Using Data Loss and Duplication Maps to Study Data Use Associated With Nearest-Neighbor Resampling

By

Matthew Perigny

A research paper submitted to

**The Department of Geosciences,
Oregon State University**

In partial fulfillment of the
Requirements for the
degree of

**Master of Science
Geography Program**

December, 2003

Directed by
A. Jon Kimerling

Acknowledgments

There are many people that I would like to thank for providing support and encouragement to me while at Oregon State University. I will not attempt to list all of these people since I would undoubtedly omit some of them. However, there are a few individuals who have greatly contributed to my experience here and have made my completion of studies here possible.

I would especially like to thank my major professor Dr. Kimerling for directing me through my research and providing valuable insights into the development of my studies as well as my research paper. I would also like to thank my commitee members, Dr. Jackson, and Dr. Wright, for their support and patience during the course of my studies at OSU.

Finally, I would like to thank my parents, Ralph and Shirleen Perigny for giving me the support and encouragement needed to accomplish this goal as well as many others in my life.

Table of Contents

Using Data Loss and Duplication Maps to Study Data Use Associated with Nearest-Neighbor Resampling

ABSTRACT

Nearest-neighbor resampling is a common procedure used by cartographers to reproject raster data sets. This type of resampling is useful because it retains the data from the original raster data set. However, during nearest-neighbor resampling some of the original data will be lost as it is reprojected, while other data will be duplicated in the reprojected data set.

This paper discusses a method for studying and displaying the amount of data loss and duplication associated with nearest-neighbor resampling on equal-angle data sets, such as the ETOPO5 data set. The method uses a computer program, developed by the author, which performs nearest-neighbor resampling on equal-angle data sets. The program produces a tool for studying data loss and duplication, called the Data Loss and Duplication Map, and gives visual and statistical information about the data loss and duplication associated with reprojecting equal-angle data sets using various map projections.

This paper will examine the development of this method and will serve as a manual for using the computer program written for studying Data Loss and Duplication Maps. Finally, the paper will discuss the results of using this method, the limitations of the method and the program, and possible direction for future studies using this method.

INTRODUCTION

Resampling of raster data sets and images is a procedure used in many cartographic and geographic information system (GIS) applications to fit the data or image to a raster grid of different size or map projection. A raster data set has values, such as elevations, that are regularly spaced over a geographic area. Each data value is indexed by its (x,y) location on the grid and its geographic position on the ground. An example of a raster data set is the ETOPO5 which has elevation values on the earth spaced at every 5 minutes of latitude and longitude. This data set is organized into a grid of 4320 columns and 2160 rows with each column representing a line of equal longitude and each row representing a line of equal latitude. Each (column, row) cell in the data set holds the elevation at the corresponding latitude and longitude.

Resampling raster data sets differs from regular map projection transformations used on continuous data sets (such as vector data sets). Continuous data sets can be transformed exactly from one projection to another by using map projection transformation equations to change the x,y coordinate of each vector point. However, this cannot usually be done with raster data sets because they need to completely fill a new regular grid. Because of this, the distortion and error properties associated with raster resampling transformations is different than those associated with continuous data sets.

Nearest-neighbor resampling, called resampling in the rest of this paper, is used to fit a raster data set to a map projection grid. It involves first selecting a map projection to transform the data set to, which will provide a new cartesian (x,y) grid of user defined density. Each (x,y) point in the grid represents a certain latitude and longitude, as defined by the projection. In nearest-neighbor resampling, the data value in the original grid that is closest in latitude and longitude to (x,y) is chosen to represent the elevation at that point. This results in a new grid of data values on the chosen map projection. Nearest-neighbor resampling is not a one-to-one mapping of the raster data set. There typically are parts of the data set that are never selected (data loss) and other parts that are used more than once (data duplication). There has been little research devoted to studying the distribution of data loss and duplication with different map projections.

There have been a few studies of how data loss and duplication due to raster resampling occurs on different map projection transformations. A fairly straightforward tool recently developed to look at data loss and duplication is called the Data Loss and Duplication Map (DLDM) (Kimerling, 2002). The DLDM is simply a map that shows which raster cells from the original data set were sampled in the resampling process, and how many times each cell was sampled. The DLDM gives a variety of information about the sampling distribution of a map projection, including information about overall data loss and duplication, and the pattern of cell selection, as well as being a method for comparing projections and comparing the effects of sampling density and scale changes.

DLDM's will also be useful for studying the efficiency of different global data structures. Currently, global elevation data sets such as ETOPO5, ETOPO2, and GLOBE are stored as equal-angle grids, meaning that elevation data are stored in grid cells defined by equal increments of latitude and longitude. The equal angle grid is a computationally efficient method for storing data, since the earth's surface can be represented as a rectangular grid. However, for processing data, this grid may be inefficient, because meridians converge toward the poles so that grid cells terminally decrease in area from the equator to the poles. For example, in the ETOPO5 grid (5' intervals between data points), there are 4320 rectangular cells approximately 221,615 km^2 in area, whereas next to the poles there are 4320 triangular-shaped cells approximately 194 km^2 in area.

DLDM's can serve as benchmarks for studying the effectiveness of different global data set designs. DLDM's show potential for giving raw statistical information on the geometrical nature of resampling onto different map projections. All DLDM's are expected to have different spatial patterns of data loss and duplication, and each projection studied so far has shown distinctly different patterns. Some DLDM's appear to have fractal-like Moire intereference patterns when viewed at certain scales. Studying the patterns of data loss and duplication on DLDM's can lead to increased efficiency and accuracy for projecting data since it will allow for concentrating on improving the accuracy of grid cells that are more likely to be sampled with a particular map projection.

This paper looks at the development of the Resampler program as a tool for creating and studying Data Loss and Duplication Maps. The Resampler program was developed to aid in studying DLDM's and the data loss and duplication associated with resampling the ETOPO5 global elevation data set onto various map projections. A description of what the program was written to do and how this aids in studying DLDM's, as well as a description of how to use the program, comprise the main sections of this paper.

THE RESAMPLER PROGRAM

Resampler was written in Visual C++ version 4.0 as a tool for resampling equal-angle grid data sets, and for creating Data Loss and Duplication Maps for viewing and analysis by the user. The program was designed to study the effects of resampling on the ETOPO5 global elevation data set. ETOPO5 is a row-major order, two-byte binary integer data set with no header. Resampler should work for any data set with these properties.

There are three main objectives of the Resampler program. The first is to serve as a tool for reprojecting equal-angle data sets onto different map projections. The second objective is to create a Data Loss and Duplication Map for each resampling, and to provide statistical data about the DLDMs produced. The final objective is to provide a display of the DLDM for visual analysis by the user.

The user can choose from nine different map projections provided in the program to reproject the source data. The reprojected output file is created in the same format as the input file (row-major order, binary two-byte integers), and can be used (with a header file) in many GIS and image processing programs, with an appropriate header file. The user has many options to choose from for each map projection, including changing the scale, the earth's radius, and the standard parallels of the projection, variables that affect the size and density of coverage for the output file.

Creating the Data Loss and Duplication Map for each resampling performed with the program is a fairly simple, but important feature of the Resampler program. The DLDM is created using the same procedure as described in Kimerling (2002). DLDMs are created by starting with a regular grid of the same size as the original data set, with all cells initialized to zero. Each time a cell in the original data set is sampled, the corresponding cell in the DLDM grid is increased by 1. When resampling is complete, each cell in the DLDM will show the number of times it was sampled in the resampling process. Important statistics relevant for the DLDM are given, such as overall data loss and duplication at the end of each resampling. The user also has the option of saving the

DLDM for further analysis or display as either a two-byte integer raster file or a portable pixel map (.ppm) file.

Resampler also allows the user to look at the DLDM at three different scales, and gives a glimpse of the order in which cells are selected as the DLDM is created. The user has the option of seeing the DLDM as it is created, which can be helpful in studying the pattern of data loss and duplication. Because showing the DLDM in its entirety (1 pixel represents 1 grid cell) is difficult for most computer monitors (a 1024x768 monitor could only show eight percent of an ETOPO5 DLDM), the user is provided with three different zoom levels, providing an overall view of the DLDM as well as the ability to examine individual cells in the DLDM.

USING RESAMPLER

Resampler was designed to be an easy to use program for performing nearest-neighbor resampling on equal-angle grid data sets. There are nine different map projections available to the user for reprojecting from equal-angle data sets. There are several output options for a user to choose from depending on his or her needs, including the resampled output file, a Data Loss and Duplication Map, and statistics describing the DLDM. This paper gives instructions on how to use the Resampler program, as well as its capabilities and limitations.

Starting Resampler

To start the Resampler program, open the program file Resampler.exe. This file is located on the CD that accompanies this paper, and can be moved to any location on your hard drive. When this program first starts, the main Resampler window appears as shown belo (Figure 1):

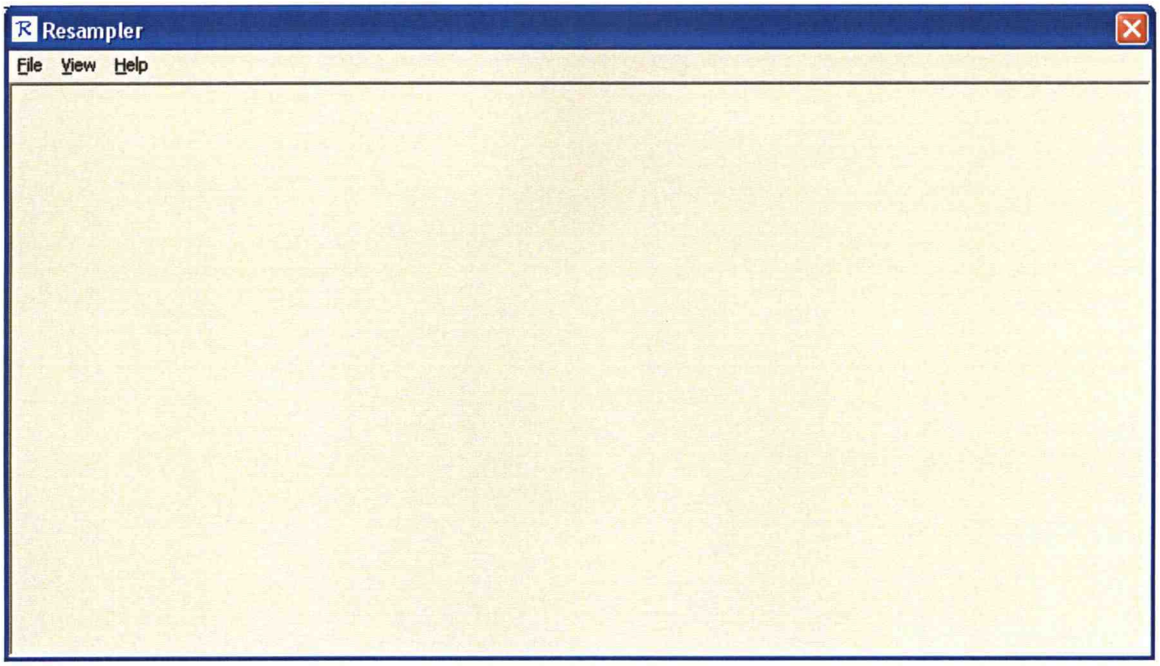


Figure 1. Resampler window at start up.

The main program window has two function areas: the control menu and the display area. The control menu is used for loading data files to be resampled, running the Resampler program, and adjusting the view options. The display area is used to show progress in the resampling procedure, display a preview of the reprojected data set, and display the Data Loss and Duplication Map produced during resampling. The display area will be blank (gray) until a resampling operation is performed.

The Control Menu

The control menu is used for operating Resampler. This menu is divided into three sub-menus: the File menu, the View menu, and the Help menu. Clicking on one of these sub-menus will allow you to do several things, such as opening data files for resampling, starting the resampling procedure, saving DLDMs produced by Resampler, and changing your display settings for the display area. The functions available in the menu are given in Table 1.

Table 1. – Menu Commands

Menu	Name	Function
File Menu	Open Data File	Select data file to resample
	Run Resampler	Starts resampling procedure
	Save DLDM	Saves DLDM as .ppm graphics file
	Save DLDM as .bin	Saves DLDM as binary file
	Exit	Exits program
View Menu	Map	Shows a preview of the resampled data
	DLDM	Shows DLDM in display area (default)
Help Menu	Instructions	Lists instructions for using Resampler
	About	Displays information about the program

Resampling

This program was designed to resample the ETOPO5 global elevation data set, an equal-angle, row-major order, two-byte low-byte-first order integer binary data set, which is often labeled as ETOPO5.dos. This file is produced by the National Geophysical Data Center (NGDC) and consists of 2,160 rows and 4,320 columns for a total of 9,331,200 integer data values or 18,662,400 bytes. Any equal angle grid of less than 6,500 rows and 6,500 columns should work with the program, but it has only been tested with the ETOPO5 data set.

To run Resampler, you need to either use ETOPO5 or provide your own equal-angle data set. Once you have this data set on your computer and Resampler.exe is opened, select from the File menu Open Data File (Figure 2). This will open a window where you can select the data file that you wish to resample. Select your data file by clicking on the file name (you can use the controls available on this window to navigate to different locations on your computer) and then clicking the Open button.

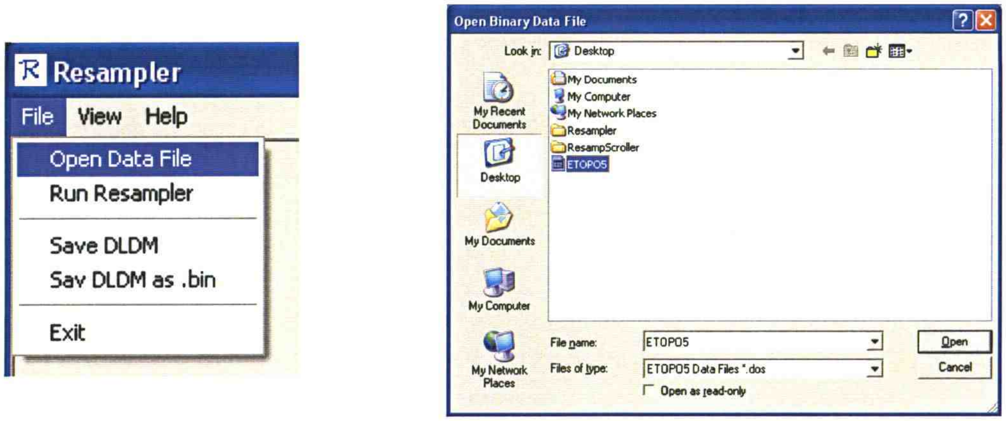


Figure 2. File Menu and Open Data File Window

Once you have opened your input file, you are ready to begin the resampling process. Select the Run Resampler option from the file menu to begin resampling. This will open a window where you can choose the name of the output file that you wish to create. This file will store the output in the same format as the input file (row-major order, two-byte integers). Here you have the option of saving over an existing file or creating a new file for the output. The size required for this output file will depend on the variables you select for resampling later on.

From this point on, a series of dialog boxes will appear asking you to enter information about the input file, the output file, and the map projection that you wish to resample to. You can navigate through each edit box (where you input information) by using the Tab key to cycle through the edit boxes and enter information by typing numbers into the boxes. Be careful to enter the correct information, since once the information is entered you must complete or terminate the procedure. After you have selected your map projection, you cannot end the resampling process. However, prior to this you can end the proceess by not selecting a map projection in the map projection selection dialog box.

Input File Window

The Input File Window (Figure 3) asks for variables related to the input data file. These are standard properties of all equal-angle grid data sets and can be found in the header file of you data set if included. The ETOPO5 global grid variables have been selected as the default variables for this program, so they only need to be changed if a different grid is used. After you click on the OK button to enter you input grid properties, you may have to wait for a short amount of time before the next window appears, depending on the speed of your computer.

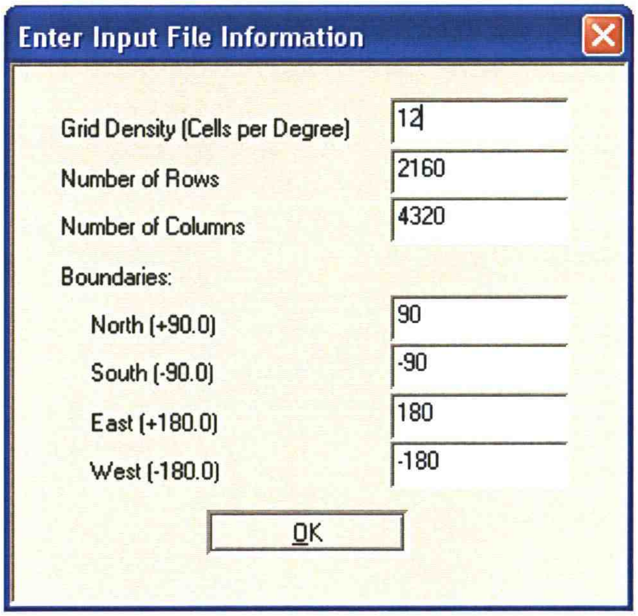


Figure 3. Input File Window

The required inputs here are the number of rows and columns in the grid, and the geographic boundaries in the grid. Since the grid is an equal-angle grid, it should be "rectangular" in latitude and longitude, meaning that all values in each row have the same latitude and all the values in each column have the same longitude. The boundary values must be between –180 and +180 longitude and –90 and +90 latitude with ranges of less than or equal to 360 and 180 for longitude and latitude, respectively.

Select Projection Dialog Box

Once the properties of the input data set have been entered, the Select Projection dialog box (Figure 4) will appear. You can select a projection from the projection list in the drop-down menu on the dialog box. There are nine projections available, all of which use the spherical equations found in Map Projections: A Working Manual (Snyder, 1987). The available projections are the Azimuthal Equidistant, Cylindrical Equal Area, Lambert Azimuthal Equal Area, Mercator, Mollweide, Orthographic, Stereographic, Sinusoidal, and Transverse Mercator. You must select one of the projections from the drop-down menu and make sure it is typed exactly as in the menu or you will need to restart the resampling process. Select OK to continue to the next dialog box.

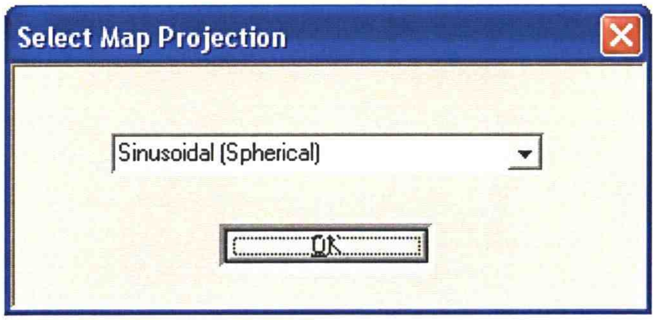


Figure 4. Select Projection Dialog Box

Projection Properties Dialog Box

The last dialog box that appears after you select "Run Resampler" is the Projection Properties Dialog Box (Figure 5). This dialog box is separated into two areas: 1) variables that affect the density of resampling for the output file, and 2) variables that define the map projection that was selected in the previous dialog box. The available projection variables that appear are dependent on the projection that was selected.

There are two variables that affect the resampling density of the output file: the map scale, and the image density in dots per inch (dpi) for the output file. The map scale is the true scale at the point of the line of tangency of the map projection used. For this program, the map scale is entered as a ratio of 1:n where n is the scale denominator. For example, enter 63600 in the scale edit box for a scale of 1:63,600.

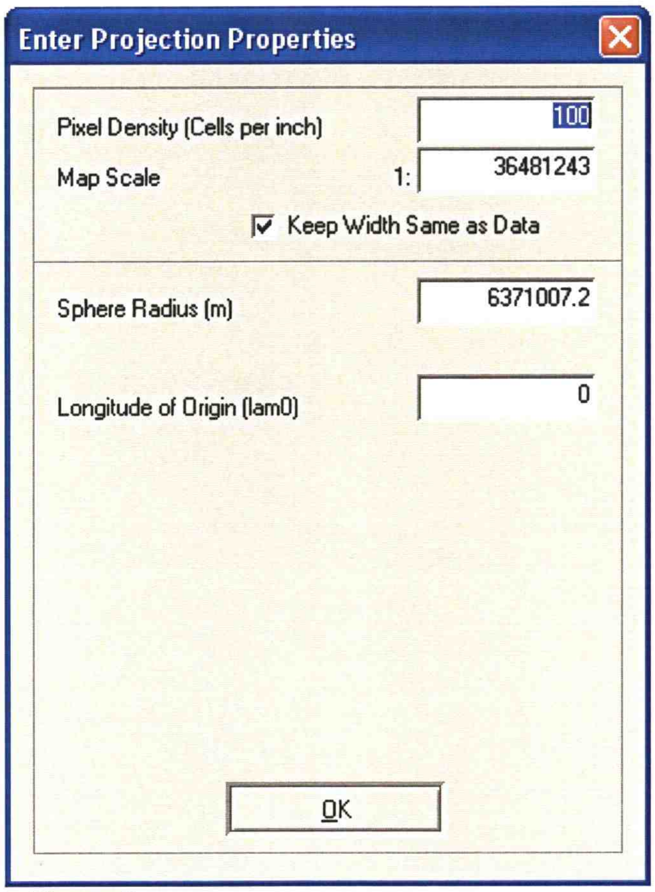


Figure 5. Projection Properties Dialog Box.

The Pixel Density edit box reflects how the output file will be displayed by the user. Pixel Density and map scale are related in terms of what the output file will look like. Increasing the pixel density has the same effect on the output file as decreasing the map scale (increasing the map scale value in the edit box). For example, the output file with a pixel density of 100 dpi and map scale of 1:32,844,650 will be the same size as a file with pixel density 10 dpi and map scale of 1:3,284,465.

For analyzing the DLDMs produced by different map projections, it is important to have consistency among the DLDMs produced. The program allows the user to give all output files the property that the number of columns in the output file are equal to the number of columns in the input file. This property was chosen to be consistent with the study done by Kimerling (2002) on the viability of using DLDMs for analyzing map projection reampling. To ensure this property in the program, there is a "Keep Width Same as Data" check box. When this box is checked it will ensure that the map scale and pixel density are such that the number of columns in the output file are equal to the number of columns in the input file. The default setting for the program is that the box is checked.

The lower portion of the Projection Properties Dialog Box contains variables specific to the projection selected, as well as an edit box to define the Earth's radius in meters for the projection. Variables available depend on the type of projection and include the standard parallels, and latitudes and longitudes of the parallel of origin. Select OK after the appropriate scale values and variables have been entered.

The Resampling Procedure

After you have clicked the OK button of the Projection Properties Dialog Box, the resampling process begins. While Resampler is creating the binary output file, it will show the status of the resampling by drawing in the display area either the DLDM or the map preview. This process will likely take a few minutes, depending on your computer speed and the size of the output file.

The resampling process itself is a very direct procedure, and is outlined in pseudocode in Appendix A. The process begins by first creating an M-by-N array to store the original data set, called ingrid[M][N]. For example, with ETOPO5, ingrid is 2160x4320, and the elevation for 90°N, 180°W is stored in ingrid[0][0]. Then, using the selected projection and variables associated with it, the output file size and boundaries are determined and an array, outgrid, of size X_{out}-by-Y_{out} is created. Following this, for each x,y cell in the output grid the inverse map projection equations are used to determine the latitude and longitude of the corresponding point on the map. The latitude and longitude values are then compared to the input grid, and the input grid value m,n that is closest in terms of latitude and longitude is selected for point x,y. If the latitude and longitude of point x,y is not within the bounds of the input file, then point x,y gets the value –0. For azimuthal projections two maps will be created in the output file, one map centered at the origin specified in the Projection Properties Dialog Box and one map centered at the point on the earth farthest away from the origin specified (at the antipodal point). Each map will include half of the globe, and the two maps will combine to show the entire earth. See Appendix B for notes and descriptions of the nine map projections included in Resampler.

Creating the DLDM is a simple addition to the resampling process. An additional array of the same size as the input file (M-by-N) is created, called DLDM[M][N]. All elements of DLDM[M][N] are initialized to equal 0, and whenever a data value from ingrid[m][n] is selected, the value of DLDM[m][n] is increased by 1.

The resampling process is finished when the value for each x,y point in the output grid has been determined, and the DLDM is created. When these are finished, a message box will appear saying that the program is done resampling. When the OK button on this message box is clicked, a new message box will appear with basic statistics about the output grid and the DLDM, including the number of rows and columns in the output grid, and the data loss, use, and duplication, triplication, etc. shown in the DLDM (Figure 6).

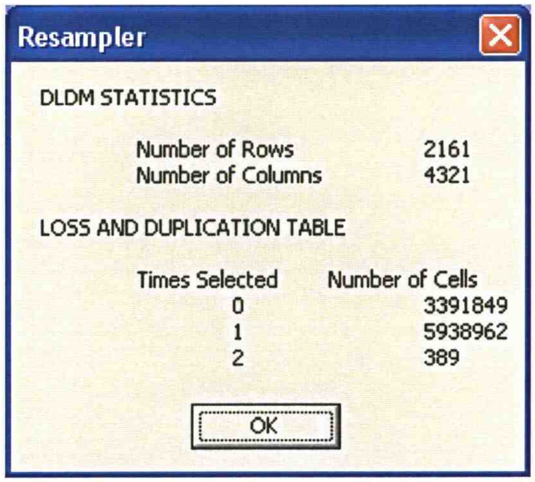


Figure 6. Message box showing DLDM statistics

The View Menu Options

Resampler's display area provides the user with displays for three reasons. The first function of the display area is as a status bar to show that the program is running and to estimate the program's progress. The second function is to preview the resampled map projection, while the third function is as a simple graphics tool to examine the Data Loss and Duplication Map.

Once the resampling process begins in the Resampler program (after clicking OK in the Projection Properties Dialog Box), there is usually a lengthy delay while the resampling is being performed before the next window appears. To let the user know that the program is still running, and to allow the user to check on the progress of the resampling, the display area will show the progress being made in creating the new map as each pixel in the new map is resampled. The changes will be shown as either map updates or DLDM previews in real time, meaning the user can "see" the map or DLDM being built. The user has the option of displaying either the DLDM (Figure 7), or the map preview (Figure 8) during the resampling process, but must choose which to show before selecting Run Resampler (DLDM is default).

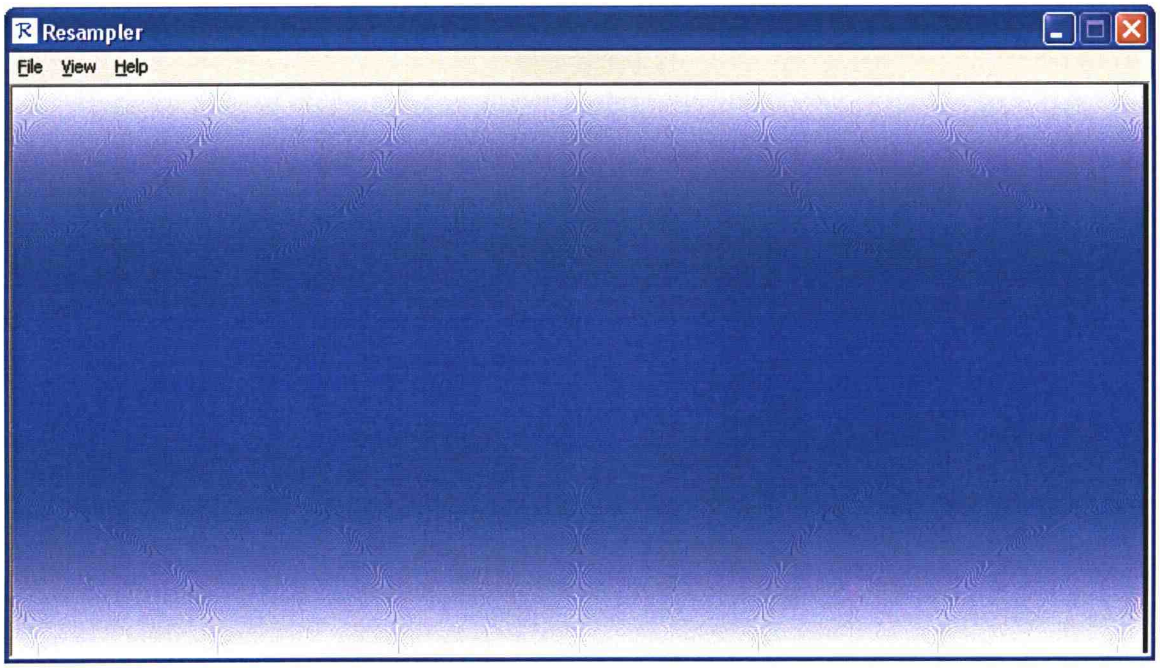


Figure 7. Sinusoidal ProjectionDLDM

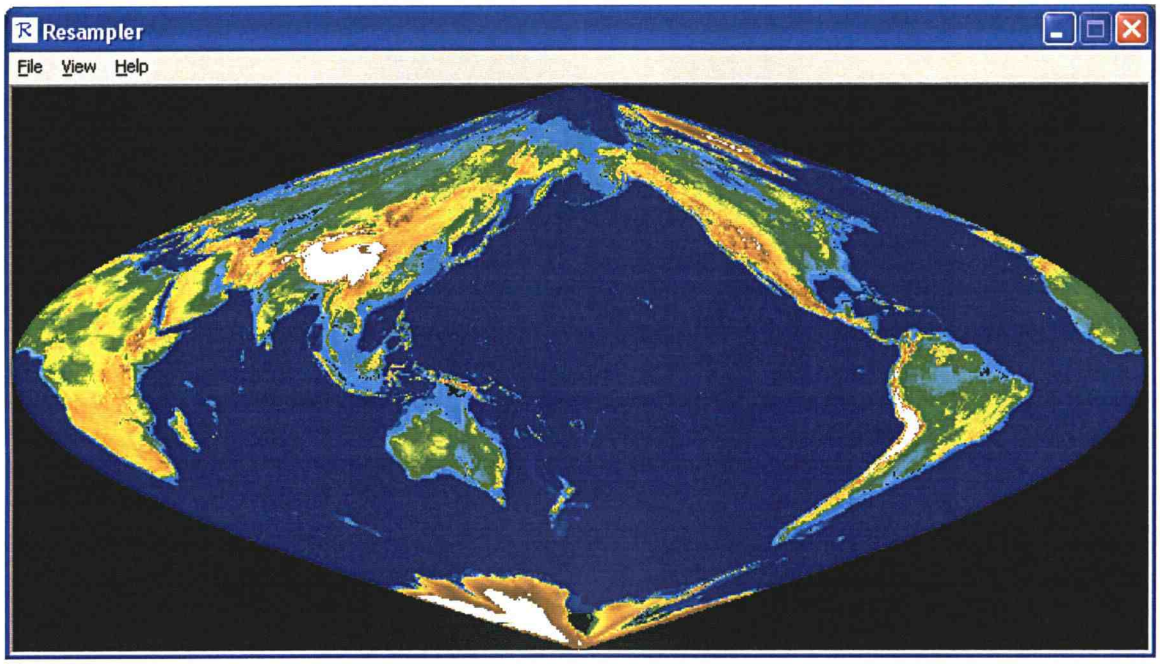


Figure 8. Sinusoidal Projection map preview

As stated above, the user has the option of seeing either a preview of the resampled map or the DLDM. To select the item (map or DLDM) that you wish to display in the display area, click on the View menu in the menu bar and select either DLDM or Map. Whichever view was selected last will be displayed while the resampling is taking place. Once the resampling is completed, the user may use the View menu to toggle between the two views.

The DLDM option in the View menu displays the DLDM as it is being created, and allows for viewing the DLDM at three different zoom levels. The DLDM preview in the display area is color coded for ease of user interpretation. Cells in the DLDM are colored by how many times they were used to create the output map according to Table 2.

Table 2. – DLDM Colors

Cell Usage	**Color (Binary RGB Color)**
Not used	White (1,1,1)
Used once	Blue (0,0,1)
Used twice	Red (1,0,0)
Used three times	Green (0,1,0)
Used four times	Yellow (1,1,0)
Used five times	Magenta (1,0,1)
Used more than five times	Gray (0.5,0.5,0.5)

During resampling, the DLDM is displayed at a 1/36th scale, where each pixel in the display area represents a 6-cell by 6-cell area in the DLDM (Figure 7). This scale allows the entire DLDM to be represented in the display area, and each pixel in the display area is colored by the average value of the 36 cells it represents.

After resampling is completed, the user can zoom into various sections of the DLDM. Counting the initial 1/36th scale zoom stage, there are three zoom levels available: 1/36th, 1/9th, and 1/1 scale. At the 1/9th scale, you can look at any one of the four quadrants of the DLDM. Single left clicking anywhere in the display area at the initial 1/36th scale will show the locations of the quadrants by displaying white lines that divide the display area (Figure 9a). To zoom in to any of the quadrants, move your mouse pointer over the quadrant and double click. This will show the same area at 1/9th

scale with each pixel representing a 3-by-3 area of the DLDM (Figure 9b). From the 1/9th scale level, you can zoom to the 1/1 scale level in a similar fashion, except there are 9 zones to zoom to instead of four. Single left clicking on the display area shows the zones, double left-clicking zooms to the zone, where each pixel represents a cell in the DLDM (Figure 10). Right clicking at any point while viewing the DLDM will automatically return you to the 1/36th scale image of the entire DLDM.

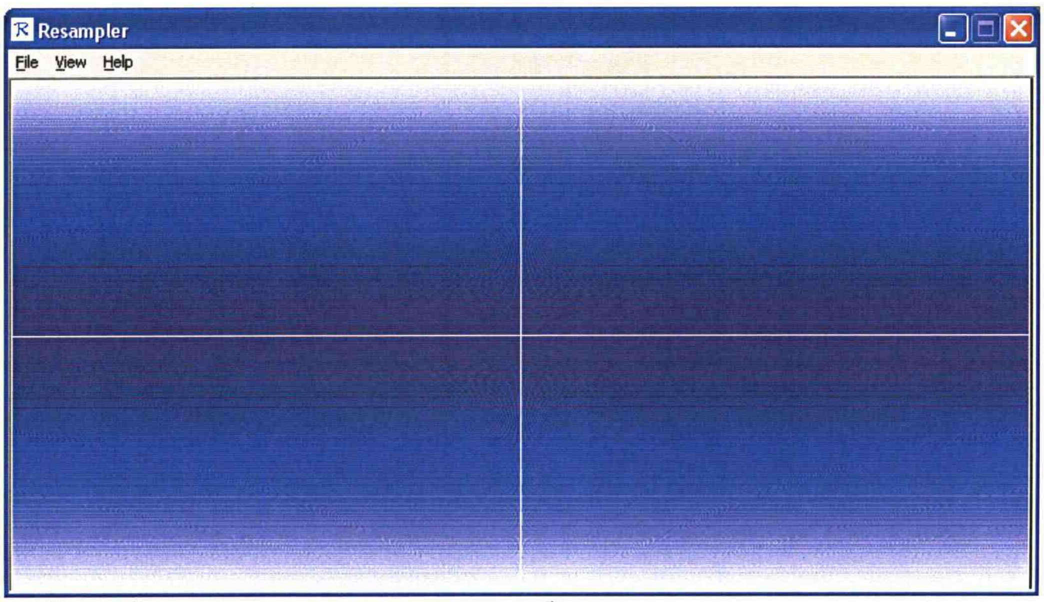


Figure 9a. Mollweide Projection DLDM 1/36th scale, showing quadrants after left-clicking

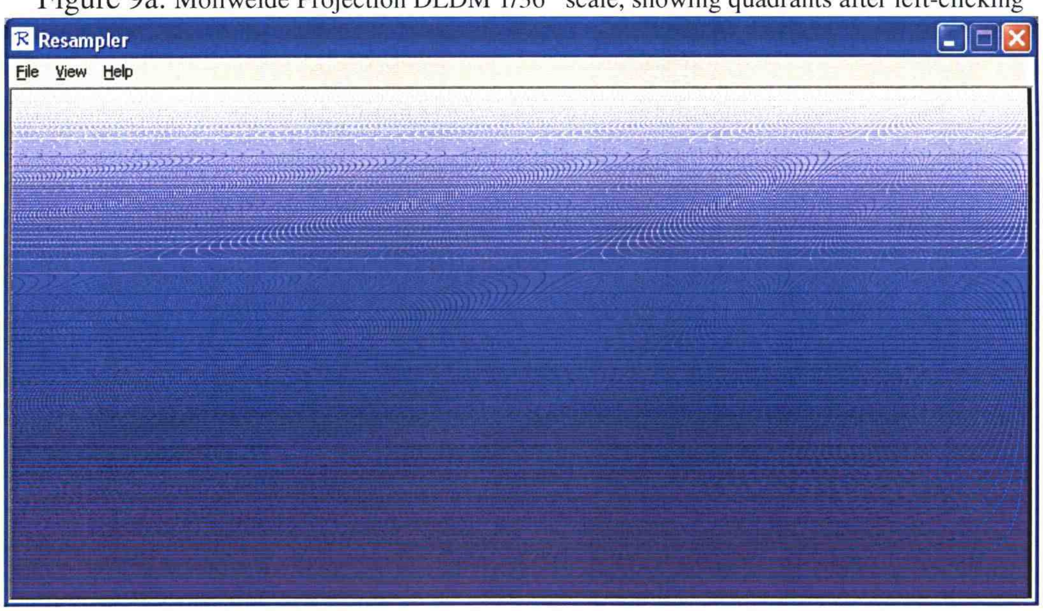
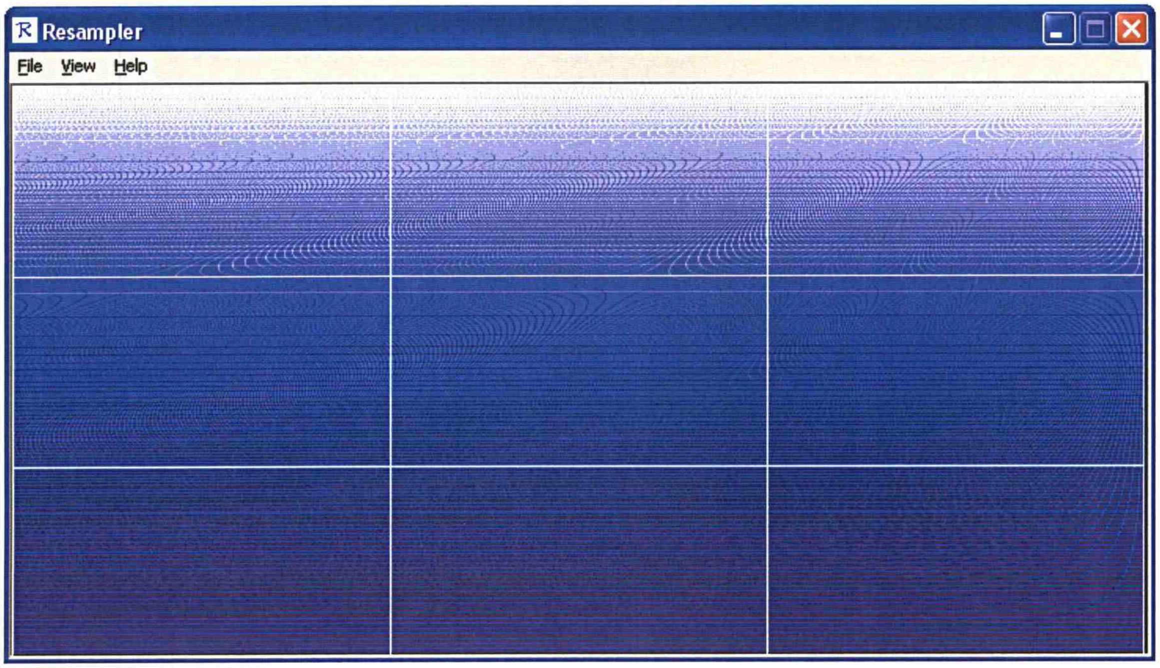


Figure 9b. Upper left quadrant, Mollweide Projection



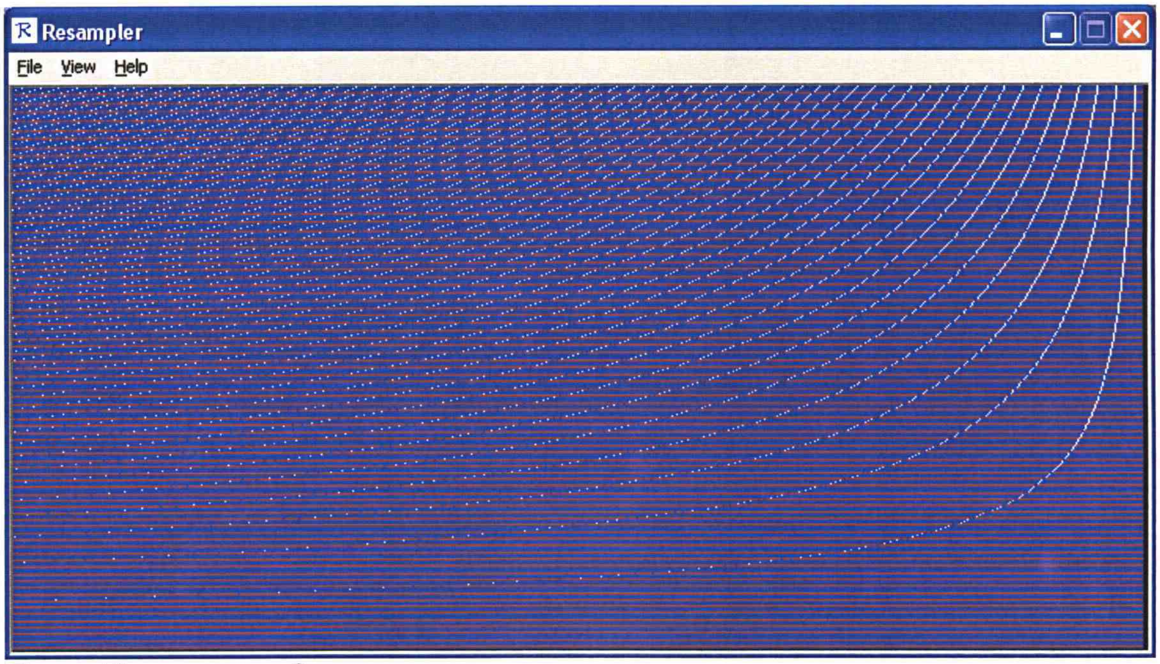


Figure 10. Zones for 1/9th scale DLDM for the Mollweide Projection (top) and 1/1 scale DLDM, lower right zone (bottom)

Selecting the Map option of the View menu works similarly to the DLDM option, except this will show a map preview instead of the DLDM. The map preview will fit entirely within the display area, and may not cover the entire display area. While

Resampler is resampling, the map preview will also be drawn as it is being created line-by-line from the top of the display area to the bottom. The upper-right-hand corner of the map preview represents the first row and column of the binary output file. The map preview has been colored using a continuous color scheme designed for this program to use on the ETOPO5 elevation data set (blues for water, greens, reds, and white for land, and black for missing data). Other data sets may not be well suited for this color scheme. There is no zoom function for the map preview.

Using Files Created From Resampler

Resampler can produce, depending on the users needs, a binary map output file, and two different DLDM files. All three files are uncompressed and designed to be available for use by a wide variety of programs. However, these files may need to be compressed or have headers added to them due to size restrictions or program header requirements.

The binary map output file will be in the same format as the input file: row major order, two-byte integer, with no header. The map file is always created whenever any resampling is done under a name chosen by the user. This file is of a map projection so the equal-angle grid characteristics will not be present in the file cells. Also, since few of the output files will be rectangular map projections, while the output file will always be rectangular, there will be cells in these output files with a value of –0 which are not part of the map, but are necessary to keep the file rectangular. To use this file in another program, you will need to create a header file in the same format as the header file for the input data file. Most of the information for the header file will be the same as for the input data file, except for the number of rows and columns, which will be given in the output statistics message box at the end of the resampling.

There are two available options for saving the DLDM to a file. The two Save DLDM options in the File menu are used for saving the DLDM as either a portable pixel map file (.ppm), or as a binary output file. DLDMs can only be saved when the resampling is complete. DLDMs can be saved as binary output files by selecting the "Save DLDM as .bin" from the file menu. This will produce a file in the same format as the binary map output file, except it will have the same number of rows and columns as the input file and the cell values will represent the number of times the cell was resampled.

The portable pixel map (.ppm) option will produce a DLDM file that is well suited for graphical display, and is what will be created when the "Save DLDM" command is executed. Portable pixel maps are simple graphics files that have a header and one-binary-byte per color RGB color values for every pixel in row major order. The Portable Pixel Map format was chosen for this program because it is simple to create, and can be converted to many different graphics formats using graphics programs, such as IrfanView (Irfan View can be found on the internet at http://www.irfanview.com). This format produces fairly large files, but retains data for every cell in the DLDM. The header in a .ppm file consists of four parts: a Magic Value, which identifies the file as a .ppm file (in this case it is "P6" for binary .ppm file), an image height value, an image width value, and a number representing the maximum color brightness value. The header is written at the beginning of the file in ASCII with each item separated by ASCII spaces. The color values for each pixel follow after the last space in the header with RGB binary values between 0 and 255 in row major order: red first, then green, then blue.

RESULTS

The DLDMs produced by this program show that they can be useful for studying data loss and duplication errors caused by nearest-neighbor resampling. One result learned from testing this program is that similar map projections produce similar DLDMs. For example, all pseudocylindrical and cylindrical map projections included in this program have produced DLDMs with similar characteristics, while the azimuthal projections in this program have DLDMs similar to each other, but distinctly different in character from the other projections (Figure 11).

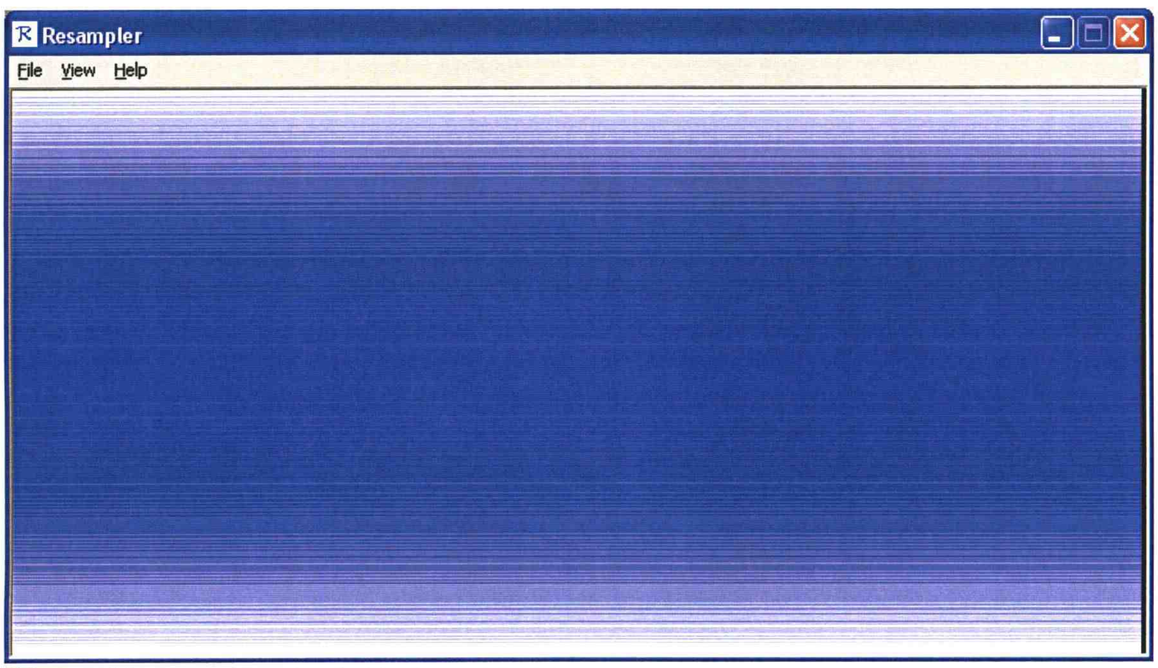


Figure 11a. Example of cylindrical type DLDM

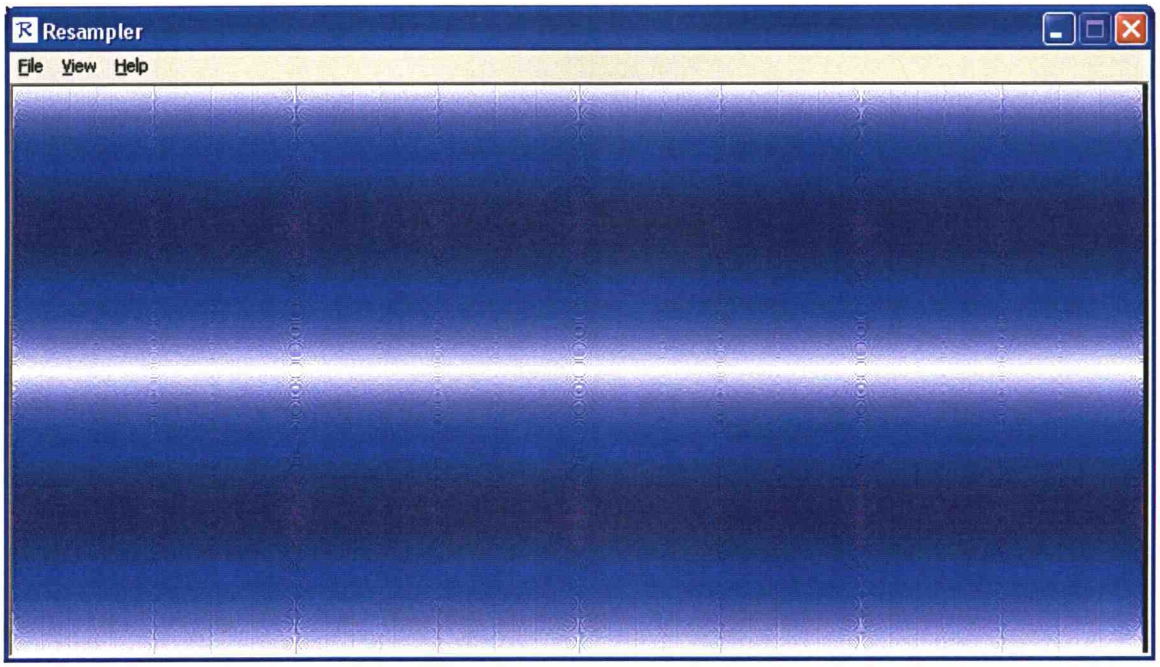


Figure 11b. Example of azimuthal type DLDM

The Resampler program demonstrates its usefulness in quickly resampling equal-angle data sets, producing DLDMs and providing simple statistics for analysis. The numerical statistics provided at the end of resampling can give the user a quick idea of how much of the original data is retained in the output map, and gives quantitative data for comparison to other map projections. The graphical output of the program allows for quick visual interpretation of the DLDM, provided that most of the data cells of the original data set were resampled less than 5 times. The display provided by the program also allows the user to quickly check for errors in the resampling procedure itself, in the case that there are still programming errors in Resampler.

CONCLUSION

The conclusion will focus on the overall effectiveness of the Resampler program as a nearest-neighbor resampling tool and DLDM creator, the limitations of the program, and discussion of future studies relating to the program.

Effectiveness

The Resampler equal-angle data set resampling program has shown to be successful in meeting its objectives. It can successfully perform nearest-neighbor resampling on the ETOPO5 equal-angle data set for any of the nine map projections available in the program and create Data Loss and Duplication Maps for each resampling procedure performed. This program also performs well in displaying the DLDMs for visual analysis and providing useful statistics about the DLDM as well as producing a graphics file and a binary file of the DLDM for more in-depth analysis and viewing outside of the program.

Limitations

Resampler has limitations in a few key areas that need to be enhanced for more in-depth studies on resampling. The limitations of Resampler can be divided into two areas: computational limitations and display limitations. Computational limitations include limits on the availability of map projections and limits brought onto the program due to memory and speed of computers. Display limitations include visual display issues and statistical information available about the DLDMs.

The projections selected for Resampler were selected because they are all commonly used. However, only nine projections were included, and only the spherical form of each projection was included. Including more projections is possible, but may require significant changes to the program due to slight differences in the procedure for resampling to different projections. Also, since there are many possible map projections, a limit on projections had to be set to keep program developmental times reasonable.

As with most programs, there are computational limitations that reduce the functionality of Resampler. Because this program needs to store the entire input data set at once, the size limitations for input data sets has been capped at 6,500 rows and 6,500 columns, meaning data sets larger than this size may need to be broken into smaller pieces to be resampled.

The statistical information provided at the conclusion of resampling may need to be enhanced for more complex resampling studies. Resampler currently only includes the most basic statistics on data loss and duplication. More complex statistics were left out of Resampler until more research is done on DLDMs to determine which statistics will be most useful. A routine to calculate other statistics could be added to Resampler in the future, or the user could write another program to analyze the .bin or .ppm files produced during resampling.

Directions for Future Work

Future studies building on this work involve three related areas: 1) further studies of data loss and duplication patterns, 2) finding which statistics are important to calculate in relation to DLDMs and 3) using data from these studies and the Resampler program to develop better global grid systems and compare these systems to the equal-angle grid system. Studies in each of these areas will benefit from use of the Resampler program, and may lead to or require improvements in the program.

Studying the data loss and duplication patterns of DLDMs with various map projections will lead to greater understanding of how efficient the equal-angle grid is for different projections. DLDMs produced in Resampler will aid in understanding the mathematics behind the patterns of loss and duplication in each DLDM. They can also be used to identify the data ceels that are more likely to be sampled, which can help cartographers get a better idea of the accuracy of the new map produced. From pattern studies may also come the need for statistical pattern analysis, specifically spatial statistical analysis, and thus the need for adding spatial statistical calculations to Resampler.

Finally, as global grid systems other than equal-angle grids are developed, researchers will need a tool for comparing the efficiency of the new grids to the equal-angle grid. DLDMs may be one tool that will be useful for comparison as they describe the data loss and duplication and pattern of loss and duplication. Presumably, other grid systems can also have DLDMs created for them, using similar principles as used by the equal-angle grid.

REFERENCES

Gurewich, O. and Gurewich, N. 1996. Teach Yourself Visual C++ 4 in 21 Days, Indianapolis, IN, Sams Publishing.

Kelley, A. and Pohl, I. 1996. C by Dissection: The Essentials of C Programming, Menlo Park, CA, Addison-Wesley Publishing.

Kimerling, A. J. 2002. Predicting Data Loss and Duplication when Resampling from Equal-Angle Grids, *Cartography and Geographic Information Science* 29(2) 111-126.

Snyder, J. P. 1987. Map Projections: A Working Manual, Washington, DC, US Government Printing Office.

APPENDIX A: Pseudo-Code for the Resampling Process

```
Read input file INFILE and place into array INGRID[rows][columns]
for(i=0;i<rows;++i){
        for(j=0;j<columns;++j){
                fread(&INGRID[i][j],2,1,INFILE);
                DLDM[i][j] = 0;
        }
}
```

Get Projection Information (Type, Variables, Scale)

Determine the number of Rows (w_{out}) and Columns (h_{out}) needed for the output file by

1) (For Cylindrical and Pseudo-cylindrical projections) – Determining the width and height (in degrees) of the input file, w_{in} and h_{in}, and using the forward projection equations to find the x,y values for $(-w_{in}/2,-h_{in}/2),(w_{in}/2,-h_{in}/2),(-w_{in}/2,h_{in}/2),(w_{in}/2,h_{in}/2)$. From the resulting set of x and y values, take the maximum difference in the x values to be the width w_{out}, and the maximum difference in the y values to be the height, h_{out}. The minimum x value becomes the initial x value ($init_x$), and the minimum y value becomes the initial y value ($init_y$).

2) (For Azimuthal projections) – Assume for this process that the projection is centered on 0°N,0°E. Use the forward projection equations to determine the distance from 0°N,0°E to 0°N,180°E. Twice this distance is equal to w_{out} and h_{out}.

Resampling Algorithm

```
for(i=0; i<h_out;++i){
        yp = h_out + init_y – i;
        for(j=0;j<w_out;++j){
                xp = j + init_x;
                Apply xp and yp, along with the projection variables, to the inverse
                projection to obtain the latitude (ϕ) and longitude (λ) of (xp,yp).
                If(ϕ and λ are within the bounds of the input file){
                        Set integer variables m and k to:
                                m = (int)(((northbound - ϕ°)*inputdensity) + 0.5);
                                k = (int)(((λ° + eastbound)*inputdensity) + 0.5);
                        outgrid[i][j] = ingrid[m][k];
                        ++DLDM[m][k];
                }
                Otherwise,
                        Set m and k both equal to –0, outgrid[i][j] = -0.
                Write outgrid[i][j] to the end of the output file.
        } //end for j
} //end for i
```

APPENDIX B: Map Projection Summaries

Each projection will be described in terms of the inverse projection formula used, method of determining the number of rows and columns (in parenthesis next to name, see Appendix A), variables associated with the projection, and DLDMs of each projection at 100 dpi and 141 dpi. All projection formulas were taken from Snyder, 1987

Symbols used:

x = column position of cell on new map
y = row position of cell on new map
R = adjusted radius of projection
k_0 = scale factor along central meridian
ϕ_0 = latitude of origin in radians
λ_0 = longitude of origin in radians
ϕ = latitude in radians
λ = longitude in radians

Azimuthal Equidistant (Azimuthal)

Inverse Projection Formula:

$$\rho = (x^2 + y^2)^{1/2}$$

$$c = \rho/R$$

$$\phi = \arcsin[\cos(c)\sin(\phi_0) + (y\sin(c)\cos(\phi_0)/\rho)]$$

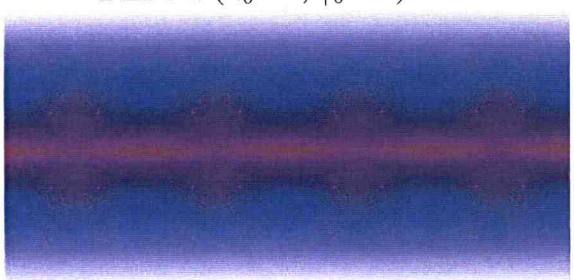
if $\phi_0 \neq \pm\pi/2$

$$\lambda = \lambda_0 + \arctan[x\sin(c)/(\rho\cos(\phi_0)\cos(c) - y\sin(\phi_0)\sin(c))]$$

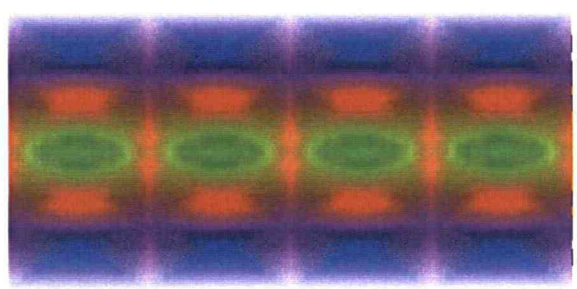
if $\phi_0 = \pm\pi/2$

$$\lambda = \lambda_0 + \arctan[x/(\pm(-y))]$$

DLDMs ($\lambda_0 = 0$, $\phi_0 = \pi$)



100 dpi



141 dpi

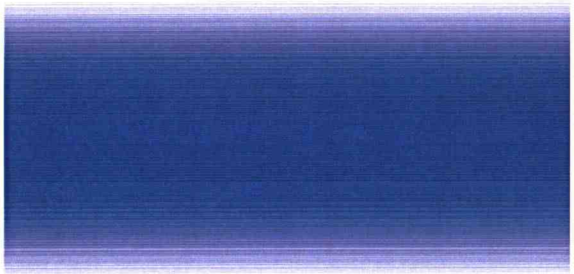
Cylindrical Equal-Area (Cylindrical)

Inverse Projection Formula

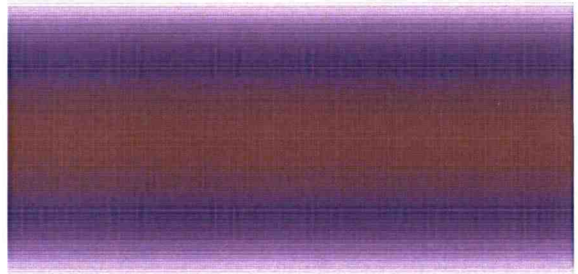
$$\phi = \arcsin[(y/R)\cos(\phi_0)]$$

$$\lambda = \lambda_0 + x/(R\cos(\phi_0))$$

DLDMs ($\lambda_0 = 0$, $\phi_0 = 0$)



100 dpi



141 dpi

Lambert Azimuthal Equal-Area (Azimuthal)

Inverse Projection Formula

$$\rho = (x^2 + y^2)^{1/2}$$

$$c = 2 \arcsin[\rho/(2R)]$$

$$\phi = \arcsin[\cos(c) \sin(\phi_0) + (y \sin(c) \cos(\phi_0)/\rho)]$$

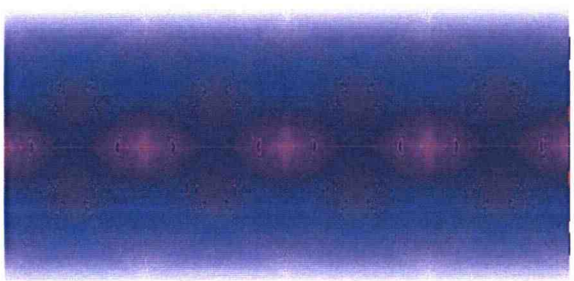
if $\phi_0 \neq \pm\pi/2$

$$\lambda = \lambda_0 + \arctan[x \sin(c)/(\rho \cos(\phi_0) \cos(c) - y \sin(\phi_0) \sin(c))]$$

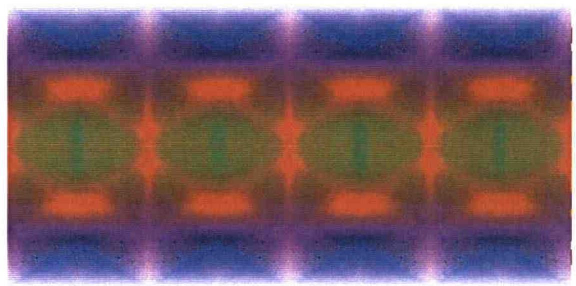
if $\phi_0 = \pm\pi/2$

$$\lambda = \lambda_0 + \arctan[x/(\pm(-y))]$$

DLDMs ($\lambda_0 = 0$, $\phi_0 = \pi$)



100 dpi



141 dpi

Mollweide (Pseudo-Cylindrical)

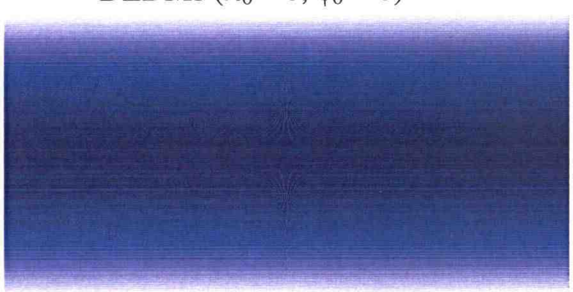
Inverse Projection Formula

$$\theta = \arcsin[y/(\mathrm{sqrt}(2)*R)]$$

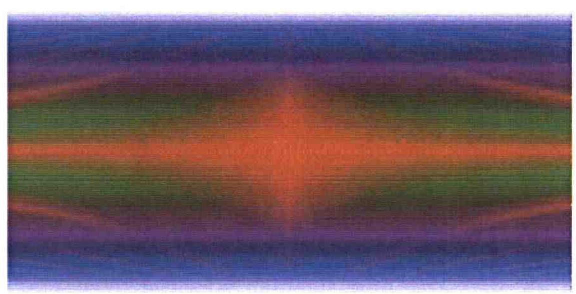
$$\phi = \arcsin[2\theta + \sin(2\theta))/\pi]$$

$$\lambda = \lambda_0 + \pi x/(\mathrm{sqrt}(8)*R*\sin(\theta))$$

DLDMs ($\lambda_0 = 0$, $\phi_0 = 0$)



100 dpi



141 dpi

Orthographic Azimuthal (Azimuthal)

Inverse Projection Formula

$\rho = (x^2 + y^2)^{1/2}$

$c = \arcsin[\rho/R]$

$\phi = \arcsin[\cos(c)\sin(\phi_0) + (y\sin(c)\cos(\phi_0)/\rho)]$

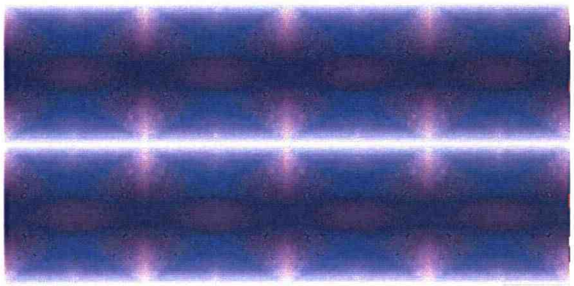
if $\phi_0 \neq \pm\pi/2$

$\lambda = \lambda_0 + \arctan[x\sin(c)/(\rho\cos(\phi_0)\cos(c) - y\sin(\phi_0)\sin(c))]$

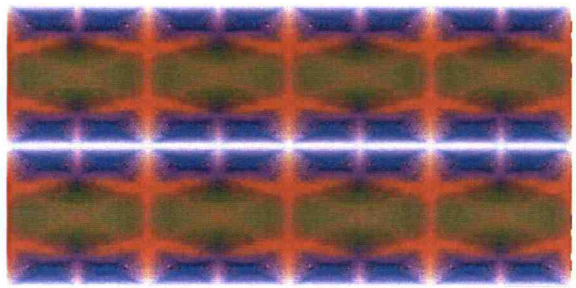
if $\phi_0 = \pm\pi/2$

$\lambda = \lambda_0 + \arctan[x/(\pm(-y))]$

DLDMs ($\lambda_0 = 0$, $\phi_0 = \pi$)



100 dpi



141 dpi

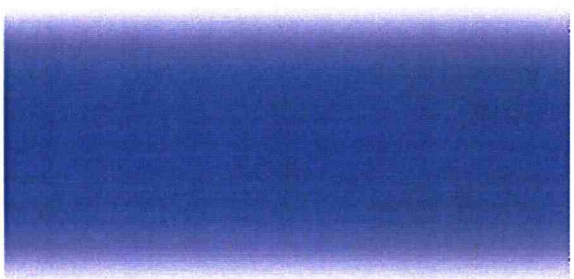
Sinusoidal (Pseudo-Cylindrical)

Inverse Projection Formula

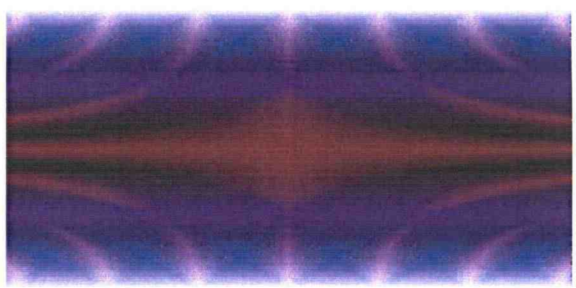
$\phi = y/R$

$\lambda = \lambda_0 + x/(R\cos(\phi_0))$

DLDMs ($\lambda_0 = 0$, $\phi_0 = 0$)



100 dpi



141 dpi

Stereographic (Azimuthal)

Inverse Projection Formula

$\rho = (x^2 + y^2)^{1/2}$

$c = 2\arctan[\rho/(2R*k_0)]$

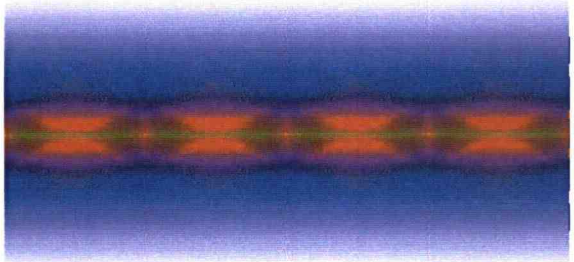
$\phi = \arcsin[\cos(c)\sin(\phi_0) + (y\sin(c)\cos(\phi_0)/\rho)]$

if $\phi_0 \neq \pm\pi/2$

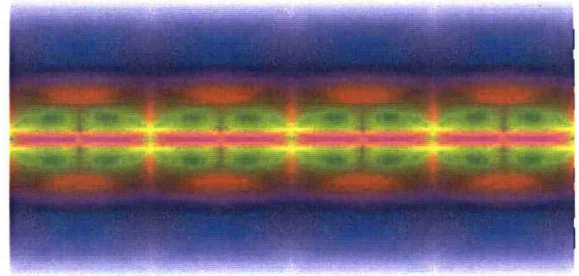
$\lambda = \lambda_0 + \arctan[x\sin(c)/(\rho\cos(\phi_0)\cos(c) - y\sin(\phi_0)\sin(c))]$

if $\phi_0 = \pm\pi/2$

$\lambda = \lambda_0 + \arctan[x/(\pm(-y))]$ DLDMs ($\lambda_0 = 0$, $\phi_0 = \pi$)



100 dpi



141 dpi

Mercator (Cylindrical)

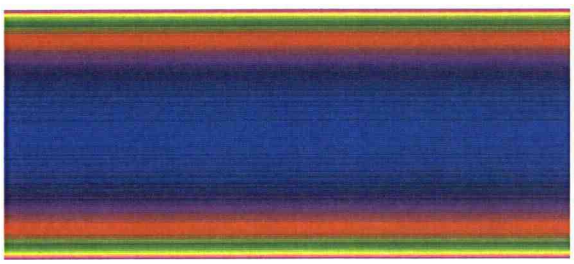
Inverse Projection Formula

$$\phi = (\pi/2) - 2 \arctan(\mathbf{e}^{-(y/R)})$$

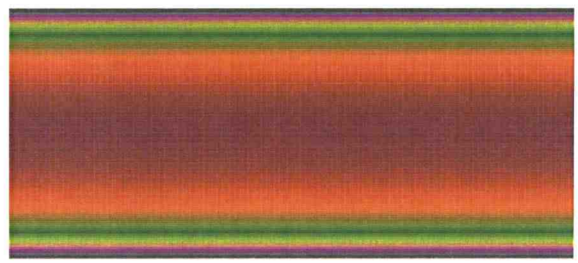
$$\lambda = \lambda_0 + (x/R)$$

Note: Mercator Projection ends at 85° N and 85° S

DLDMs ($\lambda_0 = 0$, $\phi_0 = 0$)



100 dpi



141 dpi

Transverse Mercator (Cylindrical)

Inverse Projection Formula

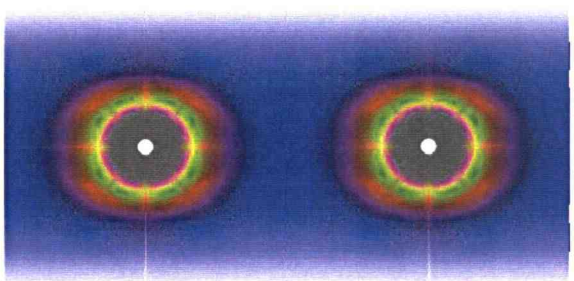
$$D = y/(R\ k_0) + \phi_0$$

$$\phi = \arcsin[\sin(D)/\cosh(x/(R\ k_0))]$$

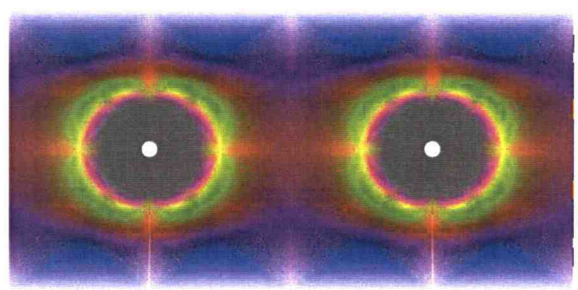
$$\lambda = \lambda_0 + \arctan[\sinh(x/(R\ k_0))/\cos(D)]$$

Note: Transverse Mercator Projection bounds are selected by the user (<90° from the central meridian), here at 85° N and 85° S

DLDMs ($\lambda_0 = 0$, $\phi_0 = 0$)



100 dpi



141 dpi