

Quantifying uncertainty in biodiversity data for monitoring and reporting indicators



Landcare Research
Manaaki Whenua

Quantifying uncertainty in biodiversity data for monitoring and reporting indicators

Norman W. H. Mason, Robert J. Holdaway, Sarah J. Richardson

Landcare Research

Prepared for:

Department of Conservation

70 Moorhouse Avenue
Christchurch 8011
New Zealand

April 2015

*Landcare Research, Gerald Street, PO Box 69040, Lincoln 7640, New Zealand,
Ph +64 3 321 9999, Fax +64 3 321 9998, www.landcareresearch.co.nz*

Reviewed by:

Approved for release by:

Duane Peltzer
Scientist
Landcare Research

Peter Bellingham
Portfolio Leader – Measuring Biodiversity Change
Landcare Research

Landcare Research Contract Report:

LC 2190

Disclaimer

This report has been prepared by Landcare Research for the Department of Conservation. If used by other parties, no warranty or representation is given as to its accuracy and no liability is accepted for loss or damage arising directly or indirectly from reliance on the information in it.

© Department of Conservation 2015

This report has been produced by Landcare Research New Zealand Ltd for the New Zealand Department of Conservation. All copyright in this report is the property of the Crown and any unauthorised publication, reproduction, or adaptation of this report is a breach of that copyright and illegal.

Contents

Summary	v
1 Introduction.....	1
2 Objectives	2
3 Methods	2
3.1 Quantification of measurement error.....	2
3.2 Comparison to existing Data Quality Limits (DQLs).....	5
3.3 Framework to incorporate measurement error in National Biodiversity Monitoring and Reporting Measures	6
4 Results	8
4.1 Measurement error distributions.....	8
4.2 Data Quality Limits	11
4.3 Incorporation of uncertainty into Tier 1 measures	15
5 Discussion	18
5.1 Data quality limits.....	18
5.2 Consequences of measurement error for biodiversity measures.....	19
6 Recommendations.....	21
7 Acknowledgements	21
8 References.....	22

Summary

Project and Client

- We analysed data from seven repeat-measured standard forest biodiversity monitoring plots to (i) identify areas of high uncertainty to focus on during staff training; (ii) assist the development of objective field audit standards, and (iii) quantify the effect of measurement error on uncertainty associated with biodiversity indicators and biodiversity reporting more widely. This work contributes to the Department of Conservation's Biodiversity Monitoring and Reporting programme.

Objectives

- Quantify measurement error uncertainty associated with key biodiversity indicators, such as species richness, species turnover, and cover scores.
- Use these data to assess the suitability of the current biodiversity field audit standards (data quality limits; DQLs) employed by Department of Conservation (DOC).
- Develop a framework to incorporate measurement error into relevant Indicators and Measures for national-scale reporting on biodiversity, focusing on the following two Measures: (i) M 5.1.1. (size-class structure of canopy dominants) by calculating the basal area, size class structure and occupancy of 10 common tree species; and (ii) M 5.2.1. (representation of plant functional types) by calculating the richness of palatable and unpalatable plant species on plots and their occupancy nationally.

Methods

- Measurement errors were estimated by comparing data recorded by different teams on the same plot.
- Errors were modelled either using a single distribution applied generally (e.g. DBH) or using predictive models (e.g. species richness of palatability groups).
- We simulated new data for 500 randomly selected re-measured LUCAS plots, based on modelled errors, to obtain estimates of confidence intervals for the indicators we examined that account for measurement error.

Results

- Eight of the most widespread tree species showed a significant change in size distribution in the observed data. However, when measurement error was included only four species differed significantly between measurements.
- There was a significant increase in the proportional richness of "Avoided" species even when measurement error was included in estimating confidence intervals.
- Many of the plot measures varied more between teams than the amount of variation allowed for in DOC's DQLs.

- Measurement considerably increased uncertainty for some key indicators, which means the power of the LUCAS plot network to detect changes is not as great as previously assumed.
- Non-detection probabilities for individual species can be estimated using a few simple predictors. This may help identify species and functional groups where greater standardisation in search effort within plots and species identification is required.

Conclusions

- Our results demonstrate the importance of quantifying and integrating measurement error into biodiversity measures. Our data clearly indicate that measurement error is an unavoidable component of biodiversity data.
- Measures of community-level species composition are particularly vulnerable to measurement error, and thus may have limited power to detect change through time. However, we demonstrate that key monitoring and reporting measures are robust to this measurement error when applied to a large plot network with high statistical power.
- Inclusion of uncertainty can alter the statistical significance of changes in measures, underscoring the need to account for uncertainty in biodiversity reporting in order to minimise the risk of Type I errors (i.e. reporting a ‘false’ change in a measure).

Recommendations

- We recommend incorporating measurement uncertainty into monitoring and reporting of biodiversity measures across Public Conservation Lands. To achieve this, estimates of measurement uncertainty are required from a larger sample of forest ecosystems and from non-forest ecosystems.
- We recommend extending this work to assess the uncertainty in all measures used to report on Ecological Integrity.
- We recommend that indices of species turnover and measures based on raw species richness be avoided in Monitoring and Reporting programmes due to the effect of (unavoidable) measurement error.
- We recommend using the RECCE data to determine the consequences of measurement error on classification robustness using the Wiser et al. (2011) classification.
- We recommend applying the uncertainty framework to existing data from experimental settings (e.g. exclosure studies) to test whether reported effects are statistically robust once measurement uncertainty is accounted for.
- We recommend quantifying detection probabilities for a representative suite of the indicator species proposed by Monks et al. (2013) across a range of growth forms, assumed abundance and habitats, and developing a framework for reporting on those species with uncertainty (specifically false absences).
- Measurement error is an unavoidable component of biodiversity data. We recommend a two-stage solution that first minimises error through comprehensive training programmes, and second, accommodates remaining error through quantification and integration of measurement error into reported measures.

1 Introduction

National scale biodiversity reporting relies on the use of highly derived metrics to synthesise across large datasets and spatial scales (Pereira et al. 2013). For example, the valuation of ecosystem services requires integration of data on ecosystem function (mechanisms, fluxes, pools), land use (maps, classifications, area estimates), and economic or social estimates of the value provided by that service (Costanza et al. 1997; Cardinale et al. 2012). In New Zealand, the Department of Conservation (DOC) is currently developing a derived metric of 'Ecological Integrity' that aggregates multiple data sources on animal and plant occupancy and abundance (Lee et al. 2005; Bellingham et al. 2013).

Each component data source will have uncertainty. Uncertainty arises from several sources: the inability to perfectly measure key variables (i.e. differences among observers), the necessary use of models to make predictions (e.g. the use of allometric relationships to model tree biomass from tree diameter), and the natural variability of ecosystem processes across the landscape (sampling uncertainty; Bolker 2008). Although sampling uncertainty is usually reported (i.e. among-plot variability in the derived metric), other sources, such as measurement error and model uncertainty, are generally not incorporated into the reported metrics (Clark and Kellner 2012; Muller-Landau et al. 2013).

Misrepresentation of uncertainty in derived metrics can lead to false assessment of significance and biased results. For example, Phillips et al. (1998) analysed long-term plot data and reported that tropical forests were a net carbon sink; however, re-analysis by Clark (2002) showed that this result was biased by 'artefacts' associated with measurement of buttressed trees. It is therefore important for researchers to have quantitative estimates of the uncertainty associated with derived metrics (Chave et al. 2004; Yanai et al. 2010; Butt et al. 2013; Holdaway et al. 2014). It is essential to show the correct level of uncertainty in derived metrics so that management implications and policy decisions can be assessed with the appropriate level of confidence. Understanding the major determinants of uncertainty can also be a powerful tool for improving methodology and the accuracy of the resulting estimates (for example, Baker et al. 2004).

To quantify measurement error associated with standard field methodologies used by the Land Use and Carbon Analysis System (LUCAS) and DOC for national biodiversity monitoring and reporting (DOC 2014), Landcare Research, with support from DOC, made three fully independent measurements of seven 20 × 20 m standard biodiversity monitoring natural forest plots in 2011. These data have been analysed in relation to uncertainty in carbon estimates, in particular uncertainty in plot area, tagged stems, tree height, tree species identification, diameter at breast height (DBH), and coarse woody debris (CWD) (Holdaway et al. 2014). Here, we analyse the biodiversity components of the data (e.g. species cover scores in height tiers collected using the RECCE method) and use these alongside the analyses of stem diameter components of the data (from Holdaway et al. 2014) to (i) identify areas of high uncertainty to focus on during staff training; (ii) assist the development of objective field audit standards, and (iii) quantify the effect of measurement error on uncertainty associated with biodiversity indicators and biodiversity reporting more widely.

2 Objectives

The objectives of this investigation are to use data from seven standard biodiversity monitoring plots with three repeat-measurements (hereafter referred to as the ‘uncertainty plots’) to:

1. quantify measurement error uncertainty associated with key biodiversity indicators, such as species richness, species turnover, and cover scores;
2. use these data to assess the suitability of the current biodiversity field audit standards (DQL’s) employed by DOC;
3. develop a framework to incorporate measurement error into relevant Indicators and Measures for national-scale reporting on biodiversity, focusing on the following two Measures: (i) M 5.1.1. (size-class structure of canopy dominants) by calculating the basal area, size class structure and occupancy of 10 common tree species; and (ii) M 5.2.1. (representation of plant functional types) by calculating the richness of palatable and unpalatable plant species on plots and their occupancy nationally.

3 Methods

3.1 Quantification of measurement error

3.1.1 Data sources

We used data collected by Holdaway et al. (2014) to quantify observed measurement error distributions. In March 2012, seven existing 20 × 20 m LUCAS natural forest plots were measured three times using independent field teams following the standard LUCAS field protocols (Payton et al. 2004; MfE 2011). These seven ‘uncertainty plots’ were located in the central North Island of New Zealand, and were selected to encompass a broad range of temperate broadleaved forest types and stem densities (Holdaway et al. 2014). While these plots are a geographically restricted sample of New Zealand’s forests, they contain examples of most of the dominant forest types (e.g. beech forest, podocarp forest, broadleaved forest, and regenerating forest). In the absence of further data from elsewhere in New Zealand, these uncertainty plots provide the best available data source to explore the implications of biodiversity uncertainty at a national scale. Each field team comprised four people and included at least one skilled botanist familiar with the local species, and two people with reasonable (>5 years) field experience. Plots typically took one day to complete, and, to represent standard field conditions and time expectations, each team had a 10-day period in which to measure all seven plots. Variation among teams therefore reflected typical measurement error expected from experienced field teams under standard field conditions (with, for example, weather and time constraints). All field teams had the same information prior to arriving at the plot (i.e. plot-sheets and species lists from previous measurements) and used the same field manual. All field staff undertook additional training prior to fieldwork, to standardise interpretation of the field manual. Care was taken to minimise disturbance on the plot and no communication among teams occurred during the measurement period. Individual stems for which species identification was uncertain in the field were collected and identified

by independent expert botanists for each team. Data were entered into the national vegetation survey (NVS) databank and all species were classified according to their palatability to ungulates. Species were assigned to one of four groups based on their palatability to ungulates ('avoided', 'not selected', 'preferred' and 'unclassified' following Forsyth et al. 2002).

3.1.2 Statistical determination of measurement error distributions

Error distributions for measurements of stem diameter, tree height, and tagged species identity were sourced from Holdaway et al. (2014). Error distributions for biodiversity variables derived from the plant species RECCE (species richness and species turnover, species cover scores, and species identification) were quantified as follows.

Species richness and species turnover

We used the Jaccard dissimilarity index to measure species turnover between teams (i.e. observer error for species presence-absence). We did this for individual tiers, growth forms and palatability classes within plots. Jaccard dissimilarity incorporates both differences in species richness and changes in species identity between samples. To separate the influence of these factors on dissimilarity values we decomposed the Jaccard index into species richness difference and species name difference components as follows:

$$Jaccard = 1 - \frac{a}{a + b + c}$$

$$Jaccard_{Richness} = 1 - \frac{\min(S_1, S_2)}{\min(S_1, S_2) + |S_1 - S_2|}$$

$$Jaccard_{Namediff} = Jaccard - Jaccard_{Richness}$$

Where: a is the number of shared species, b is the number of species in sample 1 that do not occur in sample 2, c is the number of species that occur in sample 2, but not in sample 1, S_1 is species richness of the first sample and S_2 is species richness of the second sample. In practical terms, splitting Jaccard into richness and name difference components indicates the contribution to overall species turnover made by (i) differences in search effort or taxonomic resolution of species identities and (ii) differences in taxonomic treatment.

Representation of palatability groups

We initially modelled changes in raw palatability group species richness between measurements using the 500 Tier One / LUCAS forest plots described in section 3.3.1. However, this revealed that all groups increased in species richness between measurements (Figure 1). Raw species richness values are often difficult to use as indicators for newly-established plot networks, as there tends to be an across-the-board increase in species richness from the first measurement to subsequent measurements. This is believed to arise due to a combination of extra emphasis on biodiversity during the second measurement and the fact that teams in subsequent years search for all the species on the previous measurement's list and then find extra species. To overcome this artefact in the data, we

chose to focus on the proportional richness of palatability groups, instead of raw richness values.

We estimated differences in proportional richness on a pairwise basis, with the absolute difference between RECCE samples for each pairwise combination of teams for each plot being recorded. We recorded the mean observed proportional richness as the mean of the values for both teams.

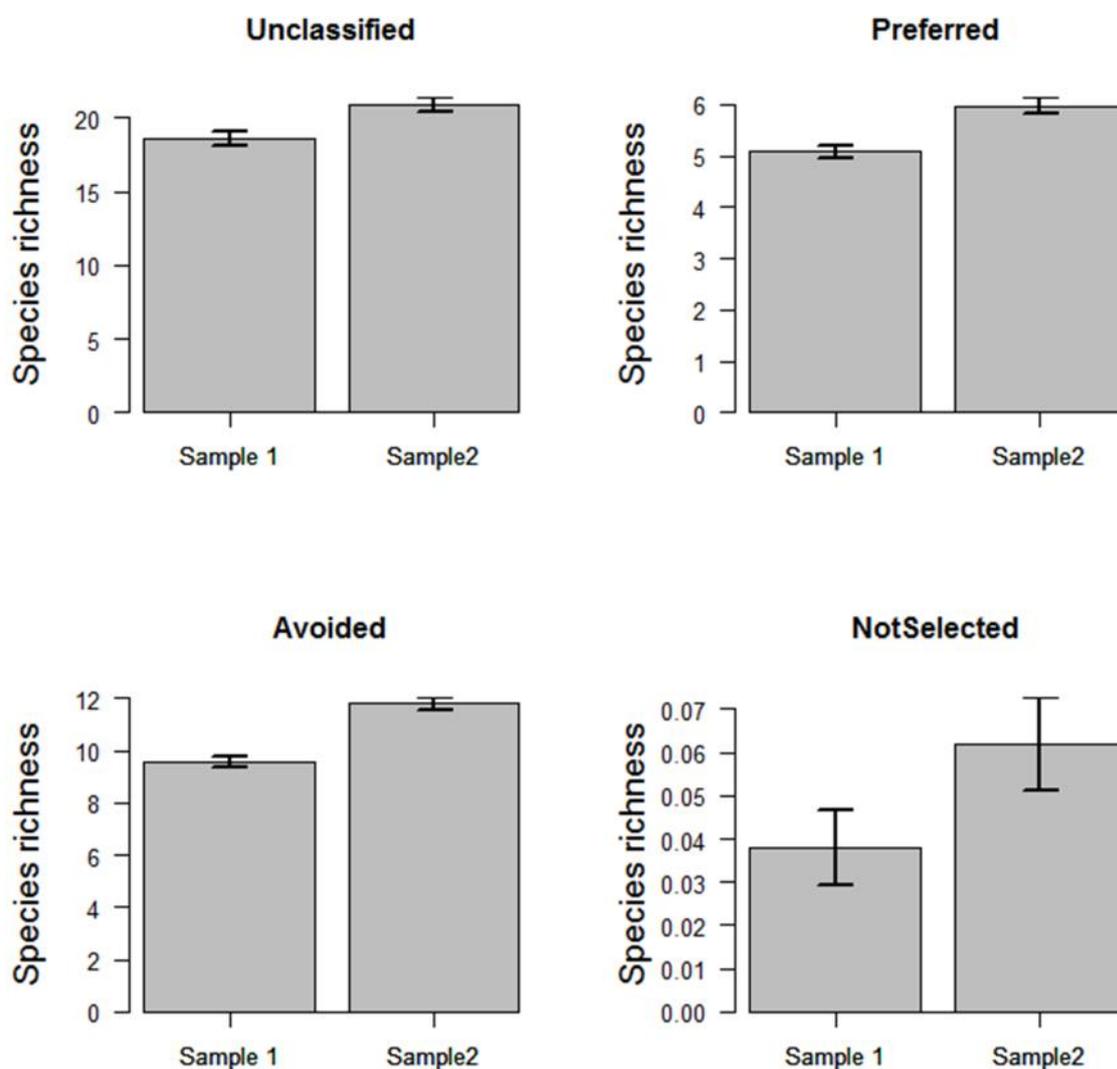


Figure 1 Mean observed richness of palatability groups in the first (2002-2007) and second (2009-2014) LUCAS measurement period (see Methods section 3.3.1 for further details of dataset used). The results show that each group has increased in mean richness between measurements, which is probably an artefact. Error bars show the standard error of the mean, and the mean interval between measurements was seven years.

We modelled uncertainty in proportional richness as a function of palatability group identity and the mean observed proportional richness using generalised linear models (GLM). We used a quasibinomial distribution with a logit link function to ensure that predicted values were bound between 0 and 1. GLM analyses were achieved using the *glm()* function in R.

The resulting GLM was used to simulate errors in the proportional richness of each group. This model included observed species richness, observed proportional richness and palatability group identity as predictors.

Species cover score

We assessed uncertainty in species cover scores by comparing cover classes assigned to species that occurred in the same plot and same tier of both measurements, based on pairwise combinations of the three repeat-measurements of the uncertainty plots. We used this to build a confusion matrix, which documented the probability of a species being assigned a certain cover score in the second measurement given the cover score it was assigned in the first measurement. For example, if a species was assigned cover class 1 in the first sample, the confusion matrix gives the probability of it being assigned to the same or a different class in the second measurement.

Species identification (RECCE)

We sought to generate species-specific ‘non-detection’ models for predicting the likelihood of individual species being detected (in the same tier) in one sample of a plot, but not another. We termed this a non-detection probability, although strictly, it could arise either from non-detection or assigning a different name. We used generalised linear mixed effect (GLME) models to predict non-detection probabilities. Plot and team identity were included as random factors, while growth form, tier cover class, species richness of the genus and species richness of the family were included as fixed effects. These last two terms were to test whether the probability of assigning a different name was related to the number of species in a genus or family. A binomial distribution was used with a logit link function. We used AIC values to select the most parsimonious model from all possible combinations of predictors. GLME modelling was performed using function *glmer()* in R package *lme4*.

3.2 Comparison to existing Data Quality Limits (DQLs)

We used the data from the seven uncertainty plots to assess whether recommended data quality limits (DQLs) were based on reasonable expectations of experienced field staff. To do this, we compared the predictions of our quantified measurement error distributions (described above) with DOC’s current DQLs (Hawcroft et al. 2009). We build upon previous work that reviewed the audit procedure previously employed by the Ministry for the Environment (Affeld & Allen 2011) by comparing audit DQLs to actual field-quantified measurement error distributions. We focussed on objectives that were directly relevant to the stems (live and dead standing), the RECCE vegetation description, and the RECCE site components of the data (equivalent to field audit objectives 1-3, 5, 11-16 and 21-27 in Hawcroft et al. 2009), since these relate to botanical information derived from the RECCE and stem diameter data.

3.3 Framework to incorporate measurement error in National Biodiversity Monitoring and Reporting Measures

To incorporate measurement error uncertainty in assessing state and change of the two biodiversity monitoring and reporting measures we used a Monte Carlo simulation approach with 1000 simulations based on the observed measurement error distributions. In essence, we generated 1000 new datasets for each measurement period, where new estimates of our indicators were generated using uncertainty models developed on the seven remeasured plots. We describe the simulations in detail below.

3.3.1 Data sources

Estimates of measurement error uncertainty were applied to a random subset of 500 permanent forest plots on Public Conservation Lands (PCL) selected from the Tier One / LUCAS plot network. We started with the sample of forest plots used in Bellingham et al. (2014) to report on changes in tree size class structures on PCL. We discarded any plots that did not have two measurements of the RECCE data as we required both stem diameter and RECCE data for the uncertainty analyses. From the remaining plots, we randomly selected 500 (using the function *sample()* in R) as an objective sample of forest plots. This dataset was then used to evaluate the effect of measurement error uncertainty on confidence limits around the following two Tier 1 metrics.

3.3.2 M 5.1.1. Size-class structure of canopy dominants

Using the subset of 500 Tier 1 / LUCAS forest plots, we modelled changes in size structure as changes in plot-level mean diameter for the 26 most abundant species (Allen et al. 2013; Peltzer et al. 2014). This measure summarises the size-class structure of a species (or species group) and can be used to make inferences about population-level recruitment and mortality and hence, the maintenance of a species (or species group) (MacLeod et al. 2012; Bellingham et al. 2014; Peltzer et al. 2014). These analyses simulated errors in stem diameter measurements and species identification following the method of Holdaway et al. (2014). Diameter errors were simulated by first randomly sampling from a log-normal distribution of the co-efficient of variation (CV), with mean = $\log(0.0105)$ and SD = 0.8286 (Table 1). Using the sampled CV, stem-level diameter errors were obtained through a second random sampling from a normal distribution with mean of 0 and standard deviation equal to the random CV \times the observed diameter measurement. Thus absolute measurement error increased with increasing stem diameter, as per the observed error distribution. Simulated diameter measurements were then obtained by adding the randomly-generated error to the observed measurement.

Errors in species identification were simulated assuming that, on average, 2.18% (SD = 2.08%) of tagged stems per plot are misidentified (Table 1) (Holdaway et al. 2014). The simulated percentage of misidentified stems per plot was obtained through random sampling from a normal distribution with the total number of misidentified stems in a plot found by multiplying the randomly generated percentage by the number of stems observed. Misidentified stems were assigned the identity of a species occurring on the same plot from the same functional group, following Holdaway et al. (2014).

We calculated the change in mean plot-level stem diameter for the 26 most abundant species both with and without uncertainty.

3.3.3 M 5.2.1. Representation of plant functional types

Using the subset of 500 Tier 1 / LUCAS forest plots, we modelled changes in the species richness of groups that were palatable to introduced ungulates. Browsing pressure from introduced ungulates can reduce the occurrence of palatable species and hence this indicator is intended to report on maintenance of palatable species, relative to other species, across forest plots (Allen et al. 2013). New proportional richness values for each group were simulated using the GLM for uncertainty in proportional richness described in Section 3.1.2. To do this we first calculated the mean and standard error for predicted uncertainty for each observation (i.e. each palatability group in each plot) in the Tier 1 / LUCAS data. We then randomly sampled from a normal distribution defined by these parameters to obtain uncertainty. Finally, we randomly assigned the simulated uncertainty value as a positive or negative difference to the observed value.

3.3.4 Incorporating uncertainty in assessing state and change

We first examined the effect of uncertainty on the estimates of our test statistics, proportional richness of palatability groups (mean and standard deviation across plots) and plot-level species mean diameter (mean and standard deviation across plots) in each sampling period. To do this we documented estimates of mean and standard deviation across plots for each simulation, and calculated the mean of these values across simulations. We also calculated the mean upper and lower 90% confidence intervals for estimates of the mean (across simulations). This method is similar to that used by Holdaway et al. (2014) to simulate the effect of uncertainty on forest carbon stock estimates. For changes in mean plot-level diameter within the 26 most abundant species, we only used plots where the species was present at both measurements.

To assess the effect of uncertainty on power to detect change between measurement periods, we compared the mean and standard deviation of pairwise differences between plots in the observed data with those obtained when uncertainty was incorporated. This involved recording values for our test statistics for each plot in each simulation for both measurement periods. For each simulation, pairwise differences between measurement periods were calculated for each plot. This allowed the mean and standard deviation of pairwise differences to be calculated for each simulation. We then estimated confidence intervals for pairwise differences by taking the mean (across simulations) of the mean and standard deviation for pairwise differences. This allowed us to directly compare the confidence intervals obtained from the observed data (among plot variability only), with the confidence intervals obtained when the effect of measurement error was also included.

Using our simulation results, we ran a power analysis to identify and compare the minimum detectable effect size for the two measures, with and without the incorporation of measurement error. This analysis used a repeated measures design (paired t-test) with a power of 0.90, a significance level of 0.05, and standard deviations derived from either the observed or simulated data. All statistical analyses were conducted in R version 3.1.2 (R Development Core Team 2010).

4 Results

4.1 Measurement error distributions

Quantified distributions for sources of uncertainty are summarised in Table 1 for stem diameter, tree height, tree species identity, and the proportional richness of palatability groups. For species detection and species identity from the RECCE, we present the turnover among teams for palatability groups, growth forms and height tiers in Figure 2.

Table 1 Sources of uncertainty in carbon estimates and proportional richness of three palatability groups and their quantified distributions (adapted from Holdaway et al. 2014)

<i>Source of uncertainty</i>	<i>Parameter</i>	<i>Parameter distribution</i>	<i>Mean value</i>	<i>Standard deviation</i>
*Stem diameter (cm)	CV	Log-normal	-4.5543	0.8286
*Tree height (m)	CV	Log-normal	-3.1664	0.8356
*Species misidentified (N stems)	%	Normal	2.18%	2.08%
**Palatability group proportional richness				
Avoided	GLM	Quasi-binomial	0.029	0.017
Not selected	GLM	Quasi-binomial	0.024	0.013
Preferred	GLM	Quasi-binomial	0.017	0.016

* Taken from Holdaway et al. 2014.

** Parameter values are given for the observed response scale rather than the logit link function scale.

Turnover (Jaccard dissimilarity) of palatability groups was <30%, for all groups except the unclassified group (Figure 2). Turnover due to differences in richness (i.e. detection) was less than 20% for the avoided, not selected and preferred groups, but was greater than 20% for the unclassified group. Turnover due to name differences was >30% for the unclassified group.

Turnover (Jaccard dissimilarity) was >50% for forbs, graminoids and vines/other species (Figure 2). Turnover due to differences in richness (i.e. detection) was >30% for forbs and was >20% for graminoids and vines/other species. These high levels of turnover suggest either great variation in search effort or taxonomic resolution (i.e. lumping or splitting taxonomic concepts), or both. Ferns, graminoids, shrubs and vines all had > 20% species turnover due to name differences. Turnover in shrubs was driven equally by differences in richness and differences in names, while turnover in trees was driven more by differences in names.

In Tiers 6A and 7A, the name difference component of turnover was more important than the richness difference component, suggesting that misidentification is a larger problem than non-detection (Figure 2). However, turnover due to differences in richness was still 25% for Tier 7A and 20% for Tier 6A.

Turnover due to species richness was low in Tiers 3, 4 and 5, indicating that field teams reliably detect species in these tiers. However, turnover due to species richness was 25% in Tier 2 which might indicate that field teams miss species in the highest tiers (canopy and emergent canopy) or misjudge the height of species and thus assign species to the wrong tier. Finally, since turnover due to name differences is high in Tiers 5, 6 and 7a, efforts to standardise taxonomy would have greatest benefit if focussed on these tiers.

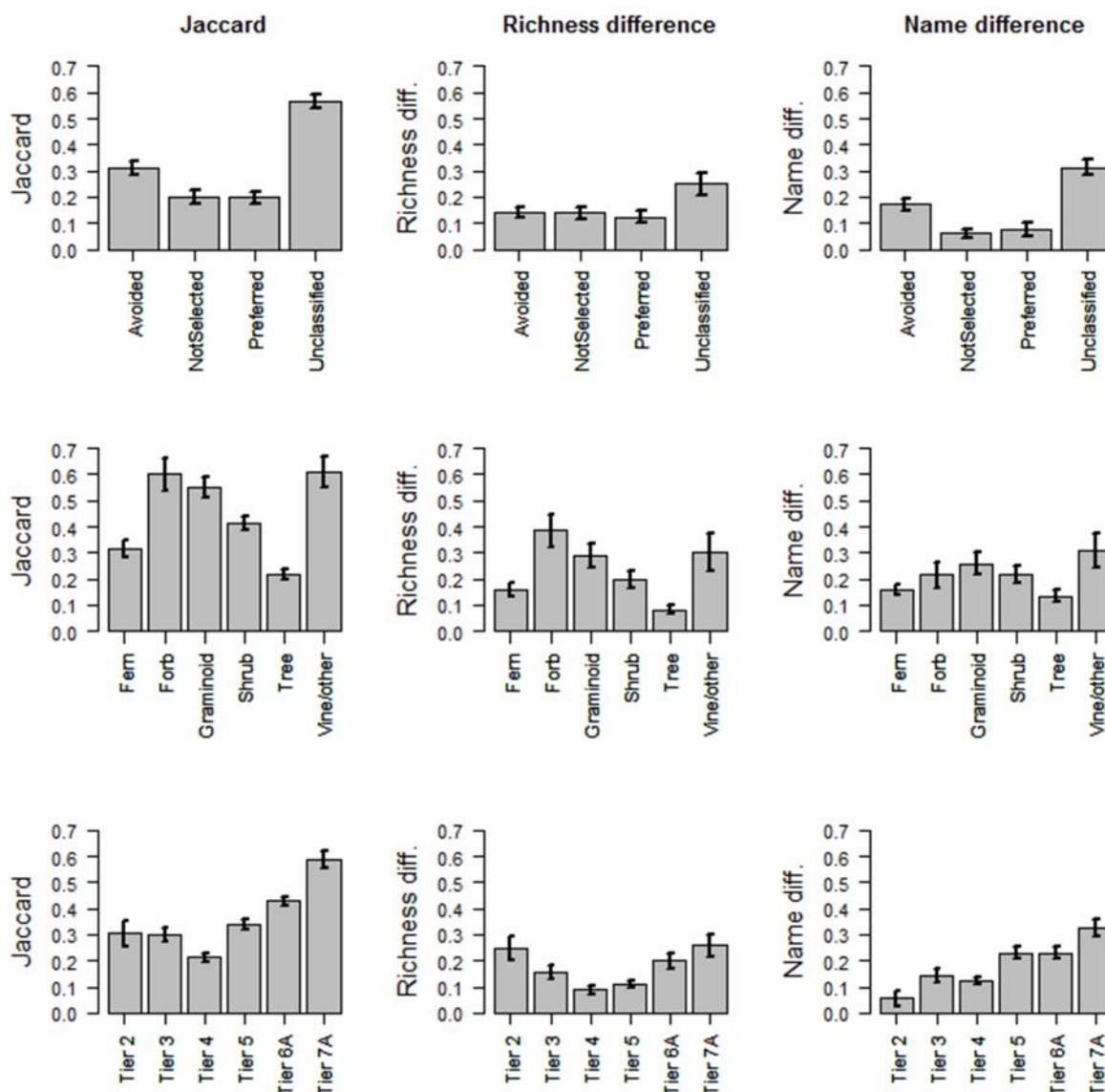


Figure 2 Jaccard dissimilarity (left column), and its components, Jaccard richness difference (middle column) and Jaccard name difference (right column), for each palatability group, growth form, and tier. Each observation represents a pairwise comparison of RECCE species lists for different teams on the same plot. Error bars show the standard error of the mean.

Height tier, cover class and growth form were retained in the most parsimonious GLME model predicting species non-detection. In general, species in Tier 6a and 7a, with a cover class of < 1% and belonging to the forb, graminoid or vine/other groups were the least likely

to be found in both samples in paired comparisons of different teams on the same plot (Figure 3). Conversely, trees over 5m with a cover class >1 % were highly likely to be found in both samples in pair comparisons of different teams (Figure 3). The best model predicted >50% of the variation in mean non-detection probability across tiers within plots (Figure 4).

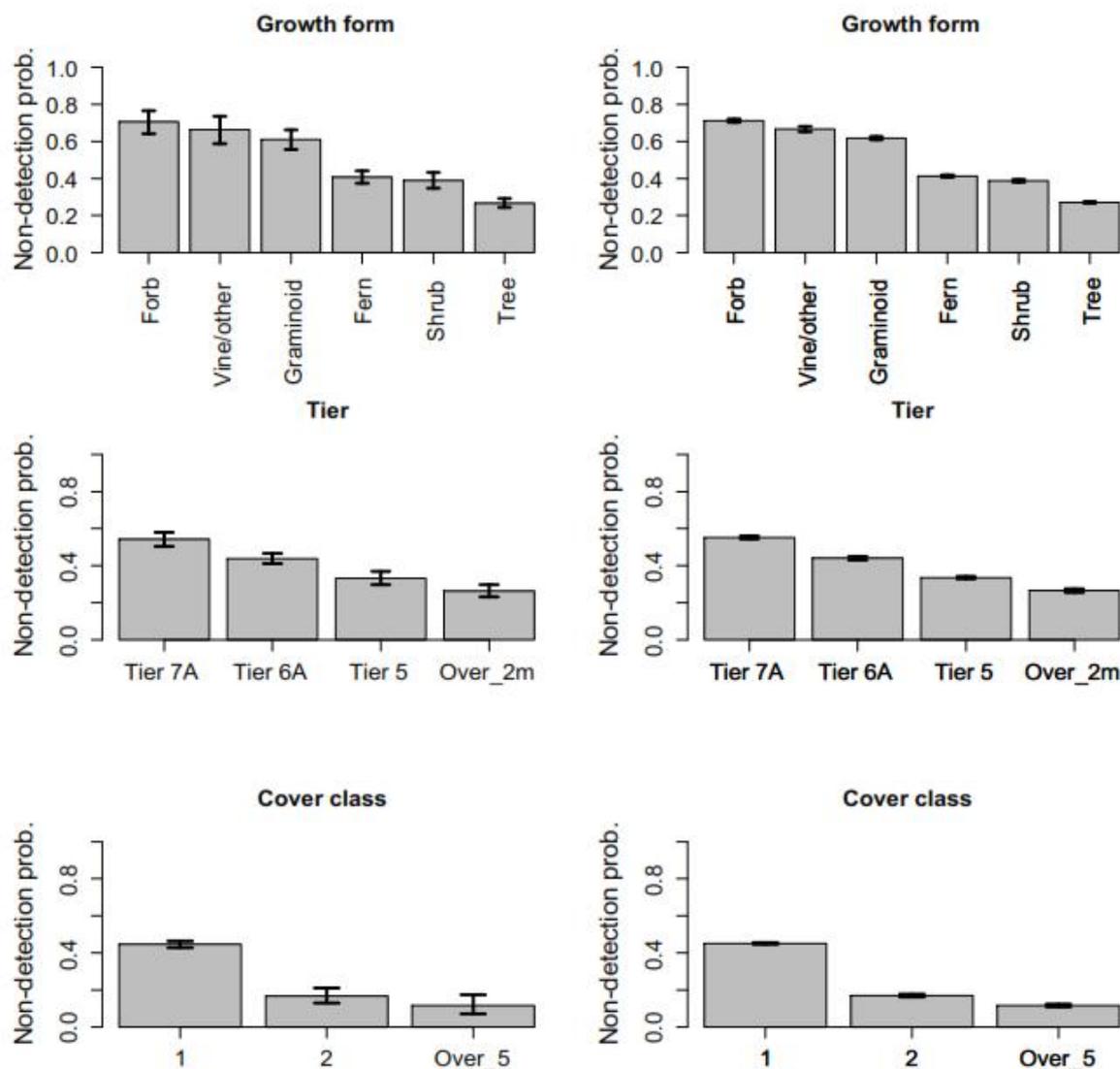


Figure 3 Observed ‘raw data’ (left hand column) and fitted (right hand column) non-detection probabilities for each growth form, height tier and cover class. Note that tier and cover class have been reclassified because there was minimal variation in non-detection probabilities for tiers above 2m and for cover classes above 5% (cover classes are shown as 1 (<1 %), 2 (2–5%) and over 5 (>5 %)). Each observation represents a pairwise comparison of RECCE species lists for different teams in the same tier on the same plot. Error bars show 95% confidence intervals for observed and fitted non-detection probabilities.

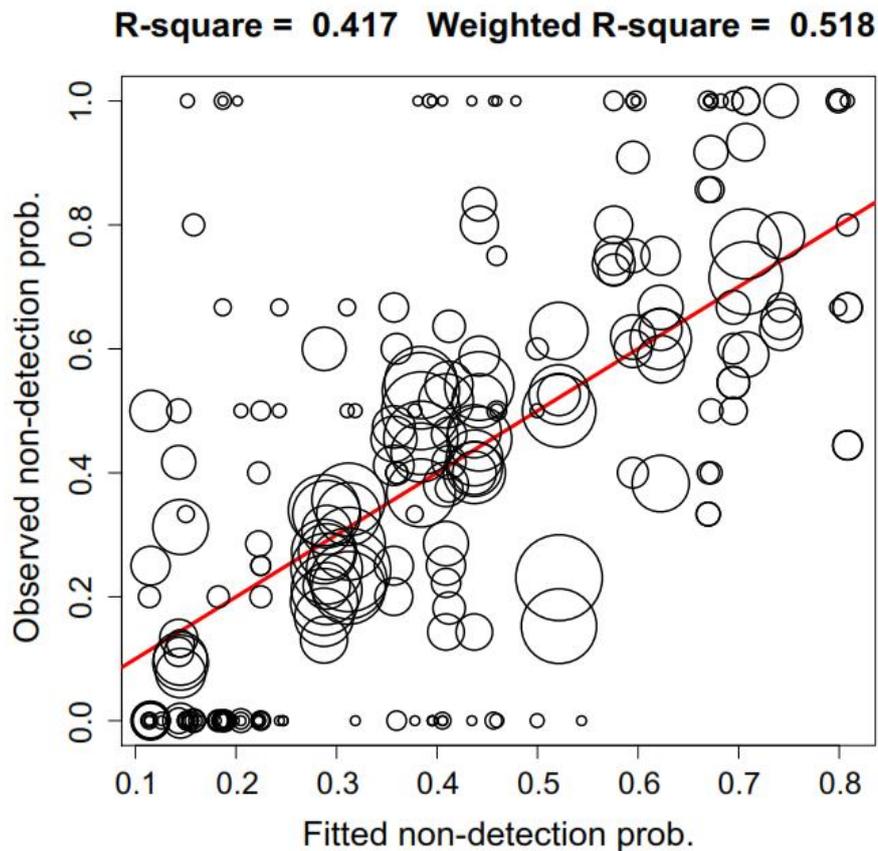


Figure 4 Mean observed and fitted non-detection probabilities of RECCE species lists for different teams in the same tier on the same plot. Each circle represents a comparison between two teams in a single tier in a single plot. The size of circles is proportional to species richness in each pairwise comparison (taken as the mean richness across the two teams). The red line represents a 1:1 relationship between observed and expected.

4.2 Data Quality Limits

Comparison of the observed error distributions against the current DQLs used by DOC revealed that many of the current limits appear to be narrower than the observed measurement error when entire plots are measured blind by independent field teams (Table 2). For example, the current DQL for stem diameter measurements for stems >10.0 cm is for 95% of the stems to be within $\pm 1\%$ of the diameter value (i.e. 0.2 cm for a 20.0 cm diameter stem), whereas the quantified measurement error distribution suggests 95% of the stem diameter measurements are within $\pm 5\%$. This discrepancy may result, in part, from the fact that our quantified measurement error distributions include all stems on the plot, whereas the field audit procedure makes allowances in the field for trees where it is genuinely difficult to get a standardised measure (Oliver Gansell, pers. comm.). Tree height measurements showed a similar pattern, with an observed 95% CI of $\pm 21.6\%$ compared to a current DQL of $\pm 10\%$ (Table 2).

The DQLs associated with the RECCE vegetation description were also generally much stricter than the observed measurement error distributions. In particular, the vascular species

identification DQL (DQL 14 in Table 2, 90% agreement of the plot-level vascular species list) was much stricter than the observed variability among the three fully-blind teams in this project ($63\% \pm 2\%$). In part, this may be because field audit teams are assessing a species list, whereas our measurements were fully-blind to the species recorded by other teams. Similarly, the DQLs for presence in height tiers were strict relative to measured error distributions from fully-blind teams: mean shared species within \pm one tier was 89% for Tiers 2–5 and 71% for Tier 6.

Table 2 Comparison of DOC forest plot measurement biodiversity data quality objectives with observed measurement error distributions

<i>DQL</i>	<i>Component</i>	<i>Variable</i>	<i>Reporting unit</i>	<i>Data quality range</i>	<i>Predicted range from observed measurement error distributions</i>	<i>Conclusion</i>
1	RECCE site	Slope	Nearest 1°	±5°	90% within ±5°	DQL OK
2	RECCE site	Aspect	Nearest 5°	±5°	24% within ±5°	May be too strict
3	RECCE site	Ground cover	Nearest 5%	±15%	92% within ±15% across 7 ground cover categories	DQL OK
5	RECCE site	Plot layout: horizontal distance	Nearest 0.1 m	±0.5 m	This DQL translates into maximum 5% error on total plot area. Observed 95% CI of total plot area was ±4.6%	DQL OK
11	RECCE vegetation description	Vascular species presence per height tier	Presence in tiers 1–5	95% agreement species are in ± 1 height tier	89% agreement species are in ± 1 height tier	May be too strict
12	RECCE vegetation description	Vascular species presence per height tier	Presence in tier 6	90% agreement species are in ± 1 height tier	71% agreement species are in ± 1 height tier	May be too strict
13	RECCE vegetation description	Cover class per vascular species per height tier	Cover class 1–6	95% agreement of species in ± 1 cover class within a tier	83–99 % agreement species in ± 1 cover class within a tier (on average across cover classes 1-5)†. Lowest for classes 3–5 (see Table 3).	May be too strict for some tiers
14	RECCE vegetation description	Vascular species identification	Lists as compiled in the field, or updated post field collection verification	90% agreement to species	63% agreement to species (on average)	May be too strict
16	Stems (live)	Species identification	Lists as compiled in the field, or updated post field collection verification	95% agreement to species	Mean difference 2.18% (97.82% agreement)	DQL OK

21	Stems (dead standing & alive)	Stem DBH 2.5–10 cm	Nearest 0.1 cm	±0.1 cm (currently being tested)	95% of stems ±5% of measured DBH	May be too strict
22	Stems (dead standing & alive)	Stem DBH 10.1–60 cm DBH	Nearest 0.1 cm	± 1% (currently being tested)	95% of stems ±5% of measured DBH	May be too strict
23	Stems (dead standing & alive)	Stem DBH > 60.1 cm DBH	Nearest 0.1 cm	±1%	95% of stems ±5% of measured DBH	May be too strict
25	Stems (dead standing & alive)	Stem height < 20 m	Nearest 0.1 m	±10%	95% of stems ±21.6%	May be too strict
26	Stems (dead standing & alive)	Stem height 20.1–30 m	Nearest 0.1 m	±10%	95% of stems ±21.6%	May be too strict
27	Stems (dead standing & alive)	Stem height > 30.1 m	Nearest 0.1 m	±10%	95% of stems ±21.6%	May be too strict

† Note that there were no cover classes of 6 recorded in the data used to assess these DQLs.

For cover classes 1 and 2, there was a 99% chance of obtaining a cover score within ± 1 cover class of the observed, which satisfies the DQL of 95% (Table 3). However, all other cover classes (3-5 in these data) were below 95%, suggesting the field audit standards may be too strict.

Table 3 Confusion matrix showing the percentage chance of a species being assigned a second cover score given a first cover score. The percentage of observations that are within one cover class either side of the other observed cover class is also presented (% within ± 1 class)

		<i>First cover score</i>				
		0.5	3	15.5	38	63
<i>Second cover score</i>	0.5	88	39	8	2	0
	3	11	46	33	4	0
	15.5	1	14	52	26	17
	38	0	0	6	62	33
	63	0	0	1	5	50
<i>% within ± 1 class</i>		99	99	91	93	83

4.3 Incorporation of uncertainty into Tier 1 measures

4.3.1 Tree size class structures by species

Including measurement error increased the uncertainty around the mean estimate of change in mean stem diameter by 10.2% on average across all species (Table 4), but there was large variability across species. There was no obvious pattern in the effect of measurement error on uncertainty. For instance, there was no trend for species occurring in fewer plots to be more strongly affected by measurement error. Importantly, the inclusion of measurement error altered the significance of a paired t-test testing for a change in mean diameter between measurements. Mean diameter differed significantly in eight species based on the observed mean change and observed standard deviation. However, four of those eight species were not significant once measurement error was included. This result is critical; analyses based on the observed data would have drawn false conclusions on the demographic processes of four common tree species.

Table 4 Observed and simulated means and standard deviations of changes in plot-level mean stem diameter for the 26 most widespread tree species. Observed values include sampling uncertainty only, whereas simulated values include both sampling uncertainty and measurement error. Species are presented in three groups according to whether or not there was a significant difference in mean diameter (see footnote), and then by the number of plots in which the species was present

<i>Species code</i>	<i>N plots present</i>	<i>Mean observed change in mean stem diameter over 7 years (cm)</i>	<i>Standard deviation of observed mean (cm)</i>	<i>Simulated mean change in mean stem diameter over 7 years (cm)</i>	<i>Standard deviation of simulated mean (cm)</i>	<i>Change in uncertainty when measurement error included (%)</i>
CARSER*	132	0.50	1.82	0.48	2.00	9.9
PSECRA*	120	0.45	1.55	0.44	1.73	11.6
MELRAM*	119	0.32	1.44	0.31	1.55	7.6
KUNERI*	61	0.89	2.73	0.89	2.78	1.8
NOTSOL†	106	0.56	2.69	0.54	2.99	11.2
COPFOE†	87	0.13	0.56	0.12	0.66	17.9
RAUSIM†	63	0.38	1.47	0.38	1.59	8.2
KNIEXC†	63	0.59	2.05	0.56	3.09	50.7
WEIRAC	235	0.25	2.77	0.24	2.91	5.1
NOTMEN	178	0.36	4.14	0.34	4.45	7.5
DICSQU	137	-0.02	1.14	-0.02	1.25	9.6
CYASMI	130	-0.02	1.45	-0.02	1.51	4.1
GRILIT	129	0.2	2.13	0.20	2.41	13.1
PSECOL	116	0.02	0.64	0.03	0.77	20.3
PODCUN	114	0.12	2.63	0.11	3.05	16.0
NOTFUS	97	-0.48	11.76	-0.46	12.15	3.3
MYRDIV	89	0	0.67	0	0.70	4.5
CYADEA	87	0.16	1.63	0.17	1.76	8.0
BEITAW	84	-0.02	5.4	-0.01	5.54	2.6
HEDARB	82	0.27	1.3	0.27	1.72	32.3
PRUFER	80	0.14	4.03	0.13	4.38	8.7
DACCUP	80	-0.92	9.2	-0.92	9.56	3.9
QUISER	78	0.03	1.99	0.03	2.06	3.5
METUMB	74	-1.40	11.49	-1.41	11.94	3.9
LEUFAS	67	0.10	0.62	0.10	0.63	1.6
LEPSCO	66	-0.21	2.93	-0.21	2.92	-0.3

*Significant difference in mean diameter when measurement error is included in estimating SE.

†Significant difference in mean diameter using standard paired t-test but not when measurement error was included.

4.3.2 Palatability group proportional species richness

Including measurement error increased the uncertainty around the mean estimate of net change in proportional species richness by 23% for preferred species, 37% for avoided species, 40% for unclassified species, and 141% for the not selected group (Table 5). Model simulations were generally not biased; however simulations overestimated mean proportional richness for the not selected group. This is probably because values for this group were strongly zero-inflated (most plots had no species in this group), making it difficult to model from available distribution types. Indeed, the confidence intervals in predicted values for this group were very large.

Based on the random subset of 500 Tier One / LUCAS plots used in this report, the proportional richness of avoided species increased between measurement periods and this increase was statistically significant based on the observed data (mean net difference = 1.24, 95% CI = 0.82 to 1.66) and after the effect of measurement error was included (mean net difference = 1.24 95% CI with measurement error = 0.66–1.82) (Table 5).

Table 5 Observed means and standard deviations of observed and simulated means of proportional palatability group richness. Observed values include sampling uncertainty only, whereas simulated values include both sampling uncertainty and measurement error

<i>Group</i>	<i>Observed mean change in proportional richness (%)</i>	<i>Standard deviation of observed mean (%)</i>	<i>Standard deviation of simulated mean (%)</i>	<i>Change in uncertainty when measurement error included (%)</i>
Preferred	-0.22	3.67	4.52	+23
Avoided	1.24	4.81	6.57	+37
Not selected	0.06	0.78	2.65	+141
Unclassified	-0.28	6.17	8.66	+40

The increase in uncertainty when measurement error is included translates into an equivalent increase in the minimum detectable effect size at any given level of replication. This is illustrated for the preferred group in Figure 5, which presents the results of a power analysis based on the observed and simulated standard deviations. The decrease in power due to measurement error with a sample size of 750 plots is roughly the same magnitude as the increase in power obtained by increasing the replication from 750 to 1250 plots.

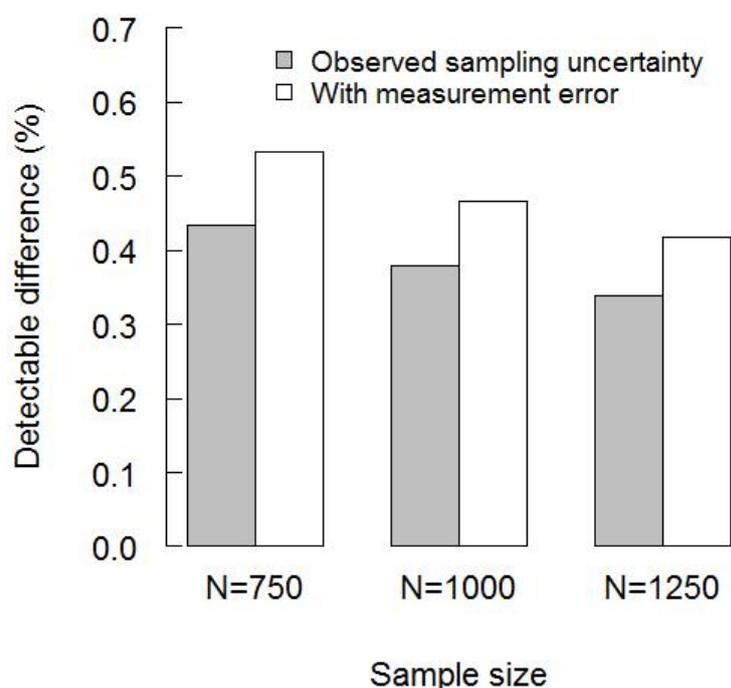


Figure 5 Effect of sample size and inclusion of measurement error on the minimum detectable change in proportional richness (%) for the preferred palatability group. Results apply at the national scale assuming a 7-year measurement interval, and are based on power analysis (`power.t.test` in R) with power = 0.9, significance level = 0.05, and standard deviations from Table 5.

5 Discussion

This report has demonstrated the importance of quantifying and integrating measurement error into biodiversity measures. Our data clearly demonstrate that measurement error is an unavoidable component of biodiversity data: highly-qualified field teams differed in the identity and abundance scores given to species in the field. Measures of community-level species composition are particularly vulnerable to measurement error, which in turn can limit the power to detect change through time. However, we demonstrate that key monitoring and reporting measures are robust to this measurement error when applied to a large plot network with high statistical power. Our analyses demonstrate that inclusion of uncertainty can alter the statistical significance of changes in measures, underscoring the need to account for uncertainty in biodiversity reporting in order to minimise the risk of Type I errors (i.e. reporting a false change in a measure).

5.1 Data quality limits

Our error distributions are larger than those of the DQLs in many instances. However, due to the nature of the field audit procedure, this is not necessarily a reason to change the field audit DQLs. Some stems are excluded from the DQL calculation in the field audit due to exceptional circumstances (e.g. misshapen and difficult to measure). In our calculations, all stems on a plot are included in the measurement error distributions. Our data are also for ‘blind’ variability between measurements. In contrast, the field audit teams have the most

recent measurement data that they are auditing, and so can recheck their measurements if they vary from the observed. Thus our error distributions are not directly transferable into DQLs. Alignment of our error distributions with the DQLs would require the field audit procedure to be changed to a fully-blind whole plot measurement. However, the current field audit procedure has two advantages over a fully-blind whole plot measurement: firstly, it is much quicker, and secondly, it provides scope for the field audit team to identify how and why errors occurred, and this maximises the opportunity to provide specific feedback to field teams. In light of this, the current procedure is the most useful for field teams and training purposes, although the actual measurement errors are likely to be larger than the current DQLs.

Detection probabilities vary according to the type of plant and the traits of that species (this report; Garrad et al. 2013) and are inversely proportional to abundance (McCarthy et al. 2012; this report). Clustered individuals are more easily detected, even at low abundance, than scattered individuals (McCarthy et al. 2012). Interestingly, we found no evidence that species were harder to detect if they were from a species-rich genus or family. Detection was lowest in forbs and grasses which reflect that many field botanists are less confident at identifying these groups, relative to other growth forms. Detection probabilities increase with the amount of time spent looking, particularly for scattered or rare species (Garrad et al. 2008) and thus one way of increasing detection is to allocate sufficient time to structured searching by expert field botanists. However, an important finding from our study is that measurement error is an unavoidable component of biodiversity data. While training and adequate resourcing can reduce error, they cannot eliminate it. The solution is minimise measurement error and to quantify and integrate its impact into reported measures, as has been done here.

5.2 Consequences of measurement error for biodiversity measures

There was an almost universal increase in the standard error of mean values of biodiversity measures when including uncertainty. Our analyses have demonstrated that inclusion of uncertainty can alter the statistical significance of changes in measures (specifically, changes in the mean diameter of common tree species). This strongly suggests that uncertainty should be accounted for in all Tier 1 measures to minimise the risk of Type I errors, that is, reporting a false change in a measure.

Our measures of uncertainty include three sources of measurement error: (1) a plant was not detected (2) a plant was detected by all teams but each team gave that plant a different name and (3) each team established the plot with different boundaries and hence plants were not included equally by all teams. Training can reduce each of these three sources of error, although it is widely recognised that detection is imperfect, even by the most experienced and highly-trained field botanists (e.g. McCarthy et al. 2012). Additional progress beyond comprehensive training can be made by quantifying detection probabilities and integrating those into measures and indicators.

Recent work in Australia (Garrad et al. 2008; 2013; in press) has used data on the time taken for trained botanists to find known individuals in 1 ha plots to quantify the probability of a given survey effort detecting a species' presence or true absence, and demonstrated how these probabilities can be incorporated into environmental impact assessments. Similar procedures could be adopted in New Zealand, particularly for rare species currently proposed as indicator

species (e.g. Monks et al. 2013). The measures in this report either use common species (e.g. mean diameter of common tree species) or aggregate across species groups (e.g. the proportion of palatable species) and these are likely to be less vulnerable to detection errors than measures based on single species with limited distributions (e.g., the distribution and abundance of *Anisotome haastii* as an indicator of browse pressure in alpine terrestrial habitats, Monks et al. 2013).

The non-detection model developed in this report successfully predicted non-detection in >50% of instances (Figure 4). This model could be applied to all species in the flora to predict the probability of non-detection. These probabilities could be used to guide training and field effort towards species with a high probability of non-detection. Broadly, these are likely to be forbs, graminoids, vines or unclassified groups that are commonly found with a low cover. However, we emphasise that our detection probabilities are based on a sample of seven forest plots from a geographically-restricted area (western North Island) and a more robust sample of forests, and data from non-forest ecosystems would be highly desirable to extend this uncertainty framework to all ecosystems, nationally. Lastly, standardising nomenclature in forbs and graminoids could reduce uncertainty as the turnover in these groups was high (>20%). While some of this turnover will be due to detection, a component will be due to variation among teams in whether taxonomic concepts are ‘split’ or ‘aggregated’, and this source of error can be reduced through use of standardised floras.

Species richness increased across the sample of 500 forest plots between the first and second measurements (Figure 1). This artefact of biodiversity sampling is well-established but presents challenges when reporting on temporal trends in richness. The solution applied in this report was to relativise richness (e.g. the proportion of avoided species, rather than absolute richness of avoided species) and we recommend similar solutions for other measures based on temporal trends in richness.

The percentage increase in uncertainty around proportional richness in palatability groups was similar to the percentage increase in uncertainty in net carbon change estimates (+35%, Holdaway et al. 2014). However, the significance of this increase depends on the size of the uncertainty relative to the expected effect size. Encouragingly, our sample of 500 plots detected a significant increase in the proportion of avoided species, even though the size of this effect was small (+1.24%). Likewise, differences of < 0.6% are detectable for the preferred group. Given the turnover in compositional data from the RECCE (Figure 2), it is reassuring that measures which aggregate across palatability groups can detect such small differences, underscoring the value of a large plot network (with > 500 plots), and the advantages of measures which are based on many species. Based on our estimates of species turnover, we suggest that measures based on few plots, few species, a specific species, species-turnover or on total richness are likely to be unreliable. Our framework could be applied to other biodiversity measures, but further data would be needed, both from non-forest situations and from a wider sample of forest situations, and for other data sources used (e.g. pellet counts) to estimate ecological integrity.

6 Recommendations

1. We recommend incorporating measurement uncertainty into monitoring and reporting of biodiversity measures across Public Conservation Lands.
2. In order to incorporate measurement uncertainty into monitoring and reporting of biodiversity measures across all Public Conservation Lands, estimates of measurement uncertainty are required from a larger sample of forest ecosystems and from non-forest ecosystems. We recommend quantifying measurement uncertainty in a random sample of 15 Tier One forest plots and 15 Tier One non-forest and shrubland plots to provide a nationally-robust sample of biodiversity uncertainty.
3. We have demonstrated that measures vary widely in their measurement uncertainty. We recommend extending this work to assess the uncertainty in all measures used to report on ecological integrity. A logical starting point would be the measures that are used to report on the composite index of indigenous dominance, namely, pest animal pellet counts, possum trap catch index, bird count data, and cover and richness of exotic plant species.
4. We recommend that indices of species turnover and measures based on raw species richness be avoided in Monitoring and Reporting programmes, due to the effect of (unavoidable) measurement error.
5. We recommend using the RECCE data to determine the consequences of measurement error on classification robustness using the Wisser et al. (2011) classification.
6. We recommend applying the uncertainty framework to existing data from experimental settings (e.g. exclosure studies) to test whether reported effects are statistically robust once measurement uncertainty is accounted for.
7. We recommend quantifying detection probabilities for a representative suite of the indicator species proposed by Monks et al. (2013) across a range of growth forms, assumed abundance and habitats, and developing a framework for reporting on those species with uncertainty (specifically false absences).
8. Measurement error is an unavoidable component of biodiversity data. We recommend a two-stage solution that first minimises error through comprehensive training programmes, and second, accommodates remaining error through quantification and integration of measurement error into reported measures.

7 Acknowledgements

We thank Meredith McKay and Elaine Wright at the Department of Conservation for their support; and Duane Peltzer and Peter Bellingham for peer review.

8 References

- Affeld K, Allen RB. 2011. Review of LUCAS Natural Forest plot measurement programme 2009–2010. Landcare Research Contract Report LC 245. Prepared for the Ministry for the Environment, Wellington.
- Allen RB, Wright EF, MacLeod CJ, Bellingham PJ, Forsyth DM, Mason NWH, Gormley AM, Marburg AE, MacKenzie DI, McKay M. 2013. Designing an inventory and monitoring programme for the Department of Conservation’s Natural Heritage Management System. Landcare Research Contract Report LC1730 prepared for the Department of Conservation.
- Baker TR, Phillips OL, Malhi Y, Almeida S, Arroyo L, Di Fiore A, Erwin T, Higuchi N, Killeen TJ, Laurance SG, Laurance WF, Lewis SL, Monteagudo A, Neill DA, Vargas PN, Pitman NCA, Silva JNM, Martinez RV. 2004. Increasing biomass in Amazonian forest plots. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 359: 353–365.
- Bellingham PJ, Richardson SJ, Gormley AM, Husheer SW, Monks A. 2014. Department of Conservation biodiversity indicators: 2014 assessment. Landcare Research Contract Report LC1957 prepared for the Department of Conservation.
- Bellingham PJ, Richardson SJ, Gormley AM, MacLeod CJ, Forsyth DM. 2013. Department of Conservation biodiversity indicators: 2013 assessment. Landcare Research Contract Report LC1621. Landcare Research Contract Report LC1621 prepared for Department of Conservation.
- Bolker B. 2008. *Ecological models and data in R*. Princeton, NJ: Princeton University Press.
- Butt N, Slade E, Thompson J, Malhi Y, Riutta T. 2013. Quantifying the sampling error in tree census measurements by volunteers and its effect on carbon stock estimates. *Ecological Applications* 23: 936–943.
- Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, Narwani A, Mace GM, Tilman D, Wardle DA, Kinzig AP, Daily GC, Loreau M, Grace JB, Larigauderie A, Srivastava DS, Naeem S, 2012. Biodiversity loss and its impact on humanity. *Nature* 486: 59–67.
- Chave J, Condit R, Aguilar S, Hernandez A, Lao S, Perez R. 2004. Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 359: 409–420.
- Clark DA. 2002. Are tropical forests an important carbon sink? Reanalysis of the long-term plot data. *Ecological Applications* 12: 3–7.
- Clark DB, Kellner JR. 2012. Tropical forest biomass estimation and the fallacy of misplaced concreteness. *Journal of Vegetation Science* 23: 1191–1196.
- Costanza R, d’Arge R, de Groot R, Farber S, Grasso M, Hannon B, Limburg K, Naeem S, O’Neill RV, Paruelo J, Raskin RG, Sutton P, van den Belt M. 1997. The value of the world’s ecosystem services and natural capital. *Nature* 387: 253–260.

- Department of Conservation. 2014. Field protocols for DOC Tier 1 Inventory & Monitoring and LUCAS plots; Department of Conservation.
- Forsyth DM, Coomes DA, Nugent G, Hall GMJ. 2002. The diet and diet preferences of ungulates (Order: Artiodactyla) in New Zealand. *New Zealand Journal of Zoology* 29, 323–343.
- Garrard GE, Bekessy SA, McCarthy MA, Wintle BA. 2008. When have we looked hard enough? A novel method for setting minimum survey effort protocols for flora surveys *Austral Ecology*, 33, 986–998.
- Garrard GE, Bekessy SA, McCarthy MA, Wintle BA. in press. Incorporating detectability into environmental impact assessment for threatened species. *Conservation Biology*. doi: 10.1111/cobi.12351
- Garrard GE, McCarthy MA, Williams NSW, Bekessy SA, Wintle BA. 2013. A general model of detectability using species traits. *Methods in Ecology and Evolution*, 4, 45–52.
- Hawcroft A, McKay M, Kusabs A. 2009. DOC DM-1115350 – DOC I&M/LUCAS Internal Field Protocol. Internal DOC Report.
- Holdaway R, McNeill S, Mason NH, Carswell F 2014. Propagating uncertainty in plot-based estimates of forest carbon stock and carbon stock change. *Ecosystems* 17: 627–640.
- Lee WG, McGlone MS, Wright EF. compilers 2005. Biodiversity Inventory and Monitoring: A review of national and international systems and a proposed framework for future biodiversity monitoring by the Department of Conservation. Landcare Research Contract Report LC0405/122. 216 p.
- MacLeod CJ, Affeld K, Allen RB, Bellingham PJ, Forsyth DM, Gormley AM, Holdaway RJ, Richardson SJ, Wisser SK. 2012. Department of Conservation biodiversity indicators: 2012 assessment. Landcare Research Contract Report LC1102, prepared for Department of Conservation, Christchurch.
- McCarthy MA, Moore JL, Morris WK, Parris KM, Garrard GE, Vesik PA, Rumpff L, Giljohann KM, Camac JS, Bau SS, Friend T, Harrison B, Yue B. 2013. The influence of abundance on detectability. *Oikos* 122: 717–726.
- MfE (Ministry for the Environment). 2011. Land Use and Carbon analysis system Natural Forest Data collection Manual. In: Ministry for the Environment Wellington, New Zealand.
- Monks JM, O'Donnell CFJ, Wright EF. 2013. Selection of potential indicator species for measuring and reporting on trends in widespread native taxa in New Zealand. DOC Research and Development Series 338. Department of Conservation, Wellington. 18 p.
- Muller-Landau HC, Detto M, Chisholm RA, Hubbell SP, Condit R. 2013. Detecting and projecting changes in forest biomass from plot data. Coomes DA, Burslem DFRP, Simonson WD, editors. *Forests and global change*. Cambridge University Press, Cambridge, UK. Pp. 381–415.

- Payton IJ, Newell CL, Beets PN. 2004. New Zealand Carbon Monitoring system indigenous forest and shrubland data collection manual. Christchurch, New Zealand: Caxton Press for the Ministry for the Environment. 67 p.
- Peltzer DA, Mason NWH, Bellingham PJ, Richardson SJ, Knightbridge P, Wright EF, Allen RB. 2014. Disentangling drivers of tree population size distributions. *Forest Ecology and Management* 331: 165–179.
- Pereira HM, Ferrier S, Walters M, Geller GN, Jongman RHG, Scholes RJ, Bruford MW, Brummitt N, Butchart SHM, Cardoso AC, Coops NC, Dulloo E, Faith DP, Freyhof J, Gregory RD, Heip C, Höft R, Hurtt G, Jetz W, Karp DS, McGeoch MA, Obura D, Onoda Y, Pettorelli N, Reyers B, Sayre R, Scharlemann JPW, Stuart SN, Turak E, Walpole M, Wegmann M. 2013. Essential biodiversity variables. *Science* 339:277–278.
- Phillips OL, Malhi Y, Higuchi N, Laurance WF, Nunez PV, Vasquez RM, Laurance SG, Ferreira LV, Stern M, Brown S, Grace J. 1998. Changes in the carbon balance of tropical forests: Evidence from long-term plots. *Science* 282:439–442.
- R Development Core Team. 2010. R: A language and environment for statistical computing, Vienna, Austria, R Foundation for Statistical Computing.
- Yanai RD, Battles JJ, Richardson AD, Rastetter EB, Wood DM, Blodgett C. 2010. Estimating uncertainty in ecosystem budget calculations. *Ecosystems* 13: 239–248.