

Exploring the CA FIA dataset

Jacob Bukoski

5/27/2020

Notes

In: The Forest Inventory and Analysis Database: Database Description and User Guide for Phase 2, Version 7.0.1 (2018)

Index of Tables (filtered to those available through rFIA) - page 569

- **COND** - information on the discrete combination of landscape attributes that define the condition (condition will have same land class, reserved status, owner group, forest type, stand size class, regeneration status, and stand density)
- **COND_DWM_CALC** - condition down woody material calculation table. this table contains calculations used to create estimations on the down woody material indicator.
- **INVASIVE_SUBPOT_SPP** - invasive subplot species table. provides percent cover data of invasive species identified on the subplot.
- **PLOT** - information relevant to the entire 1-acre field plot. links to most other tables via the linkage "CN" (i.e., "sequence number")
- **POP_ESTN_UNIT** - an estimation unit is a geographic area that can be drawn on a map. It has known area and is field-sampled at the same intensity. Generally estimation units are contiguous areas, but exception are made when certain ownerships, usually national forests, are sampled at different intensities. One record in the POP_ESTN_UNIT table corresponds to a single estimation unit.
- **POP_EVAL** - an evaluation is the combination of a set of plots (the sample) and a set of phase 1 (remote sensing) data (a stratification) that can be used to produce population estimates for a State (an evaluation may be created to produce population estimates for a region other than a State - such as the Black Hills National Forest). A record in the POP_EVAL table identifies one evaluation and provides some descriptive information about how the evaluation may be used.
- **POP_EVAL_GRP** - information on the suite of evaluations that were used to generate a complete set of reports for an inventory. In a typical State inventory report, one evaluation is used to generate an estimate of the total land area; a second evaluation is used to generate current estimates of volume, number of trees and biomass; and a third evaluation is used for estimation growth, removals and mortality. One record in the POP_EVAL_GRP record identifies all the evaluations that were used in generating estimates for a State inventory report.
- **POP_EVAL_TYP** - population evaluation type table. Provides information on the type of evaluations that were used to generate a set of tables for an inventory report.
- **POP_PLOT_STRATUM_ASSGN** - stratum information is assigned to a plot by overlaying the plot's location on the phase 1 imagery. Plots are linked to their appropriate stratum for an evaluation via the POP_PLOT_STRATUM_ASSGN table.
- **POP_STRATUM** - area within an estimation unit is divided into strata. The area for each stratum can be calculated by determining the proportion of phase 1 pixels/plots in each stratum and multiplying that proportion by the total area in the estimation unit. Information for a single stratum is stored in a single record of the POP_STRATUM table.
- **SEEDLING** - count of number of live trees of a species found on a microplot that are less than 1 inch in diameter but at least 6 inches in length for conifer species or at least 12 inches in length for hardwood species.
- **SUBP_COND** - information about the proportion of a subplot in a condition
- **SUBP_COND_CHNG_MTRX** - subplot condition change matrix table. Contains information about the mix of current and previous conditions that occupy the same area on the subplot.

- **SUBPLOT** - describes features of a single subplot. Multiple subplots per 1-acre field plot and can be multiple conditions sampled on each subplot
 - **SURVEY** - a record for each yr an inventory is conducted in a State for annual inventory or one record for each periodic inventory
 - **TREE** - information for each tree 1 inch in diameter and larger found on a microplot, subplot, or core-optional microplot
 - **TREE_GRM_BEGIN** - Tree net growth, removal, and mortality begin table. Contains information for remeasured trees where values have been calculated for the beginning of the remeasurement period. Only those trees where information was reculated for time 1 (T1) are included. The information in this table is used to produce net growth, removal, and mortality estimates on remeasured trees.
 - **TREE_GRM_COMPONENT** - Tree net growth, removal, and mortality component table. Stores information used to compute net growth, removals, and mortality estimates for remeasurement trees. Each remeasurement tree has a single record in this table.
 - **TREE_GRM_MIDPT** - Tree net growth, removal, and mortality midpoint table. Contains information about a remeasured tree at the midpoint of the remeasurement period. It does not contain a record for every tree. Midpoint estimates are computed for trees that experience mortality, removal, or land use diversion or reversion. The information in this table is used to compute net growth, removal, and mortality estimates on remeasurement trees.
-

Data exploration with CA FIA dataset

Load FIA library

```
library(rFIA)

## Registered S3 methods overwritten by 'ggplot2':
##   method      from
##   [.quosures   rlang
##   c.quosures   rlang
##   print.quosures rlang
```

```
library(tidyverse)
```

```
## Registered S3 method overwritten by 'rvest':
##   method      from
##   read_xml.response xml2
```

```
## -- Attaching packages -----
## v ggplot2 3.1.1    v purrr  0.3.2
## v tibble  3.0.1    v dplyr  0.8.5
## v tidyr   1.1.0    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.4.0
```

```
## -- Conflicts -----
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

Download all FIA data for California, uncomment to download data

```
# dat <- getFIA( states = "CA", dir = './data/')
```

After downloading data and saving to file, can simply load data

```
dat <- readFIA("../data/")
```

View the names of the tables within the object

```
sort(names(dat))

## [1] "COND" "COND_DWM_CALC"
## [3] "INVASIVE_SUBPLOT_SPP" "PLOT"
## [5] "POP_ESTN_UNIT" "POP_EVAL"
## [7] "POP_EVAL_GRP" "POP_EVAL_TYP"
## [9] "POP_PLOT_STRATUM_ASSGN" "POP_STRATUM"
## [11] "SEEDLING" "SUBP_COND"
## [13] "SUBP_COND_CHNG_MTRX" "SUBPLOT"
## [15] "SURVEY" "TREE"
## [17] "TREE_GRM_BEGIN" "TREE_GRM_COMPONENT"
## [19] "TREE_GRM_MIDPT"
```

Extract some basic information

```
plotDat <- dat$PLOT

nPlts <- length(unique(plotDat$PLOT))
invyrs <- sort(unique(plotDat$INVYR))
```

Number of plots in CA: 17971

Years of inventories: 1994, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018

Count number of plots with repeat measurements

```
filt <- dat$COND %>%
  filter(INVYR != 1994) %>%
  group_by(PLOT) %>%
  mutate(nyrs = length(unique(INVYR))) %>%
  ungroup() %>%
  filter(nyrs == 2)
```

```
plts_w_repeats <- nrow(filt)
```

Note, plot IDs < 50000 are from inventory year 1994. Drop them from the analysis.

There are 32774 plots with repeat measurements.

Look at how many plots by different ownership groups

10 = Forest Service 20 = Other federal 30 = State and local government 40 = Private and native american

```
filt %>%
  group_by(OWNCD) %>%
  summarize(n = n()) %>%
  ungroup() %>%
  mutate(percent = round(n / sum(n) * 100, 1)) %>%
  select(OWNCD, n, percent)
```

```
## # A tibble: 11 x 3
##   OWNCD      n percent
##   <int> <int>   <dbl>
## 1     11  7292    22.2
## 2     12     1     0
## 3     21  1980     6
## 4     22  4195   12.8
## 5     23    32    0.1
## 6     24   167    0.5
## 7     25  1099    3.4
## 8     31  2255    6.9
## 9     32   504    1.5
## 10    33    13     0
## 11    46 15236   46.5
```

OWNCD - Ownership codes:

```
$ 11 = National Forest
$ 12 = National Grasslands and/or Prairie
$ 13 = Other Forest Service Land
$ 21 = National Park Service
$ 22 = Bureau of Land Management
$ 23 = Fish and Wildlife Service
$ 24 = Departments of Defense/Energy
$ 25 = Other Federal
$ 31 = State, including State public universities
$ 32 = Local (County, Municipality, etc.) including water authorities
$ 33 = Other non-federal public
$ 46 = Undifferentiated private and Native American
```

(Additional ownership codes that need to be requested)

```
$ 41 = Corporate, including native corporations in Alaska and private universities
$ 42 = Non-governmental conservation/natural resources organization
$ 43 = Unincorporated local partnership/association/club
$ 44 = Native American
$ 45 = Individual and family, including trusts, estates, and family partnerships
```

```
filt %>%
  group_by(OWNGRPCD) %>%
  summarize(n = n()) %>%
  ungroup() %>%
  mutate(percent = round(n / sum(n) * 100, 1)) %>%
  select(OWNGRPCD, n, percent)
```

```
## # A tibble: 4 x 3
##   OWNGRPCD      n percent
##   <int> <int>   <dbl>
## 1      10  7293    22.3
## 2      20  7473    22.8
## 3      30  2772     8.5
## 4      40 15236   46.5
```

OWNGRPCD - Owner Group Code:

\$ 10 = Forest Service (OWNCD 11, 12, 13)

\$ 20 = Other federal (OWNCD 21, 22, 23, 24, 25)

\$ 30 = State and local government (OWNCD 31, 32, 33)

\$ 40 = Private and native american (OWNCD 41, 42, 43, 44, 45, 46)

FORINDCD - Private owner industrial status code

Code indicating whether landowner owns and operates a primary wood-processing plant. A primary wood-processing plant is any commercial operation that originates the primary processing of wood on a regular and continuing basis (e.g., pulp or paper mill, sawmill, panel board mill, post, or pole mill). **(Need to request this data)**

\$ 0 = Land is not owned by industrial owner with wood-processing plant

\$ 1 = Land is owned by industrial owner with wood-processing plant