Exploring Spatial Interpolation and Terrain Analysis

Prepared for

Lawrie Keillor-Faulkner

Environmental Modeling / GEOM 105

By

Douglas Piper &

Tristan Gingras-Hill

February 8th, 2018

Fleming College Frost Campus

School of Environmental & Natural Resource Sciences

Table of Contents

List of Tables:

List of Figures:

1 Introduction

Drumlins are elongated, teardrop-shaped hills of rock, sand, and gravel, formed by a number of processes, including incremental accumulation of till, erosion of previously deposited sediment, meltwater floods, and sediment deformation (nsidc.org, 2018; Maclachlan & Eyles, 2013). The Peterborough Drumlin Field is amongst the biggest drumlin fields in North America and is the foundation for numerous surrounding communities which depend on the groundwater located below the various substrates (Lotimer, 2014). Population growth in the area has raised concerns with respect to the long-term sustainability of this precious fresh-water resource, thereby prompting considerable interests in determining the location, extent and potential productivity of subsurface aquifers in the region (Lotimer, 2014). The particular shape of drumlins makes them easily identifiable when looking at topographical maps as they consist of streamlined hills, several hundred meters wide, and sometimes more than a kilometer in length (geocaching.com, 2017).

In order to conduct spatial analyses and environmental modeling on the Peterborough Drumlin Field, it is important to have accurate topographic digital data to understand how the morphology of the area could be contributing to environmental factors such as hydrological processes (Maclachlan & Eyles, 2013; Lotimer, 2014). Digital Elevation Models (DEMs) are used to predict cell values of unknown raster cells by interpolating them from known sampling points to generate 3-D terrain models (ESRI, 2004; Avrun, 2013a). The accuracy of the generated terrain model depends on the interpolation mechanism used, which may or may not be suitable for the model depending on the type of analysis required and the quality/ quantity of the sampling data (Avrun, 2013b). Interpolation methods are based on the principle of spatial autocorrelation or spatial dependence. If autocorrelation is present in the data, the correlation can be used to measure: 1) the similarity of objects within an area, 2) the degree to which a spatial phenomenon is correlated to itself, 3) the level of interdependence between the variables and 4) the nature and strength of the interdependence. Different interpolation methods will almost always produce different results (ESRI, 2004). There are two categories of interpolation techniques: 1) Deterministic interpolation (e.g. Inverse Distance Weight), which creates surfaces based on measured points or mathematical formulas and 2) Geostatistical interpolation (e.g. Kriging), based on statistics and are used for more advanced prediction surface modeling (ESRI, 2004). Because the most appropriate method will depend on the distribution of the sampling points and the phenomenon being studied, it is crucial to understand which interpolation method will be optimal for addressing the Peterborough Drumlin Field DEM with the quality of data presented.

The objectives of this assignment were to 1) Model/ create continuous elevation surfaces by interpolating values from sample points, 2) compare different interpolation models to each other and evaluate how well these conform to the ‘real’ world and, 3) compare two different extensions available in ArcGIS 10.5 (ESRI®) for environmental modeling purposes, the Spatial Analyst extension and the Geostatistical Analyst extension. This assignment will determine the most accurate interpolation method for the DEM for terrain mapping the Peterborough Drumlin Field.

2 Methods

2.1 Data acquisition

The data used to conduct the various interpolation methods was downloaded from the National Topographic Database (NTDB) 031D8 data representing Peterborough, Ontario (Table 1). The data was clipped to the extent of the study area prior to processing. The projection used for the data was the Universal Transverse Mercator (UTM) and the Datum was the North American Datum 1983. A UTM projection is an ideal projection to use in spatial analyses because it maintains accurate measurements of areas and distances, minimizing distortions in latitudes between 80°S and 84°N, which is ideal for the Peterborough area and the narrow width of each zone ensures minimal distortions (University of Toronto, 2018; Geokov, 2018). Furthermore, UTM projection is favorable for interpolation as it uses meters and therefore there is a known distance between each sample point. The interpolations were based on the *elevation* fields of both the clip_031D08_contours_L shapefile (553 records) and the elev_pt shapefile (486 records). The contour lines shapefile contained 10-meter intervals.

Table 1 *List of feature classes used for the testing of the various interpolation methods and their corresponding properties.*

Feature class	**Classification**	**Source**	**Coordinate System: Projection, Datum, Zone, extent**	**Feat type (pt, line, poly, raster, etc.)**	**Fields**	**Max/ Min Value elevation**	**# of records**	**Description/comments**
clip_031D08_contours_L.shp	Clipped Contours	NTDB 031D8	Projection: Transverse Mercator. Datum: NAD 83 Zone: 17N Extent: Top (4919411.890610 m) Bottom (4902685 m) Right (717294.091445 m) Left (699160 m)	Polyline	Shape, EntityName, NTS, Version, Theme, ID, Code_Gener, ATC, ATG, ATZ, ATE, Accuracy, Type, Generation, Elevation	Max:330 Min: 190	553	This has been clipped from 031D08_contours_L.shp. Contour intervals were specified at 20 m.
Clip_031D08_elev_pt_p.shp	Clipped Spot heights	NTDB 031D8	Projection: Transverse Mercator. Datum: NAD 83 Zone: 17N Extent: Top (4919336.000000 m) Bottom (4902863.000000 m) Right (717197.000000 m) Left (699333.000000 m)	point	Shape, EntityName, NTS, Edition, Version, Theme, ID, Code_Gener, ATC, ATG, ATZ, ATE, Accuracy, Type, Elevation, Angle	Max: 188 Min:333	486	This has been clipped from 031D08_elev_pt_p.shp
clip_031D08_water_b_a.shp	Clipped Water (Lakes)	NTDB 031D8	Projection: Transverse Mercator. Datum: NAD 83 Zone: 17N Extent: Top (4919330.462095 m) Bottom (4902824.000000 m) Right (717183.697379 m) Left (699090.426099 m)	polygon	Shape, EntityName, NTS, Edition, Version, Theme, ID, Code_Gener, ATC, ATG, ATZ, ATE, Accuracy, Type, Elevation, Centroid_X, Centroid_Y.	Max: -9999 Min: -9999	107	This has been clipped from 031D08_water_b_a.shp
clip_031D08_water_b_L.shp	Clipped Water (Streams)	NTDB 031D8	Projection: Transverse Mercator. Datum: NAD 83 Zone: 17N Extent: Top (4919406.653594 m) Bottom (4902687.000000 m) Right 717289.638155 m) Left (699180.000000 m)	Polyline	Shape, EntityName, NTS, Edition, Version, Theme, ID, CODE_Gener, ATC, ATG, ATZ, ATE, Accuracy, Type, Elevation	Max: 183 Min: 307	376	This has been clipped from 031D08_water_b_L.shp
clip_031D08_road_L.shp	Clipped road	NTDB 031D8	Projection: Transverse Mercator. Datum: NAD 83 Zone: 17N Extent: Top (4919400.112413 m) Bottom (4902705.000000 m) Right (717292.000000 m) Left (699171.000000 m)	Polyline	Shape, EntityName, NTS, Edition, Version, Theme, ID, Code_Gener, ATC, ATG, ATE, Accuracy, Classifica, Support, surface, status, NB_Lanes, FIR_ROADINO, Sec_Roadino, THI_Roadino, elevation	Max: -9999 Min: -9999	3016	This has been clipped from 031D08_road_L.shp

2.2 Data observation and pre-processing

Figure 1 *(a-d) Figure demonstrating the differences between contour lines and spot heights in the city of Peterborough.*

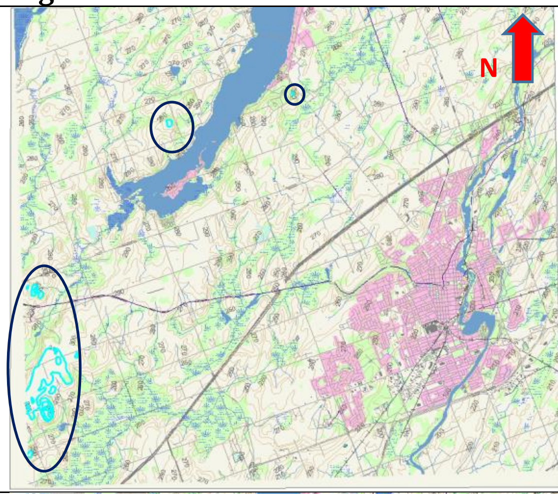
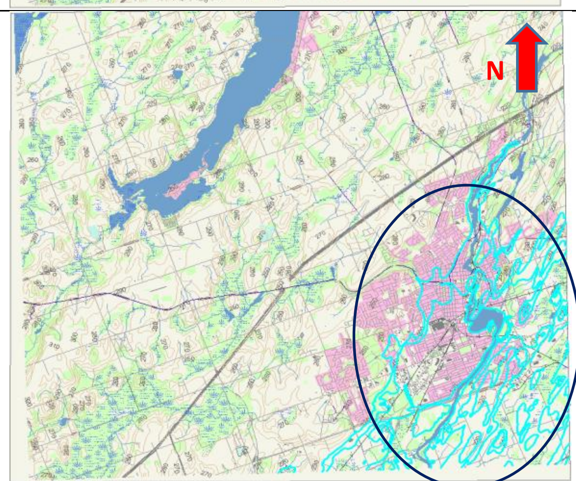
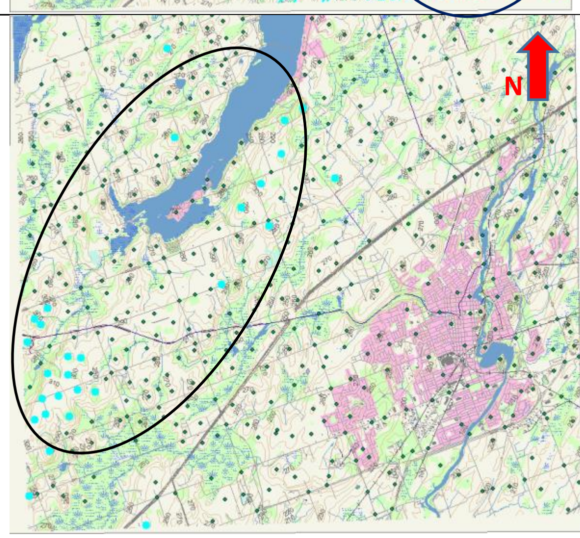
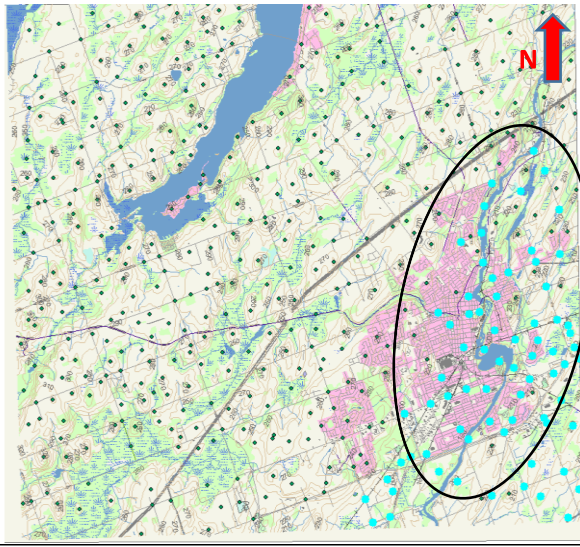
Name of Figure	Figure	Observations
Figure 1.a Contour lines shapefile displaying the highest elevations (>300 m) within the city of Peterborough.		The contours display the area of highest elevation in the southwest corner of the city with a maximum elevation of 330 m.
Figure 1.b Contour lines shapefile displaying the lowest elevations (<200 m) within the city of Peterborough.		The contours display that the area of lowest elevation is in the southeast corner of the city with a minimum elevation of 190 m.
Figure 1.c Spot heights shapefile displaying the highest elevations (>300 m) within the city of Peterborough.		The spot heights display that the areas of highest elevation are on the west side of town. The area of highest elevation is 333 m.

Figure 1.d Spot heights shapefile displaying the highest elevations (<200 m) within the city of Peterborough.		The spot heights display that the areas of lowest elevation are on the east side of town. The lowest elevation is 188 m.

The pre-processing data observation indicates that contour lines describe the surrounding landscape much better than spot heights (Figure 1). With the contour lines, we can visually see in which direction the terrain is slopping as they connect points of equal elevation and thereby do a better representation of the topography (Chang, 2016). Spot heights, however, are not recorded in intervals and therefore give a better value if we are interested in precise measurements. Contour lines were converted to points using the *Feature Vertices to Point* tool. This tool creates a feature class containing points generated from the specified vertices or locations of the input feature. This tool was more suitable for the interpolations comparatively to *Feature to Point*, which represents the midpoints of the input feature classes (ESRI, 2010). The input feature class for every interpolation method was the raw data transformed using the *Feature Vertices to Point tool* (. i.e. *ContPoint*), with the exception of *TopoToRaster*, which can interpolate directly from contour lines and *Spline*, which required simplification prior to running the model (ESRI, 2018). Additional details on data preparation, selected parameters and the reasoning behind their selection are also discussed in Table 2.

2.3 Interpolating using the Spatial Analyst extension

Table 2 *Table containing the various interpolation methods used with the Spatial Analyst extension and their selected parameters.*

Method	**Input/ Data Preparation**	**Parameters**	**Comments**
Inverse Distance Weighting (IDW)	Contour lines converted to points using the **'Feature Vertices to Point'** tool.	**Input Point Features:** ContPoint **Z value field:** ELEVATION **Output raster:** ContPoint_IDW **Output cell size:** 20 **Power:** 3 **Search radius:** Variable **Number of points:** 12 **Maximum distance:** N/A **Input barrier polyline features:** Rivers	Default parameters were left mostly as-is, with the exception of power which was changed to 3 to improve the accuracy of the surface.
***Method description:** The weight assigned to an un-sampled cell is a function of the distance of an input point from the unsampled cell location. The greater the distance, the less influence the cell has on the unsampled cell value (ESRI, 2004).			
Spline (Tension)	Contour lines simplified using the **'Simplify Line'** tool using a 50m simplification tolerance and then converted to points using the **'Feature Vertices to Point'** tool.	**Input Point Features:** ContPoint_Simplofy **Z value field:** ELEVATION **Output raster:** ContPoint_Spline_Tension **Output cell size:** 20 **Spline type:** TENSION **Weight:** 0.1 **Number of points:** 12	Tension was used in order to more accurately generate a surface that is faithful to the original data. Other parameters were left as default.
Spline (Regularized)	Contour lines simplified using the **'Simplify Line'** tool using a 50m simplification tolerance and then converted to points using the **'Feature Vertices to Point'** tool.	**Input Point Features:** ContPoint_Simplofy **Z value field:** ELEVATION **Output raster:** ContPoint_Spline_Regularized **Output cell size:** 20 **Spline type:** REGULARIZED **Weight:** 0.1 **Number of points:** 12	Regularized was used to see what sort of output it would generate.
***Method description:** This approach uses a mathematical function to minimize the surface curvature. It is the best method for representing the smoothly varying surfaces of phenomena such as temperature (ESRI, 2004).			
TopoToRaster	Raw contour lines	**Input feature data:** • Contour Lines // ELEVATION// Contour • Spot Heights // ELEVATION // PointElevation • Lakes // N/A // Lake • Rivers // N/A // Stream **Output surface raster:** ContLine_T2R **Output cell size:** 20 **Output Extent:** Same as layer study area **Margin in cells:** 20 **Smallest z value:** N/A **Largest z value:** N/A **Drainage enforcement:** ENFORCE **Primary type of input data:** CONTOUR **Maximum number of iterations:** 20 **Roughness penalty:** N/A **Profile curvature roughness penalty:** N/A **Discretization error factor:** 1 **Vertical standard d error:** 0 **Tolerance 1:** 2.5 **Tolerance 2:** 100 **Optional outputs:**	Barrier layers were added in order to boost the accuracy of the generated surface. Other parameters were left with their default settings.

***Method description:** The TopoToRaster method imposes constraints that ensure a hydrologically correct DEM that contains a connected drainage structure and correctly represents ridges and streams from input contour data. This method was specifically designed to work efficiently with contour inputs (ESRI, 2004).			
Kriging	Contour lines converted to points using the **'Feature Vertices to Point'** tool.	**Input Point Features:** ContPoint **Z value field:** ELEVATION **Output raster:** ContPoint_Kriging **Kriging method:** Ordinary **Semivariogram model:** Spherical **Output cell size:** 20 **Search radius:** Variable **Number of points:** 12 **Output variance:** N/A	Parameters were left with their default settings in order to generate favorable results.
***Method description:** Kriging assumes that the distance or direction between sample points reflects a spatial correlation that can be used to explain variation in a surface. This method is most appropriate when a spatially correlated distance or directional bias in the data is known and is often used for applications in soil science and geology (ESRI, 2004). There are various types of kriging, again varying along data type.			
Natural Neighbour	Contour lines converted to points using the **'Feature Vertices to Point'** tool	**Input Point Features:** ContPoint **Z value field:** ELEVATION **Output raster:** ContPoint_Natural Neighbor **Output cell size:** 20	No optional parameters other than cell size available to experiment with.

2.4 Geostatistical Analyst extension and Exploratory Spatial Data Analysis (ESDA).

Table 3 *Table demonstrating the various ESDA methods used to test and explore the data prior to the selection of an ideal interpolation method for the Geostatistical Analyst extension.*

Validation method	Description
Normality **Representation:** Histogram & QQ-Plot	Histograms and QQ plots were used to verify the normality of the data. Data was considered normally distributed if the mean and median were similar, skewness was near zero, and the kurtosis was near the value of 3 (ESRI, 2017).
Global trend **Representation:** Trend analysis graph	A trend analysis graph was used to determine if a directional trend occurred in the data. Directional trends (. e.g. North-South, East-West…) can skew the results of geostatistical interpolations (Chang, 2016).
Stationarity **Representation:** Voronoi map	A Voronoi map was used to test stationary data, meaning that the relationship between two points and their values depends on the distance between them, not their location. Two methods were attempted on the contour points data, (1) simple, which assigns the cell value recorded within each individual cell, and (2) entropy, which places cells in 5 classes based on the natural grouping of data values (ESRI, 2017).
Spatial Autocorrelation **Representation:** Semi-Variogram cloud	A Semi-variogram cloud was used to determine if the data was spatially auto-correlated, meaning that locations closer in distance are assumed to be closer in value than locations further apart (ESRI, 2017). The angle of the semi-variogram was adjusted to determine the directions of maximum and minimum sill and range values to determine in which direction the data ceases to correlate spatially. **Subsets (3.5 %) of the contour point data was used for this analysis due to difficulties analyzing the results with all the sampling points.**

Prior to conducting interpolation with the geostatistical analyst extension, the Exploratory Spatial Data Analysis (ESDA) tool was used (Table 3). This tool allows for more informed decisions on how the interpolation model should be constructed (ESRI, 2017). In order to understand the type of interpolation model suitable for our data, a number of validation methods were conducted. Difficulties with interpretation were experienced when verifying spatial autocorrelation with the semi-variogram cloud. Geostatistical Analyst is optimized for small sampled datasets (. i.e. <500 points), and therefore a subset of data was used for the spatial autocorrelation analysis (GEOM105, 2018).

2.5 Geostatistical Analyst extension and Interpolation models (Post-ESDA)

Table 4 *Table displaying the interpolation methods applied, including geostatistical kriging interpolation and a deterministic non-kriging method.*

Interpolation Method	Min/ Max elevation	Layers & Parameters	Description
Simple Kriging	196-304m	**Layers:** Spot Heights **Parameters:** Default	Simple kriging assumes that the mean of the data set is known and was, therefore, a better solution with spot heights as these are exact values (Chang, 2016).
Universal Kriging (1)	189-324m	**Layers:** Spot heights **Type:** Universal **Transformation type**: none **Order of trend removal:** constant (default) **Semi-variogram modeling:** k-bessel **anisotropic:** On **Searching Neighborhood:** 1 max neighbor 4 sectors	This method assumes a spatial correlation between sampling points and a directional trend in the data (Chang, 2016).
Universal Kriging (2)	188-333m	**Layers:** Contour Points **Type:** Universal **Transformation type**: none **Order of trend removal:** constant (default) **Semi-variogram modeling:** k-bessel **anisotropic:** On **Searching Neighborhood:** 1 Max neighbor: 1 4 sectors	
Empirical Bayesian kriging	190-330m	**Layers:** Contour points **Subset size:** 100 **Overlap Factor:** 1 **Number of Simulations:** 100 **Output surface type:** Prediction **Transformation:** none **Semi-variogram type:** Power **Neighbourhood type:** standard circular **Maximum neighborhoods:** 8 **Minimum neighbors:** 1 **Sector type:** 8 **Angle:** 0 **Radius:** 106.5918 **Predicted X:** 708227 **Predicted Y:** 4911048 **Value:** 250	The third method was Empirical Bayesian kriging (EBK). EBK automatically calculates these parameters through a process of sub-setting and simulations (ESRI, 2018).

Inverse Distance Weighting (IDW)	190-330m	**Layer:** Contour points **Power:** 100 **Neighborhood:** Standard **Min/Max neighbors:** 1 **Sector:** 1 **Angle:** 0 **Major/ minor semi-axis:** 6167.628 **Anisotropy factor:** 1	See table 2 for additional details on this interpolation method.

Multiple geostatistical interpolation methods were applied and compared to identify which of the models were better suited for creating a DEM with the available data. The parameters (. i.e. direction, shape, anisotropy…) for each model were modified until the best outcome was produced by visualizing the “fit” associated with the semi-variogram (Table 3). Cross-validation was used to compare the outcomes of each interpolation method by comparing the Root Mean Square (RMS)-standardized value nearest to one, mean standardized value nearest to 0, the lowest average standard error and the smallest difference when subtracting the RMS value from the average standard error.

2.6 Quality Assurance and Quality Control (QA/QC)

QA/QC was used to check for the accuracy of the data and to explore how well the models compare to each other and predict values at unknown locations (. i.e. observed vs predicted). The first method utilized was to zoom in on a number specific X, Y coordinate locations and identify the differences in elevation in relation to the various kriging models used (see table 4 for additional details). The second method used the raster calculator tool to compare the best output surfaces against some of the least-representative interpolation results. The outputs produced by the raster calculator were the areas of greatest variations amongst both interpolation methods, which was represented by the standard deviation. This enabled us to quickly locate areas where the elevation was exaggerated or underestimated.

3 Results

3.1 Interpolating using the Spatial Analyst extension

Figure 2 *(a-f) Figures demonstrating the results from using the " Feature Vertices to Point" tool on the contour lines (with the exception of topo To raster) and the corresponding interpolated DEM models. Figure 2.a is an overview of the area and the inset map displays the precise study location.*

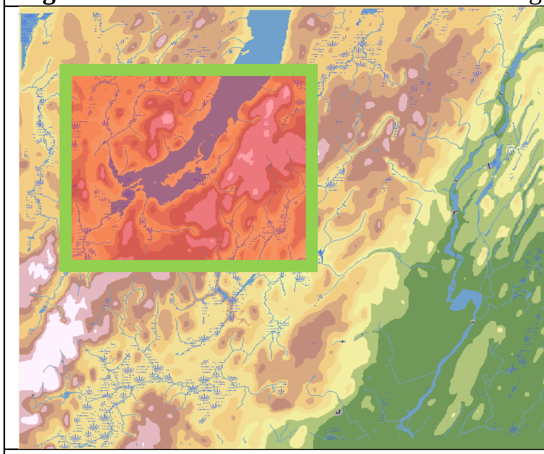
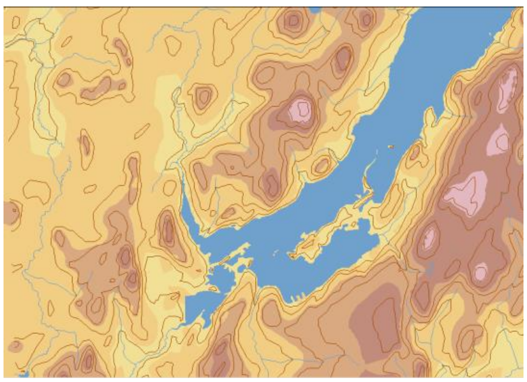
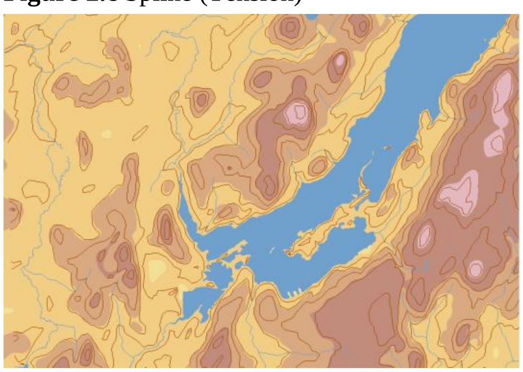
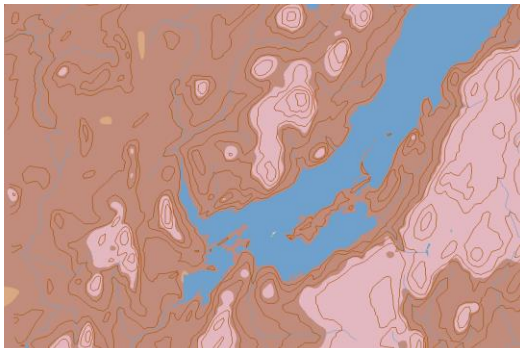
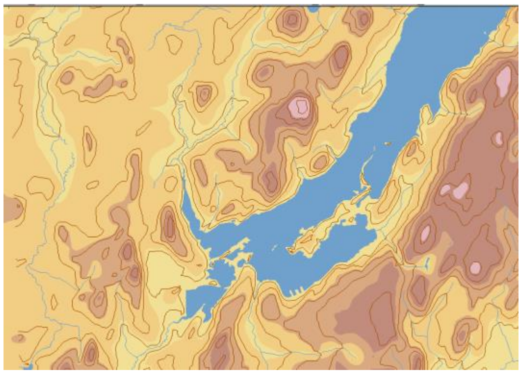
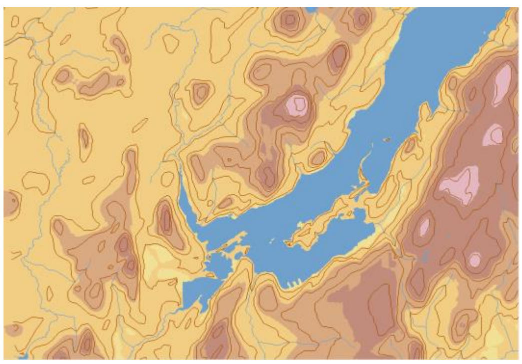
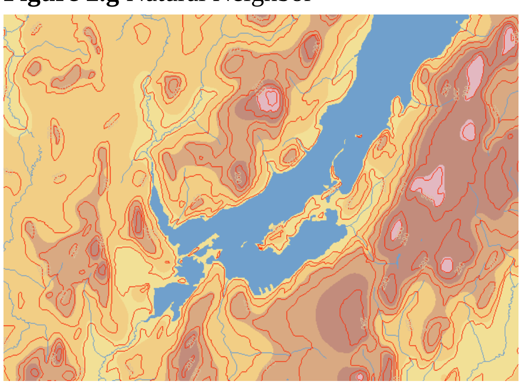
<table>
<tr><td colspan="3">Figure 2.a Grid 031D8 total overview with insight map to demonstrate the study areas of figure 2.(b-g)</td></tr>
<tr><td></td><td>Legend (m)
Extremely Low
Very Low
Low
Moderately Low
Medium
Moderately High
High
Very High
Extremely High</td><td></td></tr>
<tr><td>Figure 2.b Inverse Distance Weighting
</td><td colspan="2">Result
IDW produced a surface that, while not perfect, was quite accurate. Some of the smaller areas of heightened/lowered terrain were lost, and others were enlarged slightly, but overall the results were largely faithful to the source data.</td></tr>
<tr><td>Figure 2.c Spline (Tension)
</td><td colspan="2">Result
Spline required the most amount of data preparation (. e.g. Simplifying contour lines, converting them to points, and then running the process) for mediocre results. Elevation was replicated decently, but some areas were visually off. In particular, the tension spline produced low points in the northern lakebed that were ~40m deeper than they are in reality.</td></tr>
</table>

Figure 2.d Spline (Regularized)	**Result**
	Using a regularized spline produced the worst results by far. Areas were shown as being up to 300m lower than they are in reality, appearing as low as -150m.
Figure 2.e TopoToRaster (T2R)	**Result**
	The simplest tool and also the best. By running the T2R with contour lines, and using rivers and lakes as boundaries, we were able to get a surface that was highly accurate to the original data.
Figure 2.f Kriging (ordinary)	**Result**
	Kriging provided poor results. Loss of accuracy was present across the surface, with non-existent high- and low-points being created, and existing variations in elevation being misrepresented.
Figure 2.g Natural Neighbor	**Result**
	Nearest neighbor provided results that did a terrific job of maintaining the shape of elevation features, although the size of some shapes was exaggerated or understated.

3.2 Geostatistical Analyst extension and Exploratory Spatial Data Analysis (ESDA).

***Figure 3** (a-d). Histograms (a-b) and QQ-Plots (c-d) used to verify the normality of the contour points (a-c) and the spot heights (b-d) to determine potential and optimal interpolation models.*

Figure 3.a Histogram of the contour points.

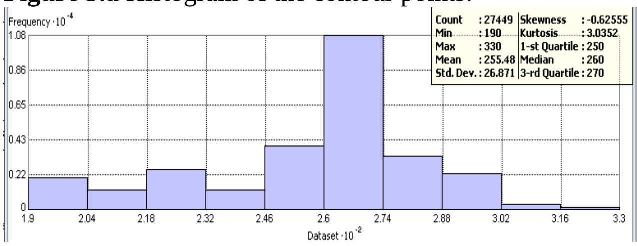


Figure 3.b Histogram of the spot heights.

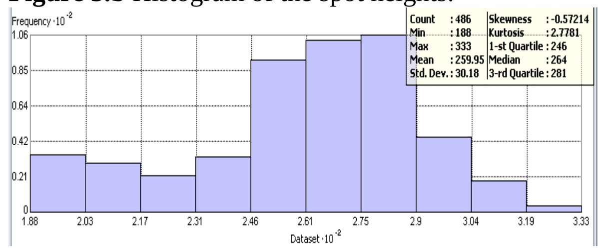


Figure 3.c QQ-plot of the contour points.

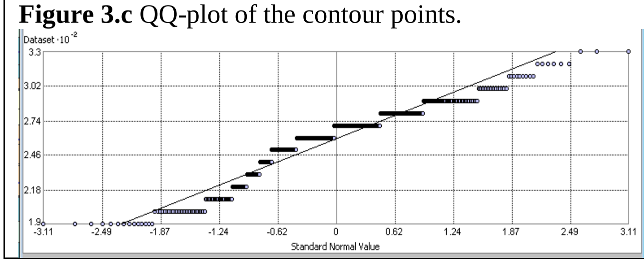
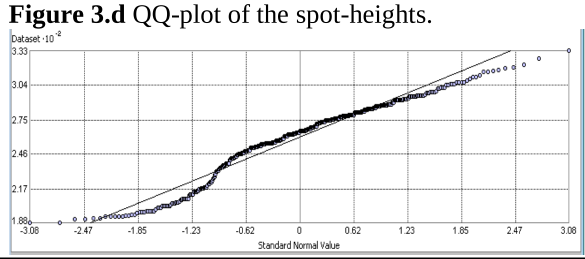


Figure 3.d QQ-plot of the spot-heights.



Although both spot heights and contour points appear to be moderately skewed towards areas of higher elevation (. i.e. > 260 m), both feature classes seem to be displaying normally distributed data. This can further be confirmed when looking at the statistics displayed in figure 1 (a, b). The mean (255.48 m) and median (260 m) of the contour points (Figure 1.a.) were fairly similar, the skewness was near 0 (Skewness = -0.625) and the kurtosis-value was approximately 3 (Kurtosis = 3.0352). A similar statement can be said for the spot height data (Figure 1.b.) (mean = 259.95 m; median = 264 m; Skewness = -0.572; Kurtosis = 2.7181). Comparable results were produced with subsets of the data (. i.e. 50%), however, they will not be presented in this report. The 10-meter intervals associated with the contour points (Figure 3.c.) created linear QQ-plot, comparatively to the spot heights which used precise values (Figure 3.d.).

Figure 4 (a-b) Trend analysis graph to determine the presence of any directional trend using contour points (a) and spot heights (b).

Figure 4.a Trend analysis graph of the contour points.

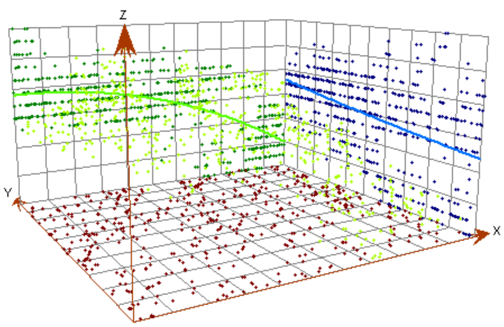
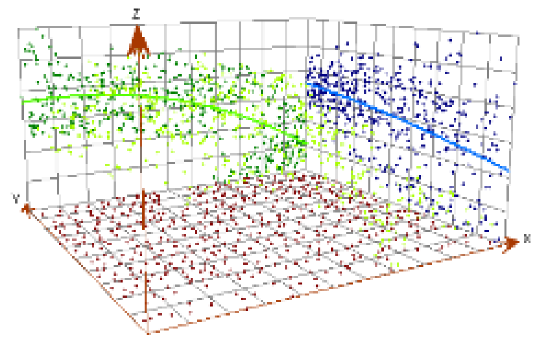


Figure 4.b Trend analysis graph of the spot heights.



Both contour points (Figure 4.a.) and spot heights (Figure 4.b.) displayed directional trends. On the X-axis (West-East), a gradual decline in elevation towards the eastern portion of the 031D8 OBM grid was evident (or incline towards the western portion of the grid), and on the Y-axis (North-South), a gradual increase in elevation towards the northern section of the grid was evident (or decrease towards the southern portion of the grid). Once again, the contour points (Figure 4.b.) displayed more linear results.

Figure 5 (a-b) *Voronoi map displaying the stationarity of the data for contour points. Two methods were attempted, simple (a) and entropy (b).*

Figure 5.a. Voronoi map using the simple method.

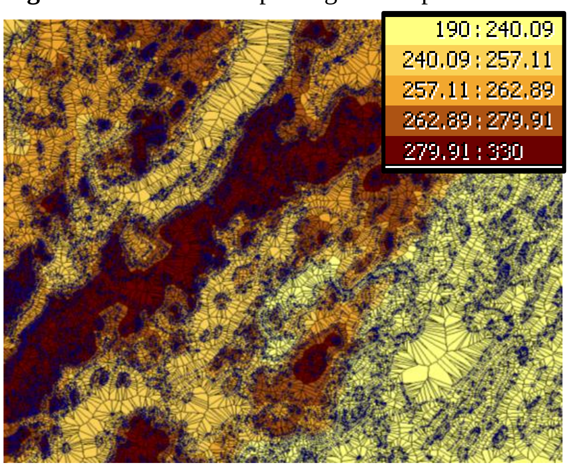
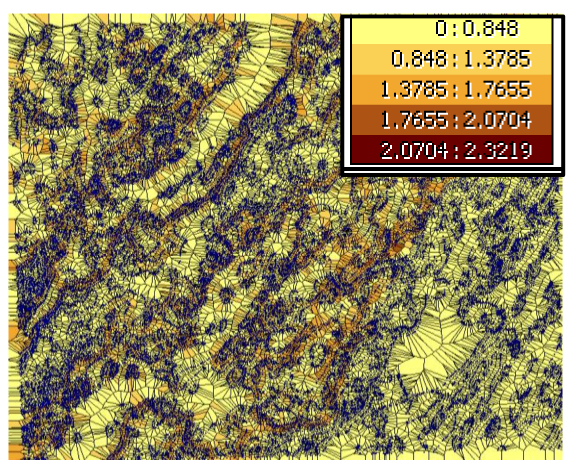


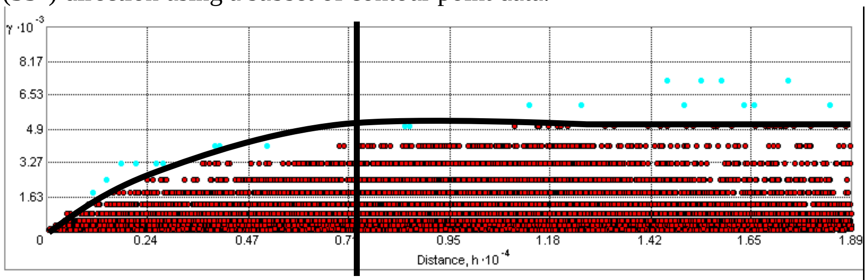
Figure 5.b. Voronoi map using the entropy method.



The Voronoi map with the simple method (Figure 5.a.) did a fairly good representation of the 031D8 grid as we can clearly identify where the drumlin is located (darkest portion of the map). However, this demonstrates that values are dependant on their exact location and not necessarily the distance between them. This could potentially be considered a violation of stationarity. The entropy method (Figure 5.b) did not produce representative results and were difficult to interpret.

Figure 6 *(a-c) A semi-variogram cloud used to determine spatial autocorrelation. The angle of the semi-variogram was adjusted to determine the directions of minimum (a) and maximum (b) sill and range values. Clouds were created with a subset (. i.e. 3.5 %) of the contour points. This figure also displays areas with the greatest outliers (c).*

Figure 6.a. Semi-variogram with the lowest sill (~4.9 γ -10^{-3}) and range (~0.71 h -10^{-4}) in a southwest-northeast (35°) direction using a subset of contour point data.



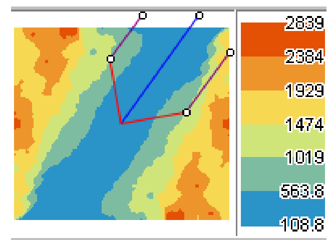
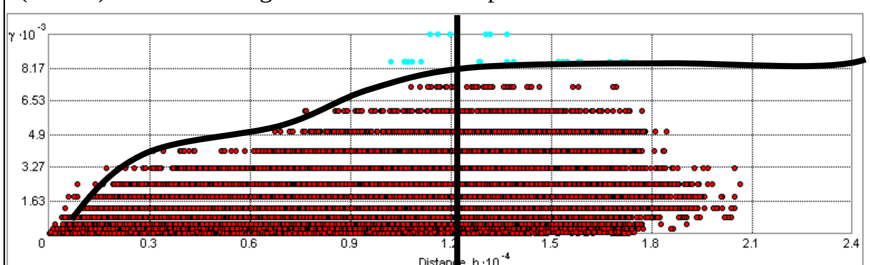


Figure 6.b. Semi-variogram with the highest sill (~8.1 9 γ -10^{-3}) and range (~1.171 h -10^{-4}) in a northwest-southeast (284.5°) direction using a subset of contour point data.



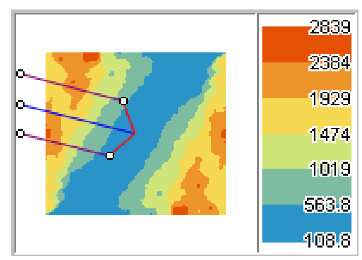
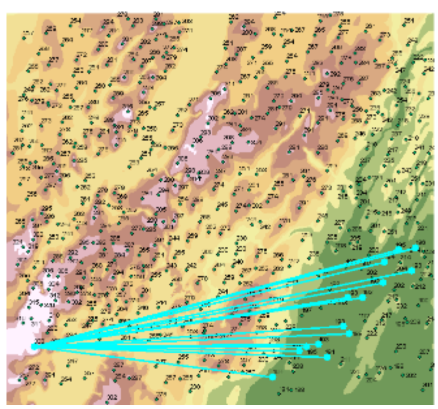


Figure 6.c. Figure displaying where the outliers are located in the data, affecting the spatial autocorrelation.



The semi-variogram displayed a direction trend in the data, which was previously confirmed with the trend analysis (Figure 4). Figure 6.a demonstrated spatial autocorrelation in the data until the range levels-out (~0.71 h -10^{-4}) and represented the lowest sill (~4.9 γ -10^{-3}) in a southeast to northwest direction (35°). Figure 6.b also appeared to display a directional trend, however, the direction appeared to be in a northeast-southwest direction, where the highest sill was observed (~8.1 9 γ -10^{-3}). This observation is further confirmed by Figure 6.c, which displays the majority of outliers being in the southwest (284.5°) portion of the grid.

3.3 Geostatistical Analyst extension and Interpolation models (Post-ESDA)

Figure 7 *(a-e) Results from geostatistical analyses (a-c) and deterministic model (d).*

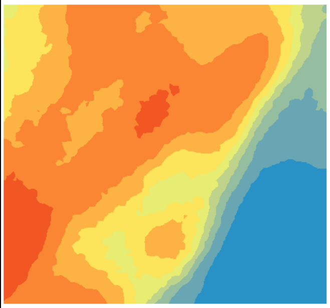
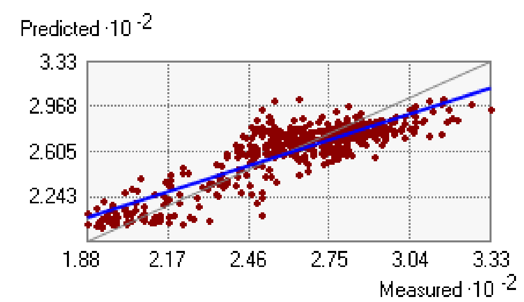
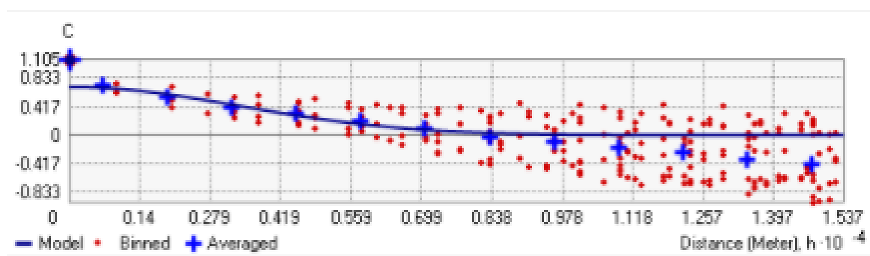
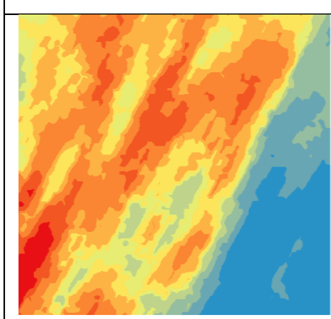
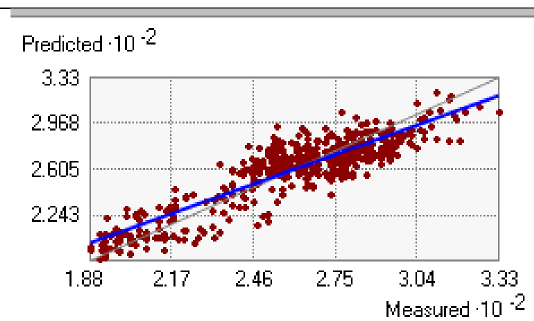
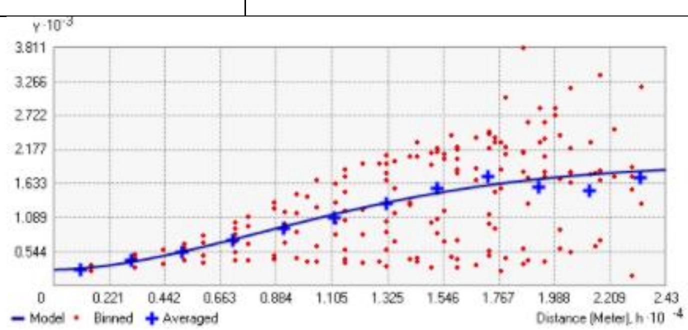
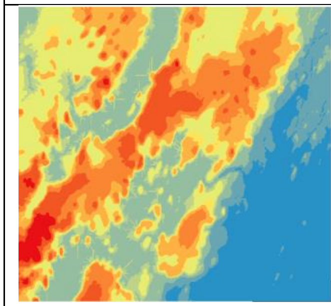
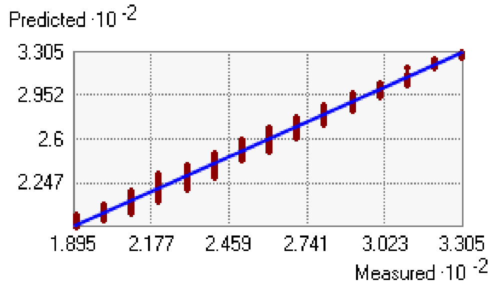
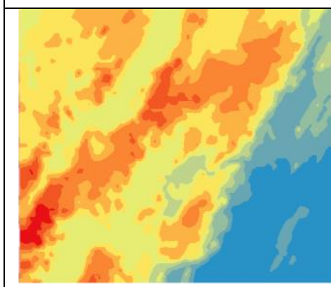
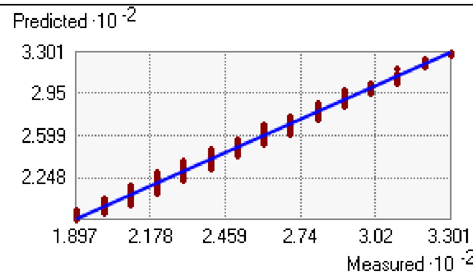
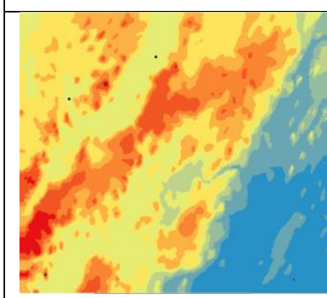
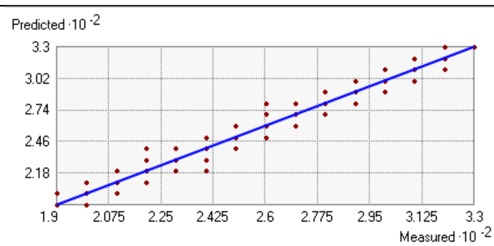
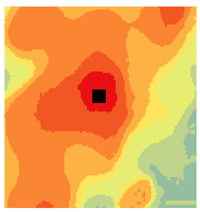
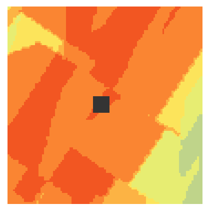
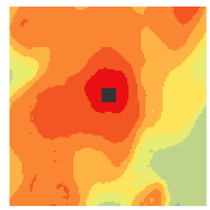
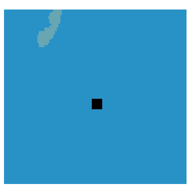
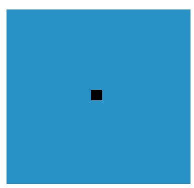
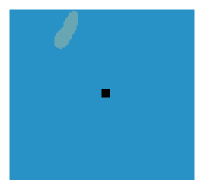
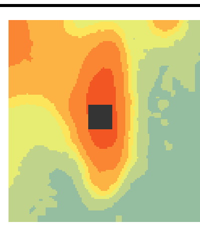
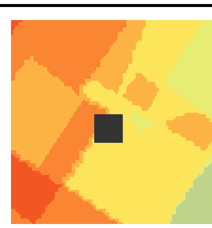
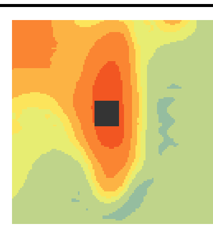
<table>
<tr><td colspan="3">Figure 7.a Simple kriging with data from spot heights.</td></tr>
<tr><td></td><td>
</td><td>RMS-Standardized: 0.882383
Mean Standardized: -0.01263158
Avg Standard error: 18.36772
RMS: 15.20299
Difference between Root-Mean-Square and Average Standard Error: -3.16473</td></tr>
<tr><td colspan="2">
</td><td>Lag Size: 1280.735
Number of lags: 12</td></tr>
<tr><td colspan="3">Figure 7.b Universal Kriging with data from spot heights.</td></tr>
<tr><td></td><td>
</td><td>RMS-Standardized: 0.8942589
Mean Standardized: 0.009754824
Avg Standard error: 15.25399
RMS: 13.63839
Difference between Root-Mean-Square and Average Standard Error: -1.6156</td></tr>
<tr><td colspan="2">
</td><td>Lag Size: 2024.987
Number of lags: 12</td></tr>
</table>

Figure 7.c Universal Kriging with data from contour points.		
		**RMS-Standardized:** 0.8965142 **Mean Standardized:** 0.01371116 **Avg Standard error:** 1.455773 **RMS:** 1.328321 **Difference between Root-Mean-Square and Average Standard Error:** -0.127452
Figure 7.d Empirical Bayesian Kriging with data from contour points.		
		**RMS-Standardized:** Na **Mean Standardized:** Na **Avg Standard error:** 1.623188 **RMS:** 1.179419 **Difference between Root-Mean-Square and Average Standard Error:** -0.443769
Figure 7.e Inverse Distance Weighting (deterministic) with data from contour points.		
		**Mean:** 0.1092936 **RMS:** 2.35165

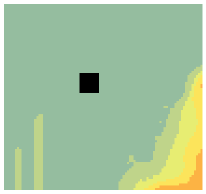
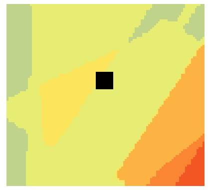
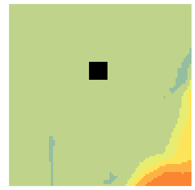
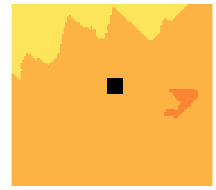
When visualizing the RMS standardized values, universal kriging with the contour points (Figure 7.c) appears to have the closest value to 1 (RMS standardized = 0.8965142). Universal Kriging with spot heights (Figure 7.b) had the best standardized mean value (standardized mean value = 0.009754824) with a value nearest to 0. The lowest average standard error (1.455773) was attributed to Universal Kriging with contour points data (Figure 7.c.), however, the lowest RMS value (1.179419) was attributed to the Empirical Bayesian Kriging (Figure 7.d.). Finally, the smallest difference between RMS and average standard error (-0.127452) was detected with Universal Kriging using the contour points (Figure 7.c.). The results indicate that universal kriging is likely the best available model to be used with the current data and simple kriging (Figure 7.a.) was the worst fit, hosting the least favorable values for nearly each cross-validation statistic. Although difficult to compare due to unproduced cross-validation statistics (.i.e. RMS-standardized, Mean-standardized), Empirical Bayesian Kriging produced a favorable output (Figure 7.d.) when interpolating contour points. Although it is difficult to compare IDW with the geostatistical methods used, visually it appears to be rather similar to Empirical Bayesian Kriging.

3.4 Quality Assurance and Quality Control (QA/QC)

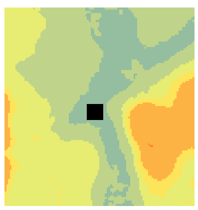
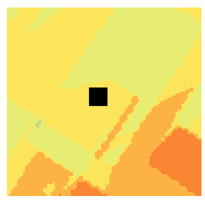
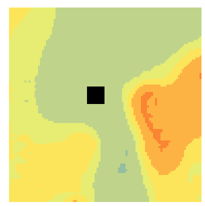
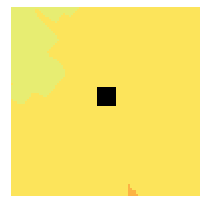
Table 5 *First QA/QC method used to compare each kriging method to one another and compare interpolation methods using spot height and contour points.*

Spot	**Z- Universal (contour points)**	**Z - Universal (spot heights)**	**Z - Empirical (contour points)**	**Z - Simple (spot heights)**
A X: -78.437764 Y: 44.361716	 310.28m	 287.72m	 310m	 278.56m
Looking at the peak of a drumlin, universal and empirical kriging both produce similarly high-quality results when it comes to shape and height. Universal kriging done with spot heights loses the shape of the drumlin and is less accurate when it comes to height. The default simple kriging shows poor interpolation of shape and height.				
B X: -78.301792 Y: 44.269765	 200m	 205.37m	 200m	 205.22m
Examining the low areas south of Peterborough gives more consistent results with the various methods tested. With no significant variation in contours, there wasn't much height to be lost in the interpolation process. The two methods run using spot heights (the second universal kriging and the default simple) have a slight edge up here in terms of accuracy though, given that the contour point-interpolated rasters are based on data with even, 10m intervals.				
C X: -78.486729 Y: 44.259996	 280m	 280.63m	 291.14m	 284.60m

As with the previous drumlin, the methods that used contour points produced favorable results. Interestingly though, while the shapes are lost upon interpolation with the other methods, the disparity between the heights given from all the methods is much less noticeable here, with only 11m of difference between the best and worst results.

D				
X: -78.486729 Y: 44.377403	 250m	 266.76m	 250m	 278.26m

The lake in the northwestern corner of the map produced diverse results across the methods. While both universal (contour points) and empirical recorded the exact same heights, their shapes differ slightly, with universal having some strange linear contour shapes created during the interpolation process. Universal (spot heights) was less accurate in terms of shape and elevations, and the default kriging raster was inaccurate in both regards.

E				
X: -78.462526 Y: 44.356582	 250m	 266.18m	 250m	 271.37m

Lastly, the river valley in the northwestern corner of the map produced some interesting results that showcased the accuracy of universal kriging.

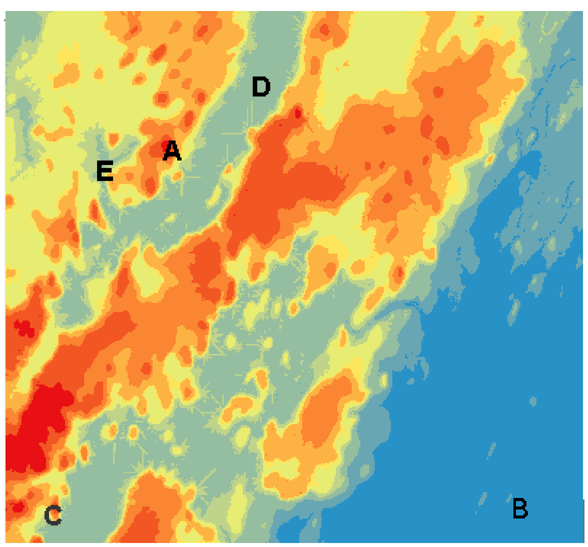
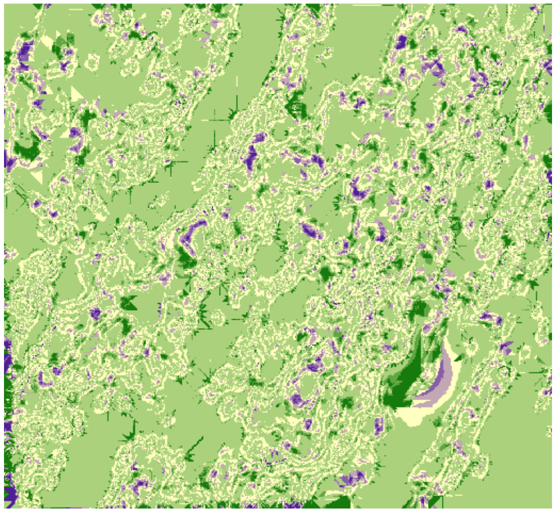
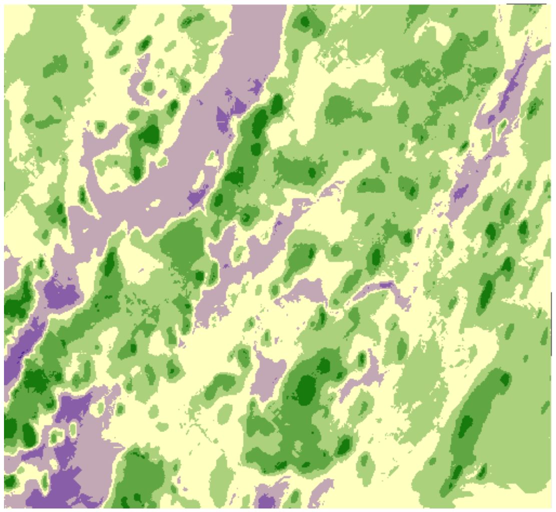
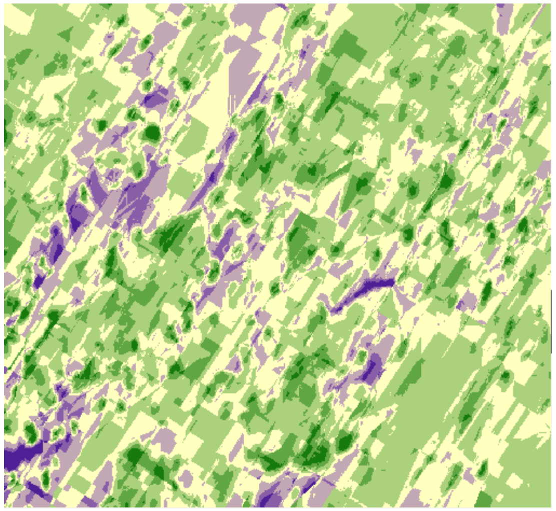
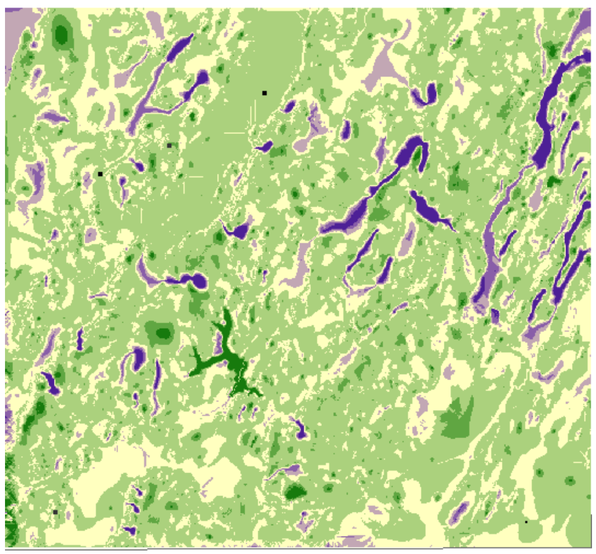


Table 6 Second of the QA/QC methods used to compare the better of the interpolation methods to the worst interpolation methods.

Comparison	Image	Differences	Values
Universal (contour points) vs. Empirical Bayesian (contour points)		In comparing our two best results (a universal kriging grid vs. an empirical Bayesian grid), noticeable differences occur near the south-western side of Peterborough, in both positive and negative deviations. Overall the map is largely consistent, with the majority of surfaces being within 1m of each other when comparing the two methods. Other areas of deviation occurred along the drumlins, and along the edges of the surface beyond the contour lines and spot heights used to generate the raster.	Most of the surface generated from the raster calculator came out with fairly even results, with the majority of cells coming back between -0.5 and +0.5. The areas of the most extreme deviation occurred near Peterborough, with deviations as low as -12 and as high as 16
Universal (contour points) vs. Simple (spot heights)		Comparing the best surface with our worst (a modified universal kriging grid vs. a simple kriging grid with default parameters) showed just how much of a difference using the right interpolation method can make. Large swaths of land in the lake beds exhibited significant deviation, as did most of the highest and lowest parts of the map.	Significant deviation was present in the resulting calculated raster, with values ranging from -42 to 21. The highest deviations occurred in the valleys in the south-western corner of the map, and along the tops of the drumlins throughout the raster. The majority of the map was between -19 and 1.4.

Universal (contour points) vs. Universal (spot heights)		The contrast between using universal kriging with spot heights vs. contour points, while not quite as strong as it was with universal vs. simple, is still quite apparent. The results generated by using spot heights rather than contour points has led to strong deviations, located primarily in the lakebed in the north-east and with occasional deviations along the tops of the drumlins.	Deviation ranged from -56 to 17, and was strongest along the valleys and tops of the drumlins. The rest of the map sat somewhere between 0.6 and -16.
Topo to Raster vs. Universal (contour points)		In comparing the best results from the Spatial Analyst and Geostatistical Analyst toolkits, the difference that having properly defined contour lines can make becomes quite apparent. With Topo to Raster allowing the degree of precision that it does (and the ability to use lakes and rivers as barriers), the contrast between the two methods becomes readily apparent in the fine details of the surface: namely, the lows around the rivers and the ridges atop the drumlins.	Deviation ranged from -46 to 10, with the most significant lows occurring in along the rivers in the northern half of the surface.

4 Discussion

4.1 Interpolating using the Spatial Analyst extension

Out of all the DEM models interpolated with the spatial analyst extension, TopoToRaster produced the greatest output (Figure 2.e). TopoToRaster imposes constraints that ensure a hydrological connectivity within the output by correctly representing ridges and streams from input contour data, and therefore benefiting from multiple inputs to ensure precise data (ESRI, 2004). The spline tool was the worst of the interpolation methods and required the most preparation (Figure 2.c-d). This is likely because spline uses mathematical functions that minimize overall surface curvature (ESRI, 2004). For this reason, the DEM's produced by both spline methods produced extremely inaccurate results, with certain areas interpolated into negative values. Spline interpolations are best suited for surfaces that vary smoothly, such as water table heights (ESRI, 2009).

4.2 Geostatistical Analyst extension and Exploratory Spatial Data Analysis (ESDA).

The exploratory spatial data analysis (ESDA) tool provided by the geostatistical analyst extension allows for greater data observation and exploration than that of the spatial analyst extension. This tool allowed for a proper analysis prior to conducting any kriging interpolations to know if we should apply deterministic models or geostatistic models. The first assumption to check for a geostatistic model was looking at normality. Although our histogram appeared slightly skewed, the skewness, kurtosis value, the similarity between mean and median and, the QQ-plot, confirmed that our data was normally distributed. The second assumption is that all the values are spatially auto-correlated, meaning that sampling points in proximity of one another should have a greater similarity than points which are further apart (University of Omaha, 2018). This rule was likely violated, as sampling points located at the top of the Peterborough drumlin field caused for outliers with areas located at the lower elevations (. i.e. the city of Peterborough). A similar study conducted by Maclachlan & Eyles (2013), also found minimal spatial randomness within the Peterborough drumlin field and found directional trends with similar values as we did. However, geographical data is generally affected by its location (GEOM105, 2018). Knowing that directional trends were present in the data allowed for a better understanding of which kriging method would be better suited for the DEM model.

4.3 Geostatistical Analyst extension and Interpolation models (Post-ESDA)

Universal kriging appeared to be the better of the interpolation methods used in the geostatistical analyst extensions. This is because universal kriging accounts for a directional trend in the data and therefore uses a different mathematical approach to compensate (ESRI, 2004; GEOM105, 2018). Simple kriging was the worst of the DEM models utilized and this is likely because it does not assume any trends in the data (GEOM105, 2018). Although certain assumptions were violated prior to using geostatistical interpolation models, the deterministic model used (. i.e. IDW) did not represent the data adequately. Again, this value does not account for any directional trend and predicts values within proximity to be more similar (ESRI, 2004). Surprisingly, however, IDW appeared visually as being a better fit for the model than Simple Kriging. This might suggest that when geostatistical assumptions are violated, which is often the case with geographical data, deterministic models might be a more suitable fit.

4.4 Quality Assurance and Quality Control (QA/QC)

QA/QC is definitely a good method for comparing results next to one another to gain further insight. One of our primary observations was that conducting interpolation with spot heights is not as beneficial as contour points as spot heights lose a lot of essential details (Figure 8). Another interesting finding from this project was that comparing interpolation methods at low elevation (Figure 8.b) does not allow for efficient visual representation. Comparisons should be conducted at the highest elevations (Figure 8.a) or areas of rapid change (. e.g. sharp slope) to gain a better perspective on the differences between interpolation methods.

The second QA/QC method (Figure 9) is also a great method for comparing interpolation methods as it allows for a global view of the entire study area, however, it has the disadvantage of only displaying 2 models at a time. This method was particularly useful when displaying drastic differences in the data, such as Figure 9.b and determining what are the exact causes of issues with the interpolation model.

5 Conclusion

Overall, the objectives of this assignment were all met. We modeled continuous elevation surfaces from two sets of sampling points (. i.e. spot heights and contour points) and concluded that contour points are the better option if contour lines are available, as they display greater precision. We found that *TopoToRaster* was the better and simpler of all interpolation methods when conducting elevation models, being able to not only directly input contour lines, but also hydrological data for the greatest precision. And finally, we determined that the spatial analyst extension, although simple and easy to use for pre-processing and interpolation models, does not have the flexibility of the geostatistical analyst extension which allows much more in terms of pre-processing the data and modifying parameters when interpolating the data. That being said, the increase in options with the geostatistical analyst tool implies that the user should be more comfortable with the statistics used in interpolation and the effects of modifying the parameters on the final product.

6 References

1. Avrun, P.V., 2013a. A comparative analysis of different DEM interpolation methods. The Egyptian Journal of Remote Sensing and Space Science, 16, 133-139. Doi.org/10.1016/j.ejrs.2013.09.001.

2. Avrun, P.V., 2013b. A terrain-based hybrid approach towards DEM interpolation. Geodesy and Cartography, 39, 245-252. DOI: 10.1080/19475683.2013.843590.

3. Chang, K., 2016. Introduction to Geographic Information Systems, 8th Edition (New York, NY: McGraw-Hill), 425 pages with companion CD.

4. ESRI., 2004. Interpolating Surfaces in ArcGIS Spatial Analyst. http://webapps.fundp.ac.be/geotp/SIG/interpolating.pdf.

5. ESRI., 2009. Geoprocessing and surface creation. http://resources.esri.com/help/9.3/arcgisengine/java/gp_toolref/geoprocessing/surface_creation_and_analysis.htm 2009

6. ESRI., 2010. Feature Vertices to Point. http://pro.arcgis.com/en/pro-app/tool-reference/data-management/feature-vertices-to-points.htm.

7. ESRI., 2017. Performing Spatial Interpolation Using ArcGIS. https://www.esri.com/training/catalog/5763042d851d31e02a43ed97/performing-spatial-interpolation-using-arcgis/

8. Geocaching., 2018. https://www.geocaching.com/geocache/GC15AM7_peterborough-drumlin-fields?guid=d4c79ae4-ebee-4778-9062-6a073ff201f3.

9. Geokov., 2018. http://geokov.com/education/map-projection.aspx.

10. GEOM105., Week 2 lecture-Geostatistical Methods of Spatial Interpolation. https://fleming.desire2learn.com/d2l/le/content/78185/viewContent/880579/View.

11. Lotimer, L., 2014. Subsurface Stratigraphy and Hydrogeology of the Peterborough Drumlin Field, Southern Ontario, Canada. McMaster University. URL: http://hdl.handle.net/11375/16072.

12. Maclachlan, J.C. and Eyles, C.H., 2013. Quantitative geomorphological analysis of drumlins in the Peterborough Drumlin Field, Ontario, Canada. Geografiska Annaler: Series A, Physical Geography, 95, 125–144. DOI:10.1111/geoa.12005.

13. NSIC., 2018. https://nsidc.org/cryosphere/glaciers/gallery/drumlins.html

14. University of Omaha., 2018. The principles of geostatistical analysis. http://maps.unomaha.edu/Peterson/gisII/ESRImanuals/Ch3_Principles.pdf

15. University of Toronto., 2018. Selecting the right projection. https://mdl.library.utoronto.ca/technology/tutorials/selecting-right-projection.