Current Research (1992) Newfoundland Department of Mines and Energy, Geological Survey Branch, Report 92-1, pages 321-330

SOME USEFUL MAPPING UTILITIES FOR GEOSCIENTIFIC DATA USING A PERSONAL COMPUTER

G. Kilfoil and L. Nolan
Geochemistry and Geophysics Section

ABSTRACT

In recent years, computer programs and algorithms have been developed by the Geological Survey Branch to assist in the processing and presentation of existing geoscientific databases. In an effort to promote the use of the databases held by the Branch, several of the more useful routines are described here. These PC-based utilities range in complexity from a simple coordinate transformation routine to more elaborate software for data processing and display of gridded geochemical and geophysical data.

INTRODUCTION

Several PC-based computer utilities have been developed at the Geological Survey Branch to streamline map production. In most cases, these utilities were originally written to perform a specific task for which no program existed at the Branch. Later, as their day-to-day usage increased, certain utilities were broadened and updated to handle a variety of computing situations or tasks and thereby became more general in application.

The utilities can be divided into two categories: general-purpose mapping utilities and those that perform a particular task with gridded datafiles. A brief description of the function of each utility is listed in Table 1 in the order that they appear in the text.

All utilities are compiled for use on an IBM-compatible personal computer running under the MS-DOS or PC-DOS operating systems. Although the presence of a math co-processor is not strictly required by these programs, the more computation-intensive utilities, such as **DISPLAY** or the gridded-data filtering routines, will operate significantly more slowly in the absence of a math co-processor, particularly if executed on a PC, based on the Intel 8086/8088 or 80286 micro-processors. **GRIDVIEW** and **DISPLAY** require at least a VGA graphics board and colour monitor; all other utilities will work with monochrome or CGA graphics capabilities. Disk storage space is required for input (and output) datafiles for most of the utilities. Although all utilities will function from floppy disk drives, the presence of a hard disk (with a disk cache) in the PC is recommended when executing the gridded-data manipulation routines, as disk access is quite intensive for these utilities.

The source code for all the general purpose utilities was written in Basic and compiled with the Microsoft QuickBASIC optimizing compiler (Microsoft, 1988). In contrast, the utilities that operate on gridded datafiles have been written and compiled in FORTRAN (Microsoft, 1989), due to its superior data-handling capabilities. The general-purpose utilities tend to be smaller and each performs a specific task, but such tasks are generally required more often than those that the gridded data-handling utilities fulfil.

The gridded data-handling utilities can be further classified into general-purpose grid utilities and grid-filtering utilities. Among the tasks performed by the general-purpose grid utilities are: creating VGA colour images from gridded datafiles (in real time), merging gridded datafiles from adjacent areas, and converting gridded datafiles to alternative formats. In contrast, the grid-filtering utilities alter the contents of a gridded datafile by passing a small matrix of filter values over the gridded data elements in order to extract or remove a particular trend or component from the data. The dimensions and formats of filtered grid files remain unchanged relative to the unfiltered versions.

GENERAL-PURPOSE MAPPING UTILITIES:

UTMCON: Converts geographic coordinates of Latitude and Longitude to Eastings and Northings within a user-specified UTM projection, or the reverse.

This handy utility is designed to convert sets of geographic coordinates between the spherical (i.e., Latitude and Longitude) and Universal Transverse Mercator (UTM)

This project is partially funded by the Canada–Newfoundland Cooperation Agreement on Mineral Development (1990-1994); project carried by Geological Survey Branch, Department of Mines and Energy, Government of Newfoundland and Labrador.

Table 1. List of documented utilities for mapping geoscientific data and their functions

General Purpose Utilities:	
UTMCON:	Converts coordinates of Latitude and Longitude to Eastings and Northings within a user-specified UTM projection, or the reverse.
UTMCON2:	Converts Easting and Northing coordinates within one UTM zone to equivalents within a second zone.
MAP_CORN:	Automatically determines map area titles and UTM corner coordinates from their NTS designations
UNISTAT:	Provides basic univariate statistics on numerical data sets and generates publication-quality cumulative frequency plots.
GRIDVIEW:	Examines the spatial distribution of data points to aid in selection of gridding parameters.
Grid Utilities (General Purpose):	
DISPLAY:	Interactively displays the contents of binary gridded data files as images on a PC equipped with a VGA graphics card and monitor.
MERGEO:	Merges two binary files of gridded data and/or extracts a rectangular window of the input data for output.
GRD_ASC:	A conversion utility to dump all (or a portion) of a binary grid file to an ASCII file, or vice-versa.
MASK:	Compares two binary gridded data files, and replaces data values within the second with null values at all grid cell locations where null values exist within the first.
GRD_LEG:	Replaces data within a rectantular area of a binary gridded data file with a range of values corresponding to a grid legend.
Grid Utilities (Filtering):	
MEAN:	Applies a Mean filter of variable-width to a binary gridded data file.
MEDIAN:	Applies a Median filter of variable-width to a binary gridded data file.
ATM:	Applies an Adaptive Trim Mean filter of variable-width to a binary gridded data file.
FENCE:	Applies a Fence filter of variable-width to a binary gridded data file.

coordinate systems. The coordinates can be input and/or output either to the computer screen or to a user-specified datafile.

The user is first given the choice of converting Latitudes and Longitudes to UTM Easting and Northing coordinates, or vice-versa. Then the default central meridian, UTM zone and scale factor are displayed and the user is provided the opportunity to change any of these. By choosing a non-standard central meridian and/or scale factor the user can select a specific Transverse Mercator (as opposed to UTM) coordinate system for transformation. If the UTM zone is specified, the central meridian for that zone is taken as the default value.

The user is then queried as to whether input from or output to a file is desired. If either is selected, the name of the file is requested. If input from a file is selected, the first fields of each input line must be UTM Easting, Northing and zone when the UTM conversion option is specified or Latitude and Longitude for spherical coordinate conversion. On output, these fields will be replaced with their equivalents in the new coordinate system and all remaining character data on the input line will be output unchanged.

UTMCON will continue to convert pairs of geographic coordinates until either the end-of-file is reached on input from a data file or zeros are typed as coordinates during keyboard entry.

UTMCON2: Converts Northing and Easting coordinates within one UTM zone to equivalents within a second zone.

UTMCON2 is actually a special version of **UTMCON** that converts coordinates from one UTM zone to another in a single step. This utility has proven useful for converting the coordinates measured from maps in one UTM zone to those for an adjacent zone before the datasets are merged to produce a regional map.

The user is queried for the UTM zones to be used on input and output. As in **UTMCON,** the input data source can be selected either as a datafile or the keyboard; output data destination can be specified as a datafile or the screen.

```
--- Program to Calculate UTM's for Corners of NTS Map Sheets ---

Do you want printer output? n
Do you want to output polygons of corner coordinates to a file? n
Do you want to display corner coordinate windows on the screen? n

Do you want to round map corner extremes (UTM) to a nearest factor? y
Enter a rounding factor for the extremes of UTM corners (e.g. 100)? 100

(1) Central Meridian: 57 ; (2) UTM Zone: 21 ; (3) Scale Factor: .9996
To Change a Parameter, Type Number in Brackets; Otherwise Hit <Enter>?

Input NTS Map: e.g., 'NTS 12A'; '12A' (0 to quit; -1 to alter params)? nts 2d/4

        Map Title:  Twillick Brook
       Map Number:  2D/4
            Scale:  1:50 000
   Latitude range:  48.00 to 48.25 degrees North
  Longitude range:  56.00 to 55.50 degrees West

       (48.25, 56.00)                                     (48.25, 55.50)
  ( 574236.32,5344351.72)  -----------------------------  ( 611353.68,5344955.94)
           |                                                    |
           |                                                    |
           |                                                    |
       (48.00,|56.00)                                     (48.00,|55.50)
  ( 574597.46,5316565.14)  -----------------------------  ( 611895.46,5317169.95)

Rounded UTM Window of Map Corner Extremes:
  ( 574200.00,5316500.00)  -----------------------------  ( 611900.00,5345000.00)

Inputs NTS Map: e.g. 'NTS 12A'; '12A' (0 to quit; -1 to alter params)? 0
```

Figure 1. *Example use of* **MAP_CORN** *showing results for the Twillick Brook map area, NTS 2D/4, central Newfoundland. (User responses are in bolder text).*

MAP_CORN: Automatically determines map-area titles and UTM corner coordinates from their NTS designations.

This utility provides the title and corner coordinates for a map area by simply referring to the 1:250 000- or 1:50 000-scale map sheet by its (Canadian) National Topographic System (NTS) reference. **MAP_CORN** is a hybrid version of **UTMCON** that determines and reports the four corner coordinates of the map area in both spherical and Easting/Northing of a user-specified UTM zone.

MAP_CORN searches a list of map-area titles in a DOS file 'MAPS.LST', which should be located in the current working directory. If a file by this name does not exist in the current directory, the user will be prompted for the path to its location. Presently, this list contains only the map-area titles for the area covered by the Province of Newfoundland and Labrador, but the list may be expanded or replaced with a list for any other province/territory or for the whole of Canada. In practice, the map list should be kept to a minimum as the response time for **MAP_CORN** may be slowed by searches through for very long lists of map titles. If a map title for a particular NTS designation is not located within 'MAPS.LST', the user will be prompted for the map title.

Individual 1:50 000-scale map areas are distinguished from 1:250 000-scale map areas by including the slash character (/) in the NTS designation. For example, the 1:50 000-scale map area for Twillick Brook in central Newfoundland would be specified by typing 'NTS 2D/4' (or simply '2D/4'), whereas the 1:250 000-scale map for Gander Lake (which contains the Twillick Brook map sheet) would be specified by typing 'NTS 2D'. Figure 1 shows an example session in which the corner coordinates for the Twillick Brook map area have been calculated within UTM zone 21. Note that **MAP_CORN** iterates the calculation of corner coordinates for additional map areas until a '0' is typed at the prompt for the next map area.

Additional features of **MAP_CORN** are: the ability to specify the central meridian and scale factor of a specific Transverse Mercator projection; an option to calculate the smallest UTM coordinate window, rounded to the nearest multiple of a factor that contains the all four corners of the map area; an option to output (to a user-specified ASCII datafile) the map title followed by five Easting/Northing data pairs defining a closed polygon of the map-area corner coordinates; and optional output to the default printer. The rounded UTM coordinate window option is useful for determining the range of Easting and Northing coordinates for a particular map area when gridding or windowing (see description of **MERGEO** below) a large dataset to a particular cell size. For example, before interpolating a dataset for the Twillick Brook map area to a 100-m grid cell, a user may wish to execute **MAP_CORN** as in Figure 1 (specifying the 100-m rounding factor) in order to determine the Easting and Northing coordinate ranges that would limit the output grid size to this map area.

UNISTAT: Provides basic univariate statistics on numerical datasets and publication-quality cumulative frequency plots.

UNISTAT (UNIvariate STATistics) is an interactive computer program for the IBM-PC and closely compatible microcomputers designed to assist in the process of analyzing and interpreting geochemical data, element by element (univariate) by providing basic dataset statistics, combined with publication-quality graphical output (Nolan, 1990). Descriptive statistics (minimum, maximum, mean, median, standard deviation and coefficient of variation) are computed using selected options of upper/lower data cut-off limits and logarithmic transformation to produce histograms and cumulative frequency plots. Symmetry, population distributions, range and outliers are readily interpreted from the histogram, and the cumulative frequency plot enables breaks between populations of element concentrations to be uniquely identified. The datafile used by UniStat has a fixed ASCII format (Table 2), which can be created with any text editor or with the file-preparation module, which will allow the user to interactively label variables and flag missing cases in a data matrix. Datafiles created by this module or a text editor can contain up to 40 variables (columns) and 10,000 samples (rows).

Table 2. Example data set showing **UNISTAT** ASCII file format.

Title Line—Test Data set			
3			
Arsenic	ppm		
Copper	ppm		−3 variables (cols)
Nickel	ppm		
179.0	6	11	
811.0	3	4	
37.2	5	6	−473 samples (rows)
34.1	7	7	
.	.	.	
.	.	.	
.	.	.	

The histogram provides information at a glance on the frequency distribution of the data. It determines if the frequency distribution is symmetrical (normal) or skewed, unimodal (one population) or polymodal (several populations) and whether outliers are present, and also the range of the data.

The cumulative frequency plot provides a graphical presentation of the frequency distribution (cfd) on a normal probability scale. A straight line plot on a normal probability scale indicates a dataset with a single, normal population. If the line is curved or plots as several straight lines, a non-normal distribution or multiple populations in the data is indicated. The points of inflection on the plot can be read off the concentration axis and used as distinct population breaks for plotting geochemical element distributions on contour maps (Sinclair, 1983).

GRIDVIEW: Examines the spatial distribution of data points to aid in the selection of gridding parameters.

The presentation of point observations as a continuous surface for display as a raster image, or for contouring, requires that the irregularly distributed values be transformed into a regular grid. The extent to which this computed grid represents the original data points is controlled by several parameters that must be selected by analysis of the dataset. The grid cell size is the most important parameter that controls the quality of the output grid. A cell size too large will result in the grid being overly smooth with the loss of fine detail and one too small will result in increased processing (not critical) with no improvement in grid detail.

GRIDVIEW displays the set of data points on an x-y grid of user-specified grid cell size, on which a circular search area with a user-specified radius can be displayed. This search area can be moved over the grid allowing visual analysis of sample density with respect to the grid cells to aid in the selection of gridding parameters. The cell size and search area radius can be interactively changed. As well, the circular search area may be moved around the dataset while calculating the points that fall within. A histogram of the number of data points per cell is displayed for each selected grid cell size. An efficient grid cell size is one where the number of cells containing one point is a maximum while keeping the overall grid within a reasonable size for processing.

GRIDVIEW has been designed as an interactive tool to aid in the selection cell size and search radius for gridding-point sample datasets by combining visual displays of the data with informative histograms. Other point data-analysis techniques such as variograms give an indication of the spatial continuity of data-point measurements over distance, which helps determine the search radius parameter but not grid cell sizes or search methods (nearest neighbour, octant, etc.). Variograms are more complex to apply, and typically depend on the data behaving in certain consistent, mathematically describable ways, which is often not the case.

Grid Utilities (General Purpose):

DISPLAY: Interactively displays the contents of binary-gridded datafiles as images on a PC equipped with a VGA graphics card and monitor.

The operation of program DISPLAY, complete with examples of colour and grey-scale images, has been outlined previously (Kilfoil, 1991a). The more comprehensive description contained in the User's Manual (Kilfoil, 1990) has been supplemented by upgrade notes as the utility was debugged and retrofitted with more functions. Only those recent major revisions that have added options to **DISPLAY** are outlined here.

Aside from minor bug fixes, two functions have been added: the ability to produce monotone images in shades of one of several colours, and the capability of overlaying vector files on the images for geo-referencing purposes.

In previous versions of **DISPLAY**, monotone shaded-relief or total-field images could only be generated in 64 shades of grey. Monotone images can now be generated in any one of the additional colours: red, green, blue, cyan, magenta, yellow or grey. The monotone colour appears as an extra option in the **DISPLAY** menu. The monotone colour can be selected by typing 'G' at the menu prompt.

The major drawback to previous versions of **DISPLAY** was the inability to overlay information that would provide geo-referencing on images generated. Presently, **DISPLAY** allows only vector files in Tydac SPANS format to be plotted over images for reference. Vector files consist merely of a series of points defined by their X-Y locations. These points are grouped in the vector file into line segments that define a geographic feature, such as a lake, geological boundary, survey line or a boundary of a map sheet or claim block. Appendix A describes the structure of a Tydac SPANS vector file in more detail.

An extra option line was added to **DISPLAY**'s menu screen to allow the user to toggle the vector overlays off or on by typing 'V'. The colour used in plotting the overlay vectors can be selected from the prompt when vector overlays are turned on. Any of the monotone colours listed above, as well as black and white, can be selected independently for each vector file. Thus, the user may overlay one vector file, such as a file of coastline, in grey and another vector file, such as map boundaries, in black on a colour image. The vector overlays are not stored in memory when the user switches between the image display screen and the menu screen. Instead, the vector overlays are plotted after the image(s) have been generated, each time that the image screen is selected. The user may wish to temporarily turn the vector display off when selecting optimum images for two reasons: the vector overlays plot over the image(s), masking the image information contained beneath and thereby detracting from the image(s) and the plotting of vector overlays can be time consuming, particularly if the vector files are quite large. For both of these reasons, the user is advised to keep vector overlays to a minimum; use only sufficient information to properly geo-reference images generated.

Previous versions of **DISPLAY** contained errors that impaired the operation of the program. In particular, an annoying programming bug, which affected the refresh of the right-hand image in the two-image display mode, often resulted in a partially blank image screen. Several bugs have been corrected in Version 2.0.

Since the ability to overlay geo-referencing information represents a major revision to program **DISPLAY**, the program has been released as Version 2.0.

MERGEO: Merges two binary files of gridded data and/or extracts a rectangular window of the input data for output.

The utility, **MERGEO**, was originally designed to merge the contents of two input binary-gridded datafiles from contiguous map areas, resulting in a larger binary-gridded datafile as output. The utility was later modified to allow a window (defined by a range of geographic coordinates) of the merged dataset to be selected for output.

The binary grid-file type can be selected separately for the input and output datafiles. Currently, Geosoft (Versions I and II) and Geopak binary grid-file formats are supported. Thus, by specifying a different grid-file type for input than for output, **MERGEO** can be used as a data conversion utility.

In addition to the output file, which is specified on the command line, **MERGEO** allows the user the option of specifying either one or two input file names. If only one input file name exists on the command line, no data merging is performed and either the entire contents or that within a data window of the input file will be transferred to the output file. If two input files are specified, **MERGEO** merges data values from the two and outputs the result (or a portion within the specified window) to the output file. **MERGEO** uses the geographic coordinates of the grid origin, stored in the grid header line, to determine the relative positions of data from the two input files and the location of the optional window selected for output. Data values in areas of overlap on the two input grids are averaged before output.

The output data window is specified on the command line by the two: **-x=*;*** and **-y=*;*** flags, where **x** and **y** indicate respective grid dimensions and *;* specifies the minimum and maximum extremes of the data window expressed in geographic coordinates. Alternatively, the output data window may be specified in grid cells by using the **-nx=*;*** and **-ny=*;*** flags. If a coordinate of the specified data window falls outside the coordinate extremes of the input or merged gridded data, that window coordinate is set to the respective data extreme.

Several other features have been added to **MERGEO**. Basic statistics (i.e., minimum, maximum) and a frequency histogram are accumulated during the merging process and these are stored in the headers of the output grid files. **MERGEO** will merge and/or convert the formats of data stored in either of the row or column orientations used by Geosoft. A gap (represented by null values) in the input and/or merged datafile may be filled by applying a minimum curvature (Akima, 1972) interpolation from the values of points to either side of the gap. Interpolation (applied along the grid storage direction) can be selected by the command line flag **-s=n**, where **n** specifies the maximum gap width in grid cells.

The user's manual provides further details on the use of the **MERGEO** utility.

GRD_ASC: A conversion utility to dump all (or a portion) of a binary grid file to an ASCII file, or vice-versa.

GRD_ASC is a reformatting utility that provides a two-way link between binary-gridded datafiles and their ASCII equivalent. The routine was originally written in response to requests from those who purchased gridded aeromagnetic data released as digital Open Files and were viewing those files with **DISPLAY**. Several people expressed an interest in having a means to dump the data contents of a Geosoft or Geopak binary-gridded file to ASCII characters, so that data values could be viewed with any ASCII editor or sent to a printer.

During the design of **GRD_ASC**, its functionality was extended to include the ability to import existing datafiles stored in ASCII to a Geosoft or Geopak binary-storage format, so that the data may be viewed with the **DISPLAY** image generating utility. In its present form, **GRD_ASC** may be used to:

- Dump the entire data contents of a Geosoft or Geopak format binary-gridded datafile to ASCII.
- Select a window of a Geosoft or Geopak grid for export to ASCII. This reduces storage requirements and allows the user to quickly view data values only from the particular area of interest.
- Import a datafile, stored in ASCII, of gridded data values to a binary file in Geosoft or Geopak format. The ASCII input file could be data dumped from any other gridding or viewing software.
- Select a portion of an ASCII datafile for import to a binary-gridded datafile, in order to reduce size or processing time. Program **MERGEO** may later be used to piece smaller binary grid files into a large data file.
- Alter most of the descriptive grid parameters that are stored in the header lines of a Geosoft or Geopak binary datafile.

When operated in the data import mode, **GRD_ASC** also performs simple statistics on the selected data window and stores these along with a 32-interval frequency histogram of data values in the header file of the output binary grid file.

The functions and operation of **GRD_ASC** are controlled by a series of command line flags, a simple GXF header command file (in ASCII), or through a combination of both. The command line usage for **GRD_ASC** can be obtained by typing **'GRD_ASC'** at the DOS prompt. By default **GRD_ASC** operates in the data export mode. Users are cautioned that binary-gridded datafiles require considerably less storage space than their ASCII equivalent. Therefore, if data-storage space is limited, only smaller data windows from large binary grid files should be selected for export to ASCII.

The data import mode can be selected on the command line with the **-i** flag. When **GRD_ASC** is used in the import mode, parameters used for control are assigned default values. This enables the utility to be used to create a binary grid file from a generic ASCII datafile with minimal preparation by the user.

The format of all ASCII files exported from or imported to **GRD_ASC** should conform to the Grid eXchange File (GXF) format as described in Appendix G (Revision 1.0) of the Geosoft 2-D Mapping System user's manual (Geosoft, 1990). The GXF format was chosen because of its adaptability to a wide range of existing ASCII data formats. The GXF format was further revised for **GRD_ASC** to include such features as the ability to skip any number of descriptive header text lines in the ASCII datafile, or to skip any number of text lines at the start of each row of data.

The **GRD_ASC** utility is included with all requests for the **DISPLAY** utility. A nine-page, comprehensive user's reference guide (Kilfoil, 1991b) describes the various functions of **GRD_ASC** and the formats of the GXF and binary-gridded datafiles.

MASK: Compares two binary-gridded datafiles and replaces data values within the second with null values at all grid-cell locations where null values exist in the first.

MASK simply masks or blanks out areas in a gridded datafile whose locations correspond to null values in a second gridded datafile. Null values are usually placed in a gridded datafile to designate areas where no data exist. The interpolating routines used to create gridded datafiles often pad the areas outside the input data with null values in order to ensure that a gridded datafile is rectangular.

In the process of grid generation, grid values may be over- or under-estimated when interpolated across areas where no input data exists, such as across inlets, straits, large lakes or recreational areas. If a gridded datafile exists with such areas blanked out, **MASK** can reference null values in these areas to blank out portions of other interpolated data grids.

MASK is compatible with Geosoft or Geopak format gridded datafiles.

GRD_LEG: Replaces data within a rectangular area of a binary-gridded datafile with a range of values corresponding to a grid legend.

This utility is designed to insert a rectangular legend into a Geosoft or Geopak gridded datafile. **GRD_LEG** was originally developed as an addendum to the Geosoft mapping system, as a means of labelling colour maps, produced from gridded datafiles, with a bar-type legend showing the range and scale of data.

The bar-type legends are created by overwriting areas in a gridded datafile with false data values; therefore, the process is irreversible in that any data that previously existed in the area of the legend will be destroyed. Bar-type legends can be placed in the gridded datafile with their long axis oriented horizontally or vertically.

Two types of legends can be generated and inserted into a gridded datafile by **GRD_LEG**: one with discrete cells of a user-specified size or a continuous strip legend having data varying linearly from one end to the other. The data values to be placed in a legend with discrete cells are input from a small (ASCII) control file (equivalent to Geosoft's zone file), one value per line and usually arranged in ascending order. Each value in the control file will produce one cell in the legend to a maximum of 50 cells. In the case of a continuous strip legend, the values placed in the legend will be linearly stretched from the values of the first to the last data entries in the control file, since these are assumed to be the approximate minimum and maximum data values, respectively, for the grid. The size of each cell in the legend is determined by the command line flag **-c=x,y,** where **x** and **y** are sizes expressed in grid cells for the two dimensions. Therefore, the long dimension of a legend (in grid cells) is determined by multiplying the length of each legend cell by the number of values in the control file. If the legend is longer than the corresponding grid dimension, the user is notified and no legend will be created.

The locations of legends within the gridded datafile is controlled by the **-o=x,y** flag, specifying the legend origin (**x** and **y** refer to numbers of grid cells from the grid origin) as a command line option. Legends are usually placed in the gridded datafiles within areas of null values. If a location outside the dimensions of the gridded datafile is specified, the output grid file will be automatically expanded to accommodate the legend. This option is useful in cases where a grid file contains little or no areas of null values.

Specification of an output grid-file name when using **GRD_LEG** is optional. If only one grid-file name (the input file) is included on the command line, the legend will be inserted into the input grid file; if an output file is also named, the contents of the input grid will be written to the output grid file together with the legend, and the input file will remain unchanged.

Gridded Data Filtering

The amount of local variation (including noise) contained in a dataset can be varied during the grid interpolation process by the gridding parameters chosen. Grid appearance and accuracy (to the original data) can be controlled by grid cell size, search radius, neighbourhood selection pattern and interpolation algorithm. Gridded datasets that are destined for a specific purpose, such as contouring, can be created using the initial gridding parameters to produce smooth contours. An alternative method of producing smooth contours from gridded data is to choose initial gridding parameters that will produce an output grid that honours input data points and retains as much of the original information as possible. This detailed grid can be filtered to produce output grid files that can be used for contouring, outlier detection, anomaly detection, smoothing of local variation, edge detection as well as other specialized applications. The aim in most filtering situations is to minimize the background noise while enhancing the data signal.

A set of filter programs were written in FORTRAN-77 on a IBM-PC and are designed to read and write the binary grid-storage formats, which are commonly used by geochemical and geophysical software packages. These programs do not require the computer system to have graphics capabilities but, due to the mathematical calculations involved, a math co-processer is recommended for speed considerations.

MEAN: Applies a Mean filter of variable-width to a binary-gridded datafile.

The most common grid filter is the moving average or rolling mean filter, which calculates a new grid based on the mean of a n by n window for each cell in the detailed grid. There are several variations of the mean filter (low pass, hanning, etc.) but they are all very similar in operation and results. The amount of smoothing is controlled by the size of the window (3 by 3 is common) and the number of times it is passed over the grid. Mean filters diminish localized high/low values to enhance regional patterns in the grid. Mean filters are used to improve the appearance of the display data in cases where the loss of localized information is not critical. Mean filters may be strongly affected by extreme outliers in the data, which in many cases are due to errors.

MEDIAN: Applies a Median filter of variable-width to a binary-gridded datafile.

The median (the midpoint value of a dataset) is replacing the mean (the average of a dataset) as a descriptive statistic that better describes the central tendency of a given set of data values in many cases. Similarly, in grid filtering methods the median and mean filters are commonly used to filter grids to remove spurious values. The median filter does have some advantages over the mean. Edges and sudden changes in data values between grid cells are retained by the median filter, but small scale variation is largely removed.

ATM: Applies an Adaptive Trimmed Mean (ATM) filter of variable-width to a binary-gridded datafile.

The mean and median filters each have common and unique features that are optimum in specific situations. The ATM filter (Chork and Mazzucchelli, 1989) combines features from both techniques to remove background noise, retain edges and sudden changes in data values and strip outliers from the data. As the name suggests, the filter calculates the mean using the median as a robust measure to remove outliers from the active window. Only cell values that fall within approximately 2 standard deviations (computed from the median absolute deviation about the median) of the median are used to calculate the mean for the cell window. The degree of generalization produced can be controlled by increasing the size of the active window area (3 by 3, 5 by 5, 7 by 7, etc.) beginning with the average sample density of the original survey as the smallest window that will have any distinguishable effects. This approach can be quite useful for producing geochemical grids, which emphasize geological structure (Davenport and Nolan, 1991) while still retaining anomalous signatures.

FENCE: Applies a Fence filter of variable-width to a binary-gridded datafile.

The majority of values in a regional geochemical data set comprise the geochemical background due to variations in bedrock lithology, and a minority of these values are anomalies related to mineralization. Depending on the element, this signal-to-noise ratio will vary. For instance, gold data is generally noisy due to the small-scale variance and most samples fall below the analytical detection limit. The filtering routines discussed so far remove the outliers from the dataset, whereas the Fence filter (Chork and Mazzucchelli, 1989) highlights the anomalous regions. Using basic Exploratory Data Analysis (EDA), the sample median of a central block of cells with **n** by **n** area is compared to the upper and lower quartiles of a annulus ring of cells of given radius (fence width) from the central cell being calculated (Figure 2.). In general, the **FENCE** filter measures the strength of a localized anomaly compared to its surroundings. If the median of the central cells is greater than the upper quartile (75th percentile) of the fence cells, based on robust statistical measures of the fence cells, then the central cell is flagged as anomalous. The Fence filter will flag anomalous cells in areas of relative low as well as high data values. The degree of generalization produced by the filter can be altered by varying the size of the central block of cells and the number of cells in the fence area. The cells that fall between the central cells and the fence ring are excluded from the computations.

Filtering Procedures

The filtering utilities are implemented and controlled through a command line parameter list of input binary grid file, output binary grid file, file type (Geosoft Version I or II, or Geopak), window size in grid cells and, for the **FENCE** filter, the size of the exclusion fence in grid cells. Each filtering utility will provide a help screen if the filter name is entered without a command line parameter list. Experience with these filters has shown that combinations of different filters may be passed over a grid in succession to obtain the desired results. For example, an **ATM** having a wide window followed by a **MEAN** with a narrow window will result in

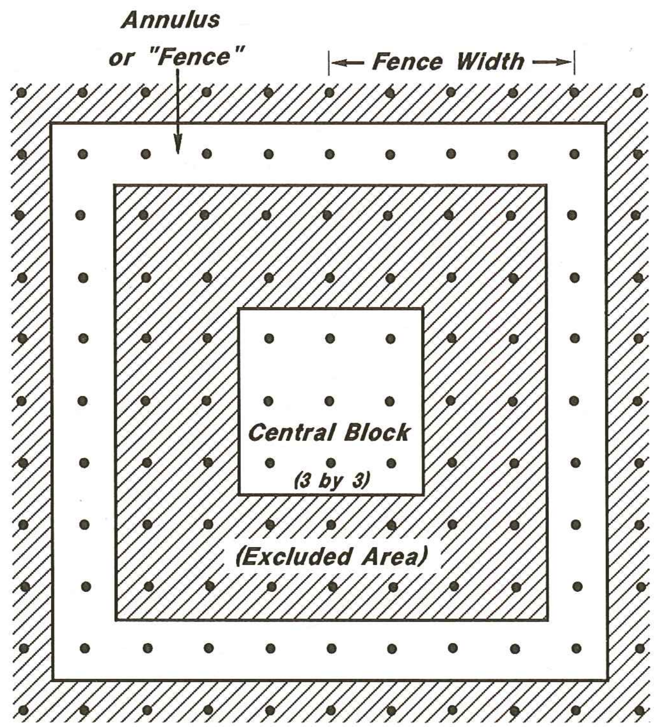

Figure 2. *Schematic diagram of a Fence Filter where dots represent the centres of grid cells. The central block has an effective radius of 1 (3 by 3) and the fence width is 4 grid cells.*

a output grid that retains large scale features (**ATM**) in the data where the transitional areas between features will be smoothed to improve the appearance (**MEAN**). When choosing the filter window size, consideration should be given to the sample density of the original data. Using a window that is smaller than the original sample density will result in a grid that is very similar to the input grid. The selection of the window size can be aided by using programs such as **GRIDVIEW** to view sample density (smallest window) and variograms to show variation of sample values over distance. However, some experimentation with window widths and filter types may be required before an optimum is reached.

The principles of filtering logic as applied to exploratory data analysis have been discussed by Chork and Mazzucchelli (1989).

CONCLUSION

The functions of several PC-based computer utilities developed at the Geological Survey Branch and adapted for general usage are described above. Many other data-handling utilities have been developed, but their functions are very specific and tailored to the software and hardware that exists within the Branch. These routines have not yet been generalized to handle a variety of computer configurations that may exist in the public domain. As those utilities are updated and as others are developed, their descriptions will be described in future Current Research papers.

The utilities listed in this article can be made available in executable form on a floppy diskette at a nominal cost of reproduction. Requests for copies of many of these utilities will be accompanied by brief documentations describing the function and usage of each.

ACKNOWLEDGMENTS

Dr. Peter Davenport is thanked for critically reading and suggesting improvements to this manuscript.

Appendix A

The TYDAC SPANS GIS Vector Interchange Format

The TYDAC SPANS Vector Interchange Format requires that the first line of an ASCII vector file consist of the left justified string 'ARCS'. This is followed by each vector line segment, one after another. The first line of each vector line segment representation is a header line, which must have a minimum of four integer parameters: the vector segment number, the number of points (n), the level designation, and the feature code. The header line parameters must be separated from each other by blank spaces. This header line is followed by coordinate pairs (**X** and **Y**), one coordinate pair per line. Thus, each vector line segment requires **n+1** lines in the ASCII file, where **n** is the number of points in the line segment.

The example that follows is a SPANS Vector Interchange Format file containing two closed polygons: the corner coordinates, in metres within UTM Zone 21, for the 1:50 000-scale NTS 2D/4 and 2D/12 map areas, respectively. The comments (within brackets) in this example file are not required.

```
ARCS                          (Indicates a Vector)
    1      5     0   9963     (Segment Header for NTS
                              2D/4 Outline)
574597.46  5316565.14
574236.32  5344351.71         (Last Coordinates in
611353.68  5344955.94         Segment are Equal to the
611895.46  5317169.95         First: Closed Polygon)
574597.46  5316565.14
    2      5     0   9963     (Segment Header for NTS
573873.75  5372139.46         2D/12 Outline)
573509.76  5399928.40
610263.73  5400531.36
610809.77  5372743.08
573873.75  5372139.46
```

The above file could be used by **DISPLAY** to overlay the two map-area outlines on colour images generated from a gridded datafile. To enable **DISPLAY** to overlay vector files properly, the geographic coordinates of points in vector line segments must be consistent with the coordinate system to which the grid file is referenced. For instance, the above vector file would overlay correctly on images generated from the gridded aeromagnetic data for Newfoundland, released as a digital Open File of binary-gridded datafiles (Kilfoil and Bruce, 1990), which were also referenced to the UTM coordinate system in zone 21.

REFERENCES

Akima, H.
1972: Interpolation and smooth curve fitting based on local procedures. Communications of the ACM, Volume 15, Number 10, pages 914-918.

Chork, C.Y. and Cruikshank, B.I.
1984: Statistical map analysis of regional stream-sediment data from Australia. Journal of Geochemical Exploration, Volume 21, numbers 1-3, pages 405-419.

Chork, C.Y. and Mazzucchelli, R.H.
1989: Spatial filtering of exploration geochemical data using EDA and robust statistics. Journal of Geochemical Exploration, Volume 34, number 3, pages 221-243.

Davenport, P.H. and Nolan, W.L.
1991: Shaded-relief maps of geochemical data to delineate geological structure. *In* Current Research. Newfoundland Department of Mines and Energy, Geological Survey Branch, Report 91-1, pages 263-269.

Davenport, P.H., Nolan W.L. and Kilfoil, G.J.
1991: Image-processing of geochemical data to map geological structures and alteration zones. *In* Exploration Geochemistry, 1990. Czech Geological Survey, Prague, pages 57-61.

Geosoft Inc.
1990: Geosoft 2-D Mapping System, User's Manual and Appendices. Geosoft Inc., Toronto, Canada.

Herkommer, M.A.
1986: Techniques devised to improve contour mapping by microcomputer. Geobyte, Spring 1986, pages 38-40.

Hoffman, S.J.
1986: Writing geochemical reports. The Association of Exploration Geochemists, Special Volume No. 12, 29 pages.

Kilfoil, G.J.
1990: DISPLAY—a computer program for the interactive display of files of gridded data on a VGA monitor, computer program (Version 1.0) and User's Manual. Newfoundland Department of Mines and Energy, Geological Survey Branch, Open File NFLD (2064), 24 pages.

1991a: DISPLAY: An interactive program to create colour images of spatial data on a PC. *In* Current Research. Newfoundland Department of Mines and Energy, Geological Survey Branch, Report 91-1, pages 271-277.

1991b: GRD_ASC: A routine to convert gridded datafiles stored in binary formats to ASCII, or vice-versa (Version 1.0), Reference Manual. Newfoundland Department of Mines and Energy, Geological Survey Branch, (unpublished manuscript), 9 pages.

Kilfoil, G.J. and Bruce, P.A.
1990: Regional aeromagnetic data grids (200 m grid cell), in digital form, for all of insular Newfoundland, Version 1.0. Newfoundland Department of Mines and Energy, Geological Survey Branch, Open File NFLD (2063).

Microsoft Corporation
1988: Microsoft QuickBASIC Programmer's Manual, Version 4.5. Microsoft Corporation, Redmond, WA, USA, 458 pages.

1989: Microsoft FORTRAN Programmer's Reference, Version 5.0. Microsoft Corporation, Redmond, WA, USA, 523 pages.

Nolan, W.L.
1990: UNISTAT: Statistical and graphics package for geochemists. *In* Current Research. Newfoundland Department of Mines and Energy, Geological Survey Branch, Report 90-1, pages 129-130.

Rock, N.M.S.
1987: ROBUST: An interactive FORTRAN-77 package for exploratory data analysis using parametric, robust and non-parametric location and scale estimates, data transformations, normality tests, and outlier assessment. Computers and Geosciences, Volume 13, pages 463-494.

Sinclair, A.J.
1981: Applications of Probability Graphs in Mineral Exploration. The Association of Exploration Geochemists, Special Volume No. 4, 95 pages.

1983: Univariate analysis. *In* Handbook of Exploration Geochemistry, Volume 2, Statistics and Data Analysis in Geochemical Prospecting. *Edited by* R.J. Howarth. Elsevier Scientific Publishing Company, Amsterdam, 1983, pages 59-81.

Stanley, C.R.
1987: Instruction Manual for Probplot: A computer program to fit mixtures of normal (or log-normal) distributions with maximum likelihood optimization procedures. The Association of Exploration Geochemists, Special Volume No. 14.