FINAL REPORT

SOIL/FOREST PRODUCTIVITY RELATIONSHIPS IN ST. LOUIS COUNTY RESULTS OF ANALYSIS OF WOODS5 DATA

By

William E. Berguson, Scientist
Karen Updegraff, Consultant

January 1990
NRRI/TR-90/01

Prepared for:

St. Louis County

Natural Resources Research Institute
University of Minnesota, Duluth
5013 Miller Trunk Highway
Duluth, MN 55811

EXECUTIVE SUMMARY

The goal of this project is to organize and analyze the Wood5 data set to aid in the formulation of forest management recommendations in St. Louis County. Collection of Wood5 data has been ongoing since the start of the soil survey. The purpose of the Wood5 project is to collect detailed soil and forest growth data on a variety of soil and forest cover types to develop relationships between soil characteristics and forest growth.

The first objective of this project was to organize the Wood5 data into computer format to allow easy retrieval and analysis. At the beginning of the project, Wood5 data existed in a variety of formats in a variety of places. Some of the soil data had been entered into computer format, while much of the forest growth information existed in hard copy form. The first objective was to organize the Wood5 data to produce a database that contained all of the detailed soil and forest growth data. This was viewed as a necessary step to permit statistical analyses of the data. Also, the organization of these data in computer format will allow rapid retrieval and update of the database as information is collected in the future.

Forest growth data (Wood5 forms) existed primarily in hard copy form and were entered into computer format by hand. The next step was to match the Wood5 forms containing the detailed forest growth information to soil profile descriptions on these sites. This process produced a complete database containing approximately one hundred Wood5 plots with soil profile descriptions at each site. This merged database was translated to Paradox database software, which supports translation of data to a variety of formats.

The merged soil and forest database was then used to calculate soil attributes that were expected to relate to forest production. A number of soil and site attributes were used in statistical analyses. Some of these are available water capacity, percent coarse fragments, depth to water table, pH, slope, and landform. The completeness of these data were found to vary from plot to plot. Missing data on plots prevented some analyses from being done. In particular, data on slope and presence/absence of a shallow water table were found to be potentially useful but extensive analyses were not possible due to missing data.

Published site index equations were applied to each tree of the dominant species (based on basal area) and an average plot site index calculated. In addition to site index, an index of volume increment was calculated using the basal area and height of all sampled trees. This volume index was then divided by the average age of the site index trees to produce an approximate mean annual volume increment.

Statistical analyses of these data were done to investigate relationships of soil properties to the two indices of forest production mentioned above. After initial analysis, it became apparent that detailed analysis of specific combinations of cover type and soil series would not be possible due to a lack of data for any one combination of soil series and forest cover type. Further analyses proceeded using forest growth information aggregated into two groups; aspen/birch and upland conifers. Analyses were done to assess the relationship of soil properties to aspen/birch and upland conifer growth indices both within and across soil great groups.

Analyses showed a number of statistically significant relationships. Particularly, the relationship of upland conifer site index to great group, slope, drainage class, and landform. These showed that Dystrochrepts, Eutroboralfs, Udipsamments, and Glossoboralfs differed

significantly with drainage class, slope, and landform, further influencing growth. Growth relationships for the aspen/birch group were not as evident with only the relationship of decreasing aspen site index with increasing slope being statistically significant. Equations explained approximately 68 percent of the variation in upland conifer site index.

Based on results of these analyses and conversations with Land Department personnel, analyses were done to assess the number of plots required to achieve a predetermined accuracy using site index as the growth variable of interest. Indications are that the use of soils data will be used for management at the soil series or mapping unit level. Because of this, a calculation of the number of plots needed to accurately calculate site index for the species groups at the series level was done. These calculations are based on a 90 percent Confidence Limit for a range +/- 5 percent of the mean site index. Overall, a total of approximately 400 plots would be needed to develop equations to the desired accuracy, four times as many plots as are presently in the database. Data collection efforts would have to be increased to achieve this accuracy.

Given the results of analyses, recommendation for further sampling are discussed. Obviously, the most cost-effective method to increase the number of plots is to use existing forest inventory information. Most of the variables found to be influencing forest growth are site related and do not require detailed soil profile descriptions. Verification of the soil series could be done by survey personnel without a detailed profile description.

Site index in the county forest inventory program is based on one site index tree per plot. Reservations regarding using the existing forest inventory centered on the potential loss in accuracy in determining site index by using only one site index tree per plot. Phase I (FIA) data were used to address these concerns. One tree was randomly selected from a group of site index trees on a plot and compared to the average site index for the plot. These analyses showed that little accuracy is lost in the determination of site index based on one tree.

Based on this, a pilot data collection program is proposed using the existing forest inventory information. This inventory would be supplemented with soil and site descriptions concentrating on predominant commercial species and those soil series on which these species are likely to be most productive. This pilot project would concentrate on recording the time invested in each phase of the data collection procedure to calculate costs and maximize the effectiveness of further sampling.

INTRODUCTION

The purpose of this project is to collect and organize data related to soil/forest production information (Wood5 data) and statistically analyze this data set. The ultimate goal of these analyses is to better understand the influence of soil factors on forest production in St. Louis County and use these relationships in forest management where applicable. Results of these analyses will also be useful to serve as a guide for further soil/forest production sampling to streamline data collection procedures and maximize the benefit derived from further sampling.

This project is comprised of three main tasks as follows:

1. Organize existing soils and forest growth information (Wood5 data) into an integrated computer framework to allow easy retrieval and analysis.

2. Conduct statistical analyses of the Wood5 data to determine the potential to develop soil/productivity relationships.

3. Report results of statistical analyses and formulate recommendations to improve practical application of soil survey data for forest management in the county.

This report, along with other information previously submitted by Don Prettyman, will more clearly define issues related to the use of soils information in the County.

TASK I. COMPUTERIZATION OF SOILS/WOOD5 DATA

Two databases were constructed using data compiled by the St. Louis County Soil Survey. The first database, the county soils database, was converted from those maintained by the Soil Survey (SiteInfo Database). This data contains complete horizon descriptions for pedons throughout the county. The second, the Wood5 database, is comprised of a select subsampling of survey plots for which detailed vegetation and productivity data had been collected. Survey plots in these databases were matched and the complete data set used to test soil/site productivity relationships for the county. The data were analyzed using standard regression and graphical analysis methods.

Methods

The plot productivity data (Wood5 plots) were entered by hand into R-Base database format. These data were later converted to ASCII format and to Paradox® database format. The database structure and field descriptions are shown in Appendix 1. The Wood5 data were then merged with the corresponding soil data, which was selected out of the larger Soil Survey database. Plot matching was achieved by location (township, range indicators) and further cross-checking the plot notes for references to Wood5 plot numbers. The Wood5 numbers could not be used throughout this procedure as they were not included consistently and no obvious correlation to the Soil Survey's soil identification codes existed.

The net result of the merging process was a data set of 100 plots having complete soils and forest growth information. For each plot, information on soil depth, texture, pH, parent material, slope, aspect, and landform, was available with varying degrees of consistency. In addition, overstory species, basal area, cover density, age, and height, were included. Data had also been collected for understory species but these were not in a form relevant to the present analysis. The plot basal area (BA) is based on a variable-radius prism plot, however, dbh, height, and age data, on at least 5 dominant-codominant trees, enabled the estimation of mean plot age and mean annual increments of BA and volume growth. While volume was averaged only over the sampled trees, plot BA X mean height was used to generate an index of standing volume; this was divided by mean age of sampled trees to obtain an index of mean volume production for the plot. Finally, site index was calculated for individual trees using equations published by Laidly (1979). Basal area of each species on each plot were then summed to determine the dominant species. Using this indicator of dominance, the average site index of the trees comprising the dominant species for each plot was used to represent plot site index.

Soil data was also manipulated to produce variables more relevant to forest production. Most prominent among these calculated variables was available water capacity (AWC). Since no direct measurements for this variable were available, average AWC values for each textural class published by the Soil Conservation Service were used to estimate AWC. These values were applied to the textural classifications for each horizon and weighted by horizon thickness. The resulting estimates (in inches/inch of soil) were summed over the entire profile and for the top 30 inches of each profile. An adjusted AWC was produced using total percent coarse fragments for each horizon to reduce the AWC proportionately. Coarse fragments were also entered as an independent variable by averaging percent coarse fragments over the profile, weighting the top 30 cm at twice the value of the lower horizons. Total depth of sampling was assumed to represent depth to bedrock or some other restrictive layer (such as a pan or the water table) and therefore meaningful with respect to rooting volume. The limited data collected on rooting depth was useful to the extent that it was possible to state the presence or absence of roots at a given depth. As it was not quantitative and not consistently present in the data, no more analyses were attempted using root data. Only data on fine roots were included in analyses as data on the presence/absence of medium and large roots was too sparse to be useful.

Following the calculation, coding and selection of variables, the data were translated to the SYSTAT (Wilkinson 1986) statistical package for analysis. In general, the number of variables collected in the Wood5 data set were far greater than statistically useful as true relationships become obscured when the data set is dissected into a large number of groups containing fewer and fewer cases (plots). Because of this, data were aggregated into higher levels as analyses progressed.

The plots used in the analyses included most of the species normally found in northern forests. Data were then divided into species groups, aspen/birch (species group 1), upland conifers (species group 2) and others including wetland species and miscellaneous hardwoods (species group 3 - only three plots). Due to the lack of plots, species group 3 was not used for analyses. The aspen/birch group (Populus tremuloides, balsamifera and grandidentata, and Betula papyrifera,) included 43 plots, while 54 plots were assigned to the upland conifer group. The latter were comprised largely of plots dominated by white spruce (Picea glauca), white pine (Pinus strobus), red pine (P. resinosa), jack pine (P. banksiana), and balsam fir (Abies balsamea).

Landforms were classified into "rolling plains or uplands" and "level uplands" with the exception of one peatland site which was later excluded. While the Soil Survey data included rather detailed landform indicators, these contained too many categories to be useful. Slopes fell mostly in the one to ten percent range, with only a few sites with slopes greater than fifteen percent. The majority of the plots had drainage classes ranging from three to six, available water capacities of less than 15 inches, and were of medium texture. Of the 13 soil Great Groups represented, four groups (the Eutroboralfs, Dystrochrepts, Glossoboralfs, and Udipsamments) contained 82 of the plots. With the exception of one Borohemist (a Greenwood series), all of the other soils were also classified as Entisols, Inceptisols, and Alfisols.

Four different approaches to grouping were used. First, the entire data set was tested using soil series. Individual soil series were then aggregated to the great group level with seven great groups represented. Other descriptors used as indicator variables included parent material, presence/absence of a shallow water table (less than 40 inches), and depth to restrictive layer as indicated by a "very firm" moist consistency (bulk density of at least 1.8).

TASK II. STATISTICAL ANALYSES

Data Analysis Procedure And Model Development

Initial analyses were run for group means and variance where possible to characterize the data set as a whole. Each independent variable was graphed individually against each dependent variable (mean annual basal area and volume production, site index, plot volume index, and mean annual production) to show basic trends in the data. No data transformations were done during these analyses as no clear curvilinear relationships were identified on any of the scatter plots. Individual variables were then tested using simple linear regression or ANOVA analysis.

Full multiple regression models combining all variables were then constructed. At this point it became necessary to eliminate some independent variables such as pH and depth to resistant layer, due to the small sample size resulting when these variables were included in the analyses. These reduced data sets were due primarily to the fact that data were missing from plots. During this process, the large number of variables (see Appendix 1) were reduced for inclusion in "full" model analysis.

The final form of the full model contained independent variables for slope (percent), cosine of degrees azimuth, drainage class (an integer), total sampling depth (inches), weighted average percent coarse fragments, AWC and AWC for 30 inches, depth of fine root presence (inches), soil great group and presence/absence of shallow water table. Again, due to missing data for some variables, this model only had a sample size of 34, which included all species together. Although analyses including all species were done, these were used to characterize the data as a whole and not to develop practical relationships for management at this point in the analyses.

The coefficient of determination (R^2) and individual p-values for variables were used to assess the accuracy and relative value of each model. The R^2 is the amount of variation in the dependent variable (in most cases site index) explained by the independent variable. That portion of unexplained variation ($1-R^2$) is due to sampling error, the limits of accuracy of calculated variables (such as AWC) and the omission of potentially important variable influencing production (e.g., chemical characteristics). In no case was an R^2 of greater than 0.50 obtained at this point in the analyses. Stepwise regression was performed on the full model, excluding categorical variables. Categorical variables (e.g., great group) are those which do not have a natural ordering. A categorical variable can be thought of as a "family" as opposed to a continuous variable such as percent slope.

Results of Statistical Analyses

In order to correctly perform statistical analyses, tests were done to confirm that data reasonably conformed to normality (a bell-shaped distribution). The assumption of normal distribution was tested graphically using probability plots and box plots. On the whole, the dependent variables conformed to a normal distribution pattern.

Using more simplified regression models, detailed analyses of data were performed using a level of significance of 0.05. This level of significance can be interpreted to mean that the same result would be obtained 95 times out of 100 if the Wood5 sampling were done repeatedly. Analyses were broken down using the two species groups, aspen/birch and upland conifers. The short models shown in Table 1 are the result of numerous iterations using combinations of all the independent variables. Stepwise regression was used to reduce the data set down to those soil factors likely to be significantly effecting forest growth. The resulting simplified models were then re-tested using various combinations of categorical variables. These included indicators for soil great group, parent material, presence of a shallow water table, and landform. The continuous variables included in the "full" model statements were slope (in degrees), aspect, drainage class (numerical indicator), total depth of sampled profile (inches), percent coarse fragments (mean for the profile, weighted to double the importance of the top 30 inches), available water capacity (inches), and maximum observed depth of fine roots (inches).

Best Models

Using the data reduction steps mentioned above and further re-testing of all combinations of potentially significant soil factors, a number of statistically significant relationships were found. Based on analyses shown in Table 1, taxonomic groups differentiated only to the great group level (e.g., Eutroboralf, Dystrochrept) and do relate significantly to site index when the sample size is sufficiently large. These relationships are shown in detail in Table 1. The criteria for retaining the equations listed in Table 1 were: size of the standard error of the estimate, individual p-values for the variables (<0.05), R^2 for the model and p-value for the model. Equations are broken down by model form and species group with the actual coefficients and groups shown in Table 2. For example, using model 1, Mean Site Index = constant + (Soil Great Group) for species group 2 (upland conifers), the predicted site index for the Eutroboralf great group would be 61.7 (the constant) -5.349 (the coefficient for Eutroboralf from Table 2). Admittedly these tables are somewhat complicated but an attempt was made to include all significant relationships to show the reader all productivity relationships. The most immediately useful values in these tables are the model form, the value of the constant, the R^2, and the coefficients for each model in Table 2. Further in this report, the most significant relationships are presented in text form for simplicity.

Table 1
MODEL COEFFICIENTS AND STATISTICS

Spp. Group	Mean	Variable	Coefficient	p-value	MS	MSE	R^2	DF (error)	SE(est)[1]
1. Mean Site Index = constant + (Soil Great Group)									
All (N=91)	66.0	Constant	67.1						
		Soil Group	*	0.01					
		(F-test)[2]		0.01	317.8	112.2	0.11	84	10.6
2 (N=48)	60.2	Constant	61.7						
		Soil Group	*	0.003					
		(F-test)		0.003	244.1	60.5	0.28	41	7.8
2. Mean Site Index = constant+ (Soil Great Group) + Slope									
2 (N=48)	60.2	Constant	60.3						
		Soil Group	*	0.000					
		Slope	0.38	0.04					
		(F-test)		0.00	963.7	56.1	0.40	40	7.5
3. Mean Site Index = constant + Slope									
1 (N=40)	71.8	Constant	76.1	0.000					
		Slope	-1.34	0.007					
		(F-test)		0.007	721.2	90.0	0.15	38	9.5
4. Individual MAVI[3] = constant + %Coarse Fragments + AWC									
All (N=91)	4.1	Constant	5.30	0.00					
		% C.F.	-0.03	0.02					
		AWC	-0.12	0.02					
		(F-test)		0.01	9.46	2.05	0.07	88	1.4
2 (N=50)	4.3	Constant	5.43	0.00					
		% C.F.	-0.03	0.04					
		AWC	-0.12	0.07					
		(F-test)		0.04	7.1	1.99	0.09	47	1.4
5. Individual BAI[4] = constant + %Coarse Fragments + AWC									
All (N=92)	0.14	Constant	0.17	0.00					
		%C.F.	-0.001	0.02					
		AWC	-0.003	0.05					
		(F-test)		0.02	0.006	0.002	0.06	89	0.04
6. Standing Volume Index[5] = constant + drainage class + landform									
All (N=84)	282	Constant	66.0						
		Dr. Class	39.8	0.004					
		Landform	*	0.001					
		(F-test)		0.000	133137	14589	0.19	81	121
2 (N=48)	311	Constant	66.7						
		Dr. Class	43.7	0.009					
		Landform	*	0.025					
		(F-test)		0.008	68583	12900	0.19	31	114
7. Plot MAVI[6] = constant + (Shallow WT) + (Soil Great Group)									
All (N=88)	5.3	Constant	5.6						
		Sh. WT	*	0.008					
		Soil Group	*	0.002					
		(F-test)		0.001	20.9	5.51	0.21	80	2.35
2 (N=48)	5.2	Constant	5.2						
		Sh. WT	*	0.000					
		Soil Group	*	0.000					
		(F-test)		0.000	24.2	4.12	0.50	40	2.03
8. Plot MAVI = constant + (Shallow WT) + (Soil Great Group) + Dr. Class									
All (N=88)	5.3	Constant	3.4						
		Sh. WT	*	0.037					
		Soil Group	*	0.001					
		Dr. Class	0.52	0.08					
		(F-test)		0.001	20.5	5.36	0.23	79	2.32
2 (N=48)	5.2	Constant	1.2						
		Sh. WT	*	0.025					
		Soil Group	*	0.003					
		Dr. Class	0.88	0.05					
		(F-test)		0.001	17.6	3.82	0.35	39	1.78

[1] Standard error of estimate calculated by $\sqrt{MSE}$.

[2] F-test results for whole model or through testing of Anova parameters.

[3] Mean annual volume increment of sampled dom/codom trees (ft^3).

[4] Mean annual basal area increment of sampled dom/codom trees (ft^2).

[5] Index is plot basal area x mean height (ft^2).

[6] Plot volume index / mean age of sampled trees (ft^3).

Table 2
Coefficients for Effects in ANOVA Models

Eq. No.:	1		2	6		7		8	
Spp. Group:	All	2	2	All	2	All	2	All	2
I. Effect of Soil Group.									
Eutroboralf	-3.187	-5.349	-5.718			-0.428	-0.173	-0.486	-0.141
Glossoboralf	2.280	-0.094	0.409			-0.655	-0.180	-0.819	-0.295
Glossaqualf	-6.289	-0.889	0.118			-0.136	0.319	0.568	1.676
Ochraqualf	9.323	0.633	0.295			2.306	-0.985	3.114	0.561
Dystrochrept	-6.506	-5.891	-6.348			-1.508	-1.303	-1.953	-2.154
Eutrochrept	1.987	3.581	3.455			-1.215	-0.509	0.521	-0.919
Udipsamment	2.392	8.009	7.789			1.636	2.830	0.191	1.272
II. Effect of Landform Type.									
Hilly				-51.456	-44.885				
Level				51.456	44.885				
III. Effect of Shallow Water Table.									
Absent						3.465	6.271	2.790	4.239
Present						-3.465	-6.271	-2.790	-4.239

It should be noted that although 13 soil great groups were included in the original sample, it was necessary to eliminate nearly half due to an uneven sampling distribution that resulted in many of the groups being represented by only one or two plots. This process of elimination was also applied to most of the other categorical variables resulting in the use of landform as a predictor in equation (6), and presence of shallow water table in equations (7) and (8). Eventually, the number of great groups were reduced to four, Eutroboralf, Glossoboralf, Dystrochrept, and Udipsamment, for use in final analyses. Table 3 shows the soils series included in analyses and their associated great group.

There has been considerable discussion in the literature regarding the value of using volume production indices rather than site index to assess productivity (e.g., Schmoldt et al. 1985, Esu and Grigal 1979). For this reason several different measures were used. Site index does appear to correlate as well or better than volume-based measures to site factors in this study. Certainly the smallest prediction errors (12-16 percent) were obtained for the site index equations. The reason for this may be related to the attention given to site selection; only reasonably undisturbed, even-aged plots, with mature dominant tree cover were chosen for assessment. Site index is the most unbiased site quality indicator when the stands are undisturbed and even-aged. The volume data, by contrast, are rather approximate in nature; the individual mean basal area and volume increments were based on a non-random subsampling of trees in each plot.

Most of the variable coefficients obtained behaved fairly consistently in all the equations. The relationship of productivity as measured by mean volume and basal area increment (equations 4,5) to percent coarse fragments and available water capacity (AWC) is always negative. Coarse fragments may reduce rooting volume and possibly cause root damage due to frost action. The negative relationship of AWC to site index indicates that water limitations are not generally a problem in the county; drought stress is less likely to be a problem in this area than is poor drainage. The positive relationship between drainage class and productivity index (equations 6,8) supports that theory. The effect of slope on site index differed between species

Table 3.--Soil series classification and associated great groups.

Series	Great Group	Series	Great Group
Duluth Tax	Eutroboralf	Vermilion	Dystrochrept
Itasca	Eutroboralf	Eveleth	Dystrochrept
Taylor	Eutroboralf	Unnmd 1330	Dystrochrept
Ontonagon	Eutroboralf	Cloquet	Dystrochrept
Cutaway	Eutroboralf	Unnmd1335	Dystrochrept
Hibbing	Eutroboralf	Cromwell	Dystrochrept
Grygla	Eutroboralf	Unnmd 1334	Dystrochrept
Duluth	Eutroboralf	Toimi	Dystrochrept
Nashwauk	Glossoboralf	Ahmeek	Eutrochrept
Dusler	Glossaqualf	Menahga	Udipsamment
Indus	Ochraqualf	Sartell	Udipsamment
Unmd 1335	Dystrochrept	Redby	Udipsamment
		Hiwood	Udipsamment

groups; the coefficient is positive for upland conifers (equation 2) and negative for aspen (equation 3). The negative relation of slope to site index for aspen makes sense in the context of flat outwash, lakebed or floodplain sites being the richest. Upland conifers, of which the pines in particular, tend to be more productive on well drained, even xeric sites, including those with steeper slopes; hence the positive slope coefficient.

The differences in fit obtained between species groups are noteworthy. In only equation 3 was the fit for aspen statistically significant, with an R^2 of only 0.15. However, the fact that aspen site index relates better to slope than to any other soil factor seems to reflect its rather broad site requirements; more level terrain may imply a better water supply due to decreased runoff. The fact that AWC did not relate to aspen site index is not explainable using the existing data. Therefore it is difficult to make inferences about specific site requirements of aspen based on this one significant relationship.

Increased productivity among the upland conifers seems chiefly related to drainage and rooting volume; note the preference for Alfisols or Entisols over Inceptisols. Other illustrations of this tendency are the positive relationship of conifer productivity to slope and drainage class, and the negative relationship to percent coarse fragments and AWC. The seven great groups and their predicted conifer site index are shown in Figure 1.

Based on the significant relationship of soil Great Group to conifer site index, further analyses were done to assess these relationships in more detail. Four of the great groups having the highest number of sampling plots were used including Eutroboralfs (29 plots), Glossoboralfs (11 plots-all Nashwauk series), Dystrochrepts (26 plots), and Udipsamments (15 plots). This subset of data was then re-analyzed to determine if the accuracy of equations could be enhanced by reduction of the data into fewer groups having a greater number of cases in each group.

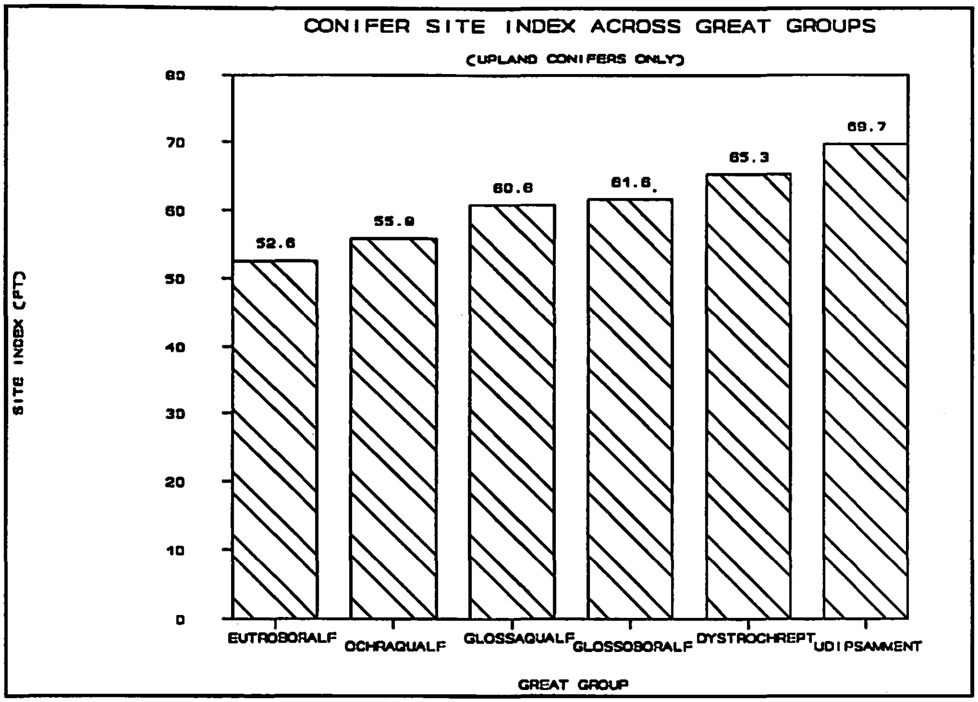

Figure 1.--Conifer site index across great groups.

Analyses of the aspen/birch group showed the same low correlations as previously found using the entire data set. The relationship between slope and aspen site index was nearly identical to previous analyses with a coefficient for slope of -0.126. The equation for site index across all sites is as follows:

Site Index = 77.575 - (1.256 X slope).

For example, a site having a slope of 10 percent would be reduced by approximately 12.5 feet down to a site index of 65. This relationship explained approximately 15 percent of the variation in site index (R^2=0.15), still too low to be considered of practical value. Also, this effect of lower slope may relate more to the fact that higher site indexes are usually found on the flatter, lacustrine soils. Again, the data set was not large enough to fully explore these relationships.

The lack of fit for aspen could be due to two primary factors:

1. The relatively small sample size in the data set as a whole (49 total aspen plots) or,

2. The sampling scheme is not incorporating those attributes about forest soils which are the primary drivers of productivity.

A group of these variables are likely chemical in nature. Based on research reported by Pastor et al. (1985) in the Lake States, nitrogen mineralization rates have been shown to be directly related to forest production rates. Stand history, and as a result, forest floor organic matter composition, will exert a strong influence on nitrogen availability. This is thought to be due to chemical differences between foliage of conifers and hardwoods. In general, hardwood leaf litter is more easily decomposed and will release nitrogen at a higher rate than conifer litter. High concentrations of resins and tannins in conifer litter has been shown to depress nitrogen availability. Attempts were made to approximate nitrogen supplying capacity of the soils by using the thickness of the O and O + A horizons but no significant correlations were found.

As was the case with the full data set, analyses of the reduced conifer data (four of the larger great groups) produced higher correlations of soil factors to forest production than the aspen data set. Specifically, soil great group, drainage class, slope, and landform, were shown to be significantly affecting conifer site index. Analyses of the four great groups combined showed a significant relationship between great group, slope (%), drainage class (1=very poor to 7=excessive), and landform (1=rolling and hilly plains, plateaus and uplands, 2=level and undulating uplands, plains, plateaus). The equations for conifer site index are as follows:

if Eutroboralf: Site Index (ft) = 19.5 - (2.35) + (SLOPE(%) X 0.356) + (DRAINAGE CLASS (1 to 7) X 5.27) + (LANDFORM (1 or 2) X 6.852)

if Glossoboralf: Site Index (ft) = 19.5 + (5.23) + (SLOPE X 0.356) + (DRAINAGE CLASS (1 to 7) X 5.27) + (LANDFORM (1 or 2) X 6.852)

if Dystrochrept: Site Index (ft) = 19.5 - (7.24) + (SLOPE X 0.356) + (DRAINAGE CLASS (1 to 7) X 5.27) + (LANDFORM (1 or 2) X 6.852)

if Udipsamment: Site Index (ft) = 19.5 + (4.35) + (SLOPE X 0.356) + (DRAINAGE CLASS (1 to 7) X 5.27) + (LANDFORM (1 or 2) X 6.852)

The above relationship explained approximately 68 percent of the variation in site index across the four great groups represented. Although this equation was among the highest found in the entire data set, some implications of this relationship are difficult to explain. Specifically, the increasing site index going from the "rolling and hilly plain" landform to a "level and undulating uplands, plains or plateau" landform. The fact that slope remains as a significant site factor, after landform is included seems to be contradictory. However, this relationship has immediate value in assessing conifer site index. In addition, this relationship of great group, slope, drainage class, and landform, indicates that increased attention should be paid to these factors in future data collection efforts.

Analysis of the conifer data set using the presence or absence of a shallow water table data showed that the presence of a shallow water table may decrease site index by approximately 16 feet. However, the data on shallow water table was extremely sparse (5 plots) and could not be included in the other analyses.

Further analysis of the Udipsamment great group (containing Hiwood, Menahga, Redby and Sartell series) showed that drainage class had a significant positive effect on site index accounting for 73 percent (R^2=0.73) of the variation in site index (the strongest relationship in the entire data set). This great group is dominated by jack pine plots (10 plots out of 12 total plots, the remainder red pine). The equation to predict site index within the Udipsamments is as follows:

Site Index (ft)= 29.385+(DRAINAGE CLASS X 6.8)

Figure 2 shows this relationship graphically with the predicted site index line and the actual data points.

RECOMMENDATIONS

The practical application of soil survey information to assess forest productivity will likely fall along two primary lines:

1. Soils data will provide a means to predict stand site index in cases where direct measurement has not been done or is likely to be inaccurate (high-graded stands, young stands or those not yet visited by forest inventory crews).
2. Soils data will be used to aid in the decision-making process in plantation establishment on recently harvested or old field sites.

Ultimately, the soils mapping unit boundaries will be digitized and overlaid on the tree stand boundary coverage in REMIS to derive a soils/forest stand map. This map will be useful to predict site index across the county and project stand yields and harvests. Also, soils data may be used to assess site operability and formulate recommendations for season of harvest.

Given these basic objectives and results of Wood5 data analysis, some recommendations for future sampling can be made. First, problems in the present Wood5 data set will be briefly discussed.

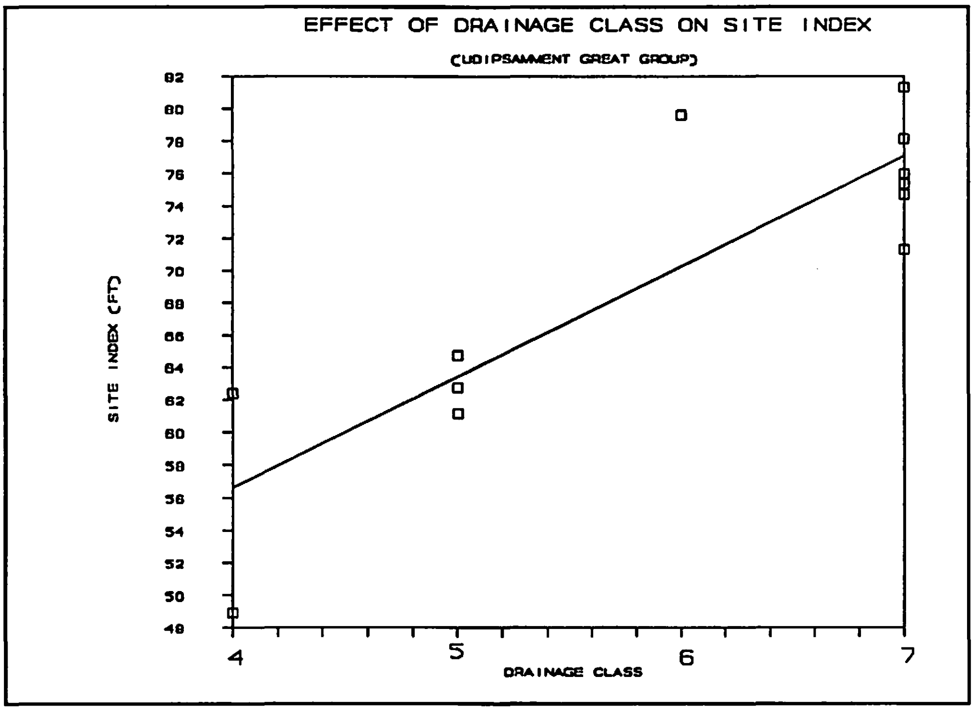

Figure 2.--Effect of drainage class on site index.

PROBLEMS IN THE DATA

Most of the problems in the data set have already been touched upon. The primary problems in the data are the relatively low number of plots and the inconsistency of data collection from plot to plot. The incompleteness of the data collection often resulted in the elimination of a plot from the analysis, further reducing an already marginal sample size. On the other hand, much superfluous data such as understory composition and cover density was included with the Wood5 plot forms. Unless some other use is envisioned for these data, it appears the collection of some data could be reduced while more critical factors should be measured more consistently (slope, presence/absence of shallow water table, etc.).

The relatively low number of plots for each species/series group is an evident problem. The intensity of sampling required will depend on the specific end-use of the soils data. If management decisions must be made at the series and species level, data collection efforts must be increased significantly. After analyzing the data, it appears that managing at the series or mapping unit level for aspen/birch and the upland conifer species groups is more realistic.

Based on discussions with county personnel, indications are that the use of soils data will be for management at the soil series or mapping unit level. Because of this, a calculation of the number of plots needed to accurately calculate site index for the species group at the series level was done. These calculations are based on a 90 percent Confidence Limit for a range +/- 5 percent of the mean site index. This means, for example, in the case of the Dusler series for the aspen/birch group (species group 1), that 90 times out of 100, the average site index would fall between 73.1 +/- 3.65 feet if sampling were done repeatedly. In the case of the Dusler series for the aspen/birch species group, a total of 27 plots are needed to achieve this accuracy, 25 more than are presently in the database. Those plots marked with asterisks in the "Number of Plots Needed" category in Table 4 indicate that only one plot was present for that series/species group combination. As a result, the mean has no variance and the "Number of Plots Needed" cannot be calculated.

RECOMMENDATIONS FOR FUTURE SAMPLING

As the current Wood5 data collection procedure and analysis is time-consuming and produced little readily usable information for practical forest management, attention should be paid to future sampling. Obviously, the current Wood5 data has value in that it has allowed the construction of some useful relationships, particularly for upland conifers and provided data to guide the design of a future sampling plan. Using Table 4, an average of 12 plots per series/species group combination are required. Assuming all series are of equal acreage distribution across county ownership and therefore of equal priority for sampling (this is likely not the case), a total of 396 plots or approximately 300 additional plots would be required to quantify site index to the accuracy desired. Obviously, if the number of plots suggested in Table 4 are to be collected, data collection procedures must be streamlined and series for sampling must be prioritized. Based on the analysis of the data, it appears that a considerable amount of streamlining in the data collection procedure is possible, concentrating on the consistent collection of visually discernible site features.

Table 4.--Site index (SI) by soil series and species group with number of sampling plots needed to achieve a 90 percent confidence interval (CI) on 5 percent of mean site index.

Soil Series	Avg SI	Current Number of Plots	Number of Plots Needed
		FOR ASPEN/BIRCH SPECIES GROUP	
Ahmeek	84.8	1	**
Duluth	81.5	2	3
Dusler	73.1	2	27
Eveleth	52.6	1	**
Grygla	76.2	1	**
Hibbing	68.7	5	28
Hiwood	61.0	2	6
Indus	74.7	4	10
Nashwauk	74.2	6	20
Ontonogan	57.2	2	17
Redby	85.0	1	**
Taylor	79.8	2	5
Toimi	72.0	5	11
Unmd 1323	64.9	3	5
		FOR UPLAND CONIFERS	
Ahmeek	65.3	1	**
Cloquet	52.1	6	6
Cromwell	67.5	1	**
Cutaway	59.2	1	**
Duluth	55.1	2	2
Duluth Tax	87.6	1	**
Dusler	60.8	2	6
Emmert	48.0	1	**
Hibbing	57.8	8	29
Hiwood	62.4	1	**
Menahga	75.1	5	5
Nashwauk	61.6	3	8
Ontonogan	49.9	1	**
Redby	60.4	3	23
Sartell	67.1	4	15
Taylor	64.0	2	43
Toimi	55.7	5	9
Unmd 1323	58.3	3	9
Vermillion	56.3	2	3

The first issue in designing a modified sampling scheme is plot selection criteria. The current method used to select Wood5 plots requires that an even-aged, fully stocked stand be selected on a site representative of the series. These relatively stringent criteria prevent forest production sampling from being done at the same time the survey crew is in the field mapping soils and prevent the use of most of the existing forest inventory data. If stand site index is the only production-related variable of interest, a fully stocked stand is not necessary as site index is independent of stocking. If the intent of the Wood5 program is to identify "maximum productivity stands" for sampling, site index appears to be the most useful variable to measure assuming stand history includes no drastic site disturbance. Also, site index is generally assumed to be relatively independent of stocking density. Based on our results, productivity relationships to site index seem to be stronger compared to other volume-based indices in assessing potential site productivity. For this reason, future stand sampling should concentrate less on locating fully stocked stands. Obviously stands should be selected that appear to be healthy and relatively undisturbed. These relaxed site-selection criteria allow the possibility for the survey crew to collect data while mapping without selecting the Wood5 plots as a separate step. Also, the possibility to use current forest inventory information exists assuming those stands that have been high-graded or are very poorly stocked are excluded. This option will be discussed in more detail at the end of the report.

An additional concern regarding the Wood5 program is the selection of sites representative of the series. Site selection should be done with enough plots to characterize both those sites representative of the series as well as those that appear to be on the extreme ends of the series, again, being careful not to select stands suspected to have a history of disturbance. In this way, the variation in site index as well as the average site index within a series can be calculated. Individual plot site index is more likely to fall near the mean site index for the series, however, assuming site index is normally distributed within the series. Sampling sites could be selected using existing soils maps or at the same time soils are being mapped. In the case of mapping units or complexes which have significant inclusions of series within the mapping units, care should be taken to ensure that sampling sites are located on the correct series. By keeping the sampling at the series level, site indices for larger mapping units and complexes can be calculated using the weighted average site index (weighted by acreage) for the mapping unit or complex.

After relaxing site selection criteria to allow more cost-effective location of sampling sites, streamlining of the data collection procedures can be done. Based on analysis of the Wood5 data, those variables having the greatest effect on forest production are **great group** (to be later substituted for by soil series in future sampling), **slope, drainage class, landform** and the **presence/absence of a shallow water table**. These site factors are visually discernible and do not require a detailed profile description or minimal profile characterization. Because of this, the maximum benefit per unit of sampling effort expended would be obtained by concentrating on the consistent and accurate data collection of these site factors. Detailed soil profile descriptions should be done only when absolutely necessary. Also, detailed data on understory species should not be collected unless some specific future use is intended (site classification). In sum, the data collection effort should focus on the ultimate goal of the project and compile the relevant data in a consistent and useable form.

The detailed sampling plan should be developed to prioritize sampling and statistically monitor progress. Those series projected to be more intensively managed in the future should receive the highest priority for sampling. Once priorities are defined, sampling sites should be located on all landforms on which the series occur.

If chemical characteristics are to be incorporated into forest management in the future, a system of using foliar concentrations of nutrients on a site-specific basis may be the most promising. Based on preliminary forest fertilization trials on aspen in St. Louis County (to be reported) and results published in the literature, the potential exists to increase yields by 15 to 30 percent. The recommended method to collect data would be to stratify by soil series and collect foliage samples for analysis. The quantification of chemical characteristics across all series is likely to be prohibitively expensive and a more focussed program on a few of the most extensive series is recommended. Based on preliminary cost estimates for forest fertilization, an aspen stumpage price of approximately $20/cord would be needed to break even. However, integration of forest fertilization with other operations such as thinning may be cost-effective in the future.

Pilot Project

Based on the above analyses and discussions with Land Department personnel, a pilot data collection project is proposed. This project makes use of previously collected forest inventory information and the concentration of data collection efforts on visually discernible site factors (slope, drainage class, landform, and the presence/absence of a shallow water table).

A crucial factor in this pilot project is the status and detail of the existing forest inventory. This existing forest inventory includes cover type, stand size, stand density, high-grading notes, and a measurement of height and age for one site index tree per plot. Reservations in using this data for soils/forest production work centered on the re-location of the plot and the potential for inaccurate measurement of site index due to one measurement. After discussion with Land Department personnel, it was felt that general relocation of the plot (plus/minus 200 feet) is possible through the use of aerial photographs.

Analyses were done using Phase I data for St. Louis County to determine the potential loss of accuracy in estimating site index based on only one site index tree per plot. The Phase I data contains three site index trees per plot. Site index information for 1000 plots in St. Louis County were stripped off of the state-wide Phase I (1977) data and the average site index for each plot determined by calculating the site index for each site index tree individually and averaging these values. Then one site index tree per plot was randomly selected. Regression analyses were done to determine the accuracy lost from using one site index tree per plot. Across the sixteen species represented, the average loss of accuracy (reduction in R^2) was approximately 9.5 percent with the average R^2 being 0.905. Based on these analyses, relatively accurate determinations of site index can be done using the existing forest inventory data.

Given the above information, a pilot-scale study could begin concentrating on a limited number of species (aspen, and/or a conifer) on a particular soil series (possibly Hibbing or Toimi). The purpose of this project would be to determine the potential to develop accurate productivity ratings for soil series in a cost-effective manner. As existing forest production data will be used and relatively little soil profile description done, it should be possible to collect enough plots to allow characterization of a variety of species/soil series combinations.

In addition to site index information, basal area data should be recorded at each plot. Given that the majority of time involved in revisiting the sites will likely be location and travel to the site, the additional investment in time necessary to collect basal area information is expected to be minimal. This could be done using a variable radius plot with relatively little additional time. Basal area data will be used to approximate stand volume. Using the age determinations existing in the current forest inventory, mean annual volume increment could be calculated. Mean annual volume increment could then be used as an additional productivity indicator to supplement site index.

During the process of plot location and data collection, records should be kept to track the amount of time taken to locate the plot and collect the soils/site data. Assuming the pilot-scale project is a success, this information could be used to estimate the cost of an expanded sampling program. During the pilot-scale phase, data should be immediately entered into computer format and the status of each species/series combination site index estimate statistically updated. These programs could be provided in Lotus 123 or stand-alone programs by NRRI or others. By simplifying the data collection procedure and the number of variables to be collected, it will be possible to keep a running status of the sampling program during the summer months. This continually-updated sampling method will provide immediate feedback to field personnel and notify the field staff when the desired accuracy for a specific species/soil series has been achieved.

CONCLUSIONS

The analysis of the Wood5 data has produced some practically useful relationships for upland conifers and, perhaps more importantly, helped to identify those variables likely to have the greatest impact on forest production in the county. The sampling method outlined above is designed to maximize the potential to accurately calculate site index while, at the same time, being cost-effective enough to allow an adequate number of plots to be measured. Discussions should take place within the Land Department and with other soil survey staff to refine the above outline into a workable plan to ensure that the objective of developing soils-based forest productivity indices are met. An expansion of a sampling program to describe chemical characteristics is likely to be most cost-effective using foliar analyses within predominant series. A detailed evaluation of future data collection procedures and the time needed for each step should be done and the relative value of data collected in each step weighed. In summary, the primary direction of sampling should be to streamline the data collection procedure to allow the sampling of more plots through the collection of less data per plot to allow a more accurate determination of productivity for each species/soil series combination.

REFERENCES

Esu, I.E., and D.F. Grigal. 1979. Productivity of quaking aspen (P. tremuloides Mich.x) as related to soil mapping units in northern Minnesota. Soil Science Soc. Am. J. 43: 1189-1192.

Laidly, P.R. 1979. Metric site index curves for aspen, birch and conifer in the Lake states, USDA Forest Service. Gen. Tech. Rep nc-54. US Dept. of Agric, For. Serv., North Cent. For Exp. Stn. St. Paul, MN. 15 p.

Pastor, J., J.D. Aber, C.A. McClaugerty, and J.M. Melillo. 1985. Aboveground Production and N and P cycling along a nitrogen mineralization gradient on Blackhawk Island, Wisconsin. Ecology 65(1): 256-268.

Schmoldt, D.L., G.L. Martin, and J.G. Bockheim. 1985. Yield-based measures of northern hardwood site quality and their correlation with soil-site factors. Forest Science. 31: 209-219.

Wilkinson, L. 1986. SYSTAT; The System for Statistics. Systat, Inc., Evanston, IL.

APPENDIX 1. STRUCTURE OF SOILS.DB DATABASE

(matching soils data for Wood5 plots)

Field Name	Data Type	Field Description
soilid	TEXT	State ID No.
horiznum	INTEGER	No. of horizon.
drycons	TEXT	Dry consistence.
moicons	TEXT	Moist consistence.
updepth	INTEGER	Upper limit of horiz.
lodepth	INTEGER	Lower limit of horiz.
fhorizon	TEXT	Horizon label (e.g., Oa)
moicolor	TEXT	Moist color.
motabb	TEXT	Abundance of mottling.
textmod	TEXT	Texture modifier.
ftexture	TEXT	Texture (e.g., SL)
grade1	TEXT	Describes structural elements.
size1	TEXT	Describes structural elements.
shape1	TEXT	Describes structural elements.
ph	REAL	pH
roots	TEXT	Letter codes for root abund.
cfrag1	INTEGER	Small coarse fragments.
cfrag2	INTEGER	Large coarse fragments.
pedseres	TEXT	Soil series.
cfc	INTEGER	Cfc number (refers to Lab No.)
pederosn	INTEGER	Erosion grade.
landfrm	TEXT	Landform code.
pm1	TEXT	Primary parent material.
pm2	TEXT	Secondary parent material.
location	TEXT	Detailed location description.
class	TEXT	Taxonomic classification.
notes	NOTE	Field notes.
legal	TEXT	Legal location (township/range).
soilno	TEXT	Soil number.
LNDFRM	TEXT	Landform code.
SOILSER	TEXT	Series name.
SLOPE	INTEGER	Slope %.
ASP	INTEGER	Aspect-degrees.
LFM	TEXT	Landform.
MICROL	TEXT	Microrelief code.
REMARKS	NOTE	Plot notes.
BA	INTEGER	Basal area (prism plot).
CDEN	INTEGER	Overstory cover density %.
DRCL	TEXT	Drainage class.
awi	REAL	Av. water cap. in./in. soil.
hthick	REAL	Horizon thickness.
awh	REAL	In. of AWC for each horizon.

tcfrag	REAL	Total % coarse fragment per horizon.
depthtot	INTEGER	Total sampling depth of profile.
cumpctcf	REAL	Cumulative % coarse frag.s.
wtpctcf	REAL	% C.frag weighted (x2) top 30 in
avwat	REAL	Total AWC for profile.
adawc	REAL	AWC adjusted for % coarse fragments.
DepFR	INTEGER	Max. depth of fine roots (inches).
DepMR	INTEGER	Max. depth of medium roots. (in.).
rdepth	INTEGER	Depth to resistant layer (BD>1.8).
athick	REAL	Thickness of O + A horizon (in.)

APPENDIX 2. STRUCTURE FOR SITEPRD2.DB DATABASE

(Wood5 plots)

Field Name	Data Type	Field Description
SOILNO	TEXT	Wood5 soil number.
SOILSER	TEXT	Soil series name.
TSPECIES	TEXT	Species code.
DBH	REAL	Diam. breast ht. (in.).
tage	INTEGER	Tree age.
HT	INTEGER	Tree height (ft.).
LNDFRM	TEXT	Landform code.
basar	REAL	Basal area (sq. ft.)
cfvol	REAL	Total cubic foot volume.
mbai	REAL	Mean Plot BAI.
mvi	REAL	Mean Plot vol. inc.
BA	INTEGER	Prism plot BA (ft sq.)
CDEN	INTEGER	Overstory cover density.
aveht	REAL	Mean height for plot (ft).
plbaxht	REAL	Plot vol. index: BA x Mean Ht.
sampbai	REAL	Mean BAI of sampled trees.
sampvi	REAL	Mean Vol. inc. of sampled trees.
sifm	REAL	Individual site index.

Current number of rows: 575