
Challenges of Working With FIADB17 Data: The SOLE Experience

Michael Spinney¹ and Paul Van Deusen²

Abstract.—The Southern On Line Estimator (SOLE) is an Internet-based Forest Inventory and Analysis (FIA) data analysis tool. SOLE is based on data downloaded from the publicly available FIA database (FIADB) and summarized by plot condition. The tasks of downloading, processing, and summarizing FIADB data require specialized expertise in inventory theory and data manipulation. The FIADB is an important FIA product that should be made as easy to use and as error free as possible. Some of the errors that have been found in the FIADB are outlined here along with the incremental steps made toward improving it over the past year.

Introduction

The Southern On Line Estimator (SOLE) (<http://ncasi.uml.edu/SOLE/>) is an Internet-based database analysis tool developed cooperatively by the National Council for Air and Stream Improvement (NCASI) and the U.S. Department of Agriculture Forest Service Southern Research Station. SOLE performs tabular, chart, linear model, and map analyses on annual Forest Inventory and Analysis (FIA) data made publicly available through the FIA database (FIADB).

SOLE's data is based on publicly available annual FIA data downloaded from the FIADB data download facility and summarized by plot condition. This largely automated process requires a comprehensive program that must be modified to accommodate changes in the FIADB. This paper illustrates some of the challenges that NCASI has faced in using the FIADB by giving several examples of major events that have impacted the utility of the FIADB.

FIADB Revisions

The FIADB has undergone several revisions. The East/Westwide databases were consolidated into FIADB Version 1.0 in 2001 (Miles *et al.* 2001) to produce a consistent structure and repository for FIA data. FIADB1.0 was revised to FIADB1.7 in 2004 (Alerich *et al.* 2004), which substantially changed the database structure. Each of the nine tables carried over from FIADB1.0 doubled in size with new variables describing biological conditions/properties (plot location, forest health data, regional variables, etc.) and database administration (estimation methods, data modification dates, etc.).

The upgrade to Version 1.7 was disjointed because data were released before the supporting documentation. Substantial changes in the database caused users to struggle to understand new variables and tables. The most notable change was the redefinition of condition proportion (CONDPROP, which describes the proportion of a plot defined by a particular suite of site characteristics) to be adjusted over stratum. All values in the CONDPROP field were missing, and close inspection of the condition table revealed five new variants of CONDPROP. An inquiry to the Northeastern regional FIA office confirmed that one of these variants was a close approximation to the old definition of CONDPROP. This approximation was incorporated into SOLE. Documentation explaining this change in FIADB1.7 was released later that month.

In July 2005 FIA upgraded FIADB to Version 2.1 to use the National Information Management System to collect, compile, summarize, and distribute the data. The 2.1 revision was introduced as “development” files (available alongside Version 1.7 data) with a draft user manual. Implementing this newest FIADB data into SOLE has not shown any substantial difference in the database structure or variable definitions. Version 2.1 has since moved into production, yet the user manual and data file names are still labeled as “draft.”

¹ Quantitative Forestry Analyst, National Council for Air and Stream Improvement, Inc., 600 Suffolk Street, Fifth Floor, Lowell, MA 01854. E-mail: mspinney@ncasi.org.

² Principal Research Scientist, National Council for Air and Stream Improvement, Inc., 600 Suffolk Street, Fifth Floor, Lowell, MA 01854.

The FIADB Data Dump

FIADB data is delivered via the FIADB data dump, which is the only public access point to access FIADB data. Users navigate to this Web page and select a data file for download. Recent revisions to the FIADB data download facilities have improved on some, but not all, aspects of data delivery. Because the data dump is the singular public gateway to FIA data, complete and accurate documentation is critical.

In 2005, data files began receiving a time stamp when they are uploaded to the FIADB. The intent of the time stamp is for users to determine whether they have the most recent data. Unfortunately, this time stamp is not always accurate. For example, the AL_04 compressed file has a timestamp of April 2005. Downloading and decompressing the file reveals that the individual files were actually modified in November 2005. FIA should ensure that the time stamp is accurate.

Major changes to the FIADB data are not always documented. Georgia's annual data disappeared from the FIADB for several months in 2004. The data file reappeared without explanation. Concurrently, California and Oregon's data had internal formatting errors that prevented proper file import. The Pacific Northwest station confirmed the error, then applied a correction over the next few months. The errors were never noted on the Web site. In December 2005, all annual data from 1998 to 2002 for Indiana and Missouri were removed without explanation. Errors must be carefully and completely explained to enable users to assess the integrity of past analyses.

The latest incarnation of the FIADB data dump, "FIA Data Mart," improves on the former data download facility. State selection has been condensed from multiple pages of text to a single drop down list. Multiple data formats can be selected for each State and a direct link leads to the user guide for each format. Unfortunately, notation of survey type (periodic/annual) has been removed. This information is available only by viewing the summary statistics link on the MapMaker homepage.

NCASI's FIADB Assessment

NCASI created a FIADB Assessment to characterize the following properties of annual FIADB data by State:

1. Plot coordinate type.
2. Number of plots by measurement year.
3. Proportion of plots containing key variables.

Changes in these characteristics are tracked over time. The assessment reports are updated on a quarterly basis and posted at <http://ncasi.uml.edu/SOLE/>.

Summary

FIA data is essential to monitor the forests of the United States. A tremendous amount of effort goes into collecting, compiling, and delivering data to the public through the FIADB. The public accesses FIADB data through a single portal, thus it is critical that the portal be error free and easy to use. The essential information a user needs to perform an analysis includes survey type, FIADB manual, an accurate time stamp, and explanation of past data errors. All of this information should be accessible from one location.

Literature Cited

- Alerich, C.L.; Klevgard, L.; Liff, C.; Miles, P.D. 2004. The Forest Inventory and Analysis database: database description and users guide. Ver. 1.7. http://www.ncrs2.fs.fed.us/4801/FIADB/fiadb_documentation/FIADB_DOCUMENTATION.htm.
- Miles, P.D.; Brand, G.J.; Alerich, C.L.; Bednar, L.F.; Woudenberg, S.W.; Glover, J.F.; Edwards, E.N. 2001. The Forest Inventory and Analysis database: database description and users manual. Ver. 1.0. GTR-NC-218. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 130 p.