**Recommended FIA filter criteria for BIEN**

Brad Boyle

8 Feb. 2013

**Introduction**

This document provides a list of criteria and suggest SQL statements that can be used to filter FIA plots (USDA Forest Service 2013) to produce a subset representing only natural vegetation on public land, sampled using a single protocol. The FIA tables and columns names should be up to date with respect to the schema documented in the November 2012 version of the FIA database user's manual (O'Connell et al. 2012). The filter criteria were assembled in consultation with Jes Coyle, Bob Peet, Margaret Evans and Brian McGill. I particularly thank Jes for sharing her R script used to compile FIA forest plots for the eastern USA.

**Specific goals of the filter**

1. Construct unique ID for a plot and for an individual census of a plot
2. Identify repeat censuses of the same plot, filtering out all but the most recent census.
3. Filter out all non-standard plot methodologies, including only plots that conform to the standard FIA four subplot sampling protocol for trees >= 1 in (2.5 cm) dbh (see Appendix 1 for details).
4. Filter out plots representing non-natural vegetation (plantations, logging treatments, etc.).
5. Remove plots with excessively fuzzed or swapped coordinates. Although all FIA plots are fuzzed to 0.5 miles, some plots on private land also have coordinates swapped between adjacent plots, and may have greater fuzzing. Therefore all plots on private land should be excluded.

**Filtering steps**

I illustrate these steps using MySQL-flavored SQL. Of course, the same criteria can be adapted to other languages and platforms. I illustrate the main queries only. If working in SQL, be sure to index all columns involved in joins or where clauses, otherwise these queries will be extremely slow with such a large dataset.

Two caveats. One, I still do not know if and how FIA differentiates between stems and individuals. In other words, we have yet to determine how an individuals with two stem measurements (i.e., branches below breast height) is distinguished from two individual trees with single stems bearing the same measurements. Second, with the exception of the Pacific Northwest-specific codes STUMP\_CD\_PNWRS and STND\_COND\_CD\_PNWRS, my examples do not allow NULL values. As NULL is the commonest value within many FIA fields, the queries below may exclude more records than necessary. However, in the absence of detailed information about the meaning of NULLs, I thought it best to be more rather than less stringent.

**1. Download state-specific tables**

FIA database tables are separated by state, and can be downloaded as csv files from <http://apps.fs.fed.us/fiadb-downloads/datamart.html>. The naming convention is [state\_code]\_[table\_name]; e.g., AL\_PLOT. For our purposes, the only plots we need are those with suffixes \_COND, \_PLOT, \_SUBPLOT, \_TREE. Table TREE gives species and measurements of all trees >=2.5 cm dbh. Because of uncertainty regarding methodology, I exclude the seedling stratum, which resides in table SEEDLING

**2. Create unique plot code (plotCode) and plots census code (plotCensusCode) while combining state tables into the four tables COND, PLOT, SUBPLOT, TREE**

As the default codes in FIA are only guaranteed to be unique within counties, a globally unique ID must be constructed by concatenating state, county and plot codes. A unique census is identified by adding the inventory year.

plotCode = CONCAT\_WS("\_",STATECD,COUNTYCD,PLOT)

plotCensusCode = CONCAT\_WS("\_",STATECD,COUNTYCD,PLOT,INVYR)

The above operation should be performed for each of the four tables.

**3. Remove all but the most recent census of each plot in each table**

One approach is to mark the most recent census of each plot and delete the remainder, as follows (example using table COND):

ALTER TABLE COND

ADD COLUMN deleteme INT(1) DEFAULT 1 NOT NULL,

ADD INDEX(deleteme);

UPDATE COND p JOIN

(
SELECT plotCode, MAX(INVYR) as maxINVYR

FROM COND

GROP BY plotCode

) AS p2

ON p.plotCode=p2.plotCode AND p.INVYR=p2.maxINVYR

SET deleteme=0;

DELETE FROM COND WHERE deleteme=1;

Plots in the remaining tables can be marked and unwanted plot censuses deleted by joining to table COND by plotCensusCode. Or extract a "black list" prior to deletion from the first table and delete unwanted plots directly in the remaining table by joining to the black list.

**4. Prepare a white list of good plots representing natural vegetation, standard methodology, on public land**

This can be done based on the following criteria:

Table 1. Criteria used to filter FIA plots to include only natural vegetation on public land, sampled using using four subplots and no macroplots (see Appendix 1).

|  |  |  |  |
| --- | --- | --- | --- |
| **Table** | **Condition** | **Purpose** | **Comments** |
| COND | STDORGCD==0 | Include plots without artificial regeneration | Exclude STDORGCD==1, -> planted |
| COND | DSTRBCD1<>80 AND DSTRBCD2<>80 AND DSTRBCD3<>80 | Exclude plots with evidence of human disturbance | Exclude DSTRBCD 80, possibly 30-32, and Include TRTCD=0 (check meaning of NULLs) |
| COND | TRTCD1==0 AND TRTCD2==0 AND TRTCD3==0 | Include only plot without observable silvicultural treatments | TRTCD=0 means no silvlcultural treatment. |
| COND | STDAGE>=80 | Include only plots 80 years old or older | Using an older stand age would exclude most eastern US forests plots |
| COND | (STUMP\_CD\_PNWRS="N" OR STUMP\_CD\_PNWRS IS NULL) | Include only plots with no evidence of stumps or cutting | Applies to PNW plots only. The NULL condition avoids excluding plots to which this criterion does not apply |
| COND | (STND\_COND\_CD\_PNWRS=7 OR STND\_COND\_CD\_PNWRS IS NULL) | Include only plots flagged as "Old growth" | Applies to PNW plots only. The NULL condition avoids excluding plots to which this criterion does not apply |
| COND | OWNGPCD<>40 | Public land only | OWNGPCD==40 is private land; these plots have coordinates fuzzed |
| PLOT | DESIGNCD IN (1,115,311,312,313,314) | Include only plots with national design | Limits to single, standard design of 4 subplots |
| PLOT | SAMP\_METHOD\_CD==1 | Include only plots that were visited in the field | Excludes remotely-censused plots |
| PLOT | PLOT\_STATUS\_CD==1 | Include only sampled plots | Possibly same as preceding criterion, perhaps stricter |
| PLOT | QA\_STATUS<> 5 | Exclude botched plot files | QA\_STATUS==5 -> botched! |
| PLOT | MACRO\_BREAKPOINT\_DIA IS NULL | Exclude plots that were sampled using a macroplot | missing value means plot was sampled with regular 24 ft radius subplot only. These are the ones we want. |
| TREE | SUBP>=4 | Remove bad subplot numbers | Each plot should consist of 4 subplots, number 1-4. Ideally, filter should ensure that the plot has exactly four subplot numbered 1 to 4. Neither more nor less, and with no other numbering system. |

Using SQL, the above filter criteria can be applied in three steps as follows.

**4.a. Prepare white list of plots containing exactly 4 subplots numbered 1 to 4**

These are the standard design plots we want to keep. The goal is to remove plots which have greater or fewer than this number of subplots, or have subplot numbers other than 1-4. One approach to this is shown below.

CREATE TABLE goodplots1

SELECT DISTINCT s1.plotCode

FROM

(SELECT DISTINCT plotCode FROM TREE WHERE SUBP=1) AS s1

JOIN

(SELECT DISTINCT plotCode FROM TREE WHERE SUBP=2) AS s2

JOIN

(SELECT DISTINCT plotCode FROM TREE WHERE SUBP=3) AS s3

JOIN

(SELECT DISTINCT plotCode FROM TREE WHERE SUBP=4) AS s4

ON s1.plotCode=s2.plotCode AND s2.plotCode=s3.plotCode AND s3.plotCode=s4.plotCode;

-- the following assumes SUBP is an INTEGER data type; change "(1,2,3,4);" TO

-- "('1','2','3','4')" if SUBP is string/text/varchar

CREATE TABLE badplots

SELECT DISTINCT plotCode

FROM TREE

WHERE SUBP NOT IN (1,2,3,4);

DELETE FROM g1

USING goodplots1 g1 JOIN badplots b

ON g1.plotCode=b.plotCode;

**4.b. Prepare a second white list of plots based on the COND and PLOT tables**

CREATE TABLE goodplots2

SELECT DISTINCT plotCode

FROM COND c JOIN PLOT p

ON c.plotCode=p.plotCode

WHERE

c.STDORGCD==0 AND

(c.DSTRBCD1<>80 AND c.DSTRBCD2<>80 AND c.DSTRBCD3<>80) AND

(c.TRTCD1==0 AND c.TRTCD2==0 AND c.TRTCD3==0) AND

c.STDAGE>=80 AND

(c.STUMP\_CD\_PNWRS="N" OR c.STUMP\_CD\_PNWRS IS NULL) AND

(c.STND\_COND\_CD\_PNWRS=7 OR c.STND\_COND\_CD\_PNWRS IS NULL) AND

c.OWNGPCD<>40 AND

p.DESIGNCD IN (1,115,311,312,313,314) AND

p.SAMP\_METHOD\_CD==1 AND

p.PLOT\_STATUS\_CD==1 AND

p.QA\_STATUS<> 5 AND

p.MACRO\_BREAKPOINT\_DIA IS NULL;

**4.c. Prune tables by joining to the two white lists:**

UPDATE COND

SET deletme=1;

UPDATE COND c JOIN goodplots1 g1 JOIN goodplots2 g2

ON c.plotCode=g1.plotCode and g1.plotCode=g2.plotCode

SET deleteme=0;

DELETE FROM COND WHERE deleteme=1;

The remaining tables are pruned in a similar manner. Or extract a black list prior to deleting, and delete unwant plots directly from the remaining tables, joining on plotCode.

**5. Remove individual records of dead trees from TREES table**

DELETE FROM TREE

WHERE STATUSCD==1;

**References**

USDA Forest Service. 2013. Forest inventory and Analysis Program. <http://www.fia.fs.fed.us/>

O'Connell, BM, LaPoint, EB, et al. 2012. The Forest Inventory and Analysis Database: Database Description and Users Manual Version 5.1.4 for Phase 2. Nov. 2012. <http://www.fia.fs.fed.us/library/database-documentation/current/ver5.1.4/FIADB_user%20manual_5-1-4_p2_11_2012.pdf>

**Appendix 1. Standard FIA forest plot protocol**

Most FIA forest inventory plots use 4 circular subplots designed to subsample roughly 1 acre (Fig. A1.1). Many variations to this sample design exist. BIEN filter criteria exclude plots with macroplots, with more or less than four subplots, or with subplot numbering other than 1 though 4.

Note that the total area sampled (in four subplots) for trees ≥ 5 in dbh is 7238 ft2, or 672.4 m2 (0.067 ha). For trees ≥ 1 in dbh, total area sampled (in four microplots) is 581 ft2, or 54.0 m2 (0.0054 ha).



Fig. A1.1. Standard four-subplot ayout of a FIA forest inventory plot. Subplots are numbered as shown. Trees ≥ 5 in dbh (12.7 cm) are measured in four subplots of 24.0 ft (7.315 m) radius, numbered as shown. In some inventories, these measurements are extended to a radius of 58.9 ft (17.953 m) to increase sampling of large trees. Within each subplot, trees ≥ 1 in (2.5 cm) are sampled within a circular microplot of 6.8 ft radius (2.073 m).