Canadian Journal of
Forest Research

Measurement uncertainty in a national forest inventory: results from the northern region of the USA

Ruth D. Yanai[a], **Alexander R. Young**[a], **John L. Campbell**[b], **James A. Westfall**[c], **Charles J. Barnett**[b], **Gretchen A. Dillon**[a], **Mark B. Green**[b,d], **and Christopher W. Woodall**[b]

[a]Department of Sustainable Resources Management, SUNY College of Environmental Science and Forestry, Syracuse, NY 13210, USA; [b]Northern Research Station, USDA Forest Service, Durham, NH 03824, USA; [c]Northern Research Station, USDA Forest Service, York, PA 19073, USA; [d]Department of Earth, Environmental, and Planetary Sciences, Case Western Reserve University, Cleveland, OH 44106, USA

Corresponding author: **Ruth Yanai** (email: rdyanai@syr.edu)

Abstract

Statistical confidence in estimates of timber volume, carbon storage, and other forest attributes depends, in part, on the uncertainty in field measurements. Surprisingly, measurement uncertainty is rarely reported, even though national forest inventories routinely repeat field measurements for quality assurance. We compared measurements made by field crews and quality assurance crews in the Forest Inventory and Analysis program of the U.S. Forest Service, using data from 2790 plots and 51 740 trees and saplings across the 24 states of the Northern Region. We characterized uncertainty in 12 national core tree-level variables; seven tree crown variables used in forest health monitoring; three variables describing seedlings; and 11 variables describing the site, such as elevation, slope, and distance from a road. Discrepancies in measurement were generally small but were higher for some variables requiring judgment, such as tree class, decay class, and cause of mortality. When scaled up to states, forest types, or the region, uncertainties in basal area, timber volume, and aboveground biomass were negligible. Understanding all sources of uncertainty is important to designing forest monitoring systems, managing the conduct of the inventory, and assessing the uncertainty of forest attributes required for making regional and national forest policy decisions.

Key words: carbon accounting, Forest Inventory and Analysis, forest inventory, measurement uncertainty, forestry

1. Introduction

Forest inventories have historically been performed for many purposes (Labau et al. 2007), including estimating merchantable timber, determining risk of fire, evaluating habitat for wildlife, and assessing biodiversity. Forest inventories are increasingly important to forest carbon monitoring both for field-derived baselines (Woodall et al. 2015) and for development of carbon monitoring via remote sensing (Harris et al. 2016). In addition to traditional measurements of live and standing dead trees, other important forest attributes can be monitored, such as downed dead wood, soil carbon, invasive understory species, tree regeneration, and forest health.

Quantifying uncertainty in forest inventories is important to applications ranging from stand management to national and global scales. Reporting uncertainties associated with forest resource inventories is required for greenhouse gas accounting, for example, by the United Nations Framework Convention on Climate Change (Penman et al. 2000, 2003) and REDD+ (Reducing Emissions from Deforestation and forest Degradation plus the sustainable management of forests and the conservation and enhancement of forest carbon stocks) programs (Yanai et al. 2020). Clearly, making programmatic improvements to reduce uncertainty requires information about the magnitude of the sources that contribute to it.

The uncertainty of forest inventory estimates stems from multiple sources: spatial variation (sampling error), model fit, model selection, and uncertainty in measurements (Cunia 1987; Schreuder et al. 1993; Köhl et al. 2006). Measurement uncertainty typically receives less attention than sampling or model error (Chave et al. 2004; Hunter et al. 2013; Berger et al. 2014). We broadly define measurement uncertainty to include typical forest inventory data such as tree species, tree classification (e.g., crown class, tree class, tree grade, and decay class), and judgments that affect whether a tree should be included in the inventory, such as whether it falls within the plot or meets size specifications. Measurement uncertainty stems from sources that include incorrect use of instruments and mistakes in data entry, which can be reduced with training but not eliminated completely (Butt et al. 2013; Larjavaara and Muller-Landau 2013).

The Forest Inventory and Analysis (FIA) program of the U.S. Department of Agriculture, Forest Service has conducted and maintained a comprehensive inventory of U.S. forests since the 1930s (Bechtold and Patterson 2005; Woudenberg et al. 2010; USDA 2018). Like other national forest inventory programs, the FIA uses a Quality Assurance (QA) program to achieve acceptable standards of consistency and quality throughout the workflow of the inventory program, from plot selection to analytical reporting (Pollard et al. 2006). To assess the repeatability and quality of field measurements, a subset of field measurements is remeasured by QA field crews, who have extensive supervisory, training, and field experience, ideally within 3 weeks of the original measurements, on at least 4% of plots (Gormanson et al. 2017; Pollard et al. 2006; USDA 2019). These are conducted as "blind checks", meaning that the QA crew does not view the measurements conducted during the earlier plot visit by the field crew. Although the primary purpose of the QA program has been to evaluate and improve field crew performance, the difference between the two measurements is a rich source of data for quantifying uncertainty in a wide variety of forest measurements.

The goal of this study was to assess the uncertainty of forest inventory measurements in the 24 states of the Northern Region of the FIA program. We compared data collected from field crews and QA blind checks for one inventory cycle (2011–2016) involving 2790 plots, 51 740 trees and saplings of 145 species, and counts of 39 301 seedlings from 19 301 seedling plots. We characterized uncertainty in 12 national core tree-level variables, including diameter, height, and species identification; seven tree crown variables used in forest health monitoring; three variables describing seedlings; and 11 variables describing the site, such as elevation, slope, and distance from a road. Finally, we aggregated tree-level measurement differences to the plot level and larger spatial scales, to evaluate the importance of measurement uncertainty at scales relevant to forest inventory and management.

2. Material and methods

2.1. Data collection

FIA plots are systematically distributed across the conterminous United States, with a national base sample intensity of approximately one FIA plot per 2400 ha (Woudenberg et al. 2010); some states and national forests have invested in greater intensities. These plots are remeasured on a cycle of 5–7 years in the Northern and Southern Regions of the eastern United States (10 years in the western United States). This study was limited to forests of the Northern Region, which spans 70 million ha across 24 states with the FIA inventory comprising more than 40 000 forested plots and nearly 1.4 million sampled trees (USDA 2021). We used data from a recent inventory cycle from 2011 to 2016 with measurements occurring from October 2010 to April 2017, depending on the state.

The FIA plot design consists of a cluster of four fixed-area circular subplots, each with a radius of 7.3 m (Bechtold and Scott 2005; USDA 2018). Trees at least 12.7 cm (5 in.) in diameter at breast height (DBH) or, for woodland species (Burrill et al. 2021), at the root collar (DRC), are measured. We assessed data from all 2790 plots that were visited by both QA and field crews and that included at least one tree, sapling, or tree seedling (Fig. 1). In total, 51 740 trees and saplings had at least tree status and species reported by both crews. Additional variables related to forest health are measured on a subset of plots during the growing season (e.g., tree crown health) (USDA 2005; Woodall et al. 2011*a*), providing 1780 plots for our analysis.

Each subplot contains a circular microplot with a radius of 2.1 m on which sapling and seedling data are collected. Saplings are defined as having a DBH or DRC of at least 2.5 cm (1 in.) and less than 12.7 cm (5 in.), while seedlings are smaller but at least 0.15 m in height for conifers or 0.3 m for hardwoods (USDA 2018). The QA crew evaluated a subset of these microplots depending on the time available and their judgment of the likelihood of field crew measurement issues. There were 3862 microplots evaluated by both crews.

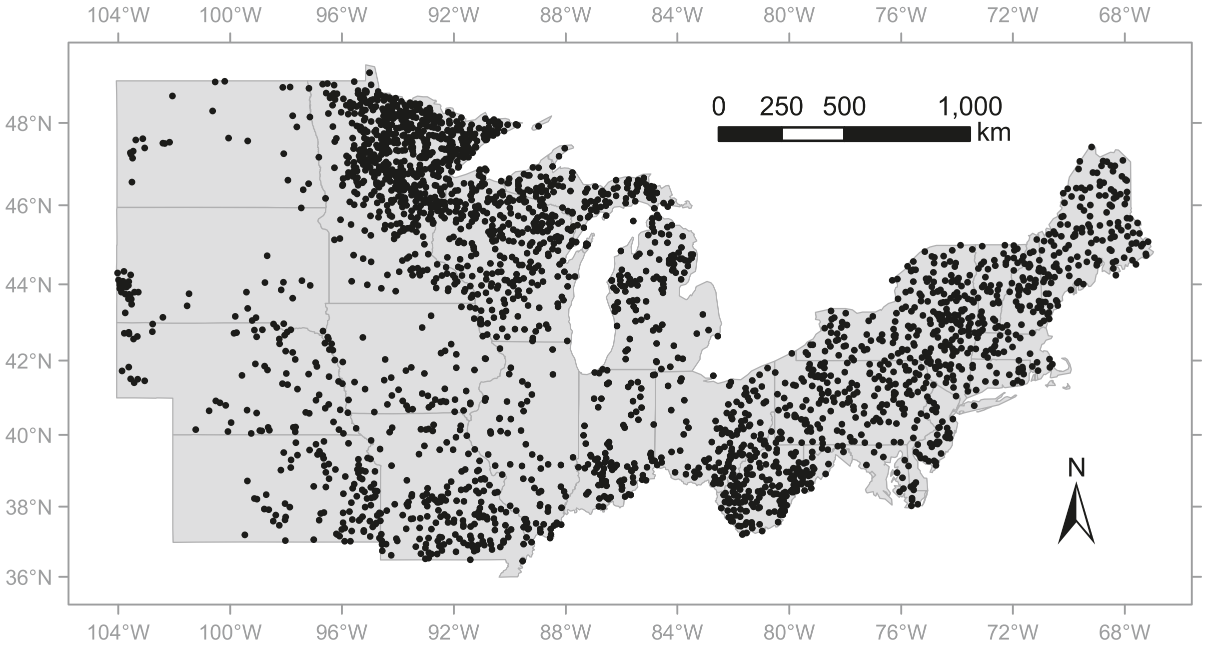
The time between visits for field and QA crews ranged from 0 to 507 days, with a median of 25 days and an interquartile range from 7 to 61 days.

National core tree-level measurements collected were tree status, species, diameter, height, broken height, rotten and missing cull, tree class code, crown class code, compacted crown ratio, tree grade code, decay code, standing dead, and mortality agent (Table S1 in the Supporting Information). Tree crown measurements collected on a subset of plots for forest health monitoring were uncompacted crown ratio, foliage transparency, crown dieback, crown position, light exposure, crown density, and vigor (Randolph 2009). Stand-level measurements were stand age, forest type, stand origin, stand size class, the land use if not forested, type of water body, and the dominant artificially regenerated species. These can differ within a subplot, in which case the boundaries are noted. Finally, site characteristics were slope, aspect (recorded at the subplot level), stand age (recorded at the stand level), and elevation and distance to an improved road that is maintained for use by motor vehicles (recorded at plot center). We did not compare measurements of the depth of water or snow, because these could change between visits by the two crews.

2.2. Data analysis

Because the QA crew does not necessarily measure the entire plot and because they can make different judgments than the field crew, an algorithm to pair QA and field crew observations is usually required at the stand- and tree-level (Pollard et al. 2005; Barnett 2022). From these paired data, the differences between QA and field crew observations can be calculated. For a continuous variable x measured by both crews for tree i, the disagreement between crews is $d_{xi} = x_{\mathrm{QA}i} - x_{\mathrm{FC}i}$, where $x_{\mathrm{QA}i}$ is the QA crew measurement of attribute x on tree i and $x_{\mathrm{FC}i}$ is the corresponding field crew measurement. The mean disagreement $\bar{d}_x$ evaluated over the n_x paired trees in our data set is $\frac{\sum_{i=1}^{n_x} d_{xi}}{n_x}$, and the standard deviation of the disagreement is the square root of the variance, which is $\frac{\sum_{i=1}^{n_x}\left(d_{xi} - \bar{d}_x\right)^2}{n_x - 1}$. Ignoring the direction of the differences, the mean "absolute disagreement" is $\frac{\sum_{i=1}^{n_x}|d_{xi}|}{n_x}$. We also report percentiles of the distribution of d_{xi} and the "relative

Fig. 1. Forest Inventory and Analysis (FIA) plots in the northeastern United States that were visited by both field crews and quality assurance (QA) crews and that included at least one tree, sapling or seedling. In forested areas, there is one FIA plot per ∼2.400 ha, with greater sampling intensities in some states and national forests.



disagreement", the absolute disagreement as a fraction of the measurement made by the QA crew $\frac{\sum_{i=1}^{n_x} \frac{|d_{xi}|}{x_{QAi}}}{n_x}$, expecting that some differences would scale with the magnitude of the measurement. For some values, a relative disagreement would not be meaningful as the true value could be zero. For example, for crown dieback and transparency, and site slope and aspect, relative disagreement was not reported.

For ordinal variables, we characterized the distribution of disagreements using the number of classes separating the QA crew and field crew observations. For categorical variables, the disagreement rate was the percentage of observations for which the QA crew and the field crew reported different values.

Slope, aspect, stand age, and distance to an improved road were sometimes not collected independently by the two crews because they are not part of the metrics for compliance. For these variables, we report uncertainty statistics both for all observations (which underestimates the true uncertainty, due to cases where agreement was due to the absence of an independent measurement) and for only the observations without agreement (which overestimates the true uncertainty, due to cases where the two independent measurements agreed).

2.3. Scaling up

To describe the effect of measurement uncertainty on estimates of basal area, timber volume, and aboveground biomass, we calculated these outputs for each tree based on both QA and field crew measurements. The measurements that affect these estimates depend on geographic location and may include tree species, diameter, total height, merchantable height, and cull (defect) (Woodall et al. 2011*b*). In some states, site index and basal area of the stand contribute to volume and biomass calculations (Miles and Hill 2010), but uncertainty in these variables does not contribute to the differences we report because they are not calculated in the FIA QA process. All trees with diameters of at least 2.5 cm have a biomass value; minimum tree size thresholds and tree grade are used to determine which trees have merchantable timber volume. We followed the methods currently used by FIA in generating congressionally mandated resource reports by state (Public Law 105-185).

Of the outputs describing timber volume estimates, we selected the net cubic foot volume of sawlogs (VOLCSNET) because it was the most restrictive and required the most variables to compute. At the other extreme, we chose the most inclusive output for aboveground biomass, namely the total dry biomass (DRYBIOT), which includes boles, tops, limbs, and stumps (Woodall et al. 2011*b*).

We compared these outputs for individual trees (38 514 for basal area, 38 507 for biomass, 9132 for volume) and for the 4663 subplots that were measured in full by the QA crew. We did not conduct this analysis at the plot level because only 86 plots had all four subplots evaluated by the QA crew. There were 1209 plots with two subplots evaluated, 819 with one subplot, and 419 with three subplots.

To compare the subplot totals of a calculated value y (basal area, volume, or biomass) on subplot j based on measurements taken by the two crews, the difference (d_{yj}) between crews is based on the sum of the y_{QAi} over the n_{QAyj} trees measured by the QA crew and the sum of the y_{FCi} over the n_{FCyj} trees measured by the field crew: $d_{yj} = \sum_{i=1}^{n_{QAyj}} y_{QAi} - \sum_{i=1}^{n_{FCyi}} y_{FCi}$, where n_{QAyj} and n_{FCyj} are not necessarily equal, being based on independent judgments by the two crews as to which trees should be measured.

For basal area and biomass, which require summing the contributions of saplings, measured on 13.5 m^2 microplots, as well as trees, measured on 168 m^2 subplots, we divided by the relevant area before summing to obtain the total basal area or biomass of the subplot on a per hectare basis.

We reported the average difference between the field and QA estimates for all trees, all subplots, and subplots in each of 24 states, each of seven forest type groups, and the Northern Region as a whole. We excluded the following forest types, which are not very common in the region we analyzed: oak–gum–cypress (4 subplots), exotic hardwoods (6), exotic softwoods (8), pinyon–juniper (6), other eastern softwoods (20), loblolly shortleaf pine (34), and ponderosa pine (60). We reported results for the following forest types: oak–pine (124 subplots), white–red–jack pine (307), elm–ash–cottonwood (361), spruce–fir (393), aspen–birch (463), maple–beech–birch (744), and oak–hickory (1337). To evaluate the importance of field and QA agreement on which trees to measure, we also reported rates of disagreement for just the subplots that were regarded as entirely forested by both crews and for just the subplots where both crews measured exactly the same trees.

3. Results

3.1. Tree measurements

3.1.1. Tree diameter

Tree diameters were measured with very high agreement between the QA crew and the field crew (Table 1, Fig. 2). For trees at least 12.7 cm (5 in.) in diameter, most measurements (62%) agreed perfectly ($d_{xi} = 0$); the resolution of the measurement was 0.25 cm (0.1 in.). For live trees, only 5% of the 29 042 stems disagreed by more than 0.25 cm; for dead trees, this proportion was slightly larger at 8%. Disagreement was higher in cases where one or both crews estimated the diameter due to vines along the stem or other circumstances that precluded access to the measurement location at the stem; the mean absolute disagreement for estimated diameters of live trees, which represented 0.4% of the diameter measurements (121 trees) was 0.80 cm, compared to 0.13 cm for trees that were measured by both crews. For dead trees, the mean absolute disagreement was 0.68 cm for the 23 trees that were estimated and 0.17 cm for the 2478 trees that were measured.

For saplings, which are defined as trees $\geq$2.5 cm (1 in.) and <12.7 cm (5 in.) in diameter, differences were even smaller, averaging 0.11 cm for live stems and 0.16 cm for dead ones. But because these are small stems, the differences were larger relative to the diameter, amounting to 2% for live and 3% for dead saplings, whereas the mean relative disagreement for trees at least 12.7 cm in diameter was 0.7% for live and 0.8% for dead trees. Diameters were estimated, not measured, for only 10 of the 7764 saplings in the data set; these trees had diameter measurement differences averaging 1.0 cm.

Some trees (6824 live trees, 716 dead trees, and 1625 saplings) were not measured at breast height, due to a variety of factors, related to form (e.g., swelling, forking) or growth habits and situations (e.g., hillsides, leaning trees, growing on rocks). The height at which these trees were measured is recorded, but no correction is made to these diameters in calculations of timber volume or tree biomass. The average absolute disagreement for trees measured at another height by either or both crews was 0.19 cm for live trees, 0.24 cm for dead trees, 0.17 cm for live saplings, and 0.06 cm for dead saplings. This uncertainty is greater than that for trees measured at breast height but less than the uncertainty incurred by estimating rather than measuring the diameters.

The breakdown of diameter measurement differences for trees and saplings, live and dead, measured or estimated or measured at another height than breast height, by one crew or the other or both, is provided in excruciating detail in Table S2 in the Supporting Information.

3.1.2. Tree height

Tree heights were measured with somewhat less agreement between the QA crew and the field crew than was tree diameter (Table 1, Fig. 3). Only 14% of all trees had identical heights; 36% disagreed by 0.3 m (1 ft), the resolution of the measurement, and 53% agreed within 0.6 m. The average absolute disagreement in the height of all trees was 0.97 m. As a proportion of tree height, this amounted to 6.8% of the height of the trees, on average, which is an order of magnitude more than the proportional disagreement in diameter measurement.

Broken trees present challenges for height measurement; 1370 trees and 91 saplings were reported as broken by one or both crews. In addition to the height to the highest point (called the "actual" height), the height including the broken length is reported. Differences in heights reported by the field crew and the QA crew are greater for broken than for unbroken trees, averaging 1.75 m (13%) when either or both rate the tree as broken, compared to 0.96 m (6.1%) when neither crew sees it as broken (Table S3 in the Supporting Information). When one crew sees that a tree is broken and the other does not, there is a mean disagreement of ~2 m in the total height reported, due to one crew including the length of the broken portion. The mean absolute disagreements for trees inconsistently rated as broken by the two crews were the highest in the height data set: 2.9 m for live trees and 3.2 m for dead trees that the QA crew recorded as broken but the field crew did not.

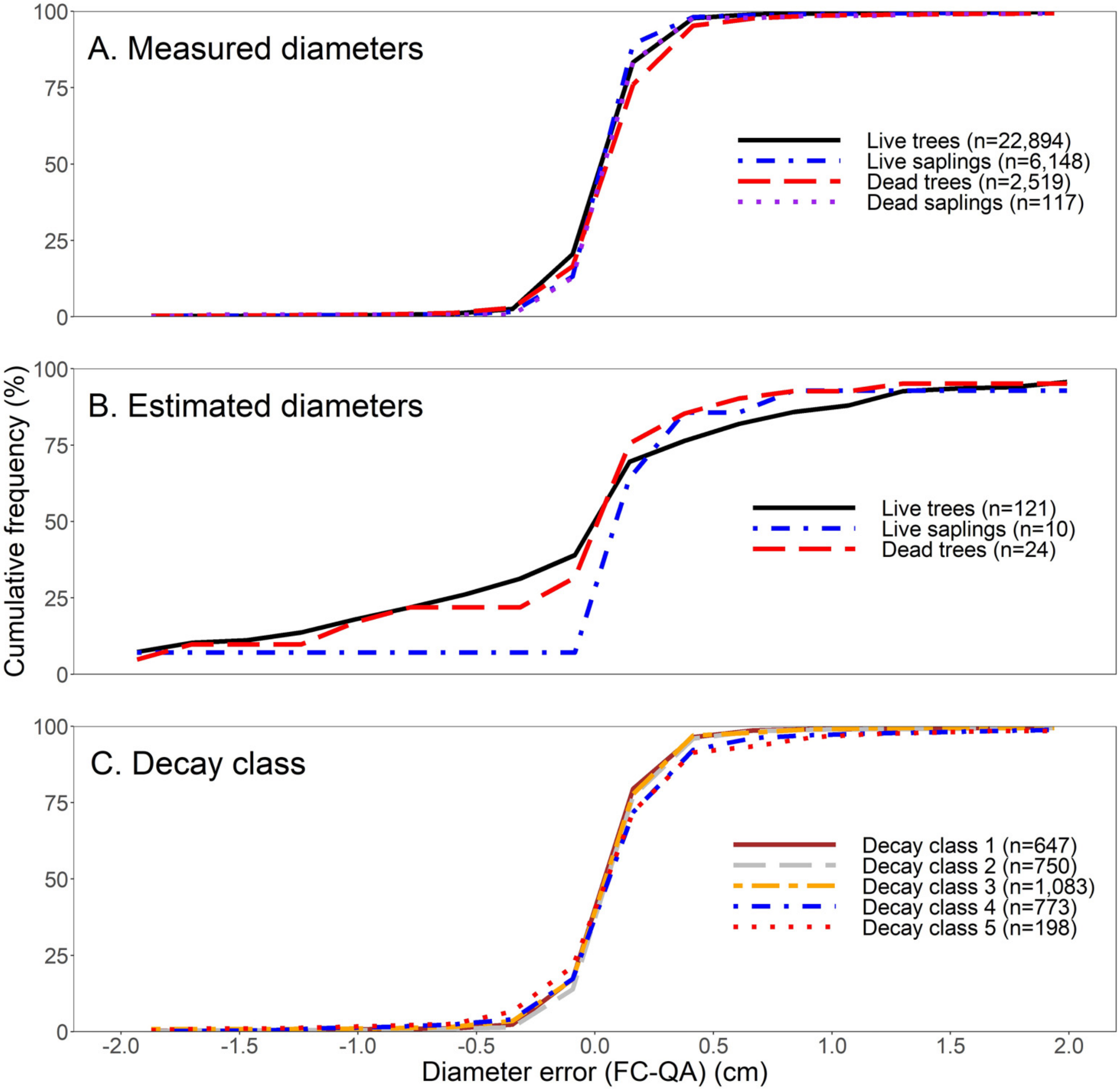
Field crews often estimate the height of trees, rather than measuring them with a hypsometer. Of the 35 033 tree heights in the data set, 8186 trees were estimated by one or both crews (7167 by the field crew and 1792 by the QA crew, including 772 estimated by both crews). Estimated heights were only slightly more prone to disagreement than measured heights, with average absolute disagreements of 1.05 m or 7.4%. When the field crew estimated the height and the QA crew measured it (the appropriate roles for a QA check), disagreement was only 1.02 m or 6.6% of height, which is impressively close to the 0.97 m (6.1%) disagreement for the 22 651 trees that were measured by both crews.

Table 1. Measurement uncertainty in tree (including sapling), seedling, and site attributes that are continuous variables.

Attribute (units)	Median	IQR	Number of observations	Mean disagreement	SD of the disagreements	Mean absolute disagreement	Relative absolute disagreement (%)	Percentiles of the disagreements			
								25	50	75	95
Tree level											
Diameter (cm)	17.5	11.7	41 182	− 0.004	0.55	0.14	1.0	0	0	0.25	0.51
Compacted crown ratio (%)	38	20	37 717	− 0.09	9.62	6.96	NA	2	5	10	20
Height with broken portions (m)	14.6	7.9	35 033	− 0.03	1.49	0.97	1.0	0.3	0.6	1.2	3.0
Height (m)	15.2	7.3	32 298	− 0.05	1.52	1.02	6.7	0.3	0.6	1.2	3.0
Rotten and missing cull (%)	0	0	29 925	0.1	3.5	0.8	NA	0	0	0	5
Uncompacted crown ratio (%)	55	30	2620	0.5	11.3	7.6	NA	2	5	10	25
Crown dieback (%)	0	5	2106	0.1	6.1	1.9	NA	0	0	5	5
Crown density (%)	50	15	144	− 2.7	11.5	8.9	19	5	5	15	25
Foliage transparency (%)	20	10	144	− 0.9	6.6	5.0	22	0	5	5	15
Seedling											
Seedlings by species (count)	2	3	8015	0.13	3.2	1.2	0.28	0	0	1	5
Seedling total (count)	6	11	3862	0.15	5.8	2.5	0.25	0	1	3	10
Seedling species per plot (count)	2	2	3862	− 0.16	0.9	0.9	0.20	0	0	2	4
Site variables											
Slope (%)	6	16	9183	0.01	3	1	NA	0	0	0	4
Slope excluding perfect agreement	3	9	1545	0.08	8	4	NA	1	2	4	12
Aspect (degrees)	47	200	9183	− 0.08	17	3	NA	0	0	0	13
Aspect excluding level subplots	180	178	5128	0.4	12	3	NA	0	0	0	15
Aspect excluding perfect agreement	170	177	691	3	33	19	NA	5	10	22	65
Stand age (years)	61	39	3081	− 0.03	14	2	NA	0	0	0	10
Stand age excluding perfect agreement	60	42	525	− 0.2	35	12	NA	1	5	12	49
Elevation (m)	351	178	2263	− 9.9	34	20	NA	5	12	23	63

Note: The mean disagreement reflects opposing positive and negative differences between the measurements by the field crew and the quality assurance crew and is thus smaller than the average absolute value of the disagreements. The relative absolute disagreement is the disagreement divided by the estimate, which is not applicable (NA) for variables with values that include zero. The mean of the relative absolute disagreement for all trees is given (this differs from the mean disagreement divided by the mean estimate). Percentiles are given for the absolute disagreement. Crown ratio and rotten and missing cull are estimated to the nearest 1%, but crown density, crown dieback, and foliage transparency are estimated to the nearest 5%, and this is reflected in the resolution of the disagreement percentiles. The number of observations varies because not all attributes were measured on all trees.

Fig. 2. Tree diameter measurement errors. (A) Diameters measured by both the field crew and quality assurance (QA) crew. (B) Diameters estimated by either the field crew or the QA crew. (C) All diameters, measured and estimated, divided by decay class as defined by the QA crew.



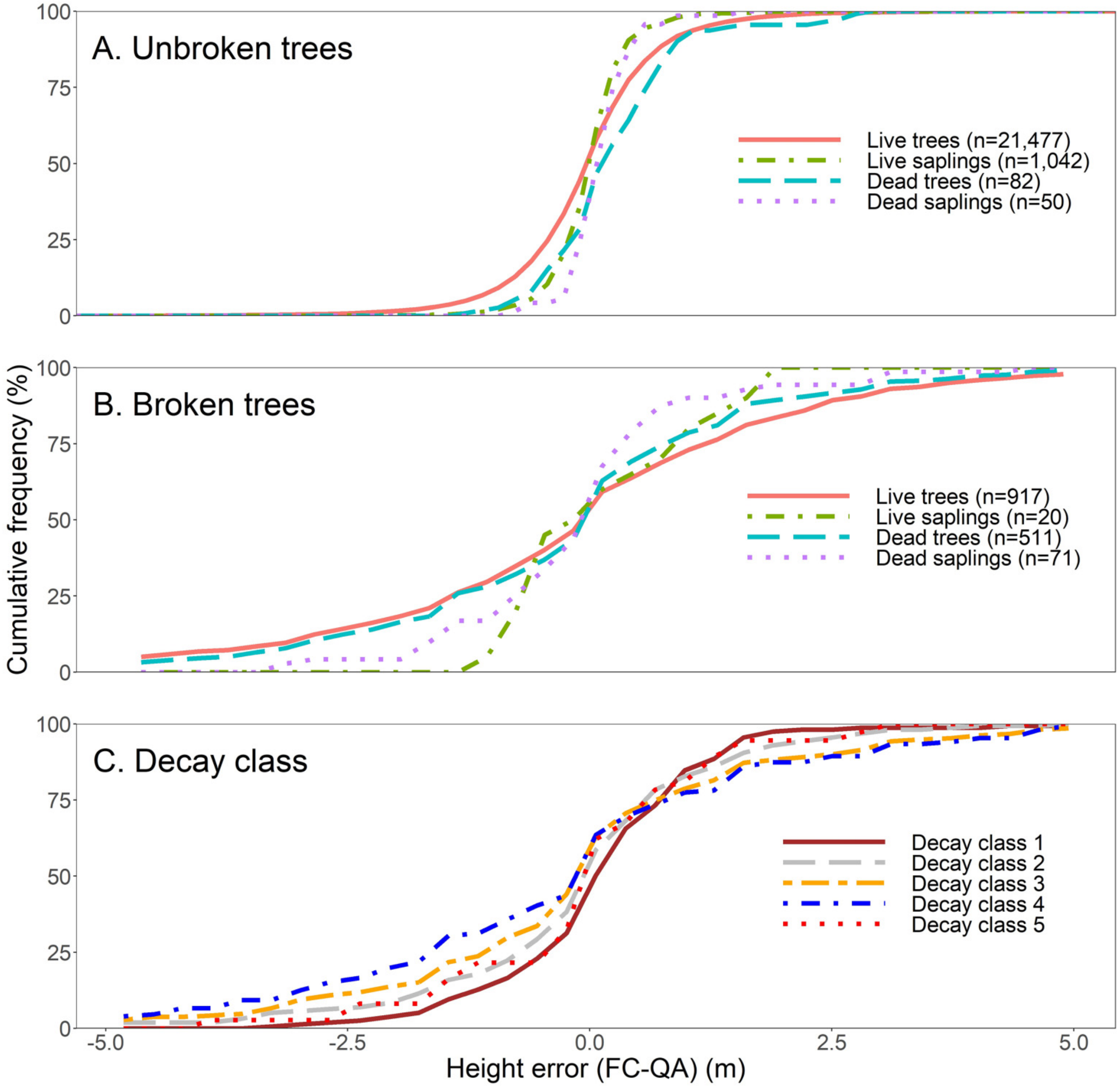
Dead trees were no harder to measure than live trees: both live and dead unbroken trees had absolute disagreement averaging 6.1% of tree height when trees were measured by both crews, and 7.2% (live) or 7.8% (dead) of tree height when heights were estimated by either or both crews. Saplings, being smaller, had measurement disagreements that were smaller in units of height (0.51 m for unbroken and 0.34 m for broken stems) but larger relative to their heights (7.6% for unbroken and 8.5% for broken stems).

3.1.3. Tree inclusion

A rarely reported source of uncertainty in forest inventory is the omission of trees in the plot or the inclusion of trees outside the plot. There were 271 live trees and 43 dead trees that both crews agreed were missed during the previous inventory (based on "reconcile codes"). There were 95 live trees and 5 dead trees that were regarded as missed by one crew but as ingrowth by the other crew—this is a matter of judging whether the tree was large enough to be tallied during the previous inventory. There were 84 trees tallied at the previous inventory that should not have been included, which most often happens on steep slopes, because inclusion depends on the horizontal distance to the tree. All together, these 498 instances represented 1.5% of the trees in our sample.

Differences in land-use determination (forest vs. nonforest) by the two crews also affected tree inclusion. Because trees are measured only if they are considered to be on

Fig. 3. Tree height measurement errors. (A) Trees not viewed as broken by either the field crew or the quality assurance (QA) crew. (B) Trees viewed as broken by either the field crew or the QA crew. (C) All trees, broken and unbroken, divided by decay class as defined by the QA crew.



forested land, disagreements about the forested condition result in discrepancies in measurements. For example, crews may disagree as to whether a tree is in a forest because it is in a residential landscape or because tree stocking or forest area requirements are not met. Of the 5410 subplots for which at least one crew assigned a forested status, there were 2% in which the other crew did not agree. Disagreements about tree inclusion affect estimates at the sub-plot scale and above (Section 3.4, below).

3.1.4. Species identification

Of the 51 512 trees that were identified to species, the field crew and QA crew disagreed in their species assignments for 1.8% of the trees (including saplings) observed by both crews. The rate of disagreement was similar for live trees (1.8%) and dead trees (1.6%). Saplings, which accounted for 7924 stems, had a disagreement rate of 2.7%. There were a total of 145 tree species reported (Table S4 in the Supporting Information).

Trees were not always identified to species: 671 trees, or 1.3%, were assigned only to genus by the field crew; 243 of those trees were identified by the QA crews. It was rare for genus not to be identified: there were 10 trees identified as “unknown dead hardwood” by the QA crews and 17 by the field crews.

The most common species were easy to identify, and rare species were more likely to have low agreement between the two crews (Fig. 4).

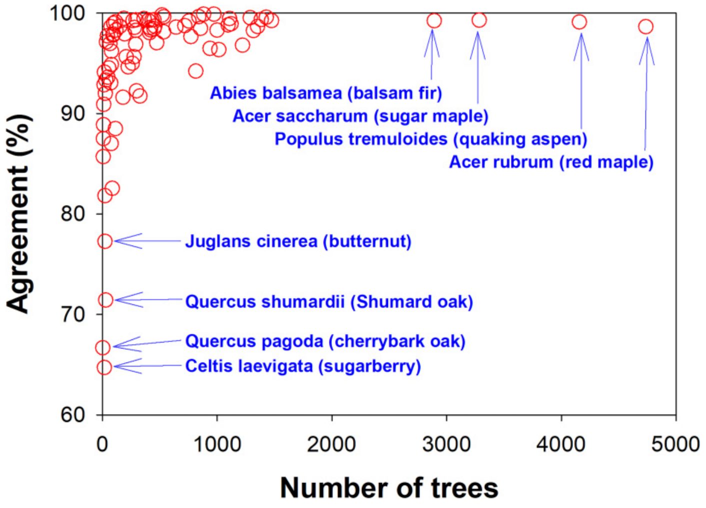


Fig. 4. Agreement in species identification between the field crew and the quality assurance crew. Species that were rarely observed were less likely to be correctly identified than the more common species.

3.1.5. Tree classification

Trees are classified by status, tree class, crown class, tree grade, decay class, crown ratio, and crown position. For these variables, too, we determined measurement uncertainty by comparing the results from the QA crew and the field crew. For categorical variables, we report the disagreement rate; for ordinal variables, we report the number of classes by which the two measures disagreed.

Tree status (live, dead, or harvested) was in agreement for 99.1% of trees (including saplings) for the 45 736 trees that were evaluated by both crews (Fig. 5A). The greatest agreement was for live trees (99.8%) and the worst was for harvested trees (91%); dead trees were intermediate (97%). Removed trees that are thought to not have been utilized are coded as dead trees, whereas if they are thought to have been utilized they are recorded as harvested. Thus, differing assessments of utilization contribute to the relatively low agreement on harvested trees. It is also likely that some trees were harvested between the field crew and QA crew measurements.

Tree class (growing stock, rough cull, or rotten cull) had 94% agreement between the field crew and the QA crew (Fig. 5B). Of the trees classed as growing stock by the QA crew, 97% were classed as growing stock by the field crew. For rough cull, there was only 84% agreement, with 13% of these trees classed as growing stock by the field crew. For rotten cull, similarly, there was 83% agreement, and 13% of these trees were classed as rough cull by the field crew. Cull is notoriously difficult to assess accurately.

Crown class assignments agreed exactly for only 83% of trees, which was worse than tree status and tree class, in part because there are more crown classes than status or class classes (Fig. 5C). The greatest agreement (90%) was for codominant trees, which were the most common in our data set (52% of trees), and agreement declined with the prevalence of the crown classes, with 83% for overtopped and 69% for intermediate crown classes. Dominant trees were only 1% of the data set, and agreement was 50% for these; the worst agreement (28%) was for open-grown trees, which were very rare (0.1%). Unlike tree class, discrepancies in crown class were quite symmetrical between the QA and the field crew.

Tree grade is assigned only to merchantable sawtimber trees, which are commercial species classified as growing stock (not rough or rotten cull) of sufficient diameter (12.7 cm) to qualify as sawtimber. The two crews agreed on the grade of 79% of the 7063 trees that they both evaluated (Fig. 5D). Agreement was highest for the best trees (Grade 1), at 86%, and lowest for the worst trees (Grade 5) at 60%.

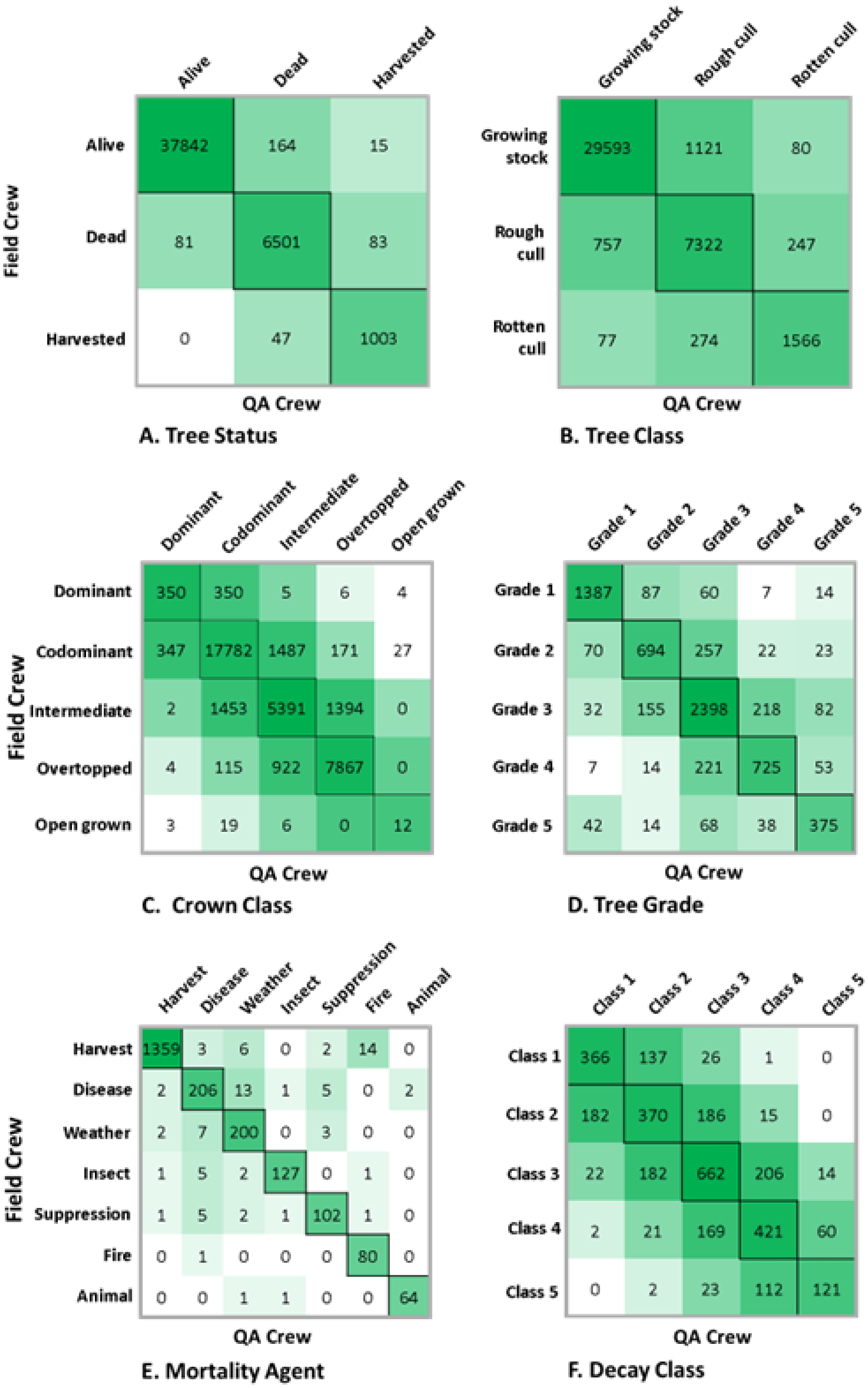
The cause of mortality is assessed for trees that died since the last inventory, but the cause is often unclear, or there may be multiple contributing factors. The field crew declined to assign a cause for 39% of the 4453 trees that were evaluated by both crews. The QA crew called an even higher proportion, 44%, of the trees "unknown" as to the cause of mortality. For the trees that were assigned a cause by both crews, there was 96% agreement as to the cause (Fig. 5E). The greatest agreement was for harvest (98%), which was the most common cause, and animal (93%), which was the least common cause. The worst agreement was for suppression (56%) and disease (64%).

Decay class was also difficult to rate exactly, with an average agreement of only 59% for the 3300 standing dead trees that were rated by both crews (Fig. 5F). However, 96% of trees agreed within one class, which is the tolerance considered acceptable by the FIA program. Only 0.2% of trees disagreed by three classes; none disagreed by four classes. The best agreement was for the least decayed trees (Class 1) at 64%, although Class 3 was the most common decay class. The worst agreement was for Class 2 at 52% and Class 4 at 56%.

3.1.6. Other crown characteristics

Crown ratio is the proportion of the stem length supporting foliage after the crown is visually compacted to account for areas of missing foliage. The estimates of the two crews disagreed by 7.0%, on average, for the 37 717 trees that were observed by both crews (Table 1). On a subset of plots, additional measurements are taken for purposes of forest health assessment (Westfall 2009). These include uncompacted crown ratio (the portion of tree height occupied by foliage without accounting for areas of missing foliage), crown density, crown position, crown dieback, crown vigor, crown light exposure, and foliage transparency. The mean absolute disagreement for crown dieback was only 1%, with the high level of repeatability owing to many trees having zero crown dieback (25%) and the resolution of the measurement being 5% (Table 1). Agreement was somewhat less for foliage transparency (5% average absolute disagreement) and crown density (9% disagreement). Crown position, crown light exposure, and crown vigor are categorical variables with three to five possible ordered values. Exact agreement was most

Fig. 5. A comparison of determinations of tree attributes by the field crew and the quality assurance (QA) crew. (A) Tree status, (B) tree class, (C) crown class, (D) tree grade, (E) mortality agent, and (F) decay class. Black outlines around cells indicate agreement between the two crews. The colors reflect the number of trees in each cell. Saplings are included for all attributes for which they are measured.



Field Crew

	Alive	Dead	Harvested
Alive	37842	164	15
Dead	81	6501	83
Harvested	0	47	1003

QA Crew

A. Tree Status

	Growing stock	Rough cull	Rotten cull
Growing stock	29593	1121	80
Rough cull	757	7322	247
Rotten cull	77	274	1566

QA Crew

B. Tree Class

Field Crew

	Dominant	Codominant	Intermediate	Overtopped	Open grown
Dominant	350	350	5	6	4
Codominant	347	17782	1487	171	27
Intermediate	2	1453	5391	1394	0
Overtopped	4	115	922	7867	0
Open grown	3	19	6	0	12

QA Crew

C. Crown Class

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Grade 1	1387	87	60	7	14
Grade 2	70	694	257	22	23
Grade 3	32	155	2398	218	82
Grade 4	7	14	221	725	53
Grade 5	42	14	68	38	375

QA Crew

D. Tree Grade

Field Crew

	Harvest	Disease	Weather	Insect	Suppression	Fire	Animal
Harvest	1359	3	6	0	2	14	0
Disease	2	206	13	1	5	0	2
Weather	2	7	200	0	3	0	0
Insect	1	5	2	127	0	1	0
Suppression	1	5	2	1	102	1	0
Fire	0	1	0	0	0	80	0
Animal	0	0	1	1	0	0	64

QA Crew

E. Mortality Agent

	Class 1	Class 2	Class 3	Class 4	Class 5
Class 1	366	137	26	1	0
Class 2	182	370	186	15	0
Class 3	22	182	662	206	14
Class 4	2	21	169	421	60
Class 5	0	2	23	112	121

QA Crew

F. Decay Class

Table 2. Measurement uncertainty in tree attributes that are ordinal variables.

Attribute	Number of observations	Number of classes	Mean disagreement (classes)	Mean absolute disagreement (classes)
Crown class	37 717	5	– 0.02	0.18
Tree grade	7063	5	– 0.02	0.29
Decay class	3300	5	0.03	0.45
Distance from a maintained road	2588	9	0.04	0.25
Crown position	198	3	– 0.13	0.16
Crown light exposure	198	5	0.12	0.69
Crown vigor	54	3	0.07	0.11

common (91%) with crown vigor (three classes) and least common (50%) with crown light exposure, which has five classes (Table 2).

3.2. Seedling measurements

The two crews agreed exactly on the number of seedlings in 2561 of the 3862 microplots visited by a QA crew (Table 1). The average difference in the total count was 2.5 seedlings, a 32% disagreement; the average number of seedlings per microplot was 10, and the median was 6. The number of species observed ranged from 1 to 11; the two crews agreed on the number of species in 36% of the microplots, and the average absolute disagreement was 0.46 species. Finally, the two crews agreed exactly on the number of seedlings reported for each species in each microplot for 35% of the 8015 cases of species by microplot. The average difference was about 1 seedling, a 28% disagreement; the average number of seedlings of a single species in a microplot was 4, and the median was 2.

3.3. Site characteristics

Inventory plots are assessed for a variety of site attributes such as forest type, stand size class, and management activities. Agreement between the two crews for the identification of artificial or natural regeneration silvicultural systems was very high (98.8%), while identifying the land use following forest conversion had lower agreement (84.6%) (Table 3). Repeatability of assessments for forest type and stand size class were intermediate, with rates of agreement being 94.0% and 92.5%, respectively.

For some attributes, it was clear that the two crews did not always make independent measurements. For the 9183 subplots for which there were values of slope, which is measured with a resolution of 1%, 83% were in perfect agreement (Table 1), not because measurement uncertainty was <1%, but because slope is an attribute that the QA crew is not required to check. Considering only the cases where the two measurements differed (1545 subplots), disagreements in slope were still quite small, with 26% differing by only 1%, 55% differing by up to 2%, and 85% agreeing within 5%.

For aspect, perfect agreement was even more common (90%) than for slope, because aspect is not measured where slope is <5%, but is instead assigned a value of 0; this was the case on 3958 subplots. Of the 5225 subplots with sufficient slope to be assigned as an aspect by both crews, 80% were in perfect agreement, but this includes an unknown number of cases where independent measurements were not made. For the 786 cases where the two crews reported different aspects, the mean absolute disagreement was 19 degrees, with an interquartile range of 5–22 degrees. The true uncertainty in aspect measurements lies between this and 4 degrees, the mean absolute disagreement excluding subplots with low slope.

Stand age was reported with perfect agreement for 83% of cases; this is often not estimated by either crew, but obtained by adding the number of years since the last inventory. There were 525 cases in which the two crews disagreed as to stand age, with a median disagreement of 5 years and a mean of 12 years.

Elevation was measured independently by the two crews, as a strategy for improving the estimates using GPS technology. Perfect agreement occurred on only 1% of plots, but the mean absolute disagreement in elevation measurements was only 6 m, with or without the plots in perfect agreement.

Distance from an improved road was evaluated categorically, with nine classes delineated at 30 m, 91 m, 152 m, 305 m, 805 m, 1.6 km, 4.8 km, and 8.0 km. There was perfect agreement in 82% of the plots, some of which were undoubtedly not evaluated independently. In most cases of disagreement, crews disagreed by only one class (14% of plots); 3% of plots disagreed by two classes, 1% by three or four classes, and 0.2% and <0.1% were off by five or six classes, respectively.

3.4. Scaling up

We quantified the effect of uncertainty in tree inventory data on estimates of basal area, aboveground forest biomass, and merchantable volume by making these calculations using as inputs either the values from the QA crew or the field crew, at tree, subplot, and larger scales (Table 4).

At the tree level, for basal area, which depends only on tree diameter, 64.9% of the 38 514 trees and saplings were in perfect agreement. The average relative disagreement was larger for saplings (3.5%) than for trees (1.3%). For biomass, which depends primarily but not entirely on tree diameter, 64.6% of trees and saplings were in perfect agreement. Relative disagreement was larger for saplings (4.2%) than for trees (1.7%). Agreement was lower for net merchantable volume, as this calculation depends on a form cull calculation, in addition to rotten and missing cull. Of the 9132 trees that were

Table 3. Measurement uncertainty in tree and plot attributes that are categorical variables.

Attribute	Number of observations	Number of classes	Disagreement rate (%)
Tree level			
Tree species	51 512	139	1.6
Tree status	45 736	3	0.9
Tree class	41 037	3	6.2
Standing dead	6501	2	1.1
Mortality agent	4453	8	18.8
Salvage	3165	2	8.0
Site variables			
Site status	7498	5	2.9
Land use if not forest	3756	12	15.4
Forest type (by dominant species)	3081	67	7.5
Forest type group (broader classes)	3081	14	3.6
Stand origin	3081	2	1.2
Field size	3081	5	7.5
Type of water body on the plot	2588	7	13.1

Note: Disagreement describes the percentage of observations for which the field crew and the quality control crew disagreed.

evaluated by both crews for the necessary attributes, only 37.8% of volume estimates were in perfect agreement. The average relative disagreement for tree-level volume was 5.3%. Saplings are too small to have any merchantable volume.

Scaling up to the subplot level, average disagreements are generally less than individual tree disagreements, because positive and negative disagreements for individual trees are summed across the plot. Specifically, basal area disagreements were reduced by 0.5% (from 1.7% to 1.2%), biomass disagreements by 0.7%, and merchantable volume disagreements by 0.05%, when scaling from tree to subplot level, when both crews measured the same trees (Table 4).

Additional uncertainty is introduced by disagreement as to whether trees should be included in the sum. Crews can disagree as to whether a tree is large enough to be measured, tall enough, in the case of a snag, or close enough to plot center, as opposed to being out of the plot. They can also disagree as to whether the site is a forest. Tree inclusion was a smaller source of disagreement than tree measurement: assuming that these sources are independent and the variances sum in quadrature, at the subplot level tree inclusion contributed 0.3% error in basal area, 0.4% error in biomass, and 1.3% error in volume, comparing all subplots to those in agreement about which trees to measure (Table 4). Comparing only subplots that both crews considered to be entirely forested (4525 of the 4663 subplots), disagreements were intermediate between those for all subplots and those where they agreed which trees to measure (Table 4). This is the rate of measurement uncertainty that corresponds to most forestry applications, where crews are not asked to determine whether a tree is in a forest but do need to make judgments about size thresholds and plot boundaries.

Aggregated at larger scales, measurement uncertainties become even less important. Summed or averaged for each of the 24 states, basal area, biomass, and timber volume had average relative disagreements of 0.4%, 0.4%, and 1.1%, respectively, based on all subplots (i.e., including those where crews did not agree on which trees to measure). At the scale of forest types, disagreements were reduced to 0.1%, 0.1%, and 0.5%, averaging the discrepancy between estimates based on field crew vs. QA crew measurements for seven forest types.

Finally, at the scale of the entire Northern Region, disagreements were reduced to 0.1% or less for all three variables, in spite of including subplots in which crews did not measure the same trees.

4. Discussion

Tree diameter is arguably the single most important attribute in forest inventories. Our comparison of blind checks by QA personnel with data collected by field crews in the U.S. Forest Service FIA program revealed that the average absolute disagreement in tree diameter measurement was only 0.1 cm, with an average disagreement of essentially 0.0 cm (Table 1), meaning that diameter measurements likely contribute trivial uncertainty in a wide range of applications.

Also of importance are tree height, which is needed for estimates of gross tree volume and site index, and rotten and missing cull, which are needed for estimates of sound volume and biomass. Both of these variables also exhibited near zero average disagreement (bias), but mean absolute disagreement was higher than that of tree diameter (as shown by Westfall and Patterson 2007). Tree heights are more difficult to measure accurately than diameters because the observer is necessarily at a distance from the tree and the view of the top may be obstructed. Tools used in the FIA program include clinometers and lasers in addition to ocular estimation; the uncertainty in measurements was not very sensitive to the approach used, as has been found elsewhere (Stereńczak et al. 2019).

Table 4. Measurement uncertainty in basal area, aboveground biomass, and timber volume, calculated from the difference between estimates based on measurements by the field crew and the quality assurance crew, at the scale of individual trees, subplots, and subplots averaged for 24 states, seven forest types, and the entire Northern Region.

Level		*n*	Median	Mean disagreement	SD disagreement	Mean absolute disagreement	Relative absolute disagreement (%)
Basal area (cm^2 for trees) (m^2·ha^{-1} for plots)							
Tree	All trees	38 514	255	−0.267	22.63	4.96	1.68
Subplot	All subplots	4663	22	−0.017	0.47	0.23	1.16
	Entirely forested	4211	23	−0.020	0.47	0.24	1.16
	Same trees	3964	23	−0.023	0.47	0.23	1.11
State	All subplots	24	26	−0.006	0.11	0.09	0.36
Forest type	All subplots	7	25	0.001	0.03	0.02	0.09
Region	All subplots	1	24	−0.017	NA	0.02	0.07
Biomass (kg for trees) (Mg·ha^{-1} for plots)							
Tree	All trees	38 507	98	−0.0003	0.02	0.003	2.09
Subplot	All subplots	4663	99	−0.176	3.09	1.29	1.38
	Entirely forested	4211	102	−0.194	3.17	1.31	1.37
	Same trees	3964	103	−0.211	3.21	1.30	1.31
State	All subplots	24	145	−0.010	0.71	0.55	0.41
Forest type	All subplots	7	115	−0.080	0.17	0.12	0.09
Region	All subplots	1	122	−0.176	NA	0.18	0.14
Merchantable volume (m^3 for trees) (m^3·ha^{-1} for plots)							
Tree	All trees	9132	0.5	0.000	0.10	0.03	5.26
Subplot	All subplots	4663	45	0.005	8.80	2.67	5.21
	Entirely forested	4211	48	0.024	8.93	2.72	5.19
	Same trees	3964	48	0.062	8.81	2.67	5.06
State	All subplots	24	93	0.206	1.20	0.95	1.09
Forest type	All subplots	7	85	0.075	0.48	0.38	0.49
Region	All subplots	1	79	0.005	NA	0.01	0.01

Note: At the subplot scale, we report the results for all subplots, subplots judged to be entirely forested by both crews, and subplots for which the two crews measured exactly the same trees.

Tree crown variables varied in repeatability, with crown dieback having near zero (0.1%) mean and mean absolute disagreements. Uncompacted crown ratio, crown density, and foliage transparency had mean absolute disagreements ranging from about 5%–9% with mean differences suggesting some bias for crown density and foliage transparency. As with heights, crown variables must be measured at a distance from the tree and may be difficult to assess in closed-canopy stands. These variables also require more judgment than measurements of height and diameter. A comparison of multiple field crews assessing trees in two stands in Michigan, likewise, found much higher variability in estimates of crown ratio than tree diameter (McRoberts et al. 1994). Accounting for uncertainty in measurements of crown attributes may be important where these variables are used in predictive models of tree growth and ecosystem services.

The repeatability of tree classification variables such as species, status, class, standing dead, salvage, and crown vigor was generally high (Tables 2, 3; Figs. 4, 5), such that little concern is warranted when making forest resource estimates using these classifications. However, mortality agent, tree grade, decay class, crown class, crown position, and crown light exposure were more difficult to assess. Determination of seedling counts and species presence were surprisingly consistent, with relative absolute disagreements <0.3% for all evaluations (Table 1), suggesting that measurement uncertainty has minor effects on regeneration indicators (McWilliams et al. 2018).

Only minor disagreements in assessments of key site variables were noted for slope, aspect, stand age, forest type, and stand origin (natural or plantation). These types of variables are often important for growth and yield models and for determining practices that promote specific forest management objectives. Slight variations in these variables are likely not important, but their implications could be evaluated in specific contexts.

The QA data that we analyzed were not collected to quantify uncertainty in the measurements, but rather to evaluate whether objectives set by the FIA for measurement quality are met (USDA 2018). For example, the objective for decay class is to have 90% of measurements agree within one class, and the objective for seedling count is to have 90% of measurements agree within 20%. Analytical compliance results for most data collected by FIA can be found in Pollard et al. (2006) and Westfall (2009).

The very low measurement uncertainty reported in this study might be difficult to achieve in other geographic regions. In the U.S. Pacific Northwest, which includes tree species that attain very large stature, a comparison of QA and field crew measurements of DBH revealed mean absolute differences of 0.5 cm (Melson et al. 2002), compared to only 0.1 cm in our analysis of similar data from the northeastern United States, and mean absolute differences for tree height were 1.5 m in the Pacific Northwest, compared to 1.0 m in our study. The variability of diameters among multiple crews measuring plots dominated by sugar maple in the Upper Peninsula of Michigan (McRoberts et al. 1994) was less than that of multiple crews measuring plots of broadleaved temperate forest types in the North Island of New Zealand (Holdaway et al. 2014). Both these studies reported standard deviations of tree diameter as a function of tree diameter (0.4 cm in Michigan and 1.2 cm in New Zealand for a 50-cm tree). Disagreement in species identification was 2.2% in the New Zealand study (Holdaway et al. 2014), slightly higher than the rate we found (1.6%); crews in both studies referenced earlier species lists. Some studies are difficult to compare directly to ours, as there are many ways to quantify measurement quality, but it seems likely that the high rates of reproducibility that we report would be difficult to achieve in other parts of the world, for example, where buttress roots are common (Metcalf et al. 2009).

Forest monitoring depends on recognizing which trees to measure. Crews can disagree about whether a subplot, or a portion of a subplot, was classified as a forest. This source of uncertainty is common in national forest inventories (Pollard et al. 2006) but does not pertain to most forest management applications. In an additional 1% of subplots, crews did not measure the same trees because of judgments about whether they were in or out of the plot, greater or less than the threshold diameter, or, in the case of snags, greater or less than the threshold height for measurement. Improving the repeatability of forest inventory depends on judgments about tree inclusion as well as on tree measurement.

High rates of reproducibility require effort. FIA field crews undergo annual training and certification, are subject to ongoing QA checks, and utilize sophisticated data recording software (USDA, no date) that reduces the incidence of errors. The field guide that defines measurement protocols is well developed and highly specific to ensure consistency in observation. For example, the section describing tree diameter measurement is 15 pages long (USDA 2018).

It should be noted that uncertainties in forest measurement are generally small relative to natural variation across plots (sampling error); the uncertainty in models used to predict biomass or volume is also relatively small (Gertner and Köhl 1992; McRoberts and Westfall 2014). When the quality of measurements at the level of the tree and plot is very high, the most important decision controlling uncertainty in forest inventory is the number of plots to be sampled. The sampling intensity of the FIA inventory was based on sampling error targets for estimates of forest area and tree volume, removals, and growth (USDA 2008). Sampling error is the source of uncertainty most commonly reported in ecological studies (Yanai et al. 2021), and it is the only source routinely reported by the FIA. A full accounting of uncertainty in forest inventories requires attention to the magnitude of model and measurement errors, including uncertainty as to which trees are included.

5. Conclusions

Many national forest inventories have QA procedures for ensuring compliance with target tolerances; we demonstrate how these data can be used to evaluate measurement uncertainty. In the Northern Region of the U.S. FIA, we found that uncertainty in measurements of individual trees was small and the uncertainty associated with which trees to measure was even smaller. The accuracy of forest inventory measurements will be needed to evaluate alternative measurement methods, whether involving ground-based sensing, such as terrestrial laser scanning or automated dendrometers, or remote sensing, such as satellite imagery, terrestrial LiDAR, or sub-canopy unmanned aerial vehicles (drones). In addition to providing measures of confidence in the results, quantifying uncertainty in long-term monitoring programs makes it possible to improve the allocation of limited resources (Levine et al. 2014). Objective measures of uncertainty are essential to providing confidence in estimates of biodiversity, timber value, and fuel loads; parameterizing models to predict forest composition, growth, and wildlife habitat; and meeting national and international goals for carbon accounting.

Acknowledgements

The field crews, regional trainers, QA staff, and database managers of the USDA Forest Service, Forest Inventory and Analysis program collected the forest inventory data and developed the products used in this study. Javier Garcia Perez Gamarra provided global context and helpful suggestions. This publication is a product of QUEST (Quantifying Uncertainty in Ecosystem Studies), a working group dedicated to advancing uncertainty analysis in ecosystem studies (www.quantifyinguncertainty.org), and QUERCA (Quantifying Uncertainty Estimates and Risk for Carbon Accounting), which is funded by the US Department of State and US Agency for International Development.

Article information

History dates

Received: 4 March 2022
Accepted: 13 September 2022
Accepted manuscript online: 25 November 2022
Version of record online: 2 February 2023

Copyright

Data availability

Barnett, C.J. 2022. Paired Quality Assurance and Operational Field Data from the Northern Region of the US Forest

Inventory and Analysis (FIA) Program. Fort Collins, CO: Forest Service Research Data Archive doi:10.2737/RDS-2022–0056

Author information

Author ORCIDs

Ruth D. Yanai https://orcid.org/0000-0001-6987-2489
John L. Campbell https://orcid.org/0000-0003-4956-1696

Authors' contributions

Conceptualization: RDY, ARY, JC, GAD, MBG, CWW
Data curation: ARY, CB
Formal analysis: RDY, ARY, JC, JAW, CB, MBG
Investigation: CB, MBG, CWW
Methodology: JAW
Project administration: RDY
Supervision: RDY
Visualization: RDY, ARY, JC
Writing – original draft: RDY, JAW, GAD, CWW
Writing – review & editing: RDY, JC, JAW, MBG

Competing interests

The authors declare that they have no competing interests.

Supplementary material

Supplementary data are available with the article at https://doi.org/10.1139/cjfr-2022-0062.

References

Barnett, C.J. 2022. Paired quality assurance and operational field data from the Northern Region of the US Forest Inventory and Analysis (FIA) program. Forest Service Research Data Archive, Fort Collins, CO. doi:10.2737/RDS-2022-0056.

Bechtold, W.A., and Scott, C.T. 2005. The forest inventory and analysis plot design. *In* The enhanced forest inventory and analysis program—National sampling design and estimation procedures. *Edited by* W.A. Bechtold and P.L. Patterson. Gen. Tech. Rep. SRS-80. U.S. Department of Agriculture, Forest, Asheville, NC. pp. 27–42. doi:10.2737/SRS-GTR-80.

W.A. Bechtold and P.L. Patterson, (eds). 2005. The enhanced forest inventory and analysis program - national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. U.S. Department of Agriculture, Forest, Asheville, NC. doi:10.2737/SRS-GTR-80.

Berger, A., Gschwantner, T., McRoberts, R.E., and Schadauer, K. 2014. Effects of measurement errors on individual tree stem volume estimates for the Austrian national forest inventory. For. Sci. **60**: 14–24. doi:10.5849/forsci.12-164.

Burrill, E.A., DiTommaso, A.M., Turner, J.A., Pugh, S.A., Menlove, J. Christiansen, G., et al. 2021. The Forest Inventory and Analysis Database: database description and user guide version 9.0.1 for phase 2. U.S. Department of Agriculture, Forest Service, Asheville, NC. 1026p. [Online]. Available from https://www.fia.fs.usda.gov/library/database-documentation/current/ver90/FIADB%20User%20Guide%20P2_9-0-1_final.pdf [accessed December 23, 2022].

Butt, N., Slade, E., Thompson, J., Malhi, Y., and Riutta, T. 2013. Quantifying the sampling error in tree census measurements by volunteers and its effect on carbon stock estimates. Ecol. Appl., **23**: 936–943. doi:10.1890/11-2059.1. PMID: 23865241.

Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., and Perez, R. 2004. Error propagation and scaling for tropical forest biomass estimates. Philos. Trans. R. Soc., B. **359**: 409–420. doi:10.1098/rstb.2003.1425. PMID: 15212093.

Cunia, T. 1987. On the error of continuous forest inventory estimates. Can. J. For. Res. **17**(5): 436. doi:10.1139/x87-075.

Gertner, G., and Köhl, M. 1992. An assessment of some nonsampling errors in a national survey using an error budget. Forest Science **38(3)**: 525–538.

Gormanson, D.D., Pugh, S.A., Barnett, C.J., Miles, P.D., Morin, R.S., Sowers, P.A., and Westfall, J.A. 2017. Statistics and quality assurance for the Northern Research Station Forest Inventory and Analysis Program, 2016. Tech. Rep. NRS-166. US Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA. 23p. doi:10.2737/NRS-GTR-166.

Harris, N.L., Hagen, S.C., Saatchi, S.S., Pearson, T. R.H., Woodall, C.W. Domke, G.M., et al. 2016. Attribution of net carbon change by disturbance type across forest lands of the conterminous United States. Carbon Balance Manage. **11**: 24. doi:10.1186/s13021-016-0066-5. PMID: 27909460.

Holdaway, R.J., McNeill, S.J., Mason, N.W.H., and Carswell, F.E. 2014. Propagating uncertainty in plot-based estimates of forest carbon stock and carbon stock change. Ecosystems, **17**: 627–640. doi:10.1007/s10021-014-9749-5.

Hunter, M. O., Keller, M., Victoria, D., and Morton, D.C. 2013. Tree height and tropical forest biomass estimation. Biogeosciences, **10**(12): 8385–8399. doi:10.5194/bg-10-8385-2013.

Köhl, M., Magnussen, S., and Marchetti, M. 2006. Sampling methods, remote sensing and GIS multiresource forest inventory(**Vol. 2**). Springer, Berlin. Section 3.4.

LaBau, V.J., Bones, J.T., Kingsley, N.P., Lund, H.G., and Smith, W.B. 2007. A history of the forest survey in the United States: 1830–2004. FS-877. U.S Department of Agriculture, Forest Service, Washington, DC. 82p. Available from https://www.fs.usda.gov/research/treesearch/14550 [accessed December 23, 2022].

Larjavaara, M., and Muller-Landau, H.C. 2013. Measuring tree height: a quantitative comparison of two common field methods in a moist tropical forest. Methods Ecol. Evol. **4**(9): 793–801. doi:10.1111/2041-210X.12071.

Levine, C.R., Yanai, R.D., Lampman, G.G., Burns, D.A., Driscoll, C.T. Lawrence, G.B., et al. 2014. Evaluating the efficiency of environmental monitoring programs. Ecol. Indic. **39**: 94–101. doi:10.1016/j.ecolind.2013.12.010.

McRoberts, R.E., and Westfall, J.A. 2014. Effects of uncertainty in model predictions of individual tree volume on large area volume estimates. For. Sci. **60**(1): 34–42. doi:10.5849/forsci.12-141.

McRoberts, R.E., Hahn, J.T., Hefty, G.J., and Van Cleve, J.R. 1994. Variation in forest inventory field measurements. Can. J. For. Res. **24**(9): 1766–1770. doi:10.1139/x94-228.

McWilliams, W.H., Westfall, J.A., Brose, P.H., Dey, D.C., D'Amato, A.W. Dickinson, Y.L., et al., 2018. Subcontinental-scale patterns of large-ungulate herbivory and synoptic review of restoration management implications for midwestern and northeastern forests. Gen. Tech. Rep. NRS-182. US Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA. doi:10.2737/NRS-GTR-182.

Melson, S., Azuma, D., and Fried, J. 2002. A first look at measurement error on FIA plots using blind plots in the Pacific Northwest. *In* Proceedings of the Third Annual Forest Inventory and Analysis Symposium. Gen. Tech. Rep. NC-230. U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN. pp. 11–20. https://www.fs.usda.gov/treesearch/pubs/14422.

Metcalf, C., Clark, J., and Clark, D. 2009. Tree growth inference and prediction when the point of measurement changes: modeling around buttresses in tropical forests. J. Trop. Ecol. **25**(1): 1–12. doi:10.1017/S0266467408005646.

Miles, P.D., and Hill, A.D. 2010. Volume equations for the Northern Research Station's Forest Inventory and Analysis Program as of 2010. Gen. Tech. Rep. NRS-74. U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA. p. 50doi:10.2737/NRS-GTR-74.

O'Connell, B.M., Conkling, B.L., Wilson, A.M., Burrill, E.A., Turner, J.A. Pugh, S.A., et al. 2016. The Forest Inventory and Analysis Database: database description and user guide version 6.1 for Phase 2. U.S. Department of Agriculture, Forest Service, St. Paul. 892p. Available from https://www.fia.fs.usda.gov/library/database-documentation/historic/ver6/FIADB_user%20guide_6-0_p2_5-6-2014.pdf [accessed December 23, 2022].

Penman, J., Gytarsky, M., Hiraishi, T., Krug, T., Kruger, D. Pipatti, R., et al. 2003. Good practice guidance for land use, land-use change and forestry. *In* Good practice guidance for land use, land-use change and forestry. Institute for Global Environmental Strategies (IGES) for the IPCC, Hayama, Japan. Available from https://www.ipcc-nggip.iges.or.jp/public/gpglulucf/gpglulucf_files/GPG_LULUCF_FULL.pdf

Penman, J., Kruger, D., Galbally, I. E., Hiraishi, T., Nyenzi, B. Emmanuel, S., et al. 2000. Quantifying uncertainties in practice. *In* Good practice guidance and uncertainty management in national greenhouse gas inventories. Intergovernmental Panel on Climate Change, Kanagawa, Japan. Chapter 6. https://www.ipcc.ch/publication/good-practice-guidance-and-uncertainty-management-in-national-greenhouse-gas-inventories/

Pollard, J.E., Westfall, J.A., Patterson, P.A., and Gartner, and D., L. 2005. FIA Quality Assurance Program: evaluation of a tree matching algorithm for paired forest inventory data. *In* Proceedings of the Fifth Annual Forest Inventory and Analysis Symposium; November 18-20, 2003, New Orleans, LA. Gen. Tech. Rep. WO-69. U.S. Department of Agriculture Forest Service, Washington, DC. pp. 117–122. Available from https://www.nrs.fs.usda.gov/pubs/gtr/gtr_wo069/gtr_wo069_117.pdf [accessed December 23, 2022].

Pollard, J.E., Westfall, J.A., Patterson, P.L., Gartner, D.L., Hansen, M., and Kuegler, O. 2006. Forest inventory and analysis National Data Quality Assessment Report for 2000 to 2003. Gen. Tech. Rep. RMRS_GTR181. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO. Available from https://www.fs.usda.gov/pnw/publications/forest-inventory-and-analysis-national-data-quality-assessment-report-2000-2003 [accessed December 23, 2022].

Randolph, K.C. 2009. Crown condition. *In* Forest health monitoring: 2006 national technical report. General Technical Report SRS-117. U.S. Department of Agriculture Forest Service, Southern Research Station, Asheville, NC. pp. 65–110. Available from https://www.fs.usda.gov/treesearch/pubs/55901 [accessed December 23, 2022].

Schreuder, H.T., Gregoire, T.G., and Wood, G.B. 1993. Sampling methods for multiresource forest inventory. Wiley, New York. Section 7.

Station, Fort Collins, CO. Available from

Stereńczak, K., Mielcarek, M., Wertz, B., Bronisz, K., Zajączkowski, G. Jagodziński, A.M., et al. 2019. Factors influencing the accuracy of ground-based tree-height measurements for major European tree species. J. Environ. Manage. **231**: 1284–1292. doi:10.1016/j.jenvman.2018.09.100. PMID: 30297223.

U.S. Department of Agriculture, Forest Service, Northern Research Station. 2019. Data collection quality assurance system. USDA Forest Service, Washington, DC. Available from https://www.nrs.fs.fed.us/fia/data-collection/docs/Q80%2020190711.pdf [accessed 25 May 2022].

U.S. Department of Agriculture, Forest Service. 2005. Forest inventory & analysis phase 2 and phase 3: Ground measurements FIA Fact Sheet Series. U.S. Department of Agriculture, Forest Service., Washington, DC. Available from https://www.fia.fs.usda.gov/library/fact-sheets/data-collections/Phase2_3.pdf [accessed December 23, 2022].

U.S. Department of Agriculture, Forest Service. 2008. Accuracy standards. In: Forest Survey Handbook. FSH 4809.11, chapter 10. U.S. Department of Agriculture, Forest Service, Washington, DC. Available from www.fs.fed.us/im/directives/fsh/4809.11/4809.11_10.doc [accessed December 23, 2022].

U.S. Department of Agriculture, Forest Service. 2018. Forest Inventory and Analysis (FIA) National Core Field Guide Volume I: Field data collection procedures for phase 2 plots (version 7.2). U.S. Department of Agriculture, Forest Service, Washington, DC. Available from https://www.fia.fs.usda.gov/library/field-guides-methods-proc/index.php [accessed December 23, 2022].

U.S. Department of Agriculture, Forest Service. 2021. Forest Inventory and Analysis Datamart. U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis Program, Washington, DC. https://apps.fs.usda.gov/fia/datamart/datamart.html [accessed 16 July 2021].

U.S. Department of Agriculture, Forest Service. No date. Mobile integrated data acquisition system (MIDAS) installation & users guide.. U.S. Department of Agriculture, Forest Service, Arlington, VA. Available from https://apps.fs.usda.gov/fia/midas-v2/Documentation/MIDAS_UsersGuide.pdf [accessed 25 May 2022].

Westfall, J. A.(ed.). 2009. FIA national assessment of data quality of forest health indicators. Gen. Tech. Rep. NRS-53 U.S. Department of Agriculture, Forest Service, Northeastern Research Service, Newtown Square, PA. 80p. doi:10.2737/NRS-GTR-53.

Westfall, J.A., and Patterson, P.L. 2007. Measurement variability error for estimates of volume change. Can. J. For. Res. **37**(11): 2201–2210. doi:10.1139/X07-082.

Woodall, C.W, Heath, L.S., Domke, G. M., and Nichols, M.C. 2011b. Methods and equations for estimating aboveground volume, biomass, and carbon for trees in the U.S. forest inventory, 2010. Gen. Tech. Rep. NRS-88. U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA. pp. 1–30. doi:10.2737/NRS-GTR-88.

Woodall, C.W., Amacher, M. C., Bechtold, W.A., Coulston, J. W., Jovan, S. Perry, C. H., et al. 2011a. Status and future of the forest health indicators program of the USA. Environ. Monit. Assess. **177**: 419–436. doi:10.1007/s10661-010-1644-8. PMID: 20711860.

Woodall, C.W., Coulston, J.W., Domke, G. M., Walters, B.F., Wear, D.N. Smith, J. E., et al. 2015. The U.S. forest carbon accounting framework: stocks and stock change, 1990-2016. Gen. Tech. Rep. NRS-154. U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA. 49p. doi:10.2737/NRS-GTR-154.

Woudenberg, S.W., Conkling, B.L., O'Connell, B.M., LaPoint, E.B., Turner, J.A., and Waddell, K.L., 2010. The forest inventory and analysis database: Database description and users manual version 4.0 for Phase 2. Gen. Tech. Rep. RMRS-GTR-245. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO. 336p. doi:10.2737/RMRS-GTR-245.

Yanai, R.D., Mann, T.A., Hong, D., Pu, G., and Zukswert, J.M. 2021. The current state of uncertainty reporting in ecosystem studies: a systematic evaluation of peer-reviewed literature. Ecosphere, **12**(6): e03535. doi:10.1002/ecs2.3535.

Yanai, R.D., Wayson, C., Lee, D., Espejo, A.B., Campbell, J.L Green, M.B., et al. 2020. Improving uncertainty in forest carbon accounting for REDD+ mitigation efforts. Environ. Res. Lett., **15**: 24002. doi:10.1088/1748-9326/abb96f.