# Near-infrared spectrometry allows fast and extensive predictions of functional traits from dry leaves and branches

Flávia R. C. Costa [iD],[1,5] Carla Lang,[1] Danilo R. A. Almeida,[2] Carolina V. Castilho,[3] and Lourens Poorter[4]

[1]*Coordenação de Pesquisa em Biodiversidade, Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, Brazil*
[2]*USP/ESALQ, University of São Paulo, Av. Pádua Dias, 11, 13418-900 Piracicaba, SP Brazil*
[3]*EMBRAPA, Boa Vista, Roraima Brazil*
[4]*Forest Ecology and Forest Management Group, Wageningen University, P.O. Box 47, 6700 AA Wageningen, The Netherlands*

*Abstract.* The linking of individual functional traits to ecosystem processes is the basis for making generalizations in ecology, but the measurement of individual values is laborious and time consuming, preventing large-scale trait mapping. Also, in hyper-diverse systems, errors occur because identification is difficult, and species level values ignore intra-specific variation. To allow extensive trait mapping at the individual level, we evaluated the potential of Fourrier-Transformed Near Infra-Red Spectrometry (FT-NIR) to adequately describe 14 traits that are key for plant carbon, water, and nutrient balance. FT-NIR absorption spectra (1,000–2,500 nm) were obtained from dry leaves and branches of 1,324 trees of 432 species from a hyper-diverse Amazonian forest. FT-NIR spectra were related to measured traits for the same plants using partial least squares regressions. A further 80 plants were collected from a different site to evaluate model applicability across sites. Relative prediction error ($RMSE_{rel}$) was calculated as the percentage of the trait value range represented by the final model RMSE. The key traits used in most functional trait studies; specific leaf area, leaf dry matter content, wood density and wood dry matter content can be well predicted by the model ($R^2 = 0.69$–$0.78$, $RMSE_{rel} = 9$–$11\%$), while leaf density, xylem proportion, bark density and bark dry matter content can be moderately well predicted ($R^2 = 0.53$–$0.61$, $RMSE_{rel} = 14$–$17\%$). Community-weighted means of all traits were well estimated with NIR, as did the shape of the frequency distribution of the community values for the above key traits. The model developed at the core site provided good estimations of the key traits of a different site. An evaluation of the sampling effort indicated that 400 or less individuals may be sufficient for establishing a good local model. We conclude that FT-NIR is an easy, fast and cheap method for the large-scale estimation of individual plant traits that was previously impossible. The ability to use dry intact leaves and branches unlocks the potential for using herbarium material to estimate functional traits; thus advancing our knowledge of community and ecosystem functioning from local to global scales.

*Key words: Amazon; individual level traits; leaf spectra; tropical forests; up-scaling; wood spectra.*

## Introduction

Functional traits are defined as any morphological, physiological, or phenological characteristic that affect organism fitness (Violle et al. 2007). Leaf and wood traits are being widely used to understand plant performance (e.g., Poorter et al. 2008), species distribution ranges (Stahl et al. 2014), community structure and dynamics (e.g., Paine et al. 2011), ecosystem functioning and services (e.g., Garnier and Navas 2012) and to model vegetation changes in Dynamic Global Vegetation Models (e.g., Sakschewski et al. 2015). Measuring traits however, is a laborious process, and few studies (Paine et al. 2011) have collected trait data for all individuals or even all species at a site. Traits measured at the individual level represent realized values and should be closer to the performance of the same individual. Although traits are conceptually defined as the characteristics of individuals, most studies use average trait values for species obtained from a small number of individuals per species. According to trait protocols these traits should be collected from

individuals under "optimal" growth conditions, so that the maximum trait expression and growth potential of the species is obtained (Perez-Harguindeguy et al. 2013). Although this strategy is reasonable for comparing species, it ignores the potential of individuals to acclimate. Violle et al. (2012) recently suggested including trait variance when studying plant communities, and that we should measure individual-based traits. The implication of this is that we are faced with the burden of obtaining thousands of tedious or expensive measurements in order to advance the field of functional ecology. Finding proxies for these traits is a possible solution if the proxies are easy, fast, and accurate enough.

In hyper-diverse regions, even the collection of species is difficult because identification to species level is laborious, time consuming, and prone to a large amount of errors. For example, in Amazonia, there are an estimated 16,000 species (ter Steege et al. 2013) and a single site may contain more than a thousand species (Ribeiro et al. 1999). The estimated error rate in identifications by para-taxonomists and other non-experts is around 40% (Hopkins 2007). Species delimitation, the first step in the identification process, varies widely from person to person, and can lead to just 52% similarity between the delimitations provided by different para-taxonomists (Gomes et al. 2013). Estimating species-level

trait values can therefore be problematic without the support of an army of good botanists.

Spectrometry, the recording of light properties after irradiation of an object or substance, allows trait data to be inferred, since the reflectance, transmittance, and absorbance of light depend on the size, density, and shape of plant tissues and the content of chemical components. Spectrometry has already been used successfully to determine leaf chemical composition (Asner et al. 2014, Serbin et al. 2014, Ramirez et al. 2015), physiological (Doughty et al. 2011) and phenological traits (Wu et al. 2017), and wood mechanical properties (reviewed in Tsuchikawa and Kobori 2015). Spectrometry is fast and the cost of purchasing the equipment quickly pays off, and so far, it has been very accurate for determining chemical traits for the conditions studied (Asner et al. 2014, Serbin et al. 2014). Spectrometers can be carried on satellites or aircraft, allowing a wide range of plant traits to be measured from canopies (Asner et al. 2017). FT-NIR spectrometry is also highly effective for identifying plant species (Krajšek et al. 2008, Durgante et al. 2013) even in different ontogenetic stages (Lang et al. 2015). We therefore reasoned that this method could be widely applied for determining the leaf and wood properties of species, using easy protocols that would allow measurements to be carried at the individual level. This would liberate researchers from the need to identify the plants for the calculation of species-level trait values since the spectral profiles can be used to differentiate between species (Durgante et al. 2013, Lang et al. 2015) thereby maximizing work cost benefits.

Spectrometers use two methods: (1) the dispersive method, where the light source shines on a diffraction grating or prism and then on the sample (VNIR spectroscopy) or (2) the interferometric method, where the light source shines on the input of an interferometer and then on the sample, the recorded interference pattern is transformed back into a spectral signal with a Fourier transformation (FT-NIR spectroscopy). FT-NIR has a high throughput, because it eliminates the entrance or exit slits that limit the amount of radiation reaching the detector. The advantages of FT-NIR are the higher resolution ($<0.1$ cm$^{-1}$) and the much higher (about one order of magnitude) signal to-noise ratio and scan speed (Skoog et al. 2007). On the other hand, the advantage of the dispersive method is that instruments can be portable while the need for tight control of the mirror movements in interferometry requires bench-fixed instruments. Most applications in ecosystem ecology so far have being using classical spectrometry (Asner et al. 2014, Serbin et al. 2014, Wu et al. 2017), while industry and food research use predominantly FT-NIR (e.g., Nicolaï et al. 2007), due to the need for high precision and accuracy in the evaluation of very heterogeneous materials. However, comparisons of NIR and FT-NIR did not find large differences in results (Armstrong et al. 2006, Igne et al. 2010).

Potential limitations in some previous spectrometry applications were that measurements had to be carried out on fresh leaves soon after collection (e.g., Doughty et al. 2011, Wu et al. 2017), or that dry leaves needed to be prepared as powder (e.g., Serbin et al. 2014, Ramirez et al. 2015). These requirements limit the extensive use of this technique in hyper-diverse regions or under difficult field conditions, given the higher time and cost of preparing samples as powder and the need for energy sources to power equipment in the field. Also, the only morpho-structural trait that has been evaluated with extensive data sets so far was LMA (reviewed by Serbin et al. 2014). Here we examine the ability of FT-NIR to estimate a series of 14 functional traits using dry and intact leaves and branches. We selected traits related to the global leaf and wood economics spectrum (Wright et al. 2004, Chave et al. 2009), important for the carbon, water, and nutrient balances of the plant. We also collected an independent data set to evaluate equations derived from FT-NIR for their ability to estimate traits for a different site. If the spectra collected from dry materials are able to estimate plant traits, this opens up a unique opportunity to obtain a large amount of trait data just by scanning the already collected plant material currently stored in herbaria.

Many trait-based studies use a community-wide mean trait value, the variance of trait values or the community trait frequency distribution to test for community assembly or to scale community properties up to the ecosystem level, so we also tested how well FT-NIR derived trait values can approximate these community metrics to see if this could provide an easy shortcut to these types of studies.

We asked the following questions: (1) What traits are best predicted at the individual level? (2) How well are the community weighted means and trait frequency distributions described? (3) What chemical properties may affect the estimates of traits by FT-NIR spectrometry? and (4) How many samples are needed in order to build an accurate local FT-NIR based model?

## Methods

### Study sites

Trait data were collected in the Reserva Florestal Ducke, located 26 km north-west of Manaus (02°55′ S/59°59′ W at the reserve headquarters). Reserva Ducke covers 10,000 ha (10 × 10 km) of terra-firme tropical rain forest. The mean annual temperature is around 26°C, and the mean annual rainfall around 2,400–2,700 mm, with the dry season occurring between July and October (Marques-Filho et al. 1981). Altitude varies from 30 to 120 m. Soils are derived from tertiary marine sediments from the Alter-do-Chão formation. There is a continuum of soil change from the upper flat terraces (plateaus) to the valleys around watercourses (Chauvel et al. 1987). Vegetation is dense terra-firme forest, with a closed canopy 30–37 m high and emergent trees of up to 40–45 m (Ribeiro et al. 1999).

Independent data for the external validation of models was collected in the Bosque da Ciência (BC), an urban forest located within the campus of the National Institute for Amazonian Research (INPA) in Manaus, which is 30 km south of Reserva Ducke. BC is also covered by dense terra-firme forest, but has some patches of secondary forest and has higher temperatures and lower precipitation than Reserva Ducke.

### Sampling design and data collection

We collected trait data for 1,324 tree individuals distributed in 17 permanent plots covering a gradient of

topographic conditions, from the high clayey plateaus to the sandy valleys. The plots are widely distributed across Reserva Ducke covering an area of 30 km$^2$ and they are at least 1 km apart. Each plot follows an altitude contour, so within each plot, altitude, distance to water table, and soil are consistent (Magnusson et al. 2005). The trees represent 432 named species (36% of the 1,176 tree species registered in the Reserve), plus 295 either unidentified individuals or identified to genus level.

We sampled all trees having a stem diameter at breast height (DBH) between 1 cm and 5 cm in a strip of 250 × 1 m per plot, which resulted on average in 74 ± 22 individuals per plot (mean ± SD). For each individual, we collected a branch at least 40 cm long. For plants with only short branches, the longest were collected. Branches were taken from the most brightly illuminated side of the plant canopy while trying to avoid damaged, sick, and epiphyll-covered leaves.

We measured 14 anatomical and morphological leaf and branch traits that are important for the carbon, water, and nutrient balance of the plant for each collected individual. Samples were processed within 24 h after collection and maintained in closed, humid, black, plastic bags in an air-conditioned room before processing. In the laboratory, we took (1) the best two leaves (i.e., mature and healthy) for wet/dry mass and leaf area, (2) a 4–5 cm terminal piece of the branch for wet/dry mass and volume, (3) a 1 cm length of branch next to the first for macro-anatomical measurements, and (4) another 0.5-cm piece for FT-NIR spectrometry. Leaves were measured for thickness with a micrometer (middle part, avoiding main veins), for chlorophyll content with a SPAD, scanned, weighed to a precision of 0.0001 g, and dried for 48 h at 60°C to obtain the dry mass. Petioles and rachises were weighted separately from the leaf lamina. Force to punch was measured for one leaf per individual, using a home-made penetrometer constructed using the general design as per Aranwela et al. (1999).

Branch pieces were first weighed to a precision of 0.001 g, the volume was determined with and without bark by the water displacement method, and then dried at 105°C for 3–4 d and weighed again. Anatomical measurements of the proportions of pith, xylem, and bark were made with a caliper under a stereo-microscope (magnification 40×), or by using images taken with a camera coupled to the stereomicroscope. The remainder of the branch and leaves, were made into a voucher specimen and dried following common practice (i.e., newspaper sheets, in an oven at 60°C for 2–3 d or until dry). The branch piece cut for FT-NIR was stored in a paper envelope with silica, which was replaced until it was dry.

Three dry leaves per plant were scanned with FT-NIR, taking one reading at the adaxial and another at the abaxial surface, in the middle part of the leaf, avoiding main veins and where there were any visible signs of disease. For the branches, one reading was taken from each side of the dry transversal branch cut, which included the bark. Leaf and branch spectra were collected with a Thermo Nicollet spectrophotometer (Thermo Fischer Scientific, Waltham, Massachusetts, USA), using the Antaris FT-NIR II Method Development System (MDS). The spectral readings are expressed as absorbance values from the wavelengths of 1,000–2,500 nm in the near-infrared and the spectrum for each sample reading consists of 1,557 absorbance values, equivalent to a resolution of 0.96 nm. Each reading produced by the instrument was the average of 16 measurements of absorbance at a certain wavelength. A black, opaque material was placed over the point where the spectral readings were collected to avoid light scattering. A background calibration reading was performed before each reading was taken. The external data set collected at BC followed the same protocols described above for RFAD.

### Trait calculations

Based on the initial measurements, we calculated the following traits for each individual. The density and dry matter content of wood, bark, and branch were calculated based on the small branch sample (~5 cm long) that we had taken. Wood density was calculated as branch wood dry mass over branch wood volume (WD, g/cm$^3$), bark density (BarkD, g/cm$^3$) as bark dry mass over bark volume, branch density (BranchD, g/cm$^3$) as branch dry mass (including wood and bark) over branch volume, and wood dry matter content (WDMC, mg/g) and bark dry matter content (BDMC mg/g) were calculated as the dry mass divided by the fresh mass of the material. Proportions of bark (BarkProp), xylem (XylProp), and pith (PithProp) were calculated as the proportion occupied by the area of each tissue on the transversal branch cut.

Specific leaf area (SLA, cm$^2$/g) is the leaf area per unit of leaf dry mass. It was calculated by pooling the two leaves that were collected per branch, and by dividing their leaf area over their dry mass. Petioles were not included in the SLA calculation, because they contribute very little area for photosynthesis, but inflate the leaf mass. For compound leaves, the SLA was based on all the leaflets of the two leaves; the rachis was not included in the SLA calculation, for the same reason as for the petioles. Chlorophyll content per unit leaf area (Chl, µg/cm$^2$) was calculated based on the SPAD values using an equation for rainforest trees (Coste et al. 2010): Chl = (117.1 × SPAD)/(148.84 − SPAD). The leaf dry matter content (LDMC, mg/g) was calculated as the dry mass over the fresh mass of these tissues. Leaf thickness (LT, mm) is the average of the measurements of two leaves per plant. Leaf density (LD, g/cm$^3$) is the leaf dry mass per unit leaf volume, and it was calculated as the reciprocal of (SLA × LT). Force to punch (FP, N/mm) is the force in Newtons to perforate the leaf divided by the circumference of the punch.

Leaf traits included here are indicators of light capture and photosynthetic capacity (SLA, LD, Chl) and leaf defense and persistence (LDMC, LT, FP). WD and xylem proportion are related to biomechanical support, defense and tolerance to drought; pith proportion is also related to biomechanical support in soft wood, and to water and reserves storage; WDMC, bark density, and BarkDMC are related to branch defense and persistence.

### Analyses

Models to predict traits from leaf or branch FT-NIR spectra were built with partial least squares regressions

(PLSR). PLSR is a widely used method for analyzing the spectral data of biological materials, and is very well suited to the high multi-colinearity present in this kind of data. Being a linear method, a potential drawback is the loss of non-linear information. Non-linear methods have been proposed to deal with that, such as neural networks (ANN) or Least squares support vector machines (LS-SVM), but so far there is not much evidence that these nonlinear techniques perform better than classical linear algorithms (Nicolaï et al. 2007). This is due to the fact that NIR spectroscopy, both classical or FT-NIR, is essentially a very linear technique. Differences in spectral data between classical NIR or FT-NIR do not affect the performance of analytical methods used for chemometrics. The high resolution of FT-NIR instruments provide more spectral information, and less reliance on sophisticated chemometric algorithms, FT-NIR can thus be considered more robust to variations in analytical tools.

The spectra used for all models comprised of the average of readings taken on both sides of the leaves or branches, and all 1,557 variables (i.e., absorbance at each of 1,557 wavelengths) read for each sample. The first step in the modelling approach was to scan and clean the data set from outliers; i.e., values outside the known range of each trait or outside the 95% range of the observed values. We also evaluated biplots among traits to localize points too far from the expected correlation and then checked for potential errors. The optimal number of components for an effective model for each trait, at the same time avoiding over-fitting, was selected by minimizing the prediction residual sum of squares (PRESS). PRESS was calculated for successively more complex models (including a larger number of components), through a 10-fold cross-validation. Each of the cross-validations used 90% of the data points for model construction and 10% for validation, these subsets were selected at random. The models for which PRESS was a minimum were then selected as the best for each trait, and for these the root mean squared error of prediction (RMSE) and $R^2$ reported are the averages of the cross-validation. We also report the relative prediction error (%RMSE) as the percentage of the trait value range represented by the final model RMSE (as in Serbin et al. 2014). Data from BC was used as an external test data set. Estimates of trait values were derived for BC samples from the FT-NIR based PLSR equations developed for the Ducke data, and the quality of the predictions was assessed by the $R^2$ and RMSEP of the relationship between observed and predicted trait values.

In addition to the direct test of how well trait values can be estimated by NIR-spectra, we asked if community level trait values could be well estimated. Three community descriptors were evaluated: the average trait value for the site, which is equivalent to the community weighted mean (CWM), the variance of this average, and the whole community trait frequency distribution. The Komolgorov-Smirnov test was used to determine how well the estimated trait frequency distribution matched the observed one, $t$ tests compared means, and $F$ tests compared variances of the observed and predicted trait distributions.

The variable importance of projections (VIP), which describes the contribution of each variable, in this case the absorption at each wavelength, to the PLSR model, was used to determine which regions of the spectra were more associated to the prediction of each trait. We also ran a matrix of correlations among traits, to evaluate which of them may have been good predictions due to correlations with others.

To determine how many samples (plant individuals) are needed to build a NIR-based model for local use, i.e., to estimate traits for all individuals of a site based on direct measurements of a smaller sample, we used a structured sub-sampling and ran the PLSR model for each sub-sample. We sampled individuals non-randomly from the total pool, simulating what could be the actual procedure in an ecological study site. We took samples from all of our 17 plots, starting with the first 10 individuals and increasing continuously in amounts of 10 until the full sample. The results from each run were used to build the sampling effort curve.

## Results

### What traits are best predicted at the individual level?

We obtained moderate ($r^2 = 0.5$–$0.7$) to good ($r^2 > 0.7$, error level <15%) predictions for most of our traits at the individual level (Table 1, Fig. 1). Wood properties had, in general, higher coefficients of determination and lower errors than leaf properties. The key traits (SLA, LDMC, WD, WDMC), used in most functional trait studies can be reasonably well predicted ($R^2 = 0.69$–$0.78$, $RMSE_{rel} = 9$–$11\%$) using the simple and crude approach of taking NIR spectra from dry leaves and branches without any processing of samples (Fig. 2). A fair estimation can be obtained for xylem proportion ($R^2 = 0.61$, $RMSE_{rel} = 16\%$), bark density ($R^2 = 0.53$, $RMSE_{rel} = 17\%$) and BarkDMC ($R^2 = 0.61$, $RMSE_{rel} = 14\%$), even if bark represents only an average 36% ($\pm13\%$ [SD]) of the transversal area of the branch samples. Leaf Density (LD) can also be moderately well estimated, with around 15% of error ($R^2 = 0.54$, $RMSE_{rel} = 15\%$). Chlorophyll had the worst prediction ($R^2 = 0.30$, $RMSE_{rel} = 16\%$), and the branch proportions of pith and bark had the higher relative errors of prediction ($RMSE_{rel} > 20\%$).

The external validation with samples from Bosque da Ciência achieved good predictions for WD, WDMC and Branch Density ($R^2$ from 0.64 to 0.77, $RMSE_{rel} < 15\%$), a reasonable prediction of SLA ($R^2 = 0.69$, $RMSE_{rel} = 19\%$) and poor predictions (errors above 20%) for the other traits (Table 1). For SLA, LD, and chlorophyll, the prediction errors were concentrated in the larger trait values (Fig. 2; Appendix S1).

### How well are the community trait distributions described?

There was a good agreement between observed and predicted trait frequency distributions for the Ducke community (Fig. 2; Appendix S1: Fig. S1); although for most traits, the predicted distributions were more peaked around the mean. None of the 14 traits' observed and predicted frequency distributions differed significantly in their mean, but all differed significantly in their variance (Table 2). The shape of the frequency distributions of SLA, LDMC, branch density, WD, and WDMC did not differ significantly between observed and predicted values.

TABLE 1. Performance of models to determine traits from Fourier-Transformed Near-Infrared Spectrometry (FT-NIR) spectra, given by cross-validated coefficients of determination ($R^2$) and root mean square errors of predictions (RMSE).

| Organ and trait | Units | $N$ | Range | Ncomp | Internal validation (Duke) | | | External validation (BC) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | $R^2$ | RMSE | RMSE$_{rel}$ (%) | $R^2$ | RMSE | RMSE$_{rel}$ (%) |
| Leaf | | | | | | | | | | |
| Specific leaf area (SLA) | cm$^2$/g | 1064 | 6.01–25.34 | 25 | 0.78 | 1.48 | 11 | 0.69 | 2.77 | 19 |
| Leaf dry matter content (LDMC) | mg/g | 1067 | 208.9–667.3 | 28 | 0.69 | 40.67 | 9 | 0.36 | 64.63 | 15 |
| Leaf thickness (LT) | mm | 1081 | 0.09–0.349 | 22 | 0.50 | 0.03 | 16 | 0.06 | 0.05 | 25 |
| Leaf density (LD) | g/cm$^3$ | 1071 | 0.148–0.698 | 28 | 0.54 | 0.06 | 15 | 0.32 | 0.32 | 10 |
| Chlorophyll (Chl) | µg/cm$^2$ | 1089 | 32.46–102.3 | 18 | 0.30 | 10.45 | 16 | 0.37 | 13.2 | 20 |
| Force to punch (FP) | N/mm | 1070 | 1.16–7.84 | 16 | 0.35 | 0.88 | 21 | – | – | – |
| Branch | | | | | | | | | | |
| Branch density (BranchD) | g/cm$^3$ | 1056 | 0.18–0.83 | 31 | 0.78 | 0.05 | 9.9 | 0.72 | 0.07 | 13.3 |
| Wood density (WD) | g/cm$^3$ | 1066 | 0.18–0.93 | 27 | 0.76 | 0.06 | 10.5 | 0.77 | 0.07 | 12.3 |
| Wood dry matter content (WDMC) | mg/g | 1070 | 256.7–727.2 | 26 | 0.79 | 37.55 | 7.0 | 0.63 | 70.72 | 12.6 |
| Bark density (BarkD) | g/cm$^3$ | 1062 | 0.16–0.75 | 30 | 0.53 | 0.07 | 16.9 | 0.34 | 0.11 | 23.9 |
| Bark dry matter content (BarkDMC) | mg/g | 1086 | 148.7–630.3 | 28 | 0.61 | 54.44 | 13.7 | 0.43 | 82.57 | 19.1 |
| Pith proportion (PithProp) | | 900 | 0–0.36 | 26 | 0.51 | 0.05 | 47.3 | 0.31 | 0.09 | 71.4 |
| Xylem proportion (XylProp) | | 919 | 0.1–0.81 | 20 | 0.61 | 0.08 | 15.7 | 0.36 | 0.10 | 18.5 |
| Bark proportion (Bark Prop) | | 936 | 0.1–0.78 | 20 | 0.56 | 0.08 | 21.7 | 0.15 | 0.12 | 35.9 |

*Notes:* RMSE$_{rel}$ gives the percentage of the trait value range represented by the error in estimate. Ncomp gives the number of components retained in the best model. Both internal validation (within the Ducke forest) and external validation (using the Ducke equations to predict traits in Bosque de Ciência (BC)) are given. FP was not measured for BC plants, so there is no external validation for this trait. Dashes indicate missing data and analyses.

Applying the model developed at Ducke to BC; we see that the means of the observed and predicted trait distributions of the BC community did not differ for 10 out of the 14 traits, and did not differ in variance for 5 out of 8 branch traits (Table 2). The shape of the trait distributions also did not differ significantly between observed and predicted values for 8 out of 14 traits (Table 2).

*What is the mechanistic basis of predictions?*

The variable importance of projections (VIP) shows that the more important wavelengths for SLA prediction are related to the bands where cellulose is strongly absorbed (between 2,300 and 2,500 nm), and to protein (N-H bonds) at around 1,000 and 2,180 nm (Appendix S1: Fig. S2). In fact, the VIP profile of SLA looks very similar to the NIR spectra of pure cellulose (Workman and Weyer 2007), with an overlapping peak at around 2,180 nm, which corresponds to protein absorption. The LDMC VIP profile has important peaks in the same regions as SLA, but also has stronger peaks than SLA at around 1,780 nm and 1,940 nm (corresponding to cellulose) and around 2,060 nm (corresponding to protein/nitrogen; Appendix S1: Fig. S2). Leaf thickness and leaf density VIP profiles resemble the ones for LDMC a lot, but with proportionally greater contributions of wavelengths >2,400 nm, and lower wavelength contributions at around 1,780 m (Appendix S1: Fig. S2). For the best predicted branch traits (WD, WDMC, BranchD), some absorption peaks correspond to starch, oil, and proteins, besides the important peaks for cellulose, but some regions specifically associated with high absorption by cellulose did not have the highest VIPs (Appendix S1: Fig. S2).

To evaluate how correlations among traits may have influenced the capacity to predict them, we present a table of their correlations (Appendix S1: Fig. S3). Among branch traits, moderate correlations (0.50–0.64) were found between WD and WDMC, between WDMC and bark density, BarkDMC, and pith proportion. Stronger correlations (0.77–0.83) were found between bark proportion and xylem proportion and BarkDMC and bark density. Among leaf traits, LD had moderate correlation with SLA and LDMC.

*How many samples are needed for a local model?*

The evaluation of sampling effort needed to build a good local model indicated that around 400 individuals are necessary to attain the lowest relative error level (RMSE$_{rel}$) for most of the traits (Fig. 3). However, for most traits, the improvement is a modest 2–5% when moving from 100 to more samples, and small samples of 100–200 individuals spread across the range of trait values may work well.

*A summary of performance of NIR-based models at different levels*

Table 3 summarizes the performance of FT-NIR-based models in predicting individual traits and community level traits for the same site or across sites. The key traits (SLA, LDMC, WD, WDMC) have good predictions at any level, and also across sites. Other traits perform well at one or another level, but are less consistent. Among these, LD, BarkDMC, and xylem proportion can still be recovered reasonably within the site where the model was developed.
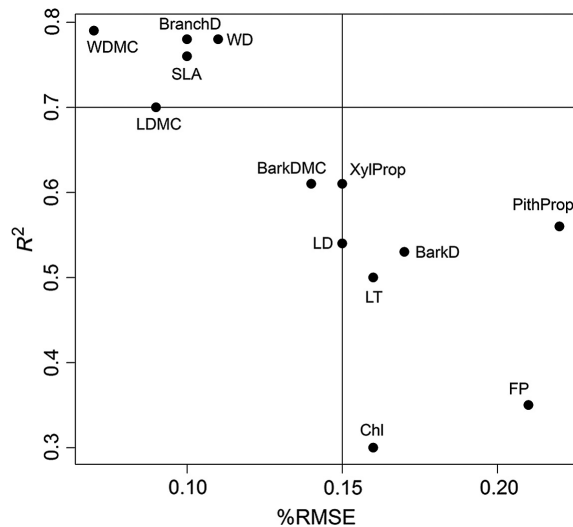
FIG. 1. Accuracy of Fourier-Transformed Near-Infrared Spectrometry (FT-NIR)-based models to estimate values of 14 traits for Reserva Ducke, showing the coefficient of determination ($R^2$) and relative error rate (%RMSE). Lines indicate thresholds for good models: $R^2 > 0.70$ and relative error rate (%RMSE) <15%. For trait abbreviations, see Table 1.

## DISCUSSION

We asked whether FT-NIR spectrometry can be extensively used to estimate trait values for individuals and communities, and if the models developed from NIR spectra for one site can be used to estimate traits in another site. Our results indicate that FT-NIR spectrometry has good potential as a tool for estimating the range of functional trait values of sites, and for predicting values for individuals for 8 out of the 14 traits that were evaluated. We also found that good local models can be built using around 400 or fewer individuals. This provides an unprecedented opportunity for rapidly screening communities to find their trait values, especially in remote sites where trait measurements are logistically challenging because of the lack of electrical power and a basic infrastructure. Samples from these sites, dried as traditional botanical vouchers, can be read in the spectrometer when back in the lab. Also, sites that are too difficult to access, but have previously been collected for herbaria can be included in functional analyses, thus unlocking the potential of herbaria for trait-based ecology and global trait screening efforts (e.g., Kattge et al. 2011).

### The core traits can be estimated rapidly and accurately

Our results indicate that the core traits (SLA, WD, and WDMC) reflecting the main global spectra in plant strategies (Wright et al. 2004, Chave et al. 2009) can be estimated accurately at the individual level, and should therefore provide better estimates of trait values at the individual level rather than using a mean value for the species.

The model developed at our core site (Ducke) was good for estimating individual trait values at a different site (BC) for some of the wood traits (BranchD, WD, and WDMC), and had acceptable prediction errors, but wasn't good for most leaf traits. For SLA, the $R^2$ between

predicted and observed values was high (0.69), but the mean relative error was 19%, because of a greater underestimation at higher observed values. This points to a need for the inclusion of a bigger representation of extreme values over a wider geographic range to produce a generally transferable model.

### Community-level properties can be estimated with NIR-spectrometry

Even if trait estimates at the individual level are not considered to be sufficiently accurate in some cases, we were able to demonstrate that the means and distributions of community traits can generally be estimated more accurately. The trait frequency distributions of local sites allows for the mapping of trait mean, variance, range, and other statistics in large scale studies, or to feed global vegetation models (Sakschewski et al. 2015), and large-scale estimations of traits made possible by NIR could therefore improve the use of traits in large-scale applications. Community means and trait frequency distributions of most of the traits were well estimated for the external site, indicating that a good model developed for one site can be used to estimate community level properties of others. FT-NIR scans of vouchers from permanent plots can therefore provide a good estimation of ecosystem and community level properties without the need of new collections. This would also be a significant improvement over using trait distributions derived from trait databases (e.g., Kattge et al. 2011), which still cover only a few intensively studied sites and capture only a little of the global biodiversity (Jetz et al. 2016).

### Chemical/structural basis of NIR–traits relationships

Our VIP analyses have shown that chemical properties are important determinants of relationships between NIR spectra and SLA, LDMC, LT, and LD. In particular, the contents of cellulose and protein are associated with peaks of absorbance and the general VIP profile resembles the absorbance profile of cellulose (Workman and Weyer 2007). Cellulose contents can be directly related to all the leaf properties measured here, as it is the major constituent of cell walls, and wall thickness and the proportion of cell walls to empty spaces provide much of the anatomical basis for these traits (Poorter et al. 2009). Proteins, starch, and structural carbohydrates such as cellulose were shown to vary consistently across a range of LMA values (Poorter et al. 2009), given their association to growth strategies, i.e., fast growth is associated with leaves with less cellulose and proportionally higher content of proteins and starch.

The same can be said of the branch traits with best predictions (WD, WDMC, BranchD, xylem proportion, and BarkDMC). It was surprising, however, that lignin, an important component of xylem, did not have an important contribution to the association of wood traits with NIR spectra. Given the young stage of development of wood in terminal branches, as those used here, we may expect that the content of lignin was smaller than that of cellulose, and therefore its contribution to differentiation of traits smaller.
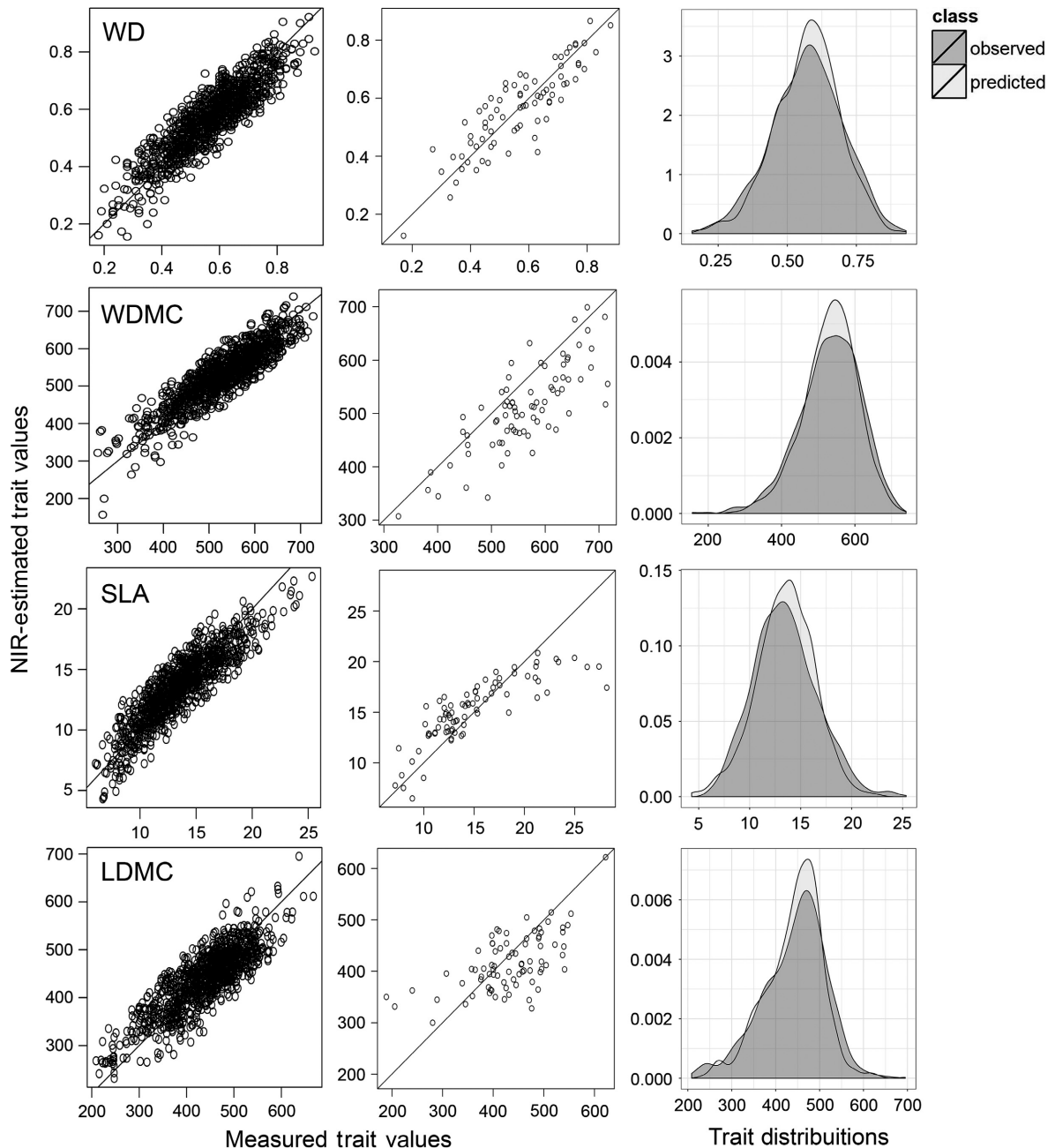
FIG. 2. Evaluation of performance of NIR-derived models in predicting trait values for individuals in the same site (left column), in a different site (middle column), and to predict community trait distributions (right column). Left column shows the relationships between trait values observed and estimated with the PLS-NIR model developed for samples from Reserva Ducke, $N = 900–1,089$ samples. The middle column show the relationships between trait values observed at Bosque da Ciência and trait values estimated with the PLS-NIR model developed for samples from Reserva Ducke, $N = 77–80$ individuals. Lines show the expected 1:1 relationship between measured and estimated values. The right column shows the distribution of trait values observed and estimated by the NIR-PLS model for Reserva Ducke, $N = 900–1,089$ samples. For trait abbreviations and units, see Table 1.

Despite these associations, there were also strong absorbance peaks that do not correspond to the known bands of the major components of plant tissues evaluated here. These may correspond to minor basic compounds or secondary compounds, but given the macro-structural properties evaluated here, most probably these peaks correspond to internal or external structural properties of materials. Surface rugosity affects light reflectance (Cooper et al. 2011), and the arrangement of cells and empty spaces affects the scattering of light within tissues, and therefore reflectance (Slaton et al. 2001). Without direct evaluation, we can only speculate, but our VIP results suggest that a large fraction of the important absorbance bands that predict the leaf and wood traits evaluated here is given by structural rather than basic chemical properties.

TABLE 2. Prediction accuracy of the observed trait frequency distributions of a site.

| Organ and trait | Internal validation (Ducke) | | | External validation (BC) | | |
|---|---|---|---|---|---|---|
| | Mean | Variance | Distribution | Mean | Variance | Distribution |
| Leaf | | | | | | |
| SLA | **0.01 (0.98)** | 0.83 (<0.001) | **0.05 (0.10)** | **0.52 (0.60)** | 0.44 (<0.001) | 0.29 (<0.001) |
| LDMC | **0.08 (0.93)** | 0.75 (<0.001) | **0.06 (0.04)** | 1.90 (0.06) | 0.47 (<0.001) | **0.21 (0.05)** |
| LT | −0.14 (0.97) | 0.57 (<0.001) | 0.10 (<0.001) | **1.27 (0.19)** | 0.54 (0.01) | 0.31 (<0.001) |
| LD | **0.05 (0.95)** | 0.62 (<0.001) | 0.09 (<0.001) | **0.07 (0.93)** | 0.36 (<0.001) | **0.20 (0.08)** |
| Chlorophyll | **0.04 (0.95)** | 0.35 (<0.001) | 0.16 (<0.001) | 1.61 (0.10) | 0.24 (<0.001) | 0.34 (<0.001) |
| Branch | | | | | | |
| BranchD | −0.06 (0.95) | 0.86 (0.02) | **0.06 (0.05)** | 0.79 (0.42) | **1.00 (0.99)** | **0.13 (0.56)** |
| WD | **0.04 (0.97)** | 0.83 (0.003) | **0.05 (0.16)** | **0.44 (0.65)** | **0.75 (0.22)** | **0.13 (0.53)** |
| WDMC | **0.06 (0.97)** | 0.82 (0.001) | **0.06 (0.06)** | −4.3 (<0.001) | **0.93 (0.75)** | 0.42 (<0.001) |
| BarkD | **0.06 (0.95)** | 0.65 (<0.001) | 0.10 (<0.001) | **−1.48 (0.14)** | **0.75 (0.22)** | **0.19 (0.14)** |
| BarkDMC | −0.14 (0.88) | 0.70 (<0.001) | 0.08 (<0.001) | −2.37 (0.02) | **1.03 (0.90)** | **0.18 (0.18)** |
| PithProp | −0.08 (0.94) | 0.61 (<0.001) | 0.13 (<0.001) | **1.28 (0.21)** | 0.32 (<0.001) | **0.22 (0.05)** |
| XylProp | −0.03 (0.98) | 0.65 (<0.001) | 0.10 (<0.001) | **−1.16 (0.24)** | 0.62 (0.04) | **0.15 (0.37)** |
| BarkProp | **0.13 (0.89)** | 0.60 (<0.001) | 0.14 (<0.001) | **0.60 (0.54)** | 0.33 (<0.001) | 0.26 (0.01) |

*Notes:* We tested whether the observed and predicted community values differ in their means (two-sample $t$ test), variance (variance-ratio test) and whole distribution fit (Komolgorov-Smirnov test). Values are the test statistics and $P$ values (in parentheses). Values in boldface type indicate comparisons that do not differ significantly between observed and estimated traits. Number of samples and units of measurement as in Table 1. Both internal validation (within the Ducke forest) and external validation (using the Ducke equations to predict traits in Bosque de Ciência, BC) are given.
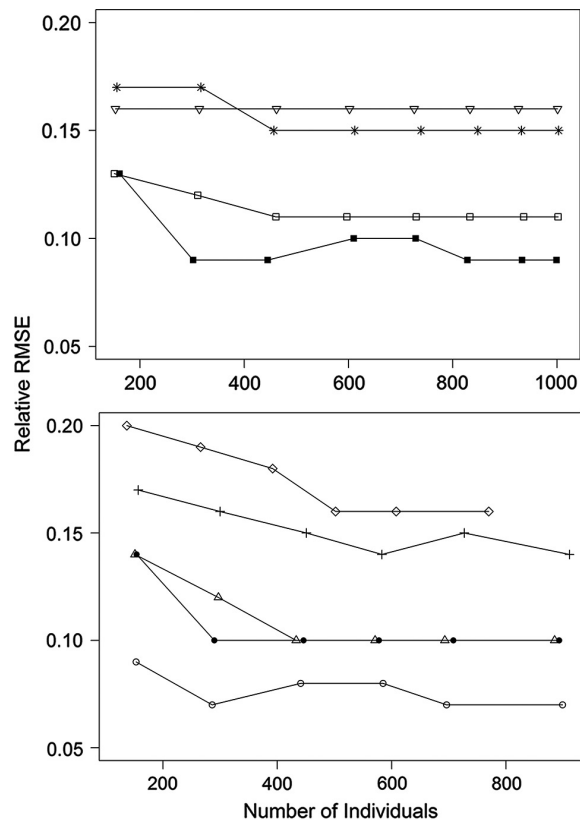


FIG. 3. Evaluation of the effort, given in number of individuals, to reach the lowest level of relative error (RMSE_rel) for trait values. In the upper panel are leaf traits: open squares, SLA; filled squares, LDMC; stars, LD; inverted triangle, LT. In the bottom panel are branch traits: filled dots, WD; open dots, WDMC; triangles, Branch Density; crosses, BarkDMC; diamonds, Xylem Proportion.

Some of the evaluated traits (the proportions of pith, xylem, and bark, and even the bark density and dry matter content), may have been relatively well recovered due to the correlations to traits that have a more clear structural/chemical link to NIR absorbance patterns. This is not unusual, other studies of wood NIR were able to recover mechanical properties due to indirect relationships with anatomical properties (Tsuchikawa and Kobori 2015).

*Methodological limitations and how to move forward*

The poor estimates for some traits are probably related to limitations in equipment and in measurement protocols, and not necessarily to the incapacity of spectrometry to recover them. Chlorophyll, for example, is well known to have absorbance peaks in the visible spectra (400–700 nm), which is not included in the spectral range read by our equipment (1,000–2,500 nm). The anatomical proportions of bark, xylem, and pith were determined using only two radial measurements across the transversal branch cut, and assuming a circle or ellipse to convert them to area, but many plants have irregular formats, so the imprecision in the original measurements may be the real cause of the lack of good predictions.

Wood traits were better predicted than leaf traits, and this may be related to the inclusion of the entire wood sample in the reading, while only a spot (1 cm$^2$) of the whole leaf can be scanned with our equipment. Therefore, the readings represent all the properties of the wood sample, while they represent only a variable proportion of the leaf, depending on leaf size and number of readings per leaf. Given that leaf structure is not homogenous, a different amount of veins may be included in each reading, affecting the absorption of cellulose, and increasing the variance between readings of t-

TABLE 3. Summary of the performance of NIR spectrometry in predicting trait values for individual plants and trait frequency distributions of communities.

| Organ and trait | Individual | | Community | |
|---|---|---|---|---|
| | Within site | Across sites | Within site | Across sites |
| Leaf | | | | |
| SLA | Yes | Yes† | Yes | |
| LDMC | Yes | Yes† | Yes | Yes |
| LT | | | | |
| LD | Yes | | | |
| Chlorophyll | | | | Yes |
| FP | | | | |
| Branch | | | | |
| BranchD | Yes | Yes | Yes | Yes |
| WD | Yes | Yes | Yes | Yes |
| WDMC | Yes | Yes | Yes | |
| BarkD | | | | |
| BarkDMC | Yes† | | | Yes |
| PithProp | | | | Yes |
| XylProp | Yes† | | | Yes |
| BarkProp | | | | |

*Notes:* Performance at the individual level was considered good (Yes) when model fit ($R^2$) between observed and predicted values was above 0.7 and prediction errors were less than 15%. Yes† denotes cases when the general fit was smaller than 0.7, but error levels are below 15%. At the community level, performance was considered good when the shape and mean did not differ between observed and predicted trait distributions. Within sites indicates evaluations using NIR-based models developed for a site to predict traits for the same site, while across sites indicates evaluations using the model developed at one site to predict traits of another site.

he same plant. Higher variance in the estimation of leaf traits may also be linked to the presence of fungi and bacteria, not detected by visual inspection (Pinto et al. 2000), which may affect leaf properties (Friesen et al. 2011).

More precise estimations of trait values can be obtained with more complex and time-demanding protocols than used here, such as transforming leaves or wood into powder (Serbin et al. 2014). The protocol suggested here is a simple way of obtaining trait estimations with the minimum processing requirements, instead of the time-consuming process of turning samples into powder. It may not even be possible to take material from herbaria for destructive processing, in which case the analysis of entire leaves is the only alternative. The branch sample needed for estimation of wood properties is very small (0.5 cm) and it should not be a problem to obtain sufficient material from herbaria samples. It should be remembered that the particle size distribution of ground samples has a large effect on reflectance (Foley et al. 1998), and ensuring that ground samples are homogeneous may create another difficulty for expediting analyses and ensuring comparability among studies.

Some improvements on the protocols used here should also provide better results, for example, we did not attempt to identify the exact spot where the leaf thickness was measured in order to use the same spot for the NIR reading. This could reduce the noise and also improve the leaf density model, which depends on leaf thickness values. Using a mean based on more readings per leaf, including the

gradient of venation size, or spectrometers with a larger scan area may also provide better estimates of leaf traits, and equipment that covers the light spectra from visible to infrared should help decrease other potential sources of error.

We focused on dry materials rather than fresh leaves or wood, to provide the basis of trait estimations from botanical collections. In the future, a more general development of spectrometry derived traits will be equations to make dry and fresh measurements comparable, by the removal of the water absorption profile.

### Spectrometry as an alternative to direct measurements

The estimation of traits by NIR spectrometry is faster and cheaper (not considering the cost of equipment purchase) than by direct measurement. The cost of the equipment will quickly pay for itself. The time necessary for processing and measuring one sample for all 14 leaf and wood traits by well-trained people was around 40 min (or approximately 2.7 min per measurement, not counting the time required for drying the samples), while the measurement of leaf and wood spectra for one sample takes around 1.5 min. This makes it possible to obtain trait values for a larger number of individuals and traits than are normally obtained during typical trait studies.

We envision that FT-NIR may provide an easy, fast, and cheap way to estimate individual traits and thus will advance our knowledge of community and ecosystem functioning. The model developed in this study is still not adequate for widespread application to Amazonia, because it used samples from a restricted range of environments, but does demonstrate that such a model is possible and this will be the next step in our study. Chemical traits can also be added to the list of traits evaluated here, so from a single NIR reading a large suite of morphological and chemical traits can be derived, and if a library of species identities is built, also these can be obtained or checked. This would free up time for the measurement of the more difficult, "hard" traits that are more time consuming to measure and have typically been measured only for a few species or individuals.

We conclude that NIR based models can make it possible to create extensive databases of traits of a scale that was hitherto impossible. It also unlocks herbarium materials for trait values estimation, thus expanding our knowledge of community and ecosystem functioning from local to global scales.

### LITERATURE CITED

Aranwela, N., G. Sanson, and J. Read. 1999. Methods of assessing leaf-fracture properties. New Phytologist 144:369–393.

Armstrong, P. R., E. B. Maghirang, F. Xie, and F. E. Dowell. 2006. Comparison of dispersive and Fourier-transform NIR instruments for measuring grain and flour attributes. Applied Engineering in Agriculture 22:453–457.

Asner, G. P., R. E. Martin, L. Carranza-Jimenez, F. Sinca, R. Tupayachi, C. B. Anderson, and P. Martinez. 2014. Functional and biological diversity of foliar spectra in tree canopies throughout the Andes to Amazon region. New Phytologist 204:127–139.

Asner, G. P., R. E. Martin, D. E. Knapp, R. Tupayachi, C. B. Anderson, F. Sinca, N. R. Vaughn, and W. Llactayo. 2017. Airborne laser-guided imaging spectroscopy to map forest trait diversity and guide conservation. Science 355:385–389.

Chauvel, A., Y. Lucas, and R. Boulet. 1987. On the genesis of the soil mantle of the region of Manaus, Central Amazonia, Brazil. Experientia 43:234–241.

Chave, J., D. Coomes, S. Jansen, S. L. Lewis, N. G. Swenson, and A. E. Zanne. 2009. Towards a worldwide wood economics spectrum. Ecology Letters 12:351–366.

Cooper, P. A., D. Jeremic, S. Radivojevic, Y. T. Ung, and B. Leblon. 2011. Potential of near-infrared spectroscopy to characterize wood products. Canadian Journal of Forest Research 41:2150–2157.

Coste, S., C. Baraloto, C. Leroy, É. Marcon, A. Renaud, A. D. Richardson, J.-C. Roggy, H. Schimann, J. Uddling, and B. Hérault. 2010. Assessing foliar chlorophyll contents with the SPAD-502 chlorophyll meter: a calibration test with thirteen tree species of tropical rainforest in French Guiana. Annals of Forest Science 67:607.

Doughty, C. E., G. P. Asner, and R. E. Martin. 2011. Predicting tropical plant physiology from leaf and canopy spectroscopy. Oecologia 165:289–299.

Durante, F. M., N. Higuchi, A. Almeida, and A. Vicentini. 2013. Species spectral signature: Discriminating closely related plant species in the Amazon with near-infrared leaf-spectroscopy. Forest Ecology and Management 291:240–248.

Foley, W. J., A. McIlwee, I. Lawler, L. Aragones, A. P. Woolnough, and N. Berding. 1998. Ecological applications of near infrared reflectance spectroscopy a tool for rapid, cost-effective prediction of the composition of plant and animal tissues and aspects of animal performance. Oecologia 116:293–305.

Friesen, M. L., S. S. Porter, S. C. Stark, E. J. VonWettberg, J. L. Sachs, and E. Martinez-Romero. 2011. Microbially mediated plant functional traits. Annual Review of Ecology and Systematics 42:23–46.

Garnier, E., and M. L. Navas. 2012. A trait-based approach to comparative functional plant ecology: Concepts, methods and applications for agroecology. A review. Agronomy for Sustainable Development 32:365–399.

Gomes, A. C. S., et al. 2013. Local plant species delimitation in a highly diverse Amazonian forest: Do we all see the same species? Journal of Vegetation Science 24:70–79.

Hopkins, M. J. G. 2007. Modelling the known and unknown plant biodiversity of the Amazon Basin. Journal of Biogeography 34:1400–1411.

Igne, B., J. B. Reeves, G. McCarty, W. D. Hively, E. Lundc, and C. R. Hurburgh. 2010. Evaluation of spectral pretreatments, partial least squares, least squares support vector machines and locally weighted regression for quantitative spectroscopic analysis of soils. Journal of Near Infrared Spectroscopy 18:167–176.

Jetz, W., et al. 2016. Monitoring plant functional diversity from space. Nature Plants 2:16024.

Kattge, J., et al. 2011. TRY—a global database of plant traits. Global Change Biology 17:2905–2935.

Krajšek, S. S., P. Buh, A. Zega, and S. Kreft. 2008. Identification of herbarium whole-leaf samples of Epilobium species by ATR-IR spectroscopy. Chemistry and Biodiversity 5:310–317.

Lang, C., F. R. C. Costa, J. L. C. Camargo, F. M. Durgante, and A. Vicentini. 2015. Near infrared spectroscopy facilitates rapid identification of both young and mature Amazonian tree species. PLoS ONE 10:e0134521.

Magnusson, W. E., A. P. Lima, R. Luizao, F. R. C. Costa, C. V. de Castilho, and V. F. Kinupp. 2005. RAPELD: A modification of the Gentry method for biodiversity surveys in long-term ecological research. Biota Neotropica 5:bn01005022005.

Marques-Filho, A. O., M. N. G. Ribeiro, H. M. Santos, and J. M. Santos. 1981. Estudos climatológicos da Reserva Florestal Ducke. Acta Amazonica 11:759–768.

Nicolaï, B. M., K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K. I. Theron, and J. Lammertyn. 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biology and Technology 46:99–118.

Paine, C. E. T., C. Baraloto, J. Chave, and B. Hérault. 2011. Functional traits of individual trees reveal ecological constraints on community assembly in tropical rain forests. Oikos 120:720–727.

Perez-Harguindeguy, N., et al. 2013. New handbook for standardized measurement of plant functional traits worldwide. Australian Journal of Botany 61:167–234.

Pinto, L. S. R. C., J. L. Azevedo, J. O. Pereira, M. L. C. Vieira, and C. A. Labate. 2000. Symptomless infection of banana and maize by endophytic fungi impairs photosynthetic efficiency. New Phytologist 147:609–615.

Poorter, L., et al. 2008. Are functional traits good predictors of demographic rates? Evidence from five neotropical forests. Ecology 89:1908–1920.

Poorter, H., Ü. Niinemets, L. Poorter, I. J. Wright, and R. Villar. 2009. Causes and consequences of variation in leaf mass per area (LMA): a meta-analysis. New Phytologist 182:565–588.

Ramirez, J. A., J. M. Posada, I. T. Handa, G. Hoch, M. Vohland, C. Messier, and B. Reu. 2015. Near-infrared spectroscopy (NIRS) predicts non-structural carbohydrate concentrations in different tissue types of a broad range of tree species. Methods in Ecology and Evolution 6:1018–1025.

Ribeiro, J. E. L., et al. 1999. Flora da Reserva Ducke. INPA/DFID, Manaus, Brazil.

Sakschewski, B., W. von Bloh, A. Boit, A. Rammig, J. Kattge, L. Poorter, J. Peñuelas, and K. Thonicke. 2015. Leaf and stem economics spectra drive diversity of functional plant traits in a dynamic global vegetation model. Global Change Biology 21:2711–2725.

Serbin, S. P., A. Singh, B. E. McNeil, C. C. Kingdon, and P. A. Townsend. 2014. Spectroscopic determination of leaf morphological and biochemical traits for northern temperate and boreal tree species. Ecological Applications 24:1651–1669.

Skoog, D. A., F. J. Holler, and S. R. Crouch. 2007. Principles of instrumental analyses. Sixth edition. Thomson Brooks/Cole, Singapore. http://zeus.qui.ufmg.br/~valmir/livros/analitica/Principles%20of%20Instrumental%20Analysis%20(6ed)%20-%20Skoog,%20Holler%20&%20Crouch.pdf

Slaton, M. R., E. R. Hunt, and W. K. Smith. 2001. Estimating near-infrared leaf reflectance from leaf structural characteristics. American Journal of Botany 88:278–284.

Stahl, U., B. Reu, and C. Wirth. 2014. Predicting species' range limits from functional traits for the tree flora of North America. Proceedings of the National Academy of Sciences USA 111:13739–13744.

ter Steege, H., et al. 2013. Hyperdominance in the Amazonian tree flora. Science 342:325–334.

Tsuchikawa, S., and H. Kobori. 2015. A review of recent application of near infrared spectroscopy to wood science and technology. Journal of Wood Science 61:213–220.

Violle, C., M. L. Navas, D. Vile, E. Kazakou, C. Fortunel, I. Hummel, and E. Garnier. 2007. Let the concept of trait be functional! Oikos 116:882–892.

Violle, C., B. J. Enquist, B. J. McGill, L. Jiang, C. H. Albert, C. Hulshof, V. Jung, and J. Messier. 2012. The return of the variance: intraspecific variability in community ecology. Trends in Ecology & Evolution 27:244–252.

Workman, J., and L. Weyer. 2007. Practical guide to interpretive near-infrared spectroscopy. CRC Press, Taylor & Francis, Boca Raton, Florida, USA.

Wright, I. J., et al. 2004. The worldwide leaf economics spectrum. Nature 428:821–827.

Wu, J., et al. 2017. Convergence in relationships between leaf traits, spectra and age across diverse canopy environments and two contrasting tropical forests. New Phytologist 214:1033–1048.

## Supporting Information

Additional supporting information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/eap.1728/full

## Data Availability

Data available from the Brazilian Program for Biodiversity Research Data Repository: https://ppbiodata.inpa.gov.br/metacatui/#view/PPBioAmOc.226.2