Urban Forestry & Urban Greening 22 (2017) 124–135

Contents lists available at ScienceDirect

Urban Forestry & Urban Greening

journal homepage: www.elsevier.com/locate/ufug

Original article

Data quality in citizen science urban tree inventories

Lara A. Roman[a,*], Bryant C. Scharenbroch[b,c], Johan P.A. Östberg[d], Lee S. Mueller[e,f], Jason G. Henning[g], Andrew K. Koeser[h], Jessica R. Sanders[i], Daniel R. Betz[a,j], Rebecca C. Jordan[j]

[a] US Forest Service, Philadelphia Field Station, 100 N. 20th St., Suite 205, Philadelphia, PA 19103, USA
[b] University of Wisconsin – Stevens Point, College of Natural Resources, 800 Reserve Street, Stevens Point, WI 54481, USA
[c] The Morton Arboretum, 4100 Illinois Route 53, Lisle IL 60532, USA
[d] Swedish University of Agricultural Sciences, Department of Landscape Architecture, Planning and Management, P.O. Box 66, SE-230 53 Alnarp, Sweden
[e] Friends of Grand Rapids Parks, PO Box 3199, Grand Rapids, MI 49501, USA
[f] Davey Resource Group, 295 S Water Street, #300, Kent, OH 44240, USA
[g] The Davey Institute and US Forest Service, Philadelphia Field Station, 100 N. 20th St., Suite 205, Philadelphia, PA 19103, USA
[h] Department of Environmental Horticulture, CLCE, IFAS, University of Florida – Gulf Coast Research and Education Center, 14625 County Road 672, Wimauma, FL, 33598, USA
[i] Technical Services and Research, Casey Trees, 3030 12th St. NE, Washington, DC 20017, USA
[j] Human Ecology, School of Environmental and Biological Sciences, Rutgers University, New Brunswick, NJ 08901, USA

ARTICLE INFO

Article history:
Received 1 July 2016
Received in revised form
18 November 2016
Accepted 1 February 2017
Available online 4 February 2017

Keywords:
Diameter at breast height
Observation error
Species misidentification
Tree monitoring
Urban forest
Volunteer monitoring

ABSTRACT

Citizen science has been gaining popularity in ecological research and resource management in general and in urban forestry specifically. As municipalities and nonprofits engage volunteers in tree data collection, it is critical to understand data quality. We investigated observation error by comparing street tree data collected by experts to data collected by less experienced field crews in Lombard, IL; Grand Rapids, MI; Philadelphia, PA; and Malmö, Sweden. Participants occasionally missed trees (1.2%) or counted extra trees (1.0%). Participants were approximately 90% consistent with experts for site type, land use, dieback, and genus identification. Within correct genera, participants recorded species consistent with experts for 84.8% of trees. Mortality status was highly consistent (99.8% of live trees correctly reported as such), however, there were few standing dead trees overall to evaluate this issue. Crown transparency and wood condition had the poorest performance and participants expressed concerns with these variables; we conclude that these variables should be dropped from future citizen science projects. In measuring diameter at breast height (DBH), participants had challenges with multi-stemmed trees. For single-stem trees, DBH measured by participants matched expert values exactly for 20.2% of trees, within 0.254 cm for 54.4%, and within 2.54 cm for 93.3%. Participants' DBH values were slightly larger than expert DBH on average (+0.33 cm), indicating systematic bias. Volunteer data collection may be a viable option for some urban forest management and research needs, particularly if genus-level identification and DBH at coarse precision are acceptable. To promote greater consistency among field crews, we suggest techniques to encourage consistent population counts, using simpler methods for multi-stemmed trees, providing more resources for species identification, and more photo examples for other variables. Citizen science urban forest inventory and monitoring projects should use data validation and quality assurance procedures to enhance and document data quality.

Published by Elsevier GmbH.

1. Introduction

Citizen scientists have been involved with ecological monitoring across a range of programs, expanding public engagement in research (Dickinson et al., 2012). In the ecological sciences, citizen science engages the public in authentic research, typically through volunteers collecting field data (Dickinson et al., 2012), which promotes environmental awareness, scientific literacy, and social capital (Cooper et al., 2007; Bonney et al., 2009; Conrad and Hilchey 2011; Crall et al., 2013). While the data generated by citizen scientists has been used for research and natural resource management (Dickinson et al., 2010; Tulloch et al., 2013), concerns have

* Corresponding author.
E-mail address: lroman@fs.fed.us (L.A. Roman).

http://dx.doi.org/10.1016/j.ufug.2017.02.001
1618-8667/Published by Elsevier GmbH.

been raised about data quality (Bird et al., 2013; Lewandowski and Specht 2015).

Assessments of observation error in citizen science have had mixed results concerning both the level of volunteer accuracy and implications of those findings for applying citizen science to research and management. Species misidentification and incomplete taxonomic resolution in citizen science projects can lead to interpretation problems, such as overestimation of species diversity (Gardiner et al., 2012), and limited research utility of volunteer-generated species lists beyond community-level assessments (Kremen et al., 2011). For example, Kremen et al. (2011) found that volunteers missed half the bee groups recorded by researchers. However, citizen science studies focused on coral reefs, crabs, and plants have concluded that data collected by volunteers was mostly accurate, and sometimes of comparable quality to data collected by professionals (Delaney et al., 2008; Edgar and Stuart-Smith 2009; Crall et al., 2011; Butt et al., 2013; Danielsen et al., 2013). For example, Crall et al. (2011) found that volunteer species accuracy for invasive plants was 72%, compared to 88% for professionals, with both groups having lower accuracy for difficult-to-identify species. These varied studies of citizen science data quality also had widely different task complexity, with species identification involving less than 10 to dozens or even hundreds of species. With the quality of volunteer data as well as task complexity varying by case, and each case having particular data quality needs, new implementations of citizen science should include pilot testing and accuracy evaluations.

While data quality from volunteers is sometimes questioned, field data collected by researchers and their paid crews is not free of errors. When forest monitoring is conducted by researchers, examining the extent and sources of error helps to identify best practices for training crews, conducting field work and managing data (van Doorn 2014). Whether data are produced by paid or unpaid field crews, observation errors can be documented and potentially minimized through quality assurance and data validation (Ferretti 2009; Wiggins et al., 2011), and quantified error can be accounted for in statistical models (Chave et al., 2004). Evaluations of citizen science data quality can therefore be viewed in the larger context of best management practices for ecological monitoring (Lindenmayer and Likens 2010). As with any ecological research, assessing observation error in citizen science is critical to both designing effective programs and determining appropriate uses of the data.

In this paper, we present a pilot study about data quality in urban tree inventories collected by volunteers. We focused on street tree inventories, as street trees are on the front lines of engagement and management in municipal forestry. Street tree inventories record the locations and particular attributes of trees in sidewalks and other street-side environments. Such inventories are used for a wide range of purposes, including managing tree risk, prioritizing maintenance, mapping storm-damaged trees, charting species diversity and size class distribution, and estimating ecosystem services (Jim and Liu 2001; Harris et al., 2004; McPherson et al., 2005; Sjöman et al., 2012; Bond 2013; McPherson and Kotow 2013; Östberg and Sjögren 2016). Researchers and managers also use repeated inventories and systematic monitoring to explore trends in street tree populations, such as composition changes and mortality rates (Dawson and Khawaja 1985; Roman et al., 2013; Roman et al., 2014). Depending on the particular objectives of urban forest inventories, the data quality necessary and qualifications of those collecting the data may differ.

While street tree inventories are traditionally carried out by professional arborists, citizen scientists are now used in many cities. Examples of citizen science in urban forest management include the street tree census in New York City, NY (Silva et al., 2013; Campbell 2015), the Tree Inventory Project in Portland, OR (St. John 2011), the OpenTreeMap software, which has been used in cities in the United States, Canada and the United Kingdom (www.opentreemap.com, Kocher 2012), and survival monitoring for planting programs across the United States (Roman et al., 2013; Silva and Krasny 2014). Citizen science can improve volunteer knowledge about trees (Cozad 2005) and some authors have suggested that engaging volunteers in data collection can build support for municipal and nonprofit programs (McPherson 1993; Bloniarz and Ryan 1996). The application of citizen science in urban forestry builds on a rich tradition of volunteerism in urban forest management, with volunteers engaging in tree planting and other forms of stewardship (Romolini et al., 2012). Such activities can deepen participants' civic engagement and cultivate a sense of empowerment (Westphal 2003; Fisher et al., 2015; Ryan 2015).

Yet, even with urban foresters already using volunteers to collect data, integrating volunteer data into urban forest research and management has been met with skepticism due to lack of information about observation errors in the urban forest context (Roman et al., 2013). The complexity of tasks in urban tree inventories may make such work particularly challenging for citizen scientists who lack prior experience. Specifically, field crews must contend with high species diversity, with, on average, 77 tree species across 38 cities world-wide (Yang et al., 2015), and substantially higher in some municipalities, such as 161 species in Chicago, IL (Nowak et al., 2013). This includes native and exotic species, and identification guides for novices are not widely available. Urban tree inventories also typically involve measuring diameter at breast height (DBH), and if monitoring DBH change is desired, this requires field crews to make consistent measurements that allow for longitudinal tracking of individual tree growth over years or even decades.

There are only two previous studies about volunteer data quality in urban tree inventories. Although both studies conclude that there is potential for relying on field data collected by volunteers, accuracy rates for certain variables do not seem tenable for research and management applications. Cozad (2005) studied volunteer accuracy for a street tree inventory in Minneapolis, MN, and found 76% accuracy for DBH tree size class reported by volunteers, and 80% accuracy for species identification. Bloniarz and Ryan (1996) studied volunteer accuracy in Brookline, MA, and found that 94% of volunteers agreed with arborists for genus identification, and 80% for species identification; although that study considered only the most common species. Both studies found relatively low data quality for volunteers reporting maintenance needs (49% in Cozad 2005; 75% in Bloniarz and Ryan 1996), indicating that such evaluations should be performed to professionals. As public participation in urban tree inventories and monitoring expands, it is essential to build upon these studies with evaluations of data quality in more locations, and to make explicit connections between observed data quality and appropriate data uses.

Our study compared street tree data collected by experts to data collected by field crews with novice and intermediate levels of prior experience as a pilot test of new tree monitoring protocols. The goals of our study were to (1) identify the magnitude and frequency of inconsistencies in urban forestry field data; (2) determine whether novice and intermediate crews differ in their performance for genus and DBH; and (3) generate suggestions to revise training and data collection procedures in ways that may enhance data quality. We then draw lessons learned for volunteer data collection in urban forestry, with comparisons to prior studies (Bloniarz and Ryan 1996; Cozad 2005), and provide recommendations for designing field methods suitable to volunteers as well as appropriate applications of citizen science in urban forestry.

2. Methods

2.1. Background

Our study grew from the Urban Tree Growth and Longevity (UTGL) working group, a community of practice affiliated with the International Society of Arboriculture (Scharenbroch et al., 2014; Campbell et al., 2016). UTGL members – including researchers, university students, municipal and consulting arborists, and nonprofit staff – developed a customizable framework for urban tree monitoring protocols that provides technical guidance for local urban forestry organizations in tracking tree growth and mortality. The new protocols responded to needs identified from a prior survey of practitioner-driven tree monitoring programs (Roman et al., 2013). Respondents to that survey noted challenges in training field crews and producing reliable data, especially with volunteers and interns. The new monitoring protocols emphasize locational accuracy for longitudinal studies and precise re-measurement of tree size using a field guide with many photo examples.

To evaluate data quality in citizen science urban tree inventories using these new protocols, we conducted a pilot test that compared observations from experts to other field crews. Data collected by experts was considered correct; a similar approach has been employed in other citizen science accuracy studies (e.g., Bloniarz and Ryan 1996; Cozad 2005; Delaney et al., 2008; Edgar and Stuart-Smith 2009; Kremen et al., 2011; Crall et al., 2011; Butt et al., 2013) and this assumption is backed by the high consistency rates among arborists in Bloniarz and Ryan (1996). In addition to field data collection, we surveyed the non-expert participants (hereafter "participants") to get details about their relevant prior experiences as well as to solicit their suggestions to improve training and data collection.

We carried out this study in four cities: Lombard, IL; Grand Rapids, MI; Philadelphia, PA; and Malmö, Sweden (Table 1, Fig. 1). These are small, medium and large cities with varied climates, which provided a range of tree species and site conditions. The cities were included based on interest from UTGL members who were involved in developing the protocols and worked in those cities; this was done to pilot test the protocols with interested parties, and cities were not chosen based on any other criteria.

2.2. Data collection

2.2.1. Street tree inventory

Street tree inventory data included the variables described in Table 2. These variables were identified by the UTGL working group as the minimum data set: the core group of variables essential to any long-term monitoring study of urban tree population changes. We did not include conventional tree condition ratings (i.e., good, fair, poor, dead or dying) from i-Tree Streets (www.itreetools.com) because these are subjective ratings that require knowledge of how particular species perform in a given region. Our variables for crown dieback, crown transparency, and wood condition were meant to mirror the four classes in the condition variable from i-Tree Streets while providing crews with more objective means of evaluating condition. Crews also recorded time spent per tree (not including transportation time to the neighborhood).

Within each city, the sampled trees were located within neighborhoods selected to include a range of common and rare species, as well as span many size classes and growth forms. We selected sampling areas in each city to encompass roughly 150 trees, the amount we expected crews could record in 1.5–2 days. Crews were instructed to record every street tree within the pre-defined geographic boundary. Data was collected Jul.–Sep. 2014.

Expert crews were urban forest researchers and certified arborists with extensive prior experience conducting tree inventories; specifically, most of the co-authors on this paper served as experts. Participants self-identified as novice or intermediate based on the following descriptions. Novice crews had little to no prior urban forestry field work experience (1 year or less) and little prior knowledge of essential skills (e.g., species identification, DBH measurements). Intermediate crews had relevant past experience in urban forestry field work (at least 1–3 years) and some prior knowledge of essential skills. These participants were mostly volunteers, except for a few paid interns in Philadelphia who had comparable skill levels to the volunteers. Participants were recruited via email through local urban forestry nonprofit organizations (e.g., tree stewardship and gardening programs) and universities (e.g., environmental studies departments).

Trees were observed independently by one expert, three intermediate, and three novice crews. There were seven independent observations for each tree in each city except Malmö. In that city, the expert recorded all trees in the study area, but novices and intermediates recorded a subset of 93 trees per crew on average. This was due to challenges getting participants in Malmö to commit to more than one day of field work. Although Scandinavian countries have a rich history of volunteerism, volunteering in urban green space management is uncommon (Molin and Konijnendijk van den Bosch 2014). Novice and intermediate field crews worked in pairs within the same experience level. Experts operated alone in all cities except Philadelphia, where the experts were paired.

All participants received a standardized in-person 6–7 h training, led by the experts in each city, then carried out field work without direct supervision. For comparison, volunteer training time was 12 h in Bloniarz and Ryan (1996) and 1 day in Cozad (2005). For practitioner-run citizen science projects, volunteer tree inventory training is 3.5 h followed by 3 h of closely supervised data collection in New York City, NY (Silva et al., 2013; C. Cochran, pers. comm.), and 3.5 h followed by 4 half-day field days with on-site arborist support in Portland, OR (A. diSalvo, pers. comm.). Training is typically 2–3 days (R. Hoehn, pers. comm.) for the more intensive i-Tree Eco protocol (www.itreetools.org), which is generally for interns with prior field experience.

For each city, the training timing broke down roughly as follows: introduction to the project (15 min); field crew information, site type, and land use (30 min); recording tree location (30 min); species identification (1.5 h); mortality status and condition (45 min); DBH (1 h); outdoor practice (1 h); and closing discussion, including brainstorming potential sources of error (30 min). The species identification lesson reviewed botanical terms and 10–12 common local street trees, including a "species face-off" for similar leaf shapes (e.g., how to distinguish *Acer* spp. from other trees with similar palmate leaf shapes, like *Platanus* x *acerifolia* and *Liquidambar styraciflua*). Regionally-relevant species identification resources were made available to participants, including handouts and apps which they could use during field work. This included a two-page street tree identification guide from New York City, NY Department of Parks and Recreation; the Stikky tree guide (Holt 2010); the website "What Tree is That" from the Arbor Day Foundation (www.arborday.org/trees/whattree); and the Leafsnap app (Kumar et al., 2012). Crews were instructed to write "unknown" when they did not know genus or species, and for cases of known genus but unknown species, crews were instructed to record the genus only (e.g., *Acer* sp.) to acknowledge their uncertainty. Genus-only identification was deemed acceptable for *Crataegus*, *Malus*, and *Prunus*, as even expert crews may have difficulty distinguishing those hybrids and cultivars. Participants practiced with leaf samples and DBH measurement indoors, and practiced the entire protocol on several trees outdoors during the training session. DBH training stressed common mistakes and best practices for using diameter tape. Crews were given a field guide explaining the data collection protocols that included many

Table 1
Characteristics of cities and trees used in the pilot study. Population is from the US Census (2010) and Statistics Sweden (2010). Climate is based on Köppen-Geiger climate classification system (Peel et al., 2007). Number of genera, species, and observed trees (n) reflect the expert tree inventory conducted for this study.

City	Popn.	Popn. density (per km^2)	Climate	Neighborhood	No. genera	No. species	n
Lombard, IL	43,165	1626	humid continental	Downtown	17	26	157
Grand Rapids, MI	188,040	1635	humid continental	East Hills	16	23	147
Philadelphia, PA	1,526,006	4394	humid subtropical	University City	20	32	152
Malmö, Sweden	298,963	1922	oceanic	Central	6	8	154

Fig. 1. Photos of the study sites, clockwise from top left: Lombard, IL; Grand Rapids, MI; Malmö, Sweden; Philadelphia, PA.

photos and diagrams; they paged through this field guide during the training session and carried it during field work. Crews recorded data on standardized data collection sheets and research interns or the experts in each city subsequently entered that data.

For DBH, crews were instructed to follow procedures from the Urban Forest Inventory and Analysis (UFIA) guide from the US Forest Service (US Forest Service 2016). Multiple stems at DBH height were recorded as a single tree; rules about where to record DBH on each stem relate to where the stems fork and piths join. DBH was recorded with DBH tape to the nearest 0.254 cm for all cities except Malmö, where it was recorded to the nearest 0.25 cm.

All cities used the same field guide but there were differences in how street tree was defined, the kind of neighborhood where the study took place, and crew access to a pre-existing tree map. Crews were told a definition of street tree relevant to their city: 1) all trees located between the sidewalk and the curb (Grand Rapids, Philadelphia), or 2) all trees within a set distance from road center (i.e., the right-of-way), and such trees could be located in lawns but still considered street trees (Lombard, Malmö). In most cities,

Table 2
Street tree inventory data collected by field crews using protocols from UTGL. Tree location protocols were adapted from common methods employed in various urban tree asset management systems. Site type and land use were adapted from i-Tree Streets and i-Tree Eco, respectively (www.itreetools.com); only site type and land use categories relevant to street trees are listed below. Crown dieback, transparency and DBH methods were adapted from Urban Forest Inventory and Analysis (UFIA) procedures (US Forest Service 2016) and urban tree health metrics system (Bond 2012).

Variable	Description	Categories
location	tree location in the landscape, recorded with address and site code plus street segment identifiers (on, to and from streets)	
site type[b]	characteristics of the tree's immediate location, reflects planting environment	sidewalk cut-out, sidewalk planting strip, median, other hardscape, front yard, maintained park[a], other maintained landscaped area
land use[b]	how the property around or adjacent to the tree is used by humans (parcel level)	single-family residential attached, single-family residential detached, multi-family residential, commercial, industrial, institutional, maintained park[a], cemetery, other
species	tree species	genus and species when possible, genus only (if species not identified, unknown (if genus not identified)
mortality status	whether the tree is alive or dead	alive, standing dead (no green leaves or live buds visible)
crown dieback	recent dieback in the upper and outer portion of the crown	0–25%, 26–50%, 51–75%, 76–100%
crown transparency	skylight visible through the live, normally foliated part of the crown	0–25%, 26–50%, 51–75%, 76–100%
wood condition	structural stability of the tree	good, fair, poor, critical
diameter at breast height (DBH)	diameter of the main stem at 1.37m, with multiple stems recorded as a single tree	

[a] "Maintained park" is both a site type and a land use. As a site type, this describes a park-like setting, and as a land use, it describes how the land is used by people. Trees located in park land uses do not necessarily have park site types and vice versa. For example, a tree in a side-walk cut-out could be located in a park land use, or a tree within the right-of-way of a landscaped school yard would be in a park site type and institutional land use.

[b] Additional site type and land use categories are possible particularly for non-street tree projects; we only list categories here that were encountered for the neighborhoods in this pilot study.

the selected areas spanned residential and commercial land uses. In Malmö, the pilot study area was a cemetery, which has rows of trees within the right-of-way that have the same form and function as street trees. Also in Malmö, pre-existing maps of tree locations were already available from other projects. These maps were provided to participants indicating a unique location number for every tree; participants were not provided with any information besides location. Participants in other cities were not provided with tree location maps.

2.2.2. Field crew questionnaire

We distributed an online survey using Qualtrics (Qualtrics version Dec. 2015, Qualtrics Labs, Provo, Utah, US) to all participants (12 per city) to learn about their sociodemographic characteristics and prior relevant experiences, and solicit feedback about the protocols. We explicitly told participants that their field work was part of a pilot test, and that their responses would be used to improve the protocols. After an initial email solicitation, we sent reminders twice to non-respondents. Survey formatting and methods were adapted from Dillman (1999). We asked which variable participants felt most confident reporting; participants selected only one choice from a list of all variables in the protocol. That question was repeated regarding the variable participants felt the least confident reporting. Participants also reported their educational attainment achieved, employment and financial status, and prior urban forestry experience. Lastly, participants were asked, with open-ended questions, to share their suggestions to improve the training and field guide. Specifically, those questions were, "How would you change the training to make it better" and "How would you change the field guide to make it better".

2.3. Data analysis

For each variable in each city, we summarized the proportion of observations consistent with expert observations (hereafter "consistency rates"). Because of the possibility that crews could mistakenly skip trees or record extra trees, we calculated the percent omitted and percent extra out of the correct street tree count, with expert tree counts considered the true value. For species, we reported percent 1) unknown genus, 2) consistent genus (after excluding unknown genera) and 3) consistent species within correct genus (for trees that had species listed by both the experts and participants). We reported these metrics for all trees within a city, and for genera that constituted 5% or more of the sample.

For DBH, we noted disagreement regarding number of stems per tree (e.g., expert recorded a single-stem tree but participants recorded a multi-stemmed tree). For trees that participants and experts recorded as a single stem, we assessed exact DBH agreement with expert measurements and agreement within thresholds of 0.254 cm and 2.54 cm. We propose that values within 0.254 cm of the expert would be appropriate for DBH growth research (e.g., for carbon sequestration or tree growth studies) whereas values within 2.54 cm of the expert would be appropriate for management-focused inventories to understand overall size class structure.

We also compared species, genus and DBH consistency rates to the minimum quality standards from UFIA, which are rigorous standards used to evaluate professional field crew performance. Professional seasonal field crews are checked by highly experienced crews, and those professional crews must meet the standards to get certified to continue working in the program (US Forest Service 2014). The acceptable tolerance in UFIA for genus is correct (i.e., agrees with highly experienced crew) 99% of the time, and for species, 95% of the time. For DBH, values should be within 0.254 cm for every 50.8 cm of diameter at least 95% of the time (US Forest Service 2016).

We present results for novice and intermediate participants within each city for genus, species and DBH, which are the most essential parameters for any urban tree inventory. For genus and DBH, we tested for significant differences among participants using the McNemar exact test, which tests for differences in proportions (in our case, proportion consistent with experts) among paired observations. Analyses were performed in SAS 9.3 (SAS Institute, Cary NC), using the PROC FREQ procedure for the McNemar exact test. This test was done for each pairwise comparison of participant crews within Grand Rapids, Lombard and Philadelphia (e.g., in a given city, intermediate crew 1 vs. every other intermediate and novice crew), for trees which were observed by each pair of crews and the expert. If intermediates were different than novices in DBH and genus performance, we would expect to see significant ($\alpha < 0.05$) p-values for most of the intermediate vs. novice tests. We

Table 3
Field crew performance overview, averaged (with standard deviation) for all six field crews within each city. Extra and omitted trees are presented as percent of the expert tree count. For all columns to the right of omitted trees, values reported are percent agreement with expert observations. Within mortality results, consistency is reported in terms of expert-reported alive and standing dead status (i.e., participants agreed with experts on mortality status for an average of 99.8% of expert-reported alive trees and 100% of expert-reported dead trees). The grand mean is across all cities (without weighting by number of trees per city).

City	Min./tree	Extra trees	Omitted trees	Percent agreement with expert						
				Site type	Land use	Mort.		Dieback	Transp.	Wood cond.
						Alive	Dead			
Lombard, IL	3.4 (2.1)	2.2 (1.7)	2.5 (3.8)	93.1 (3.0)	90.6 (10.4)	99.8 (0.4)	100.0 (0.0)	84.7 (15.3)	63.2 (5.9)	22.9 (7.3)
Grand Rapids, MI	2.3 (0.6)	0.3 (0.6)	0.5 (0.6)	97.9 (1.1)	79.1 (5.7)	100.0 (0.0)	100.0 (0.0)	87.5 (12.5)	74.1(3.9)	55.1 (2.8)
Philadelphia, PA	3.0 (0.4)	0.5 (1.3)	0.7 (0.9)	85.8 (3.1)	89.4 (5.9)	99.4 (0.7)	100.0 (0.0)	91.5 (3.8)	72.6 (4.5)	71.6 (7.3)
Malmö, Sweden	3.3 (1.0)	n/a	n/a	83.3 (40.8)	96.1 (6.9)	100.0 (0.0)	100.0 (0.0)	96.9 (2.7)	76.1 (9.3)	71.3 (9.7)
Mean	3.0	1.0	1.2	90.0	88.8	99.8	100.0	90.2	71.5	55.2

did not apply the McNemar test to Malmö because most trees were not observed by all participants.

The remainder of our analysis focused on evaluating descriptive statistics and graphical data summaries as recommended by Bland and Altman (1996a,b). For variables besides genus, species and DBH, where summaries indicated no practical differences among experience levels (i.e., consistency rates across experience levels have similar ramifications for potential data uses), we reported only descriptive statistics. We also averaged consistency rates across all cities for each variable. Outliers were included in these descriptive statistics. We noted specific causes of observation disagreements wherever possible, based on experts' experiences with training as well as our time spent doing the inventory ourselves with trees that crews found challenging.

From the participant survey, we summarized participant characteristics for each city. Open-ended responses were qualitatively assessed for common themes by a single analyst, counting the number of times participants mentioned similar ideas (Babbie 2007). Themes were not pre-determined. Results for the multiple-choice questions are presented as a percent of the total number of survey respondents per city, and results for the open-ended questions are reported across all cities.

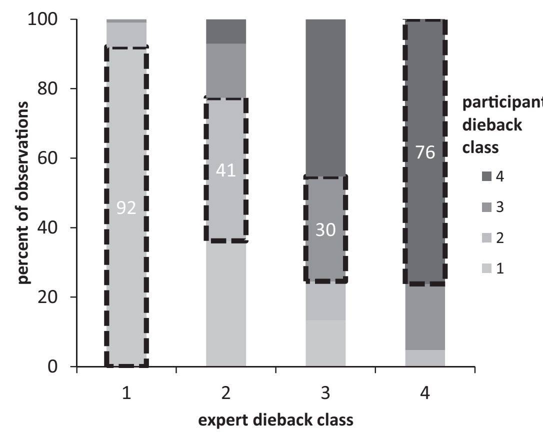

Fig. 2. Percentages indicate relative of number of observations falling into each combination of expert and participant dieback classes. The highlighted bar sections (black dashed outline), with associated percentages, indicate where the participants and experts were in agreement. There were 2782 observations in expert dieback class "1" (0%–25% dieback), 323 observations in class "2" (26%–50%), 53 observations in class "3" (51%–75%), and 21 observations in class "4" (76%–100%).

3. Results

3.1. Participant field work performance

Average time per tree was 3.0 min (Table 3). Trees were omitted and/or extra trees were counted by 6/6 crews in Lombard, 5/6 crews in Grand Rapids and 4/6 crews in Philadelphia, with 1.0% omitted trees and 1.2% extra trees on average. Most commonly, a few trees were accidentally skipped in the midst of other trees recorded properly. In Lombard, one intermediate crew missed 14 trees on two entire city blocks, and also counted several extra trees that were not within the scope of their assigned area. In another example of extra trees, an intermediate crew in Philadelphia recorded small seedlings that had naturally sprouted within old decaying tree trunks as separate trees. The same crew also recorded a multi-stemmed tree (*Laegerstromia indica*) whose piths joined below ground as three individual trees. The expert crew and all other crews did not record the seedlings and recorded that multi-stemmed tree as a single tree. In Malmö, we could not compare overall tree counts since each crew went to a subsample of the area.

Participants' classifications of site type and land use were consistent with expert classifications for 90.0% and 88.8% of trees, respectively (Table 3). The study areas in most cities had a mix of residential, commercial, institutional and industrial, except Malmö. All trees in Malmö had site type "maintained park" (because the cemetery was a park-like setting, see Table 2) with land use "cemetery". Site type had the lowest level of consistency for this city, due to one crew misclassifying all trees as "other maintained landscaped area." Philadelphia's lower than average site type agreement was largely due to classification discrepancies between "sidewalk cut-out" and "planting strip."

With respect to mortality status, there were very few standing dead trees through which to evaluate inconsistencies: experts recorded only two standing dead in Lombard and two in Philadelphia, and none in Grand Rapids and Malmö. Mortality status inconsistencies (Table 3) occurred rarely for misclassifications of live trees as standing dead (99.8% average consistency); the reverse misclassification did not happen.

Dieback rating (Table 2) consistency was 90.7% overall (Table 3), and was most consistent for the lowest and highest dieback classes (Fig. 2). However, for the middle dieback classes, participants agreed with the expert on less than half of the trees. Participants were more likely than experts to place trees in the lowest and highest dieback classes. It should also be noted that most trees had little dieback, with 88% of observed trees classified by experts as having 0–25% dieback, so the agreement percentages of higher dieback classes are based on relatively few trees. Transparency and wood condition had the least consistency between participants and experts in every city (Table 3).

Participants recorded unknown genera for 1.0% of trees on average, and agreed with expert evaluation of genera for 90.7% of trees, and species agreement (where genera was in agreement) was 84.8% (Table 4). Among the more common genera (Table 5), *Acer*, *Aesculus*, *Crataegus*, *Gleditsia*, *Platanus*, and *Tilia* had particularly high performance (>96% consistency in at least one city) whereas *Amalanchier*,

Table 4
Participant performance for genus, presented as percent agreement with expert observations, averaged across novice (N), intermediate (I) and all participants within a city. The grand mean is across all participant pairs in all cities (without weighting by number of trees per city).

City	Unknown genus			Genus[a]			Species[b]		
	N	I	All	N	I	All	N	I	All
Lombard, IL	1.5 (2.1)	2.1 (1.8)	1.8 (1.8)	82.1 (2.1)	76.2 (6.2)	79.1 (5.2)	74.3 (13.1)	61.6 (11.7)	67.9 (13.1)
Grand Rapids, MI	0.9 (1.6)	0.0 (0.0)	0.5 (1.1)	91.9 (4.1)	89.9 (5.6)	90.5 (4.4)	84.5 (1.7)	88.4 (2.0)	86.4 (2.7)
Philadelphia, PA	2.0 (3.4)	1.5 (0.4)	1.8 (2.2)	90.3 (3.7)	98.0 (1.4)	94.1 (4.9)	89.9 (5.2)	90.7 (4.4)	90.3 (4.3)
Malmö, Sweden	0.0 (0.0	0.0 (0.0)	0.0 (0.0)	99.5 (0.8)	98.9 (1.1)	99.2 (1.0)	95.5 (1.2)	93.6 (2.7)	94.6 (2.1)
Mean	1.1	0.9	1.0	91.0	90.8	90.7	86.1	83.6	84.8

[a] Percent genus agreement, excluding unknown genera.
[b] Percent species agreement, within correct genera, for trees that had species listed by both experts and participants.

Table 5
Participant performance for genera within each city that comprised 5% or greater of that city's measured trees. Performance is presented as percent agreement with expert observations, averaged across novice (N), intermediate (I) and all participant pairs within a city. The (n) for each genus represents the count of individual trees with that genus for each city.

City	Genus (n)	N	I	All
Lombard, IL	*Acer* (38)	100.0 (0.0)	93.9 (4.0)	96.9 (4.2)
	Tilia (28)	91.2 (9.0)	89.7 (3.6)	90.5 (6.2)
	Syringa (19)	68.3 (22.8)	40.2 (36.3)	54.3 (31.2)
	Gleditsia (15)	97.8 (3.8)	100.0 (0.0)	98.9 (2.7)
	Fraxinus (14)	100.0 (0.0)	61.9 (10.9)	81.0 (22.0)
	Amalanchier (9)	0.0 (0.0)	50.9 (44.1)	25.5 (39.5)
	Other (34)	68.2 (18.6)	68.5 (12.3)	68.4 (14.1)
Grand Rapids, MI	*Acer* (98)	100.0 (0.0)	99.3 (1.2)	99.7 (0.9)
	Prunus (8)	60.1 (24.0)	37.5 (12.5)	48.8 (21.1)
	Tilia (8)	100.0 (0.0)	100.0 (0.0)	100.0 (0.0)
	Other (33)	68.8 (15.2)	72.7 (18.4)	70.8 (15.3)
Philadelphia, PA	*Acer* (59)	98.3 (1.7)	99.4 (1.0)	98.9 (1.4)
	Platanus (20)	100.0 (0.0)	100.0 (0.0)	100.0 (0.0)
	Crataegus (12)	100.0 (0.0)	100.0 (0.0)	100.0 (0.0)
	Prunus (12)	55.6 (38.5)	100.0 (0.0)	77.8 (34.4)
	Quercus (9)	100.0 (0.0)	100.0 (0.0)	100.0 (0.0)
	Tilia (8)	95.8 (7.2)	62.5 (54.5)	79.2 (39.3)
	Other (32)	75.7 (10.2)	91.9 (7.3)	83.8 (11.9)
Malmö, Sweden	*Tilia* (109)	100.0 (0.0)	99.3 (1.1)	99.7 (0.8)
	Aesculus (28)	100.0 (0.0)	98.2 (3.0)	99.1 (2.1)
	Acer (12)	93.3 (11.5)	96.3 (6.4)	94.8 (8.5)
	Other (2)	100.0 (0.0)	100.0 (0.0)	100.0 (0.0)

Prunus, and *Syringa* had the lowest performance (<60% consistency).

For DBH, participants agreed with the number of stems per tree reported by the expert for 91.1% of trees on average, but 100.0% of the time in Malmö, which had only single-stem trees. For trees recorded as single-stem by the expert and all other participants, DBH matched exactly for 20.2% of trees. Participant DBH values were within 0.254 cm of expert DBH for 54.4% of trees, and within 2.54 cm for 93.3% of trees. Using the UFIA acceptable tolerance thresholds, 66.0% of participant single-stem DBH measurements met data the quality requirement. Outliers appear to be due to transcription or recording error; for example, transposing numbers, mistyping, and dropped decimal points. Participants generally recorded DBH as slightly higher than experts (Fig. 3) with an overall median +0.25 cm, mean +0.33 cm.

In the McNemar tests for DBH and genus performance differences among crews, there were no patterns of consistent differences among intermediate and novice crews (see supplemental materials). Most pairwise comparisons of crews were not statistically significant, but a few crews stood out as particularly distinctive from the others. For example, one intermediate crew in Philadelphia was significantly different from all other crews in that city for DBH (0.254 cm and UFIA thresholds), and that crew had the highest level of consistency in Philadelphia (79.5% of trees within 0.254 cm of experts) and, indeed, across all cities. For genus, one

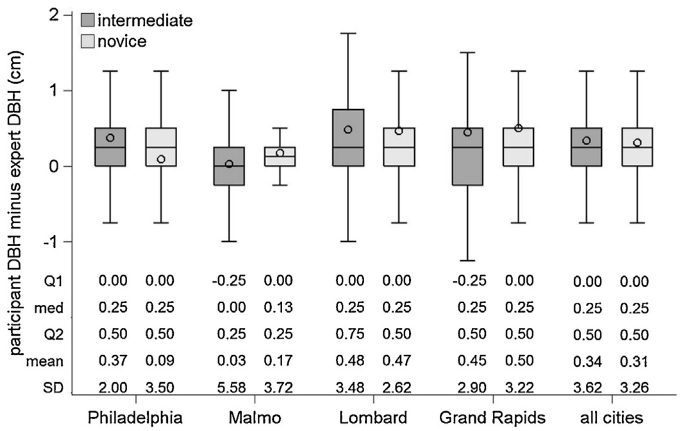

Fig. 3. Box plots summarizing the distributions of differences between participant and expert measurements on the same trees. The boxes indicate the first, second, and third quartiles while the whiskers indicate the minimum and maximum values within 1.5 times of the interquartile range of the first and third quartiles respectively. Circles reperesent the mean. Outliers were included in the calculations but are not shown on this graph.

intermediate crew in Grand Rapids (83.4% consistent with expert), one intermediate crew in Lombard (69.1%), one novice crew in Philadelphia (86.0%), and one intermediate crew in Philadelphia (99.3%) were significantly different from several other crews; the former three crews had the lowest genus performance in their respective cities, while the latter had the highest.

Across other variables with the descriptive statistics, we also did not see patterns that intermediate participants were consistently performing differently than novice participants.

3.2. Participant questionnaire

Response rates for the questionnaire were 83% in Lombard (n = 10), 92% in Grand Rapids (n = 11), 83% in Philadelphia (n = 10), and 67% in Malmö (n = 8). Across all cities, participants felt most confident in reporting DBH, with those in Philadelphia also feeling confident about mortality status. Crews in Lombard and Grand Rapids had the least confidence in species identification, while crews in Philadelphia and Malmö had the least confidence in transparency and wood condition.

Participants had high levels of educational attainment, with a majority having a bachelor's degree or higher in all cities (Tables 6 and 7). Most participants in Lombard were retired, while most in Grand Rapids and Philadelphia were employed full-time. In terms of racial and economic diversity, field crews were almost entirely white, with a range of self-reported financial status. Participants in Philadelphia had the most prior experience with tree care and tree identification courses, whereas participants in Lombard had the least (Table 8).

Table 6
Participant performance for diameter at breast height (DBH), averaged across novice (N), intermediate (I) and all participants within a city. Stem count reflects percent agreement with the expert about the number of stems for each tree. DBH consistency evaluation was limited to trees deemed single-stem by both the expert crew and the other crews. DBH exact reflects percent exact agreement with expert measurements and tolerances within the specified thresholds of expert measurements. The grand mean is across all cities (without weighting by number of field crews per city).

City	stem count			DBH exact			DBH within 0.254 cm			DBH within 2.54 cm		
	N	I	All	N	I	All	N	I	All	N	I	All
Lombard, IL	82.6 (1.5)	81.9 (0.3)	82.3 (1.0)	9.1 (2.6)	5.7 (3.1)	7.4 (3.2)	47.3 (6.1)	45.6 (4.2)	46.4 (4.8)	90.3 (4.9)	89.9 (3.2)	90.1 (3.7)
Grand Rapids, MI	90.9 (2.6)	90.6 (1.7)	90.8 (1.9)	22.5 (2.2)	18.8 (1.8)	20.7 (2.7)	50.2 (4.3)	47.2 (0.9)	48.7 (3.2)	92.2 (2.0)	91.7 (2.3)	91.9 (1.9)
Philadelphia, PA	90.7 (2.5)	91.9 (4.4)	91.3 (3.2)	27.6 (2.3)	28.4 (8.9)	28.0 (5.8)	59.3 (5.3)	65.2 (12.9)	62.3 (9.4)	94.9 (0.8)	96.4 (2.1)	95.6 (1.6)
Malmö, Sweden	100.0 (0.0)	100.0 (0.0)	100.0 (0.0)	24.9 (7.1)	24.2 (11.3)	24.6 (8.5)	60.6 (8.6)	60.1 (10.6)	60.3 (8.6)	95.2 (0.8)	96.0 (10.6)	95.6 (0.8)
Mean	91.1	91.1	91.1	21.0	19.3	20.2	54.3	54.6	54.4	93.1	93.5	93.3

Table 7
Participant socioeconomic characteristics (% participants within each city). Educational attainment refers to the proportion with a bachelor's or master's degree or higher. Respondents could select more than one option for employment status and race/ethnicity.

City	Education		Employment status				Race/ethnicity		Financial status		
	Bach.	Mast.	Stud.	Empl. part-time	Empl. full-time	Retired	White	Non-white	Struggl-ing	Stable	Comfortable-Affluent
Lombard, IL	80	50	20	20	10	60	100	0	0	20	80
Grand Rapids, MI	72	27	9	9	73	18	100	0	0	64	36
Philadelphia, PA	90	30	40	20	50	0	80	20	20	50	30
Malmö, Sweden	63	13	38	38	38	0	88	12	25	50	25

Table 8
Prior urban forestry experiences of the participants (% participants within each city).

City	Tree care course	Tree ID course	Volunteer	Job	Hobby
Lombard, IL	30	0	70	20	60
Grand Rapids, MI	82	91	73	9	64
Philadelphia, PA	45	60	36	30	18
Malmö, Sweden	38	76	0	38	75

To improve training, 45% of participants suggested clarifying protocol methods, particularly DBH, condition, and site type; 33% suggested spending more time on species identification; and 30% suggested adding more time to practice data collection in the real world. To improve the protocols themselves and accompanying field guide, 45% suggested adding more examples and pictures, especially about species identification and condition; 23% suggested alternative ways to organize the field guide, such as tabs to help find key information during field work or a one page "cheat sheet", and 23% thought that the field guide was good. A majority (81%) would prefer a mobile interface to paper data collection, however some participants still prefer paper. Reasons for preferring paper include concerns that not all volunteers would have access to a smartphone or tablet, and that paper may be easier and more reliable in terms of workflow.

4. Discussion

Our results indicate that minimally trained field crews produce data that are not entirely consistent with expert observations, with some variables being more inconsistent than others. We will discuss implications of our findings in terms of revisions to the field protocol, participant training, and program models for citizen science in urban forestry. We also raise appropriate uses of volunteer data for the particular variables we evaluated, because urban tree inventories have a wide range of purposes that demand different levels of accuracy.

The average time per tree in our study, 3 min, compares to 2 min in Bloniarz and Ryan (1996) and 6 min in Cozad (2005). The latter appears to be larger because it included transportation time as well as several additional variables related to street tree management. We therefore conclude that, for the field work itself, data collection using a limited suite of variables is quite fast.

Issues of omitted and extra trees were small on average, yet with some notable outliers of participants that missed substantial numbers of trees. With repeated monitoring, issues of omitted and extra trees would compound, as tree counts would not be consistent across years and across crews; such problems would be especially problematic to calculating rates of change such as population growth or mortality. We contend that there is a need to be stricter about how "tree" and "street tree" are defined. Our training included street tree definitions for each city, therefore other inventories with no explicit "street tree" definition may result in greater tree count problems, as was documented in a street tree census in Oakland, CA (Roman et al., 2014). In the prior study by Cozad (2005), volunteers missed 2.2% of trees recorded by professionals, an omission rate higher than our average; that study did not indicate whether any particular definition of street tree was given to volunteers. In our study, we did not use size-based criteria for inclusion (e.g., minimum DBH 2.54 cm) because such criteria would exclude newly planted trees highly relevant to management. However, size-based criteria would be appropriate if consistent repeated tree population counts over many years are a priority.

Although stricter definitions of "street tree" may help, we noted that some inconsistencies were actually because of accidental omission. For example, participants may omit trees when they are distracted talking to pedestrians. These problems could potentially be addressed through quality control: have participants double-check that they have recorded every tree on a block before considering that block's data final. Additionally, participants could be given more photo examples of what trees should be considered "in" and "out" of the study scope. Another potential solution would be for experts to affix temporary flagging or permanent identification tags to trees to indicate which are included in an inventory. Tree identification tags have been employed with young tree survival monitoring in cities (Roman et al., 2015). Indeed, tagged tree inventories are standard practice in forest ecology research (van Doorn 2014) and could be more widely applied in urban forests, yet the convenience of tree tags for data collection should be weighed against risks of vandalism and perceived aesthetic problems. Indeed, none of these solutions are foolproof. Even in permanent forest research plots with tagged trees, there can be occasional problems of extra and omitted trees, particularly at plot edges (van Doorn 2014).

Site type and land use inconsistencies could be addressed in several ways. Crews asked for more examples, particularly for site type. The specific misclassifications that we observed in Philadelphia (e.g., "sidewalk cut-out" vs. "sidewalk planting strip") could be dealt with by providing more photo examples in the field guide, as requested by some participants, and by making the definitions for those categories clearer with exact size delineations for available surface soil. For urban tree inventory projects seeking to connect tree growth and health with site conditions, additional site measurements are needed (Sanders and Grabosky 2014; Scharenbroch and Catania 2012 Scharenbroch and Catania 2012) and their suitability for citizen science has not been evaluated. For land use, cities may also merge field tree inventories with municipal land use data. Yet such municipal records may not always be available or up-to-date. Our site type and land use categories, while they need some clarifications, seem to be relatively simple for volunteers to record, and would be useful at a coarse level to understand the trees' physical setting.

With respect to mortality, we assert that having a firm definition of standing dead is critical for monitoring projects. We used a strict definition of standing dead: no green leaves or live buds visible (Table 2). A few participants misclassified live trees as standing dead; these are trees that had substantial missing foliage and we suspect that participants did not examine the crown thoroughly enough for leaves. Participants should be shown more examples of nearly dead trees during training – unhealthy trees that should be classified alive – to further promote consistency. Because mortality analysis is often a goal of monitoring programs (Roman et al., 2013), this issue is especially relevant for repeated inventories.

Crown transparency and wood condition had the poorest performance and participants expressed concerns with these variables. Therefore, we conclude that these variables should be dropped from future citizen science projects. Wood condition may also be troublesome for municipal arborists to include in citizen science inventories because of liability implications of tree risk, and our evidence that volunteers vary widely in their evaluation of structural stability. We suggest that dieback should be the preferred metric for tree condition in future citizen science tree inventories. Dieback is meant to evaluate recent stress (US Forest Service 2016), and we specifically defined this parameter as focusing on recent growth in the upper and outer portions of the crown (Table 2). However, further research is needed to determine participant consistency with dieback depending on the classes used: 25% classes (as our project did), 20% (following Vogt and Fischer 2014), or 5% (following i-Tree Eco and UFIA methods). Additionally, our dieback findings suggest that participants are proficient at evaluating the high and low dieback categories but could not be relied upon to consistently classify trees in the middle dieback categories. Participants' most frequent disagreements with experts came by classifying trees as having high dieback where the experts recorded less dieback. In practical terms for urban forest managers, this error would lead to more trees classified as being stressed and potentially requiring attention. We suggest that additional photo examples should be offered since participants encountered relatively few trees outside lowest dieback class in the field. Additionally, it is possible that field crews were including large branch mortality in their dieback evaluations, so we suggest that during training, instructors should emphasize the meaning of dieback as being about recent stress in the upper and outer portions of the crown.

Genera consistency rates were 90.7% across all cities, with species consistent 84.6% within consistent genera. In general, crews in Malmö had the best performance for genera identification, yet their task was also relatively simpler, with roughly one third as many genera as the other cities. Crews in Lombard and Grand Rapids felt particularly under-prepared for species identification, which aligns with the lower rates of consistent genus and species identification in those cities. This suggests the potential for self-evaluation by citizen scientists after their field work, to help flag crews that may have more observation errors with species identification. Participants also asked for more species identification guides and training. There is a need for better regionally-relevant species identification resources (e.g., online training materials, field guides) for urban trees that are accessible to novices.

Our species and genera consistency levels did not meet UFIA tolerance thresholds (which are for paid and professionally certified field crews), but were similar to Bloniarz and Ryan (1996), which found genus consistency rates of 91–96% for common genera. For the same genera reported in that study (*Acer*, *Fraxinus*, *Gleditsia*, *Quercus*, *Platanus*, *Tilia*), our results were as good or better for all genera except *Tilia* in Philadelphia. Our species consistency rates were also similar to the 80.2% volunteer accuracy reported in Cozad (2005). We therefore conclude that our levels of consistency for tree identification are in line with those reported in these prior street tree studies even though our participants received less training time. Direct comparisons with volunteer species accuracy in other studies are challenging because of the different nature of participant tasks and different ways of reporting results. For example, volunteers had 72% species accuracy for six invasive species after a day of training (Crall et al., 2011), and 79–92% species accuracy for five invasive species depending on training mode (Starr et al., 2014).

If correct species identification is paramount for an urban tree inventory project, then relying on minimally trained volunteers may not be appropriate; but if mostly correct genera identification is sufficient – for common genera in particular – then volunteer data may be adequate. Additionally, volunteers may be appropriate for monitoring recently planted trees, since species information would be available from planting records and therefore volunteers would not need species identification skills. For volunteer inventories that do ask participants to carry out species identification, alternate program models may lead to improved performance. For example, while citizen scientists collect tree inventory data in Portland, OR, arborists circulate on bicycle to assist with questions (A. diSalvo, pers. comm.). Volunteers could also take photos for species validation by experts (Wiggins et al., 2011) and/or evaluation with automated visual recognition (Kumar et al., 2012). Such an approach would require a streamlined mobile data collection and management system (Boyer et al., 2016) as well as specific protocols regarding taking photos of the tree for identification purposes.

Finally, participants struggled with DBH protocols for multi-stemmed trees, as revealed by their survey feedback as well as the consistency rates for per-tree stem counts (Tables 6 and 7). The multi-stemmed DBH method we borrowed from the UFIA Field Guide (US Forest Service 2016), which involves measuring several stems according to where piths join, appears to be too challenging for minimally trained participants. The DBH protocols from i-Tree Eco (www.itreetools.com) call for up to six stems of at least 2.54 cm DBH to be recorded. Based on our feedback from participants, we suggest simplifying multi-stemmed DBH protocols based on tree growth habit. Tree types that tend to fork around 1.37 m (e.g., *Crataegus*, *Malus, Prunus*, *Pyrus*, *Zelkova*) should have their DBH recorded lower than 1.37 m, where there is a single stem. This DBH method may promote consistency across volunteer field crews and would be appropriate to group multi-stemmed trees by size class. Indeed, this simplified DBH approach is already used by professional arborists (Bernhardt and Swiecki 2001). However, further research is needed to evaluate implications of various multi-stemmed DBH methods for urban trees for objectives such as radial growth, allometry and biomass estimation, as has been done for mangrove forests (Clough et al., 1997).

For trees recorded as a single stem by all crews, participant DBH values were within 0.254 cm of expert values 54.4% of the time, within 2.54 cm 93.3% of the time, and volunteer DBH performance met UFIA data quality standards for 66.0% of trees. Cozad (2005) reported 76.5% accuracy for DBH size class (15.24 cm bins), and our performance was higher with the narrower threshold of within 2.54 cm of expert values. We cannot be certain why our DBH performance was better than Cozad (2005), but it is possible that our approach to training may have helped: we highlighted common mistakes and best practices for using diameter tape, and concluded the training with an open discussion about possible sources of error.

Research applications of street tree inventories (e.g., ecosystem services evaluations, tree growth studies) generally require greater DBH accuracy and precision than management applications (e.g., understanding overall size class distribution). Given that our participants had only one hour of training on DBH followed by practice on a few trees, we think that these findings are encouraging for the widespread use of volunteers to measure urban tree DBH for applied management purposes, where coarse precision (e.g., nearest 2.54 cm) is acceptable. Our findings indicate that volunteer DBH measurements may not be adequate for radial growth monitoring at the sub-centimeter level. Volunteer DBH data might be acceptable for ecosystem services estimates, provided that observation error rates are propagated into the models (Chave et al., 2004). In general, we could not distinguish between DBH errors because of mismeasurement, recording error in the field, or transcription error during data entry, except for a few outliers where the cause was obvious (e.g., 2.1 became 21). Lastly, participant DBH was slightly larger than expert DBH on average (+0.33 cm), indicating systematic bias. We suspect this may be due to experts pulling DBH tape more snugly around the trunk and/or experts being more careful about holding the tape perpendicular to the trunk. Yet this bias is small enough to have negligible practical impact on volunteer DBH measurements used for management purposes.

In general, we did not find evidence that novice and intermediate crews performed differently in terms of data consistent with the experts. In prior studies, higher levels of prior knowledge and education have sometimes been associated with higher data quality, and sometimes not (Lewandowski and Specht 2015). Self-reported high level of comfort was related to better identification of invasive plant species (Crall et al., 2011). Yet in a study of interobserver variability for forest pest detection, participants with higher experience levels had more false detections (Fitzpatrick et al., 2009), and volunteers with higher skill level had more false positives for rare bird detection (Farmer et al., 2012). Indeed, even when inexperienced volunteers generally perform worse than experienced researchers, certain volunteers can produce higher quality data than some researchers (Bernard et al., 2013). In our study, the lack of consistent patterns of intermediates differing from novices suggests that both types of volunteers can produce similar data quality in general for street tree inventories. However, the existence of low-performing crews for genus identification among both intermediates and novices suggests that, to promote higher data quality, all crews should be tested for identification skills, to note those that need additional training. Indeed, in Malmö, we saw that novices had better species identification than intermediates. In this particular case, there is a straightforward explanation for the better performance among novices. Most *Tilia* sp. trees were *Tilia x europaea*. Novice participants knew *Tilia x europaea* best, and wrote it down for nearly all cases of this genus, therefore they had high performance for correct species. We think that intermediate crews were familiar with more species but also appeared to be more cautious; they sometimes guessed incorrectly and other times wrote *Tilia* sp. to acknowledge their uncertainty.

Future research could evaluate predictors of species and genera misidentification as well as other observation errors, such as participant characteristics and genus or species rarity. Our analysis assumed that expert data is "correct" which follows the precedent of previous citizen science research (Bloniarz and Ryan 1996; Cozad 2005; Lewandowski and Specht 2015), but nonetheless could be evaluated in-depth in future studies. New research could investigate the interobserver variation of expert vs. participant data to determine whether experts are consistent among themselves. Controlled experiments contrasting different methods for species identification training, day-of support and validation (Starr et al., 2014) could reveal optimal approaches for urban forestry. Finally, new studies could evaluate the impact of citizen science on the volunteers, their communities, and urban forestry programs. While some authors have asserted that volunteering for street tree inventories and monitoring can build constituencies for urban forestry programs (McPherson 1993; Bloniarz and Ryan 1996), this has not yet been evaluated. New research could reveal whether volunteer outcomes found in the urban environmental stewardship literature for activities such as tree planting (Westphal 2003; Fisher et al., 2015; Ryan 2015) are similar to volunteer outcomes in citizen science for tree inventories, and also evaluate strategies for engaging underrepresented groups in citizen science for urban forestry (Pandya 2012). As urban forest managers decide whether or not to pursue citizen science, both data quality and social outcomes are relevant.

There are several limitations to our study. First, the study sites were chosen based solely on interest from UTGL members and were limited geographically to temperate cities in the United States and Europe. Urban forest and community characteristics in different cities (e.g., species diversity, volunteer education level) may impact data quality. Second, our participants used paper data collection, and it is possible that some transcription errors were introduced during data entry, which we did not evaluate. Third, as in the real world of urban tree inventories, there were slight differences in how our pilot study manifested across the four cities, including different definitions of "street tree" and different situations with field crews. Nevertheless, we did find many areas of consensus with Bloniarz and Ryan (1996) and Cozad (2005), so we think that the conclusions here are relevant to volunteer street tree inventories more broadly.

Quality assurance is paramount to the success of volunteer-collected inventories. Our study provides raw quantified rates of observation error that can be used to help urban forest researchers and managers decide whether citizen science might be appropriate for their needs. For any urban forestry programs that wish to use volunteer data, it is important to document and report accuracy rates, as well as to use that data quality information in larger analyses. Project-specific accuracy rates should be regularly evaluated for citizen science protocols to ensure that volunteer performance meets the desired uses of the data (Danielsen et al., 2005; Crall et al., 2011).

5. Conclusion

Citizen science shows promise for urban forestry management and research, with important caveats about matching data quality with intended uses of the data. We conclude that citizen science is a viable option for some urban tree inventory and monitoring projects, but not for projects that require extremely high accuracy with species identification and DBH. Additional resources should be developed to support volunteer engagement in urban tree data collection, such as species identification guides, streamlined mobile data collection systems, and best practices for training. As with any field-based data collection effort, it is critical to understand field crew performance, refine methods based on evaluations of that performance, and implement quality assurance and data validation procedures to document and enhance data quality. Because task complexity, volunteer backgrounds and intended data uses can vary so widely across citizen science projects, it is essential to evaluate data quality within each project.

Acknowledgements

We thank many UTGL members for their contributions to the monitoring protocols, especially those who developed the minimum data set: E.G. McPherson, P.P. Peper, J.R. Mills, J. Bond, E. King, B. Fischer, M. Bigger, J. Karps, D. Wildman and S. Kremske. We thank L. Shafer for designing the draft field guide and E. Desotelle for assistance wi$th training materials. The participant survey was carried out and analyzed with assistance from H. Finley and T. Omolo. We are grateful to the many participants in this pilot study and many more volunteers and interns who contribute to other urban tree inventories. We also thank two anonymous reviewers whose comments strengthened the manuscript. This project was financially supported by the US Forest Service Philadelphia Field Station and The Morton Arboretum.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ufug.2017.02.001.

References

Östberg, J., Sjögren, J., 2016. The linear index of tree appraisal (LITA) model for economic valuation of large urban trees in Sweden. Arboriculture Urban For. 42, 21–30.
Babbie E.R. 2007. The basics of social research. Wadsworth: 576 pp.
Bernard, A.T.F., Götz, A., Kerwath, S.E., Wilke, C.G., 2013. Observer bias and detection probability in underwater visual census of fish assemblages with independent double-observers. J. Exp. Mar. Biol. Ecol. 443, 75–84.
Bernhardt, E.A., Swiecki, T.J., 2001. Guidelines for developing and evaluating tree ordinances. Developed by Phytosphere Research. Available at http://www.isa-arbor.com/education/resources/educ_treeordinanceguidelines.pdf.
Bird, T.J., Bates, A.E., Lefcheck, J.S., Hill, N.A., Thomson, R.J., Edgar, G.J., Stuart-Smith, R.D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J.F., Pecl, G.T., Barrett, N., Frusher, S., 2013. Statistical solutions for error and bias in global citizen science datasets. Biol. Conserv. 173, 144–154.
Bland, J.M., Altman, D.G., 1996a. Measurement error proportional to the mean. BMJ 313, 106.
Bland, J.M., Altman, D.G., 1996b. Statistical notes: measurement error and correlation coefficients. BMJ 313, 41.
Bloniarz, D.V., Ryan, H.D.P., 1996. The use of volunteer initiatives in conducting urban forest resource inventories. J. Arboric. 22, 75–82.
Bond, J., 2012. Urban Tree Health: A Practical and Precise Estimation Method. Urban Forest Analytics, Geneva, NY, 107 pp.
Bond, J., 2013. Best Management Practices –Tree Inventories, 2nd ed. International Society of Arboriculture, Champaign, IL, 35 pp.
Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V., Shirk, J., 2009. Citizen science: a developing tool for expanding scientific knowledge and scientific literacy. BioScience 59, 977–984.
Boyer, D.J., Roman, L.A., Henning, J.G., McFarland, M., Dentice, D., Low, S.C., Thomas, C., Abrams, G., 2016. Data Management for Urban Tree Monitoring –software Requirements. Azavea, Philadelphia, PA, Available online at http://www.azavea.com/research/urban-tree-monitoring, 124pp.
Butt, N., Slade, E., Thompson, J., Mahli, Y., 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.
Campbell, L.K., Svendsen, E.S., Roman, L.A., 2016. Co-production of knowledge at the research-practice interface: case studies from urban forestry. J. Environ. Manage. 57, 1262–1280.
Campbell, J., 2015. NYC's decennial street tree census will enter the digital age this year. The Village Voice, Apr. 30, 2015. Available at http://www.villagevoice.com/news/nycs-decennial-street-tree-census-will-enter-the-digital-age-this-year-6713834.
Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., Perez, R., 2004. Error propagation and scaling for topical forest biomass estimates. Philos. Trans. R. Soc. B 359, 409–420.
Clough, B.F., Dixon, P., Dalhaus, O., 1997. Allometric relationships for estimating biomass in multi-stemmed mangrove tress. Aust. J. Bot. 45, 1023–1031.
Conrad, C.C., Hilchey, K.G., 2011. A review of citizen science and community-based environmental monitoring: issues and opportunities. Environ. Monit. Assess. 176, 273–291.
Cooper, C.B., Dickinson, J., Phillips, T., Bonney, R., 2007. Citizen science as a tool for conservation in residential ecosystems. Ecol. Soc. 12, 11.
Cozad, S., 2005. STRATUM case study evaluation in Minneapolis, Minnesota. In: Masters Thesis. University of California, Davis, 81 pp.
Crall, A.W., Newman, G.J., Stohlgren, T.J., Holfelder, K.A., Graham, J., Waller, D.M., 2011. Assessing citizen science data quality: an invasive species case study. Conserv. Lett. 4, 433–442.
Crall, A.W., Jordan, R., Holfelder, K., Newman, G.J., Graham, J., Waller, D.M., 2013. The impacts of an invasive species citizen science training program on participant attitudes, behavior, and science literacy. Public Underst. Sci. 22, 745–764.
Danielsen, F., Burgess, N.D., Balmford, A., 2005. Biodiv. Conserv. 14, 2507–2542.
Danielsen, F., Adrian, T., Brofeldt, S., van Noordwijk, M., Poulsen, M.K., Rahayu, S., Rutishauser, E., Theilade, I., Widayati, A., An, N.T., Bang, T.N., Budiman, A., Enghoff, M., Jensen, A.E., Kurniawan, Y., Li, Q., Mingxu, Z., Schmidt-Vogt, D., Prixa, S., Thoumtone, V., Warta, Z., Burgess, N., 2013. Community monitoring for REDD+: international promises and field realities. Ecol. Soc. 18, 41.
Dawson, J.O., Khawaja, M.A., 1985. Change in street-tree composition of two Urbana, Illinois neighborhoods after fifty years: 1932–1982. J. Arboric. 11, 344–348.
Delaney, D.G., Sperling, C.D., Adams, C.S., Leung, B., 2008. Marine invasive species: validation of citizen science and implications for national monitoring networks. Biol. Invasions 10, 117–128.
Dickinson, J.L., Zuckerberg, B., Bonter, D.N., 2010. Citizen science as an ecological research tool: challenges and benefits. Annu. Rev. Ecol. Evol. Syst. 41, 149–172.
Dickinson, J.L., Shirk, J., Bonter, D., Bonney, R., Crain, R.L., Martin, J., Phillips, T., Purcell, K., 2012. The current state of citizen science as a tool for ecological research and public engagement. Front. Ecol. Environ. 10, 291–297.
Dillman, D.A., 1999. Mail and Internet Surveys: The Tailored Design Method, 2nd ed. Wiley & Sons, New York, NY, US, 480 pp.
Edgar, G.J., Stuart-Smith, R.D., 2009. Ecological effects of marine protected areas on rocky reef communities – a continental-scale analysis. Mar. Ecol. Prog. Ser. 388, 51–62.
Farmer, R.G., Leonard, M.L., Horn, A.G., 2012. Observer effects and avian-call-count survey quality: rare-species biases and overconfidence. Auk 129, 76–86.
Ferretti, M., 2009. Quality assurance in ecological monitoring – towards a unifying perspective. J. Environ. Monit. 11, 726–729.
Fisher, D.R., Svendsen, E.S., Connolly, J., 2015. Urban Environmental Stewardship and Civic Engagement: How Planting Trees Strengthens the Roots of Democracy. Routledge, New York, NY, 164 pp.
Fitzpatrick, M.C., Preisser, E.L., Ellison, A.M., Elkinton, J.S., 2009. Observer bias and the detection of low-density populations. Ecol. Appl. 19, 1673–1679.
Gardiner, M.M., Allee, L.L., Brown, P.M.J., Losey, J.E., Roy, H.E., Smyth, R.R., 2012. Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. Front. Ecol. Environ. 10, 471–476.
Harris, R.W., Clark, J.R., Matheny, N.P., 2004. Arboriculture: Integrated Management of Landscape Trees, Shrubs, and Vines, 4th ed. Prentice Hall, Upper Saddle River, NJ, 592 pp.
Holt, L., 2010. Stikky Trees: Learn to Recognize at a Glance the 15 Most Common Trees in the United States –in Just One Hour, Guaranteed. Laurence Holt Books, 256 pp.
Jim, C.Y., Liu, H.T., 2001. Species diversity of three major urban forest types in Guangzhou City, China. For. Ecol. Manage. 146, 99–114.
Kocher, W., 2012. New software for tree inventory and urban forestry outreach. Arborist News 21, 50–51.
Kremen, C., Ullmann, K.S., Thorp, R.W., 2011. Evaluating the quality of citizen-science data on pollinator communities. Conserv. Biol. 25, 607–617.
Kumar, N., Belhumeur, P.N., Biswas, A., Jacobs, D.W., Kress, W.J., Lopez, I.C., Soares, J.V.B., 2012. Leafsnap: a computer vision system for automatic plant species identification. Proceedings of the 12th European Conference on Computer Vision ECCV 2012 vol. 7573, 502–516.
Lewandowski, E., Specht, H., 2015. Influence of volunteer and project characteristics on data quality of biological surveys. Conserv. Biol. 29, 713–723.
Lindenmayer, D.B., Likens, G.E., 2010. The science and application of ecological monitoring. Biol. Conserv. 143, 1317–1328.
McPherson, E.G., Kotow, L., 2013. A municipal forest report card: results for California USA. Urban For. Urban Greening 12, 134–143.
McPherson, G., Simpson, J.R., Peper, P.J., Maco, S.E., Xiao, Q., 2005. Municipal forest benefits and costs in five US cities. J. For., 411–416.
McPherson, E.G., 1993. Monitoring urban forest health. Environ. Monit. Assess. 26, 165–174.
Molin, J.F., Konijnendijk van den Bosch, C.C., 2014. Between big ideas and daily realities –the roles and perspectives of Danish municipal green space managers on public involvement in green space maintenance. Urban For. Urban Greening 13, 553–561.
Nowak D.J., Hoehn III R.E., Bodine A.R., Crane D.E., Dwyer J.F., Bonnewell V., and Watson G. 2013. Urban trees and forests of the Chicago region. Res. Bull. NRS-84. Newtown Sq., PA: USDA Forest Service, Northern Research Station. 106 pp.
Pandya, R.E., 2012. A framework for engaging diverse communities in citizen science in the US. Front. Ecol. Environ. 10, 314–317.
Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11, 1633–1644.
Roman, L.A., McPherson, E.G., Scharenbroch, B.C., Bartens, J., 2013. Identifying common practices and challenges for local urban tree monitoring programs across the United States. Arboriculture Urban For. 39, 292–299.
Roman, L.A., Battles, J.J., McBride, J.R., 2014. The balance of planting and mortality in a street tree population. Urban Ecosyst. 17, 387–404.

Roman, L.A., Walker, L.A., Martineau, C.M., Muffly, D.J., MacQueen, S.A., Harris, W., 2015. Stewardship matters: case studies in establishment success of urban trees. Urban For. Urban Greening 14, 1174–1182.

Romolini, M., Brinkley, W., and Wolf, K.L., 2012. What is urban environmental stewardship? Constructing a practitioner-driven framework. USDA Forest Service Research Note PNW-RN-566.

Ryan, R.L., 2015. The transformative power of engaging in urban greening projects. Cities Environ. 8 (1), article 6.

Scharenbroch, B.C., Catania, M., 2012. Soil quality attributes as indicators of urban tree performance. Arboriculture Urban For. 38, 214–228.

Scharenbroch, B.C., Roman, L.A., McPherson, E.G., Bartens, J., Boyer, D., 2014. Taking the urban forest's pulse: working group focused on tree growth and longevity. Arborist News 23, 54–55.

Silva, P., Krasny, M.E., 2014. Parsing participation: models of engagement for outcomes monitoring in urban stewardship. Local Environ.: Int. J. Justice Sust. 21, 157–165.

Silva, P., Barry, E., Plitt, S., 2013. TreeKIT: measuring, mapping, and collaboratively managing urban forests. Cities Environ. 6 (1), article 3.

Sjöman, H., Östberg, J., Bühler, O., 2012. Diversity and distribution of the urban tree population in ten major Nordic cities. Urban For. Urban Greening 11, 31–39.

St. John, N., 2011. Street Tree Inventories Being Conducted in Five Portland Neighborhoods. Neighborhood Notes, Jun. 29, 2011, Available at: www.neighborhoodnotes.com/news/2011/06/street_tree_inventories_being_conducted_in_five_portland_neighborhoods/.

Starr J., Schweik C.M., Bush N., Fletcher L., Finn J., Fish J., and Bargeron C.T. 2014. Lights, camera... citizen science: assessing the effectiveness of smartphone-based video training in invasive plant identification. PLoS ONE 9(11): e111433.

Statistics Sweden, 2010. Statistical database, 2010, accessed 6/22/2016 by J.P.A. Östberg. Available at: www.scb.se.

Tulloch, A.I.T., Possingham, H.P., Joseph, L.N., Szabo, J., Martin, T.G., 2013. Realising the full potential of citizen science monitoring programs. Biol. Conserv. 165, 128–138.

US Census, 2010. American FactFinder, Census 2010, accessed 6/22/2016 by L.A. Roman. Available at: factfinder.census.gov.

US Forest Service, 2014. Urban Forest Inventory: Monitoring the Trees Where People Live. USDA Forest Service Northern Research Station, Available at: http://www.nrs.fs.fed.us/fia/urban.

US Forest Service, 2016. Forest Inventory and Analysis National Urban Plot Field Guide: Field Data Collection Procedures for Urban FIA Plots, Version 7.0. USDA Forest Service, Northern Research Station, Available at: http://www.nrs.fs.fed.us/fia/urban.

Vogt, J.M., Fischer, B.C., 2014. A protocol for citizen science monitoring of recently-planted urban trees. Cities Environ. 7 (2), article 4.

Westphal, L.M., 2003. Urban greening and social benefits: a study of empowerment outcomes. J. Arboric. 29, 137–147.

Wiggins, A., Newman, G., Stevenson, R.D., Crowston, K., 2011. Mechanisms for data quality and validation in citizen science. Proceedings of the 2011 IEEE Seventh International Conference on e-Science Workshops, 14–19.

van Doorn, N.S., 2014. Best practices and quantified error rates for long-term tagged-tree inventories: examples from a temperate forest. In: Patterns and Processes of Forest Growth: The Role of Neighborhood Dynamics and Tree Demography in a Northern Hardwood Forest PhD Dissertation. University of California, Berkeley, Berkeley, CA, Chapter 3.