A hyperspectral image can predict tropical tree growth rates in single-species stands

T. Trevor Caughlin,1,6 Sarah J. Graves,1 Gregory P. Asner,2 Michel van Breugel,3,4 Jefferson S. Hall,4 Roberta E. Martin,2 Mark S. Ashton,5 and Stephanie A. Bohman1,4

1 School of Forest Resources and Conservation, University of Florida, Gainesville, Florida 32601 USA
2 Department of Global Ecology, Carnegie Institution for Science, Stanford, California 94305 USA
3 Yale-NUS College, Singapore, Singapore
4 Smithsonian Tropical Research Institute, Apartado, Panama, 0843-03092 Panama
5 Yale School of Forestry & Environmental Studies, New Haven, Connecticut 06511 USA

Abstract. Remote sensing is increasingly needed to meet the critical demand for estimates of forest structure and composition at landscape to continental scales. Hyperspectral images can detect tree canopy properties, including species identity, leaf chemistry and disease. Tree growth rates are related to these measurable canopy properties but whether growth can be directly predicted from hyperspectral data remains unknown. We used a single hyperspectral image and light detection and ranging-derived elevation to predict growth rates for 20 tropical tree species planted in experimental plots. We asked whether a consistent relationship between spectral data and growth rates exists across all species and which spectral regions, associated with different canopy chemical and structural properties, are important for predicting growth rates. We found that a linear combination of narrowband indices and elevation is correlated with standardized growth rates across all 20 tree species ($R^2 = 53.70\%$). Although wavelengths from the entire visible-to-shortwave infrared spectrum were involved in our analysis, results point to relatively greater importance of visible and near-infrared regions for relating canopy reflectance to tree growth data. Overall, we demonstrate the potential for hyperspectral data to quantify tree demography over a much larger area than possible with field-based methods in forest inventory plots.

Key words: canopy biology; field planting trials; forest dynamics; hyperspectral; light detection and ranging; Panama; plantation; precision forestry; reforestation; remote sensing; tree demography; tropical forest.

Introduction

Measuring forest dynamics over large areas is needed to understand ecological processes, including the terrestrial carbon cycle (Antonarakis et al. 2014) and forest succession (Holl and Aide 2011), at landscape and regional scales. In particular, tree growth and survival rates are keys to understanding population and community dynamics that lead to changes in biomass and species composition over time (Purves and Pacala 2008, Caughlin et al. 2016). Tree growth is most often measured in field-based forest inventory plots, using repeated stem measurements of diameter at breast height ($\Delta$DBH). Reliance on forest inventory plots for models of forest dynamics is problematic when applied to large spatial extents, because logistical constraints limit field measurements of $\Delta$DBH, and extrapolating small-scale plot data to larger scales can propagate spatial errors (Marvin et al. 2014). Accurate estimates of tree demographic rates, including $\Delta$DBH, over large spatial extents could greatly improve forest dynamics models (Antonarakis et al. 2014) and ultimately contribute to more informed forest management at large scales (Holl and Aide 2011). For example, large (>100,000 ha) tree plantations are an increasingly common land use across the tropics (Fagan et al. 2015). While improved management of these plantations could benefit biodiversity conservation and limit environmental degradation (Hartley 2002), their large spatial extent challenges the ability of plot-based inventories to monitor program success and allocate site-level interventions (Kellndorfer et al. 2003).

Remote-sensing data could provide a solution to the spatial mismatch between land management and forest
inventory data. For example, time series data on forest structure from high-spatial resolution aerial images have detected emergent tree mortality (Clark et al. 2004), quantified gap dynamics (Kellner and Asner 2014) and forest height and biomass dynamics (Dubayah et al. 2010), but time series of images must be extremely well matched in terms of geolocation and data quality to measure the subtle change in height or crown size associated with tree growth (Yu et al. 2006). Alternatively, airborne visible-to-shortwave infrared (VSWIR) and light detection and ranging (LiDAR) systems can measure forest canopy properties, such as tree species composition (Fagan et al. 2015) and canopy size structure (Kellner and Asner 2009), with high spatial resolution and fidelity at a single point in time.

Rather than directly measuring changes in crown size, we propose estimating tree growth (as ΔDBH), by detecting growth-related canopy properties in a single hyperspectral aerial image. Multiple canopy properties, including foliar nutrients (Sims et al. 2013, Serbin et al. 2014, Singh et al. 2015), plant stress (Pontius et al. 2008), and plant disease (Delalieux et al. 2009, Shafri and Hamdan 2009) are measurable using hyperspectral remote-sensing data and are directly related to tree growth rates (Cornelissen et al. 1997, Ollinger and Smith 2005). These measurable properties of tree crowns may provide a link between canopy spectral reflectance and ΔDBH that could enable reliable detection of tree growth rates from a single image.

As an initial test of whether remote sensing can detect tree growth differences, we use a single image to predict ΔDBH rates from pure species plots of nine-year-old trees planted to assess the potential of numerous species for reforestation projects (Park et al. 2010). Our study has direct relevance for monitoring growth in tree plantations, particularly where resources on the ground are limited and projects are large. Spatial management plans to allocate different interventions within a single plantation can lead to biodiversity gains while maintaining acceptable profit margins (Hartley 2002). Allocating spatially targeted interventions will require measuring tree health and performance across the entire plantation, a task that could be aided by remote-sensing data (Delalieux et al. 2009, Sims et al. 2013, Vastaranta et al. 2014). Our study system—young, homogenous stands in an experimental forest—enabled us to determine whether tree growth rates can be predicted from a single aerial image across a range of species and environmental conditions. We address two questions: (1) Is there a relationship between spectral data and ΔDBH across twenty tropical tree species that vary widely in functional traits and phylogeny? (2) Which spectral regions, associated with different canopy chemical and structural properties, are important for predicting ΔDBH?

**Methods**

The ΔDBH measurements are from an experiment designed to quantify the effect of environmental variation on tree species suitability for restoration (Park et al. 2010). The experiment includes replicate plots in three completely randomized, spatially separated blocks across an elevation gradient (range 23–151 m) with different topography and edaphic conditions. Our study takes place in Los Santos Province in Southwestern Panama (7°25′14.3″ N, 80°09′49.1″ W). The study site has a pronounced dry season from December through March, with an average of 1,700 mm of rainfall mostly falling during a rainy season from late April to late November. Pure species plots were planted with 20 seedlings of 1 of 20 tree species in 2003, and thinned to 10 trees per plot in 2005. Growth for all live tree stems in plots was measured as ΔDBH between 2008 and 2010. The experimental design ensures a range of intraspecific growth rates across many species with diverse growth rates, functional traits, and taxonomy, including 13 different families (Appendix S1).

Remotely sensed data were collected in January 2012 by the Carnegie Airborne Observatory-2, using a VSWIR infrared spectrometer (380–2,510 nm) with 2-m resolution and a dual-laser waveform LiDAR scanner with 1.3-m resolution (Appendix S2; Asner et al. 2012). In December 2013 and July 2014, we mapped canopy boundaries of each plot with a handheld GPS unit and tablet displaying the aerial image (Fig. 1). In total, we mapped 87 single-species plots with plot areas ranging from 25 to 450 m². Because the individual trees in our study were relatively small, with a canopy area of 24.05 ± 23.01 m² (mean ± SD), and could not be distinguished in the images with confidence, matching pixels to individual tree canopies, rather than single species plot boundaries, was not possible.

The goal of our analysis was to predict plot-level growth as a continuous variable for all 20 species together. Absolute ΔDBH growth rates varied widely between species (Appendix S1: Table S1), from 11.11 ± 5.52 mm/yr for *Colubrina glandulosa* to 45.72 ± 22.00 mm/yr for *Erythrina fusca*. To compare fast and slow-growing plots across species, we standardized ΔDBH growth rates on individuals within each species by centering around the mean and dividing by two standard deviations (Gelman 2008). Plot-level means of the standardized growth rate were then used as the response variable in the model. After standardization, plots near the mean growth rate of each species have a value of zero, while plots with growth rates two standard deviations below and above the mean have values of −1 and 1, respectively. This standardized growth rate can be interpreted as an index of plot-level growth rate, relative to plots of the same species. Using standardized growth rate as the response variable in our analyses increases our confidence that we are detecting intraspecific variation in tree growth rate, rather than species differences across our diverse set of 20 study species. However, in Appendix S3, we demonstrate that our approach can also predict absolute growth rate on the original scale of mm/yr.
Predictor variables included narrowband hyperspectral reflectance and elevation. To match the spatial scale of the tree growth data, we aggregated elevation and hyperspectral reflectance to the plot level by taking the mean value of all pixels within each plot. Because we expected growth rates to vary with elevation (Park et al. 2010), our model also includes per-plot elevation as a predictor variable, produced using a digital elevation model (DEM) derived from the LiDAR data. We found no evidence for spatial autocorrelation in the residuals of the final model.

The high-dimensional structure of the hyperspectral data, including multicollinearity (Asner et al. 2012), presents a challenge to identifying clear predictive relationships. Narrowband indices of wavelengths represent one way to construct predictor variables that can reduce illumination differences between pixels and isolate absorption features related to canopy chemistry (Shafri and Hamdan 2009, Roberts et al. 2011). One approach to identify narrowband indices with predictive power is to evaluate the performance of all possible two-band combinations (e.g., Delalieux et al. 2009). We build upon this approach by iteratively selecting the single narrowband index, defined here as the normalized difference between two narrowbands, with the best predictive power out of all possible narrowband indices and compiling multiple narrowband indices with high predictive power using least squares regression. We used three strategies to avoid overfitting. First, we selected individual narrowband indices using the predicted residual sum of squares (PRESS) statistic, an out-of-sample validation metric that penalizes for overfitting (Chen et al. 2004). Second, we used a randomization test to determine a cut-off for the number of narrowband indices to include in the final model, based on the probability of observing a value of the PRESS statistic significantly lower than random chance. Third, we evaluate predictive power of the least squares algorithm using leave-one-out cross-validation, and present out-of-sample $R^2$ and RMSE as metrics of model fit. We
describe our least squares algorithm in more detail in Appendix S4, and compare our results to those from partial least squares regression, another approach for analyzing high-dimensional data (e.g., Serbin et al. 2014, Asner et al. 2015, Singh et al. 2015), in Appendix S3.

RESULTS

Out-of-sample $R^2$ for our final model was 53.70%, indicating that a single image can predict landscape variation in growth rates across 20 species (Fig. 2). In total, our least squares algorithm selected six narrowband indices to predict standardized growth rate (Appendix S4: Fig. S2). We found diminishing returns in predictive power as additional variables were added. Wavelengths selected for narrowband indices in our growth index include the entire range of the measured spectrum, from the visible to shortwave regions. The first band index selected, involving the near-infrared (NIR) and shortwave infrared (SWIR) regions (810 and 2,122 nm) led to the biggest increase in $R^2$ (14.21%) and biggest proportional decrease in the PRESS statistic (18.47%). However, five of these six narrowband indices had one or both bands located in the visible region (from 470 to 700 nm). Limiting the wavelengths included in the least squares algorithm to the visible region between 470 and 750 nm, a decrease in dimensionality from 150 to 31 narrow bands, only decreased $R^2$ by 5.54% (Appendix S5). Although a model with elevation alone was poor at predicting growth, with an $R^2$ of 2.71%, elevation had a significant effect in the full model (Appendix S6).

DISCUSSION

We demonstrate the potential for airborne remote-sensing data to quantify spatial variation in one metric of tree demography ($\Delta DBH$) at large spatial extents. Our methods do not depend on the acquisition of a lengthy time series of remote-sensing data or measurement of foliar traits and, if repeatable at other sites, could be immediately applied to existing and soon-to-be collected (e.g., National Ecological Observatory Network; Keller et al. 2008) hyperspectral images. While $\Delta DBH$ is a common metric of tree demography due to the ease of measuring stem diameter growth in the field, hyperspectral data could provide a wider range of metrics linked to tree physiology, such as foliar nutrients (Ollinger and Smith 2005, Axelsson et al. 2013, Sims et al. 2013, Asner et al. 2015, Singh et al. 2015) and chlorophyll fluorescence (Calderón et al. 2013). Our results extend previous studies that demonstrate that a single remotely sensed image can measure tree health and performance for one, or perhaps a few, species (Pontius et al. 2008, Delalieux et al. 2009, Shafri and Hamdan 2009), by suggesting a general correlation between narrowband indices and $\Delta DBH$ across a wide range of species with varying functional traits at a single site. Although whether tree growth can be measured remotely in mixed-age, mixed-species forests remains an open question, our results suggest that hyperspectral data can accurately predict $\Delta DBH$ in even-aged monocultures, including tropical tree plantations.

For all tree species considered together, the $R^2$ for our best-fitting model was 53.70%, with an RMSE of 2.46.
The predictive power of a model for standardized growth across species suggests that predicting whether a tree is a fast or slow grower relative to conspecifics is possible even when species identity is unknown. The wavelengths selected in our growth index include the entire range of the measured spectrum, from the visible to shortwave regions. However, repeating the analysis without the near-infrared and shortwave regions revealed that the visible and shortest wavelengths of the near-infrared regions (up to 750 nm) of the spectrum contribute the majority of predictive power. We found that LiDAR-derived elevation measurements initially contributed a majority of predictive power ($R^2$ for elevation-only model = 2.71%), but nonetheless had a strong and significant effect in the full model (Appendix S6). At our study site, topographic differences in soil quality and water availability influence tree growth rates (Park et al. 2010).

Our results suggest that elevational differences may lead to different canopy traits, such as leaf phenology or foliar nutrients that influence remote detection of tree growth rates.

The narrowband indices selected in our model suggest canopy properties that may provide a link between reflectance and growth that can form the basis for future investigations. Our study is a departure from previously used approaches to relate optical remote sensing to forest productivity that focus on the underlying mechanism of light capture and efficiency (Garbulsky et al. 2011) or foliar chemical composition (Ollinger and Smith 2005, Serbin et al. 2014, Singh et al. 2015). The advantage of our approach is that we are not constrained to specific wavelengths or measured foliar traits and can identify emergent spectral regions and traits that may be driving growth for a particular site or species. For example, the first band index selected in our species-analysis, involving the NIR and SWIR regions (810 divided by 2,122 nm), is indicative of water content, a key canopy trait affecting both growth and reflectance (Roberts et al. 2011). Similar narrowband indices related to water absorption have been correlated with vegetation water content (Chen et al. 2005), and we suggest that leaf water content is a key variable for forest dynamics in this ecosystem that can be measured with remote sensing. Another important spectral region for predicting tree growth is between 470 and 750 nm, including five of the six narrowband indices selected. These wavelengths measure pigments that play a critical role in carbon uptake and photoprotection in tropical trees (Contin et al. 2014), including chlorophyll, carotenoids, and anthocyanins (Roberts et al. 2011). The variation in pigments is likely related to leaf senescence during the transition from the wet to dry season (when the data was collected), and may be indicative of different phenologic strategies of growth across species. If images were collected at peak greenness, which is the case for many data sources, wavelengths and traits may be more likely to reflect differences in light capture and light use efficiency. The identified wavelengths point to causal explanations for the observed correlation between tree growth and remote-sensing data; hypotheses that can be tested by pairing direct measurements of foliar chemistry and other canopy properties with spectral data.

The applicability of using a single hyperspectral image to discriminate growth differences needs to be tested for other sites and forest types. Given the underlying relationship between growth and canopy chemistry (Cornelissen et al. 1997, Ollinger and Smith 2005), and the growing body of evidence that canopy chemistry can be quantified from hyperspectral data (Sims et al. 2013, Singh et al. 2015), we believe this approach has strong potential to detect growth differences between even-aged plots at other sites. Just as the strongest wavelengths for determining leaf chemistry may vary across sites (Casas et al. 2014), the exact wavelength bands that are strongly related to growth will likely vary across sites. Extending our methods to predict ΔDBH for trees in a natural, mixed-age, mixed-species forest will present several challenges. First, predicting tree growth rates in those forests may require resolving individual tree canopies with advanced segmentation techniques. Second, our results suggest that we are capturing canopy spectral properties that are important for tree growth rates of small plots of ~10 trees. Between these even-aged plots, variation in tree growth is likely due to site factors (e.g., soil moisture). In natural forests, individual variation in tree growth is higher and related to competition for light between different-sized trees, although variations in crown size and height can be quantified by LiDAR and used as an additional predictive variable (Kellner and Asner 2014).

The ability to quantify the state of a tree in fine detail over large areas could be transformational both for our basic understanding of forest dynamics and for many applied forest management projects. One application of remote-sensing detection of tree growth rates is precision forestry, i.e., silvicultural treatments at landscape scales tailored to observed growth variation. Previous remote-sensing-based precision forestry has mostly focused on assessing differences in standing stocks, typically measured with LiDAR and radar data (Kellndorfer et al. 2003). We propose that hyperspectral imagery could detect spatial differences in growth rate before differences in woody biomass are measurable. Predictive algorithms based on aerial imagery could then enable fine-scale interventions, for example, fertilizer treatments for plots with low growth rates due to nutrient deficiencies detected in leaf reflectance (Sims et al. 2013). These kinds of early detection systems are well known in agricultural systems, where remote sensing of crop performance, followed by spatially targeted interventions, can improve crop production while reducing external inputs (Gebbers and Adamchuk 2010). We show here that similar types of performance information for trees can be derived from hyperspectral data, despite the greater structural complexity and species diversity of many forests. Our results point to a future where a single flyover by an aerial observatory could provide data to
inform spatial interventions and contribute to sustainable decision making in forested ecosystems.

ACKNOWLEDGMENTS

We thank the Carnegie Airborne Observatory and PRORENA staff for data collection and processing support. Funding for PRORENA is from the Frank Levinson Family Foundation and the School of Forestry and Environmental Studies, Yale University. T. T. Caughlin was supported by NSF grant #1415297. Airborne data collection, processing and analysis was funded by the Grantham Foundation for the Protection of the Environment and William R. Hearst III. The Carnegie Airborne Observatory has been made possible by grants and donations to G.P. Asner from the Avatar Alliance Foundation, Margaret A. Cargill Foundation, David and Lucile Packard Foundation, Gordon and Betty Moore Foundation, Grantham Foundation for the Protection of the Environment, W. M. Keck Foundation, John D and Catherine T. MacArthur Foundation, Andrew Mellon Foundation, Mary Anne Nyburg Baker and John D. and Catherine T. MacArthur Foundation, and William R. Hearst III.

LITERATURE CITED


Caughlin, T. T., S. Elliott, and J. W. Lichstein. 2016. When does seed limitation matter for scaling up reforestation from patches to landscapes? Ecological Applications, in press. DOI: 10.1002/1061-9204.a1410


analysis using imaging spectroscopy, modeling and field data. Ecosystems 8:760–778.


**Supporting Information**

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/eap.1436/suppinfo

**Data Availability**

Tree growth data are available online at http://dx.doi.org/10.5061/dryad.t6md2. R code for the least squares algorithm is available online at http://dx.doi.org/10.5281/zenodo.61972.