DOCUMENTATION FOR THE

DATA FORMATS AND

PROGRAMS USED FOR

COMPUTER ASSISTED

GLACIAL STRATIGRAPHY

AT THE

NORTH DAKOTA GEOLOGICAL SURVEY

BY

KENNETH L. HARRIS

DECEMBER 1987

NORTH DAKOTA GEOLOGICAL SURVEY

GRAND FORKS, NORTH DAKOTA

OPEN FILE REPORT OF-87-2

INTRODUCTION

For about the past year, we have been working on an IBM PC-based system to streamline the handling, interpretation and correlation of our nearsurface stratigraphic data. It is used by our geologists as a source of nearsurface information for various projects, ranging from environmental evaluations of landfills to glacial stratigraphy. Steve Moran (a former NDGS geologist) started this database in the early 1970's to manage his stratigraphic data. Steve's system has been expanded and modified, and it now contains about 9,000 entries. The information in storage was generated, over the years, by Quaternary geologists working in the state. We use an IBM System 34 as a data-storage library. Project files are transferred from it to IBM PC-compatible diskettes for processing at individual work stations.

We have a number of programs to aid in manipulating these project files once they are on diskette. Subfiles can be created that consist of data generated by individual workers or individual projects; or they may contain data sorted by data type or various geographical criteria. Subfiles are routinely created to determine the well control available in a project area; or to identify and locate samples that are stored in the Wilson M. Laird Core and Sample Library.

EQUIPMENT USED

The desk-top computer that is used with these programs is an IBM PC-XT (640K RAM, AST-SIXPAC, no graphics). The printer used is an EPSON FX-85. Some printer commands are written into some programs. DISCORT.BAS uses the compressed command for making barcharts and the graphics programs use the form feed command. There may be some others used that I don't remember. You may have to change some of these commands to make things work on your system.

THE CORRELATION PARAMETERS

In order to identify different units of glacial sediment, we have to identify some characteristics, or combination of characteristics, of the glacial sediment that are internally consistent; but which differ from unit to unit. We have found six correlation parameters to be useful for correlating glacial sediment in North Dakota. These parameters are based on the textural analysis of the sediment and the lithologic analysis of the coarse-sand fraction (1-2 mm). The textural analysis gives us the percentage composition of sand (SD), silt (SL), and clay (CL). The lithologic analysis of the coarse-sand fraction gives us the percentage composition of crystalline and metamorphic rock fragments (XT), limestone and dolostone rock fragments (CO), and shale rock fragments (SH). In order to simplify data manipulation and display, these six parameters are reduced to four. The silt and clay textural parameters are normalized :

NS=((SL/(SL+CL))*100);

and the crystalline and carbonate lithologic parameters are normalized :

NX=((XT/(XT+CO))*100).

The correlation parameters that are used for the graphical analysis and correlation of glacial sediment are :

1) the percentage content of sand (SD),
2) the percentage content of normalized silt (NS),
3) the percentage content of normalized crystalline and metamorphic rock fragments (NX),
4) the percentage content of shale rock fragments (SH).

DEFINING A TILL

Two programs we developed this past year have proved to be useful tools in defining lithostratigraphic units (tills) in the glacial sediment, and then correlating these tills between work areas (Figure 1). The first one (TILSRCH) will identify, and define modal tendencies in a dataset (tills). The second one (DISCORT) will display, compare, and assist in the correlation of the definitions (tills) developed by TILSRCH. I will briefly describe these two programs and show you how they are used.

TILSRCH (AS-TS.BAS):

Modal groups in a dataset can be located by interpreting crossplots of the correlation parameters in the dataset (Figure 2). These coordinates are then used as input for TILSRCH. An interactive, iterative process is used to refine the definition of a modal group (a till), retrieve all of its member-samples, and statistically evaluate the quality of that definition. This procedure is repeated as often as necessary, until a good definition of the modal group is achieved. When a satisfactory definition is achieved, the program will label all member-samples of that group in the dataset. This procedure is repeated until all modal groups in the dataset have been defined and labeled. The output of the program can take three forms:

1) A list of the member-samples of a specific modal group (a till),
2) A list of the entire dataset identifying all member-samples of all modal groups (the correlated dataset),
3) A list of all of those samples that do not meet the definition of any modal group.

The correlated dataset lists all testholes in order of their location, and all samples at each location in order of increasing depth (Figure 3). This provides the geologist with a mappable summary of the correlated data, which can then be interpreted stratigraphically.

DISCORT (DISCORT.BAS):

This program displays the definitions developed in TILSRCH graphically and checks for overlap between the definitions. The input required is the value of the correlation parameters for each definition (till), and the number of member-samples in that till. This has proved to be a very useful program and can be used in at least three ways.

1) It will display the definitions as bar-charts (Figure 4). It can be instructed to display these charts using various arrangements of the correlation parameters in ascending or descending order.
2) It can perform a check of the quality of the definitions we developed in the dataset. Each definition can be checked against each other definition to find the degree of overlap that exists. The level of precision of that check (allowable variation between compared parameters) can be specified, so that a judgement can be made on the quality of the definitions. In this case we do not want the definitions to overlap. So, the poorer the indicated correlation is, the better our definitions are. The results of the comparison are shown in Figure 4.
3) It can be used as a correlation tool. Each definition in a dataset can be compared with each definition in another dataset or datasets. The level of precision of the comparison can be specified, and the quality of the indicated correlations can be judged. In this case we are looking for overlap between compared definitions. So, the better the indicated comparison is, the more likely we are to have a good correlation. The results of the comparisons are shown in Figure 4.

DOCUMENTATION

FORMATS

There are three data formats (ND, AS, and GF) that are used. That's terrible I know, but that's the way things evolved and I haven't taken the time to redo them. A description of each variable in each format follows. The information shown inside the square brackets indicates the format of the data element(s) and the location of the data element in the record. Alphanumerical variables are indicated by "A",and integer variables by "I". A label of "2A6" means two alphanumerical variables each with a field of six characters.

ND FORMAT

This is the format used in the main, near-surface dataset that is maintained on our IBM System 34. All ND-formatted files are identified as "ND-XXXXX.XXX" in the file name. The formats and fields indicated here are precisely what you should find. These records are treated as 82-character, string variables.

System 34 variables: Unknown [?;1-6]
Well Number: A unique, sequential file number assigned to each well or sample site [I6; 7-12]

Interval: The sample collection interval [I1; 13]
Depth: The depth of the sample or well as appropriate [I4; 14-17]
T-R-S-QQQ-County: The location of the sample site [2I3, I2, A3, I2; 18-30]
Elevation: The elevation of the sample site or samples [I4; 31-34]
SD-SL-CL: The results of the textural analysis (integer percent) [3I2; 35-43]
XT-CO-SH: The results of the lithologic analysis of the coarse sand fraction (integer percent) [3I2; 44-52]
Who: The name of the geologist or agency who collected these samples [A3; 53-55]
ID1-ID2: A project identification code that might have been used by the geologist [2A3; 56-61]
SS-AA-BB-CC-DD-EE: These are codes that identify the samples as being from subsurface or surface sections, collected using a certain type of drill rig, and other noteworthy tidbits. [2I1, 2I2, I1, I2; 62-70]
Map-Row-Column: Identifies the sample site as being located on a particular Atlas Series map sheet and a 7 1/2' topographic map sheet [I2, A1, I3; 71-76]
Box: The storage location of the samples in the core and sample library [I6; 77-82]

AS FORMAT

This is the Atlas Series data format used with all of the work programs except the graphics programs. These data consist of selected elements from the ND-formatted data. All AS-formatted files are identified as "AS-XXXXX.XXX" in the file name. The formats indicated are maximum fields for the variables. In the conversion to AS format, the data elements are read as string (alphanumeric) variables and converted to integer (whole number) variables, if appropriate, before being rewritten to disk. Consequently there is some compression of the integer variables (005467 -> 5467). Variables identified as XXX$ are string variables and those labeled XXX% are integer variables. The maximum record length is 55 characters.

QQQ$: The quarter-quarter-quarter location of the sample site [A3; 1-3]
WHO$: The initials of the geologist or agency who collected the samples [A3; 4-6]
ID$: A project identification code that might have been used by the geologist who collected the samples [A6; 7-12]
N%: A unique, sequential file number assigned to each well or sample site [I6; 13-18]
TWN%-RNG%-SEC%: The location of the well or sample site [2I3, I2; 19-26]
ELEV%: The elevation of the sample site or of individual samples [I4; 27-30]
DEPTH%: The depth of well or sample [I4; 31-34]
SD%-SL%-CL%: The results of the textural analysis (integer percent) [3I2; 35-40]
XT%-CO%-SH%: The results of the lithologic analysis of the coarse sand fraction (integer percent) [3I2; 41-46]
NS%-NX%: Normalized SL [((SL/(SL+CL))*100)] and normalized XT [((XT/(XT+CO))*100)] (integer percent) [2I2; 47-50]
SS%: A code that identifies the samples as being from a subsurface (0) or surface section (1) [I1; 51]

ASM%: The Atlas Series map on which the sample site is located [I2; 52-53]
LIG%: The integer percent Lignite content of the sample [I2; 54-55]

GF FORMAT

This is the format that is used by the graphics programs. The programs that read the GF-formatted data can be modified to read the AS-formatted data directly, but I haven't taken the time to modify them. These records are eight characters long.

SD%: Integer percentage content of sand [I2; 1-2]
NS%: Integer percentage normalized SL [I2; 3-4]
NX%: Integer percentage normalized XT [I2; 5-6]
SH%: Integer percentage content of shale [I2; 7-8]

PROGRAMS

The programs that are on your diskette are introduced here. All of these programs are considered developmental prototypes. There certainly is room for improvement in terms of operational speed and programming style. The author would appreciate hearing of your ideas for improving programming style or operational procedures. The discussion of the individual programs is organized according to the format of the data that they read. All programs are written in IBM-PC BASIC, version A3.10.

ND FORMAT

These programs all read the ND-formatted data. This is the format used for data storage on the IBM System 34. It is also the format used to transfer data from the IBM System 34 to IBM PC compatible diskettes. These programs require you to specify the name of the file you are reading and any file you may want to create. All files written are AS-formatted files and should be given names of the form "AS-XXXXX.XXX".

ND-FILES.BAS: This program inventories the ND-formatted data that has been transferred from the IBM System 34 to an IBM PC-compatible diskette. The output is a list of all of the subfiles present. The list shows the status of all of the data in the subfiles. These subfiles are defined as all the data listed under a specific geologist's or agency's code (WHO$). The data in each subfile is listed in one of three categories:

1) location only, no data,
2) uncoded data that exists in our back-up files,
3) numerical data that is in the subfile.

The data in each category is subdivided into textural and coarse sand data-types and the total number of entries in each subfile and category is shown. The output can be routed to screen or printer.

ND-WHOSE.BAS: This program generates a subfile consisting of all samples collected by a selected geologist or agency (WHO$). The output can be routed to screen, printer, or disk.

ND-CODE.BAS: This program generates a subfile of all entries that are indicated as having uncoded data available in our back-up files. The output can be routed to screen, printer, or disk.

ND-NUM.BAS: This program generates a subfile of all entries that have numerical data available for processing. The output can be routed to screen, printer, or disk.

AS-FORMAT

These programs all read AS-formatted data. They require you to specify the name of the file you are reading and any files you may want to write. All files written in AS format should be labeled "AS-XXXXX.XXX".

AS-LST.BAS: This program lists the contents of an AS-formatted file to screen or printer.

AS-CODE.BAS: This program generates a list of all file entries that contain uncoded data that is available in our back-up files. It will write this list to the screen, printer, or disk.

AS-NUM.BAS: This program lists all file entries that contain numerical data that is available for processing. It will write this list to screen, printer, or disk.

AS-WHOSE.BAS: This program generates a file consisting of all of the sample data collected by a specific geologist or agency (WHO$). It will write this file to screen, printer, or disk.

AS-BY-N.BAS: This program rewrites a file in ascending order of N-number (well number). It currently uses a really slow sorting routine that needs to be rewritten. The output can be routed to screen, printer, or disk.

AS-T&R.BAS: This program rewrites a file in descending order of township and range numbers. It currently uses a really slow sorting routine that needs to be rewritten. The output can be routed to screen, printer, or disk.

AS-GF.BAS: This program reads a file and rewrites it as a GF-formatted file. The output can be routed to the screen, printer, or disk.

AS-TS.BAS : This program (till search, TILSRCH) reads a data file, searches it for all samples that meet a user-defined definition, and labels all of those samples. A definition consists of specific values of the four correlation parameters (SD, NS, NX, and SH) and an allowable range of variation of these parameters. The definition that is used is developed by using cross-plots (GF-XPLT.BAS) of the dataset to locate and define modal tendencies in the data. An interactive, iterative procedure is used to refine the definition. Once an acceptable definition is achieved, the program will label all those samples in the dataset that meet that definition. This process is repeated until all identifiable modal groups in the dataset have been defined and labeled. This results in a dataset with all member-samples of all identifiable modal groups labeled. The output of the program is:

1) lists of member-samples of specific modal groups (the till datasets)
2) a list of the entire dataset identifying all member-samples in all modal groups (the correlated dataset)

3) a list of all those samples that do not meet the definitions of any of the modal groups (the residual dataset). This is the main program used to differentiate the various till units. The till dataset and the residual dataset can be written to screen, printer, or disk. The correlated dataset is written to disk only, in a modified AS format. This is because the definition labels add another variable. The correlated dataset can only be read by ASCORLST.BAS (see below).

ASCORLST.BAS: This program will read only the correlated datasets generated by AS-TS.BAS and rewrite them to the screen, printer, or disk.

GF-FORMAT

These programs read the GF-formatted data and generate graphical summaries of it. All files written in GF format should be labeled "GF-XXXXX.XXX".

GF-HIST.BAS: This program will generate histograms of each of the four correlation parameters (SD, NS, NX, and SH). The histograms show the variation in frequency of occurrence vs percentage content for each of the four correlation parameters. The output is routed to the printer and presented as text graphics. No special graphics equipment is required.

GF-XPLT.BAS: This program will generate crossplots of the correlation parameters. It presents crossplots of NS vs SD, NX vs SH, and SH vs SD as a standard group of correlation graphs, or it will graph any selected combination of the correlation parameters. The crossplots show coincident values of the percentage content of the compared parameters and tally the level of that coincidence. The number of coincident values is tallied up to 15 (1-9 + a-f). Levels of coincidence above 15 are indicated by an "*". The resulting graphs can be contoured to clearly show modal tendencies in the parameters being compared. A comparison of all of the crossplots of all of the correlation parameters can uniquely identify modal groups in the dataset. The output is routed to the printer and presented as text graphics. No special graphics equipment is required.

KEYBOARD INPUT

DISCORT.BAS: This program displays the definitions developed in AS-TS.BAS graphically and checks for overlap between the definitions. The prompted, keyboard input consists of the mean values for SD, NS, NX, and SH plus N (the number of samples in the modal group) for each definition in a dataset or datasets. This program has proved to be a very valuable tool for developing good definitions in a dataset, and more importantly, for correlating between datasets (work areas). It can be used in at least three ways:

1) It will display the definitions as bar-charts. It can be instructed to display these charts using various arrangements of the parameters in ascending or descending order.
2) It can perform a check of the viability of the definitions developed within a dataset.

Each definition can be checked against each other definition in the dataset to find the degree of overlap between the definitions. The level of precision of the check (allowable variation between compared parameters) can be specified, so that a qualitative judgement on the viability of the definitions can be made. In this case we do not want the definitions to overlap. So, the poorer the indicated correlation is, the better our definitions are. The results of the comparisons are shown as bar-graphs of the compared definitions and as a numerical ratio. The numerator of the ratio shows the maximum allowable variation between parameters, and the denominator shows the average variation between the parameters.

3) It can be used as a correlation tool. Each definition in a dataset can be compared with each definition in another dataset or datasets. The level of precision of the comparison can be specified, and consequently a qualitative judgement can be made on the indicated correlations. In this case we are looking for definitions that do overlap. So, the better the indicated comparison is, the more likely we are to have a good correlation. The results of the comparisons are shown as bar-graphs of the compared definitions and as a numerical ratio. The numerator of the ratio shows the maximum allowable variation between parameters, and the denominator shows the average variation between the parameters.

CONCLUDING REMARKS

Computer-assisted techniques and desktop computers have provided a new perspective to the problems of defining and correlating nearsurface stratigraphic units. Previously, progress was very slow in developing any sense of regional cohesiveness in glacial stratigraphy. The use of these programs and our large, nearsurface database has provided us with a means of rapidly developing preliminary lithostratigraphic interpretations, that are regional in scope. This gives our Quaternary stratigraphy a new, broader based perspective, which should provide valuable baseline stratigraphic information for a number of applications.