United States
Department of
Agriculture

Forest
Service

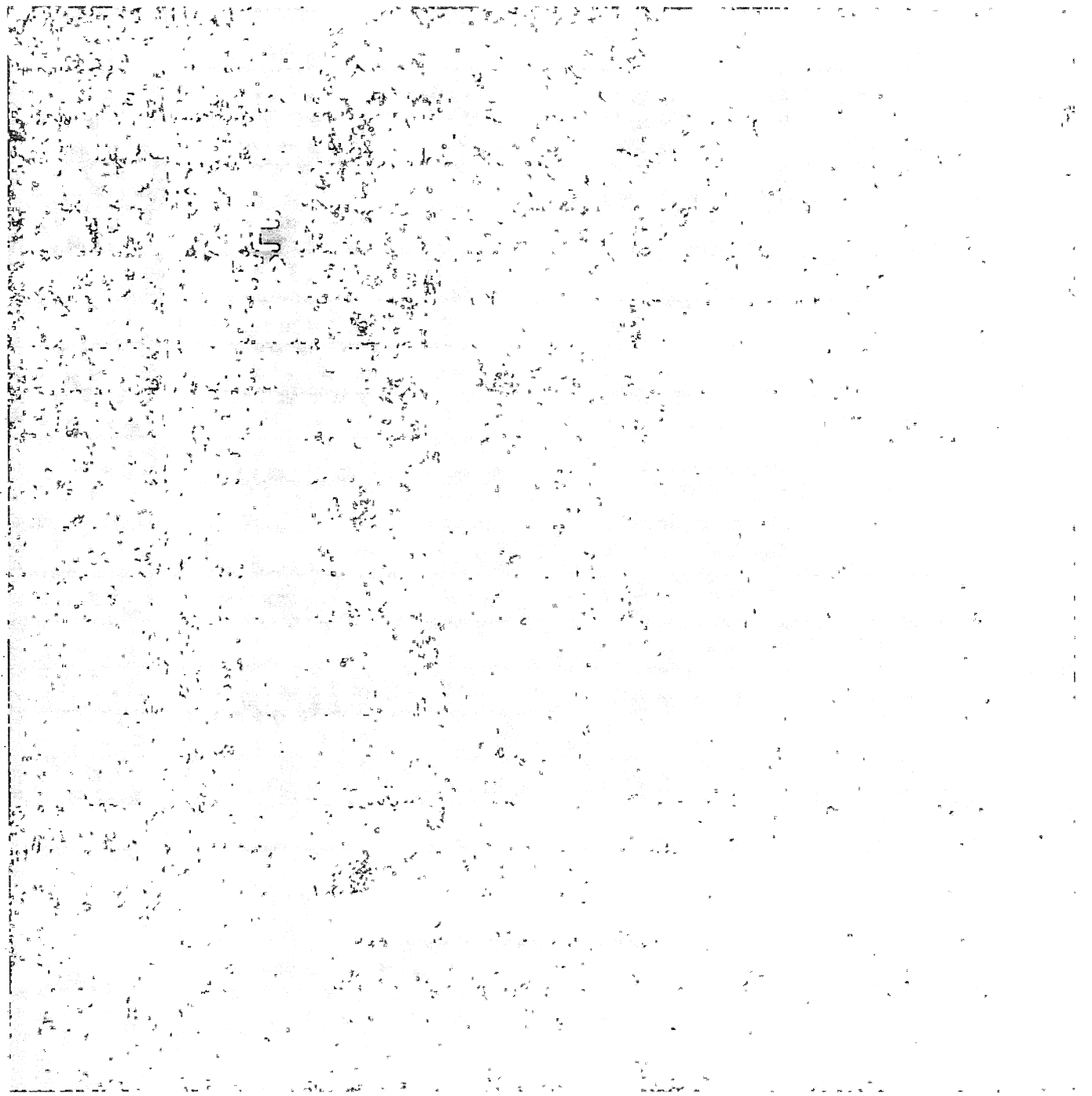
North Central
Forest Experiment
Station

General Technical
Report NC-67

Database Management:

a Forest Inventory Application

Dennis L. Murphy



CONTENTS

North Central Forest Experiment Station
Forest Service—U.S. Department of Agriculture
1992 Folwell Avenue
St. Paul, Minnesota 55108

Manuscript approved for publication July 3, 1979
1981

DATABASE MANAGEMENT: A FOREST INVENTORY APPLICATION

Dennis L. Murphy, *formerly Associate Mensurationist, (currently Senior Data Analyst, EROS Data Center, Sioux Falls, South Dakota)*

Forest Survey inventories provide data for a variety of users, including private and public forest managers, forest scientists, and the general public. Because user needs have changed in recent years, one standard set of tables no longer provides adequate information. In attempting to find a faster and more flexible method for responding to special requests for inventory data, a study of database designs was initiated using forest inventory data from Minnesota's Aspen-Birch Forest Survey Unit. Two databases were created, representing two levels of information detail. The databases were compared with each other and with standard forest inventory retrieval methods to evaluate their cost-effectiveness and ability to satisfy user demands.

The evaluation was based on the costs incurred at the University of Minnesota Computer Center. The center operates two Control Data Corporation computers: models Cyber 74 and Cyber 172.[1] Both computers utilize the NOS operating system software. Although costs at other installations will vary, the results will provide the reader with levels of the costs of creating, operating, and applying the databases.

DATABASE MANAGEMENT SYSTEMS

Many disciplines have experienced changes in user information requirements. In the 1950's and 1960's, data processors designed data files within individual programs (Sibley 1976), creating a situation of limited user access to the data. This method of designing data files also led to duplication of data, since individual programs tended to have their own files for data storage.

Problems with data access gradually led to the concept of the integrated database. Software systems were developed that provided integrated data structures, separate access languages for nonprogrammers and programmers, and data integrity functions (Everest 1976). Moser (1976) demonstrated the application of a database management system (DBMS) to an operational forest inventory, the compartment examination. Meteer (1975) argued that a second type of forest inventory, the planning inventory, could also benefit from increased user access and report flexibility. A database management system can provide the benefits of flexible access while also providing data integrity functions to control the quality and use of the data resource within an organization.

The database management system used in this study was System 2000 (MRI Corporation 1974). Two trial databases for Minnesota's Aspen-Birch Unit were developed, one from the tree detail file and one from the plot summary file. Data from the fourth Minnesota Forest Inventory (1977) supplied input to the databases. The following steps were taken to create each database.

Designing the Database

Two major factors must be considered in designing a database: (1) the logical design, and (2) the physical consequences of the design.

The former requires the analyst to identify the entities to be stored in the database, the attributes of each, and the relations among them. This step involves comparing the effectiveness of various designs in processing anticipated reports.

[1] *Mention of trade names does not constitute endorsement of the products by the USDA Forest Service.*

The second factor involves estimating the physical storage space required for the database. Every design has a physical consequence; for example, a database created in System 2000 physically exists as six random access files. The size of these files is a direct consequence of the logical design for the database, and can be estimated using ESTIMANUAL (MRI Corporation 1974). In this study a program[2] was available from the University of Minnesota Computer Center to calculate size estimates, which were then used to evaluate design alternatives.

The choice of a design for this study involved a trade-off between processing flexibility and cost of the physical storage space required. The logical design selected accommodated the major anticipated reports while using storage space most efficiently.

Defining the Database

The conceptual structure of the entities was translated into a physical data structure in the data definition language of System 2000. To the new user of a database management system, this step will not be an obvious one. However, the database must be defined before the data can be converted into the standard internal format of System 2000.

Creating the Database

A FORTRAN PLI program was written to convert the data to the internal format. PLI is a System 2000 software package that provides an interface between a FORTRAN program and the database.

Establishing Administrative Procedures

Administration of the database involves ensuring data security and integrity. User access can be restricted by assigning passwords to sensitive elements; for example, the location of an inventory plot in the case of private ownership. Providing a backup copy of the database on magnetic tape ensures against accidental destruction.

Providing Documents for User Interface

Two manuals were written for subsequent use of the database management system.[3] The first, a database manager's guide, outlines the steps required to define, create, and archive the database. With this manual, data management personnel can create databases for other survey units. The second manual is the user's guide. This manual contains introductory information for the new user, a description of the data, and reference material for access languages. Both of the manuals are essential to the implementation of the database management system.

RESULTS

Plot Summary Database

Description

The database structure selected from the available alternatives was a two-level, single-path heirarchy. The first level contains plot attributes (e.g., stand age, acres, forest type, site index). The second level contains species group attributes (e.g., species group code, number of trees per acre, growing stock volume per acre in cubic feet and board feet). A graphical representation of the data structure specifying attributes accessible at each level is called a schema diagram (fig. 1). The relation of the plot data attributes to the species group attributes is one-to-many; the number of species groups for a plot equals the number of species groups measured on that plot. When database size was estimated, it was assumed that there would be six species groups per plot. This assumption was high; the average was about four species groups per plot. The estimated size was 9.5 million characters while the actual size was 6.5 million characters.

Use

The plot summary database can be used to answer two types of requests: simple retrievals (queries) and reports. A user may be interested in a listing of data for plots meeting specific characteristics; this is a simple retrieval. For example, a resource analyst requested a listing of plot numbers, acres per plot,

[2] *S2KESTM, a FORTRAN program which estimates System 2000 database files.*

[3] *The manuals document procedures for office personnel. They are not formal publications.*

and ground land use for plots located in what was formerly known as the portal zone of the Boundary Waters Canoe Area. The results are shown below. The data were retrieved and printed at a cost of $0.15. The retrievals involved about 5 minutes of an individual's time at the terminal.

Plot No.	Acres/plot	Ground land use code[4]
528	1,246	20
533	1,330	20
535	1,371	20
536	1,371	20
540	1,395	20
541	1,161	90
542	1,395	20
543	1,625	20
661	1,537	20
662	1,537	20
664	5,449	80
2804	1,648	20
2836	1,451	20
2839	1,395	20
2842	1,395	20
2843	1,371	20
3208	1,566	65
3220	1,211	20
3223	1,451	20
3224	1,451	20
3235	1,371	20
3237	1,537	20
3238	1,625	20
3239	1,648	20
3241	1,395	20
5010	1,627	40

A second query came from a research forester who wanted to obtain the range of site index values for each of 14 forest types. The questioner constructed commands to print the minimum value, the maximum value, the average value, and the standard deviation. An example of the output for one forest type, jack pine, is shown below (18* signifies site index):

```
MIN 18* 38
MAX 18* 80
AVG 18* 56.000
SIG 18* 9.724
```

[4]*Definition of ground land uses: 20, plot previously designated as commercial forest; 40, productive reserve; 65, marsh; 80, noncensus water; 90, census water.*

The total cost of calculating these values for 14 forest types was $11.50 (56 values; an average of $0.21 for each value). The operation involved about an hour at an interactive terminal.

Formal summary reports are more appropriately produced via the Report Writer module of System 2000 or a procedural language applications program. Resource inventory data are commonly used to provide summary tables. These tables can be produced in the Report Writer language for reasonable cost and with considerable flexibility. The report commands, like the retrievals, can be written at an interactive terminal. Two examples of tables produced by Report Writer are given below.

A request was received for summary tables of acreage by site productivity, age, and basal area categories for several forest types. The results, shown in Table 1, cost $1.50 per table. Table 1 also indicates treatment opportunities in the survey unit. For example, in stands with a site productivity of 85 to 120 cu ft/acre/year, approximately 6,829 acres had sufficient basal area for thinning. The table has the potential added benefit of grouping plot data for later

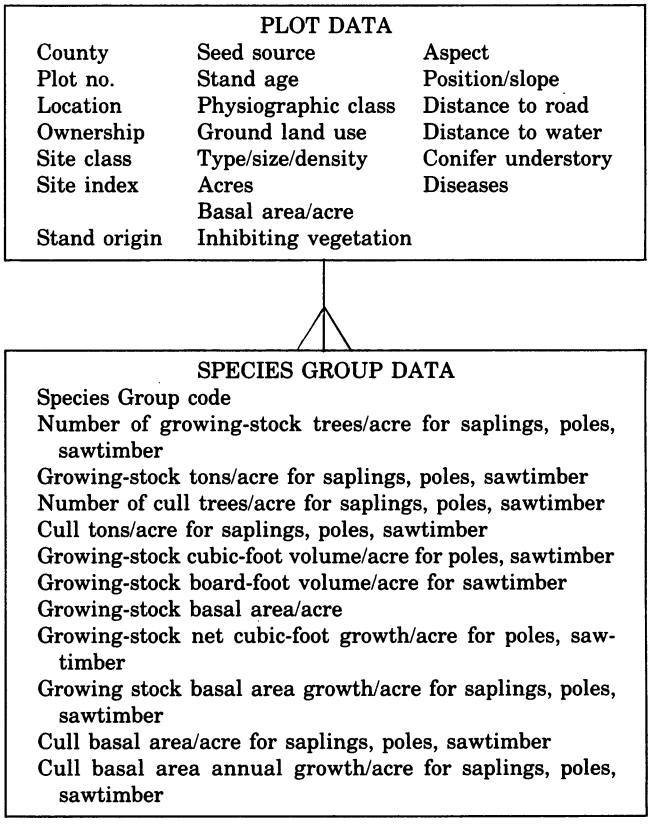


Figure 1.—*Plot summary database schema diagram.*

Table 1.—*Commercial forest area of red pine forest type, summarized by site productivity, age, and basal area classes, Minnesota Aspen-Birch Unit, 1977 data*
(In acres)

Site productivity (potential cubic-foot per acre per year (MAI) and age class (years)	Basal area (square feet/acre)						
	30	60	90	120	150	180	Total
120-165 cu ft							
31-40			1,364				1,364
Subtotal			1,364				1,364
85-120 cu ft							
21-30				1,178			1,178
31-40					2,698		2,698
41-50				1,797			1,797
51-60					2,759		2,759
81-90					1,372		1,372
Subtotal				2,975	6,829		9,804
50-85 cu ft							
0-10			1,567				1,567
11-20	2,878						2,878
31-40			1,372				1,372
41-50					2,766		2,766
51-60			1,451				1,451
61-70				4,012			4,012
71-80						7,971	7,971
81-90		1,323					1,323
91-100					1,139		1,139
101-120		1,330					1,330
Subtotal	2,878	2,653	4,390	4,012	3,905		7,971
20-50 cu ft							
41-50			1,371				1,371
61-70			1,324				1,324
81-90	1,451						1,451
121-140						1,371	1,371
Subtotal	1,451		2,695			1,371	5,517
Total	4,329	2,653	8,449	6,987	10,734	9,342	42,494

input into a linear programming module. (Data management personnel would modify the program to provide growing stock volume in addition to acreage.)

Several forest land managers responsible for managing properties within the Aspen-Birch Unit requested a table summarizing acreage by site index, forest type, and stand-age class for specific ownership and stocking levels (table 2). This is another example of a summary report using the Report Writer module. The cost of producing this table was $21.00. The Report Writer module was modified to produce tables for two stocking levels and three ownership classes, for a total of six tables.

Figure 2 demonstrates another capability of the system: an interface with an applications program. A simple Report Writer command file was written to summarize acreage by stand-age class for a specific forest type. The system created an intermediate data file which was used as an interface between the database and the graphing package. The total cost of retrieving the appropriate survey plots and plotting the histogram was $1.50.

Costs

The initial cost of computer processing for establishing the Aspen-Birch Unit[5] plot summary data-

[5]*The unit contains 4,907 plots.*

Table 2.—*Commercial forest area by forest type, age, and site index, for good/medium stocked stands, on public ownership excluding national forest, within the Minnesota Aspen-Birch Unit, 1977 data (only a portion of the table shown here)*

(In thousand acres)

Forest type and age class	Site index								
	0-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	Total
Jack pine:									
0-10				1.2					1.2
21-30			1.6		1.3				2.9
31-40				3.0	2.8	1.4			7.2
41-50				4.1		1.0			5.1
51-60				4.2					4.2
61-70		1.5	1.8	1.4					4.7
Subtotal	1.5	3.4	13.9	4.1	2.4			25.3	
Red pine:									
0-10				1.6					1.6
11-20				2.8					2.8
21-30						1.2			1.2
31-40					1.2				1.2
61-70		1.3	1.4						2.7
Subtotal		1.3	1.4	4.4	1.2	1.2			9.5
White pine:									
11-20				1.4					1.4
Subtotal				1.4					1.4
Black spruce:									
0-10	3.7	2.6	1.2						7.5
11-20	16.6	9.8	14.6	1.4					42.4
21-30	28.1	39.7	11.8	1.2					80.8
31-40	33.6	22.0	13.2	2.5					71.3
41-50	17.0	8.9	8.1	3.8					37.8
51-60	17.6	7.7	12.0	1.3					38.6
61-70	8.0	9.3	9.8	4.2	1.3				32.6
71-80	12.2	13.3	8.4						33.9
81-90	8.6	8.4	4.1						21.1
91-100	14.8	5.1	1.4						21.3
101-120	8.6	4.1							12.7
121-140	1.4	5.5							6.9
Subtotal	170.2	136.4	84.6	14.4	1.3				406.9

base was $122.50. This cost does not include the purchase price of System 2000 or the cost of initial software development for the load program.

Initial costs:	
Definition	$ 2.50
Creation	
Load	50.00
Index	50.00
Archive	20.00
Total	$122.50

A daily disk storage cost of $7.00 would be incurred if the database were maintained on disk for the entire operating day of the computer system. Considering the low cost of copying the database from magnetic tape to disk storage ($2.50 for the entire database), the database was stored on disk only when needed.

Operating costs:	
Daily disk storage	$ 7.00
Copy from tape to disk	2.50
Release	.05
Retrievals and reports	variable

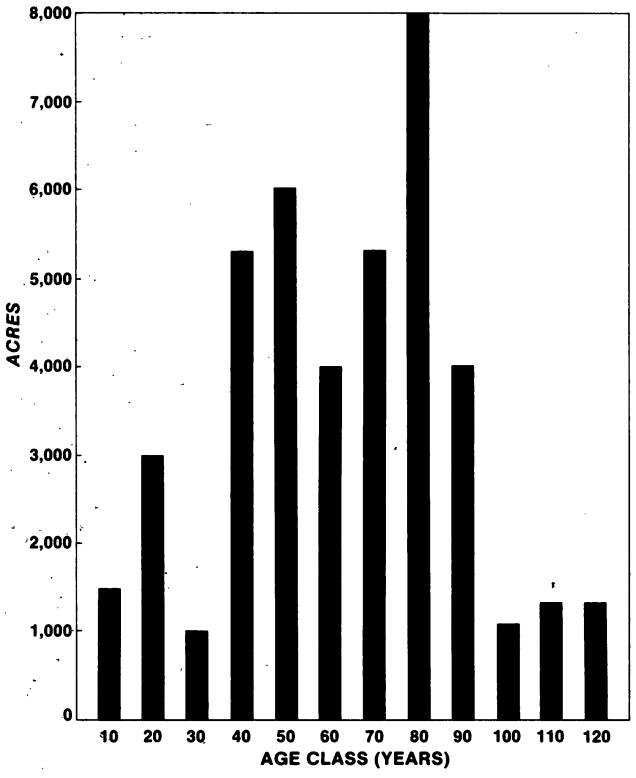


Figure 2.—*Acres of red pine forest type by stand age class, all ownerships, Minnesota Aspen-Birch Unit, 1977.*

Evaluation

The cost of the plot summary database approach was compared with that of conventional data processing for queries and summary tables. The comparison for simple queries is given below based on the example of site index ranges given previously.

	Present method	Query language (database)
Programming	$32.00	$ 8.00
Computer time	2.00	12.00
Copy to disk & 1 hour storage	—	3.50
Total	34.00	23.50
Fixed cost of creating database	—	$122.50

The major difference between the two methods is in programming time; the query language reduces programming time by 75 percent and serves the nonprogramming user. The difference in personnel time was estimated at 3 hours, or $24.00.[6] The computer time to produce the same result by use of FORTRAN program was not readily available. The cost was assumed to be no more than $2.00, based on other available programs. The use of the database involves the cost of copying the database to disk and storing it for 1 hour, as well as a fixed creation cost of $122.50. Figure 3 compares conventional costs with database costs for a varying number of requests per year. The break-even point is about 12 requests per year; after that, the database becomes the most cost-effective method.

The costs of the two alternatives for producing summary tables are presented below:

	Present method	Report Writer (database)
Programming	$128.00	$ 65.00
Computer time	5.00	10.00
Copy to disk & 1 hour storage	—	3.50
Total	133.00	78.50
Fixed cost of creating database		122.50

Survey personnel estimated it would take 2 days to write a program for a special request. A similar

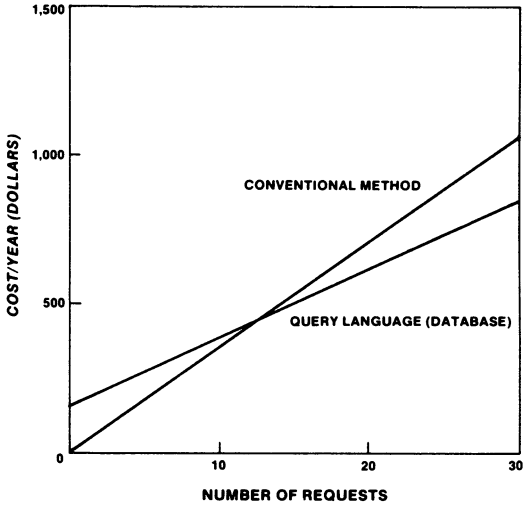


Figure 3.—*Cost comparisons for simple queries, Minnesota Aspen-Birch Unit, 1977 data.*

[6]*The estimate was made by the research forester performing the query.*

report can be written in 1 day with Report Writer. The processing cost for the present method is about $5.00 per table, based on example tables. The Report Writer costs averaged $10.00 per table for 18 relatively complex reports. Figure 4 compares conventional costs with database costs for a varying number of requests for special tables. The break-even point is about three requests per year. After that, the database becomes more cost-effective.

Tree Detail Database

Description

The tree detail database had a two-level hierarchical data structure similar to that of the plot summary database. The first level contains the same plot attributes as the plot summary database; the second level contains individual tree attributes instead of species group summaries (fig. 5). The relation between the two entities is one-to-many; that is, the lower level entity is repeated for a plot as many times as there are trees measured for that plot.

Use

The tree-detail database can be used to prepare the same formal summary reports for a survey unit or other area as those produced from the plot summary database, but test runs indicate that the cost may be 10 times higher. The size and complexity of the tree-detail database make it an inefficient means of writing such summary reports. The plot summary database provides this capability more efficiently.

Some information requests may require retrieval of individual tree detail. Examples include providing input tree lists for tree growth projection systems (Leary *et al.* 1977) and producing special summaries of information not contained in the plot summary database. To meet these needs a tree-detail database can be developed.

Costs

The direct computing costs for establishing and operating a tree-detail database for Minnesota's Aspen-Birch Unit were:

Initial costs:	
Definition	$ 2.50
Creation	
Load	330.00
Index	620.00
Archive	25.00
Total	$977.50

Operating costs:	
Daily disk storage	20.00
Copy from tape to disk	7.00
Release (purge files)	.05
Retrievals and reports	variable

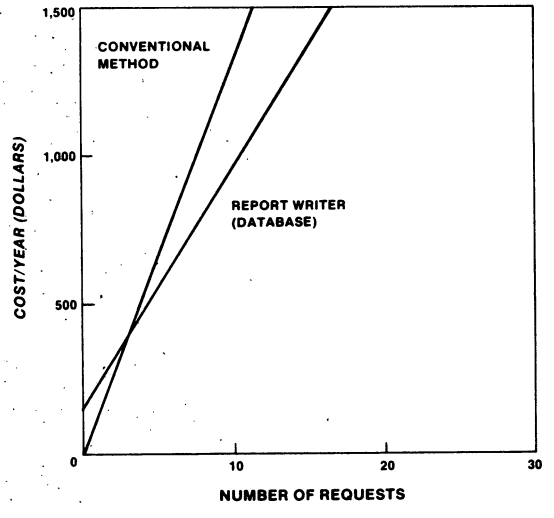


Figure 4.—*Cost comparisons for summary reports, Minnesota Aspen-Birch Unit, 1977 data.*

Point number	Two inch D.b.h. class
Tree number	Tree size class
Tree history	Softwood hardwood species group code
Species	Species group
D.b.h.	
D.b.h. increment	Point occupancy
Bole length	Ingrowth code
Bole length top diameter outside bark	Stocking size class
Cull cubic feet	Stocking percent
Saw log length	D.b.h. last measurement
Saw log top diameter outside bark	Trees/acre
Cull board feet	Net cubic foot vol/acre
Log grade	Net cubic foot saw log vol/acre
Surface defect	Net board foot vol/acre
Internal defect	Annual cubic-foot growth/acre
	Annual board-foot growth/acre
Relative bole length	Annual cubic-foot mortality/acre
Crown ratio	Annual board-foot mortality/acre
Crown class	Basal area/tree/acre
Damage cause of death	Green tons of bole/acre
Tree cover class	Green tons of top/acre

Figure 5.—*Individual tree data for tree detail database.*

This tree-detail database included 4,907 plots and 24.5 million characters.

Costs of disk storage and copying the backup tape to disk are higher here than for the plot summary database due to the increase in the total number of characters for the database.

Evaluation of Alternatives

To evaluate the usefulness of the tree-detail database, two types of requests requiring tree detail data were examined: developing tree lists for input to a tree growth projection system, and developing resource information that requires individual tree records.

Three alternatives for providing input tree lists for a tree growth processor were compared:

1. Using the present method of reading the entire file of tree detail information stored as a packed binary sequential file.
2. Using a tree-detail database to create an intermediate file from which the tree growth processor will subsequently read.
3. Using a plot summary database for selecting plots based on requested characteristics (e.g., red pine plots with basal area greater than 120 sq ft/acre), and creating a file of the selected plot numbers. The file of plot numbers is used to read the tree data from a modified format for the present sequential tree file.

The costs for the three alternatives are presented in table 3. The cost of reading the sequential file makes the present method expensive for repeated runs. The tree-detail database is less expensive. However, the plot summary database used in conjunction with an unformatted binary sequential file of tree detail data is cheaper still. For more than two requests per year, alternative three (the summary database used with a modified tree file) is the most economical method (fig. 6).

It is difficult to predict the variety of requests for tree detail information for research purposes. The potential of the database for research is just beginning to be explored. However, to provide some idea of the cost, a recent request will serve as an example.

A scientist was interested in obtaining information on tree distribution by diameter class for red pine plots within a specified range of site index values. The database was used first to obtain a list of plot numbers meeting the selection criteria; subsequently, a Report Writer command file was executed for individual plots from the list of selected plot numbers. The command file created an intermediate file which summarized the number of trees by 2-inch diameter classes. The intermediate file provided input to an applications program, which created a histogram (fig. 7). The entire sequence was performed without the user writing a program. The steps were repeated several times for different plots. Each histogram cost approximately $2.00.

CONCLUSIONS

Plot Summary Database

For both queries and reports, the System 2000 database management system had advantages over the processing of a sequential plot summary file. The savings in personnel time alone paid for the costs of database establishment. In addition, the software provided previously unavailable reporting capabilities and access to users. The plot summary database approach is a viable alternative for processing special requests and should be evaluated on a region-wide basis.

Table 3.—*Cost comparison of three alternatives for providing tree growth processor input*
(Dollars)

Activity	Present method; packed sequential file	Tree detail database	Plot summary database & modified tree rate
Programming	8.00	4.00	4.00
Computer time (create and read file)	120.00	33.00	26.00
Copy & storage (assume 1 hr storage)	—	8.00	3.30
Total	128.00	45.00	33.30
Fixed cost of database creation	—	975.00	150.00[1]

[1]Includes costs of additional tapes and their storage for the modified sequential files of tree detail data.

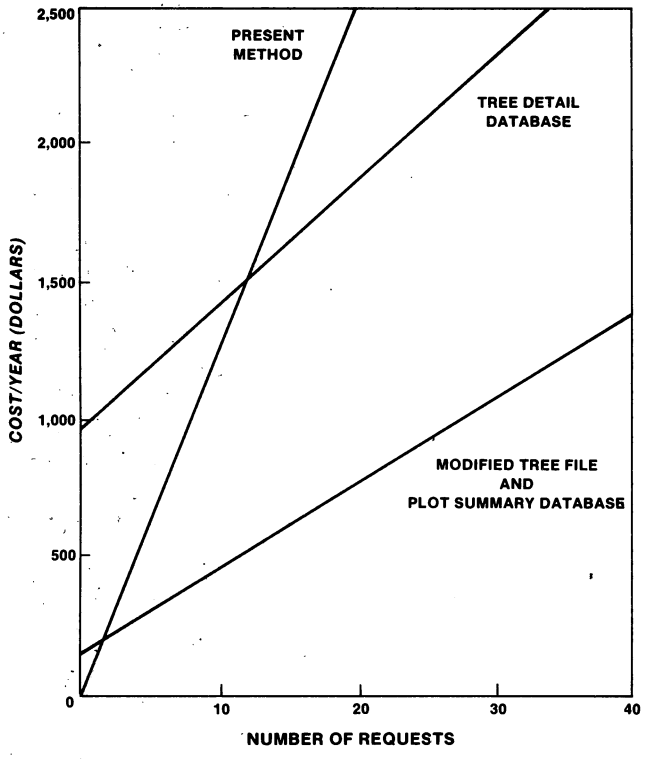


Figure 6.—*Cost comparisons of creating tree growth processor input files, Minnesota Aspen-Birch Unit, 1977 data.*

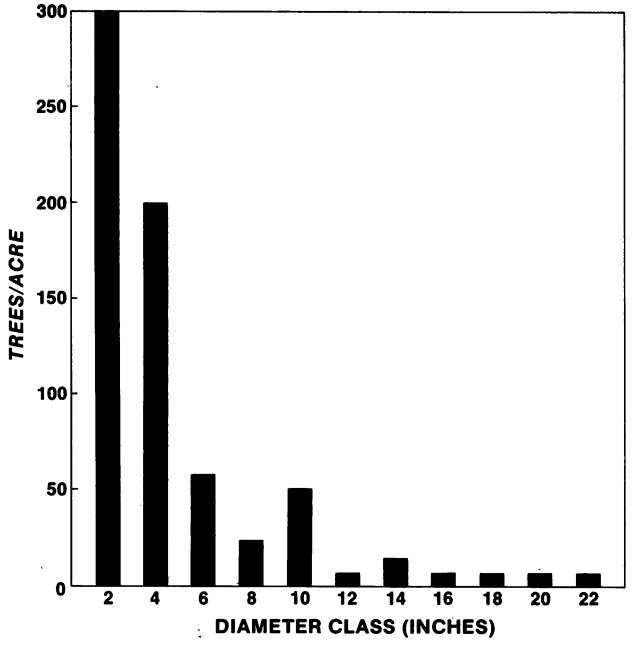


Figure 7.—*Trees/acre by 2-inch diameter class for a sample inventory plot, Minnesota Aspen-Birch Unit, 1977.*

Tree Detail Database

The tree detail database is not appropriate for unit-wide summary reports, nor is it cost-effective for creating intermediate files to be used as input to application programs. The real potential for the tree detail database lies in research applications, but additional experience will be needed to determine the cost-effectiveness. Very likely, the increased accessibility of the survey data will justify the cost of creating a tree detail database for work units that have specific research assignments.

As these database management system applications are implemented, however, it is important that the sampling error of retrieval data be evaluated. Studies are now underway to do this. Subsequent reports will also deal with the interface of the database management system with the tree growth processor to update past inventory data.

LITERATURE CITED

Everest, G. C. 1976. Database management systems tutorial. *In* Readings in management information systems. p. 164-187. McGraw-Hill, Inc., New York.

Leary, Rolfe A., Jerold T. Hahn, and Roland G. Buchman. 1977. Tests. *In* A generalized forest growth projection system applied to the Lake States region. U.S. Department of Agriculture Forest Service, General Technical Report NC-49, p. 79-89. U.S. Department of Agriculture Forest Service, North Central Forest Experiment Station, St. Paul, Minnesota.

MRI Corporation. 1974. System 2000 reference manual. MRI Corporation, Austin, Texas.

Meteer, J. W. 1975. Forest management inventories in the computer age. Forest modeling and inventory, selected papers from 1973 and 1974 meetings of midwest mensurationists. A. R. Ek, J. W. Balsiger, and L. C. Promnitz, eds. School of Natural Resources, College of Agriculture Life Science, University of Wisconsin, Madison, and Society of American Foresters p. 102-109.

Moser, J. W., Jr. 1976. Illustration of a generalized data base management system for maintaining compartment records. p. 39-52. *In* Resource data management Symposium. J. W. Moser, Jr., ed. Purdue University, Department of Forestry and Natural Resources, and Cooperative Extension Service.

Sibley, E. H. 1976. The development of data base technology. Computing Surveys 1:1-5.

Murphy, Dennis L. 1981. Database management: a forest inventory application. U.S. Department of Agriculture Forest Service, General Technical Report NC-67, 9 p. U.S. Department of Agriculture Forest Service, North Central Forest Experiment Station, St. Paul, Minnesota.

Describes the design and creation of two databases using data from a unit of the fourth Minnesota Forest Inventory, and analyzes the cost-effectiveness of each.

KEY WORDS: data processing, break-even analysis, cost-effectiveness, logical database design, schema diagram.

☆ U.S. Government Printing Office: 1981—767-252/8 Region No. 6