Forest Ecology and Management 316 (2014) 21–31

Contents lists available at ScienceDirect

Forest Ecology and Management

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

Data and database standards for permanent forest plots in a global network

Richard Condit [a,*], Suzanne Lao [a], Anudeep Singh [a], Shameema Esufali [a], Steven Dolins [b]

[a] *Center for Tropical Forest Science, Smithsonian Tropical Research Institute, Unit 0948, APO AA 34002, United States*
[b] *Department of Computer Science and Information Systems, Bradley University, Peoria, IL 61625, United States*

ARTICLE INFO

Article history:
Available online 13 October 2013

Keywords:
Forest plot database
Integrity errors
Normalized data
Tree census

ABSTRACT

The Center for Tropical Forest Science established a network of 50-ha forest inventory plots in the 1980s, and now assists local scientists with field and database methods at 44 large-scale plots across boreal, temperate, and tropical forest biomes. We published detailed field methods over a decade ago, but at that time, data were maintained in spreadsheet-like formats, most harboring design flaws that resulted in frequent errors. We since established detailed database methods and a normalized data model for housing multiple censuses of large plots. Our largest databases include >2 million measurements, and each has a master version on a server where all collaborators can access and edit data. This paper focuses on the data requirements for handling tree census data and how to design databases to meet these requirements and to ensure data integrity. There are six key elements of a tree census which the database must reflect: (1) measurements, including individual trees (genetic units), stems within trees, and multiple measurements of stems; both field and data methods must assure that every tree, stem, and measurement is precisely identified and can be relocated easily; (2) coordinates, including quadrats within a plot, because field mapping is usually done by assigning x–y coordinates relative to local quadrat markers; (3) taxonomy, carrying a species identity for every tree with a history of individual re-identifications; (4) personnel, with records of the people who performed field and data work per quadrat; (5) assessment of field error via random re-measurements; and (6) a log of changes and a system of archiving so that errors can be tracked and past versions can be reconstructed and cited in publications. A well-designed database model reduces a variety of integrity errors and improves access to data tables in identical formats across many plots, allowing data analyses to be easily replicated and results to be compared. The principal disadvantage is that complexity of the database requires experienced data managers.

Published by Elsevier B.V.

1. Introduction

Trees are perfect subjects for population biology. They are large, easy to find, and do not move so can be relocated in future censuses. Whereas population studies of mammals, birds, and insects usually involve inference from subsamples and must account for failure to detect individuals, a *tree plot* is a complete count of all living and dead individuals over a predefined area (Table 1). The ease with which tree plots are censused coupled with the rigor they lend to population studies have fostered a proliferation of plot studies in every forest biome (Ayyappan and Parthasarathy, 1999; Phillips et al., 2003; Canham et al., 2006; Coomes et al., 2009; ter Steege et al., 2006). Some have been in place for decades (Crow, 1980; Whitney, 1984; Franklin and DeBell, 1988; Condit et al., 2012; Lilleleht et al., 2014), and tree plots are the basis of national forest inventories (Tomppo et al., 2010; Álvarez-González et al., 2014).

The greatest asset of a tree census comes from the recensus, when individual trees are relocated and remeasured, leading to a rigorous and repeatable count of the population size of all species present and measurements of individual growth and death rates. These measurements and remeasurements (Table 2) are the focus of data collection in the field and data maintenance in a modern database.

The measurements collected on a single tree are brief and straightforward, but most plots in the Center for Tropical Forest Science (hereafter CTFS) network, and in many national inventories, include hundreds of thousands of records, far more than can be proofread by eye. Prior to the year 2000, CTFS data were housed in spreadsheet-like tables, and we became familiar with a variety of oft-repeated errors caused by poor design. Particularly frustrating were integrity errors resulting from mismatched records, many of which occurred after data were transcribed into the computer (Table 3). Some of these errors are magnified from census to census

---

* Corresponding author. Tel.: +507 212 8045.
E-mail address: conditr@gmail.com (R. Condit).

0378-1127/$ - see front matter Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.foreco.2013.09.011

Table 1
Definitions of trees and tree plots.

Term	Definition
Tree	Woody plant, free-standing (i.e. excluding lianas), including all parts of one genetic unit
Stem	A single woody axis of a tree (i.e., one tree can have one or more stems)
Tree plot	Every individual tree stem in a predefined region, above a minimum size, is included 1. Coordinates estimated 2. Stem diameter measured 3. Species identity determined
Other vegetation surveys	Not qualifying as tree plots 1. Enumerations of individuals within plots, but no coordinates 2. Percent cover measured in a plot instead of enumeration of individuals 3. A subset of individuals or species enumerated, instead of every one

because species names and remeasurements are connected to the wrong trees.

We relate here our efforts to apply database theories and standard practices to tree plot data (Codd, 1971; Elmasri and Navathe, 2000; Date, 2004). The database model we designed to store Barro Colorado plot data, and subsequently data from other CTFS plots, follows basic theories of normalization and eliminates many common errors (Codd, 1971). Database normalization refers to a series of design rules for data tables whose intent is to minimize data redundancy, because redundant (i.e. repeated) data are prone to anomalies during updates. An anomaly is created when a repeated datum is changed in one place but not another; omitting repetition thus prevents anomalies.

This system, referred to here as the CTFS Data Model, has been designed and implemented in a database management system, MySQL, with interfaces written in HTML and PHP for online access (Hubbell et al., 2010). These interfaces are outside the scope of this presentation, as is any discussion of field methods, which Condit (1998) cover in detail. Our goal here is to describe the data standards adhered to in the CTFS plot network and enough details of the CTFS Data Model to explain how it upholds the standards.

2. Components of a tree census database

2.1. Tree Measurements and Re-measurements

The principal data from a tree census are repeated measurements of stems. This is straightforward at its most basic, with diameter of individual stems measured at a consistent position through time. The biggest single hindrance arises from multiple stems, where a single tree (a genetic unit) has several separate stems at breast height. Indeed, if trees always had single trunks, and never forked or sprouted from the base, the most frustrating difficulties with plot remeasurements would vanish. Yet multiple-stemmed trunks cannot be ignored: there are many tree species that routinely grow clonally, with many separate stems in a genetic unit. Examples include several palm species in the American tropics, Myrtaceae in Valdivian forests of Chile, and a variety of coppicing north temperate *Quercus*.

Assuring consistent measurements requires precise identification of individual stems and measurement positions, and this leads to three hierarchies in the measurement data: (1) Single trees have multiple stems. (2) Each stem has multiple measurements through time. (3) Each measurement may have associated with it several attributes, including breaks, death, swellings at the measurement point, etc.

2.1.1. Multiple stems

Reliable and permanent identification of multiple stems within a tree is crucial for ensuring that repeat measurements can be correctly linked. A routine method for tree plots should thus be to attach individually-numbered tags to every individual stem, just as traditional forest plots assume that every tree should be tagged. The preferred tagging system includes a principal sequence of numbers for individual trees and a subordinate sequence for stems; the two types of tag differ in form so are readily distinguished. So trees are numbered 1, 2, 3, … through many thousands, and stems within trees have letters ‘A’, ‘B’, ‘C’, etc. In this system, there are both tree tags and stem tags. There is a commonly used alternative in which all stems get a single sequence and one tag type. The disadvantage to the latter method is that there is no tag defining the tree, and we consider the former the CTFS standard. In the former method, stems are readily associated with the tree they belong with, because each is identified by both a tree tag and a stem tag.

In the preferred method, a tree with a single trunk is not given a stem tag in the field, since it is not needed, but it is assigned stem tag ‘A’ in the database. In later censuses, if a trunk sprouts a second

Table 2
Standard stem measurements in CTFS tree plots. Alternatives indicate widely used variations on each measurement.

Measure	CTFS standard	Alternative
Height of diameter measure (*HOM*)	‘Breast height’, or 1.3 m	Some sites use 1.37 m as standard. *HOM* must be above buttresses. When buttresses grow upward, additional *HOM* needed
Minimum size limit	1 cm	10 cm
Death	Criteria are seldom specified but usually include 1. Leafless state in non-deciduous species 2. Trunk broken 3. Rotten wood encircling trunk 4. Trunk entirely fallen or vanished	
Species	Standard botanical taxonomy	Morphospecies (i.e., unidentified), Subspecies
Local coordinates	*X*, *Y* distance to a precisely surveyed grid post	Polar coordinates to a grid post
Stem diameter	Small stems: Largest axis (with calipers) Large stems: Circumference (with dbh tape)	Dbh tape for all sizes