Using SAS® to Navigate the USDA Forest Service's Forest Inventory and Analysis Database

Matthew Russell and Aaron Weiskittel, School of Forest Resources, University of Maine, Orono, ME

Philip Radtke, Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg,VA

ABSTRACT

The Forest Inventory and Analysis (FIA) program, administered through the USDA Forest Service, serves as the Nation's forest census. Collected over long time periods and wide geographic areas with consistent data management protocols, FIA data have been used in a variety of settings to help inform our knowledge of the Nation's forests. To use the voluminous FIA database, software that employs sound database management software along with flexible statistical modeling procedures would be advantageous to use for a database of this size and scope. This paper describes how SAS is being used to implement and put the FIA data to use in a research setting.

INTRODUCTION

The Forest Inventory and Analysis (FIA) program, administered through the USDA Forest Service, serves as the forest census for the United States. The FIA program measures variables such as changes in forest types, changes in forest ownerships, forest health, tree size, tree species composition, and total volume of harvested wood. The FIA program has existed in the US for over 80 years, which has provided a consistent, but voluminous, data structure. All of this information is used to assess and evaluate the present value of forests across the Nation, as well as to forecast the future value of forests and the services they provide.

Given the expansiveness and breadth of the FIA program, USDA manages The FIA database (FIADB) to serve as a clearinghouse for prospective users. Given the size and consistency of data collection protocols across large geographic areas, SAS has proven to be a wonderful tool for extracting data of interest, importing data from various sources in a variety of formats, and analyzing data and building models to inform forest scientists and forest managers on tree growth characteristics in the future (Figure 1).

Figure 1. General usage of SAS to analyze the Forest Inventory and Analysis database.

RETRIEVING DATA

The FIADB version 4.0 manages data primarily on a state-by-state basis with a consistent format (US Department of Agriculture, Forest Service. 2011). Since FIA data are continually updated, the SAS ability to read in the most recent data is essential. Here, reading .csv files directly into SAS using the url access method within the filename statement ensures that the most recent data are used. The FIA website does maintain Microsoft Access 2007 database files that contain empty, predefined tables that a user can readily import data into, however, scripts to read in FIA data into SAS are not found on the FIA site.

FIA archives multiple datafiles, each containing a variety of information ranging from data on the:

(1) sample plot from which the data were obtained (variables such as year of inventory, latitude/longitude of plot location),

(2) overstory trees found in that plot (variables such as tree diameter and species), and

(3) tree seedlings that are found in subplots smaller in size.

As an example, the TREE table for the state of Maine can be imported directly into SAS, using the input statement to define variable names as described in the FIA User's Manual (US Department of Agriculture, Forest Service. 2010) :

```
*FIA data can be obtained from http://199.128.173.17/fiadb4-downloads/datamart.html;

filename MEtree url 'http://199.128.173.17/fiadb4-downloads/ME_TREE.csv';

data MEtree; infile MEtree lrecl=6000 firstobs=2 missover dlm=",";
informat CN $varying15. PLT_CN $varying15. PREV_TRE_CN $varying15.;
input CN PLT_CN PREV_TRE_CN INVYR STATECD UNITCD COUNTYCD PLOT SUBP TREE CONDID
AZIMUTH DIST PREVCOND STATUSCD SPCD SPGRPCD DIA DIAHTCD HT HTCD ACTUALHT TREECLCD CR
CCLCD TREEGRCD AGENTCD CULL DAMLOC1 DAMTYP1 DAMSEV1 DAMLOC2 DAMTYP2 DAMSEV2 DECAYCD
STOCKING WDLDSTEM VOLCFNET VOLCFGRS VOLCSNET VOLCSGRS VOLBFNET VOLBFGRS VOLCFSND
GROWCFGS GROWBFSL GROWCFAL MORTCFGS MORTBFSL MORTCFAL REMVCFGS REMVBFSL REMVCFAL
DIACHECK MORTYR SALVCD UNCRCD CPOSCD CLIGHTCD CVIGORCD CDENCD CDIEBKCD TRANSCD
TREEHISTCD DIACALC BHAGE TOTAGE CULLDEAD CULLFORM CULLMSTOP CULLBF CULLCF BFSND CFSND
SAWHT BOLEHT FORMCL HTCALC HRDWD_CLUMP_CD SITREE CREATED_BY CREATED_DATE $10.
CREATED_IN_INSTANCE MODIFIED_BY MODIFIED_DATE $10. MODIFIED_IN_INSTANCE MORTCD HTDMP
ROUGHCULL MIST_CL_CD CULL_FLD RECONCILECD PREVDIA FGROWCFGS FGROWBFSL FGROWCFAL
FMORTCFGS FMORTBFSL FMORTCFAL FREMVCFGS FREMVBFSL FREMVCFAL P2A_GRM_FLG $1.
TREECLCD_NERS TREECLCD_SRS TREECLCD_NCRS TREECLCD_RMRS STANDING_DEAD_CD PREV_STATUS_CD
PREV_WDLDSTEM TPA_UNADJ TPAMORT_UNADJ TPAREMV_UNADJ TPAGROW_UNADJ DRYBIO_BOLE
DRYBIO_TOP DRYBIO_STUMP DRYBIO_SAPLING DRYBIO_WDLD_SPP DRYBIO_BG CARBON_AG CARBON_BG
CYCLE SUBCYCLE BORED_CD_PNWRS DAMLOC1_PNWRS DAMLOC2_PNWRS DIACHECK_PNWRS
DMG_AGENT1_CD_PNWRS DMG_AGENT2_CD_PNWRS DMG_AGENT3_CD_PNWRS MIST_CL_CD_PNWRS
SEVERITY1_CD_PNWRS SEVERITY1A_CD_PNWRS SEVERITY1B_CD_PNWRS SEVERITY2_CD_PNWRS
SEVERITY2A_CD_PNWRS SEVERITY2B_CD_PNWRS SEVERITY3_CD_PNWRS UNKNOWN_DAMTYP1_PNWRS
UNKNOWN_DAMTYP2_PNWRS PREV_PNTN_SRS;
run;
```

```
NOTE: 463450 records were read from the infile METREE.
      The minimum record length was 256.
      The maximum record length was 256.
      One or more lines were truncated.
NOTE: The data set WORK.METREE has 463450 observations and 142 variables.
NOTE: DATA statement used (Total process time):
      real time           25.19 seconds
      cpu time            9.06 seconds
```

MANIPULATING/EXTRACTING DATA

After data are read into SAS, the next step is to query the database to extract only the needed data. As an example, we may only be interested in analyzing all red spruce trees that are growing across the state of Maine. The User's Manual tells us that all red spruce are coded as a species code 97 (SPCD = 97).

Furthermore, we may only be interested in analyzing those trees which contain measurements of the total tree's height (variable HT). As tree heights are difficult to measure in the forest and it is a time-consuming measurement, HT is subsampled in the FIA data. Additionally, we may only be interested in analyzing those trees which are alive (dead trees are also monitored as part of the program). Hence, trees with height measurements with STATUSCD= 1 (tree that is alive) are extracted using a datastep:

```
data redspruce; set MEtree;
if SPCD=97;
run;

data redspruceHT; set redspruce;
if HT le 0 then delete;
if STATUSCD ne 1 then delete;
RUN;
```

MODEL BUILDING

Using the extracted data, we might be interested in using the data to construct a statistical model to predict tree height based on a tree's diameter at breast height (the diameter of the tree measured at 4.5 feet above the ground). This is especially important because of the difficulty and time involved in measuring tree height. Because of its flexibility, one statistical model commonly used in modeling tree growth is the Chapman-Richards function (Richards 1959). For estimating tree height (HT) using tree diameter (DIA), this model takes the form:

$$HT = 1.3 + a\left(1 - e^{-b*DIA}\right)^{c}$$

where a, b, and c are model parameters to be estimated. To estimate the parameters of this nonlinear growth equation, the MODEL procedure has served well in modeling the dynamic relationships between tree size and growth. Here, over 21,000 tree observations can be used for estimating the height of red spruce trees if tree diameter is known:

```
proc model data=redspruceHT;
var HT DIA;
parms a b c;
HT = 1.3+a*(1-exp(b*DIA))**c;
fit HT start = a 30 b -0.05 c 1.5;
title 'Estimating red spruce tree height with diameter';
run; quit;
```

```
NOTE: At OLS Iteration 28 CONVERGE=0.001 Criteria Met.

NOTE: PROCEDURE MODEL used (Total process time):
      real time           4.21 seconds
      cpu time            4.11 seconds
```

After running the procedure, results are obtained that are biologically meaningful (i.e. the magnitude and size of the parameter estimates s are appropriate) and fit statistics show strong fits to the data (as shown by low mean squared error and favorable R^2).

Generally, we have found PROC MODEL to be superior in obtaining parameter estimates when compared to other statistical modeling packages, such as those of R. For this specific model with over 20,000 observations, total processing time for obtaining parameters in PROC MODEL takes a little over 4 seconds, whereas the nonlinear least squares (**nls**) and generalized nonlinear least squares (**gnls**) functions in R takes minutes to obtain parameters. This is especially important as we're fitting individual models to a range of different species over large geographic areas, for which the number of observations can be quite large.

Figure 2. Sample output from MODEL procedure for estimating tree height using tree diameter.

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors

Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
HT	3	21300	1250508	58.7093	7.6622	0.6663	0.6663

Nonlinear OLS Parameter Estimates

Parameter	Estimate	Approx Std Err	t Value	Approx Pr > |t|
a	82.94581	1.2220	67.87	<.0001
b	-0.10757	0.00432	-24.90	<.0001
c	1.127071	0.0251	44.94	<.0001

Number of Observations		Statistics for System	
Used	21303	Objective	58.7010
Missing	0	Objective*N	1250508

CONCLUSIONS

SAS has proven to be an excellent tool in retrieving, manipulating, summarizing, and building statistical models for the Forest Inventory and Analysis database. Its ability to rapidly fit complex biological models to the large amount of data associated with the FIADB proves superior to other statistical packages.

REFERENCES

Richards, F.J. 1959. “A flexible growth function for empirical use.” Journal of Experimental Biology 10: 290-300.

US Department of Agriculture, Forest Service. 2010. The Forest Inventory and Analysis Database: Database Description and User’s Guide Version 4.0 for Phase 2 (Draft Revision 3). Washington DC. 368 pp.

US Department of Agriculture, Forest Service. 2011. http://apps.fs.fed.us/fiadb-downloads/datamart.html

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Matthew Russell
School of Forest Resources, University of Maine
Email: matthew.russell@umit.maine.edu

Philip Radtke
Department of Forest Resources and Environmental Conservation, Virginia Tech
Email: pradtke@vt.edu

Aaron Weiskittel
School of Forest Resources, University of Maine
Email: aaron.weiskittel@maine.edu