PREPARATION OF FOREST INVENTORY AND ANALYSIS (FIA) AND STATE SOIL GEOGRAPHIC DATA BASE (STATSGO) DATA FOR GLOBAL CHANGE RESEARCH IN THE EASTERN UNITED STATES

Louis R. Iverson, Anantha M.G. Prasad, and Charles T. Scott[1]

Abstract: The USDA Forest Service's Forest Inventory and Analysis (FIA) and the Natural Resource Conservation Service's State Soil Geographic (STATSGO) data bases provide valuable natural resource data that can be analyzed at the national scale. When coupled with other data (e.g., climate), these data bases can provide insights into factors associated with current and future ranges of tree species. However, a significant amount of data distillation is needed prior to such analyses. This paper describes the data base and geographic information system (GIS) processing involved with preparing the data for global change research in the eastern United States.

INTRODUCTION

To better understand the potential impacts of climate change on tree-species distributions (i.e., migration potentials), one must first understand the factors associated with current tree ranges. Then, projections of future ranges can be made assuming there are no barriers to migration. Finally, more realistic projections of migration can be modeled with proper attention to habitat and biological restrictions to migration. This paper reports on initial efforts in this overall project.

For purposes of coarse resolution analysis at the national scale, especially in the eastern United States, a county level of resolution seems appropriate. There are 3,048 counties across 37 states east of the 100th meridian. This level of detail allows several advantages over finer (e.g., individual plots or soil series) or coarser (e.g., ecoregions) levels of detail: (1) the data set is manageable in size and computing power, yet of sufficient detail for adequate sample sizes, (2) data are more readily available, as many agencies report data at the county level, and (3) precise spatial co-location of forest inventory plots and ancillary information is not necessary with a county level of resolution. However, any averaging of forest or soil information to the county will lose information, especially in counties with highly heterogeneous habitats. Nonetheless, given the advantages of such a level of analysis, a procedure to recalculate data to the county reporting unit was needed. This paper describes the procedures we used for two national-level data bases: the USDA Forest Service's Forest Inventory and Analysis (FIA) and the Natural Resources Conservation Service's State Soil Geographic data bases.

For both data sets, we used Arc/InfoAML and UNIX shell programming (especially 'awk' and 'sed') running on a workstation. We developed a series of macro programs that did the desired operations on one state's data at a time. Nearly two gigabytes of data storage were needed to store and process the information for U.S. land east of the 100th meridian.

FOREST INVENTORY AND ANALYSIS DATA BASE

The USDA Forest Service has a mandate to periodically determine the extent, condition, and volume of timber, growth, and removals of the nation's forest land. The six Forest Service FIA units conduct periodic regional surveys. Four FIA units produced a data base of standard format called the Eastwide Data Base (EWDB) for the 37 states from North Dakota to Texas and east. These data are stored in three record types as described in their user's guide (Hansen et al. 1992): county data, plot data, and tree data. These 500 megabytes of 'raw' data were summarized into the desired county-level information needed for our global change research.

---

[1]USDA Forest Service, Northeastern Forest Experiment Station, 359 Main Road, Delaware, OH 43015.

Extraction of individual tree-species information

The first step summarized the information for individual forested plots. Tree records represented observations of seedlings, saplings, and overstory trees. Tree species, tree status, and diameter at breast height were combined with information on plot size to compute a single summary record for each plot. The record contained the following information for each species that occurred in the state: number of understory trees/acre, understory basal area/acre, number of overstory trees/acre, and overstory basal area/acre. This information provided the next step in the FIA data summary.

Importance value calculation by species

Once the number of stems and basal area were available for each species (overstory vs. understory) for each plot, the relative importance of each species could be evaluated. Several importance values were calculated depending on the choice of variables considered (i.e., overstory vs. understory, basal area vs. number of stems). The overall importance value (IV) can be described as

$$IV(x) = \frac{100\,BA(x)}{BA(all species)} + \frac{100\,NS(x)}{NS(all species)}$$

where x is a particular species on a plot, BA is basal area, and NS is number of stems for both overstory and understory trees. On monotypic stands, the IV would reach the maximum of 200.

Mapping by plot

Importance values for any species could then be mapped by plot assuming that specific locations of the plots were available. Plot locations are truncated to 100 seconds (roughly 1000 m at this latitude) in the EWDB to protect the precise location of the long-term remeasurement plots. Under special arrangement with the FIA units, we obtained plot locations for Ohio, Kentucky, Illinois, and Indiana, our region of special interest for more detailed analysis. A method was devised by which importance values were divided into cartelist, with different symbols used for each cartel. Therefore, a quick visual inspection of the maps could determine where the species was found and how extensive it was relative to the other trees in the plots.

Aggregation and mapping at county level

To aggregate plot-level information to the county level, no GIS processing was necessary until final mapping since each plot has a county code associated with it. Average importance values were calculated for all forested plots, by species and by county. These values were associated with a county coverage of the United States (SRI 1992) for mapping into density slices of IV (Fig. 1 a). With these maps, biogeographical characteristics (such as absolute and optimum range) of the species can be visualized. Current research involves statistical analysis of the range relative to other variables and a better characterization of eastern forest tree species.

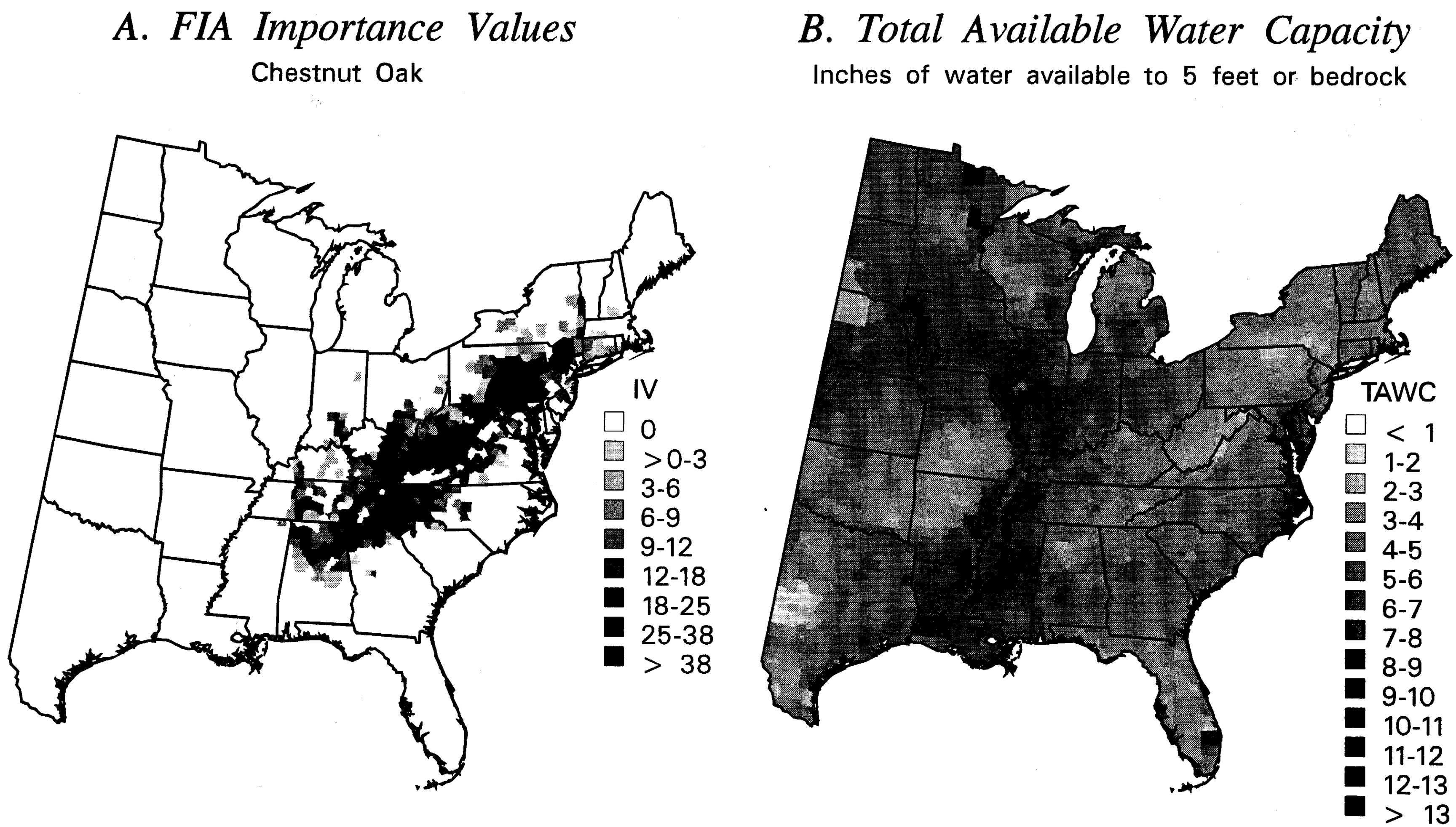

Figure 1. A. Importance values for chestnut oak, and B. Total water capacity for the eastern United States.

STATE SOIL GEOGRAPHIC DATA BASE

The State Soil Geographic Data Base (STATSGO) was developed by the USDA Soil Conservation Service (now Natural Resource Conservation Service) to help achieve their mandate to collect, store, maintain, and distribute soil-survey information for U.S. lands. STATSGO data recently became available for the entire nation on CD; it contains physical and chemical soil properties for about 18,000 soil series recognized in the nation (Soil Conservation Service 1991). STATSGO maps were compiled by generalizing more detailed soil-survey maps into soil associations in a scale (1: C) more appropriate for regional analysis. Detailed, digital soil-survey maps are available only for scattered portions of the country and requirements for data storage would be enormous. Therefore, STATSGO data currently are one of the best sources of information on the edifice landscape pattern and structure for the nation. However, the nature of the data dictates a sizable amount of preprocessing is necessary before maps of particular attributes can be produced on a national scale.

Attribute selection

STATSGO's user's guide (Soil Conservation Service, 1991) details the data structure and the myriad of files and variables contained within. For purposes of global change research, we selected 14 variables related to tree species habitat: pH, available water capacity, organic matter, permeability, bulk density, salinity, cation exchange capacity, depth to bedrock, T factor, K factor, slope, and several variables related to texture (e.g., percent clay, percent fragments > 3 inches, percent volume of soil flowing through screens with meshes of various sizes).

Weighted averaging by depth, soil-series composition, and county

Except for depth to bedrock, T factor, K factor, slope, and organic matter, each variable had estimates in STATSGO by individual layer in the horizon. A weighted average, based on the thickness of each layer to a depth of 60 inches (152 cm) or the depth to bedrock, was calculated for each soil series. A second weighted average was calculated for the horizontal dimension based on the percent composition of each soil series within each soil association. Since mapping units were associations, maps of any of the extracted variables could then be made at the association level. Finally, a third weighted average was calculated to estimate attributes by county. The proportion of each soil association within each county (an intersection of county coverage with STATSGO associations) was multiplied by the attribute weighted average for each association and summed for the county. Again, the county weighted averages can be mapped for any variable, as exemplified by available water capacity for the eastern United States (Fig. 1b).

SUMMARY

The two data bases described here provide valuable resources for natural resource evaluation and environmental assessment, including modeling impacts of potential climate change, at a regional level. Both data sets are only recently available in standard format for such a large portion of the country. Because of limitations in spatial accuracy, computer power, and data storage, a county-level analysis seems appropriate for revealing relationships between these data sets and others such as climate, elevation, or land use. At this scale, one can look a level up to understand context and a level down to understand process. Thus, the effort described here was necessary to get the data in a form useful for such analyses. Table 1 summarizes some information contained within the data bases, by state.

Table 1. Summary of selected FIA and STATSGO data by state. Column abbreviations include number of forested and total plots in the FIA data base, percent forest (from Powell et al. 1993), number of tree species recorded in FIA data, number of soil associations in STATSGO data, number of counties, and the date of the forest inventory in the FIA data base.

State	No. forest plts	No. plts	Forest, %	No. spp	No. soils	No. cnty	Date, fia
Alabama	4013	4515	67.7	117	251	67	1990
Arkansas	3158	3786	53.6	105	68	75	1988
Connecticut	286	463	58.7	61	31	8	1985
Delaware	149	250	31.1	59	19	3	1986
Florida	5377	12441	47.9	68	159	67	1987
Georgia	7152	12015	65.1	84	112	160	1988
Illinois	1132	10957	12.0	86	83	102	1985
Indiana	2146	11440	19.3	94	93	92	1986
Iowa	699	12767	5.7	63	83	99	1990
Kentucky	1995	3049	50.0	107	195	120	1988
Louisiana	2473	2893	49.7	99	336	64	1991
Maine	2161	2483	88.8	62	69	16	1982
Maryland	691	1199	42.9	98	63	23	1986
Massachusetts	379	555	63.9	71	61	14	1985
Michigan	7931	14958	50.2	81	190	83	1980
Minnesota	12260	43955	32.8	60	321	87	1990
Mississippi	3003	3618	56.6	104	215	82	1987
Missouri	3861	17270	31.8	90	106	115	1989
New Hampshire	590	697	86.8	58	45	10	1983
New Jersey	259	644	42.3	77	39	21	1987
New York	2501	4313	61.9	86	168	62	1980
North Carolina	5696	9993	61.8	90	75	100	1990
Ohio	1762	4845	30.0	104	166	88	1991
Pennsylvania	3128	5298	59.2	103	103	67	1989
Rhode Island	116	179	59.9	46	18	5	1985
South Carolina	4383	7031	63.6	85	158	46	1993
Tennessee	2350	2951	51.6	113	214	95	1989
Vermont	625	823	76.7	61	76	14	1983
Virginia	4285	7312	62.6	87	76	95	1991
West Virginia	2592	3209	78.7	109	122	55	1989
Wisconsin	6939	15908	44.6	67	128	72	1983

Future efforts entail building and enhancing data bases on elevation, climate and land use for the eastern United States, and conducting statistical analysis (e.g., regression, decision-tree, and correspondence analysis) designed to better understand the current distribution of eastern trees. Then, predictions of range changes can be made (under ideal conditions of seed source availability and habitat availability) following a climate change. Finally, models under development will help determine how well species can migrate under the real situation of fragmented habitats.

ACKNOWLEDGMENTS

This study is funded by the Northern Global Change Program of the USDA Forest Service. We thank them and the Forest Inventory and Analysis research unit of the Northeastern Forest Experiment Station for providing the FIA data. Thanks also to Rachel Hershey, Robert Brooks, and Wayne Zippered for earlier reviews of the manuscript.

LITERATURE CITED

Environmental Systems Research Institute. 1992. Arcus M, User's guide and data reference. Environmental Systems Research Institute, Redlands, CA.

Hansen M H, T. Frieswyk, J. F. Glover, and J. F. Kelly. 1992. The Eastwide forest inventory data base: user's manual. General Technical Report NC-151, USDA Forest Service, North Central Forest Experiment Station, St. Paul, MN.

Powell D S, J. L. Faulkner, D. R. Darr, Z. Zhu, and D. W. MacCleery. 1993. Forest resources of the United States, 1992. General Technical Report RM-234, Rocky Mountain Forest and Range Experiment Station. Ft. Collins, CO.

Soil Conservation Service. 1991. State soil geographic data base (STATSGO) data users guide. Miscellaneous Publication 1492, USDA Soil Conservation Service, Washington, D.C.