2 had the highest RUE overall, with a daily LUE of 0.96 gC MJ^{-1} and a WUE of 5.7 gC kg H_2O^{-1}. NE2 also had the highest variability in RUE, although this variability follows a clear pattern indicating the changes in RUE potentially emerge as a response to changes in temperature or other climatic variables. Site NW2 had the lowest overall RUE, with a daily LUE of 0.33 gC MJ^{-1} and a WUE of 2.9 gC kg H_2O^{-1}. Site NW2 had the lowest variability in LUE (0.11 gC MJ^{-1}), but the fourth highest variability in WUE (1.4 gC kg H_2O^{-1}).

3.2 Classification of structural complexity

Of the 20 LiDAR metrics originally calculated, ten unique metrics related to SC were shown through best-subsets selection to be both influential and statistically significant drivers of stand productivity when combined with RUE variables ($p \leq 0.05$), and thus were included in subsequent SEM testing (Table 3). LUE and WUE were present in all of the best-fit models regardless of spatial resolution, but the specific SC metrics included in each of the four best-fit models varied depending upon resolution, although several overarching trends stood out. SC metrics describing vertical heterogeneity were the most prevalent and existed in each of the four final model formulations. VCI_mean was the most frequently observed SC metric, and was included in three of the four models. verticalDistMax and maxZ_sd were the second most prevalent metrics, each showing up in two out of the four models. While both metrics are measures of heterogeneity in SC, verticalDistMax is associated with vertical heterogeneity while maxZ_sd is associated with outer canopy heterogeneity. The remaining seven SC metrics each only appeared in a best fit model model formulation a single time, and included rumple, meanZ_sd, sdZ_sd, LAI_sd, maxZ_mean, sdZ_mean, and LAI_mean. Of these seven SC metrics, three are related to vertical heterogeneity (sdZ_sd, sdZ_mean, and meanZ_sd), one to outer canopy heterogeneity (rumple), one is a measure of tree height (maxZ_mean), and two describe the area and density of vegetation distribution (LAI_sd and LAI_mean). Of these seven SC metrics, three are only present in the 25 m resolution model, indicating that the 25 m model has the greatest departure from the other best fit models. Fit metric ranges for the single best fit model at each resolution displayed no significant differences by resolution. Average $R^2{}_{adj}$ was 0.32 with a range of 0.02, average BIC was 4425 with a range of 20, and average MSE was 16.58 gC m¯2 day¯1 with a range of 0.44 gC m¯2 day¯1. This suggests that SC metric's viability as a driver of GPP isn't restricted to fine or coarse resolutions.

Table 3. Canopy structural complexity metrics included in SEM, isolated as highly influential through best-subsets selection for their strength as drivers of GPP.

Resolution (m)	Metric	Symbol	Units	Complexity Category
0.25				
	rumple	rumple	-	canopy heterogeneity
	verticalDistMax	VAI_{mode}	m	vertical heterogeneity
	VCI_mean	VCI_{AVG}	-	vertical heterogeneity
2				
	VCI_mean	VCI_{AVG}	-	vertical heterogeneity
	LAI_mean	LAI_{AVG}	-	area and density
	meanZ_sd	σ_H	m	height
10				
	verticalDistMax	VAI_{mode}	m	vertical heterogeneity
	maxZ_sd	R_T	m	canopy heterogeneity
	maxZ_mean	H_{AVG}	m	height
	VCI_mean	VCI_{AVG}	-	vertical heterogeneity
25				
	maxZ_sd	R_T	m	canopy heterogeneity
	sdZ_sd	R_V	m	vertical heterogeneity
	sdZ_mean	$Vert_{meanStd}$	m	vertical heterogeneity
	LAI_sd	LAI_{SD}	-	area and density

All ten of the site averaged SC metrics varied depending upon the LiDAR return spatial resolution. The majority of SC metrics decreased in value as resolution became coarser, however, three of the ten metrics (H_{AVG}, $Vert_{meanStd}$, and VCI_{AVG}) displayed the opposite trend. With the exception of rumple, metric value changes in response to shifting resolutions were approximately linear. The observed shifts in metric values with changing resolution indicated that the overall mechanistic relationships between SC metrics and productivity could be resolution dependent. The greatest differences with shifting resolution were observed in VAI_{mode} and VCI_{AVG}. VAI_{mode} decreased with decreasing spatial resolution, with values being reduced to 25 – 30% of the 0.25 m resolution value by the time a 10 m resolution was reached, and all sites had the same value (5m) upon reaching the 25 m resolution. VCI_{AVG} increased with decreasing spatial resolution, with values increasing on average by 20% with each decrease in resolution, although the difference between 10 m and 25 m was less pronounced, with an average difference of 5%.

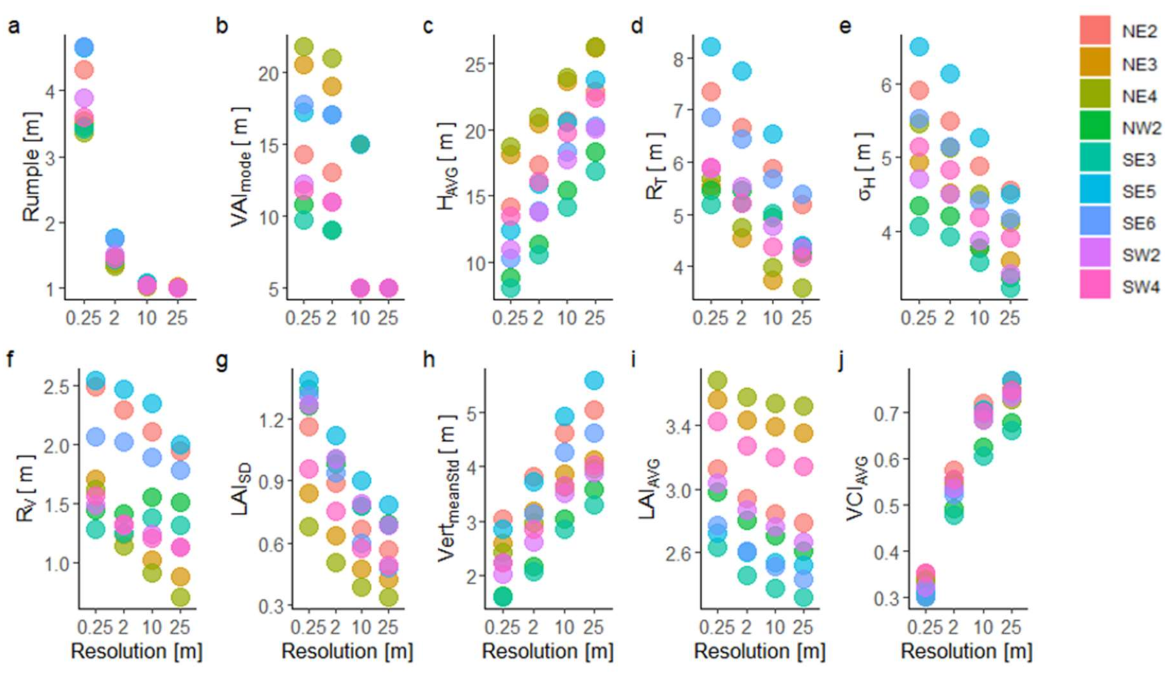


Figure 7. SC metric values by site at each of the four metric calculation resolutions explored, 0.25 m, 2 m, 10 m, and 25 m.

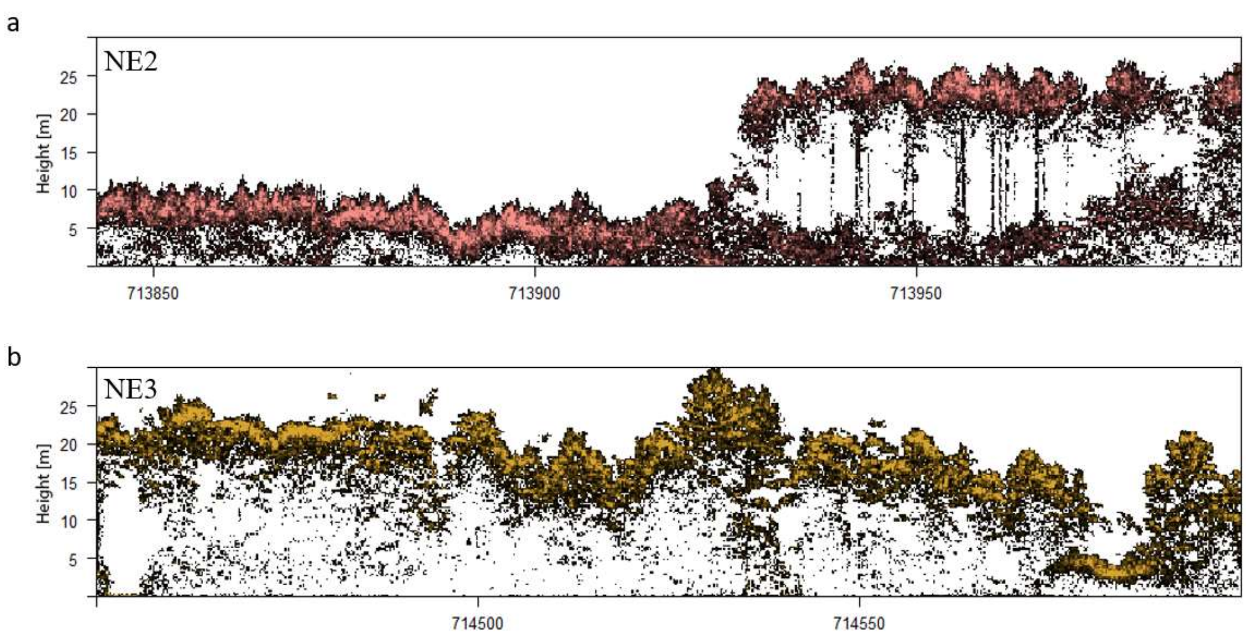
Vertical heterogeneity metrics describing the layering of the canopy such as σ_H and $Vert_{meanStd}$ generally had higher values at sites with distinct multilayered canopies such as NE2 and SE5, and lower values at sites with a more consistent single layered canopy, such as site NE4. σ_H is the standard deviation of the mean height of lidar returns for raster pixels, and conveys the variability associated with the leaf height in the canopy at which a large amount of vegetation is concentrated. $Vert_{meanStd}$ is the mean of the standard deviation of lidar return heights of raster pixels in a given transect, and it describes how vertically spread out the canopy is by way of how variable average leaf height is in a given column (Hardiman et al., 2013). A higher value indicates greater variability in how the canopy is vertically distributed. The highest values were observed at sites SW4 and NW2, and the lowest values were seen at sites SW2 and NE4, with an overall range of 1.6 m to 5.6 m across all four spatial resolutions. Variability in $Vert_{meanStd}$ values generally increases with decreasing resolution, with an increase in spread between sites of 58% from 0.25 m to 25 m resolution, whereas site-to-site variability decreased by approximately 46% with decreasing resolution for σ_H .

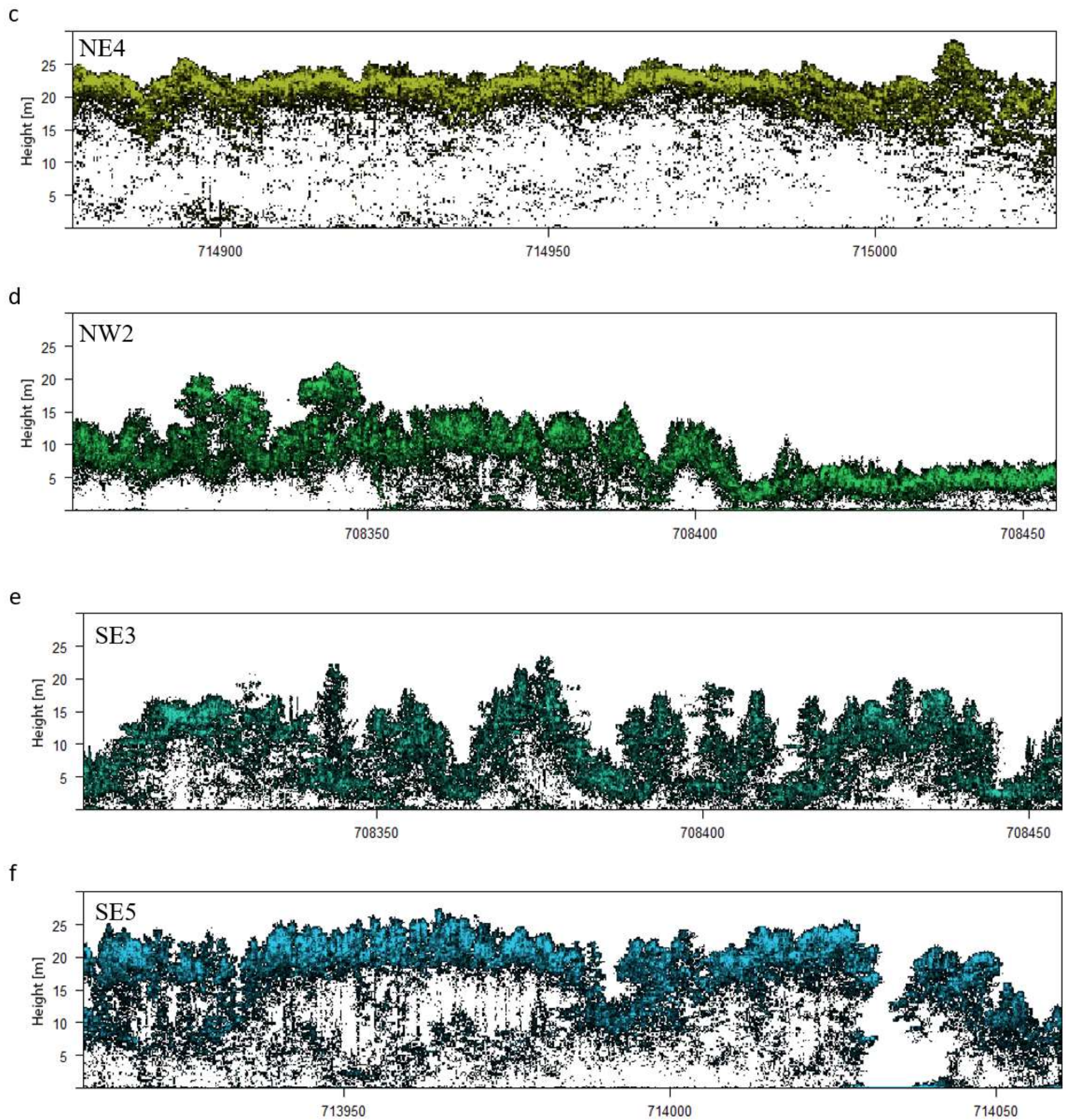
SC metrics VCI_{AVG} and vertical rugosity (R_V) offer insight as to the degree of variability in the distribution of vegetation within each vertical column. Vertical complexity index (VCI, averaged across a given stand to become VCI_{AVG}) is an ecological metric with roots in information theory (Shannon, 1948). Applied to LiDAR data, VCI_{AVG} describes how the vertical distributions of LiDAR returns differ from a uniform distribution, which is representative of the overall evenness of the vertical distribution of vegetation (van Ewijk et al., 2011), while R_V communicates the variance in each vertical column's mean leaf height variability (Atkins et al., 2018). A VCI value close to zero indicates that the distribution of points in each vertical height bin is uneven, while a value approaching or equal to one indicates an even distribution of points across height bins (van Ewijk et al., 2011). The highest VCI_{AVG} values were observed at site NE2 (9% higher than the average of the other eight sites, at a resolution of 0.25m), and the lowest values were typically seen at sites SE3 and NW2, depending upon spatial resolution. Variability between sites increased with decreasing resolution, largely in part to a widening spread between SE3 and NW2 and the remaining seven sites. The highest R_V values, indicating a less uniform vertical distribution of vegetation, were measured at sites with multilayered canopies and multiple distinct age classes present, such as SW4 and NW2, which include stands ranging in age from 7 – 110 years.

The final SC metric addressing vertical heterogeneity is VerticalDistMax. VerticalDistMax is equivalent to the variable 'mean height of vegetation area index maximum' (VAI_{mode}) described in Atkins et al., 2018, and will be referred to as such for the sake of taxonomic consistency, although it is technically a mean value, not a mode as the variable name implies. Vegetation area density (VAD) is a dimensionless variable describing the density of vegetation in a given area (Atkins et al., 2018), and vegetation area index (VAI) represents the sum of densities in a vertical column. For resolutions 0.25 m and 2 m, the highest VAI_{mode} values are seen at sites NE3 and NE4, and the lowest values are seen at sites NW2 and SE3. For 10 m resolution only two values of VAI_{mode} exist, 5 m and 15 m, with three sites (NE2, NE3, and SW4) having values of 15 m and the remaining sites having values of 5 m. By 25 m model resolution all sites have the same VAI_{mode} value of 5 m, indicating that the metric calculation resolution has a significant impact on VAI_{mode}.

Influential SC metrics representing outer canopy complexity include rumple and top rugosity (R_T). Rumple is a three-dimensional metric that describes the degree of heterogeneity associated with the outer canopy layer, where a higher value corresponds to a more complex canopy (Kane et al., 2010). Rumple is defined as the ratio of the outer canopy surface area to the underlying ground surface area (Parker et al., 2004). Average rumple values were significantly impacted by metric calculation resolution, and substantially decreased at resolutions coarser than 0.25 m, indicating that at coarser resolutions the outer canopy surface appears artificially smoothed. Variability between sites also decreased with decreasing resolution, and at resolutions coarser than 2 m, differences in rumple values between sites were negligible. For context, in a Douglas-fir and western hemlock dominated 500+ year old growth forest in Southern Washington (USA) with an extremely high level of outer canopy complexity, rumple values of 12 m were reported (Parker et al., 2004). R_T refers to the standard deviation of LiDAR column maximum return heights (Atkins et al., 2018). The highest values of R_T were observed at NE3 and NE4, with an average range of 3.6 m to 8.2 m across all four resolutions. As resolution becomes coarser the differences in values between sites becomes less pronounced, with a decrease in variability of approximately 41%.

Average tree height (H_{AVG}) serves as a simple measure of vertical stand structure, by describing the tree height averaged across all present species in a given stand. When combined with descriptive species information, H_{AVG} provides additional information regarding stand successional stage. H_{AVG} values increase with decreasing resolution, presumably because taller trees dominate and skew the average when a larger field of view is utilized. LAI_{SD} is the standard deviation of leaf area index (LAI) for each raster pixel, and LAI_{AVG} is the average LAI for each raster pixel. LAI is the ratio of the (one-sided) total leaf area per unit of ground area, and describes the amount of leaf tissue in the forest canopy. The highest values of LAI_{AVG} were observed at sites NE3 and NE4, both among the oldest sites, and the lowest values were seen at sites SE3 and SE5, both fairly young aspen sites, with an overall range of 2.3 m to 3.7 m across all four resolutions. LAI_{SD} describes the variability in LAI, and offers insight into how photosynthetic tissues are distributed in the forest canopy. The highest values were observed at site SW4, and the lowest values were seen at sites NE2 and NE3, with a total range of 0.34 to 1.4. LAI_{SD} values generally decrease with decreasing resolution, with a reduction in variability between sites (decrease in variability of 36% from 0.25 m to 25 m resolution).




c
NE4
Height [m]
25
20
15
10
5
714900
714950
715000
d
NW2
Height [m]
25
20
15
10
5
708350
708400
708450
e
SE3
Height [m]
25
20
15
10
5
708350
708400
708450
f
SE5
Height [m]
25
20
15
10
5
713950
714000
714050

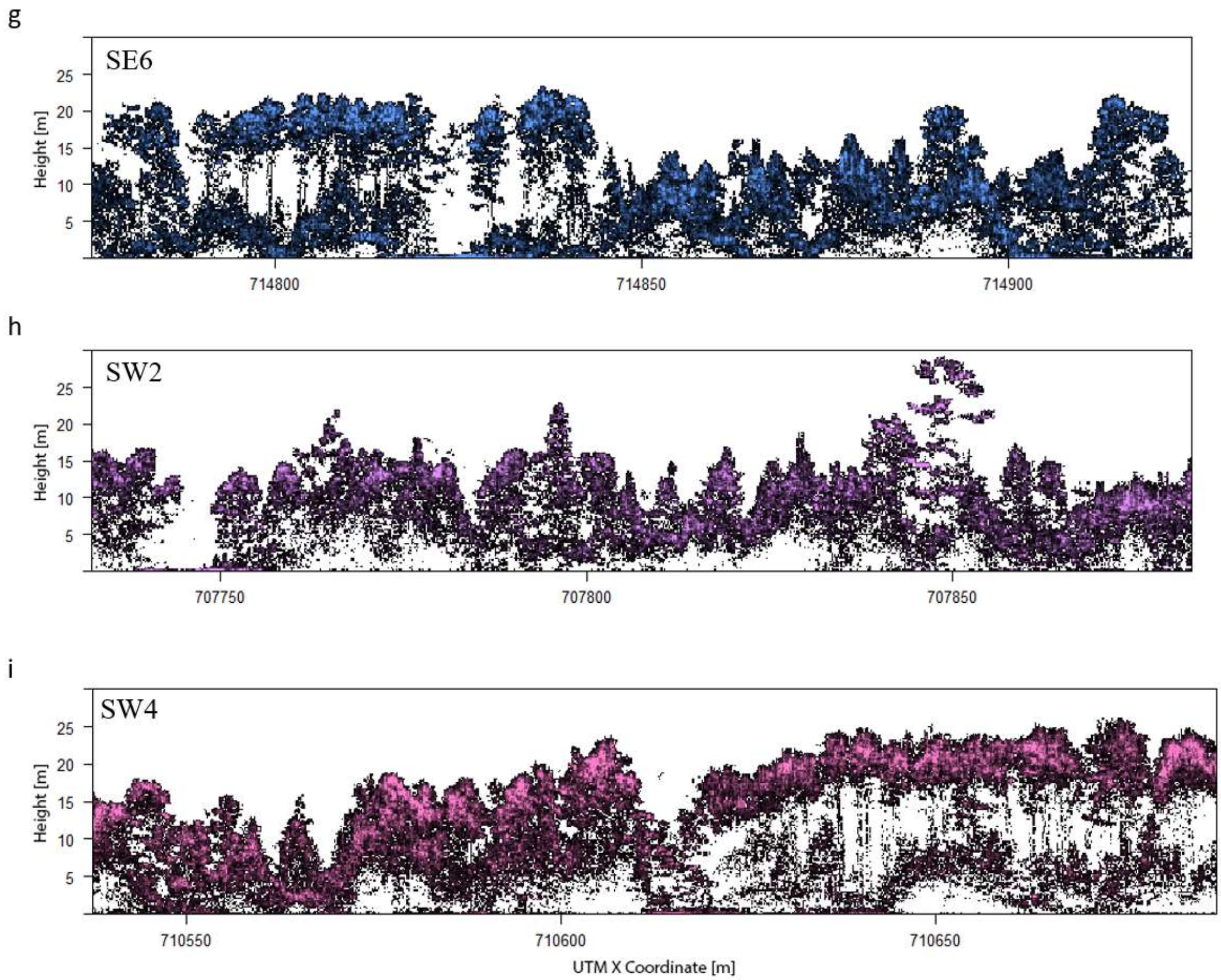


Figure 8. LiDAR point return 150 m transect images for the nine forested sites: **a)** NE2 **b)** NE3 **c)** NE4 **d)** NW2 **e)** SE3 **f)** SE5 **g)** SE6 **h)** SW2 and **i)** SW4. Color saturation represents the relative number of returns. The x-axis represents longitudinal coordinates in meters, expressed at 50 m intervals, and the y-axis is height above ground in meters.

3.3 Structural equation modeling

Comparison of SEM models showed that the reduced model, where SC metrics were restricted to influencing GPP through RUE as opposed to exerting direct influence over GPP, had a better overall fit than the fully saturated model. In other words, LUE and WUE actively mediate the mechanistic relationship between SC variables and GPP, and changes in SC result in changes in RUE and ultimately in GPP.

Across all four models 28 mediation relationships were tested in total; 14 WUE mediated relationships and 14 LUE mediated relationships. 11 of the 14 WUE mediated relationships were significant, with all cases being partial mediation, complete mediation was not observed for either WUE or LUE. WUE as a mediator between GPP and the metric VAI_{mode} was never shown to be significant, with VAI_{mode} present in both the 0.25 m and 10 m resolution best fit models. WUE as a mediator between GPP and the metric LAI_{SD}, present in the 25 m resolution model, was also shown to be insignificant. Mediation strength, characterized by the magnitude of the indirect effect of a given SC metric on GPP through WUE as a mediator was 0.10 on average, with a range of 0.14. Seven of the 14 LUE mediated relationships were significant, and metrics directly tied to light interception such as those related to LAI (LAI_{SD} and LAI_{AVG}) were always significantly mediated by LUE. LUE significantly mediated relationships between GPP and VCI_{AVG}, H_{AVG}, R_T, and $Vert_{meanStd}$ as well, but the significance of mediation was not always consistent when a given SC metric was present in different resolution models. LUE never significantly mediated relationships between GPP and VAI_{mode}, rumple, σ_H, or R_V, regardless of spatial resolution. Mediation strength of LUE on the relationship between a given SC metric and GPP was 0.02 on average, with a range of 0.08. WUE was shown to be a substantially stronger mediator between SC and GPP than LUE, with a standardized mediation strength 330% larger than that of LUE when averaged across all nine plots. Averaged across all sites, the correlation between daily WUE and daily LUE was 0.40.

Mediation analysis by resolution showed differing trends between WUE and LUE as mediators. The significance of WUE as a mediator did not appear to be resolution dependent, whereas the significance of LUE as a mediator did appear to be dependent upon the spatial resolution of the model in question. For example, LUE was a significant mediator of R_T in the 25 m resolution model, but not in the 10 m resolution model. The presence of LUE as a significant mediator was more prevalent at coarser spatial resolutions (10 m and 25 m) than at finer resolutions (0.25 m and 2 m), and cases where the structure-function relationship was mediated by both WUE and LUE were observed more frequently at courser resolutions than at finer resolutions. In summary, for SC metrics that experienced mediation (all but VAI_{mode}) the presence of a mediating factor in the overarching relationship between forest structure and function was consistent regardless of SC metric calculation resolution, but which individual relationships were significantly mediated changed with resolution shifts when LUE was the mediating variable in question. Due to the variety of measurement units involved, beta coefficients were standardized to facilitate comparison and outliers were removed. Standardized beta coefficients show that at a resolution of 0.25 m, VCI_{AVG} and rumple were the strongest drivers of GPP (β = 0.33, β = 0.11), at 2 m VCI_{AVG} was the strongest driver of GPP (β = 0.35) followed by σ_H (β = 0.16), at 10 m VCI_{AVG} and H_{AVG} were the strongest drivers of GPP (β = 0.33, β = 0.16), and at 25 m spatial resolution R_T and $Vert_{meanStd}$ were the strongest drivers (β = 0.22, β = 0.18). Additionally, all SC metrics were stronger drivers of WUE than of LUE.

4 Discussion

Our findings support the emerging consensus that a positive mechanistic relationship exists between SC and productivity in mixed temperate forests (Gough et al., 2019, Gough et al., 2016), but suggest that this is a multifaceted relationship impacted by additional factors such as the extent of species diversity and management history. Additionally, we found that this relationship is not direct but rather is mediated by the effective acquisition and assimilation of both light and water resources, and that RUE generally is enhanced by increasing SC. Furthermore, we show that in a heterogenous mixed temperate forest subject to disturbance, metrics describing the vertical profile of heterogeneity are the strongest drivers of productivity, as opposed to SC metrics that are constrained to the outer canopy. Through analysis of the structure-function relationship at four structural metric calculation resolutions ranging from 0.25 m to 25 m, we demonstrate that the scale of metric calculation has a significant impact on the metric values themselves, and thus on which SC metrics are ultimately included in predictive models of productivity. We showed that shifting the spatial resolution also changes the dynamics of the relationship between RUE and SC. Lastly, it was established that even in a study domain where sites have shared climatic and environmental conditions, differences in management and disturbance history as well as species diversity result in substantial variability in land-atmosphere exchanges of CO_2. This is likely due to changes in forest composition and trait diversity in response to disturbance.

4.1 Structural complexity

VCI_{AVG} was the most frequently observed SC metric, and consistently proved to be the most robust driver of both RUE and GPP in models where it was present, irrespective of spatial resolution. However, SC metrics related to outer canopy heterogeneity such as R_T and rumple were also imperative. Three out of the four best fit models included SC metrics related to both vertical and outer canopy heterogeneity, although vertical metrics were more prevalent in all cases. VCI_{AVG} is closely related to the degree of canopy closure as well as tree size and distribution density (Kane et al., 2010), and is known to vary with stand structure and age (Kane et al., 2008). In an example presented by van Ewijk et al. (2011), a low VCI could correspond to the stand initiation stage, where the majority of point returns are congregated in the lowest vertical bins, whereas a mid to high VCI could correspond to a stand in the midst of understory re-initiation or even a transition into old growth, where vegetation is distributed between multiple height bins. The VCI_{AVG} values observed within the study domain are consistent with the relative dominance of stands in the young to mid age classes.

The SC metric VAI_{mode} conveys important information about biomass allocation patterns. Models that did not contain VAI_{mode} did contain SC metrics related to LAI, suggesting that incorporating a variable that accounts for the complexity in arrangement of vegetative tissues is essential when describing a stand's ability to absorb incoming light. VAI is similar to the more commonly used LAI, but vegetative tissues include branches and stems in addition to photosynthesizing leaves (Scheuermann et al., 2018). However, it's worth noting that several studies have shown that the influence of LAI on production saturates in importance over time, but the same trend has not been observed in VAI (Hardiman et al., 2011), potentially making it a more reliable metric overall when describing the area-related distribution of vegetative tissues. σ_H builds on the information provided by VAI_{mode} by representing the variability associated with the height of greatest leaf density, further describing canopy layering. High values (corresponding to a multilayered canopy) were observed at sites with a variety of age classes present, where harvest practices have resulted in patches of with unique canopy features, such as site NE2 (Figure 8).

$Vert_{meanStd}$ is a reliable indicator of the spread between distinct canopy layers, high values were observed at sites such as NE2, which includes a dense canopy between 5 m – 10 m tall with an additional canopy around 25 m tall, and a fairly sparse degree of vegetation between the two canopies. Pairing this metric with R_V illustrates the variability in vertical forest profiles, and offers insight into the arrangement of the understory. For example, high values of R_V were observed at sites with dense non-uniform understories, such as site SW4. In addition to conveying information about forest successional stage when combined with species information, H_{AVG} is important to consider when interpreting the significance of observed rumple values (Kane et al., 2010), as rumple generally increases with increasing H_{AVG}. At first glance sites NE4 and NW2 could be classified as having similar levels of complexity, with rumple values of 3.4 m and 3.5 m respectively at 0.25 m resolution. However, the large differences in H_{AVG} between the sites (19 m versus 8.8 m) draws attention to the fact that the variance in complexity between the two sites is more pronounced, as a similar rumple value for a stand with less than half the H_{AVG} of NE4 indicates that NW2 has a higher degree of SC than is present at NE4.

The prevalence of vertical heterogeneity metrics focused on canopy layering and vegetation distribution in the models explored here further substantiates the claim that vertical complexity is a strong driver of productivity in mixed temperate systems (Fahey et al., 2019), and emphasizes the role of vertical variation in driving biomass growth (Stark et al., 2012). All ten influential SC metrics explored here were sensitive to changes in metric calculation resolution, highlighting the need for consistency in the spatial resolution at which SC metrics are calculated, and for the disclosure of metric calculation resolutions when reporting SC metric values and interpreting the significance of findings. For most SC metrics, values decreased as resolution became coarser (with H_{AVG}, $Vert_{meanStd}$, and VCI_{AVG} as exceptions), as did variability between sites. Moreover, differences between sites became indistinguishable for both rumple and VAI_{mode} at resolutions coarser than 10m. This signifies that for research questions centered around discerning differences in SC between sites and the potential impacts of those differences on ecosystem function, a finer resolution should be used for SC metric calculation. However, which sites are classified as most or least structurally complex overall is relatively consistent regardless of metric calculation resolution. Sites SE5 and NE2 consistently rank as the sites with the highest complexity, and sites SE3 and NW2 dependably rank as the sites with the lowest complexity. For some sites, such as NE3 and NE4, the comparative complexity ranking differs depending on which metric is being examined, for example both sites have very high complexity rankings in metrics LAI_{AVG}, VAI_{mode}, and H_{AVG} regardless of resolution, but consistently rank low in metrics R_T, LAI_{SD}, rumple, and R_V.

Ultimately SC can't be encapsulated by a single metric, and a select set of metrics will provide a more comprehensive representation. For instance, pairing a variable like σ_H that offers insight as to whether a canopy is single or multi layered with a variable like LAI_{AVG} that describes the density and arrangement of photosynthetic tissues will reveal more about a stand's potential productivity than either variable in isolation could. However, which metrics should be included in predictive models of productivity isn't a one size fits all situation, as shown here it is contingent upon spatial resolution.

4.2 The structure-function relationship

Here we showed that a positive mechanistic relationship exists between SC and forest productivity in mixed temperate forests, and that SC metrics which describe the vertical profile of heterogeneity are better predictors of GPP than metrics that are limited to the outer canopy alone. This is potentially due to vertical complexity metrics providing greater information content in terms of describing a forest's successional stage and ability to capture light as it moves beyond the outer surface of the canopy and penetrates into the forest below (Zimble et al., 2003). As early successional species overtake forest gaps created by disturbance to establish multi-canopied stands, the more biodiverse forest with greater structural complexity and range of shade tolerances will make the forest more resource efficient under variable light conditions, increasing net carbon uptake (Hardiman et al., 2011, Hardiman et al., 2013, Hooper et al., 2005). For example, NE2, which has the highest GPP, WUE, and LUE, also exhibits high levels of SC across the majority of the metrics evaluated. NE2 is predominantly pine, with aspen and paper birch intermixed (Figure 2). Due to a history of timber harvest and replanting (Figure 5) there is a significant secondary pine canopy (Figure 8) with an average age of 22 years. This multi-layered canopy is captured in the second highest values of VCI_{AVG} observed across all nine sites (at a spatial resolution of 0.25 m VCI_{AVG} = 0.35, 10% higher than the following seven sites), while H_{AVG} and rumple were also comparatively high, at 10% higher than average and 4.4% higher than average respectively.

SEM highlighted WUE as a considerably stronger driver of GPP than LUE, but it's important to pause here and consider that the temperate mixed forests of Northern Wisconsin are not water limited ecosystems, and previous studies have shown that stand-scale productivity is predominantly a function of the capacity to harvest light and fix carbon (Reich et al., 2012), so why does WUE show up as highly influential when predicting GPP? The answer lies primarily in the relationship between WUE and LUE. The tiny stomata covering the leaf surface exist in a constant tradeoff between opening and sacrificing water for the chance to take up CO_2, both of which are necessary ingredients for photosynthesis (Monteith, 1965). Regardless of available light, when plants are water stressed, stomata close in an attempt to conserve existing resources, at the cost of reducing CO_2 uptake and thus photosynthetic capacity (Hatfield and Dold, 2019, Kukal and Irmak, 2020). However, when a plant has a steady supply of water, stomata will more readily open and a greater amount of atmospheric CO_2 can be fixed per unit of incident light (Binkley et al., 2004). A recent study by Ehbrecht et al. (2021) examining climatic controls on SC at the global scale found that SC was strongly correlated with water availability across all biomes examined, and that the relationship between water availability and use and SC can be tied to mechanisms determining tree size. This is because water availability effectively controls functional diversity and shade tolerance as well as tree size following the hydrological limitation hypothesis. Shade tolerant trees are found in greater abundance in systems where growth is not limited by factors other than light, meaning non-water limited systems, as is the case in Northern Wisconsin. All three of these factors (functional diversity, shade tolerance, tree size) contribute to SC (Thom et al., 2021).

However, the importance of the relationship between SC and LUE cannot be understated, as it shows that the functional diversity driven by complexity is able to better capitalize on available resources (Williams et al., 2017, Penone et al., 2019). Additionally, although this study was limited in duration, other studies such as the Zhang et al., 2012 global meta-analysis of diversity productivity relationships showed that almost 30% of the variation in productivity between monocultures and polycultures was explained by heterogeneity of shade tolerance, and that high shade tolerance variation within a community is likely one of the most important life-history traits, leading to more efficient resource use when scaled to the ecosystem level (Stark et al., 2012).

For most SC metrics examined here, increasing SC is associated with increasing RUE, although the magnitude of the trend is dependent upon resolution. The exception is LAI_{SD}, which has a negative relationship with both WUE and LUE at all resolutions. The strongest positive relationship exists between VCI_{AVG} and WUE, and the weakest relationship exists between R_T and LUE. Mediation analysis showed that neither WUE or LUE significantly mediated the relationship between VAI_{mode} and GPP, suggesting that either the relationship is direct, or additional unaccounted for factors play the role of mediator. The most complex sites (SE5 and NE2) have differing relationships to productivity. Site NE2 has the highest GPP of all nine sites, but also has the highest R_{eco}, resulting in its classification as a small net source of CO_2 to the atmosphere. Site SE5 has the second lowest seasonal GPP as well as the second lowest R_{eco}. The two least complex sites, SE3 and NW2, have among the lowest total seasonal GPP and R_{eco}. SW2 and SW4 have the second and third highest seasonal GPP, yet consistently display only moderately levels of SC at all four spatial resolutions. However, both of these sites contain stands in a wide range of age classes (Figure 4), indicating heterogeneity in successional stages, and both sites are noted as containing very wet areas, with older (>100 years) mixed conifer swamp stands.

4.3 Disturbance impacts

Whether or not forests develop structural complexity in response to a disturbance event depends on the frequency, scale, and intensity of the event (Ehbrecht et al., 2021, Ford and Keeton, 2017). Small scale disturbances tend to increase complexity by creating favorable conditions for understory trees to establish, which results in multi-layered canopies (Wisconsin Department of Natural Resources, 2020). This amplified sub-canopy growth occurs because disturbance drives a compensatory physiological response to more readily available light, which can also help sustain overall production even in the face of frequent low intensity disturbances (Hardiman et al., 2013). In contrast, larger scale disturbances tend to simplify SC initially leading to a temporary reduction in productivity, although stands often recover to pre-disturbance carbon uptake levels within the 10 – 20 years following a major disturbance event (Amiro et al., 2010).

With the exception of NE2, sites with a record of intensive disturbance, presented as clearcutting or shelterwood harvest, exhibit lower levels of complexity across the majority of SC metrics, and across all metrics addressing vertical complexity. One reason for this could be that the harvests at NE2 were all selective harvests, and resulted in distinct structurally heterogeneous 'patches' within the site at different successional stages and with a high degree of canopy cover. In contrast to the primarily deciduous understory present at multiple other sites, several patches within NE2 feature a prominent conifer understory. As mixed conifers tend to show higher levels of vertical complexity than many purely deciduous stands do (Ehbrecht et al., 2017, Pommerening and Murphy, 2004, Zenner et al., 2012), the presence of a developing conifer understory could be contributing to a higher overall VCI_{AVG}. This is supported by the presence of a substantial conifer understory at one other site, SW4, which exhibits the highest degree of VCI_{AVG} amongst the nine sites (0.35). Again, with the exception of NE2, sites with a record of more substantial disturbance had lower levels of productivity, and lower levels of RUE. For example, site NW2, which had the highest intensity of both clear cuts and harvest events (Figure 5), had the lowest GPP of all nine sites and also had the lowest average daily LUE (0.33 gC MJ^{-1}) and WUE (2.9 gC kg H_2O^{-1}) values.

More moderate disturbances such as thinning and selective harvest could be contributing to increased SC within the study area, through assisting in the transition to uneven aged stands (Gough et al., 2021). This is observed at site SE6, which consists of a 19-year-old mixed aspen, white spruce, and balsam fir stand, a 22-year-old jack pine stand, a 75-year-old aspen stand, and a 92-year-old mixed upland hardwood stand (Figure 2). SE6 underwent species-specific commercial thinning to reduce stand density, which has been shown to impact stand growth and structure (Wisconsin Department of Natural Resources, 2020). SE6 also experienced salvage cutting to remove dead or damaged trees following a severe hail storm in 2000. Sites SE5 and NE2 consistently ranked as the most complex sites regardless of spatial resolution, and both sites have experienced moderate management disturbances such as thinning as well as manual planting.

4.5 Implications and shortcomings

A 2018 review by Fahey et al. showed that although the prevalence of complexity terminology with respect to silviculture has increased over time, the actual incorporation of complexity metrics when designing long term silviculture projects has “plateaued in the past decade or more”. This could indicate that although awareness about the importance of forest complexity has increased, a lack of understanding regarding the long-term impacts of managing to enhance complexity persists. Through the exploration of mechanistic relationships between forest SC and function, this study highlighted which complexity metrics provide important information about RUE and productivity. These metrics can then be integrated as flexible structural parameters in mechanistic ecosystem models that simulate light and water-sensitive processes. Through this, we can improve the ability of models to mimic true ecosystem responses to management, from a biogeochemical perspective. This improved representation will allow us to explore the future response of forests to a variety of management regimes and representative concentration pathways, enhancing our ability to assess mitigation and adaptation strategies beyond direct observational studies, which often take many years to produce outcomes. This would facilitate more accurate predictions of the future of a suite of ecosystem goods and services including carbon storage potential, which could significantly impact the development of natural climate solutions in the regional Midwest.

The persistent superior performance of the reduced SEM, where the relationship between SC and GPP is moderated by RUE, suggests that although specific SC metric values change slightly with metric calculation resolution shifts, the existence of a mediation effect itself is not scale dependent. This indicates that the mechanistic relationships outlined here can be scaled up from the stand to the ecosystem level to provide novel insights into forest function and carbon storage potential. While this expands the utility of observational studies, it also provides new opportunities to validate and apply information obtained from satellites, such as the Global Ecosystem Dynamics Investigation (GEDI) high resolution ecosystem LiDAR, which is capable of measuring global forest canopy height and vertical structure (Dubayah et al., 2020).

Although the EC method is the most well-established method for taking continuous measurements of energy and trace gas exchange (Desai et al., 2008), it is not without drawbacks. All measurements have associated uncertainty, and in the context of EC measurements these uncertainties can be segregated into several categories, depending on the type of error from which they are derived. These categories include instrument error, calibration error, technological limitations of the instruments themselves, inadequate sample size, and environmental conditions that violate the assumptions at the core of EC theory (Richardson et al., 2012). Some of these errors are stochastic and appear as random noise in the data, while other errors are systematic and result in a bias that is relatively constant over time. Numerous other studies have explored these uncertainties at length (Loescher et al., 2006, Hollinger and Richardson, 2005, Massman and Lee, 2002, Richardson et al., 2006), but it's worth noting general trends in overall EC uncertainty here. Random error in 30-minute fluxes ranges from 10 – 20% (Loescher et al., 2006), with annual estimates around 10% (Richardson et al., 2006), as error generally decreases with longer time series and averaging (Loescher et al., 2006). Flux uncertainty follows a strong seasonal pattern (uncertainty is generally higher during the growing season), and is sensitive to land cover type and wind speed (Richardson et al., 2006, Hollinger and Richardson, 2005). Error is also associated with the partitioning of NEE into GPP and R_{eco} and varies by partitioning method, but a survey of 23 methods conducted by Desai et. al (2008) showed that on average the difference in GPP was less than 10%, with additional uncertainty depending on the abundance of gaps in the data. In this study, there was an average of 37% gaps in measured NEE values across the nine sites.

This study primarily examined the influence of biotic forest factors, but the inclusion of prominent abiotic factors such as nutrient regimes could further enhance the study. Combining chemical analysis of leaves with the remote sensing of SC and EC measurements of land-atmosphere carbon exchange would account for the influence of factors such as nitrogen availability in determining controls on RUE and productivity (Reich et al., 2012). Another limitation of this study is the relatively short window in which data was collected. The measurement period was designated as June through October to align with the seasonal shift in the domination of the surface energy balance from latent to sensible heat flux. Although this observational window supported the primary goals of CHEESEHEAD19 related to addressing issues of energy balance closure, from a carbon cycle perspective it failed to capture winter effects on net carbon budgets. Incorporating multi-year datasets would address this problem as well as allow for a more thorough examination of the influence of stand age on RUE and productivity, whereas here analysis was inconclusive. Although the high density of EC towers in a small study domain controlled for several factors such as differences in soil type, forest type, and mesoclimate, differences in microclimate still existed between sites. This is presented as variability in temperature, latent and sensible heat flux, and wind properties including turbulence. Although heterogeneity in land cover existed, there was very little difference in topography to drive variability in air circulation or relative humidity, so the observed differences in microclimate were likely due to diversity in vegetation type and density, as well as proximity to and abundance of water. Lastly, the somewhat small site sample size involved in this study suggests caution should be exercised when evaluating SEM fit statistics.

5 Conclusions

Quantifying mechanistic relationships between forest SC and productivity is essential to advancing our ability to scale measurements from the leaf to stand to landscape level. This will greatly enhance our capacity to directly assess landscape-level ecosystem functions and implications for natural climate solutions. We approached this challenge using a combination of UAS LiDAR-derived SC metrics, land-atmosphere exchange data from nine EC towers, and SEM. Through employing a high density of EC towers across a 10 x 10 km study domain, we were able to separate variability in climate, soil fertility, and forest functional types from structural controls on productivity, allowing for a more representative physiological understanding than has been previously demonstrated. We conclude that (i) structural metrics describing the vertical complexity of a forest (specifically VCI_{AVG}) are the strongest drivers when predicting productivity in temperate mixed forests with a significant degree of heterogeneity and a long history of management; (ii) variability in the type and intensity of management and disturbance legacies contribute to substantial differences in SC metric values as well as productivity; (iii) the relationship between forest structure and function isn't direct, but is actively mediated by light and water RUE, with WUE being a stronger driver of GPP; and (iv) SC metric values change with shifts in resolution, resulting in changes to the mechanistic relationship between forest structure and function. This emphasizes the need for consistency in the spatial resolution at which SC metrics are calculated, and for the disclosure of metric calculation resolutions when reporting SC metric values and interpreting the significance of findings. These findings will allow us to improve mechanistic representation in ecosystem models of how SC impacts light and water-sensitive processes, and ultimately GPP. This will strengthen the ability of models to mimic true ecosystem responses to management, allowing for a more accurate assessment of the response of forests to various management regimes and representative concentration pathways, enhancing our ability to assess mitigation and adaptation strategies.

Acknowledgments

This project was financially supported by award AGS‑1822420 from the National Science Foundation (NSF), the University of Wisconsin-Madison's Office of the Vice Chancellor for Research and Graduate Education (OVCRGE) fall competition award and NSF Emerging Frontiers Macrosystems Biology award DEB‑1702996 supporting the Management and Disturbance in Forest Ecosystems (MANDIFORE) project. We would like to acknowledge that the study domain is located in the ancestral homeland of the Ojibwe people. We honor the indigenous caretakers of these lands before us, today, and of generations to come.

Data Availability Statement

Drone LiDAR data is available at http://co2.aos.wisc.edu/data/CHEESEHEAD-incoming/Drone_LiDAR/ . All flux and meteorological data utilized in this study, in addition to other datasets generated through CHEESEHEAD19 are publicly available through the

CHEESEHEAD19 data repository hosted by the National Center for Atmospheric Research's Earth Observing Laboratory at https://www.eol.ucar.edu/field_projects/cheesehead. Supplementary information and photographs of the tower sites are available through the CHEESEHEAD19 project website: www.cheesehead19.org. EC tower data is also publicly available through the Ameriflux website, digital object identifiers (DOI) for the nine towers utilized in this study are presented in Table 5. The authors declare no conflicts of interest pertaining to this study.

Table 4. EC flux towers included in this study, with unique DOI's

CHEESEHEAD Tower Name	Ameriflux Tower Name	Dataset DOI
NW2	US-PFc	https://doi.org/10.17190/AMF/1717851
NE2	US-PFh	https://doi.org/10.17190/AMF/1717855
NE3	US-PFi	https://doi.org/10.17190/AMF/1717856
NE4	US-PFj	https://doi.org/10.17190/AMF/1717857
SW2	US-PFl	https://doi.org/10.17190/AMF/1717859
SW4	US-PFn	https://doi.org/10.17190/AMF/1717861
SE3	US-PFq	https://doi.org/10.17190/AMF/1717863
SE5	US-PFs	https://doi.org/10.17190/AMF/1717865
SE6	US-PFt	https://doi.org/10.17190/AMF/1717866

References

Anderson-Teixeira, K. J., Herrmann, V., Banbury Morgan, R., Bond-Lamberty, B., Cook-Patton, S. C., Ferson, A. E., Muller-Landau, H. C., & Wang, M. M. H. (2021). Carbon cycling in mature and regrowth forests globally. *Environmental Research Letters*, *16*(5), 053009. https://doi.org/10.1088/1748-9326/abed01

Anten, N. P. R. (2016). Optimization and Game Theory in Canopy Models. In K. Hikosaka, Ü. Niinemets, & N. P. R. Anten (Eds.), *Canopy Photosynthesis: From Basics to Applications* (pp. 355–377). Springer Netherlands. https://doi.org/10.1007/978-94-017-7291-4_13

Arnfield, A. J. (n.d.). *Koppen climate classification | Definition, System, & Map*. Encyclopedia Britannica. Retrieved April 22, 2021, from https://www.britannica.com/science/Koppen-climate-classification

Atkins, J. W., Bohrer, G., Fahey, R. T., Hardiman, B. S., Morin, T. H., Stovall, A. E. L., Zimmerman, N., & Gough, C. M. (2018). Quantifying vegetation and canopy structural complexity from terrestrial Li DAR data using the FORESTR R package. *Methods in Ecology and Evolution*, *9*(10), 2057–2066. https://doi.org/10.1111/2041-210X.13061

Atkins, J. W., Fahey, R. T., Hardiman, B. H., & Gough, C. M. (2018). Forest Canopy Structural Complexity and Light Absorption Relationships at the Subcontinental Scale. *Journal of Geophysical Research: Biogeosciences*, *123*(4), 1387–1405. https://doi.org/10.1002/2017JG004256

Binkley, D., Stape, J. L., & Ryan, M. G. (2004). Thinking about efficiency of resource use in forests. *Forest Ecology and Management*, *193*(1–2), 5–16. https://doi.org/10.1016/j.foreco.2004.01.019

Birdsey, R., Pan, Y., Janowiak, M., Stewart, S., Hines, S., Parker, L., Gower, S., Lichstein, J., McCullough, K., Zhang, F., Chen, J., Mladenoff, D., Wayson, C., & Swanston, Chris. (2014). *Past and prospective carbon stocks in forests of northern Wisconsin: A report from the Chequamegon-Nicolet National Forest Climate Change Response Framework* (NRS-GTR-127;

p. NRS-GTR-127). U.S. Department of Agriculture, Forest Service, Northern Research Station. https://doi.org/10.2737/NRS-GTR-127

Birdsey, R., Pregitzer, K., & Lucier, A. (2006). Forest Carbon Management in the United States: 1600-2100. *Journal of Environmental Quality*, *35*(4), 1461–1469. https://doi.org/10.2134/jeq2005.0162

Bogdanovich, E., Perez-Priego, O., El-Madany, T. S., Guderle, M., Pacheco-Labrador, J., Levick, S. R., Moreno, G., Carrara, A., Pilar Martín, M., & Migliavacca, M. (2021). Using terrestrial laser scanning for characterizing tree structural parameters and their changes under different management in a Mediterranean open woodland. *Forest Ecology and Management*, *486*, 118945. https://doi.org/10.1016/j.foreco.2021.118945

Butterworth, B. J., Desai, A. R., Metzger, S., Townsend, P. A., Schwartz, M. D., Petty, G. W., Mauder, M., Vogelmann, H., Andresen, C. G., Augustine, T. J., Bertram, T. H., Brown, W. O. J., Buban, M., Cleary, P., Durden, D. J., Florian, C. R., Iglinski, T. J., Kruger, E. L., Lantz, K., … Zheng, T. (2021). Connecting Land–Atmosphere Interactions to Surface Heterogeneity in CHEESEHEAD19. *Bulletin of the American Meteorological Society*, *102*(2), E421–E445. https://doi.org/10.1175/BAMS-D-19-0346.1

Desai, A. R. (2010). Climatic and phenological controls on coherent regional interannual variability of carbon dioxide flux in a heterogeneous landscape. *Journal of Geophysical Research*, *115*, G00J02. https://doi.org/10.1029/2010JG001423

Desai, A. R., Richardson, A. D., Moffat, A. M., Kattge, J., Hollinger, D. Y., Barr, A., Falge, E., Noormets, A., Papale, D., Reichstein, M., & Stauch, V. J. (2008). Cross-site evaluation of eddy covariance GPP and RE decomposition techniques. *Agricultural and Forest Meteorology*, *148*(6), 821–838. https://doi.org/10.1016/j.agrformet.2007.11.012

Ehbrecht, M., Schall, P., Ammer, C., & Seidel, D. (2017). Quantifying stand structural complexity and its relationship with forest management, tree species diversity and microclimate. *Agricultural and Forest Meteorology*, *242*, 1–9. https://doi.org/10.1016/j.agrformet.2017.04.012

Eitel, J. U. H., Höfle, B., Vierling, L. A., Abellán, A., Asner, G. P., Deems, J. S., Glennie, C. L., Joerg, P. C., LeWinter, A. L., Magney, T. S., Mandlburger, G., Morton, D. C., Müller, J., & Vierling, K. T. (2016). Beyond 3-D: The new spectrum of lidar applications for earth and ecological sciences. *Remote Sensing of Environment*, *186*, 372–392. https://doi.org/10.1016/j.rse.2016.08.018

Fahey, R. T., Alveshere, B. C., Burton, J. I., D'Amato, A. W., Dickinson, Y. L., Keeton, W. S., Kern, C. C., Larson, A. J., Palik, B. J., Puettmann, K. J., Saunders, M. R., Webster, C. R., Atkins, J. W., Gough, C. M., & Hardiman, B. S. (2018). Shifting conceptions of complexity in forest management and silviculture. *Forest Ecology and Management*, *421*, 59–71. https://doi.org/10.1016/j.foreco.2018.01.011

Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S. R., Park, H., & Shao, C. (2016). Applications of structural equation modeling (SEM) in ecological studies: An updated review. *Ecological Processes*, *5*(1), 19. https://doi.org/10.1186/s13717-016-0063-3

Ford, S. E., & Keeton, W. S. (2017). Enhanced carbon storage through management for old-growth characteristics in northern hardwood-conifer forests. *Ecosphere*, *8*(4). https://doi.org/10.1002/ecs2.1721

Forrester, J.A., Mladenoff, D.J. & Gower, S.T. (2013). Experimental Manipulation of Forest Structure: Near-Term Effects on Gap and Stand Scale C Dynamics. *Ecosystems,* (16), 1455–1472. https://doi-org.ezproxy.library.wisc.edu/10.1007/s10021-013-9695-7

Frelich, L. E. (1995). Old Forest in the Lake States Today and before European Settlement. *Natural Areas Journal*, *15*(2), 157–167.

Gough, C. M., Vogel, C. S., Harrold, K. H., George, K., & Curtis, P. S. (2007). The legacy of harvest and fire on ecosystem carbon storage in a north temperate forest. *Global Change Biology*, *13*(9), 1935–1949. https://doi.org/10.1111/j.1365-2486.2007.01406.x

Gough, C.M., Curtis, P.S., Hardiman, B.S., Scheuermann, C.M. and Bond-Lamberty, B. (2016), Disturbance, complexity, and succession of net ecosystem production in North America's temperate deciduous forests. *Ecosphere*, 7: e01375. https://doiorg.ezproxy.library.wisc.edu/10.1002/ecs2.1375

Gough, C. M., Atkins, J. W., Fahey, R. T., and Hardiman, B. S. (2019). High rates of primary production in structurally complex forests. *Ecology,* 100(10): e02864. 10.1002/ecy.2864

Gough, C. M., Bohrer, G., Hardiman, B. S., Nave, L. E., Vogel, C. S., Atkins, J. W., Bond-Lamberty, B., Fahey, R. T., Fotis, A. T., Grigri, M. S., Haber, L. T., Ju, Y., Kleinke, C. L., Mathes, K. C., Nadelhoffer, K. J., Stuart-Haëntjens, E., and Curtis, P. S. (2021). Disturbance-accelerated succession increases the production of a temperate forest. *Ecological Applications,* 31(7):e02417. 10.1002/eap.2417

Hatfield, J. L., & Dold, C. (2019). Water-Use Efficiency: Advances and Challenges in a Changing Climate. *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/fpls.2019.00103

Hollinger, D. Y., & Richardson, A. D. (2005). Uncertainty in eddy covariance measurements and its application to physiological models. *Tree Physiology*, *25*(7), 873–885. https://doi.org/10.1093/treephys/25.7.873

Hooper, D. U., Chapin, F. S., Ewel, J. J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J. H., Lodge, D. M., Loreau, M., Naeem, S., Schmid, B., Setälä, H., Symstad, A. J., Vandermeer, J., &

Wardle, D. A. (2005). Effects of Biodiversity on Ecosystem Functioning: A Consensus of Current Knowledge. *Ecological Monographs*, *75*(1), 3–35. https://doi.org/10.1890/04-0922

Kane, V. R., Bakker, J. D., McGaughey, R. J., Lutz, J. A., Gersonde, R. F., & Franklin, J. F. (2010). Examining conifer canopy structural complexity across forest ages and elevations with LiDAR data. *Canadian Journal of Forest Research*, *40*(4), 774–787. https://doi.org/10.1139/X10-064

Kane, V. R., McGaughey, R. J., Bakker, J. D., Gersonde, R. F., Lutz, J. A., & Franklin, J. F. (2010). Comparisons between field- and LiDAR-based measures of stand structural complexity. *Canadian Journal of Forest Research*, *40*(4), 761–773. https://doi.org/10.1139/X10-024

Knauer, J., El-Madany, T. S., Zaehle, S., & Migliavacca, M. (2018). Bigleaf—An R package for the calculation of physical and physiological ecosystem properties from eddy covariance data. *PLoS ONE*, *13*(8), e0201114. https://doi.org/10.1371/journal.pone.0201114

Kukal, M. S., & Irmak, S. (2020). Interrelationships between water use efficiency and light use efficiency in four row crop canopies. *Agrosystems, Geosciences & Environment*, *3*(1), e20110. https://doi.org/10.1002/agg2.20110

Loescher, H. W., Law, B. E., Mahrt, L., Hollinger, D. Y., Campbell, J., & Wofsy, S. C. (2006). Uncertainties in, and interpretation of, carbon flux estimates using the eddy covariance technique. *Journal of Geophysical Research: Atmospheres*, *111*(D21). https://doi.org/10.1029/2005JD006932

Massman, W. J., & Lee, X. (2002). Eddy covariance flux corrections and uncertainties in long-term studies of carbon and energy exchanges. *Agricultural and Forest Meteorology*, *113*(1–4), 121–144. https://doi.org/10.1016/S0168-1923(02)00105-3

Mathias, J. M., & Thomas, R. B. (2021). Global tree intrinsic water use efficiency is enhanced by increased atmospheric CO_2 and modulated by climate and plant functional types. *Proceedings of the National Academy of Sciences*, *118*(7), e2014286118. https://doi.org/10.1073/pnas.2014286118

McElhinny, C., Gibbons, P., Brack, C., & Bauhus, J. (2005). Forest and woodland stand structural complexity: Its definition and measurement. *Forest Ecology and Management*, *218*(1–3), 1–24. https://doi.org/10.1016/j.foreco.2005.08.034

Odum, E. P. (1969). The Strategy of Ecosystem Development. *Science*, *164*(3877), 262–270.

Olson, R. J., Holladay, S. K., Cook, R. B., Falge, E., Baldocchi, D., & Gu, L. (2004). *FLUXNET. Database of fluxes, site characteristics, and flux-community information* (ORNL/TM-2003/204). Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States). https://doi.org/10.2172/1184413

Pan, Y., Chen, J. M., Birdsey, R., McCullough, K., He, L., & Deng, F. (2011). Age structure and disturbance legacy of North American forests. *Biogeosciences*, *8*(3), 715–732. https://doi.org/10.5194/bg-8-715-2011

Parker, G. G., Harmon, M. E., Lefsky, M. A., Chen, J., Pelt, R. V., Weis, S. B., Thomas, S. C., Winner, W. E., Shaw, D. C., & Frankling, J. F. (2004). Three-dimensional Structure of an Old-growth Pseudotsuga-Tsuga Canopy and Its Implications for Radiation Balance, Microclimate, and Gas Exchange. *Ecosystems*, *7*(5). https://doi.org/10.1007/s10021-004-0136-5

R Core Team. (2021). *R: A language and environment for statistical computing* (4.0.4) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Reichstein, M., Stoy, P. C., Desai, A. R., Lasslop, G., & Richardson, A. D. (2012). Partitioning of Net Fluxes. In *Eddy covariance: A practical guide to measurement and data analysis*. Springer Science and Business Media.

Rhemtulla, J. M., Mladenoff, D. J., & Clayton, M. K. (2009). Legacies of historical land use on regional forest composition and structure in Wisconsin, USA (mid-1800s–1930s–2000s). *Ecological Applications*, *19*(4), 1061–1078. https://doi.org/10.1890/08-1453.1

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Roussel, J.-R., Auty, D., Coops, N. C., Tompalski, P., Goodbody, T. R. H., Meador, A. S., Bourdon, J.-F., de Boissieu, F., & Achim, A. (2020). lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sensing of Environment*, *251*, 112061. https://doi.org/10.1016/j.rse.2020.112061

Scheuermann, C. M., Nave, L. E., Fahey, R. T., Nadelhoffer, K. J., & Gough, C. M. (2018). Effects of canopy structure and species diversity on primary production in upper Great Lakes forests. *Oecologia*, *188*(2), 405–415. https://doi.org/10.1007/s00442-018-4236-x

Schneider, F. D., Morsdorf, F., Schmid, B., Petchey, O. L., Hueni, A., Schimel, D. S., & Schaepman, M. E. (2017). Mapping functional diversity from remotely sensed morphological and physiological forest traits. *Nature Communications*, *8*(1), 1441. https://doi.org/10.1038/s41467-017-01530-3

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(4), 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x

Thom, D., Taylor, A. R., Seidl, R., Thuiller, W., Wang, J., Robideau, M., & Keeton, W. S. (2021). Forest structure, not climate, is the primary driver of functional diversity in northeastern

North America. *Science of The Total Environment*, *762*, 143070. https://doi.org/10.1016/j.scitotenv.2020.143070

Wisconsin Department of Natural Resources. (2019). Wisconsin Wiscland 2 landcover database level 4. https://data-wi-dnr.opendata.arcgis.com

Wisconsin Department of Natural Resources. (2020). Silviculture Handbook. Forest Economics and Ecology, Applied Forestry Bureau.

Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Šigut, L., Menzer, O., & Reichstein, M. (2018). Basic and extensible post-processing of eddy covariance flux data with REddyProc. *Biogeosciences*, *15*(16), 5015–5030. https://doi.org/10.5194/bg-15-5015-2018

Zhang, W., Qi, J., Wan, P., Wang, H., Xie, D., Wang, X., & Yan, G. (2016). An Easy-to-Use Airborne LiDAR Data Filtering Method Based on Cloth Simulation. *Remote Sensing*, *8*(6), 501. https://doi.org/10.3390/rs8060501

Zimble, D. A., Evans, D. L., Carlson, G. C., Parker, R. C., Grado, S. C., & Gerard, P. D. (2003). Characterizing vertical forest structure using small-footprint airborne LiDAR. *Remote Sensing of Environment*, *87*(2–3), 171–182. https://doi.org/10.1016/S0034-4257(03)00139-1