

1 Testing the effectiveness of the Forest Integrity 2 Assessment: a field-based tool for estimating the 3 condition of tropical forest

4
5 Andrew J. Suggitt^{1,2,*}, Kok Loong Yeong^{3,4}, Anders Lindhe⁵, Agnes Agama⁴, Keith C. Hamer⁶, Glen
6 Reynolds⁴, Jane K. Hill² & Jennifer M. Lucey^{7,*}.

7
8 * To whom correspondence should be addressed. Emails: jennifer.lucey@zoo.ox.ac.uk (J.M.L.),
9 andrew.suggitt@northumbria.ac.uk (A.J.S.).

10
11 1 Department of Geography and Environmental Sciences, Northumbria University, Newcastle-upon-Tyne,
12 NE1 8ST, UK.

13 2 Leverhulme Centre for Anthropocene Biodiversity, Department of Biology, University of York,
14 Wentworth Way, York, YO10 5DD, UK.

15 3 Leverhulme Centre for Climate Change Mitigation (LC3M), University of Sheffield, Department of
16 Animal and Plant Sciences, Western Bank, Sheffield, S10 2TN, UK.

17 4 South East Asia Rainforest Research Partnership, Danum Valley Field Centre, 91112, Lahad Datu, Sabah,
18 Malaysia.

19 5 High Conservation Value Resource Network, Frewin Court, Oxford, OX1 3HZ, UK.

20 6 School of Biology, University of Leeds, Leeds, LS2 9JT, UK.

21 7 Department of Zoology, University of Oxford, Mansfield Road, Oxford, OX1 3SZ, UK.

22
23 **Running title:** Testing a rapid field-based forest condition assessment tool.

24 **Article type:** Research article.

25 **Keywords:** Rapid assessment, tropical ecology, human-modified, forest set-aside, High Carbon Stock,
26 ecological integrity, forest quality.

27 ABSTRACT

- 28 1. Global targets to halt biodiversity losses and mitigate climate change will require protecting rainforest
29 beyond current protected area networks, necessitating responsible forest stewardship from a diverse range
30 of companies, communities and private individuals. Robust assessments of forest condition are critical for
31 successful forest management, but many existing techniques are highly technical, time-consuming,
32 expensive, or require specialist knowledge.
- 33 2. To make assessment of tropical forests accessible to a wide range of actors, many of whom may be
34 limited by resources or expertise, the High Conservation Value Resource Network (HCVRN), with the
35 SE Asia Rainforest Research Partnership (SEARRP), developed a South East Asian version of the Forest
36 Integrity Assessment (FIA) tool as a rapid (< 1 hour) method of assessing forest condition in the field,
37 where non-experts respond to 50 questions about characteristics of the local environment while walking a
38 site transect. Here, we examined the effectiveness of this survey tool by conducting ~ 1,000 assessments
39 of forest condition at 16 tropical rainforest sites with varying levels of disturbance in Sabah, Malaysian
40 Borneo.
- 41 3. We found good agreement (R-squared range: 0.50 – 0.78) between FIA survey scores and independent
42 measures of forest condition, including biodiversity, vegetation structure, aboveground carbon, and other
43 key metrics of ecosystem function, indicating that the tool performed well. Although there was variation
44 among assessor responses when surveying the same forest sites, assessors were consistent in their ranking
45 of those sites, and prior forest knowledge had a minimal effect on the FIA scores. Revisions or further
46 training for questions where assessors disagree, for example on the presence of fauna at a site, could
47 improve consistency.
- 48 4. We conclude that the FIA survey tool is a robust method of assessing forest condition, providing a rapid
49 and accessible means of forest conservation assessment. The FIA tool could be incorporated into
50 management practices in a wide range of forest conservation schemes, from sustainability standards, to
51 community forestry and restoration initiatives. The tool will enable more organisations and individuals to
52 understand the conservation value of the forests they manage, and to identify areas for targeted
53 improvements.

54 1 INTRODUCTION

55 Globally, forests are at increasing risk of degradation or conversion to agriculture as the need for food and
56 other resources continues to rise (Hosonuma et al., 2012). Yet, these complex ecosystems support high levels
57 of biodiversity, harbour rare, threatened and endangered species, store and sequester large amounts of
58 carbon, regulate local and global climate systems, and maintain soil, hydrological and other ecosystem
59 services (Watson et al., 2018). Provision of ecosystem services is at its highest where forests are in the best
60 condition, by which we mean that they more closely resemble intact, or primary, habitat (e.g. in terms of
61 carbon: Wang et al., 2001; biodiversity: Tawatao et al., 2014; water quality: Luke et al., 2017). Continued
62 assessment and monitoring of forests, coupled with the ongoing maintenance and enhancement of forests to
63 improve their condition, are therefore essential if global targets on biodiversity and climate mitigation are to
64 be met.

65 To curb continued deforestation and degradation, conservation initiatives incentivise companies,
66 communities and individuals to manage forest areas outside of protected areas. Industry-based certification
67 standards, such as the Roundtable on Sustainable Palm Oil (RSPO) and the Forest Stewardship Council
68 (FSC), require companies to set aside and manage natural forests within their management units. The High
69 Conservation Value (HCV) approach has been widely adopted by certification standards such as the RSPO,
70 and by corporations aiming to conserve biodiversity and meet avoided deforestation commitments
71 (www.hcvnetwork.org), while additional areas are now being set aside by the palm oil, pulp and paper, and
72 cocoa sectors, under the High Carbon Stock (HCS) Approach (www.highcarbonstock.org). Community
73 forest stewardship, for example via REDD+ (Reducing Emissions from Deforestation and Forest
74 Degradation) or ecotourism-based forest restoration schemes, is also becoming a frequent means of
75 achieving forest conservation while also benefiting local livelihoods (Holck et al., 2008; Kunjuraman &
76 Aziz, 2019). The widespread adoption of these schemes creates new opportunities for increased forest
77 protection and improved management, but constraints may exist with respect to the ongoing stewardship of
78 these vital ecosystems – which, in some cases, falls to institutions or individuals that have limited capacity in
79 forest and conservation management.

80 Existing forest monitoring techniques often require a high level of technical knowledge, and can be both
81 time-consuming and expensive (Gibbs et al., 2007). Over recent years interest has grown in remote sensing
82 (e.g. by using satellites or drones) as a means of delivering forest monitoring (Finer et al., 2018). Whilst
83 these techniques provide information about forest structure and biomass over large areas, they also require
84 technical expertise, and are thus often inaccessible to smaller operations or local communities. Data derived
85 from remotely-sensed imagery can also mask important sub-canopy aspects of conservation value and
86 disturbance (such as hunting or the presence/absence of endangered species, Green et al., 2019). Existing
87 community-based forest monitoring approaches focus largely on counting and measuring trees. While
88 relatively simple, these approaches require a degree of expert knowledge, are time-consuming, and
89 conservation managers can find it challenging to translate outputs into information of practical relevance
90 (Holck et al., 2008). Other tools such as SMART (<https://smartconservationtools.org/>), focus on monitoring
91 of threats to wildlife. SMART relies on patrols, and has been shown to be effective in enhancing protection
92 of endangered species, but is contingent on continuous on-the-ground patrolling, has far less emphasis on
93 wider ecosystem quality, and is only an option for well-funded and staffed conservation programmes (Hoette
94 et al., 2016; Critchlow et al., 2017). Many of these techniques focus on individual aspects of the forest
95 ecosystem, such as forest structure or specific species, and – as a consequence - encourage forest managers
96 to maintain a narrow focus, rather than a broader view of the whole ecosystem. Ecological integrity
97 assessment methods recognise the need to understand the multiple interacting characteristics that contribute
98 to ecosystem functioning and thus the provisioning of key ecosystem services (Tierney et al., 2009;
99 Wurtzebach & Schultz, 2016). However, these processes can be complex to assess and monitor, thus a more
100 simple, low-cost and rapid forest assessment technique is needed to assess forest condition.

101 The Forest Integrity Assessment (FIA) tool assesses multiple facets of forest condition, while also addressing
102 issues of time, resource and the need for technical expertise. The tool was developed by the HCVRN, in
103 partnership with SEARRP in the Malaysian context, ([https://hcvnetwork.org/library/forest-integrity-
104 assessment-tool/](https://hcvnetwork.org/library/forest-integrity-assessment-tool/)) as a rapid (< 1 hour to complete) means of conducting broad assessments of forest
105 condition via a cheap and efficient approach that does not require expert knowledge or extensive resources.
106 Until now, however, the robustness of scores generated by the tool has not been tested. To enable such a test,
107 we first conducted a large-scale field trial of the survey tool, completing 967 assessor surveys across 16 sites

108 in the dipterocarp rainforests of Sabah, Malaysian Borneo. We use this trial data to test three key
109 performance aspects of the tool relevant to its wider application: 1) how well the scores derived from the FIA
110 survey tool correlate with independent metrics of forest condition, 2) if the characteristics or prior knowledge
111 of assessors affect their scoring, and 3) if the survey is efficient as it can be, i.e. do all questions within the
112 survey tool discriminate between sites in good or poor condition effectively. We conclude by discussing
113 where and how the FIA tool might be deployed as a rapid and low-cost means of assessing forest condition.

114 2 MATERIALS AND METHODS

115 2.1 The survey tool

116 The FIA survey tool has been adapted for different forest types globally (Lindhe & Drakenberg, 2019) and
117 we tested the version developed for lowland dipterocarp rainforest (Lindhe et al., 2015). Until recent
118 clearances, Lowland dipterocarp rainforest was the dominant forest type across South East Asia, and it
119 represents the modal pre-clearance forest type in areas managed by RSPO member oil palm companies.
120 Dipterocarp forest has also been surveyed by a substantial number of independent research projects, and we
121 utilise the resulting datasets in this study (Table 1, see also Methods 2.3 below).

122 In the FIA questionnaire (hereafter referred to as ‘the survey’, Supplementary Info 1), assessors respond to a
123 series of 50 ‘yes’ or ‘no’ questions regarding the site they are surveying, and the final score is calculated by
124 finding the total number of ‘yes’ responses. The survey targets a range of seven criteria known to be
125 associated with forest condition, namely: landscape, topography, water, trees, flora, fauna and disturbance.
126 Features that are especially important for indicating forest condition, such as the occurrence of large trees,
127 are allocated more questions to provide greater weighting for these important aspects, and all questions are
128 designed to be insensitive to the time of day or year (i.e. wet/dry season, Lindhe & Drackenberg, 2019).

129 2.2 Collecting survey test data

130 A total of 62 people was recruited to test the survey. Among this group we recruited a range of levels of
131 expertise in forestry and conservation to test whether survey results were affected by prior knowledge,
132 experience, education or other characteristics. We recruited 14 oil palm plantation staff from estates owned
133 by Wilmar International (RSPO member since 2005), a total of 19 first year undergraduate conservation

134 biology students from Universiti Malaysia Sabah, and 17 field research assistants and 12 graduates and post-
135 graduates affiliated with SEARRP. Information on age, gender, nationality, educational background,
136 profession and prior knowledge or experience of tropical forests was obtained. Assessors could choose to
137 remain anonymous, leave the survey process at any time, or refrain from answering any of the questions if
138 they wished, in line with the ethical approval granted for the study.

139 Assessors conducted surveys at 16 forest sites in Sabah, Malaysian Borneo (Supplementary Table 1,
140 Supplementary Fig. 1), varying in size and disturbance level from continuous tracts of primary forest in
141 Danum Valley Conservation Area to small degraded and previously logged forest fragments within oil palm
142 plantations (fragment size ranged from 12ha to 120ha). These sites were chosen because: a) they represented
143 a gradient of forest degradation on similar soil types, topography and elevation typical of the region, and b)
144 they fell within the spatial coverage of datasets quantifying forest integrity (see below), to enable comparison
145 with independent measures of forest condition at survey sites. Wherever possible we used the same transect
146 or trail as that used by the original study. On the day of the survey, each assessor was given a set of
147 guidelines to study for 30 minutes (Supplementary Info 2). One of the authors (K.L.Y.) accompanied
148 assessors to the starting point of the survey, and during those 30 minutes was available to answer any
149 questions and provide clarifications. The guidelines and the survey form were available in English and
150 Bahasa Malaysia languages. Once the assessment began, the assessors were asked not to discuss answers
151 with one another or seek any further clarifications. The survey was conducted along a pre-designated 500m
152 transect at each site. Assessors spent a total of one hour walking each transect and considering their answers.

153 2.3 Test #1: Comparison of survey test scores to independent forest condition metrics

154 To ascertain the ability of the survey to correctly identify sites of good or poor forest condition, we compared
155 scores from the full survey test dataset with data from independent published studies that were conducted at,
156 or whose spatial coverage overlapped with, our 16 test sites. We used a total of 967 assessor surveys for this
157 test- note that not every assessor visited every site, but nearly all did. The validation data were derived from
158 independent studies, and encompassed a variety of aspects of forest condition or conservation value (Table 1,
159 Supplementary Fig. 2), namely: the species richness of dipterocarp trees (Yeong et al., 2016a) and of ants
160 (Tawatao et al., 2014), aboveground carbon stocks (Asner et al., 2018), and vegetation structure complexity,
161 decomposition rate, and various aspects of ecosystem function in dipterocarp trees (Yeong et al. 2016a,

162 2016b, *in review*). In almost all cases these independent data were collected on the same forest trails, plots or
163 coordinates as our assessments. The exception was the generation of aboveground carbon estimates (which
164 were provided as a 30m x 30m resolution raster), where the mean value across a circular buffer of diameter
165 1.5km was extracted using the coordinates for each site. We used these datasets to calculate R-squared
166 statistics between these components of forest condition and the survey scores generated by our assessors
167 (Table 1).

168 2.4 Test #2: Examining assessor effects on survey scoring

169 In order to establish if the FIA tool could be rolled out more widely, irrespective of the background and other
170 characteristics of the assessors involved, we needed to understand how survey scoring may have been
171 affected by prior knowledge and expertise. We also needed to assess the extent to which individual assessors
172 scored consistently across sites. We therefore used the test data to examine the effect of these factors on
173 survey scoring using a Generalised Linear Mixed Modelling (GLMM) approach. For these analyses, and to
174 link scoring of particular individuals across sites, we constructed GLMMs using test data from a subset of 34
175 assessors who: 1) chose not to remain anonymous, and 2) had identified themselves across at least 5 survey
176 sites (n= 34 assessors, n = 493 total assessor surveys). Explanatory variables in models included the highest
177 educational qualification obtained (four categories- 1) 'primary school', 2) 'secondary school', 3) 'pre-
178 university', Diploma or Malaysian Matriculation Programme, and 4) 'degree'- university undergraduate
179 degree); prior forest knowledge (two self-assessed categories: prior knowledge or no prior knowledge); age
180 (three categories: aged <31, 31-50, or 50<); gender (M/F). Candidate models were constructed using all
181 possible combinations of these variables, with individual assessor identity (assessor 1, assessor 2, etc)
182 specified as a random effect (intercept) in all models. GLMMs were constructed using the 'lme4' package in
183 R (Bates et al., 2015) , model selection was performed using AIC (Akaike Information Criterion, Burnham &
184 Anderson, 2002), and the associated marginal and conditional R-squared values were estimated using the
185 Delta method (Bartoń 2020).

186 2.5 Test #3: Determining survey effectiveness and identifying improvements

187 We tested two key performance aspects of the survey, namely: 1) survey effectiveness, i.e. how effective the
188 survey questions were at discriminating between sites of better or worse condition (hereafter 'discriminative

189 ability’), and 2) response consistency, i.e. how consistent different assessors were when surveying the same
190 sites (hereafter ‘agreement rate’).

191 To estimate the discriminative ability of questions, we first calculated the mean answer for each of the 50
192 questions at each site ($0 \leq x \leq 1$), across all assessors. We then calculated the standard deviation in these
193 means across all 16 sites to obtain a discriminative ability score for each question (hereafter referred to as the
194 ‘discriminative ability’ measure). Using the mean answer for each question at each site ensured that
195 variability associated with the assessors’ answers at each site did not contribute towards our measure of the
196 question’s ability to discriminate between sites.

197 To estimate an agreement rate score that accounts for scorers agreeing simply by chance, we calculated
198 Fleiss’s kappa score for each survey question using the ‘irr’ package in R (Gamer et al., 2019). Fleiss’s
199 kappa is a dimensionless score from zero to one, where one indicates a high level of agreement, and zero a
200 low level of agreement.

201 Both these measures were derived from a subset of the full data, wherein 30 responses- the maximum
202 number of responses available for all 16 sites- were randomly chosen and analysed for each of the 16 sites.
203 This was to ensure that each site was represented by the same number of survey responses, i.e. a balanced
204 design, with each site contributing equally to the overall scoring for each aspect of survey performance (30
205 responses per site x 16 sites = 420 total assessor surveys).

206 Finally, in the interest of optimising the survey, we tested the effect of removing survey questions where
207 agreement and/or discriminative ability was low, using the balanced subset of 420 assessor surveys.

208 Questions were removed in groups (see results for details of these), and final scores were recalculated, before
209 up-weighting them to the equivalent of a score out of 50, to enable comparisons with scores from the full
210 survey. R-squared statistics were also recalculated to test for any change in agreement with the independent
211 forest condition metrics (see Methods 2.3).

212 3 RESULTS

213 3.2 Test #1: Comparison of survey test scores to independent forest condition metrics

214 Overall, survey scores ranged from 10 to 47 out of 50, indicating that our sample of sites spanned a wide
215 range of forest condition. The survey scores generated by our assessors agreed with the majority indicators of
216 forest condition that we extracted from the independent studies specified in the methods (Fig. 1, Table 1).
217 Foremost among these was a strong positive association with vegetation structure ($R^2 = 0.75$ vs. forest
218 condition score, Yeong et al., 2016a, Fig. 1c). Scores were also positively associated with aboveground
219 carbon stocks ($R^2 = 0.69$ vs. carbon in Mg C per ha, Asner et al., 2018, Fig. 1d), although here there were
220 signs that the association could be non-linear. R-squared statistics also showed good agreement between FIA
221 scores and important aspects of ecosystem functioning in forests (decomposition $R^2 = 0.74$, Fig. 1e;
222 dipterocarp seedling prevalence $R^2 = 0.78$, Fig. 1f; Table 1). There was reasonable agreement between scores
223 and the alpha diversity of the forests they were collected in ($R^2 = 0.50$ vs. dipterocarps, 0.54 vs. ants,
224 Fig. 1a,b), considering the high variability of measured diversity at these sites (Tawatao et al., 2014, Yeong et
225 al., 2016a). Finally, some aspects of functioning did not seem to be associated with survey score (dipterocarp
226 fruiting, Fig. 1g, dipterocarp seedling survival, Fig. 1h). Overall these results offer good support for the use
227 of the FIA tool as an estimate of the relative condition of forest sites.

228 3.2 Test #2: Examining assessor effects on survey scoring

229 We found limited evidence for assessor effects on survey scoring. A GLMM containing the prior knowledge
230 variable (alone) was the ‘best’ performing model in the set (Burnham & Anderson, 2002), but five other
231 models with different formulations achieved an AIC score within two points of this model, thus also
232 achieving ‘substantial’ support (Supplementary Table 2). Importantly, the marginal R-squared statistics for
233 these six models (as for all models in the set) were low- max 0.06, minimum 0.03- indicating that assessor
234 characteristics accounted for no more than 6% of the variation in scoring (Supplementary Table 2). The
235 conditional R-squared statistic for the participant identity random effect was 0.18, indicating that 18% of the
236 variation in scoring was due to participant identity. It was therefore likely that individual-level variability

237 between assessors had a stronger influence on scoring than the other characteristics (education, prior
238 knowledge, age or gender).

239 3.3 Test #3: Determining survey effectiveness and identifying improvements

240 Discriminative ability of questions varied widely (Fig. 2). There was a large amount of variation in answers
241 to questions on landscape, topography or trees, whereas answers to disturbance questions were more
242 consistent. One question had zero spread of answers: Q4 asking if the fragment was 1ha or above. This
243 question therefore had zero discriminative power across the range of sites we visited. The average Fleiss's
244 kappa score across questions was 0.25, indicating fair agreement between scorers across sites (Fig. 2). The
245 level of agreement for two questions was worse than that expected by chance alone; these were questions
246 relating to the presence of waterfalls (Q14) and off-trail visibility (Q46). Overall, questions with a higher
247 level of agreement tended to also be better at discriminating across sites (Fig. 2).

248 Removing questions tended to have a negligible effect on the relative ranking of sites based on the average
249 survey score (Fig. 3a), and also had little effect on the value of R-squared statistics calculated against
250 independent metrics of forest condition (Fig. 3b). Removing every question appearing in the lower left-hand
251 panel of Fig. 2 (labelled 'C'), corresponding to all survey questions with below average discriminative ability
252 and below average agreement rate ($n = 23$ questions), tended to raise scores for sites in good condition, and
253 in two instances this would have changed the ordering of site rankings (by ± 1 rank, Fig. 3a). Removing Fig.
254 2 panel 'C' questions also mildly worsened R-squared scores, by a mean of -0.008 (Fig. 3b). Using only
255 those questions with above average discriminative ability and agreement rate ($n = 20$ questions, upper right
256 hand panel labelled 'B' in Fig. 2) had a similar but slightly stronger effect on site rankings (Fig. 3a), but
257 again reduced the value of r-squared statistics vs. the forest condition metrics (by a mean of -0.026 , relative
258 to the full survey). Removing the 'worst' ten questions (as per assessor agreement) had a negligible effect on
259 site rankings (Fig. 3a) and on R-squared scores (Fig. 3b, mean change: $+0.001$).

260 4 DISCUSSION

261 4.1 Agreement with independent forest condition metrics

262 The FIA tool was effective in measuring forest condition across our criteria of vegetation structure
263 complexity, biodiversity and ecosystem functioning (Fig. 1). Strong agreement with vegetation structure and
264 aboveground carbon estimates might have been expected, given the emphasis on structure-related questions
265 in the survey, such as the size and number of trees (n = 13 structure questions of 50 in total). This
266 nevertheless demonstrated that assessors were able to accurately identify critical aspects of forest structure
267 that reflect the forest condition and conservation value without spending large amounts of time taking
268 detailed measurements of tree size and identity. The strength of these associations also indicated the potential
269 applicability of this tool to REDD+ schemes and the HCS Approach, which both use carbon and vegetation
270 structure as proxies for forest 'value' (Brofeldt et al., 2014; Rosoman et al., 2017).

271 Vegetation structure and carbon stocks have been shown to correlate closely with biodiversity in tropical
272 forest habitats (Lindenmayer et al., 2000; Gao et al., 2014; Magnago et al., 2015; Deere et al., 2018), and this
273 was the case among our test sites (Tawatao et al., 2014). For the two biodiversity datasets available at our
274 study sites (dipterocarp tree diversity and ant diversity), the survey responses were reasonably well
275 correlated with these datasets, particularly given that the FIA assessors were not required to count or identify
276 species. We would expect that the survey would perform similarly against many other groups of species that
277 are forest dependent, and hence vary in occurrence in relation to forest condition. We were not able to test
278 the FIA tool against datasets of vertebrate biodiversity, which is often of concern to conservation initiatives,
279 but sections of the survey that ask questions about evidence of human disturbance including hunting, as well
280 as signs or sightings of mammals, provide potentially important insights into threat levels to vertebrates
281 (Green et al., 2019; Brodie et al., 2015).

282 The functioning of forest ecosystems is critical to their longer term ability to support biodiversity and
283 ecosystem services (Tierney et al., 2009; Wurtzebach & Schultz, 2016). Although assessors were not asked
284 to identify specific forest functions, our survey scores showed good agreement with aspects of ecosystem

285 functioning, including regeneration and decomposition rates, indicating that the characteristics covered in the
286 survey are highly relevant to ecological processes as well as structure and diversity.

287 It should be noted that the characteristics of a forest that make it of conservation value differ depending on
288 the particular conservation goal, and are ultimately a value judgement. Our aim was to measure forest
289 condition against the assumption that an intact forest ecosystem will provide the widest range of important
290 services. We believe the tool goes some way to both address the complexity of the forest ecosystem and
291 reduce the impact of value judgments that any individual metric might place on what constitutes ‘good
292 condition’ by including a range of important elements such as vegetation structure, fauna and indicators of
293 human disturbance. Therefore, although a site cannot achieve a perfect score if any of these elements are
294 reduced, a site can still score well based on a range of different criteria. Questions such as those relating to
295 saplings and fruits also indicate potential for the site to recover. It will be useful for managers to scrutinise
296 the elements of the survey that contribute to the score, rather than simply use the total score, in order to
297 understand the site condition in relation to specific conservation goals and to develop effective management.
298 Developing supporting guidance on interpreting scores for subsequent conservation management would
299 therefore be beneficial.

300 4.2 Usability and consistency

301 Overall, assessors were consistent at ranking sites by their relative condition. Accuracy and usability are key
302 criteria for the success of the tool in improving the uptake of effective forest monitoring among a wider
303 range of forest stewards with varied backgrounds and levels of expertise. Scoring by individuals had strong
304 internal consistency across sites, and there was no evidence that prior knowledge or experience of forest
305 ecology influenced the ability of assessors to discriminate between sites of different condition. Community
306 forest monitoring for REDD+, which uses more involved vegetation measurement protocols, was found to be
307 similarly reliable when undertaken by non-experts (Holck et al., 2008). The FIA tool enables the inclusion of
308 other forest properties beyond vegetation structure, with a quicker and simpler approach, and without
309 requiring additional expertise or training.

310 It should however be noted that the use of a subset of assessors who had chosen not to remain anonymous
311 could have introduced some level of bias into the derived scores. For example, those who were comfortable

312 in identifying themselves may have been more confident in their knowledge of forests, and those less
313 confident may have opted to remain anonymous. It may also be the case that the assessors that opted to self-
314 identify were less likely to deviate from what they perceived to be the consensus for the site, i.e. to provide
315 some sort of 'right' answer ('impression management', Drescher et al., 2013). It is difficult to speculate on
316 what the net effect of these biases might be, but we would however highlight that our tests on the full dataset
317 (anonymous or otherwise) showed that assessor characteristics did not affect scoring in any substantial way.

318 The assessors in our study had the opportunity to ask for clarifications from a scientist before starting the
319 survey. Questions were fully understood, on the whole, and clarifications were requested for unfamiliar
320 technical words like 'ravine', 'ephemeral', or 'cicada chimney'. This could have some effect on surveys
321 conducted without an expert present, however, we would expect that future revisions to the FIA, including
322 simpler wording, further written explanation, or provision of photos or diagrams would largely solve this
323 issue.

324 Although assessors were consistent in their ranking of sites, our results also show that there is substantial
325 variability in the absolute scores recorded by each assessor at each site, and thus scores for intact forest
326 sometimes overlapped scores for sites likely to have been in poorer condition. Therefore, individual scores
327 cannot currently be used to ascertain a specific level of forest condition. The FIA tool can, however, be used
328 to rank sites, to understand which areas may be of higher or lower conservation value, and to identify
329 suitable targets for restoration activities. It may also be possible to use the FIA to monitor the condition of a
330 site over time to understand if degradation or recovery is taking place, although more research is needed to
331 test the sensitivity of the tool to changes over time, and to determine the period of time over which changes
332 may become detectable. To make these sorts of ranks and comparisons, it is important to either use the same
333 assessor- as indicated by our results, which show strong internal consistency in survey responses- or perhaps
334 to generate an average of scores across a number of assessors for each site or time period. It should be noted
335 that we did not test the consistency of individual assessors in scoring the same site, therefore the ability of
336 the tool to accurately detect changes over time is unclear.

337 Reducing variability in responses between assessors is important if we wish to compare the site scores
338 recorded by different individuals, and to increase our confidence in assigning scores to levels of forest

339 condition (e.g. that a score of > 35 indicates primary forest). To understand whether there were aspects of the
340 survey that could be improved to reduce this variability, we investigated the discriminative ability and level
341 of agreement of the 50 survey questions. For questions where there was a low level of agreement, additional
342 explanations, graphics or training could reduce variability without compromising the range of aspects of
343 forest condition considered by the survey. Pre-survey calibration between assessors could also be employed
344 to improve agreement further. Limited discriminative ability may have been caused by characteristics that
345 are either too common at all sites, or too rare to be detected even at the most intact forest sites. One option
346 might be to remove these questions altogether, but this would have negligible effect on overall efficiency
347 (Fig. 3), or on the overall time spent surveying a site. Furthermore, although some questions may not have
348 been important across the range of sites we tested, they could become important at other sites- particularly
349 those sites of very good or very poor condition.

350 4.3 Applications

351 The FIA tool is able to rank the condition of forest areas based on our key criteria. For this reason, the tool is
352 competitive alongside alternative methods such as tree enumeration, remote sensing, or wildlife monitoring,
353 because it is able to capture vital information from across all these aspects in a fraction of the cost or time,
354 and without the need for technical expertise. These attributes would enable projects to rapidly assess forests
355 on the ground in a way that allows for engagement and participation by the wider community. Our results
356 also indicate a continuing need for field surveys, even with full access to remote sensing data, because a
357 number of aspects of forest condition for which measurement via remote sensing is either difficult or
358 impossible- such as signs of leeches (Q41), epiphytes (Q33), and sub-canopy tree metrics (Q15-25)- scored
359 highly for discriminative ability at our sites (Fig. 2).

360 The FIA tool may have applications for any project which requires information about forest condition. These
361 could include eco-tourism restoration initiatives or monitoring of conservation set-asides, such as those in
362 RSPO certified oil palm plantations. Some conservation initiatives require specific information on particular
363 aspects of forest condition, such as aboveground carbon stocks (e.g.in carbon accounting schemes) or the
364 abundance of particular focal species of conservation concern (e.g. orang-utan conservation programmes).
365 The FIA tool is no substitute for the detailed and focused measurements required for these sorts of projects.

366 It may, however, complement or supplement such measurements, enabling field staff and local communities
367 to cheaply, efficiently, and systematically capture information about the wider condition of the forest which
368 could be pertinent to the focus of the project. For example, it may provide information about vegetation
369 degradation that may affect habitat for focal species, or identify human disturbance that could impact on
370 carbon stocks in the future. We argue that regularly monitoring the full range of aspects of forest condition,
371 through the use of the FIA, could therefore contribute towards conservation goals, as well having wider non-
372 target benefits.

373 5 CONCLUSIONS

- 374 1. The FIA tool is effective in ranking sites in terms of condition, but variation among assessors means
375 that it is important that the same individual is used to conduct comparison of sites over space or time.
376 Alternatively, taking a mean score across multiple assessments of the same site is likely to improve
377 the robustness of condition estimates.
- 378 2. More information and training would enhance the accuracy of the survey. Some common sources of
379 inaccuracy could be mitigated by the provision of photos and other visual aids to help understand
380 survey questions. However, to maximise uptake it is important to balance the need for improved
381 accuracy with the need for the survey to remain quick, cost-effective and accessible for non-experts.
- 382 3. The tool was shown to be effective in discriminating among forests of varying condition, but we did
383 not test whether it is sufficiently sensitive to detect changes over time, or how repeatable scores are
384 by individuals for the same site, which are important factors for monitoring purposes. More testing is
385 needed to understand whether it can be used to monitor restoration projects, for example, and if so,
386 the requisite frequency and intensity of surveying that would be required.
- 387 4. While its simplicity may not provide the detail needed for focussed conservation projects, the FIA
388 tool provides a robust and systematic means of monitoring forest set-asides, providing rapid
389 monitoring data that are accessible to a wide range of potential users.

390 Authors' contributions

391 J.M.L., G.R. & A.L. conceived research, J.M.L, A.J.S., J.K.H, K.H. & A.A. designed research;, A.J.S.,
392 K.L.Y. & J.M.L. performed research; A.J.S., K.L.Y. & J.M.L. analysed data; J.M.L. & A.J.S. led the writing
393 of the paper, and all authors provided comments.

394 Acknowledgements

395 Thanks to Wilmar International, Yayasan Sabah, Sabah Forestry Department, the Danum Valley
396 Management Committee, Sabah Biodiversity Council and Borneo Rainforest Lodge for site access and work
397 permissions. Thanks also to SEARRP and Wilmar International for logistical support, and to the survey
398 volunteers. All statistical analyses were conducted on the R statistical platform (version 3.4.4), and we thank
399 the authors of this software and associated base packages for making this platform available and free to use.

400 Data availability statement

401 All supporting data will be archived via Dryad (<http://datadryad.org/>) and a DOI will be supplied upon
402 publication.

403 Competing interests

404 The authors have no competing interests.

405 ORCID

406 Andrew J. Suggitt: 0000-0001-7697-7633.

407 Kok Loong Yeong: 0000-0001-8193-2130.

408 Keith C. Hamer: 0000-0002-2158-2420.

409 Jane K. Hill: 0000-0003-1871-7715.

410 Jennifer M. Lucey: 0000-0001-5224-091X.

411 References

- 412 Asner, G. P., Brodrick, P. G., Philipson, C., Vaughn, N. R., Martin, R. E., Knapp, D. E., ... Coombes, D. A.
413 (2018). Mapped aboveground carbon stocks to advance forest conservation and recovery in Malaysian
414 Borneo. *Biological Conservation*, 217, 289-310. <https://doi.org/10.1016/j.biocon.2017.10.020>
- 415 Bartoń, K. (2020). MuMIn: Multi-Model Inference. R package version 1.43.17. [https://CRAN.R-](https://CRAN.R-project.org/package=MuMIn)
416 [project.org/package=MuMIn](https://CRAN.R-project.org/package=MuMIn)
- 417 Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4.
418 *Journal of Statistical Software*, 67, 1-48. <https://doi.org/10.18637/jss.v067.i01>
- 419 Brodie, J. F., Giordano, A. J., Zipkin, E. F., Bernard, H., Mohd-Azlan, J., & Ambu, L. (2015). Correlation
420 and persistence of hunting and logging impacts on tropical rainforest mammals. *Conservation Biology*, 29,
421 110-121. <https://doi.org/10.1111/cobi.12389>
- 422 Brofeldt, S., Theilade, I., Burgess, N., Danielsen, F., Poulsen, M., Adrian, T., ... Kurniawan, Y. (2014).
423 Community monitoring of carbon stocks for REDD+: does accuracy and cost change over time? *Forests*, 5,
424 1834-1854. <https://doi.org/10.3390/f5081834>
- 425 Burnham, K. P., & Anderson, D. R. (2002). Model selection and multimodel inference: a practical
426 information-theoretic approach. New York: Springer-Verlag.
- 427 Critchlow, R., Plumptre, A. J., Alidria, B., Nsubuga, M., Driciru, M., Rwetsiba, A., ... Beale, C. M. (2017).
428 Improving law-enforcement effectiveness and efficiency in protected areas using ranger-collected
429 monitoring data. *Conservation Letters*, 10, 572-580. <https://doi.org/10.1111/conl.12288>
- 430 Deere, N. J., Guillera-Aroita, G., Baking, E. L., Bernard, H., Pfeifer, M., Reynolds, G., ... Struebig, M. J.
431 (2018). High Carbon Stock forests provide co-benefits for tropical biodiversity. *Journal of applied ecology*,
432 55, 997-1008. <https://doi.org/10.1111/1365-2664.13023>
- 433 Drescher, M., Perera, A. H., Johnson, C. J., Buse, L. J., Drew, C. A., Burgman, M. A. (2013). Toward
434 rigorous use of expert knowledge in ecological research. *Ecosphere*, 4, 1-26. <https://doi.org/10.1890/ES12->
435 [00415.1](https://doi.org/10.1890/ES12-00415.1)

436 Finer, M., Novoa, S., Weisse, M. J., Petersen, R., Mascaro, J., Souto, T., ... Martinez, R. G. (2018).
437 Combating deforestation: From satellite to intervention. *Science*, *360*, 1303-1305.
438 <https://doi.org/10.1126/science.aat1203>

439 Gamer, M., Lemon, J., & Fellows, I. (2019). irr: Various Coefficients of Interrater Reliability and
440 Agreement. R package version 0.84.1. <https://CRAN.R-project.org/package=irr>

441 Gao, T., Hedblom, M., Emilsson, T., & Nielsen, A. B. (2014). The role of forest stand structure as
442 biodiversity indicator. *Forest Ecology and Management*, *330*, 82-93.
443 <https://doi.org/10.1016/j.foreco.2014.07.007>

444 Gibbs, H. K., Brown, S., Niles, J. O., & Foley, J.A. (2007). Monitoring and estimating tropical forest carbon
445 stocks: making REDD a reality. *Environmental Research Letters*, *2*, 045023. [https://doi.org/10.1088/1748-](https://doi.org/10.1088/1748-9326/2/4/045023)
446 [9326/2/4/045023](https://doi.org/10.1088/1748-9326/2/4/045023)

447 Green, E. J., McRae, L., Freeman, R., Harfoot, M. B., Hill, S. L., Baldwin-Cantello, W., & Simonson, W. D.
448 (2019). Below the canopy: global trends in forest vertebrate populations and their drivers. *PeerJ Preprints*, *7*,
449 e27882v1. <https://doi.org/10.1098/rspb.2020.0533>

450 Hoette, M. H., Kolodin, I. A., Bereznuk, S. L., Slaght, J. C., Kerley, L. L., Soutyrina, S. V., ... Miquelle, D.
451 G. (2016). Indicators of success for smart law enforcement in protected areas: A case study for Russian
452 Amur tiger (*Panthera tigris altaica*) reserves. *Integrative Zoology*, *11*, 2-15. [https://doi.org/10.1111/1749-](https://doi.org/10.1111/1749-4877.12168)
453 [4877.12168](https://doi.org/10.1111/1749-4877.12168)

454 Holck, M. H. (2008). Participatory forest monitoring: an assessment of the accuracy of simple cost-effective
455 methods. *Biodiversity and Conservation*, *17*, 2023-2036. <https://doi.org/10.1007/s10531-007-9273-4>

456 Hosonuma, N., Herold, M., De Sy, V., De Fries, R. S., Brockhaus, M., Verchot, L., ... Romijn, E. (2012). An
457 assessment of deforestation and forest degradation drivers in developing countries. *Environmental Research*
458 *Letters*, *7*, 044009. <https://doi.org/10.1088/1748-9326/7/4/044009>

459 Kunjuraman, V., & Aziz, R. C. (2019). Community-Based Ecotourism (CBE) development in Lower
460 Kinabatangan of Sabah, Malaysian Borneo. *Revista Turismo: estudos e práticas*, *8*, 13-43.

461 Lindenmayer, D. B., Margules, C. R., & Botkin, D. B. (2000). Indicators of biodiversity for ecologically
462 sustainable forest management. *Conservation Biology*, *14*, 941-950. [https://doi.org/10.1046/j.1523-](https://doi.org/10.1046/j.1523-1739.2000.98533.x)
463 [1739.2000.98533.x](https://doi.org/10.1046/j.1523-1739.2000.98533.x)

464 Lindhe, A., & Drakenberg, B. (2019). *Forest integrity assessment: A simple and user-friendly tool for*
465 *assessing and monitoring biodiversity conditions in forests and forest remnants*. Oxford: HCV Resource
466 Network.

467 Lindhe, D., Payne, J., Reynolds, G., Drakenberg, B., Lucey, J. M., & Yeong, K. L. (2015). *Forest Integrity*
468 *assessment template for Lowland Dipterocarp Rainforest*. Oxford: HCV Resource Network.

469 Luke, S. H., Barclay, H., Bidin, K., Chey, V. K., Ewers, R. M., Foster, W. A., ... Walsh, R. P. (2017). The
470 effects of catchment and riparian forest quality on stream environmental conditions across a tropical
471 rainforest and oil palm landscape in Malaysian Borneo. *Ecohydrology*, *10*, e1827.
472 <https://doi.org/10.1002/eco.1827>

473 Magnago, L. F. S., Magrach, A., Laurance, W. F., Martins, S. V., Meira-Neto, J. A. A., Simonelli, M., &
474 Edwards, D. P. (2015). Would protecting tropical forest fragments provide carbon and biodiversity
475 cobenefits under REDD+?. *Global Change Biology*, *21*, 3455-3468. <https://doi.org/10.1111/gcb.12937>

476 Rosoman, G., Sheun, S. S., Opal, C., Anderson, P., & Trapshah, R., editors. (2017). *The HCS Approach*
477 *Toolkit*. Singapore: HCS Approach Steering Group.

478 Tawato, N., Lucey, J. M., Senior, M., Benedick, S., Khen, C. V., Hill, J. K., & Hamer, K. C. (2014).
479 Biodiversity of leaf-litter ants in fragmented tropical rainforests of Borneo: the value of publically and
480 privately managed forest fragments. *Biodiversity Conservation*, *23*, 3113-3126. [https://](https://doi.org/10.1007/s10531-014-0768-5)
481 doi.org/10.1007/s10531-014-0768-5

482 Tierney, G. L., Faber-Langendoen, D., Mitchell, B. R., Shriver, W. G., & Gibbs, J. P. (2009). Monitoring
483 and evaluating the ecological integrity of forest ecosystems. *Frontiers in Ecology and the Environment*, *7*,
484 308-316. <https://doi.org/10.1890/070176>

- 485 Wang, X., Feng, Z., & Ouyang, Z. (2001). The impact of human disturbance on vegetative carbon storage in
486 forest ecosystems in China. *Forest Ecology and Management*, *148*, 117-123. <https://doi.org/10.1016/S0378->
487 [1127\(00\)00482-5](https://doi.org/10.1016/S0378-1127(00)00482-5)
- 488 Watson, J. E., Evans, T., Venter, O., Williams, B., Tulloch, A., Stewart, C., ... McAlpine, C. (2018). The
489 exceptional value of intact forest ecosystems. *Nature Ecology & Evolution*, *2*, 599.
490 <https://doi.org/10.1038/s41559-018-0490-x>
- 491 Wurtzebach, Z., & Schultz, C. (2016). Measuring ecological integrity: history, practical applications, and
492 research opportunities. *BioScience*, *66*, 446-457. <https://doi.org/10.1093/biosci/biw037>
- 493 Yeong, K. L., Lucey, J. M., Jelling, A., Benedick, S., Bernadus, M., Ola, B., ... Hill, J. K. Habitat
494 fragmentation disrupts rainforest regeneration. *In review*.
- 495 Yeong, K. L., Reynolds, G., & Hill, J. K. (2016a). Leaf litter decomposition rates in degraded and
496 fragmented tropical rain forests of Borneo. *Biotropica*, *48*, 443-452. <https://doi.org/10.1111/btp.12319>
- 497 Yeong, K. L., Reynolds, G., & Hill, J. K. (2016b). Enrichment planting to improve habitat quality and
498 conservation value of tropical rainforest fragments. *Biodiversity Conservation*, *25*, 957-973.
499 <https://doi.org/10.1007/s10531-016-1100-3>

500 [Supporting information](#)

501 Ethical approval was granted by the University of York Ethics Committee (reference number: 20160404).
502 Permissions to enter the field sites were granted by Wilmar International Ltd for privately owned sites, by
503 Sabah Forestry Department for publically managed Virgin Jungle Reserves and Malua forest reserve, and
504 Danum Valley Management Committee for the Danum Valley Conservation Area.

505 Additional supporting information may be found online in the Supporting Information section.

506 [Funding information](#)

507 A.J.S. was supported by the Natural Environment Research Council (NERC, Grant number NE/M013030/1).
508 J.M.L. was supported by a NERC KE fellowship (Grant number NE/M006840/1-2). A.A., G.R., J.M.L.,
509 K.C.H., J.K.H. & Y.K.L. were supported by the Socially and Environmentally Sustainable Palm Oil

510 Research (SEnSOR) programme, which receives funding from the Roundtable on Sustainable Palm Oil
511 (RSPO) and is facilitated by the South East Asia Rainforest Research Partnership (SEARRP).

Table 1 Indicators of forest condition, their description and reference in the ecological literature, and the value of the R-squared statistic when compared to the Forest Integrity Assessment (FIA) scores of our assessors.

Indicator of forest condition	Description	Reference(s)	Display Item	R-squared (R ²) vs. FIA scores
Alpha diversity (species richness)	Field sampling of dipterocarps (> 30cm DBH) and ants (quadrats).	Dipterocarps- Yeong <i>et al.</i> 2016a	Fig. 1a	0.50
		Ants- Tawatao <i>et al.</i> 2014	Fig. 1b	0.54
Vegetation structure	Vegetation quality score (dimensionless).	Yeong <i>et al.</i> 2016a	Fig. 1c	0.75
Aboveground Carbon	Aboveground carbon (units Mg C per Ha) mapped by combining airborne Light Detection and Ranging (LiDAR) with satellite imaging and other geospatial data.	Asner <i>et al.</i> 2018	Fig. 1d	0.69
Decomposition rate	Percentage loss of litter over 18 months.	Yeong <i>et al.</i> 2016a	Fig. 1e	0.74
Dipterocarp seedling prevalence	Presence in 1 hour searches, expressed as presence/absence per station (max 4, search radius 30m)	Yeong <i>et al.</i> <i>In review</i>	Fig. 1f	0.78
Dipterocarp fruit prevalence			Fig. 1g	0.26
Dipterocarp survival	Planted seedling survival (% over 18 months).	Yeong <i>et al.</i> 2016b	Fig. 1h	0.03.

Figure 1 Agreement between Forest Integrity Assessment (FIA) score (x-axes throughout) and independent metrics of forest condition (y-axes) sampled across a range of patch sizes in Sabah, Malaysian Borneo. The species richness values represent raw (non-bootstrapped) counts of: a) dipterocarp species (> 30cm DBH, Yeong et al. 2016a), and b) ant species (quadrats, Tawato et al. 2014). c) vegetation structural complexity or ‘forest quality’ score (dimensionless, Yeong et al. 2016a), d) aboveground carbon (Mg C per ha, Asner et al. 2018), e) Litter decomposition (% leaf litter mass lost over 120 days, Yeong et al. 2016a), f) dipterocarp seedling prevalence (n seedlings, max 4, Yeong et al. *In review*), g) fruit prevalence (n fruits, max 4, Yeong et al. *In review*), h) dipterocarp seedling survival (% , Yeong et al. 2016b). Error bars represent \pm 95% confidence intervals. Where they appear, the dark green circles represent surveys conducted in continuous forest, while the lighter green circles represent surveys conducted in forest fragments, where symbol size is in proportion to the cube root of their area.

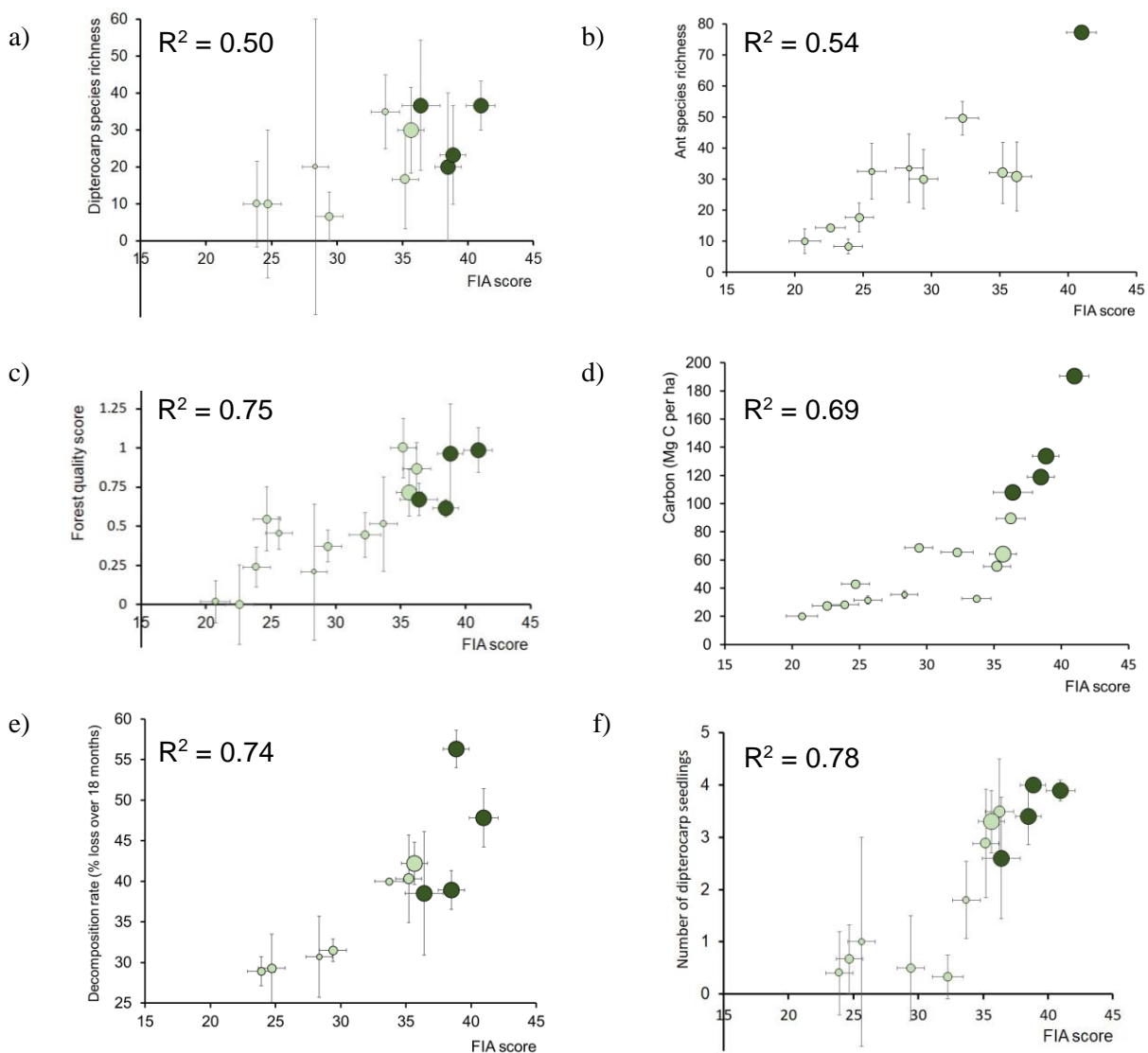
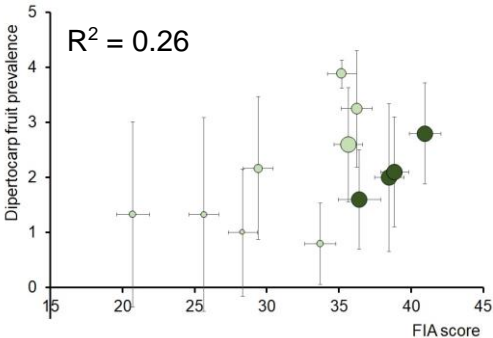


Figure 1 (continued)

g)



h)

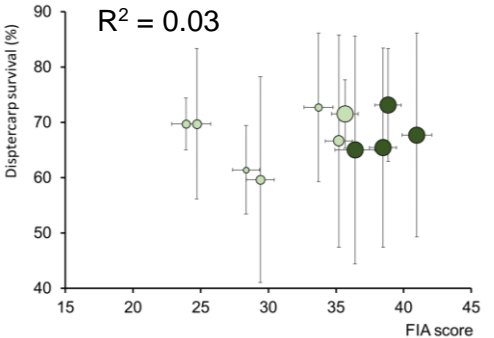


Figure 2 Performance of the Forest Integrity Assessment (FIA) questionnaire. Each question (n = 50 questions) is represented by a smaller circle, and scored based on its discriminative ability (y-axis) and the rate at which recorders agreed on the answer at each site, Fleiss's Kappa score (x-axis). Grand means for each criterion (n = 7 criteria) appear as larger circles, while the global mean score for the whole questionnaire appears as a black diamond (see key). Panels (A to D, illustrated by dotted lines) demark question scores relative to the global mean.

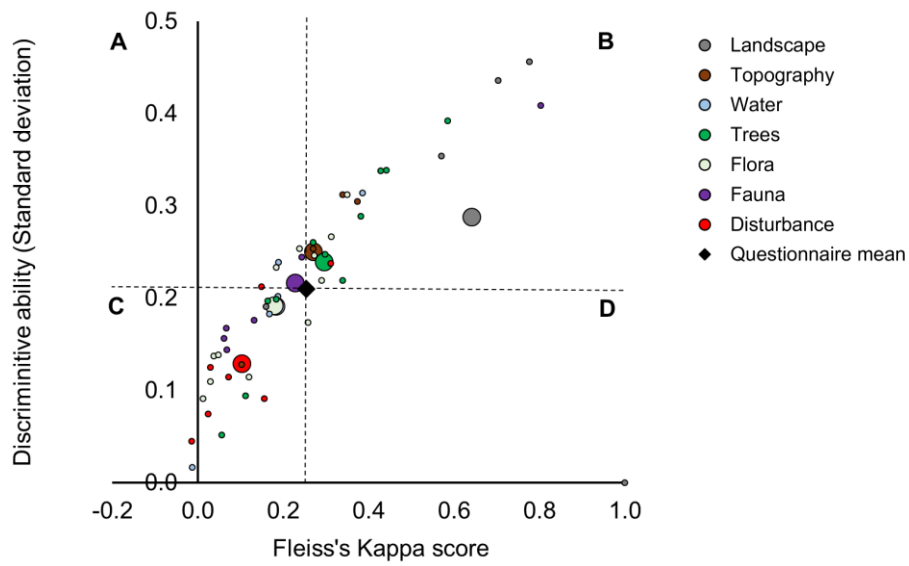


Figure 3 Effect of survey question removal on a) Forest Integrity Assessment (FIA) score and b) R-squared scores vs. independent metrics of forest condition. Neither the removal of the worst 10 questions for agreement, or the removal of all ‘Panel C’ questions (i.e. questions with low agreement and low discriminative ability, n = 23 questions), had a substantial effect on the ranking of sites by FIA score (a) or the estimates for R-squared, calculated using independent metrics of forest condition (b). Using only Panel B questions (those with high agreement rate, high discriminative ability, n = 20 questions) to calculate FIA scores had a mild effect on the order of site rankings, but an overall adverse effect on R-squared estimates.

