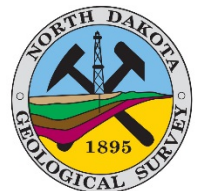# QUATERNARY STRATIGRAPHIC NOMENCLATURE, RED RIVER VALLEY, NORTH DAKOTA AND MINNESOTA:

Procedure for the Identification of Lithostratigraphic Units Described in North Dakota Geological Survey Miscellaneous Series No. 95

Kenneth L. Harris

Introduction
The data used for this project consisted of samples of glacial sediment (till to you and me) collected by various workers over the past 50 years. Samples were collected from outcrops, test borings, and other projects as part of theses work by students in the Geology Department at the University of North Dakota (UND-Geology), and geologists working for the North Dakota Geological Survey (NDGS). Realizing that the glacial units in the area should be correlatable, Steve Moran (NDGS) organized these data into a NDGS data base (N-File) and started efforts to interpret the tills in the early 1970s. Lee Clayton, Steve Moran, and Walter Moore and their graduate students explored and sampled many rivers cut banks and outcrops throughout the study area.

N-File was continued after Steve left the NDGS and has been combined with regional data from the Red River Valley of North Dakota and adjacent parts of Minnesota. This database is now called QBAS, and currently contains about 6,000 sample analyses that were used for this study. QBAS will be expanded as NDGS geologists continue Quaternary work in other parts of North Dakota.

Individual samples were analyzed for texture and coarse-sand fraction lithology. Textural analyses result in percentage content of sand, silt, and clay. The coarse-sand fraction is examined under a microscope and the percentage content of crystalline rock types, carbonate rock types, and shale grains is calculated. We have found this lithologic combination to be very useful in our area of interest. Workers in other areas will certainly find other combinations of lithologies to be more useful.

Data formats
The data recorded for each sample consist of the following:

> Identification of the geologist
> Location (T-R-S-QQQ)
> Depth and elevation
> Textural and lithologic results
> Normalized silt and normalized crystalline values
> Map area location

Detailed information on data formats is included in Appendix 1.

A working dataset is assembled using a Microsoft Excel spreadsheet. I have included digital files that include Microsoft Excel headings for the various data formats used by our programs (see Appendix 2 and Appendix 5). In general, data is entered in a Microsoft Excel spreadsheet and exported as a XXX.CSV (comma-separated variable) file which can be read by our programs. This format (ASXXX.CSV) is used in the cross-plot analysis program (AS-TS.BAS) to identify the tills present (Appendix 1, Appendix 2 and Appendix 4).

Additional software requirements
You will need DOSBOX on your computer to run the till programs contained here (www.dosbox.com). Some older versions of Microsoft Windows will allow you to run DOS programs directly (using a Windows GO.BAT statement), but not the most common versions today. The till programs were written in FORTRAN originally, and then converted to a version of Qbasic (circa 1997) to allow them to run on a PC (Appendix 5).

Procedure
Following the instructions provided in Appendix 5 (QBASIC PROGRAMS), open the cross-plot analysis program (AS-TS.BAS). Using the interactive, iterative analysis procedure with the cross-plots will help you identify and label the modal groupings present in your dataset. Once a modal group has been identified, all member samples can be labeled as till X and removed from your dataset. This procedure is repeated until all reasonable modal groupings have been identified, labeled, and removed from your dataset. This is an iterative process and can take some time before you will be happy with the viability of your results. Your results may be entered in an EXCEL spread sheet that can help you evaluate the viability of your interpretations (Appendix 2, Appendix 4, and Appendix 5).

Once you are happy with your results you can enter your interpretations in the EXCEL spread sheet HDGCORL (Appendix 5). This spread sheet allows you to label your interpretation of the till unit associated with each sample in your working database (Sample ID). This is a pretty useful spreadsheet. You can now sort the spreadsheet by

location, sample site, and depth to see each sample site in stratigraphic order, or sort by Sample ID to see where each of your interpreted till units occurs. This again is an iterative process and you may not be thrilled by the initial results – keep at it.

Now that your dataset is interpreted, you can refine the definitions of your tills. This can be done by using the CORL.XLS spreadsheet, sorted for till units (Sample ID). Calculate the mean, median, and mode for each till unit. I then break each of the till units into three groups (confidence levels) based on statistical properties:

> Group one, texture and lithology < 2 SD
> Group two; lithology < 2 SD, texture > 2 SD
> Group three, texture and lithology > 2 SD

Label each member-sample in each group:

> Group one A< 2 SD@
> Group two@ > 2 SD@
> Group three A>> 2 SD@

I also label the non-till samples. For example: sand, silt, clay, sandy-silt, silty-sand silty-clay, clayey-silt, shale, etc. This is useful when you display the sample data by location, sample-site, and depth, since the sands and lake clays are labeled as well as your interpreted tills (Appendix 5, HDG, and CORL).

Now sort the labeled samples by location, sample site, and depth to see all the samples labeled with your till units (Sample ID) and your confidence factor (< 2 SD, etc.). This is the acid test of your interpretation. Does the interpretation presented here make any sense? Are the units consistent in occurrence and distinctly separated from adjacent units?

T 151	WHO$	ID$	N%	TWN%	RNG%	SECT%	ELEV%	DEPTH%	SD%	SLT%	CLY%	XTAL%	CO3%	SH%	NSLT%	NXTAL%	RI	UNIT
add	klh	RLRC2	226	151	43	5	1078	4	30	46	24	50	35	15	66	59	>2 SD	U RLF
add	klh	RLRC2	226	151	43	5	1073	9	34	43	23	51	35	14	65	59	<2 SD	U RLF
add	klh	RLRC2	226	151	43	5	1065	17	46	38	16	60	37	3	70	62	<2 SD	L RLF
add	klh	RLRC2	226	151	43	5	1062	20	36	46	18	58	38	4	72	60	<2 SD	L RLF
add	klh	RLRC2	226	151	43	5	1056	26	43	41	16	58	39	3	72	60	<2 SD	L RLF
add	klh	RLRC2	226	151	43	5	1054	28	50	33	17	56	41	3	66	58	>2 SD	L RLF
add	klh	RLRC2A	226	151	43	5	1051	31	30	38	32	38	28	34	54	58	<2 SD	St Hilaire
add	klh	RLRC2B	226	151	43	5	1051	31	37	37	26	44	36	20	59	55	<2 SD	St Hilaire
add	klh	RLRC2	226	151	43	5	1051	31	40	39	21	59	32	9	65	65	>>2 SD	St Hilaire
add	klh	RLRC2C	226	151	43	5	1051	31	45	45	10	60	31	9	82	66	SDSL	SDSL
add	klh	RLRC2	226	151	43	5	1049	33	58	32	10	85	12	3	76	88	SDSL	SDSL
add	klh	RLRC2	226	151	43	5	1047	35	52	22	26	78	15	7	46	84	>2 SD	Marcoux
add	klh	RLRC2	226	151	43	5	1045	37	52	35	13	77	15	8	73	84	>2 SD	Marcoux
add	klh	RLRC2	226	151	43	5	1042	40	57	17	26	79	14	7	40	85	>>2 SD	Marcoux
add	klh	RLRC2	226	151	43	5	1040	42	51	40	9	85	15	0	82	85	SDSL	SDSL

This figure shows the results of the till search analysis for an outcrop on the Red Lake River near Red Lake Falls, MN. The confidence level (RI) and unit interpretation are shown in the right-hand columns.

Once you are happy with these results you are ready to map the occurrence of your till units in your area of study. This is helpful in understanding the source area of the till units, general coverage area, and the position of ice margins.

APPENDIX 1
EQUIPMENT AND FORMATS USED

EQUIPMENT USED: These programs were originally written to run on an IBM PC-XT (640K RAM, AST-SIXPAC, no graphics). The printer used was an EPSON FX-85 (a matrix printer). Some printer commands are written into some programs. DISCORT.BAS uses the compressed command for making bar-charts and the graphics programs use the form feed command. There may be some others used that I don't remember. You will have to change some of these commands to make things work on your system. The programs have run on all Windows versions through version 10, but Windows 10 requires a DOSBOX to run the programs.

FORMATS: There are three data formats (ND, AS, and GF) that are used. 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 for the main, original, near surface dataset that was maintained on an NDGS IBM System 34. All ND-formatted files are identified as ND-XXX.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.

IBM 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 [Il; l3]

Depth: The depth of the sample or well as appropriate [I4; l4-17]

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]

Sand-Silt-Clay: The results of the textural analysis (integer percent) [3I2; 35-43]

XTAL-CO3-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 7.5-minute topographic map sheet [I2, A1, I3; 71-76]

Box: The storage location of the samples in the core and sample library [I6; 77-82] (Note – all samples stored in Leonard Hall at UND were lost in the 1997 flood.)
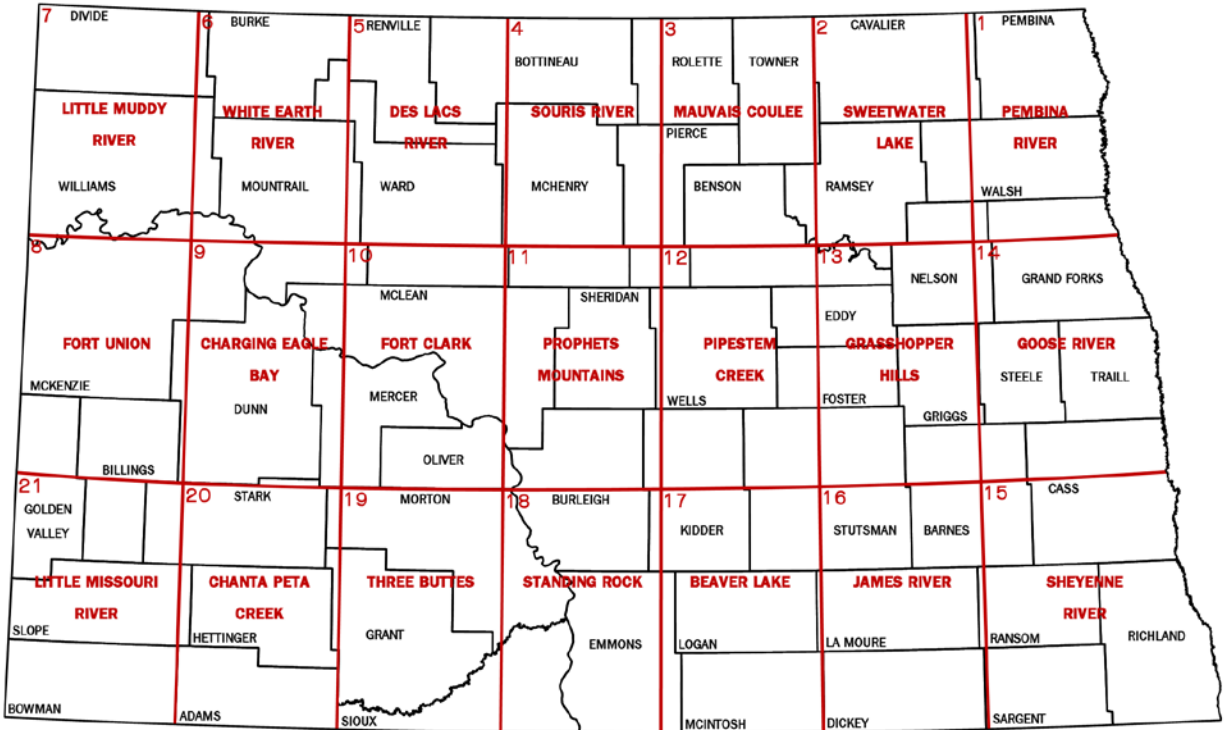
Map showing the identification numbers, names, and locations of Atlas Series map areas in North Dakota.

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 "ASXXX.CSV" 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. 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. All AS formatted files are written as comma-separated-variables.

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]

lD$: A project identification code used by the geologist who collected the samples [A6; 7-12]

N%: A unique, sequential file number (the N-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 the individual samples [I4; 27-30]

DEPTH%: The depth of well or sample [I4; 31-34]

SD%-SLT%-CLY%: The results of the textural analysis (integer percent) [3I2; 35-40]

XTAL%-CO3%-SH%: The results of the lithologic analysis of the coarse sand fraction (integer percent) [3I2; 41-46]

NSLT%-NXTAL%: Normalized SLT [NSLT = ((SLT/(SLT+CLY))*l00)] and normalized XTAL [NXTAL = ((XTAL/(XTAL+CO$_3$))*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]

LlG%: The integer percent Lignite content of the sample [I2; 54-55]

GF FORMAT: This format is used by the graphics programs. The records are eight characters long.

SD%: Integer percentage content of sand [I2; 1-2]

NSLT%: Integer percentage normalized SLT [I2; 3-4]

NXTAL%: Integer percentage normalized XTAL [I2; 5-6]

SH%: Integer percentage content of shale [I2; 7-8]

APPENDIX 2
PROGRAMS USED

PROGRAMS: The programs included in the file folder STRAT DEM 2020 are introduced here. Click on the link to download the files. The discussion of the individual programs is organized according to the format of the data that they read. All programs were written in FORTRAN and later converted to QBASIC (version A3.10). There are several generations of QBASIC that have evolved since these programs were written. Consequently, you may have to make some adjustments to the programs for more modern versions of QBASIC. Click here to

ND FORMAT: These programs all read the ND-formatted data. This is the format used for data storage on the original NDGS IBM System 34. It was 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 AS-formatted files that you write should be given names of the form "ASXXX.CSV".

> **ND-FILES.BAS:** This program inventories the ND formatted data that has been transferred from the IBM System 34 to an IBM PC compatible file. 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$). The data in each subfile is listed in one of three categories:
>
> 1) Location only, no data
> 2) Uncoded data that exists in our back-up files
> 3) Numerical data that is in the subfile
>
> The data in each category is subdivided into textural and coarse-sand lithologic data types and the total number of entries in each subfile and category is shown. The output can be routed to screen or printer.
>
> **ND WHOSE.BAS:** This program generates a subfile consisting of all samples collected by a selected geologist or agency (WHO$). The output can be routed to screen, printer, or file.
>
> **ND CODE.BAS:** This program generates a subfile of all entries that are indicated as having uncoded data available in our backup files. The output can be routed to screen, printer, or file.
>
> **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 file.
>
> **ND-T&R.BAS:** This program reads a file and rewrites it in descending order of township and range. It currently uses a really awkward and slow sorting routine that needs to be rewritten. The output can be routed to screen, printer, or file. [It still writes in an old AS-format with only two leading A-variables instead of three.]
>
> **ND-AS.BAS:** This program reads a file and rewrites it in an AS-format. The output can be routed to screen, printer, or file.

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 "ASXXX.CSV" and GF-formatted files should be labeled "GFXXX.CSV".

> **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 backup files. It will write this list to the screen, printer, or file.
>
> **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 file.
>
> **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 file.

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

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

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 file.

AS-TS.BAS: This program (Till-Search) reads an AS-formatted data file, searches it for all samples that meet a user defined definition, and labels those samples. A definition consists of specific values of the four correlation parameters (SD, NSLT, NXTAL, and SH) and an allowable range of variation of these parameters. The definitions used are 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 the member-samples of the modal group in the dataset that meet that definition and remove them from the dataset. This process is repeated until all identifiable modal groups in the dataset have been defined, labeled, and removed. This results in a dataset with all member-samples of all identifiable modal groups labeled and removed. The output of the program is:

1) <u>Till datasets</u>: lists of member-samples of specific modal groups

2) <u>Correlated dataset</u>: a list of the entire dataset identifying all member-samples in all modal groups

3) <u>Residual dataset</u>: a list of all those samples that do not meet the definitions of any of the modal groups

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 file. The correlated dataset is written to file only, in a modified AS format (because the definition labels add another variable). The correlated dataset can only be read by ASCORLST.BAS.

There are eight menu options in the Till-Search MAIN MENU that will help you choose system options. The first four will help you quickly access the programs you need to work with or create files. Options five and six call programs designed to subdivide large data files. Any file you create should contain < 450 entries or it will "choke" QBASIC. Option seven is a file directory, and option eight returns you to the system (DOS or Windows).

1) ND*.DAT AND ND*.CSV FILES: Read an ND*.DAT (NDGS N-FILE string-variable format) and:

    a) Identify workers and data types (use file: ND13-887.DAT)
    b) Create an ND*.CSV (use file: ND13-887.DAT)
    c) Create an AS*.CSV file (use file: ND13-887.DAT)
    d) Subdivide a ND*.DAT file (use file: ND13-887.DAT)

2) AS*.CSV AND COR*.CSV FILES: Read AS*.CSV or COR*.CSV formatted files and:

    a) Convert to a GF*.CSV format for graphing (use file: AS13NOLY.CSV).
    b) Convert to cross-plot files that can be posted or contoured using SURFER (use file: AS13NOLY.CSV).
    c) List an AS*.CSV file (use file: AS13NOLY.CSV or any BH*.CSV or OC*.CSV file).
    d) Identify workers and data types (use file: AS13NOLY.CSV).
    e) Display outcrop (OC*.CSV) or borehole (BH*.CSV) files (use file: OC4091.CSV, BH5001.CSV, BH6218.CSV, and BH6220.CSV).

f) List a correlated dataset (use file: AS13.COR).
g) Run a cluster analysis (use file: AS13NOLY.CSV).

3) GRAPH DATA (GF, BH, AND OC*.CSV FILES): Read a:

   a) GF*.CSV file and make cross plots (use file: GF13WK.CSV and GFR11.CSV).
   b) GF*.CSV file and make histograms (use file: GF13WK.CSV and GFR11.CSV).
   c) OC*.CSV or BH*.CSV files to graph outcrop and borehole data (use file: OC4091.CSV, BH5001.CSV, BH6218.CSV, & BH6220.CSV).

4) CORRELATED DATA: Use a:

   a) IN*.CSV file and the display-correlate program (DISCORT) to check "tills" (cluster definitions) for overlap or correlate "tills" from one area to another (use files: IN13.CSV and IN1491.CSV).
   b) SI*.CSV file and write a SO*.CSV file; creates a tally matrix that summarizes the stratigraphic relationship between your labeled "tills" (use file: SI13.CSV). The output file (SO13.CSV) can be imported into a spreadsheet and displayed using a heading (HDGSO.XLS) or printed as an unlabeled matrix using a DOS print command (TYPE SO13.CSV > LPT1).
   c) Keyboard input to convert correlation parameter notation (SD-NS-NX-SH) to conventional notation (SD-SL-CL, XT-CO-SH) and vice-versa.
   d) IN*.CSV file and list it (use file: IN13.CSV & IN1491.CSV).

5) DIVIDE AN AS*.CSV FILE INTO TWO SUBFILES: Reads an AS*.CSV file and divides it into two approximately equal subfiles.

6) DIVIDE A COR*.CSV FILE INTO TWO SUBFILES: Reads an COR*.CSV file and divides it into two approximately equal subfiles.

7) FILE DIRECTORY: Lists files of specified types (AS, ND, GF, ...) in the QBAS directory. This is useful if you forget a filename because you were far too creative in naming your files.

8) QUIT FOR NOW: Returns you to DOS or WINDOWS.

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

GF-FORMAT: These programs read the GF-formatted data and generate graphical summaries of it.

GF-HIST.BAS: This program will generate histograms of each of the four correlation parameters (SD, NSLT, NXTAL, 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 can present crossplots of NSLT vs SD, NXTAL 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 tallies the level of that coincidence. The number of coincident values is counted 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.

APPENDIX 3
DATA MANAGEMENT AND FILE MAINTENANCE

DATA MANAGEMENT AND FILE MAINTENANCE: I suggest that you use a spreadsheet/database-manager such as Microsoft Excel as an editing and formatting tool. The sorting routines available through spreadsheets are fast and easy to use. They will allow you to edit the data in order to present it in a useful format. For example, all data should be in ascending or descending order of township, range, section, QQQ to make it easy to post maps. You should also sort your data to eliminate "zero" data, and sample data should be in ascending order of sample depth (or descending order of sample elevation) for each of the sample sites (test holes or outcrop). This is necessary to interpret the CORRELATED dataset stratigraphically.

Make spreadsheets for all source files (AS13NOLY.XLS), correlated data sets (AS13.COR), and Interpreted datasets (IN13.XLS). These files contain data and interpretations for areas you have worked. You will use them later for correlation between work areas.

QBASIC programs read comma-separated-variable files (XXX.CSV). Microsoft Excel can import and export comma-separated-variable files (*.CSV).

APPENDIX 4
THE CLUSTER ANALYSIS

THE CLUSTER ANALYSIS: Select MAIN MENU-OPTION 2 and then AS*.CSV MENU-OPTION 7 to perform a cluster analysis of your dataset. Read an AS*.CSV dataset (AS13NOLY.CSV). You will be prompted to name some work files that will be generated by the program. The work files used are:

> **CORRELATED DATASET:** The completely labeled dataset you will have when you are finished (AS13.COR).

> **TILL DATASET:** Those samples that meet each of the definitions that you are about to develop (number your definitions consecutively (AS13.T01, AS13.T02, ...). I usually number the "tills" in order of increasing shale content just to have a convention to follow. Sometimes the order of selection of the clusters can have a minor effect on the number of samples in a modal group.

> **RESIDUAL DATASET:** The original dataset less those member-samples that are included in the definitions that you have developed. You can number the residual files consecutively (AS*.R01, AS13.R02 ...) or call them all AS13.RES or just RES.

This version asks you if you want only lignite-bearing samples – just say no (0).

Enter the values of the correlation parameters that you feel characterize the clusters that you want to define. A good cluster will show modal tendencies in all four correlation parameters. Just looking at the three crossplots you don't know which clusters are associated. To find out their association, you can enter just one, two, or three coordinates and the cluster analysis program will give you the value of the associated correlation parameters. You will want to enter only those you are sure of. For example, if you enter (30/6; 50/50; 50/50; 25/5) [% SD/gate width; % NS/gate width; % NX/gate width; % SH/gate width], you are saying that the sand content is 30% +/- 6%, and that the shale content is 25% +/- 5%; but that you don't know % NS or % NX (50% +/- 50%).

The cluster analysis program will return the mean and standard deviation of each of the associated correlation parameters, for each of the trial populations you define in the dataset. You simply re-enter trial definitions, of a cluster, based on the calculated mean and standard deviation values provided by the program (you want to refine the "till definition"). I usually use the indicated mean and an allowable variation of +/- two standard deviations (but not exceeding +/- 20%) for subsequent trials.

> For example:
> => first trial was - - - - - - - (20/05: 50/05: 55/05: 10/03)
> => program returns - - - - - (22/03: 48/06: 60/06: 08/04)
> => next trial might be - - -   (22/06: 48/12: 60/12: 08/08)

This process is repeated until a stable mean and standard deviation are achieved (the program returns the same values for each trial). Crossplots of the modal groups can be generated on screen using the cluster-analysis program. The crossplots can help you evaluate the "tills" that you have defined; so, you can accept them or further refine their definitions.

You now want to SAVE all of the member-samples of this modal group ("till") to disk. For a better estimate of the total population, you may want to use a different multiple of the standard deviation for the last trial run, the one you SAVE, after a stable return is achieved. Assuming a normal distribution (foolish assumption) these multiples of the standard deviation should result in the following percentages of the total population: 1 SD = 68%, 1.5 SD = 86%, 2 SD = 95%, 2.5 SD = 99%. I usually use a multiplier of 2.5 SD, but never use an allowable variation of more than +/- 20%. You may want to be more conservative on your population estimate. Your definition (AS*.T01, AS*.T02, ...) of this "till" can now be written to disk, and all of the member-samples will become part of the CORRELATED (labeled) dataset. The residual dataset consists of the remaining unlabeled samples and becomes the working dataset used for subsequent searches.

This process is repeated until all identifiable clusters have been defined and labeled. When this stage is reached, SAVE the CORRELATED dataset and the RESIDUAL dataset to disk. A crossplot of the RESIDUAL dataset (using MENU 4, option 1, or cluster-analysis program graphics) will allow you to decide if you have missed any clusters. If so, repeat the analysis procedure until you are satisfied.

Be sure to SAVE the CORRELATED dataset to disk, it will be used to develop your stratigraphic interpretations.

TEST YOUR DEFINITIONS FOR OVERLAP: Load the HDGIN.XLS spreadsheet in Microsoft Excel and generate the interpretation files (IN*.XLS and IN*.CSV). Type in the number of samples and correlation parameters that you have generated for each "till" (Number-of-Samples, % SD, % NS, % NX, % SH) and the spread sheet will calculate the conventional notation (sand-silt-clay and crystalline-carbonate-shale, and other interesting tid-bits). Save the spreadsheet for future use (IN*.XLS). Then export the portion of the spreadsheet that contains the Number-of-Samples, % SD, % NS, % NX, % SH to create a IN*.CSV file. This is an awkward procedure and you may find using the spreadsheet analysis easier (see Procedure, p. 1).

Using MAIN MENU-OPTION 4 and CORRELATED DATA MENU-OPTION 1, load the display-and-correlate program (DISCORT).

Load your IN*.CSV file(s). The program will ask for the maximum number of "tills" (for IN13.CSV it is 11), how many files you want to run (in this case it is 1), and the name of the IN*.CSV file(s) you want to read.

Make some bar-graphs to help visualize the definitions.

Search for correlatable tills; check for overlapping definitions within your dataset:

- a) Increase the GATE WIDTH from 1% to 10%.
- b) Overlapping definitions will be displayed only for the GATE WIDTH at which overlap first occurs.
- c) A CORRELATION COEFFICIENT is also provided with the printout. It is the maximum difference between any of the correlation parameters over the average difference between all the correlation parameters. This version provides correlation coefficients for all four parameters (SD; NS; NX; SH), the textural parameters (SD; NS), and the lithologic parameters (NX; SH). Sometimes the lithologic correlation may be judged to be more important than the textural correlation or vice-versa.

Cross correlate the results of this run to identify those definitions that should be combined. For example, if "till #4" and "till #5" show a correlation coefficient of 5/3 that means that there is a maximum difference of 5% and an average difference of 3% between the correlation parameters. They should be combined (if the ratio had been 9/7, a maximum difference of 9% and an average difference of 7% between the correlation parameters, they would not have been a viable cluster). To combine "tills", use DOS to add the individual disk files together (copy filename+filename new name or COPY AS13.T01+AS13.T02 AS13.TA1).

Crossplot the new file using the interactive graphics in the cluster analysis program, evaluate whether or not it should be combined, and read and record the correlation parameters for the new file, rerun the dataset using the new cluster in place of the old ones (TA1 replaces T01 and T02).

You are finished with the dataset when you are satisfied that you have defined all of the modal clusters present, and eliminated all of the overlapping definitions.

OTHER DATASETS: Repeat the above steps for other datasets.

CORRELATION BETWEEN DATASETS: Using MAIN MENU-OPTION 4 and CORRELATED DATA MENU-OPTION 1, enter the names of the IN*.CSV datasets that you want to correlate.

Make some bar graphs to help visualize the "tills".

Search for correlatable tills; check for overlapping definitions between the tills in the datasets you are correlating:

a) Increase the GATE WIDTH from 1% to 10%.
b) Overlapping definitions will be displayed only for the GATE WIDTH at which overlap first occurs
c) A CORRELATION COEFFICIENT is also displayed. A high degree of overlap between two definitions means a high probability that they correlate

Cross correlate the results of this run to order to identify those definitions that correlate. For example, if "till #4" in dataset #1 and "till #5" in dataset #2 show a correlation coefficient of 5/3 that means that there is a maximum difference of 5% and an average difference of 3% between the correlation parameters. They are probably correlative (if the ratio had been 9/7 they would not have been correlative).

STRAT DEM 2020

STRAT DEM 2020 – Stratigraphic demonstration files and programs: These files are included here as examples of working files used with Till-Search@ programs. This is a brief summary of the information they contain.

ASDAT: These are AS-Formatted files for a work area (AS13). They can be read by the Qbasic program and provide an example of the Till-Search@ procedure.

AS13.R11	AS13NOLY.CSV
AS13.T01, T02, T03, T04, T05, T06, T07,	AS13.COR
T08, TO9, T10, T11	

BH: These are Microsoft Excel spreadsheets containing AS-Formatted examples of borehole data. They can be used with the Till-Search@ programs.

BH5001.CSV	BH5295.CSV
BH5154.CSV	BH5300.CSV
BH5157.CSV	BH6218.CSV
BH5172.CSV	BH6220.CSV
BH5292.CSV	BH5167.CSV
BH5158.CSV	

CAL: These are useful Microsoft Excel spreadsheets for working with the AS-13 till data.

RRV CORL.XLSX	SI13.CSV
CSESD XXX.XLS – counting work sheet	SO13.XLS
IN13.CSV	

GFDAT: These are graphic file used with the AS-13 till data.

GF13WK.CSV
GF13.R11

HDG: These are Microsoft Excel (.XLS; .XLSX) and Quatro Pro (.WB3) spreadsheets used when working with the till data. They contain examples of AS, CORL, IN, and SO formatted data.

RRV CORL.XLSX	HDG DUP.WB3
HDGAS COLOR.XLS	HDG DUP 2.WB3
HDGAS.WB3	HDGIN.XLS
HDGAS.XLS	HDGSO.XLS
HDGCORL.WB3	
HDGCORL.XLS	

IN: These are Microsoft Excel spreadsheet examples of interpreted till data.

IN13.CSV
IN1491.CSV

NDAT: These are examples of ND-formatted till data.

ND13-887.DAT	ND1388.CSV
ND13871.CSV	ND13-88.DAT
ND13872.CSV	

OC: These are Microsoft Excel spreadsheets containing AS-Formatted outcrop data.

OC4091.CSV   OC5943.CSV
OC4093.CSV   OC5945.CSV

QBAS: These are the Qbasic application and programs used by Till-Search@. Once you have installed DOSBOX on your computer, follow this procedure to install Qbasic and give you access to the till programs:

> DOSBOX
> z:>
> z>: mount c programs
> z:>C:
> C:>cd Qbasic
> C:\Qbasic>
> C:\Qbasic>qbasic (you are now operating Qbasic in DOS)
> File> Open> StartStr>Run (This will bring up the menu to the Stratigraphic Programs)

QBASIC.EXE – Qbasic application
STARTSTR.BAS – This program is used to start Till-Search@, it calls up the main menu

QBASIC PROGRAMS: These are the programs that can be called up by the main menu to run various Till-Search@ functions. Also listed are some Qbasic utility programs not used by Till-Search@.

AS-CODE.BAS   DIRFILES.BAS
ASCORLST.BAS   DISCORT.BAS
ASFILES.BAS   GF-HIST.BAS
AS-GF.BAS   GF-XPLT.BAS
ASHACK.BAS   INLST.BAS
AS-LST.BAS   ND-AS.BAS
ASNUM.BAS   ND-CSV.BAS
AS-TS.BAS   ND-FILES.BAS
BHDRAW.BAS   NHACK.BAS
CONVERT.BAS   STR-ORD.BAS
CORHACK.BAS   AS13NOLY.CSV

SISO: Microsoft Excel spreadsheets used for working with interpreted data. Possibly useful for determining stratigraphic order of till units; read – not convinced, but written into the Till-Search@ program.

SI13.CSV
SO13.CSV

WRD: WordPerfect X9 files containing various work sheets used for collecting basic data and data interpretation.

CLUST WS.WPD – This is a cluster analysis work sheet used with Till-Search@ program.
COR DESC.WPD – This is a core description work sheet.
CSE SD B.WPD – This is a coarse-sand lithology work sheet, a Minnesota (eastern-source) model.
CSE SD C.WPD – This is a coarse-sand lithology work sheet, a Minnesota (eastern-source) model.
XCOR.WPD – This is a cross-correlation worksheet, used the identify overlapping units in your datasets.