Materials and Methods

Remote Sensing System

Large-scale analysis of forest three-dimensional (3D) structure and biological composition requires a combination of advanced airborne imaging technologies that simultaneously resolve the horizontal and vertical characteristics of the vegetation as well as the type of vegetation, even to the species level. No airborne (or space-based) technologies were available to make these simultaneous measurements until we developed the Carnegie Airborne Observatory (CAO), a new system designed specifically for mapping the biochemical, taxonomic, and structural properties of vegetation and ecosystems (http://cao.stanford.edu) (1). The CAO combines three major instrument subsystems into a single airborne package: (i) high-fidelity Imaging spectrometer (HiFIS); (ii) waveform light detection and ranging (LiDAR) scanner; and (iii) global positioning system-inertial measurement Unit (GPS-IMU).

The CAO HiFIS subsystem provides spectroscopic images of the land surface; each image pixel contains numerous spectral bands covering either the 367- to 1,058-nm or 380- to 2,510-nm range, depending upon system configuration. The CAO-Alpha configuration uses a pushbroom imaging array with 1,500 cross-track pixels, and sampling across the 367- to 1,058-nm range at up to 2.4-nm spectral resolution. The CAO-Beta system integrates the airborne visible and infrared imaging spectrometer (AVIRIS) to provide 10-nm spectral sampling across the 380- to 2,510-nm range (2, 3). Both CAO configurations use a LiDAR subsystem with an adjustable laser pulse repetition rate of up to 100 kHz (1). The GPS-IMU subsystem provides three-dimensional positioning and attitude data for the sensor package onboard the aircraft, allowing for highly precise and accurate projection of HiFIS and LiDAR observations on the ground (1).

The CAO-Alpha and -Beta configurations provide coaligned HiFIS and LiDAR data at spatial resolutions of 0.3- to 3.5-m spatial resolution, depending on the aircraft altitude above ground. The Alpha system is designed for spatial sampling at extremely high resolutions of 0.3-1.5 m. The Beta system, which incorporates the AVIRIS sensor, operates in the 2.5-3.5 m spatial sampling range. The LiDAR has a custom-designed laser beam divergence to match precisely the field-of-view of the CAO-Alpha spectrometer. When flown in the CAO-Beta mode with AVIRIS, the LiDAR collects four laser shots per spectrometer pixel.

Airborne Data Collection

In January and February 2007, we operated the CAO-Alpha and -Beta systems over 51,707 ha and 170,168 ha, respectively, of Hawaii Island (SI Fig. 4). The CAO-Alpha data were collected at 1,000 m and 2,000 m a.g.l., providing combined HiFIS and LiDAR measurements at 0.5-m and 1.0-m spatial resolution, respectively. The CAO-Beta data were collected at ~2500 m a.g.l., resulting in measurements taken at ~2.5-m spatial resolution. All flights were conducted within 2.5 h of solar noon. Data from both system configurations were collected over federal and state forest reserves and parks, including Hawaii Volcanoes National Park, Laupahoehoe Forest Reserve, Wao Kele O

Puna Reserve, Kohala Forest Reserve, and the Lower Puna reserves of Keauohana, Malama Ki, and Nanawale (SI Fig. 4).

Data Processing and Analysis

Fusion of the imaging spectroscopy and LiDAR data requires a processing stream that maximizes the sharing of information between data products. Given the enormous data volumes involved, the processing stream must be highly automated. SI Fig. 5 shows the processing stream for this study, in which raw spectral, laser, and trajectory data are integrated and analyzed in a series of higher-order products and results. The following sections briefly describe the major steps in the process.

Aircraft Positioning

The CAO uses both in-flight and post-flight data integration approaches to precisely match HiFIS and LiDAR data in three-dimensional space. The in-flight step is achieved by providing a common mount with measured offsets between instrument optical centers, as well as time stamping of spectral and LiDAR data collection streams with shared position and trajectory data. The GPS-IMU data form the common link for the detailed ray tracing of the photons between aircraft sensors and the ground. The point-for-point alignment of the LiDAR and passive image data is complicated by inherent differences in the scanning geometries of the two systems and further distortions of the ground sampling grid due to topography. Our approach is to recover best estimates for each pixel center location in three dimensions for both the LiDAR and HiFIS data (1). These pixel centers are then used for rendering of the two data sets into a single, integrated grid of HiFIS and LiDAR data for subsequent processing, analysis and product generation (SI Fig. 6).

LiDAR Data Processing

The GPS-IMU data are combined with the laser ranging data to determine the 3-D location of the laser returns. From the laser "point cloud" data, a physical model is used to estimate top-of-canopy and ground surfaces (digital elevation models; DEMs) using the REALM (Optech Inc., Vaughn, Canada) and Terrascan/Terramatch (Terrasolid Ltd., Jyväskylä, Finland) software packages. Vegetation height is then estimated by differencing the top-of-canopy and ground surface DEMs (4, 5). Vertical errors in ground heights and tree heights average 0.12 ± 0.14 m and 0.7 ± 0.2 m, respectively (1).

The large number of LiDAR points collected at high pulse rates makes it desirable to represent the data in a compact form for analysis of the vertical structure of vegetation. This data reduction is typically done to render pseudo-waveform data (6). In this case, the vertical distribution of LiDAR points is represented by binning them into volumetric pixels (voxels) of 5×5 m spatial resolution, and 1 m vertical resolution. The DEM of ground elevation was used to standardize the vertical datum of each voxel. Therefore, the heights of each vertical "slice" of a vegetation canopy are defined relative to the ground at the horizontal center of each voxel. After all LiDAR points are binned in the volume cube, each vertical column of the cube is divided by the total number of LiDAR points in

that column, yielding the percentage of LiDAR points that occurred in each voxel. This approach has the positive effect of decreasing our sensitivity to localized variations in canopy leaf density or tree branch characteristics, which can result in a different number of LiDAR returns from voxel to voxel.

HiFIS Data Processing

The HiFIS data are converted to at-sensor radiances by applying radiometric corrections developed during sensor calibration in the laboratory. Apparent surface reflectance is then derived from the radiance data using an automated atmospheric correction model, ACORN 5LiBatch (Imspec LLC). Inputs to the atmospheric correction algorithm include ground elevation (from the LiDAR), aircraft altitude (from GPS-IMU), solar and viewing geometry, atmosphere type (e.g., tropical), and estimated visibility (in km). The code uses a MODTRAN look-up table to correct for Rayleigh scattering and aerosols. Water vapor is estimated directly from the 940/1140 nm water vapor features in the radiance data.

After the preparation of the HiFIS and LiDAR data, the spectral images are masked based on illumination conditions between the sensors and canopies (SI Fig. 5). The LiDAR and GPS-IMU data provide three-dimensional maps of precise illumination conditions on each canopy, allowing for the automatic identification and masking of shaded portions of the vegetation. At the same time, a minimum vegetation height requirement of 0.5 m is applied to remove exposed ground areas and short vegetation such as grasses.

The masked HiFIS images are passed to an automated spectral mixture analysis model called *AutoMCU* (7). This algorithm uses spectroscopic signatures to quantify the fractional cover of photosynthetic vegetation (PV), nonphotosynthetic vegetation (NPV), and bare substrate within each image pixel, using "tied" spectral endmember bundles derived from field- or image-based measurements, and Monte Carlo unmixing to derive mean estimates of fractional cover along with standard deviation and root mean squared error (RMSE) data on a per-pixel basis. The PV, NPV, and bare/shade spectral bundles were derived from island-wide samples collected using 2001-2005 AVIRIS imagery as well as field-based measurements with spectroradiometers (8). For our purposes here, only the NPV data derived from the *AutoMCU* are used to mask out the standing dead trees and other nonphotosynthetic vegetation from the HiFIS data (SI Fig. 5).

The image spectra that remain after illumination, shadow, vegetation height, and NPV masking represent a well controlled set of reflectance signatures that, theoretically, should be highly indicative of unique species. The final automated step in the processing stream involves a new reformulated version of the *AutoMCU* algorithm to detect species using species-specific spectral bundles, hereafter called *AutoMCU-S* (SI Fig. 5). Whereas the previous step, running the *AutoMCU* in tied-spectral mode, isolates the fractional cover of live and dead vegetation with little sensitivity to taxonomic composition (7), *AutoMCU-S* uses the same Monte Carlo unmixing technique but adds species-specific endmember bundles to derive maps of species cover fractions per pixel, along with standard deviation and RMSE images.

In previous work, we collected image-based spectral bundles of 43 of the most common native and invasive tree species found in Hawaiian rain forests (9). Here, we

apply the *AutoMCU-S* approach with a subset of species bundles relevant to the forested regions of interest. These were selected based on our general knowledge of the likely tree taxa to be present in each study forest. For example, a spectral bundle for *Fraxinus uhdei* was only used in the forests where this species is known to occur. This geographically constrained use of the bundles simplified our detection and mapping problem, maximizing the accuracy of the approach at the landscape scale. Since the general presence of both native and invasive species is known for most forests on Hawaii Island (www.hear.org) (10), tuning the *AutoMCU-S* with regionally-specific libraries was justified from both scientific and land management perspectives.

For this study, we used endmember bundles containing one native tree species collected from all five sites (*Metrosideros polymorpha*), one native tree species from two sites (*Acacia koa*), several other less common native woody plants, and four invasive tree species (*Morella faya*, *F. uhdei*, *Falcataria moluccana*, *Psidium cattleianum*). The libraries were constructed from spectra prescreened for minimum height and illumination conditions using the coaligned LiDAR data and thus were compatible with the image data. Minimum height was set to 0.5 m, and only pixels containing canopies with unobstructed ray traces from the sun-to-canopy and canopy-to-sensor were used in the library development and image analyses. In addition, only pixels with no detected NPV from the *AutoMCU* code were selected both for library construction and image processing.

For the special case of the understory plant, *Hedychium gardnerianum*, we used a different technique to ensure accurate mapping. We combined the canopy water measurement provided by the spectroscopic analysis, as detailed by Asner and Vitousek (11), with the LiDAR near-ground laser returns as shown experimentally in SI Fig. 7, to locate large infestations of this species. The ground validation (described later) showed a detection error of less than 5% for *H. gardnerianum*, thus affording a method to map the effects of this species on forest structure.

Forest Stand Selection and Structural Analysis

Our process of searching for stands dominated by invasive tree species covered 221,875 ha of forest spanning the study zones shown in SI Fig. 4. We found thousands of vegetation patches dominated by each of the target invasive species, many of which were small infestations (SI Figs. 8 and 9). These image data were then further reduced to select for the more accessible field-verifiable forests stands. The final set of comparative study areas ranged in size from 49 to 97 ha (Table 1 of main text). To normalize for differences in the area of each delineated forest stand, the larger stands were randomly subsampled to match the smaller of the two forest stands in each pairwise comparison. We also broadly controlled for substrate age and elevation (SI Table 4). We focused on protected reserves to minimize confounding effects of land use (e.g., no logging, clearing, gathering allowed in the state and federal reserves in Hawaii).

To quantify differences in forest volumes among comparative stands, we calculated changes in LiDAR point densities among canopy strata (e.g., lower-, mid-, and upper-canopy positions). The probability of laser penetration to the forest floor is a function of flying altitude, beam divergence, laser pulse repetition frequency (PRF), vegetation density, and other factors. In comparing forest stands, we controlled for flying

altitude and PRF. CAO-Alpha measurements were collected at a PRF of 50-100 kHz. CAO-Beta measurements were collected from a higher altitude to accommodate AVIRIS, which required a PRF limited to 33 kHz. In both cases, laser beam divergence was held constant at 0.56 mrad. We held all of these factors constant when comparing specific forest stands dominated by native or invasive species.

Vegetation density varied between forest stands as well, so our approach relied on knowing that some portion of the total laser shots reached the ground, as shown in Fig. 2 of the main text. This was readily achieved in most cases, but in a few invasive stands of very high leaf area index (LAI), the laser signal was almost completely attenuated. In these cases, our field surveys indicated that little to no understory vegetation was present under the dense canopies, thus minimizing our bias caused by lost laser signal. In other words, when overstory or midcanopy vegetation density was so high as to attenuate the signal, we found no underlying vegetation that would have been missed in our airborne measurements. As such, the laser return (or lack thereof) was an excellent indicator of vegetation losses at the lowest canopy positions.

We report relative changes in the vertical canopy profiles in Fig. 3 of the main text. This was a useful approach because it allowed for a clear comparison of changes over a diverse array of canopy structures and species. However, at times, it slightly biased the results because, for example, a 50% increase in vegetation density in one canopy layer would necessarily, from a sensor standpoint, be balanced by a 50% loss from other layers. In most cases, this turned out to be a small artifact that we considered reasonable in looking for relative changes among many canopies. In a few cases, such as with the *P. cattleianum* invasion in lowland rainforests (Fig. 2e of main text), a very large increase in the midcanopy volume of this species caused an apparent decrease in upper-canopy *M. polymorpha* density that was, technically, caused by the fact that we are using relative changes in total laser returns. However, in this case, the absolute change in upper-canopy laser point returns was nonetheless 56% (also shown visually in SI Fig. 10e), so the artifact caused by using relative change as our index remained acceptable. In subsets of the mapping results, however, the bias was more pronounced, but the overall impact of using relative changes in forest profiles remained comparatively small and tolerable.

Remote Sensing Validation

We carried out field studies to evaluate the accuracy of our remote species detections in each forest stand. The analysis employed a combination of intensive plot-scale measurements, long field-based transect surveys, and low-altitude helicopter surveys that identified false-positive and false-negative detections. Transects ranged from 200 m to 3,000 m in length, with a total distance covered among sites of 16.3 km. Transect sampling was done according to Asner et al. (9) and Asner and Vitousek (11), by which a point was recorded whenever the canopy cover changed, in this case, when the overstory changed in species dominance. Each change in species cover was recorded using a survey-grade GPS for colocation of field data with the airborne measurements. A Leica GS-50+ GPS with multiple-bounce filtering and postdifferential correction was used to measure our position in the forest to average uncertainties of ~2 m (Leica Geosystems Inc., St. Gallen, Switzerland). In addition, a ruggedized tablet computer with integrated GPS (Xplore Technologies) was used to navigate in the forests,

providing a real-time analysis and logging of our position with respect to the species identified in each digital map. Helicopter-based validation measurements were carried out by entering the coordinates of detected invasive species from the digital maps into a GPS with real-time wide-area augmentation system (WAAS) corrections (Thales Navigation). Each helicopter GPS point was visited from an altitude of <75 m a.g.l., with two or three observers determining canopy species cover.

We tabulated invasive species detections at the *AutoMCU-S* threshold of \geq 75% cover fraction and compared those results with the field- and helicopter-based validation points (SI Table 2). In montane forests, we visited 164 points in the field and calculated a 0% and 4.2% error rate for false-negative and false-positive detections, respectively, for *F. uhdei*. *M. faya* was detected in submontane forests with false-negative errors of 3.0% and false-positive errors of 4.9% (n = 256). We recorded zero errors for *F. moluccana* in lowland forests (n = 64). Finally, our uncertainty in *P. cattleianum* detections in lowland and submontane forests peaked at 6.8% for false-negative detections and 0% for false positives (n = 141). Asner et al. (9) showed that the consistently low error rates for these canopy species resulted from the ability to isolate sunlit portions of each tree crown using the LiDAR data, allowing for focused spectral analysis without the confounding effects of intercrown shadows.

Understory species detections were also tabulated and reported in SI Table 3. A chi-square test showed a highly significant association between *H. gardnerianum* locations and the CAO detection ($\chi^2 = 19.0 > 3.8$; p < 0.001). The false-negative and false-positive errors averaged 5.6% and 4.1%, respectively (n = 602).

Canopy Measurements

We collected LAI measurements using a plant canopy analyzer (LAI-2000; Licor Inc.). The LAI estimates were made under diffuse sky conditions as required by the instrument data processing algorithms (12). A 50% optical block was used to mask the operator. An open-sky measurement was collected followed by 12 under-canopy measurements. The LAI-2000 sensor head was oriented in the direction of the main stem from a position 30-50 cm within the edge (drip line) of the tree crown. These measurements also included mean leaf angle, which, when combined with LAI, provided a means to estimate the fractional absorption of photosynthetically active radiation (fAPAR) by the canopy using a radiative transfer model (13). The measurements were collected in each large invasive species stand detected. Our goal here was to characterize typical LAI and light levels beneath the canopies following invasion, so only mature stands (Figs. 1, 2, and SI Figs. 8-10) were visited. In all cases, we found significantly higher LAI values in invasive canopies and concomitantly higher fAPAR values (SI Table 4). In the lowland stands containing F. moluccana, areas also containing a midcanopy infestation of P. cattleianum had significantly higher total LAI and fAPAR than those without the secondary invasion (14). Without the midcanopy, F. *moluccana* alone transmits 21 + 6% of the incoming PAR, which is essential to the P. cattleianum midcanopy trees (see SI Fig. 10d).

References

- 1. Asner GP, Knapp DE, Kennedy-Bowdoin T, Jones MO, Martin RE, Boardman J, Field CB (2007) *J Applied Remote Sensing* **1**, 10.1117/1.2794018.
- 2. Green RO, Eastwood ML, Sarture CM (2005) in *NASA Airborne Earth Science Workshop*, ed Green RO (Jet Propulsion Laboratory, Pasadena, CA).
- 3. Green RO, Eastwood ML, Sarture CM, Chrien TG, Aronsson M, Chippendale BJ, Faust JA, Pavri BE, Chovit CJ, Solis MS, et al. (1998) *Remote Sensing Environment* **65**:227-248.
- 4. Lefsky MA, Cohen WB, Acker SA, Parker GG, Spies TA, Harding D (1999) *Remote Sensing Environment* **70:**339-361.
- 5. Lefsky MA, Cohen WB, Parker GG, Harding DJ (2002) *BioScience* **52:**19-30.
- 6. Weishampel JF, Blair JB, Knox RG, Dubayah R, Clark DB (2000) *Int J Remote Sensing* **21:**409-415.
- 7. Asner GP, Heidebrecht KB (2002) Int J Remote Sensing 23:3939-3958.
- 8. Asner GP, Elmore AJ, Hughes FR, Warner AS, Vitousek PM (2005) *Remote Sensing Environment* **96**:497-508.
- 9. Asner GP, Jones MO, Martin ME, Knapp DE, Hughes FR *Remote Sensing Environment*, in press (currently available until published at http://cao.stanford.edu/publications].
- 10 Smith CW (1985) in *Hawaii's Terrestrial Ecosystems: Preservation and Management*, eds Stone CP, Scott JM (University of Hawaii, Honolulu: Cooperative National Park Resources Study Unit), pp 180-250.
- Asner GP, Vitousek PM (2005) *Proc Natl Acad Sci USA* 102:4383-4386.
- 12. Welles JM, Norman JM (1991) Agron J 83:818-825.
- 13. Asner GP, Wessman CA (1997) *Ecol Model* **103:**81-97.
- 14. Hughes FR, Denslow JS (2005) *Ecol Appl* **15**:1615-1628.

Figures

Fig. 4. Island of Hawaii with forested regions shown in green. The airborne survey areas are delineated in red boxes showing actual flight coverage. Area 1, Kohala Mountains; area 2, Laupahoehoe Forest Reserve; area 3, Lower Puna forest reserves; area 4, Wao Kele O Puna Reserve; area 5, Hawaii Volcanoes National Park.

Fig. 5. The processing stream for in-flight and post-flight integration of airborne imaging spectrometer and LiDAR observations, sunlit-live canopy detection, and invasive species mapping.

Fig. 6 The GPS-IMU system onboard the Carnegie Airborne Observatory provides (A) high-precision and accuracy 3D motion analysis of each flight (red line shows actual motion of sensors at 2000 m a.g.l. over the Laupahoehoe Forest Reserve below). These GPS-IMU data are then used to fuse (B) the hyperspectral data to the (C) LiDAR data, providing (D) 3D imaging of forest canopies.

Fig. 7. (a) Upper-canopy mapping (simple color-infrared composite) from CAO hyperspectral sub-system; (b) understory plant mapping (white shows dense vegetation 0.5-2.0 m above ground level) from CAO LiDAR sub-system. This test and validation area was once infested by *Hedychium gardnerianum* (kahili ginger) until it was manually cleared prior to overflight.

Fig. 8. (a) Mapping results of constrained AutoMCU-S analysis for the detection of invasive *M. faya* trees (reds) and native *M. polymorpha* trees (greens) in Hawaii Volcanoes National Park. (b) Invasive *F. uhdei* (orange), native *A. koa* (yellow-green) and native *M. polymorpha* (blue-purple) in Laupahoehoe Forest Reserve. Extensive field validation provided by (9).

Fig. 9. (a) Mapping results of constrained AutoMCU-S analysis for the detection of invasive *P.cattleianum* trees (reds), native *A.koa* (green) and native *M. polymorpha* (blue) in Olaa Forest Reserve. (b) Invasive *P. cattleianum* trees (reds) and native *M. polymorpha* (green) in Wao Kele O Puna Reserve. Extensive field validation provided by (9).

Fig. 10. Photographs of typical infestations of: (a) *F. uhdei*, (b) *M. faya*, (c) *H. gardnerianum*, (d) *F. moluccana* (upper canopy) and *P. cattleianum* (midcanopy), and (e) sparse *M. polymorpha* protruding through a densely populated, 8- to 10-m tall *P. cattleianum* canopy. The last photograph was taken at low altitude from a helicopter.



Fig. 4.



Fig. 10.

Montane forest	<i>F. udhei</i> (n = 164) -	Observed	Observed	Percentage Error
		No	Yes	
Detected	No	93	0	0.0%
Detected	Yes	3	68	4.2%
Sub- montane forest	<i>M. faya</i> (n = 256)	Observed	Observed	
		No	Yes	
Detected	No	130	4	3.0%
Detected	Yes	6	116	4.9%
Lowland forest	<i>F</i> . <i>moluccana</i> (n = 64)	Observed	Observed	
		No	Yes	
Detected	No	34	0	0.0%
Detected	Yes	0	30	0.0%
Lowland	P. <i>cattleianum</i> $(n = 141)$	Observed	Observed	_
		No	Yes	
Detected	No	109	8	6.8%
Detected	Yes	0	23	0.0%

Table 2. Validation results for invasive species detection at the upper-canopy level

		Understory	Understory
	Observed*	No	Yes
Detected	No	283	17
Detected	Yes	17	319

Table 3. Chi-square (χ^2) test of the validation results for understory plant detection

		Understory	Understory
	Expected	No	Yes
Detected	No	277.7	12.3
Detected	Yes	12.3	313.7

*Number of image points that met the following criteria: (1) absolute error in GPS point location was < 5 m; (2) understory was declared present when plants covered a 5 x 5 m area or greater surrounding the GPS point; (3) understory plants were 0.5-2.0 m above ground level.

Table 4. Leaf area index (LAI), mean leaf angle, and upper-canopy transmission of photosynthetically active radiation (PAR) among native and invasive tree species found in each forest. Notes after forest type indicate substrate age (s), mean elevation (e), and number of samples collected (n).

Forest Stand	Area (ha)	LAI	Canopy PAR Transmission		
Montane Forest (s = 65,000 y, e = 1600 m, n = 183)					
• F. uhdei	60.2	6.7 <u>+</u> 1.4 ^a	0.04 ± 0.01^{a}		
• Native	60.7	4.4 <u>+</u> 0.6 ^b	0.13 <u>+</u> 0.02 ^b		
Sub-montane Forest (s = 200-250 y, e = 1150 m, n = 219)					
• M. faya	67.2	8.1 <u>+</u> 1.3 ^a	0.02 ± 0.01^{a}		
• Native	64.3	5.3 <u>+</u> 0.7 ^b	0.09 ± 0.02^{b}		
Sub-montane Forest (s = 250-350 y, e = 1200 m, n = 231)					
• H. gardnerianum	61.1	8.1 <u>+</u> 1.3 ^a	0.07 ± 0.01^{a}		
• Native	63.8	5.3 <u>+</u> 0.7 ^b	0.09 ± 0.02^{a}		
Lowland Forest (s = 200-300 y, e = 75-125 m, n = 193)					
• F. moluccana		2.1 <u>+</u> 1.3 ^a	0.21 ± 0.06^{a}		
• F. moluccana + P. cattleianum	51.9	7.3 <u>+</u> 0.7 ^b	0.03 ± 0.03^{b}		
• Native	48.8	3.1 <u>+</u> 0.8 ^c	0.19 <u>+</u> 0.11 ^a		
Lowland Forest (s = 750-1000 y, e = 450m, n = 74) *					
• P. cattleianum	79.7	5.9 <u>+</u> 0.4ª	0.05 <u>+</u> 0.01 ^a		
• Native	80.8	4.1 <u>+</u> 0.4 ^b	0.12 ± 0.02^{b}		

* Measurements taken from the most accessible forest stands neighboring the study areas.