



Highlighted Student Research

Mapping distribution of woody plant species richness from field rapid assessment and machine learning

Bo-Hao PERNG¹, Tzeng Yih LAM^{1,2,*}, Su-Ting CHENG¹, Sheng-Hsin SU³, Kristina J. ANDERSON-TEIXEIRA^{4,5}, Norman A. BOURG⁵, David F. R. P. BURSLEM⁶, Nicolas CASTAÑO⁷, Álvaro DUQUE⁸, Sisira EDIRIWEERA⁹, Nimal GUNATILLEKE¹⁰, James A. LUTZ¹¹, William J. MCSHEA⁴, Mohamad Danial MD SABRI¹², Vojtech NOVOTNY^{13,14}, Michael J. O'BRIEN¹⁵, Glen REYNOLDS¹⁶, George D. WEIBLEN¹⁷, Daniel ZULETA¹⁸

1. School of Forestry and Resource Conservation, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan. 2. Department of Forest Resources Management, Faculty of Forestry, The University of British Columbia, 2045-2424 Main Mall, Vancouver, BC, V6T 1Z4, Canada. 3. Fushan Research Center, Taiwan Forestry Research Institute, No. 1, Fushan, Shuangpi Road, Huxi Village, Yuanshan Township, Yilan 264013, Taiwan. 4. Smithsonian's National Zoo & Conservation Biology Institute, 1500 Remount rd., Front Royal, VA, USA. 5. Smithsonian Tropical Research Institute, Luis Clement Avenue, Bldg. 401 Tupper Balboa Ancon, Panama, Republic of Panama. 6. School of Biological Sciences, University of Aberdeen, St Machar Drive, Aberdeen, AB24 3UU, UK. 7. Herbario Amazónico Colombiano, Instituto Amazónico de Investigaciones Científicas Sinchi, Calle 20 # 5 44, Bogotá, Colombia. 8. Department of Forest Sciences, National University of Colombia, Cra. 65 #59a-110, Medellín, Antioquia, Colombia. 9. Faculty of Applied Sciences, Uva Wellassa University, Badulla 90 000, Sri Lanka. 10. University of Peradeniya, Galaha Rd, 20400, Peradeniya, Sri Lanka. 11. Department of Wildland Resources, Utah State University, 5230 Old Main Hill, Logan, UT 84322, USA. 12. Forestry and Environment Division, Forest Research Institute Malaysia, 52109 Kepong, Selangor, Malaysia. 13. Biology Centre, Institute of Entomology of the Czech Academy of Sciences, Branišovská 1160/31 37005 České Budějovice, Czech Republic. 14. Faculty of Sciences, University of South Bohemia, Branisovska 31, 37011 Ceske Budejovice, Czech Republic. 15. Estación Experimental de Zonas Áridas, Consejo Superior de Investigaciones Científicas, Almería, Spain. 16. South East Asia Rainforest Research Partnership (SEARRP), Danum Valley Field Centre PO Box 60282, 91112 Lahad Datu, Sabah, Malaysia. 17. Bell Museum & Department of Plant & Microbial Biology, University of Minnesota, Saint Paul, Minnesota, 55018 USA. 18. Forest Global Earth Observatory, Smithsonian Tropical Research Institute, Washington, DC, USA. *Corresponding author's phone no.: +1-604-822-5478; E-mail: tylam.forest@gmail.com

(Manuscript received 13 July 2023; Accepted 29 November 2023; Online published 1 January 2024)

ABSTRACT: Sustainable forest management needs information on spatial distribution of species richness. The objectives of this study were to understand whether knowledge, method, and effort of a rapid assessment affected accuracy and consistency in mapping species richness. A simulation study was carried out with nine 25–50 ha census plots located in tropical, subtropical, and temperate zones. Each forest site was first tessellated into non-overlapping cells. Rapid assessment was conducted in all cells to generate a complete coverage of proxies of the underlying species richness. Cells were subsampled for census, where all plant individuals were identified to species in these census cells. An artificial neural network model was built using the census cells that contain rapid assessment and census information. The model then predicted species richness of cells that were not censused. Results showed that knowledge level did not improve the accuracy and consistency in mapping species richness. Rapid assessment effort and method significantly affected the accuracy and consistency. Increasing rapid assessment effort from 10 to 40 plant individuals could improve the accuracy and consistency up to 2.2% and 2.8%, respectively. Transect reduced accuracy and consistency by up to 0.5% and 0.8%, respectively. This study suggests that knowing at least half of the species in a forest is sufficient for a rapid assessment. At least 20 plant individuals per cell is recommended for rapid assessment. Lastly, a rapid assessment could be carried out by local communities that are familiar with their forests; thus, further supporting sustainable forest management.

KEY WORDS: artificial neural network, rapid biodiversity assessment, sustainable forest management, forest planning.

INTRODUCTION

Plant diversity is an important ecosystem service provided by forests that benefits human society such as providing medicinal plants (Gascon *et al.*, 2015) and protecting endangered species (Villero *et al.*, 2017). Conserving plant diversity is central to sustainably managing forest resources for meeting global demand on wood fiber. Sample (2005) suggested that sustainable forestry should include a mixture of reserves for biodiversity, intensively managed plantations for production, and moderately managed forests for multiple

objectives. This could be realized through forest landscape planning strategies such as the Biosphere Reserve Model (Pool-Stanvliet *et al.*, 2018) or the TRIAD zoning (Seymour and Hunter, 1992). To designate which forest parcels as reserves for plant diversity or for other management objectives requires information on spatial distribution of plant diversity. Thus, there is a continuing effort to refine existing approaches or develop new ones to generate the information.

Plant species richness, generally defined as the number of plant species in a forest, is adopted by the



Montréal Process and the Helsinki Process as an indicator for sustainable forestry (Hall, 2001). A map showing spatial patterns of species richness could support conservation planning. Since it is impossible to enumerate every plant individual, mapping spatial patterns of species richness relies on observed data and spatial models (Pearson and Carroll, 1998). Observed data could come from ground samples, but sample size is usually small and samples are sparsely distributed (Chong *et al.*, 2001; Haas *et al.*, 2006). Based on the assumption that heterogeneous environment could host more species, remotely sensed images have been used to derive proxy of species richness to map its spatial pattern over large area (Rocchini *et al.*, 2010). Example proxy include normalized difference vegetation index (Pau *et al.*, 2012) and fraction of photosynthetically active radiation (Coops *et al.*, 2019). Despite its many potentials, there are challenges such as spatial resolution of images, expertise needed in image processing, and complex data analyses (Rocchini *et al.*, 2016).

A proxy will theoretically predict spatial distribution of species richness well as long as it is highly correlated with the underlying species richness. This is the motivation behind developing different proxies from remotely sensed images (Rocchini *et al.*, 2010). In this study, we propose deriving proxy from field rapid assessment of plant species richness. Our proposed strategy is adapted from Lam *et al.* (2018). A forest is first tessellated into non-overlapping cells of equal size, and each cell is visited. A number of plant individuals are rapidly selected and identified to species. Once all cells are rapidly assessed, a subsample of cells are randomly selected for census of the species within. As a result, all cells have information from the rapid assessment with only some cells having true species richness from the census. Lam *et al.* (2018) then took the information and built the estimators to estimate average cell-level species richness. However, the estimators were not intended to estimate spatial distribution of species richness. Thus, no mapping was carried out in Lam *et al.* (2018). To the best of our knowledge, there has been little work on using rapid assessment to assist spatial mapping of species richness. In this study, we further the work of Lam *et al.* (2018) by extracting proxies such as cell spatial locations and diversity indices from the rapid assessment for every cell. A prediction model is built with the census cells by correlating the proxies and the true species richness. The model then predicts species richness of the cells that have not been censused and produces a final spatial map of species richness.

The proposed rapid assessment could be carried out by an expert such as a botanist or by a local community familiar with a forest. Local or traditional knowledge is increasingly recognized for its contribution to sustainable forestry (Parrotta *et al.*, 2016). For example, Thapa *et al.* (1995) found that local knowledge increased

understanding of species diversity and ecological interactions in Nepal agroforestry. There are other benefits by involving local communities in rapid assessment. For example, principles of the Forest Stewardship Council certification scheme (FSC, 2012) requires any applicant to demonstrate active involvement and consultation with local or indigenous communities in forest management activities. A challenge involving a local community in rapid assessment is their knowledge on local plant species. Knowledge of a local community is likely to be different to an expert, e.g., members of local community are more likely to identify culturally significant or abundant plant species in a near perfect manner than less common species (Lam and Kleinn, 2008).

Humphreys *et al.* (2019) calculated that the average annual extinction rate of seed plant for the past 250 years was about 2.3 species per year, which was faster than expected. Pimm and Raven (2000) warned that many plant species could disappear before they were identified due to anthropogenic influence. With many communities around the world depending on forest natural resources, this highlights the urgent need to sustainably manage forest resources while conserving plant diversity. Sustainable forest management requires adequate information and appropriate tools (Baskerville, 1986). The information and tools are the scientific bases for monitoring spatial and temporal changes in a plant community (Winter *et al.*, 2008) so that a management decision could be made. The overall goal of this study was to assess whether information from a rapid assessment improved accuracy and consistency in mapping species richness. In particular, this study aimed to understand how (1) knowledge level during a rapid assessment, (2) method and effort of a rapid assessment, and (3) sampling intensity of a census influence the accuracy and consistency. An artificial neural network model was built to predict spatial distribution of species richness. Lastly, this study was carried out as a simulation study with nine 25 to 50 ha census plots from tropical, subtropical, and temperate regions.

MATERIALS AND METHODS

Data

In this study, nine large permanent Forest Dynamics Census Plots (hereafter as sites) from the Forest Global Earth Observatory Network (ForestGEO; forestgeo.si.edu) were used for the simulation (Table 1). Six of the nine sites were in the tropics with one in the subtropics and two in the temperate zone. The areas of the nine sites ranged from 25 to 50 ha. The plant census protocols were identical between sites. All woody plant individuals with diameter at breast height ≥ 1 cm were mapped, measured, and identified to species. In this study, plant individuals of unknown species were removed, and only the main stem of an individual

**Table 1.** Characteristics of the nine long-term forest dynamics census plots.

Plot	Country	Ecological Zone	Plot Area (ha)	Plot Plant Count	Plot Species Richness	Cell Count (N)	Cell Species Richness	Cell Plant Count	Year	Reference
Amacayacu	Colombia	Tropical	25	116,570	1,233	625	102 (23.4%)	187 (23.9%)	2007	Duque <i>et al.</i> , 2017; Zuleta <i>et al.</i> , 2020
Pasoh	Malaysia	Tropical	50	335,343	820	1250	127 (11.9%)	268 (19.3%)	1985	Manokaran and LaFrankie, 1990
Danum	Malaysia	Tropical	50	233,408	693	1250	71 (19.6%)	187 (21.9%)	2011	O'Brien <i>et al.</i> , 2022
Wanang	Papua New Guinea	Tropical	50	253,653	581	1250	84 (18.8%)	192 (28.1%)	2009	Vincent <i>et al.</i> , 2015
BCI	Panama	Tropical	50	221,758	302	1250	52 (17.2%)	177 (19.9%)	2010	Hubbell <i>et al.</i> , 1999; Condit <i>et al.</i> , 2019a,b,c
Sinharaja	Sri Lanka	Tropical	25	207,469	238	625	52 (19.8%)	332 (37.5%)	1996	Anderson-Teixeira <i>et al.</i> , 2015
Fushan	Taiwan	Subtropical	25	114,354	110	625	31 (26.0%)	183 (44.3%)	2004	Su <i>et al.</i> , 2007
SCBI	USA	Temperate	25.6	29,771	68	640	10 (27.9%)	47 (103%)	2008	Bourg <i>et al.</i> , 2013
Wind River	USA	Temperate	27.2	31,722	26	680	5 (24.4%)	50 (45.2%)	2010	Lutz <i>et al.</i> , 2013, 2014

Ecological zone is defined by the ForestGEO classification (forestgeo.si.edu). Plot plant count = total number of plant individuals in a plot. Plot species richness = total number of species in a plot. Cell count = total number of non-overlapping 20 × 20 m cells in a plot. Cell species richness = average number of species in a cell with its coefficient of variation in parentheses. Cell plant count = average number of plant individuals in a cell with its coefficient of variation in parentheses. Year = year of data collection

plant was used in the simulation. Lianas were not included in the analysis. Hence, species richness in this study was defined as the number of woody plant species. In summary, species richness of the nine sites ranged from 26 to 1233 species with 1163 to 8299 plant individuals per ha.

Rapid assessment

A site was first tessellated into non-overlapping cells of 20 × 20 m (0.04 ha) (Fig. 1A). The total number of cells (N) varied between sites from 625 to 1250 (Table 1). The average species richness in a cell ranged from 5 to 127 (Table 1). As mentioned above, rapid assessment was carried out in all N cells of a site (hereafter as N rapid cells) (Fig. 1C). Simulation of rapid assessment was designed around three factors: (1) knowledge level (KN), (2) rapid assessment effort (RE), and (3) rapid assessment type (RT). KN simulated how much a person know about plant species in a forest. The factor has three levels: 50% (KN50), 75% (KN75), and 100% (KN100) assuming that a person could identify 50%, 75%, and 100% of all the species in a forest, respectively (Fig. 1C). The levels KN50 and KN75 were chosen to represent knowledge of a local community, while KN100 represented an expert with full knowledge. To simulate KN50 and KN75, a “known” species list of a site was generated by randomly selecting species without replacement and with probability proportional to their total abundance in the site. In other words, the “known” species list consisted of randomly selected locally abundant species. This assumed that the survey person was more familiar with the locally abundant species than the rare ones.

The factor RE simulated effort spent in rapid assessment represented by the number of individuals in a

cell selected for species identification. RE had three levels: 10 (RE10), 20 (RE20), and 40 (RE40) individuals (Fig. 1C). For a selected individual, its species was matched to a “known” species list generated from a KN level as above. If its species was in the list, it was recorded; otherwise, the individual was discarded because it was unknown. As a result, the recorded number of selected individuals in a cell could be less than a designated RE level depending on the “known” species list. This procedure better reflected field operation and explored the effects of incomplete knowledge on mapping species richness.

RT simulated the method by which RE plant individuals were selected in a cell for identification. Two types of selection were explored: (1) random walk (RT_{rw}), and (2) transect (RT_{tr}) (Fig. 1C). RT_{rw} mimicked randomly walking around in a cell and selecting individuals. RT_{rw} was simulated by randomly selecting individuals in a cell with equal probability. RT_{tr} mimicked selecting individuals along a transect while traversing it. To simulate RT_{tr}, a transect was randomly placed across each row of cells in a site; thus, placement of transects was independent between rows. Each cell was then subdivided into five equal intervals. Number of individuals selected for identification in each interval was one-fifth of RE (i.e., RE/5). Individuals within each interval located closest to the transect in the perpendicular distance was selected. If the total number of individuals in a cell was less than RE, all individuals were selected.

Census and sampling intensity

A subsample of n rapid cells was selected for census by simple random sampling (SRS) (hereafter as census cells) (Fig. 1B). The n census cells were randomly selected with

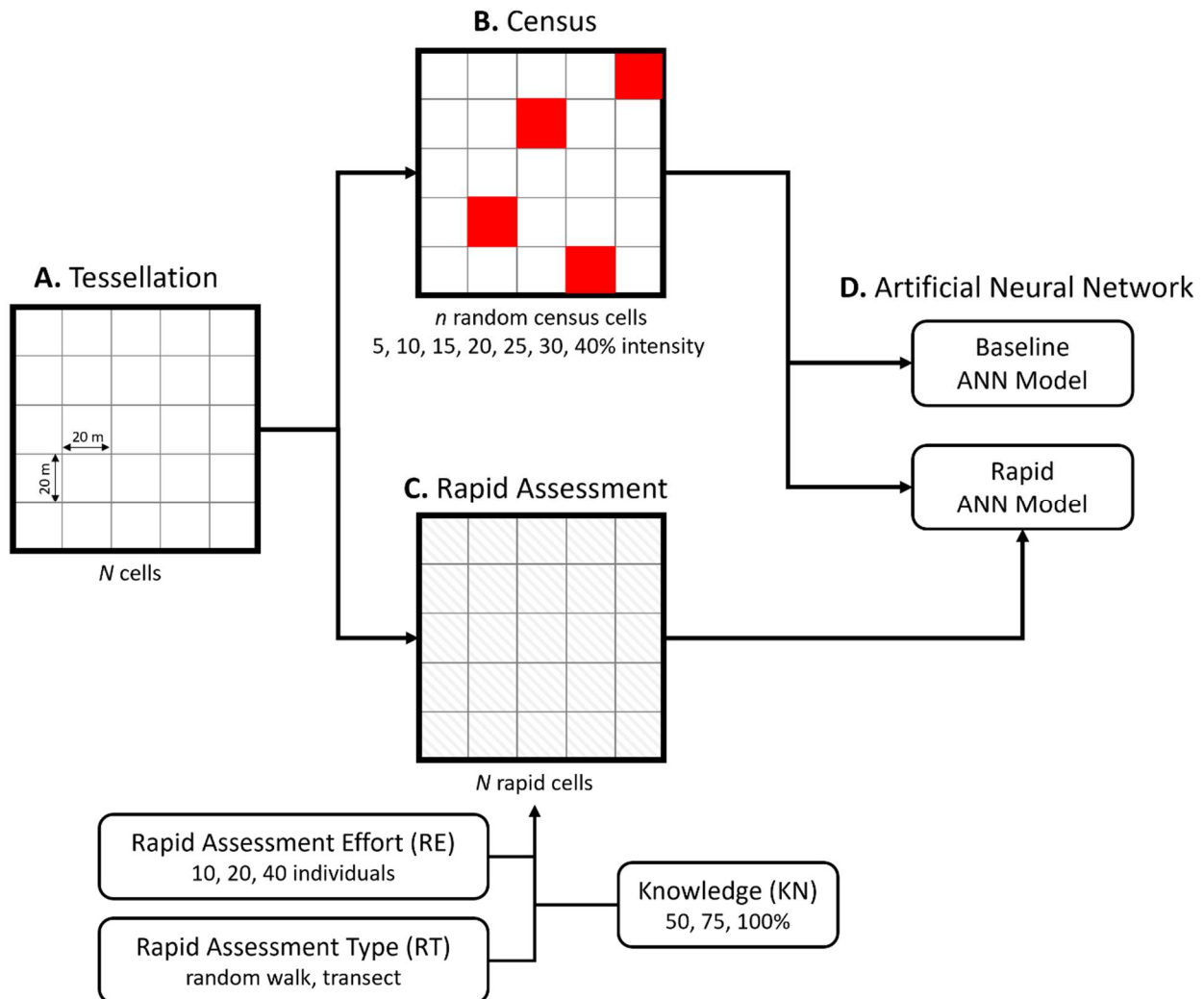


Fig. 1. Study methodology. **A.** a site is tessellated into N cells. **B.** Census is carried out in n random cells with different sampling intensities. **C.** Rapid assessment is carried out in all N cells with different combinations of knowledge level and rapid assessment effort and type; the two rapid assessment types are illustrated. **D.** Artificial neural network model building for baseline model with only census information and rapid model with both census and rapid assessment information.

equal probability and without replacement. A census was carried out by identifying species of all plant individuals in a census cell. Subsampling was carried out at the seven sampling intensities: 5%, 10%, 15%, 20%, 25%, 30% and 40% of N rapid cells in a site.

Artificial neural network

Artificial neural network (ANN) was applied to model the relationship between information from rapid assessment and census species richness and to predict underlying species richness of rapid cells that were not censused (Fig. 1D). To assess whether rapid assessment improved mapping of species richness, two ANN models were separately and independently built: baseline model and rapid model. The baseline model was trained using only the census information from the n census cells (Fig. 1D). The rapid model was trained using the information

from a rapid assessment combination and the census from the n census cells (Fig. 1D).

An ANN model had a layer-by-layer structure consisted of three main layers arranged in sequence: an input layer, a network of hidden layers, and an output layer. The process of ANN model building discussed below was illustrated in Supplement Fig. S1. The input layer organized input data and fed them to the network of hidden layers for model construction. For the baseline model, the input data consisted of (1) census species richness (S_{census}) as independent variable, and (2) cell coordinates as dependent variable. For the rapid model, the input data consisted of (1) census species richness, (2) cell coordinates, (3) rapid species richness, (4) rapid Shannon diversity index, and (5) rapid species vector. For the rapid model, the independent variable was census species richness, and the dependent variables were cell



coordinates, rapid species richness, rapid Shannon diversity index, and rapid species vector. Cell coordinates were the X-Y indices of the census cell. S_{census} was defined as the total (true) number of species found in a census cell. Rapid species richness was defined as the number of species found in the rapid assessment of a census cell. Rapid Shannon diversity index was the well-known Shannon index (Shannon, 1948) calculated with information from the rapid assessment of a census cell. Rapid species vector was a list of species name from the rapid assessment of a census cell. The input data were standardized so that they had similar distribution to maintain stability during model training. Standardization was done by subtracting values from their mean and dividing by their standard deviation.

In this study, the network of hidden layers consisted of five layers with 64 artificial neurons per layer. An artificial neuron was controlled by an activation function. The function carried out pointwise nonlinear transformation of input data into a “signal”. A strong “signal” would cause the neuron to fire its outputs to the neurons in the next hidden layer. This study applied the Mish activation function (Misra, 2020), which preserved a small amount of negative weight so that an inactive neuron could be turned on again during the training. After passing through all hidden layers, an initial ANN model was built. The output layer then predicted species richness for each n census cell (Fig. S1).

An ANN model was trained by minimizing the deviation between model prediction and true data (Fig. S1). As such, the next step was calculating the deviation between predicted species richness and S_{census} for each n census cell. The deviations were used to compute loss with the Huber loss function (Huber, 1964), which was robust in handling outliers by combining mean square error and mean absolute error. Gradient of the computed loss was then calculated (Fig. S1). Using the backpropagation algorithm by Rumelhart *et al.* (1986), the gradient was transmitted backward through the network of hidden layers (Fig. S1). The Adam gradient descent optimizer (Kingma and Ba, 2015) was used to update the parameters in the hidden layers based on the back-propagated gradient. The Adam optimizer was chosen for its adaptiveness in avoiding local minimums. To prevent overfitting, two regularizer algorithms (Reg1 and Reg2) and a dropout method by Srivastava *et al.* (2014) were applied. The training process was repeated 50 times to produce the final ANN model for statistical analyses.

Simulation

Simulation was carried out independently for each site with 100 iterations each. As a result, simulation was independent between sites and between iterations within a site. For an iteration of a site, two major steps were carried out: (1) selection of n census cells, and (2)

simulation of rapid assessment. For each of the seven sampling intensities, n census cells were selected by SRS. Thus, census cells were independent between sampling intensities. Eighteen rapid assessment combinations ($3 \text{ KN} \times 3 \text{ RE} \times 2 \text{ RT}$) were independently simulated. At first, a list of “known” species was generated according to a KN level as mentioned above. Thus, the “known” species list was randomized across iterations. Then, a RE number of individuals were selected according to a RT type. As mentioned above, the selected individuals were matched to the “known” species list, and individuals of unknown species were discarded. At the end of simulating an iteration for a site, there were 7 census datasets (7 sampling intensities) consisted of a list of plant individuals with their species and their corresponding census cell coordinates. There were also 18 rapid datasets similarly consisted of a list of plant individuals with their species and their corresponding rapid cell coordinates. Then, 7 final baseline ANN models were built from the census datasets, and 126 final rapid ANN models were built from both the census and the rapid datasets ($3 \text{ KN} \times 3 \text{ RE} \times 2 \text{ RT} \times 7$ sampling intensity). The simulation and ANN model training and building were carried out in Python. Lastly, the Python codes for ANN model training and building with example datasets were publicly available at https://github.com/Bo-Hao/ANN_example_code.

Analysis

Each final ANN model was used to predict species richness of $N - n = M$ cells that were not censused. For a final baseline ANN model, M cells were cells that were not censused. For a final rapid ANN model, M cells were rapid cells that were not censused. Let m be the m -th rapid cell that was not censused, $m = 1, \dots, M$, $S_{\text{pred},m}$ be the predicted species richness by a final ANN model for the m -th rapid cell, $S_{\text{true},m}$ be the total (true) number of species found in the m -th rapid cell. Percent absolute bias (PAB) of the m -th rapid cell was calculated as

$$PAB_m = \left| \frac{S_{\text{pred},m} - S_{\text{true},m}}{S_{\text{true},m}} \right| \times 100\%$$

Relative accuracy (Q) of the m -th rapid cell was calculated as $Q_m = S_{\text{pred},m} / S_{\text{true},m}$, where $Q_m > 1$ implied overestimation, and vice versa.

The goal of this study was to assess accuracy and consistency of rapid assessment in mapping species richness. Accuracy was defined as mean percent absolute bias (MPAB), and consistency was defined as the coefficient of variation of relative accuracy (CVQ). For a final ANN model, $MPAB_i$ of the i -th iteration was calculated by averaging PAB_{im} of all M rapid cells,

$$MPAB_i = \frac{\sum_{m=1}^M PAB_{im}}{M}$$

The CVQ_i of the i -th iteration was calculated as



$CVQ_i = SDQ_i / MQ_i \cdot 100\%$, where MQ_i was mean

relative accuracy $MQ_i = \sum_{m=1}^M Q_{im} / M$, and SDQ_i was standard deviation of relative accuracy

$SDQ_i = \sqrt{\sum_{m=1}^M (Q_{im} - MQ_i)^2 / (M - 1)}$. Hence, a

larger MPAB indicated less accuracy, and vice versa. A low CVQ suggested that relative accuracy between M rapid cells were similar due to the smaller standard deviation compared to the mean. In other words, a low CVQ implied that predicted species richness from the final ANN model will be high for rapid cells whose true species richness was high and will be low for rapid cells whose true species richness was low. Thus, CVQ indicated that predictions of the final ANN model consistently reflected the true underlying species richness. CVQ is a useful indicator if only relative species richness is needed in mapping spatial distribution of species richness for forest management decision making. If so, a rapid assessment combination that produces low CVQ is preferred.

To assess whether the information from rapid assessment improved mapping of species richness, accuracy and consistency of a final rapid ANN model were compared to those of a final baseline ANN model. For a given sampling intensity and for the i -th iteration, the difference in accuracy and consistency between a rapid ANN model (there were 18 rapid ANN models from the 18 rapid assessment combinations) and a baseline ANN model was calculated as

$$dMPAB_i = MPAB_{i,rapid} - MPAB_{i,baseline} \quad \text{and}$$

$$dCVQ_i = CVQ_{i,rapid} - CVQ_{i,baseline}, \text{ respectively. A}$$

negative value in the difference in accuracy (dMPAB) and the difference in consistency (dCVQ) suggested that the information from rapid assessment improved accuracy and consistency in mapping species richness than the baseline. To address the objectives of the influence of the rapid assessment factors on mapping species richness, a three-way Analysis of Variance (ANOVA) model was developed,

$$Y = \mu + site + KN + RE + RT$$

$$+ (KN \times RE) + (KN \times RT) + (RE \times RT) + (KN \times RE \times RT)$$

where, Y was either dMPAB or dCVQ. In the three-way ANOVA model, the site factor was included and treated as blocks to control for potential variations between sites. The three-way ANOVA model was fitted separately for each of the 7 sampling intensities. We conducted the Shapiro-Wilk test of normality for each combination of sites, sampling intensities, KN, RE, and RT. The average p-values for dMPAB and dCVQ were 0.3446 and 0.3287, respectively.

RESULTS

The three-way ANOVA analysis indicated a significant site effect on dMPAB and dCVQ consistently across sampling intensity ($p < 0.001$; Fig. 2). This suggested inherent differences among the nine sites on improving accuracy and consistency in mapping species richness. The differences in dMPAB and dCVQ between sites were evident (Fig. 3). Using the rapid assessment combination of KN50-RE10-RT_{tr} as an example, Fushan consistently produced the largest negative dMPAB compared to the other sites across sampling intensity (-2.2% to -5.9%; Fig. 3A). For dCVQ, it was Danum that produced the largest negative dCVQ for most of the sampling intensities (-7.6% to -25.7%; Fig. 3B); however, it was still Fushan that consistently produced a large negative dCVQ across sampling intensity (-2.6% to -7.1%; Fig. 3B). On the contrary, Pasoh produced a small positive dMPAB (0.36% to 0.60%; Fig. 3A) and dCVQ (0.31% to 0.59%; Fig. 3B). This suggested that information from rapid assessment had little impact on accuracy and consistency compared to the baseline for Pasoh. Nevertheless, most sites saw improved accuracy and consistency in mapping species richness by taking advantage of the information from rapid assessment.

The three-way ANOVA analysis consistently indicated an insignificant marginal effect of knowledge level (KN) on dMPAB and dCVQ across sampling intensity ($p > 0.1$; Fig. 2). The basis of comparison was the rapid assessment combination of KN50-RE10-RT_{tr}. Results showed that increasing knowledge level from 50% (KN50) to 75% (KN75) only reduced marginal dMPAB and dCVQ by about 0.03% (Tables 2 and 3). Moreover, marginal dMPAB and dCVQ were only reduced by about 0.06% with full knowledge of species (KN100; Tables 2 and 3). In short, knowing more species of a forest in a rapid assessment did not necessarily contribute to improving accuracy and consistency compared to the baseline model.

Results indicated significant marginal effects of rapid assessment effort (RE) and rapid assessment type (RT) on dMPAB and dCVQ consistently across sampling intensity ($p < 0.001$; Fig. 2). For example, at 5% sampling intensity, increasing RE from 10 (RE10) to 20 (RE20) individuals reduced marginal dMPAB and dCVQ by 0.41% and 0.55%, respectively (Tables 2 and 3). On the other hand, increasing RE from 10 (RE10) to 40 (RE40) individuals doubled the reduction in marginal dMPAB and dCVQ, i.e., by 0.91% and 1.17%, respectively (Tables 2 and 3). These results suggested that increasing the number of individuals for rapid assessment in a cell significantly improved accuracy and consistency in mapping species richness compared to the baseline model. On the contrary, applying the transect (RT_{tr}) in rapid assessment significantly increased dMPAB and dCVQ compared to the random walk (RT_{rw}) with the exception



Table 2. Estimated parameters of the fitted linear model for the effects of knowledge level (KN), rapid assessment effort (RS), and rapid assessment type (RT) on mean percent absolute bias (MPAB). Values in bold red indicates significant parameter estimates ($p \leq 0.05$). The results are for 5% to 40% sampling intensity.

Parameter	Sampling Intensity						
	5%	10%	15%	20%	25%	30%	40%
Intercept	-0.5628	-1.1758	-1.0777	-0.8186	-0.9639	-1.2457	-1.8366
BCI	0.6374	1.0536	0.8992	0.6941	0.6691	0.9573	1.6344
Danum	0.6374	1.0536	0.3488	0.6941	-0.2251	-0.3753	-0.4840
Fushan	-1.6378	-2.6803	-2.7740	-3.9128	-3.9246	-4.0120	-4.0994
Pasoh	0.9652	1.6238	1.4333	1.4195	1.3390	1.6613	2.3403
SCBI	-1.2783	-1.8817	-1.6753	-2.5293	-2.4315	-2.3930	-2.5690
Sinharaja	0.2910	0.4249	0.2655	-0.0088	-0.1048	0.1443	0.6610
Wanang	0.8574	1.3402	0.2655	0.9822	-0.1048	0.1443	1.6229
Wind River	-0.3267	-0.7217	-0.7424	-1.8265	-1.6545	-1.8973	-1.8803
KN75	-0.0301	-0.0117	-0.0008	-0.0420	0.0154	-0.0265	0.0144
KN100	-0.0604	0.0017	-0.0542	-0.0581	0.0175	0.0289	0.1166
RE20	-0.4096	-0.5734	-0.5824	-0.7592	-0.6063	-0.8227	-1.0377
RE40	-0.9102	-1.2141	-1.0985	-1.5624	-1.3743	-1.6845	-2.2103
RTtr	0.0188	0.1593	0.1653	0.1415	0.2580	0.3129	0.4773
KN75-RE20	0.0043	-0.0442	0.0029	0.0114	-0.0840	0.0623	-0.0210
KN100-RE20	0.0218	-0.0019	0.0528	0.0132	-0.0539	0.0067	-0.0921
KN75-RE40	0.0518	-0.0297	0.0023	0.0022	-0.0860	-0.0376	-0.1729
KN100-RE40	0.0267	0.0343	0.0379	-0.0211	-0.1473	-0.1591	-0.2104
KN75-RTtr	0.0451	0.0057	-0.0630	0.0483	-0.0216	0.0563	-0.0378
KN100-RTtr	0.0427	-0.0396	0.0356	0.0679	0.0220	0.0017	-0.1265
RE20-RTtr	0.0690	0.0628	0.0343	0.1251	-0.0582	0.0639	-0.0872
RE40-RTtr	0.0705	0.0534	0.0233	0.1165	0.0328	0.0392	0.0027
KN75-RE20-RTtr	-0.0243	0.0181	0.0538	0.0256	0.1783	-0.0970	0.0553
KN100-RE20-RTtr	0.0088	0.0145	-0.0524	-0.0308	0.0976	-0.0321	0.1535
KN75-RE40-RTtr	-0.0519	0.0052	0.0611	-0.0280	0.0520	-0.0569	-0.0639
KN100-RE40-RTtr	-0.0424	0.0153	-0.0055	0.0031	0.0424	0.0506	0.0446

Table 3. Estimated parameters of the fitted linear model for the effects of knowledge level (KN), rapid assessment effort (RS), and rapid assessment type (RT) on coefficient of variation of relative accuracy (CVQ). Values in bold red indicates significant parameter estimates ($p \leq 0.05$). The results are for 5% to 40% sampling intensity.

Parameter	Sampling Intensity						
	5%	10%	15%	20%	25%	30%	40%
Intercept	-1.5459	-3.2298	-2.8663	-3.5512	-3.8248	-4.7502	-6.0848
BCI	1.6998	3.1772	2.8160	3.4908	3.3359	4.1645	5.7551
Danum	1.6998	3.1772	-4.7549	3.4908	-9.3254	-13.6986	-19.6180
Fushan	-1.0457	-1.1314	-1.3995	-1.7994	-1.7579	-1.4961	-1.0015
Pasoh	2.0721	3.8159	3.3959	4.3476	4.1318	5.0326	6.6485
SCBI	-0.2257	0.3421	0.6675	0.9586	1.0140	1.6363	2.0834
Sinharaja	1.3335	2.5411	2.1316	2.6945	2.4760	3.1829	4.5185
Wanang	1.7846	3.0065	2.1316	3.1262	2.4760	3.1829	4.8540
Wind River	0.1500	0.5786	0.6498	-0.0982	0.0596	0.0491	0.1709
KN75	-0.0333	0.0407	0.0548	-0.0591	0.2632	0.1492	0.2526
KN100	-0.0564	-0.0363	-0.1458	-0.0331	0.1437	0.2682	0.3252
RE20	-0.5504	-0.8297	-0.8543	-1.0752	-0.7667	-1.0830	-1.3216
RE40	-1.1749	-1.4972	-1.4665	-2.0052	-1.5000	-1.6503	-2.8021
RTtr	0.0410	0.2022	0.2702	0.1899	0.3796	0.5427	0.8324
KN75-RE20	-0.0015	-0.0566	-0.1923	0.0303	-0.2616	0.0261	-0.3261
KN100-RE20	0.0098	0.1089	0.1994	0.0236	-0.1002	0.0444	-0.2236
KN75-RE40	0.0301	-0.1344	-0.0538	0.0795	-0.2818	-0.1657	-0.4635
KN100-RE40	0.0022	0.0192	0.1641	-0.0324	-0.2676	-0.4573	-0.1590
KN75-RTtr	0.0207	-0.0167	-0.2332	-0.0195	-0.0244	0.1345	-0.0786
KN100-RTtr	0.0319	-0.0256	0.0077	0.0455	0.0265	-0.0283	-0.4080
RE20-RTtr	0.0931	0.0814	0.0069	0.1721	-0.0621	-0.0117	-0.4696
RE40-RTtr	0.0607	0.0541	0.0066	0.1176	-0.0142	-0.0966	-0.0365
KN75-RE20-RTtr	-0.0025	0.0548	0.2620	0.0756	0.2717	-0.1002	0.5865
KN100-RE20-RTtr	-0.0236	-0.0150	0.0195	-0.0027	0.0794	-0.0352	0.8220
KN75-RE40-RTtr	0.0005	0.0300	0.1981	0.0695	0.1412	-0.1420	-0.0769
KN100-RE40-RTtr	-0.0067	0.0140	-0.0619	0.0617	0.0828	0.1972	0.0807

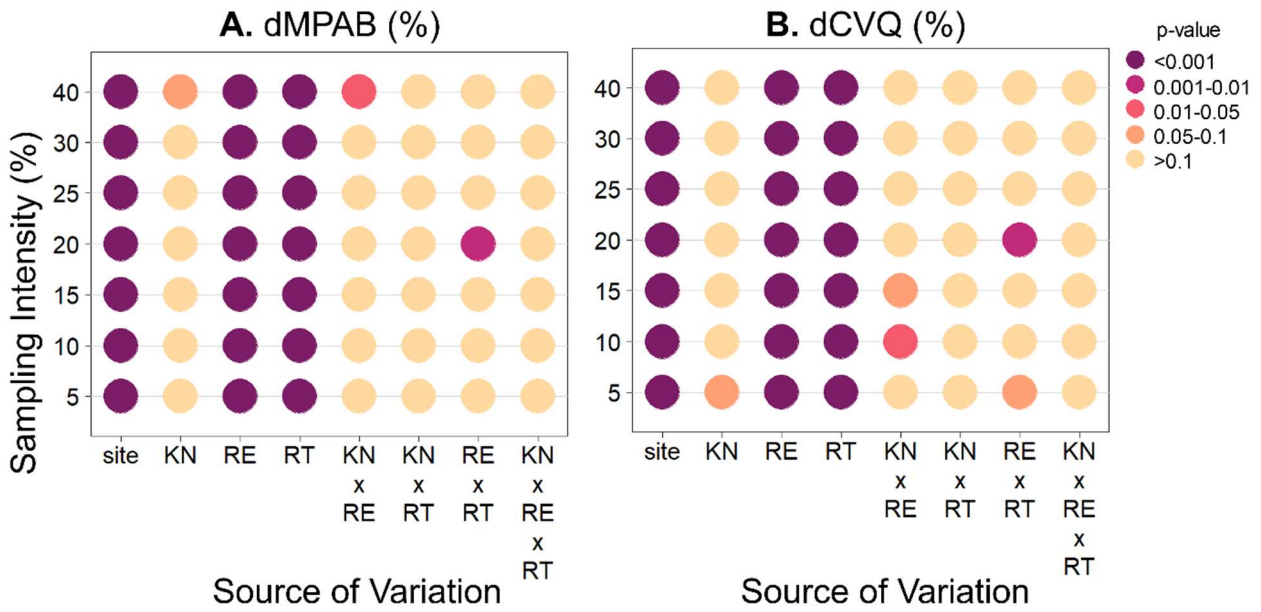


Fig. 2. Three-way Analysis of Variance (ANOVA) for **A.** mean percent absolute bias (MPAB) and **B.** coefficient of variation in relative accuracy (CVQ). Significance of sources of variation is presented across the seven sampling intensities. The factors are sites (site), knowledge level (KN), rapid assessment effort (RE), and rapid assessment type (RT). Significance is expressed as color coded p-values.

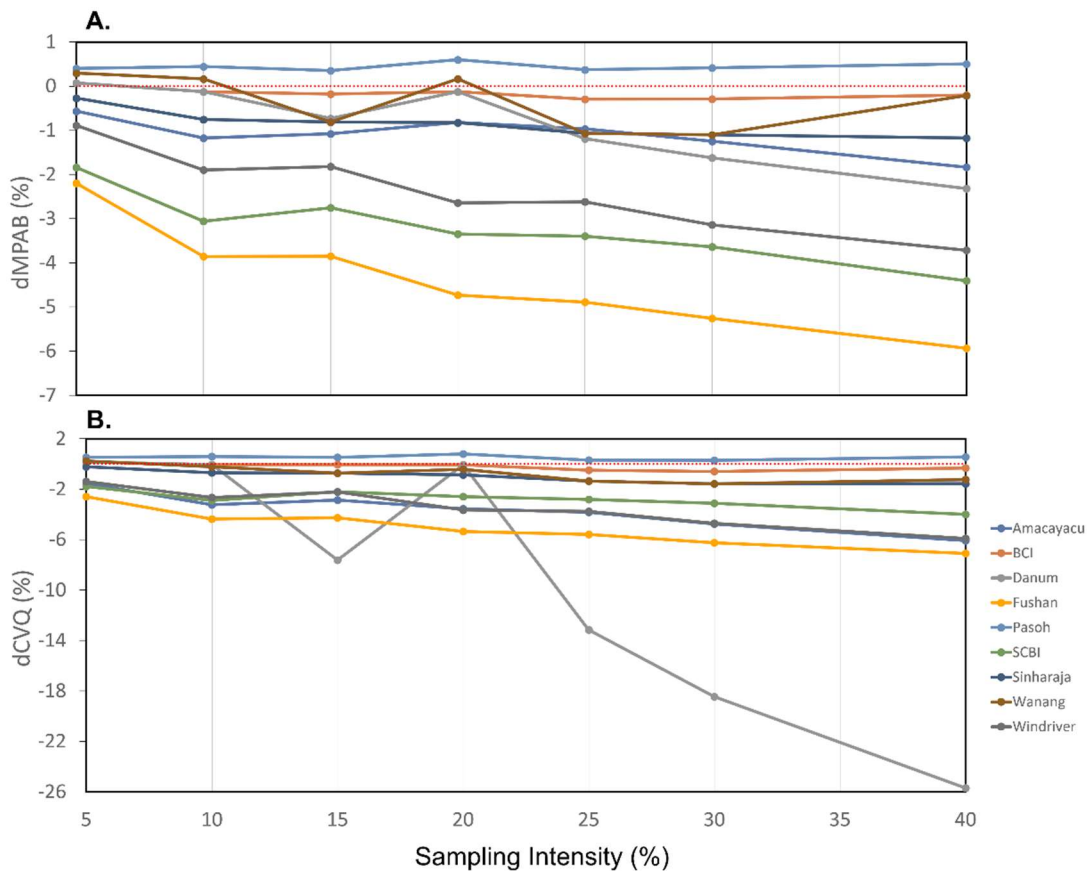


Fig. 3. Difference in **A.** mean percent absolute bias (dMPAB) and **B.** coefficient of variation in relative accuracy (dCVQ). The dMPAB and dCVQ are calculated between the rapid assessment combination KN50-RE10-RT_{rw} and the baseline across the seven sampling intensities and the nine sites. The combination KN50-RE10-RT_{rw} is rapid assessment with 50% knowledge level, 10 plant individuals, and random walk.



of 5% sampling intensity (Tables 2 and 3). For example, at 10% sampling intensity, applying RTtr would increase dMPAB and dCVQ by 0.16% and 0.20%, respectively (Tables 2 and 3). This suggested that laying out transect to select individuals in a cell during rapid assessment was less accurate and consistent in mapping species richness than randomly select individuals.

The three-way ANOVA analysis mostly suggested insignificant effects of two- and three-way interactions on dMPAB and dCVQ across sampling intensity ($p > 0.1$; Fig. 2). An exception was the RE \times RT interaction at 20% sampling intensity, which was significant for both dMPAB and dCVQ ($p = 0.005$; Fig. 2). Another exception was the KN \times RE interaction at 20% and 10% sampling intensity for dMPAB and dCVQ, respectively, but with only suggestive evidence of significance ($p = 0.01 - 0.05$; Fig. 2). The results in general suggested that the rapid assessment factors did not significantly interact with each other to influence accuracy and consistency in mapping species richness compared to the baseline model.

Increasing sampling intensity generally decreased dMPAB and dCVQ across different rapid assessment combinations (Fig. 4 and 5). For dMPAB, the rate of decrease appeared to be greater when sampling intensity increased from 5% to 10% but became more gradual from 10% to 40% sampling intensity. The trend in the rate of decrease was particularly evident for the rapid assessment combinations with RE10 (Fig. 4). For dCVQ, the rate of decrease across sampling intensity appeared to be similar between the rapid assessment combinations (Fig. 5). Despite the observed consistent trends in dMPAB and dCVQ, there were some exceptions. For example, dMPAB was higher than expected for 15% or 20% sampling intensity depending on the rapid assessment combinations (Fig. 4), while dCVQ was higher than expected at 20% sampling intensity for all the rapid assessment combinations (Fig. 5). As seen in Tables 2 and 3, increasing RE from 10 to 40 individuals decreased dMPAB and dCVQ. For dMPAB, the trends over sampling intensity were not parallel suggesting that the amount of decrease became larger at higher sampling intensity (Fig. 4). On the other hand, the parallel trends in dCVQ over sampling intensity suggested that the amount of decrease was consistent (Fig. 5). The parallel trends in dMPAB and dCVQ between RTTrw and RTtr combinations suggested that RTtr was consistently less accurate and consistent than RTTrw across sampling intensity (Fig. 4 and 5). In short, increasing sampling intensity improved accuracy and consistency in mapping species richness.

DISCUSSION

Remote sensing has contributed tremendously to mapping plant diversity for the past decades (Rocchini *et al.*, 2010). However, other alternative methodology to assist mapping could be explored. As such, this study

aims to explore how information from a rapid assessment by a local community or an expert could be used to locally map spatial distribution of species richness and to assess the methodology accuracy and consistency. Primary results suggest that simulated knowledge level of a community did not improve the accuracy and consistency in mapping species richness. On the other hand, both rapid assessment efforts and methods influenced the accuracy and consistency. Increasing rapid assessment efforts have shown to improve the accuracy and consistency. Using transect in rapid assessment is less accurate and consistent than the method of randomly selecting individuals in cells (random walk) for rapid assessment.

The proposed method of using information from a rapid assessment to map species richness is expected to perform equally well across all sites, but the results suggest otherwise. Some sites such as Fushan and Danum are more accurate and consistent than other sites such as Pasoh. We speculate the reason to be the underlying heterogeneity in the spatial distribution of species richness in a site (Fig. 6). For example, for Pasoh, majority of the cells are very similar in relative species richness, or in other words, the number of species is relatively homogeneous across cells (Fig. 6B). On the other hand, for sites such as Fushan and Danum, cells are highly variable in relative species richness (a range from light yellow to dark red; Fig. 6C and 6G), which implies that the number of species is highly heterogeneous across cells. Furthermore, there are aggregation of cells with low or high relative species richness. For example, there is a pocket of cells with low relative species richness at the lower right section in Danum (Fig. 6C). For Fushan, there is a large pocket of cells with high relative species richness diagonally across the site, which is surrounded by cells of low relative richness (Fig. 6G). Hence, we reason that for a homogenous site, information gain from a rapid assessment is very little, and information from census is sufficient to map species richness accurately and consistently. On the contrary, when species richness displays spatial clusters, rapid assessment contributes significantly to mapping the distribution. Hence, if prior knowledge of how species richness distributes spatially, it would be useful to designing a rapid assessment, but it is seldom known in advance. However, remote sensing could potentially be used to generate this prior knowledge with highly correlated proxies (Hernández-Stefanoni and Ponce-Hernandez, 2004).

The ability to identify a plant species should theoretically affect amount of information gained during a rapid assessment because unknown individuals do not contribute to the analysis. We expect that knowing 50% of the species in a forest should produce less accurate and consistent mapping compared to 100% knowledge. However, results contradict our expectation with knowledge level having minimum effect. Hence, full

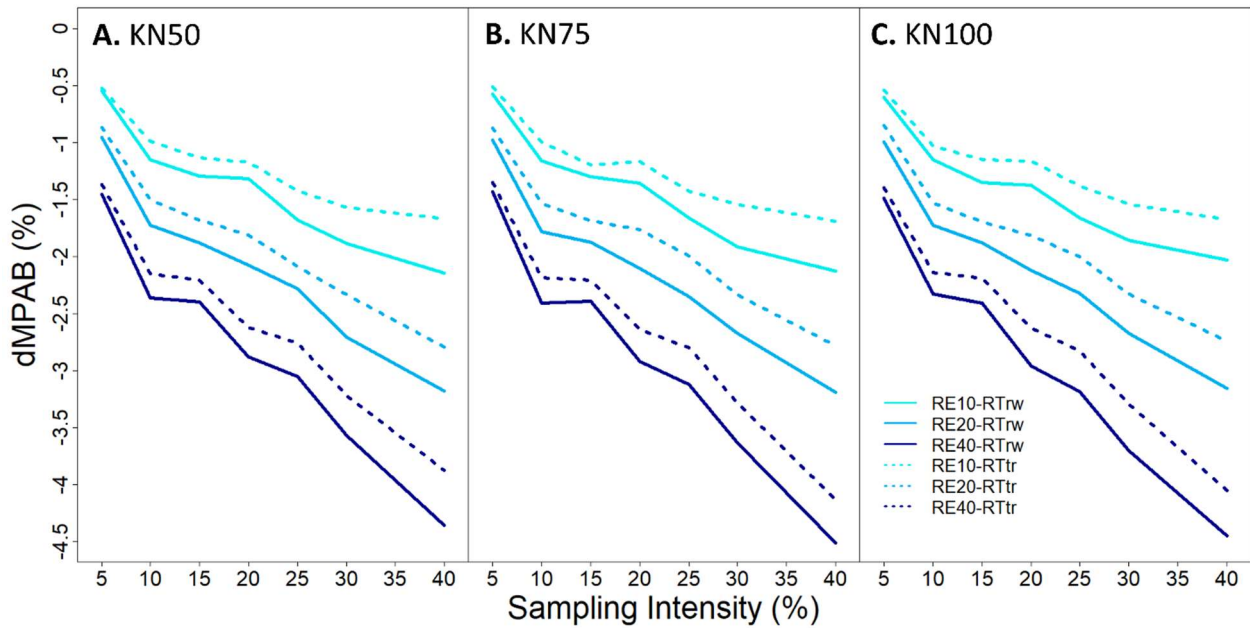


Fig. 4. Difference in mean percent absolute bias (dMPAB). The dMPAB is calculated between rapid assessment combinations and the baseline across sampling intensity averaged across the nine sites.

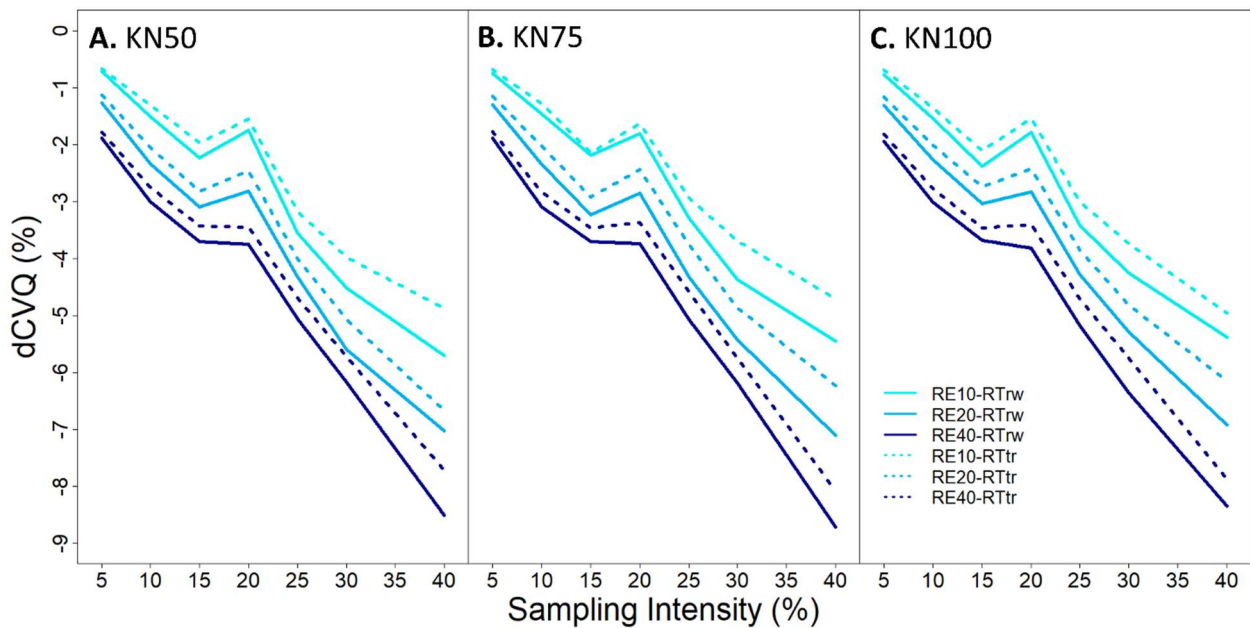


Fig. 5. Difference in coefficient of variation in relative accuracy (dCVQ). The dCVQ is calculated between rapid assessment combinations and the baseline across sampling intensity averaged across the nine sites.

knowledge is not a necessary requirement for a rapid assessment. Our study suggests at least knowing 50% of the locally abundant species. A possible explanation for the insignificant effect of knowledge level is that cells mostly consist of locally abundant plant species. Rare species with very few individuals are more sparsely distributed. These rare species might not be picked up during rapid assessment because only a small fraction of plant individuals is selected for the assessment. Thus, it is

more likely to select species that have moderate to high number of individuals in a forest, which coincides with the simulated knowledge levels in this study. In short, plant individuals selected for rapid assessment are more likely known to the survey crew because of abundance. Nonetheless, caution should be exercised with the interpretation of the results and suggestions. It is possible that some degree of correlation between knowledge levels and rapid assessment efforts exist because unidentified

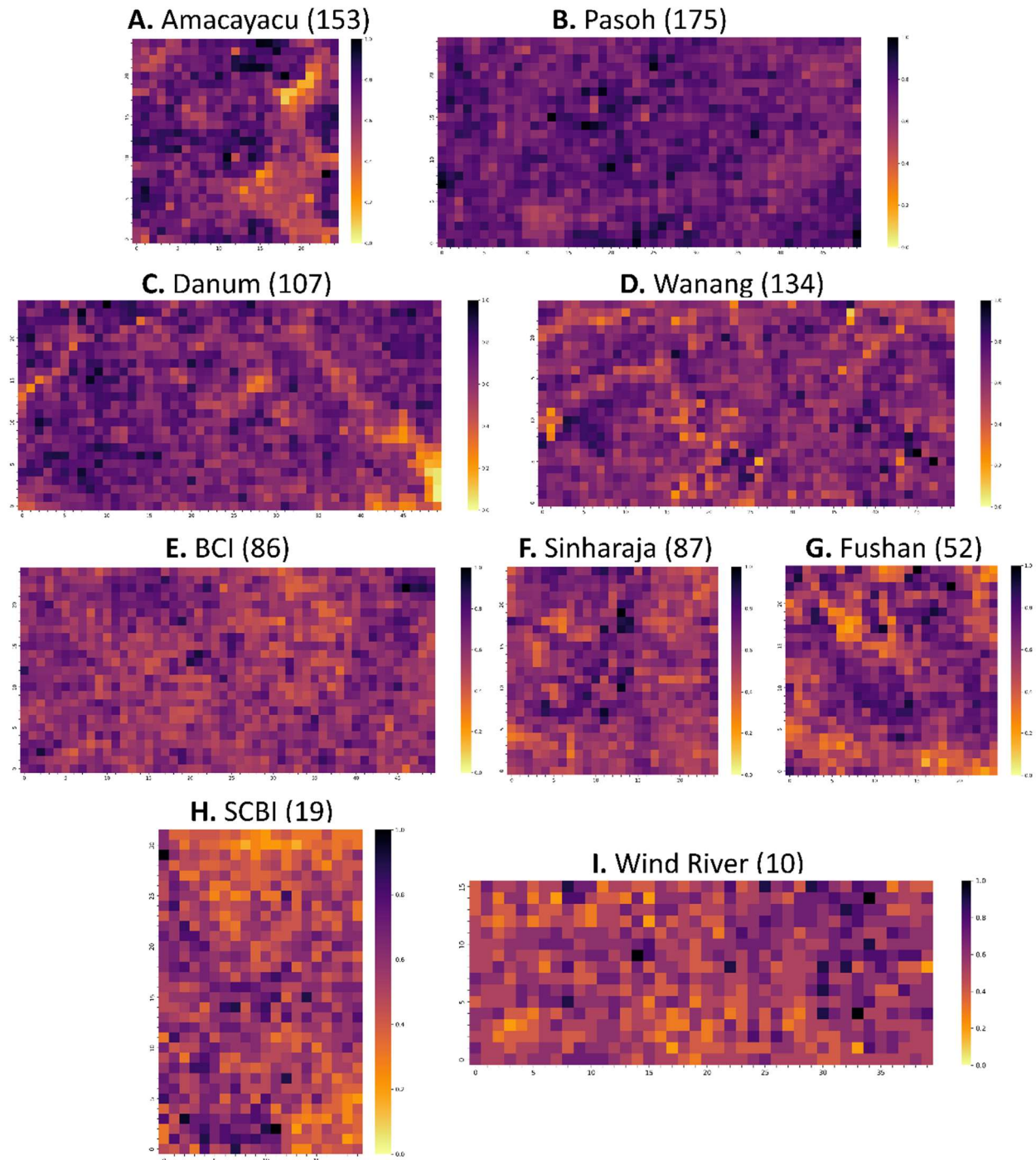


Fig. 6. Heat maps depicting spatial distribution of cell relative true species richness. The sites are: **A.** Amacayacu, **B.** Pasoh, **C.** Danum, **D.** Wanang, **E.** BCI, **F.** Sinharaja, **G.** Fushan, **H.** SCBI, and **I.** Wind River. Cell true species richness is defined as the total number of species found in a cell. Relative true species richness is the ratio of cell true species richness: maximum cell true species richness. Maximum cell true species richness of each site is given in parentheses.

individuals are dropped from modeling. The moderate to suggestive significance of $KN \times RE$ interaction may point towards this correlation. Thus, the significant effects of rapid assessment effort may have masked potential effects of knowledge level.

Nonetheless, the results on knowledge level support

engaging a local community in rapid assessment, who usually has only partial knowledge of the species in their forests. Traditional or local knowledge has contributed to conservation effort. For example, Wilder *et al.* (2016) showed that cooperation of local communities and biologists facilitated development of a biodiversity



inventory network to improve species recovery. Hernández-Stefanoni *et al.* (2006) found that indigenous Mayan knowledge was comparable to satellite imaging in assessing species diversity of a Mexican tropical forest. However, there are challenges related to the nomenclature used in a rapid assessment. For example, a species could have multiple names or multiple species could be assigned a single name (Lam and Kleinn, 2008). A possible solution is to first survey local knowledge and species identify before the start of a rapid assessment (Khasbagan and Soyolt, 2008).

Accuracy and consistency in mapping species richness are mainly driven by rapid assessment effort and type. In general, a large number of rapidly assessed individuals with the random walk method leads to more accurate and consistent mapping. The observed results with the rapid assessment effort agree with the commonly observed species-individual relationship. The relationship shows that the number of observed species increases with the number of sampled individuals up to a certain degree (Condit *et al.*, 1996). With more rapidly assessed individuals, true cell species richness is better captured. If a cell is inherently species rich, rapid species richness would be high, and vice versa. As a result, model prediction is more accurate and consistent. However, higher number of rapidly assessed individuals means higher cost and longer field time. Thus, a balance between accuracy and consistency and cost should be further studied. Our study suggests at least 20 individuals in a rapid assessment to gain at least some level of improvement over accuracy and consistency.

Better performance of the random walk (RTrw) than the transect (RTtr) could be explained by the former reducing spatial autocorrelation between sampled individuals. Some plant species are likely to exhibit local aggregation due to factors such as dispersal limitation (Seidler and Plotkin, 2006) and microclimates (Fayolle *et al.*, 2012). It is likely that RTrw reduces the chance of sampling individuals from the same species, which in turn better reflects underlying species richness in a cell. In two separate studies looking at various plot designs, Yang *et al.* (2019) and Quon *et al.* (2020) found that when plots that were in close proximity were more similar in species composition, which reduced information gain on species diversity. A possible explanation for the underperformance of RTtr could be the small cell size (20 × 20 m) that results in aggregating individuals from the same species. Increasing cell size would lead to more diverse species composition, a well-known species-area relationship (Condit *et al.*, 1996), which in turn could improve performances of RTtr. From a practical perspective, it would be difficult to implement RTrw because one would theoretically need to label every individual in a cell for a pure random selection. RTtr is easily implemented and has been widely carried out for sampling biodiversity such as the assessment of vascular

plant diversity in Switzerland (Wohlgemuth *et al.*, 2008). Thus, this study recommends RTtr despite it being less accurate and consistent. Nonetheless, simulating RTrw allows us to quantify the relative performances between the two methods.

A major cost of our proposed method comes from the census. Hence, the choice of a sampling intensity of census cells will directly impact the feasibility of a plant diversity assessment program. As expected, results suggest that increasing sampling intensity improves accuracy and consistency, but this will advertently increase the cost. The rate of decrease in accuracy and consistency over sampling intensity suggests that a sampling intensity between 5% and 10%. This is because improvement in the accuracy and consistency becomes more gradual at a higher sampling intensity, especially the former. However, it should be noted that this suggestion is observational. A more formal approach of studying the trade-off between cost, accuracy, and consistency is with a detail cost-plus-loss analysis (Lynch, 2017; Yang *et al.*, 2017). Better understanding of the trade-offs will help with designing rapid assessment and census activities. So far, this type of analysis has not been carried out in the context of diversity assessment.

A major limitation of our proposed method is spatial coverage. Mapping distribution of species richness with field rapid assessment could not match geographical coverage from remote sensing. However, it is possible to integrate remote sensing and our method in a hierarchical structure. For example, at the landscape-level, remote sensing is used to develop wall-to-wall proxies such as the ones mentioned earlier (Coops *et al.*, 2019; Pau *et al.*, 2012). The wall-to-wall coverage is then delineated into polygons of roughly homogenous values. Some of these polygons or local sites are then selected with a probability design for field rapid assessment and census as described in our method. Multilevel relationships between the wall-to-wall remote sensing proxies, the rapid assessment metrics, and the census information are modeled. The fitted models are used to predict species richness distribution across a larger geographical region. This hierarchical sampling strategy would resolve the limitation of our proposed method, and it could be studied in the future for its efficiency.

CONCLUSIONS

Conserving plant diversity is an integral part of sustainable forestry. Spatial distribution of species richness is used for management decision making such as designating conservation areas. This study develops a strategy of using rapid assessment to generate a complete coverage of proxies for the underlying species richness in a local forest. Census is carried out and linked to the rapid assessment proxies. An ANN model is built to predict the underlying distribution of species richness of the area.



Results show that information from a rapid assessment improves accuracy and consistency in mapping species richness distribution. This presents an opportunity to engage local communities in a rapid assessment with the assistance of an expert. There may be opportunities to adapt our methods to other vegetation groups such as herbs, epiphytes, and lianas, but there are foreseeable challenges such as quantifying the number of individuals. Furthermore, it is necessary to test our methods in the field in a future study to identify any practical issues. In conclusion, our study aims to diversify tools and to engage in multi-stakeholder solutions crucial to sustainable forest management because management objectives are increasingly diverse and complex to meet societal demands.

ACKNOWLEDGMENTS

We would like to thank two anonymous reviewers for their valuable comments that have significantly improved the manuscript. The funding of this project is provided by the National Science and Technology Council Taiwan (Grant no. MOST 111-2628-B-002-042 and MOST 111-2326-B-002-005-MY3). **Amacayacu:** The 25-ha Long-Term Ecological Research Project of Amacayacu is a collaborative project of the Instituto Amazónico de Investigaciones Científicas Sinchi and the Universidad Nacional de Colombia Sede Medellín, in partnership with the Unidad de Manejo Especial de Parques Naturales Nacionales and the Forest Global Earth Observatory of the Smithsonian Tropical Research Institute (ForestGEO). The Amacayacu Forest Dynamics Plot is part of ForestGEO a global network of large-scale demographic tree plots. We acknowledge the Director and staff of the Amacayacu National Park for supporting and maintaining the project in this National Park. **BCI:** The BCI forest dynamics research project was made possible by National Science Foundation grants to Stephen P. Hubbell: DEB-0640386, DEB-0425651, DEB-0346488, DEB-0129874, DEB-00753102, DEB-9909347, DEB-9615226, DEB-9615226, DEB-9405933, DEB-9221033, DEB-9100058, DEB-8906869, DEB-8605042, DEB-8206992, DEB-7922197, support from the Forest Global Earth Observatory, the Smithsonian Tropical Research Institute, the John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, the Small World Institute Fund, and numerous private individuals, and through the hard work of over 100 people from 10 countries over the past three decades. The plot project is part of the Forest Global Earth Observatory (ForestGEO), a global network of large-scale demographic tree plots. **Danum:** The Danum plot is a core project of the Southeast Asia Rain Forest Research Partnership (SEARRP). We thank SEARRP partners, especially Yayasan Sabah for their support, and HSBC Malaysia and the University of Zurich for funding. We are grateful to the research assistants who are conducting the census, in particular the team leader Alex Karolus, and to Mike Bernados and Bill McDonald for species identifications. We thank Stuart Davies and Shameema Esufali for advice and training. **Fushan:** The establishment and first census of the Fushan 25-ha plot is a collaborative project by the Taiwan Forestry Research Institute (data provider), the Taiwan Forestry Bureau, the National Taiwan University (Institute of Ecology and Evolutionary

Biology) and the ForestGEO (formerly the CTFS). We thank the agencies for providing the datasets and all field staff. S.H.S. also thanks Dr. I-Fang Sun for his long-term support for the Fushan project. **Pasoh:** Data from the Pasoh Research Forest was provided by the Forest Research Institute Malaysia-Forest Global Earth Observatory, Smithsonian Tropical Research Institute collaborative research project. Negeri Sembilan Forestry Department is the custodian of Pasoh Research Forest and I/we acknowledge the department for preserving the research forest. **SCBI:** Funding for the Smithsonian Conservation Biology Institute (SCBI) Large Forest Dynamics Plot (LFDP) was provided by the Smithsonian Institution, the National Zoological Park, and the HSBC Climate Partnership. **Sinharaja:** The 25-ha Long-Term Ecological Research Project at Sinharaja World Heritage Site is a collaborative project of the Uva Wellassa University, University of Peradeniya, the Forest Global Earth Observatory (ForestGEO) of the Smithsonian Tropical Research Institute, with supplementary funding received from the John D. and Catherine T. MacArthur Foundation, the National Institute for Environmental Science, Japan, and the Helmholtz Centre for Environmental Research-UFZ, Germany, for past censuses. The PIs gratefully acknowledge the Forest Department, Uva Wellassa University, and the Post-Graduate Institute of Science at the University of Peradeniya, Sri Lanka for supporting this project, and the local field and lab staff who tirelessly contributed in the repeated censuses of this plot. **Wanang:** The 50-ha Wanang Forest Dynamics Plot is a collaborative project of the New Guinea Binatang Research Center, the Forest Global Earth Observatory (ForestGEO) of the Smithsonian Tropical Research Institute, the Forest Research Institute of Papua New Guinea, the Czech Academy of Sciences (19-28126X) and the University of Minnesota supported by NSF DEB- 1027297 and NIH ICBG 5U01TW006671. We acknowledge the government of Papua New Guinea and the customary landowners of Wanang for supporting and maintaining the plot. **Wind River:** The Wind River Forest Dynamics Plot is a collaborative project of Utah State University and the USDA Forest Service Pacific Northwest Research Station. Funding has been provided by the Smithsonian ForestGEO, Utah State University, the Utah Agricultural Experiment Station (projects 1153, 1398, and 1423), the National Science Foundation (DEB #1542681), and private donors. We acknowledge the Gifford Pinchot National Forest and the U.S. Forest Service Wind River Field Station for providing logistical support, and the students, volunteers and staff individually listed at <http://wfdp.org> for data collection. Research was performed under long-term research permits issued by the USFS with validity from 2010 – 2040.

LITERATURE CITED

Anderson-Teixeira, K.J., Davies, S.J., Bennett, A.C., Gonzalez-Akre, E.B., Muller-Landau, H.C., Wright, S.J., Salim, K.A., Zambrano, A.M.A., Alonso, A., Baltzer, J.L., Basset, Y., Bourg, N.A., Broadbent, E.N., Brockelman, W.Y., Bunyavejchewin, S., Burslem, D.F.R.P., Butt, N., Cao, M., Cardenas, D., Chuyong, G.B., Clay, K., Cordell, S., Dattaraja, H.S., Deng, X., Detto, M., Du, X., Duque, A., Erikson, D.L., Ewango, C.E.N., Fischer, G.A., Fletcher, C., Foster, R.B., Giardina, C.P., Gilbert, G.S., Gunatilleke, N., Gunatilleke, S., Hao, Z., Hargrove, W.W., Hart, T.B.,



- Hau, B.C.H., He, F., Hoffman, F.M., Howe, R.W., Hubbell, S.P., Inman-Narahari, F.M., Jansen, P.A., Jiang, M., Johnson, D.J., Kanzaki, M., Kassim, A.R., Kenfack, D., Kibet, S., Kinnaird, M.F., Korte, L., Kral, K., Kumar, J., Larson, A.J., Li, Y., Li, X., Liu, S., Lum, S.K.Y., Lutz, J.A., Ma, K., Maddalena, D.M., Makana, J.-R., Malhi, Y., Marthews, T., Serudin, R.M., McMahon, S.M., McShea, W.J., Memiaghe, H.R., Mi, X., Mizuno, T., Morecroft, M., Myers, J.A., Novotny, V., Oliveira, A.A. de, Ong, P.S., Orwig, D.A., Ostertag, R., Ouden, J. den, Parker, G.G., Phillips, R.P., Sack, L., Sainge, M.N., Sang, W., Sri-ngernyuang, K., Sukumar, R., Sun, I.-F., Sungpalee, W., Suresh, H.S., Tan, S., Thomas, S.C., Thomas, D.W., Thompson, J., Turner, B. L., Uriarte, M., Valencia, R., Vallejo, M.I., Vicentini, A., Vrška, T., Wang, X., Wang, X., Weiblen, G., Wolf, A., Xu, H., Yap, S., Zimmerman, J. 2015 CTFS-ForestGEO: a worldwide network monitoring forests in an era of global change. *Glob. Ecol. Biol.* **21**(2): 528–549.
- Baskerville, G.L. 1986 Understanding Forest Management. *For. Chron.* **62**(4): 339–347.
- Bourg, N.A., McShea, W.J., Thompson, J.R., McGarvey, J.C., Shen, X. 2013 Initial census, woody seedling, seed rain, and stand structure data for the SCBI SIGEO Large Forest Dynamics Plot. *Ecology* **94**(9): 2111–2112.
- Chong, G.W., Reich, R.M., Kalkhan, M. A., Stohlgren, T.J. 2001 New approaches for sampling and modeling native and exotic plant species richness. *West. N. Am. Nat.* **61**(3): 328–335.
- Condit, R., Hubbell, S.P., Lafrankie, J.V., Sukumar, R., Manokaran, N., Foster, R.B., Ashton, P.S. 1996 Species-area and species-individual relationships for tropical trees: a comparison of three 50-ha plots. *J. Ecol.* **84**(4): 549–562.
- Condit, R., Pérez, R., Aguilar, S., Lao, S. 2019a Census data from 65 tree plots in Panama, 1994–2015. DataONE, Dataset.
- Condit, R., Pérez, R., Aguilar, S., Lao, S., Foster, R., Hubbell, S. P. 2019b BCI 50-ha plot taxonomy, 2019 version. Dryad, Dataset.
- Condit, R., Pérez, R., Aguilar, S., Lao, S., Foster, R., Hubbell, S.P. 2019c Complete data from the Barro Colorado 50-ha plot: 423617 trees, 35 years, 2019 version. Dryad, Dataset.
- Coops, N.C., Bolton, D.K., Hobi, M.L., Radeloff, V.C. 2019 Untangling multiple species richness hypothesis globally using remote sensing habitat indices. *Ecol. Indic.* **107**: 105567.
- Duque, A., Muller-Landau, H.C., Valencia, R., Cardenas, D., Davies, S., de Oliveira, A., Pérez, Á.J., Romero-Saltos, H., Vicentini, A. 2017 Insights into regional patterns of Amazonian forest structure, diversity, and dominance from three large terra-firme forest dynamics plots. *Biodivers. Conserv.* **26**(3): 669–686.
- Fayolle, A., Engelbrecht, B., Freycon, V., Mortier, F., Swaine, M., Réjou-Méchain, M., Doucet, J.-L., Fauvet, N., Cornu, G., Gourlet-Fleury, S. 2012 Geological substrates shape tree species and trait distributions in African moist forests. *PLOS ONE* **7**(8): e42381.
- FSC. 2012 FSC Principles and Criteria for Forest Stewardship. FSC-STD-01-001 V5-0 EN, Forest Stewardship Council.
- Gascon, C., Brooks, T.M., Contreras-MacBeath, T., Heard, N., Konstant, W., Lamoreux, J., Launay, F., Maunder, M., Mittermeier, R.A., Molur, S., Al Mubarak, R.K., Parr, M.J., Rhodin, A.G.J., Rylands, A.B., Soorae, P., Sanderson, J.G., Vié, J. -C. 2015 The importance and benefits of species. *Curr. Biol.* **25**(10): R431–R438.
- Haas, P.J., Liu, Y., Stokes, L. 2006 An estimator of number of species from quadrat sampling. *Biometrics* **62**(1): 135–141.
- Hall, J.P. 2001 Criteria and indicators of sustainable forest management. *Environ. Monit. Assess.* **67**(1): 109–119.
- Hernández-Stefanoni, J.L., Ponce-Hernandez, P. 2004 Mapping the spatial distribution of plant diversity indices in a tropical forest using multi-spectral satellite image classification and field measurements. *Biodivers. Conserv.* **13**(14): 2599–2621.
- Hernández-Stefanoni, J.L., Pineda, J.B., Valdes-Valadez, G. 2006 Comparing the use of indigenous knowledge with classification and ordination techniques for assessing the species composition and structure of vegetation in a tropical forest. *Environ. Manage.* **37**(5): 686–702.
- Hubbell, S.P., Foster, R.B., O'Brien, S.T., Harms, K.E., Condit, R., Wechsler, B., Wright, S.J., Loo de Lao, S. 1999 Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science* **283**(5401): 554–557.
- Huber, P.J. 1964 Robust estimation of a location parameter. *Ann. Math. Statist.* **35**(1): 73–101.
- Humphreys, A.M., Govaerts, R., Ficinski, S.Z., Lughadha, E.N., Vorontsova, M.S. 2019 Global dataset shows geography and life form predict modern plant extinction and rediscovery. *Nat. Ecol. Evol.* **3**(7): 1043–1047.
- Khasbagan, Soyolt. 2008 Indigenous knowledge for plant species diversity: a case study of wild plants' folk names used by the Mongolians in Ejina desert area, Inner Mongolia, P. R. China. *J. Ethnobiol. Ethnomedicine* **4**(1): 2.
- Kingma, D.P., Ba, J. 2015 Adam: A Method for Stochastic Optimization. In: Bengio, Y., LeCun, Y. (eds), 3rd International Conference on Learning Representations ICLR 2015, 7-9 May 2015. San Diego, CA, USA, 1–15.
- Lam, T.Y., Hsu, Y.-H., Yang, T.-R., Kershaw, J.A., Su, S.-H. 2018 Sampling with probability proportional to prediction: rethinking rapid plant diversity assessment. *Forestry* **91**(1): 17–26.
- Lam, T.Y., Kleinn, C. 2008 Estimation of tree species richness from large area forest inventory data: Evaluation and comparison of jackknife estimators. *For. Ecol. Manag.* **255**(3–4): 1002–1010.
- Lutz, J.A., Larson, A.J., Freund, J.A., Swanson, M.E., Bible, K.J. 2013 The Importance of large-diameter trees to forest structural heterogeneity. *PLOS ONE* **8**(12): e82784.
- Lutz, J.A., Larson, A.J., Furniss, T.J., Donato, D.C., Freund, J.A., Swanson, M.E., Bible, K.J., Chen, J., Franklin, J.F. 2014 Spatially nonrandom tree mortality and ingrowth maintain equilibrium pattern in an old-growth Pseudotsuga–Tsuga forest. *Ecology* **95**(8): 2047–2054.
- Lynch, T.B. 2017 Optimal sample size and plot size or point sampling factor based on cost-plus-loss using the Fairfield Smith relationship for plot size. *Forestry* **90**(5): 697–709.
- Manokaran, N., LaFrankie, J.V. 1990 Stand structure of Pasoh Forest Reserve, a lowland rain forest in Peninsular Malaysia. *J. Trop. For. Sci.* **3**(1): 14–24.
- Misra, D. 2020. Mish: A self regularized non-monotonic neural activation function. In *BMVC 2020*. Virtual, pp. 1–14.
- O'Brien, M.J., Hector, A., Kellenberger, R.T., Maycock, C.R., Ong, R., Philipson, C.D., Powers, J.S., Reynolds,



- G., Burslem, D.F.R.P.** 2022 Demographic consequences of heterogeneity in conspecific density dependence among mast-fruited tropical trees. *Proc. R. Soc. B.* **289(1977)**: 20220739.
- Parrotta, J., Yeo-Chang, Y., Camacho, L.D.** 2016 Traditional knowledge for sustainable forest management and provision of ecosystem services. *Int. J. Biodivers. Sci. Ecosyst. Serv. Manag.* **12(1–2)**, 1–4.
- Pau, S., Gillespie, W., Wolkovich, E.M.** 2012 Dissecting NDVI–species richness relationships in Hawaiian dry forests. *J. Biogeogr.* **39(9)**: 1678–1686.
- Pearson, D.L., Carroll, S.S.** 1998 Global patterns of species richness: spatial models for conservation planning using bioindicator and precipitation data. *Conserv. Biol.* **12(4)**: 809–821.
- Pimm, S.L., Raven, P.** 2000 Extinction by numbers. *Nature* **403(6772)**: 843–845.
- Pool-Stanvliet, R., Stoll-Kleemann, S., Giliomee, J.H.** 2018 Criteria for selection and evaluation of biosphere reserves in support of the UNESCO MAB programme in South Africa. *Land Use Policy* **76**: 654–663.
- Quon, C., Lam, T.Y., Lin, H.-T.** 2020 Designing cluster plots for sampling local plant species composition for biodiversity management. *Forest Syst.* **29(1)**: e002.
- Rocchini, D., Balkenhol, N., Carter, G.A., Foody, G.M., Gillespie, T.W., He, K.S., Kark, S., Levin, N., Lucas, K., Luoto, M., Nagendra, H., Oldeland, J., Ricotta, C., Southworth, J., Neteler, M.** 2010 Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges. *Ecol. Inform.* **5(5)**: 318–329.
- Rocchini, D., Boyd, D.S., Féret, J.-B., Foody, G.M., He, K.S., Lausch, A., Nagendra, H., Wegmann, M., Pettorelli, N.** 2016 Satellite remote sensing to monitor species diversity: potential and pitfalls. *Remote. Sens. Ecol. Conserv.* **2(1)**: 25–36.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J.** 1986 Learning representations by back-propagating errors. *Nature* **323(6088)**: 533–536.
- Sample, V.A.** 2005 Sustainable Forestry and Biodiversity Conservation. *J. Sustain. Forest.* **21(4)**: 137–150.
- Seidler, T. G., Plotkin, J. B.** 2006. Seed dispersal and spatial pattern in tropical trees. *PLOS Biology* **4(11 (e344))**: 2132–2137.
- Seymour, R.S., Hunter, Jr., M.L.** 1992 New forestry in eastern spruce-fir forests: Principles and applications to Maine. Maine Agricultural and Forest Experiment Station Miscellaneous Publication 716, Maine Agricultural and Forest Experiment Station. Orono, Maine, USA, p. 36.
- Shannon, C.E.** 1948 A mathematical theory of communication. *Bell Syst. Tech. J.* **27(3)**: 379–423.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.** 2014 Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15(56)**: 1929–1958.
- Su, S.-H., Chang-Yang, C.-H., Lu, C.-L., Tsui, C.-C., Lin, T.-T., Lin, C.-L., Chiou, W.-L., Kuan, L.-H., Chen, Z.-S., Hsieh, C.-F.** 2007 Fushan subtropical forest dynamics plot: tree species characteristics and distribution patterns. 1 edn. Taiwan Forestry Research Institute, 272 pp.
- Thapa, B., Sinclair, F.L., Walker, D.H.** 1995 Incorporation of indigenous knowledge and perspectives in agroforestry development. *Agroforest Syst.* **30(1)**: 249–261.
- Villero, D., Pla, M., Camps, D., Ruiz-Olmo, J., Brotons, L.** 2017 Integrating species distribution modelling into decision-making to inform conservation actions. *Biodivers. Conserv.* **26(2)**: 251–271.
- Vincent, J.B., Henning, B., Saulei, S., Sosanika, G., Weiblen, G.D.** 2015 Forest carbon in lowland Papua New Guinea: Local variation and the importance of small trees. *Austral Ecol.* **40(2)**: 151–159.
- Wilder, B.T., O'Meara, C., Monti, L., Nabhan, G.P.** 2016 The importance of indigenous knowledge in curbing the loss of language and biodiversity. *BioScience* **66(6)**: 499–509.
- Winter, S., Chirici, G., McRoberts, R.E., Hauk, E., Tomppo, E.** 2008 Possibilities for harmonizing national forest inventory data for use in forest biodiversity assessments. *Forestry* **81(1)**: 33–44.
- Wohlgemuth, T., Nobis, M.P., Kienast, F., Plattner, M.** 2008 Modelling vascular plant diversity at the landscape scale using systematic samples. *J. Biogeogr.* **35(7)**: 1226–1240.
- Yang, T.-R., Hsu, Y.-H., Kershaw, J. A., McGarrigle, E., Kilham, D.** 2017 Big BAF sampling in mixed species forest structures of northeastern North America: influence of count and measure BAF under cost constraints. *Forestry* **90(5)**: 649–660.
- Yang, T.-R., Lam, T.Y., Su, S.-H.** 2019 A simulation study on the effects of plot size and shape on sampling plant species composition for biodiversity management. *J. Sustain. Forest.* **38(2)**: 116–129.
- Zuleta, D., Russo, S.E., Barona, A., Barreto-Silva, J.S., Cardenas, D., Castaño, N., Davies, S.J., Detto, M., Sua, S., Turner, B.L., Duque, A.** 2020 Importance of topography for tree species habitat distributions in a terra firme forest in the Colombian Amazon. *Plant Soil* **450(1)**: 133–149.

Supplementary materials are available from Journal Website