On-Line Analysis of Southern FIA Data

Michael P. Spinney, Quantitative Forestry Analyst, and Paul C. Van Deusen, Principal Research Scientist, National Council for Air and Stream Improvement, Lowell, MA
Francis A. Roesch, Mathematical Statistician, USDA Forest Service, Southern Research Station, Asheville, NC

Abstract *—The Southern On-Line Estimator (SOLE) is a web-based FIA database analysis tool designed with an emphasis on modularity. The Java-based user interface is simple and intuitive to use and the R-based analysis engine is fast and stable. Each component of the program (data retrieval, statistical analysis and output) can be individually modified to accommodate major additions or changes. This modular structure fosters a flexible architecture to encourage outside development of analysis algorithms. SOLE produces statistical tables, figures, maps, and PDF reports based on user selected area and variables.*

Introduction

The Forest Inventory and Analysis (FIA) program of the USDA Forest Service has been collecting forest inventory data since 1930. Database structure and inventory design have evolved over time to satisfy user demand. The most recent revision of the FIA database (FIADB) has created new ecological variables and changed the definition of others. The sheer size of the FIADB can be intimidating to some users, and the database structure can confuse others. Readily accessible, easy to use tools are necessary to ensure proper analysis of FIA data.

FIA data is periodically analyzed by the Forest Service and published for the public as a standard suite of tables for a state or region. With the adoption of the annual inventory system, published reports are produced on a 5 year period. The length of this period may not be adequate for users who require more timely reports or are interested in a spatial scale different than the state or region level. These users must download data, replicate the relational database then correctly query and summarize the data. Not all users with these data needs are willing to invest this amount of effort.

Internet based FIA analysis tools bridge the gap between periodic state level reports and customized analysis by the user. Without understanding underlying database structure, users can tailor their analysis by choosing a certain area and selecting specific variables. Internet-based tools correctly analyze the data regardless of the current state of the database. FIA's MapMaker is the Forest Service's official web-based analysis tool which utilizes standardized FIA analysis algorithms. Other online tools are available.

The Southern On Line Estimator1 (SOLE http://ncasi.uml.edu/SOLE) has been cooperatively developed by the Forest Service and NCASI. It is a web-based annual FIA data analysis tool that uses a JAVA-based user interface and an R-based analysis engine. The user selects an area and variables of interest, and analysis output comes in the form of tables, charts, maps, and PDF files.

Background

SOLE Structure Overview

SOLE's structure has been developed with flexibility as the top priority. Each of the three main components of the program (data retrieval, statistical analysis, and output) have been developed to be independently modified. This open architecture facilitates outside development of analysis algorithms.

The user can specify a geographical region of interest and initiate a query by sequentially selecting tabs at the top of the interface. After choosing states in the State Selection window, the user progresses through the SOLE Map, Variable Selection, Filters (optional), Analysis and/or Mapping tabs. Once the user satisfies the requirements of each tab, SOLE makes the next tab available. Detailed instructions can be found in the help files linked to each tab.
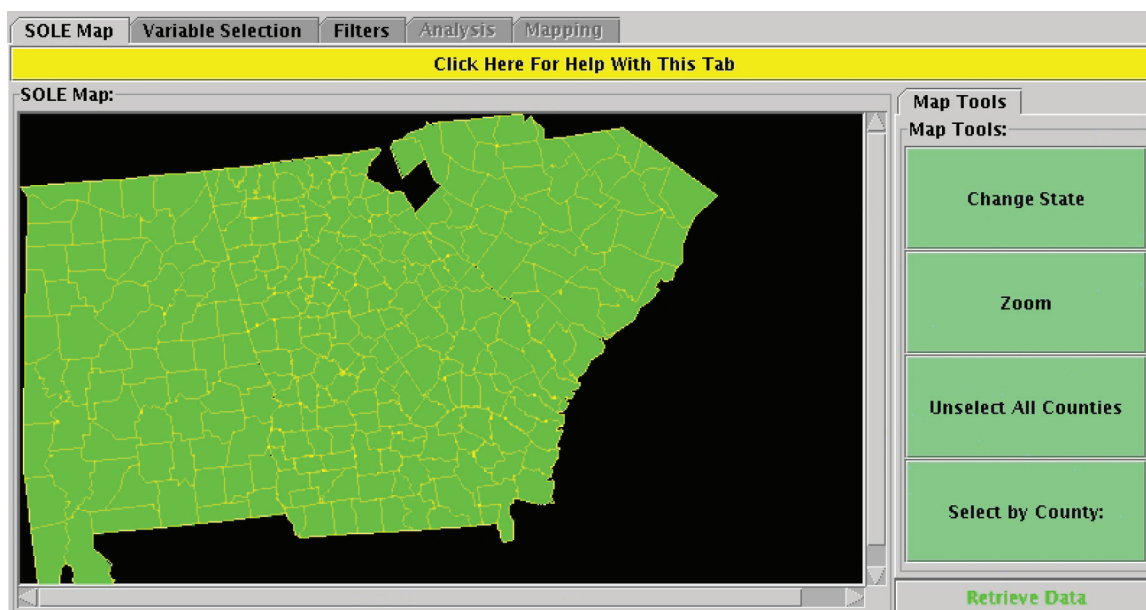
Figure 1. SOLE map tab, for county selection and data retrieval.

Using the Program

State Selection

The first window a user sees is the State Selection screen. States with annual data have black backgrounds, and states with periodic data (currently not available for analysis in SOLE) are gray. When a state is selected, the state name pops up and the state is highlighted green. When the user is satisfied with their state choices, they click the bottom of window and are brought to the SOLE Map tab.

SOLE Map

Here the user selects the counties they wish to work with (fig. 1). Individual counties can be chosen with the Select by County button, or select all counties in the state using the Select All Counties button. Clicking on the red Retrieve Data button loads the necessary data files into the statistical program, R (R Development Core Team 2003). Once the data files have been loaded, the next group of tabs, Variable Selection and Filters are made available.

Variable Selection

Currently, only FIADB volume variables are available for analysis. These are grouped in two ways. The first is Quantitative (continuous data, for example, any volume/biomass estimate) or Qualitative (categorical data, such as productivity class). In addition, Quantitative variables are grouped based on diameter greater or less than 5 inches. This preserves the precision of volume estimates by ensuring that no trees less than 5 inches d.b.h. are included in sawtimber estimates (fig. 2). For all analyses, one Quantitative and one Qualitative variable are required. Tabular analyses will examine one Quantitative and several Qualitative variables.

Filters

This tab is optional, but can be useful to select only a specific subset of data. Data availability varies by state, so we recommend determining the range of values for an analyses before setting filters. The status of a filter is displayed at the top of an analysis.

Analysis

Charts and tabular analyses are launched from this tab (fig. 3). The header of each analysis indicates which filters (if any) were used, the analysis and variable names.

Tabular analysis

There are currently two methods in SOLE for determining the mean: moving average and mixed estimators. The 5-year moving average (MA) is the default estimation method for FIA analysis; it provides a complete estimate for that 5-year period. A moving average analysis will result in the following seven tables:

1. Moving average (per acre), over measurement year or by cycle.
2. Standard error of each element of table 1.
3. Area (acres) represented by the plots used in the calculation of table 1. Summarizes EXPCURR.
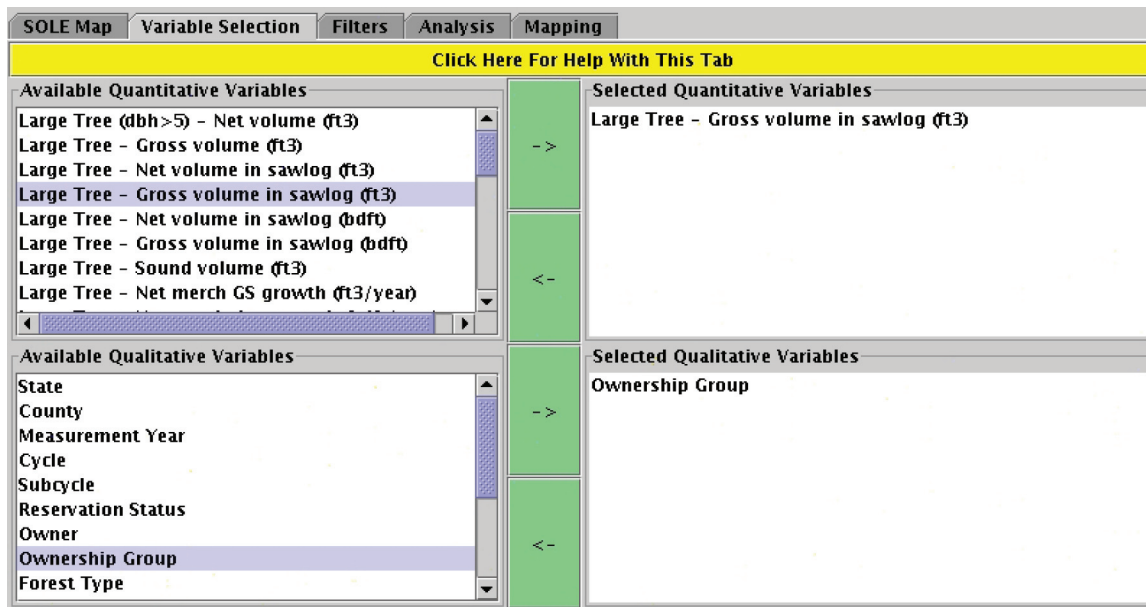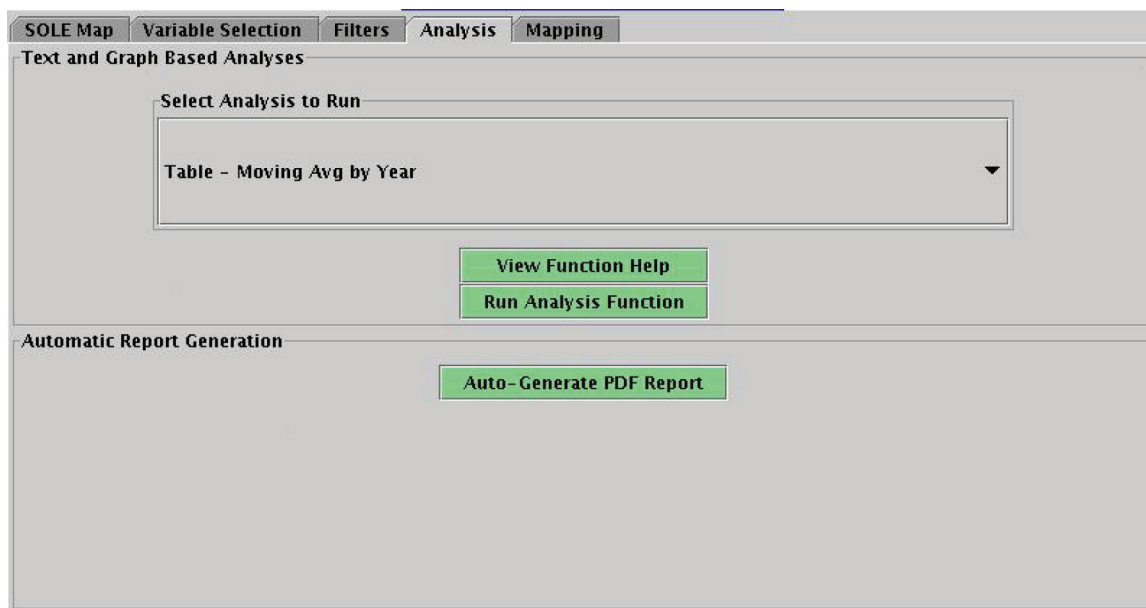
Figure 2. SOLE variable selection tab.

Figure 3. SOLE analysis selection tab.

4. Total estimate for the area sampled. Equals moving average multiplied by area.
5. Standard error of the total estimate.
6. Total number of plots used in calculation, broken down by each level of the Qualitative Variable.
7. Number of plots by measurement year.

The MA provides a complete estimate for each specific 5-year period. Using the MA as the standard estimate, change can be determined after a second MA can be calculated, in year 6. In this case the data from panel 1, measured in year 1, is dropped from the estimator, and the data from panel 1, measured in year 6, is added. Eighty percent of the data is common to the two consecutive MA estimates. Therefore, if the user wanted to compare estimates for two consecutive years, only 20 percent of the data would not be self-canceling (Van Deusen 2000). After 10 years, both the MA and mixed estimator (ME) can provide full-data change estimates, however, it is not obvious to which years the MA difference estimator should be applied. Conversely, the alternative method of mixed estimation incorporates an explicit trend model while using 100 percent of plots to determine change between any 2 years in the most efficient manner.

The mixed estimator analysis results in six tables that have measurement year as columns:

1. Mixed estimator mean by year (per acre)
2. Standard error per year
3. Acres
4. Total estimate for the sampled area
5. Standard error or the total
6. Adjusted sample size by year

Graphical analysis

Often the simplest way to visualize data characteristics is through a graph-based analysis. SOLE offers bar charts, box plots, pie charts, and XY plots. Bar and pie charts show a basic distribution of the Quantitative Variable by each level of the Qualitative Variable. XY plots offer some insight into the data distribution and frequency based on the number and dispersion of points along the axis. Box plots convey the most statistical information about the data because they show the mean, interquartile range, and outliers for each level of Qualitative Variable.

Map analysis

County-level maps can be created from the Mapping tab. Basic maps of the mean and median are supplemented by more complex ratio maps, which display the ratio of filtered to unfiltered data. For example, the user could filter for a specific ownership group and then view a map of the proportion of volume on land owned by that group to the volume on all land collectively owned by all groups.

PDF report

Any combination of tabular, graphical or map analyses; or text can be included in the PDF report. At the simplest level, the report could be completely contained such that the user only selects an area of interest, or it could be customized with the user's choice of variables. Potentially, this feature could be developed to produce an array of standard FIA report tables at the click of a button, giving the user the ability to produce real-time reports at each FIADB17 update.

Conclusions

Web based FIA analysis tools are essential for proper analysis of FIA data. SOLE provides a simple interface to allow users of all levels to obtain custom analytical results. The SOLE graphical user interface (GUI) is written in Java, and the analytical components are written in the S programming language as implemented in the OpenSource R statistical analysis program. SOLE is written in a modular fashion that makes it easy to add new capabilities. SOLE provides a wide range of options for users who wish to analyze FIA annual inventory data and obtain graphical and tabular results. Flexibility in each component of SOLE ensures that it remains highly adaptable to changes in both database structure and user needs.

References

R Development Core Team (2003). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL http://www.R-project.org.

Van Deusen, P.C. 2000. Alternatives to the Moving Average. In: Reams, Gregory A.; McRoberts, Ronald E.; Van Deusen, Paul C., eds. 2001. Proceedings of the second annual Forest Inventory and Analysis symposium; 2000 October 17-18; Salt Lake City, UT. Gen. Tech. Rep. SRS-47. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. pp. 90-93.