Report: Comparison of Methods to Produce Digital Terrain Models

Evan J Fedorko
West Virginia GIS Technical Center
27 April 2005

Introduction

This report compares Digital Terrain Models (DTM) created through two separate processes from the same source data. Using mass point and breakline data provide by a private contractor as part of the State Addressing and Mapping Board (SAMB) project, we created two versions of a DTM for the Lorado 7 ½' Quadrangle in southern West Virginia.

We first employed the traditional method of transforming mass points and breaklines into gridded elevation data. That method involves creating Triangular Irregular Networks or TINs. The TINs are transformed into lattices and eventually into digital terrain models in USGS DEM format. The final product has a 3 meter pixel size with horizontal units of meters and vertical units of feet. We will not discuss the methodology for creating these products in this report.

As a comparison test, we utilized the TOPOGRID command in Arc Info to generate a surface. TOPOGRID is an interpolation method designed to harness the efficient computing characteristics of local interpolators such as inverse distance weighting, while conserving the power of global interpolators like kriging. TOPOGRID is designed to minimize generalization of abrupt features such as streams and ridges. It is based on the ANUDEM program developed by Michael Hutchinson (ESRI 2001).

This report is essentially a comparison of a vector surface elevation model (TIN) and a raster surface elevation model. It is true that we are comparing two rasters, but one of these surfaces was derived from a TIN, which is a vector data format. Both of these formats have their advantages and disadvantages, both of which will be discussed here (Zeiler 1999).

Methods

This section will first review the TOPOGRID command and the preparation of data to run that command. Following that, we will outline our methods of comparing the TOPOGRID surface to the TIN generated surface. TOPOGRID will accept a total of 11 parameters. For the purposes of our experiment, we used only seven and we review only those seven in this document.

TOPOGRID is not specifically designed to accept elevation breaklines. For that reason, we utilized an Arc GIS toolset published by ET GeoWizards to convert the vertices of the LINE ZM type breakline shapefiles into POINT ZM shapefiles. This tool converts the nodes and vertices of each line into a point that contains the elevation information. TOPOGRID requires coverages containing a numeric input to operate. Once again utilizing ET GeoWizards, we converted each POINT ZM type shapefile to regular point shapefiles that contain an elevation attribute. The resultant mass point shapefile was converted to a coverage with Arc Toolbox. Figure 1 is a subset image of the elevation points we used as an input to TOPOGRID. This is an information rich dataset. The sheer volume and density of points is impressive. It is recommended in the TOPOGRID literature that a projection in which the vertical and horizontal units are the same is used on all data that the command will query. For this reason, we projected all data to West Virginia State Plane South with units of feet.

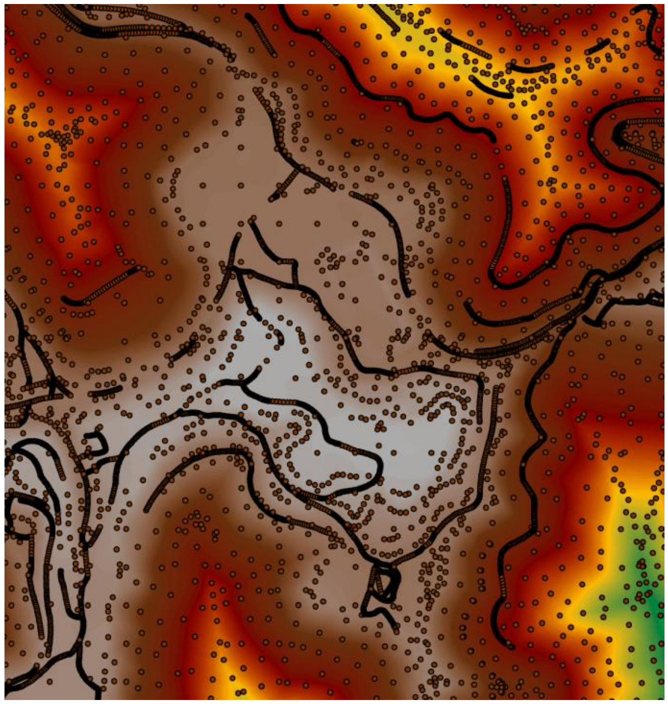

Figure 1. Mass points dataset used to generate elevation raster with TOPOGRID

Appendix 2 contains the Arc Macro Language (AML) file that we used to run the TOPOGRID command. Appendix 1 contains brief explanations taken from the Arc Info help file of each parameter or command. For the purpose of this exercise, we kept the parameters fairly simple. We used the neatline of the Lorado 7 ½' Quadrangle as our analysis extent and extended the interpolation 500' beyond that border in order to simulate a similar practice in the TIN process.

The TOPOGRID documentation advises using tight smoothing tolerances if the data quality and accuracy are very high. The documentation recommends a value of 2.5 for tolerance one when using elevation point data at a 50,000 scale. Our data is at 4,800 scale and we chose a lower value, 1, accordingly. The second tolerance, the horizontal standard error, represents the amount of inherent error in converting irregularly spaced data into a regularly spaced grid. We opted to use the default value as a smaller number may result in "spurious peaks and sinks" and a large value will result in a large amount of smoothing. Early tests with this parameter at a value of 0.5 resulted in a dataset with numerous small peaks around each point value. We chose the default value of 0 for the final tolerance, the vertical standard error.

We also chose to perform hydrologic enforcement. As part of the SAMB project, 1:4,800 scale vector stream centerlines and water body polygons were collected. TOPOGRID's STREAM parameter requires a line coverage of stream centerlines with proper flow direction. A small amount of pre-processing was required to flip some of the arcs to the proper direction. As per the TOPOGRID documentation, we also removed braiding in the streams such that there remained only one arc per section of stream. This coverage was used as an input in TOPOGRID with the STREAM parameter. We used the LAKE parameter to input the hydrologic polygon coverage, also collected at 1:4,800 scale.

Once a few unexpected technical problems were worked out, we ran an Arc Macro Language script to produce a 9.8 ft. grid cell elevation raster.

Comparison

Once the new elevation surface was created, we began the arduous task of comparing the surfaces. The most obvious method of comparing the two rasters was subtraction followed by a statistical analysis of the results. Due to the fact that we have established a preliminary NSSDA accuracy for the TIN generated DTM (6.8 feet), we were confident that comparing the TOPOGRID surface to the TIN DTM would allow us to make some conclusions about the quality of the TOPOGRID surface. Before this was possible, we had to make sure both rasters were in the same coordinate system. To achieve this uniformity, we projected the raster produced through normal methods (TIN DTM) to West Virginia State Plane South with units of feet. The resampling process introduces a small amount of error, but we did not quantify this error.

Once the projections were matched, we performed raster subtraction. We subtracted the TOPOGRID generated elevation from the TIN DTM elevation raster. In order to avoid possible errors from edge effect, we clipped the difference raster by 0.01 degrees on each side. Next, in order to calculate the Root Mean Square Error (RMSE), we squared the clipped difference raster. Below is a table of errors observed between the two elevation surfaces.

Minimum Difference	Maximum Difference	Range of Difference
-154	1168	1323
Mean Difference	RMSE	NSSDA 95% Accuracy Statistic
1.0286	10.8359	21.6284

The first red flag we see in the table above are the minimum and maximum differences and the range of values that these differences encompass. Clearly, there are some very significant differences between these two datasets. However, the mean difference is very small, only 1.03 feet. This indicates to us that these very large differences are probably localized.

Using the squared difference raster, we calculated the RMS Error. This number is in map units. The RMSE is roughly 10.8 feet. RMSE is the most widely used and accepted measure of error employed in spatial science today. The NSSDA accuracy statistic is also common. This number is calculated by multiplying the RMSE by a constant value specific to the type of data being analyzed. The NSSDA constant is 1.96, which results in a value of 21.6 feet. So, 95% of the TOPOGRID surface's elevation values are within 21.6 feet of those calculated with the TIN method.

We can also compare these surfaces visually. This method, while subjective in nature, is a revealing exercise.

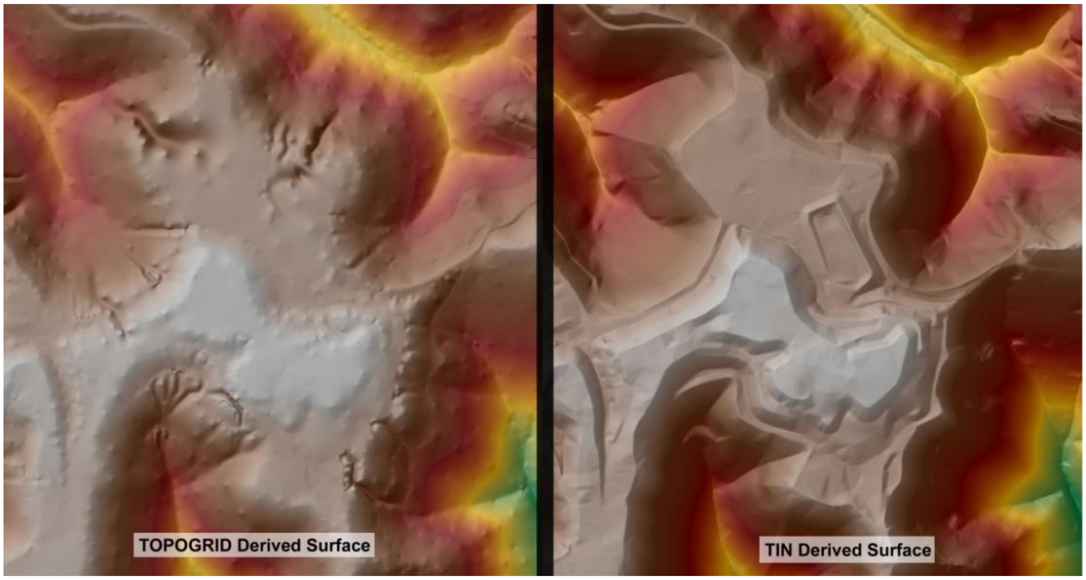

Figure 2. Comparison between TOPOGRID and TIN derived surfaces.

The above image is the same geographic extent of each raster. What is immediately apparent is the smoothing in the TOPOGRID derived surface. Where the breaklines have resulted in accurate characterization of post-coal-mining land features in the TIN derived surface, these features have become much smoother, and indeed, nearly nonexistent, on the TOPOGRID surface. The other obvious anomaly lies in the apparent representation of the point values themselves on the TOPOGRID surface. Figure 3 makes these aberrations clearer.

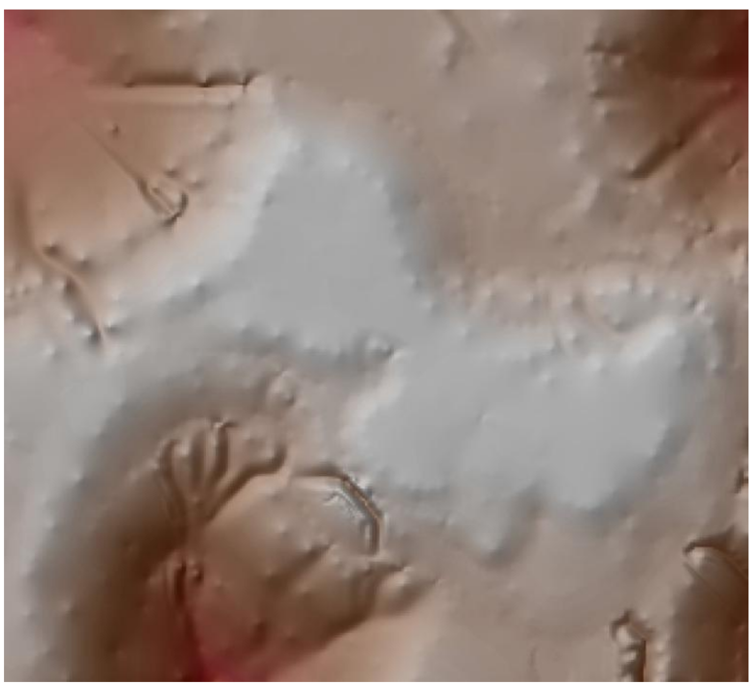

Figure 3. TOPOGRID surface subset.

A closer examination of the surface that results from TOPOGRID, here draped over a hillshade to accentuate micro-anomalies, reveals that, in areas of elevation points, there are very small peaks. While these peaks are all but completely invisible when viewing the DEM without a hillshade, even in a 3-D environment, they nonetheless do exist. Where TOPOGRID seems to be overly smoothing the surface compared to one created with breaklines, it also seems to fall off very quickly from points of known elevation.

In addition, and perhaps more alarmingly, it appears that TOPOGRID's drainage enforcement routine results in some very deep channels being carved. Figure 4 is a close up of one area on three surfaces.

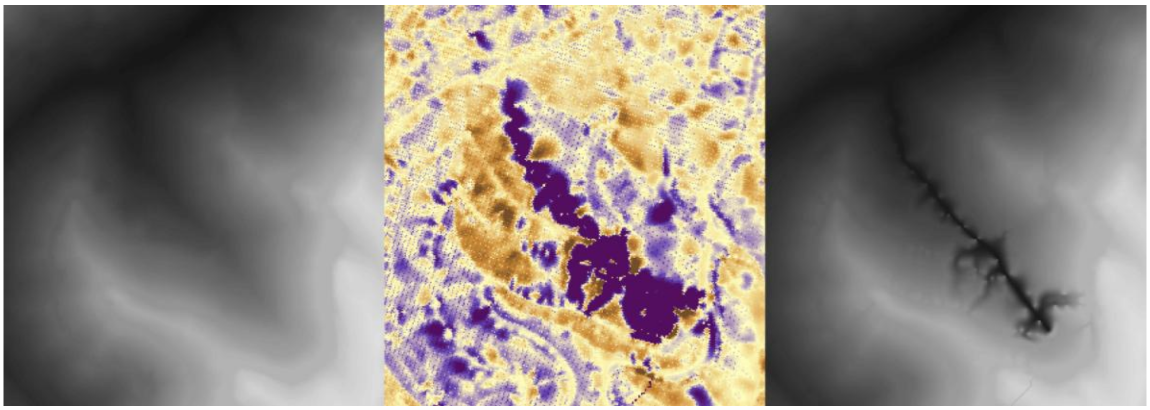

Figure 4. Difference image and comparisons.

The surface furthest to the left is that which was created with the TIN method. On the right is the surface created with TOPOGRID and the image in the center is a difference image between the two. The image is a result of subtracting the TOPOGRID raster from the TIN raster. The difference image is displayed with a diverging legend, from orange to purple. The darker purple color on the difference image indicates very high positive values while the darker oranges indicate very low vales – below zero. Those that are yellow are close to zero. What we can see from this figure is that the hydrologic enforcement routine that TOPOGRID performs results in substantial changes to the elevation values in the area around vector streams. The stippling effect that can be seen on the difference image is due to resampling distortion. In the area immediately around this particular stream valley, the TOPOGRID DTM is roughly 80 feet lower than the TIN DTM. This is clearly an error in the hydrologic enforcement routine that TOPOGRID employs and is a major cause for concern.

Discussion and Conclusion

With this report, we sought to compare elevation products produced with two interpolation methods. It is worth nothing that some of the anomalies that we observed on the TOPOGRID surface could probably be lessened with some tweaking of the input parameters. It is clear that the generalizations that TOPOGRID surfaces contain are unacceptable for creating data from these points and breaklines. In discussions with the elevation team at the United States Geological Survey office in Rolla, MO, we learned that they have more refined software for creating elevation surfaces from points. At this time, we have not utilized that software to generate a surface for comparison.

Beyond altering the TOPOGRID smoothing and hydro enforcement parameters, it would be a very worthwhile exercise to try and generate elevation surfaces at a more coarse resolution than 9.8 feet (3 meters). The data provider expressed concern about creating an elevation product of a resolution smaller than 5 meters.

TOPOGRID's main advantage over the TIN derivation method lies in its swiftness and relatively automatic process. We are satisfied that the TIN method of creating elevation surfaces will suit the source data and final product needs of this project.

References

Environmental Systems Research Institute (ESRI). *ARC/INFO Help System.* ESRI, 2001.

Federal Geodetic Control Subcommittee, Federal Geographic Data Committee. *Geospatial Positioning Accuracy Standards*. Report FGDC-STD-007.1-1998. FGDC, 1998.

Zeiler, Michael. *Modeling our world, the ESRI guide to geodatabase design.* ESRI, 1999.

Appendix 1

TOPOGRID Commands and Parameters

BOUNDARY keyword and parameter for input of a polygon coverage representing the outer boundary of the interpolated grid.

COMMANDS a listing of available subcommands.

CONTOUR keyword and parameters for input of a line coverage representing elevation contours.

DATATYPE the primary type of input data.

END keyword indicating the conclusion of data input. Entering END will start the program.

ENFORCE turns the drainage enforcement routine on or off. The default is on.

ITERATIONS the maximum number of iterations at each grid resolution. The default is 30.

LAKE a polygon coverage of lakes.

LIST lists the current setting of all subcommands.

MARGIN distance in map units to interpolate beyond the specified XYZLIMITS. The default is 0.

OUTPUTS optional outputs providing information that can be used to evaluate the output elevation grid.

POINT keyword and parameters for input of a point coverage representing surface elevations.

QUIT quits the TOPOGRID program without creating an output grid.

RESET resets all parameters to their default values.

SINK keyword and parameters for input of a point coverage representing known topographic depressions.

STREAM keyword and parameters for input of a line coverage representing streams.

TOLERANCES a set of tolerances used to adjust the calculations of the interpolation and drainage enforcement process.

XYZLIMITS the limits of input data to be used in the interpolation. The output grid is also constrained to fit within these limits.

Appendix 2

TOPOGRID AML

```
topogrid lorado3ft 9.84251968503937
boundary lorado
margin 500
datatype spot
tolerance 1 1 0
point lorallpts elev
enforce on
stream lorhyln
lake lorhypoly
outputs sink1 drain1 diag1
end
```

Appendix 3

Output Diagnostic File

```
        === TOPOGRID DIAGNOSTICS ===

 --- PARAMETER PROCESSING ---

 DRAINAGE OPTION:   1
   0 NO DRAINAGE ENFORCEMENT
   1 DRAINAGE ENFORCEMENT

 CONTOUR DATA OPTION:   0
  0 - ELEVATION DATA CONSISTS MAINLY OF SPOT HEIGHTS
  1 - ELEVATION DATA CONSISTS MAINLY OF CONTOURS

 HORIZONTAL STANDARD ERROR:   0.50000
   VERTICAL STANDARD ERROR:   0.00000

 NON-NEGATIVE ELEVATION TOLERANCE:      1.00
  TOLERANCE SHOULD BE HALF OF DATA CONTOUR INTERVAL

 MAXIMUM NO. OF ITERATIONS (NORMALLY 30-40):    30

 Z (HEIGHT) LIMITS:    592.9420   3184.6780

 X (LONGITUDE) LIMITS:   1747506.0380   1792358.4002

 Y (LATITUDE) LIMITS:    267591.0450    324234.7458

 GRID CELL SIZE:         9.842520
 NUMBER OF ROWS IN OUTPUT GRID =   5756
 NUMBER OF COLS IN OUTPUT GRID =   4558

 GRID MARGIN (NON-NEGATIVE): 501.96850

 NUMBER OF DATA FILES:    4
  TYPE 1 = POINT DATA FILE 21 = ARC/INFO POINT DATA FILE
  TYPE 2 = SINK POINT FILE 22 = ARC/INFO SINK POINT FILE
  TYPE 3 = STREAMLINE FILE 23 = ARC/INFO STREAMLINE FILE
  TYPE 4 = POLYGON FILE    24 = ARC/INFO POLYGON FILE
  TYPE 5 = CONTOUR FILE    25 = ARC/INFO CONTOUR FILE
  TYPE 6 = LAKE BDRY FILE  26 = ARC/INFO LAKE BDRY FILE

 INPUT COVERAGE  1:
   NAME: lorado
   TYPE: BOUNDARY

 INPUT COVERAGE  2:
   NAME: lorallpts
   TYPE: POINT
   ELEVATION ITEM NAME: elev

 INPUT COVERAGE  3:
```

NAME: lorhyln
TYPE: STREAM

INPUT COVERAGE 4:
NAME: lorhypoly
TYPE: LAKE

OUTPUT ELEVATION GRID NAME: lorado3ft

OUTPUT REMAINING-SINK COVERAGE NAME: sink1

OUTPUT DRAINAGE COVERAGE NAME: drain1

--- INPUT DATA PROCESSING ---

INPUT COVERAGE 1(POLYGON): lorado
NUMBER OF POLYGON POINTS = 5
NUMBER OF POINTS ACCEPTED = 5

INPUT COVERAGE 4(POLYGON): lorhypoly
NUMBER OF POLYGON POINTS = 3090
NUMBER OF POINTS ACCEPTED = 1940

INPUT COVERAGE 2 (POINT): lorallpts
NUMBER OF DATA POINTS = 584669
NUMBER OF POINTS WITHIN LIMITS = 584669
NUMBER OF POINTS WITHIN POLYGON = 584669
NUMBER OF POINTS ACCEPTED = 582840

INPUT COVERAGE 3 (STREAM): lorhyln
NUMBER OF STREAM LINE PTS = 20497
NUMBER OF POINTS WITHIN LIMITS = 20497
NUMBER OF POINTS ACCEPTED = 18261

--- INTERPOLATION PROCESSING ---

NUMBER OF GRID RESOLUTIONS = 10

RESOLUTION 1
SPACING = 0.5039E+04 NCOLS = 10 NROWS = 12

NUMBER OF DISTINCT DATA POINTS = 90
NUMBER OF SINKS AT THIS RESOLUTION = 5

RESOLUTION 2
SPACING = 0.2520E+04 NCOLS = 19 NROWS = 24

NUMBER OF DISTINCT DATA POINTS = 357
NUMBER OF SINKS AT THIS RESOLUTION = 29

RESOLUTION 3
SPACING = 0.1260E+04 NCOLS = 37 NROWS = 47

NUMBER OF DISTINCT DATA POINTS = 1353
NUMBER OF REMAINING SINKS = 68
NUMBER OF REMAINING SINKS = 67

NUMBER OF REMAINING SINKS = 67
NUMBER OF REMAINING SINKS = 67
NUMBER OF REMAINING SINKS = 67
NUMBER OF SINKS AT THIS RESOLUTION = 67

RESOLUTION 4
SPACING = 0.6299E+03 NCOLS = 74 NROWS = 93

NUMBER OF DISTINCT DATA POINTS = 5120
NUMBER OF REMAINING SINKS = 137
NUMBER OF REMAINING SINKS = 136
NUMBER OF REMAINING SINKS = 136
NUMBER OF REMAINING SINKS = 136
NUMBER OF REMAINING SINKS = 136
NUMBER OF SINKS AT THIS RESOLUTION = 136

RESOLUTION 5
SPACING = 0.3150E+03 NCOLS = 147 NROWS = 184

NUMBER OF DISTINCT DATA POINTS = 20417
NUMBER OF REMAINING SINKS = 216
NUMBER OF REMAINING SINKS = 214
NUMBER OF REMAINING SINKS = 210
NUMBER OF REMAINING SINKS = 211
NUMBER OF REMAINING SINKS = 209
NUMBER OF SINKS AT THIS RESOLUTION = 209

RESOLUTION 6
SPACING = 0.1575E+03 NCOLS = 292 NROWS = 367

NUMBER OF DISTINCT DATA POINTS = 70521
NUMBER OF REMAINING SINKS = 385
NUMBER OF REMAINING SINKS = 378
NUMBER OF REMAINING SINKS = 375
NUMBER OF REMAINING SINKS = 375
NUMBER OF REMAINING SINKS = 375
NUMBER OF SINKS AT THIS RESOLUTION = 375

RESOLUTION 7
SPACING = 0.7874E+02 NCOLS = 583 NROWS = 733

NUMBER OF DISTINCT DATA POINTS = 150884
NUMBER OF REMAINING SINKS = 666
NUMBER OF REMAINING SINKS = 607
NUMBER OF REMAINING SINKS = 596
NUMBER OF REMAINING SINKS = 589
NUMBER OF REMAINING SINKS = 590
NUMBER OF SINKS AT THIS RESOLUTION = 591

RESOLUTION 8
SPACING = 0.3937E+02 NCOLS = 1166 NROWS = 1465

NUMBER OF DISTINCT DATA POINTS = 244541
NUMBER OF REMAINING SINKS = 984
NUMBER OF REMAINING SINKS = 803
NUMBER OF REMAINING SINKS = 737

```
NUMBER OF REMAINING SINKS   =     654
NUMBER OF REMAINING SINKS   =     615
NUMBER OF SINKS AT THIS RESOLUTION =    634

RESOLUTION  9
SPACING =  0.1969E+02   NCOLS =    2331   NROWS =    2930

NUMBER OF DISTINCT DATA POINTS  = 359749
NUMBER OF REMAINING SINKS   =    2300
NUMBER OF REMAINING SINKS   =    1608
NUMBER OF REMAINING SINKS   =    1349
NUMBER OF REMAINING SINKS   =    1175
NUMBER OF REMAINING SINKS   =    1113
NUMBER OF SINKS AT THIS RESOLUTION =   1153

RESOLUTION 10
SPACING =  0.9843E+01   NCOLS =    4660   NROWS =    5858

NUMBER OF DISTINCT DATA POINTS  = 482083
NUMBER OF REMAINING SINKS   =    3348
NUMBER OF REMAINING SINKS   =    2298
NUMBER OF REMAINING SINKS   =    1714
NUMBER OF REMAINING SINKS   =    1375
NUMBER OF REMAINING SINKS   =    1174
NUMBER OF SINKS AT THIS RESOLUTION =   1214

        === END OF DIAGNOSTICS ===
```