A Multi-criteria Decision Analysis of the Risk of Injury to the Public from Rock Falls at Worbarrow Bay, Southern UK

Background Information and Objectives

This exercise outlines the key steps taken in a multi-criteria decision analysis to identify areas where members of the public may be at risk from rockfall at Worbarrow Bay, Lulworth Cove on the southern coast of the UK. Worbarrow Bay lies at the centre of the Dorset and East Devon Coast World Heritage Site, otherwise known as the Jurassic Coast, which receives over 17 million "visitor nights" each year (May, 1993). The area is popular because of its historical, spiritual and cultural significance, as well as the unique geological features here that are often the subject of school and university fieldtrips (Figure 1). In recent years, numerous rockfalls and landslides along this section of coastline have led to hazard assessment becoming a management priority for this coastal landscape.

Figure 2 illustrates a multi-criteria analysis to identify geographical areas that pose a high risk of injury to members of the public from rockfall events. High risk areas are defined as those that meet three equally weighted criteria relating to proximity to footpaths, topographic slope and the nature of the underlying geology:

- Criterion 1: Areas located within 50m of footpaths,
- Criterion 2: Areas with a slope gradient exceeding 10%, and
- Criterion 3: Areas underlain by softer, less consolidated chalk deposits.

Geographic areas meeting all three of these criteria were identified as being of high risk. The practical value of such an analysis is that it identifies geographical areas where management activities. These might include the introduction of signage and handrails to reduce risk of injury. As the Jurassic Coast is 155 km long, and a lot of this length is subject to intense pressure from tourism, identifying areas in which to focus management resources can produce significant practical efficiencies.

Figure 1 Worbarrow Bay, east of Lulworth Cove. Headland and bay coast-line exposing strata from Portland Stone to Chalk. Features include A Iron Age hill-fort on chalk ridge, B recent landslides in chalk cliffs with talus ex-tending below sea level, C coastal path at risk from cliff-top retreat, D stacks resulting from erosion of limestone steeply dipping strata which continue as a submerged ridge to the headland in the left foreground, E incised dry valley in chalk, F landslides in chalk, sands and clays, G shingle beach with occasional landslides.

Data

You are provided with an aerial photograph of Worbarrow Bay that can be used as a backdrop onto which the data layers can be overlaid (filename "Worbarrow_Bay_AOI' note: it is not used as a data layer for the analysis specified), the locations of the footpaths ('Footpaths'), a geological map ('Geology_AOI') and a map of areas where the slope exceeds 10% ('Slope_over_10_AOI').

Data Sources: The aerial photograph was supplied by Infoterra UK, footpaths were digitised by the author at a scale of 1:500 from the aerial photograph, geological classes were also digitised by the author at a scale of 1:500 from the Ordnance Survey Geological Map of the British Islands (available from DigiMap), and areas where the slope exceeded 10% were generated from a 20 m digital elevation model (also available from the topographic data layers of Digimap).

Instructions

It may be useful to work in the model maker facility of ArcGIS so that you are able to outline and implement this multi-criteria analysis as a series of sub-steps.

Create a model that implements each of the three criteria listed above, then bring those criteria together into a final step to identify areas that are "At risk" of rockfalls.

- Criterion 1: Areas that were located within 50m of footpaths:
  Drag the Footpaths layer into the dataframe

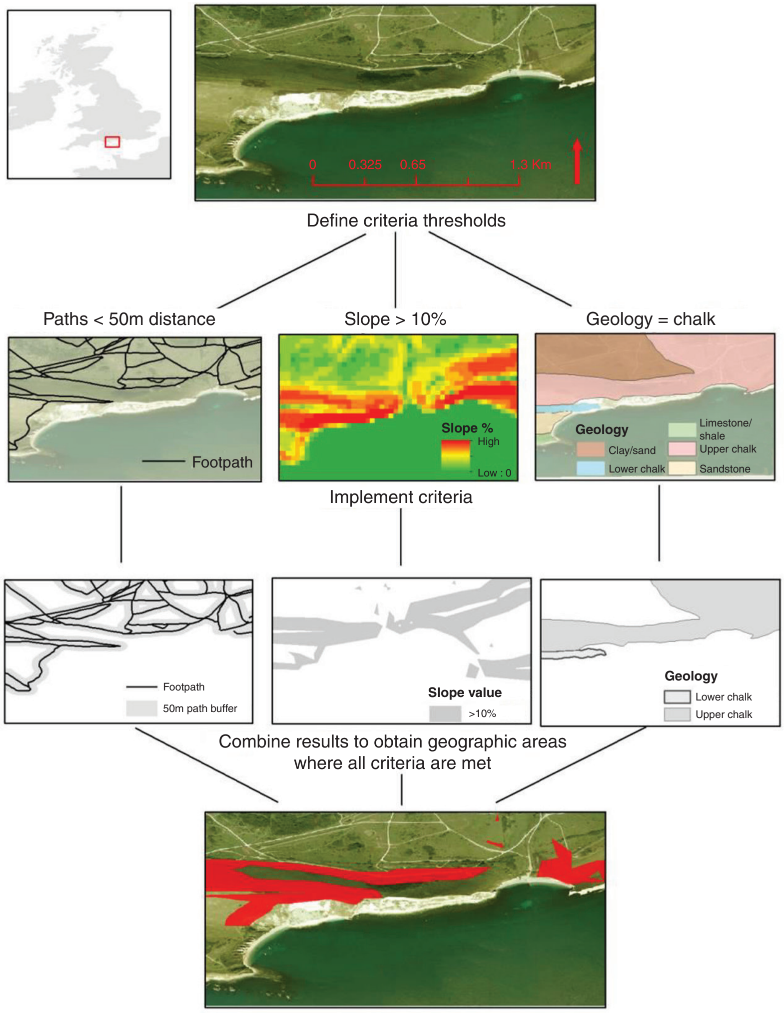


Figure 2 A hypothetical illustrative example of multi-criteria spatial analysis at Worbarrow Bay near Lulworth Cove along the coastline of southern England. The risk of injury from rock falls can be modelled based on distance to footpath, topographic slope and underlying geological character. Text boxes indicate key parts of the process of multi-criteria analysis.

Use the buffer tool (Analysis – Proximity – Buffer) to identify areas within 50 m of footpaths. Specify '50' as the distance and metres as the units from the drop down menu. This will output a vector polygon shapefile that covers all areas within 50 m of footpaths.

- Criterion 2: Areas that had a slope gradient that exceeded 10%:

  This data layer has been provided and can go straight into the final model stage that brings the multiple criteria together

- Criterion 3: Areas that were underlain by softer, less consolidated chalk deposits.

  Drag the Geology layer into the data frame. Open the attributes table and you will see an attribute called 'Geol_class' that specifies the underlying geology for each part of the map. We are interested in the 'Lower chalk' and 'Upper chalk' classes. Use the Select tool (Analysis tools – Extract – Select) to define an SQL Expression that selects the classes of interest. This will take the following form:

  "Geol_class" = 'Lower chalk' OR "Geol_class" = 'Upper chalk'

  Once the correct classes have been selected, you should have generated a new file identifying all areas underlain by chalk

- The final step of the analysis brings together the three layers relating to each of the criteria specified. Only areas that meet all three of the criteria are identified as being at risk of rockfalls.

  Use the intersect tool (Analysis – Overlay – Intersect) to identify areas on the map that meet all three criteria specified. Put in all three of the datasets as input features.

Calculating Changes to the Lagoon Volume at Diego Garcia Atoll, Chagos Islands

Background Information and Objectives

Diego Garcia is a horse-shoe shaped atoll (9 × 22 km^2) that lies 55 km south of the Great Chagos Bank in the central Indian Ocean (Figure 1). Approximately 70% of the atoll area is comprised of the extensive lagoon (11 km^2), which is enclosed by a mostly continuous land rim around the periphery, which has an open channel to the North. In the 1970s an American air base was built along the north western sector of Diego

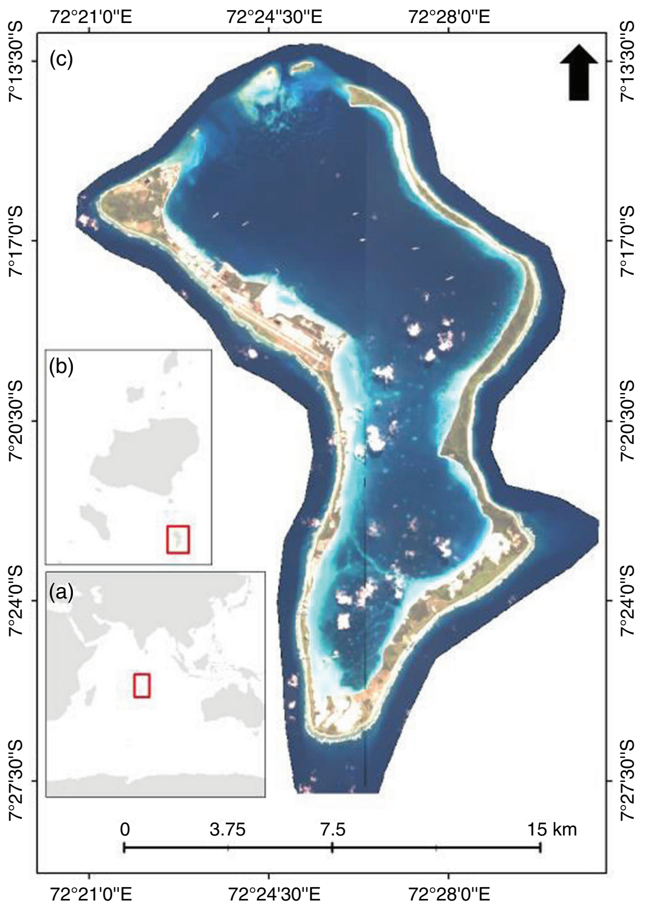


Figure 1 (**a**) The location of the British Indian Ocean Territory in the central Indian Ocean, (**b**) the location of Diego Garcia at the southern end of the Chagos archipelago (BIOT), and (**c**) Diego Garcia atoll.

Garcia atoll rim, thus, recent anthropogenic influences associated with the construction of military infrastructure include dredging and dumping of sands within the lagoon.

Data

For this exercise, two raster layers are provided that depict bathymetric models of the lagoon at Diego Garcia Atoll, Chagos Islands. These have been produced by digitising 2590 numbered points into a georeferenced point file from hydrographic charts. These were developed from the basis of two surveys of the atoll lagoon, undertaken in 1967 (H.M.S. Vidal) and 1998 (UK Hydrographic Office) (Figure 2). A kriging interpolation was applied to digitised sounding points to derive a continuous bathymetric model representing water depth across the entire lagoon floor. More detail on the collection of sounding points and production of the bathymetric model can be found in Hamylton and East (2012).

Data Sources: Hydrographic charts were accessed via Cambridge University Map Library and digitised by Holly East.

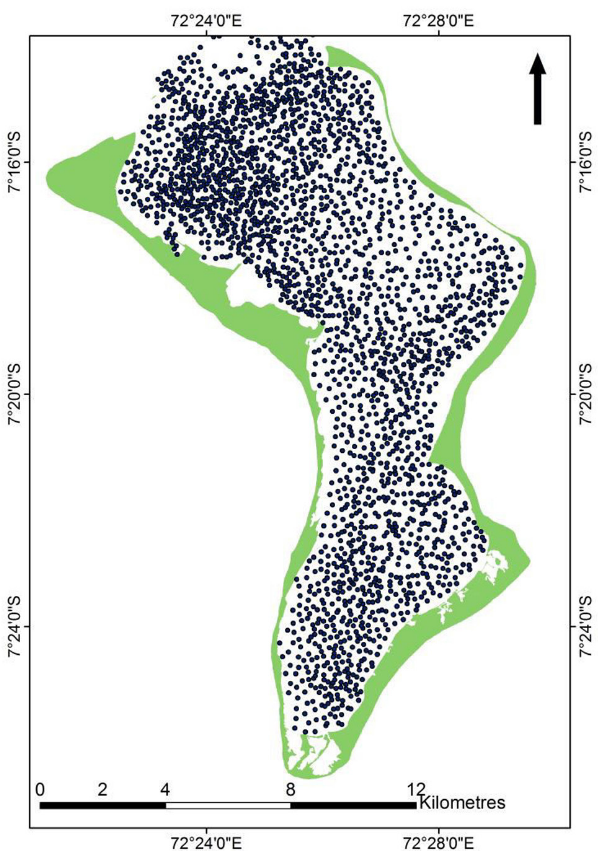


Figure 2 Point file displaying the 3,584 soundings from the 1998 UK Hydrographic Office bathymetric chart.

Instructions

Overall change in lagoon volume can be estimated via a two-step process that calculates a 'change raster' to estimate bathymetric change on a pixel by pixel basis (step one), then sums these to generate overall estimates of change for the entire lagoon (step two).

Note that conventionally, bathymetric maps are datasets that represent water depths as negative values because they are beneath the sea surface. A change raster can be calculated by subtracting the earlier survey (1967) from the later survey (1998). This can be achieved in the raster calculator (spatial analyst, map algebra). This will produce a new layer with positive values where the water depth has reduced (i.e. sediment has accumulated) and negative values where the water depth has increased (i.e. sediment has been removed) (see Figure 3).

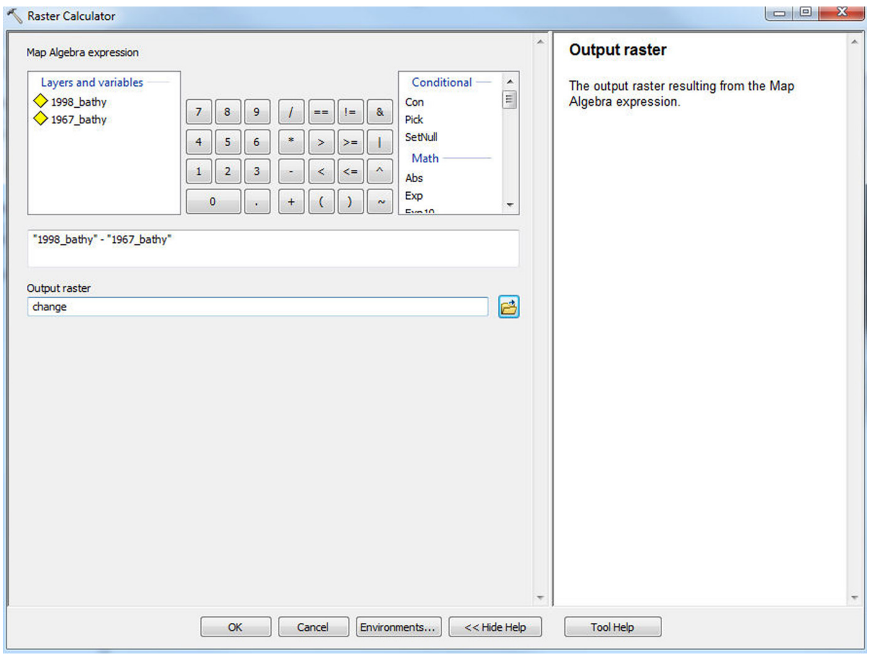


Figure 3 Calculation of net change in sediment volume between two raster bathymetry layers of the lagoon floor at Diego Garcia.

The net change in volume of sediment between the two different survey times can be calculated by summing across the raster pixels that comprised the bathymetric change model and subtracting those that represented sediment loss from sediment accumulation (Equation 1).

$$\Delta VSed = \sum VSed_{ac} - VSed_{rem} \tag{1}$$

Table 1 Per pixel change in volume of the Diego Garcia Lagoon

Sediment accumulated			Sediment removed		
VALUE	**COUNT**	**Total (value * count*pixel volume/m³)**	**VALUE**	**COUNT**	**Total (value * count*pixel volume/m³)**
1	3031	378875000	–8	4	–4000000
2	560	140000000	–7	29	–25375000
3	97	36375000	–6	72	–54000000
4	23	11500000	–5	423	–264375000
5	16	10000000	–4	828	–414000000
6	11	8250000	–3	1603	–601125000
7	11	9625000	–2	3204	–801000000
8	13	13000000	–1	7990	–998750000
9	15	16875000	0	10575	0
10	15	18750000	Total/m³		–3162625000
11	9	12375000	Total/Tonnes		–3163
12	10	15000000			
13	6	9750000			
14	5	8750000			
15	4	7500000			
16	9	18000000			
17	3	6375000			
18	3	6750000			
19	8	19000000			
20	8	20000000			
21	9	23625000			
22	5	13750000			
23	5	14375000			
24	4	12000000			
26	2	6500000			
27	1	3375000			
Total/m³		116000000			
Total/Tonnes		116			

where $\Delta VSed$ = Net change in sediment volume,

$VSed_{ac}$ = Total volume of sediment accumulated,
$VSed_{rem}$ = Total volume of sediment removed.

The values of VSed and VSed ac can be calculated by summing up the total negative and total positive pixels values from the change raster derived in step one, then multiplying through by the volume of a single pixel. Pixels where the water depth has not changed (value = 0) are excluded from the analysis. Thus, estimated change (in Tonnes) for these datasets would be 116 − 3163 = –3047 Tonnes (see Table 1 for values)

Interpolating Sediment Samples from Lady Musgrave Island

Background Information and Objectives

This dataset has been selected to illustrate the how the different interpolations and symbologies applied to a point dataset can influence the way that the dataset is visualised and interpreted. The point data represent sediment samples and their associated characteristics derived from sediment size analysis (carried out on a Malvern Mastersizer laser particle analyzer). More information on the data can be found in the following publication:

Hamylton, S. M., Carvalho, R. C., Duce, S., Roelfsema, C. M., & Vila-Concejo, A. (2016). Linking pattern to process in reef sediment dynamics at Lady Musgrave Island, southern Great Barrier Reef. *Sedimentology*, 63, 1634–1650.

Data

The data provided for this exercise include a vector polygon shapefile "Reef_flat" and a vector point shapefile "Musgrave_SedimentGrabs". The polygon shapefile is an outline of the reef platform at Lady Musgrave Island, southern Great Barrier Reef. This vector polygon file outlines the reef flat (composed of three polygons indicating the outlines of the reef flat, island and lagoon). The point shapefile is a series of 24 sediment samples collected across the reef system at Lady Musgrave. The attributes table contains results of a sediment size analysis, including percentage composition for gravel, sand and mud and a measure of overall sediment size (Phi).

Data Sources: Sediment samples were collected by Dr Chris Roelfsema and Dr Rafael Carvalho. The reef outline was digitised by the author from a Geo-Eye satellite image of the reef.

Instructions

Part 1: Create Three Interpolated Layers

(Note: You may need to turn on the Spatial Analyst extension before you are able to use this tool via *customise, extensions*).

Use the interpolation tools within the spatial analyst toolbox to apply the following three interpolations to the Musgrave_SedimentGrabs dataset:

- Inverse distance weighting
- Spline
- Kriging

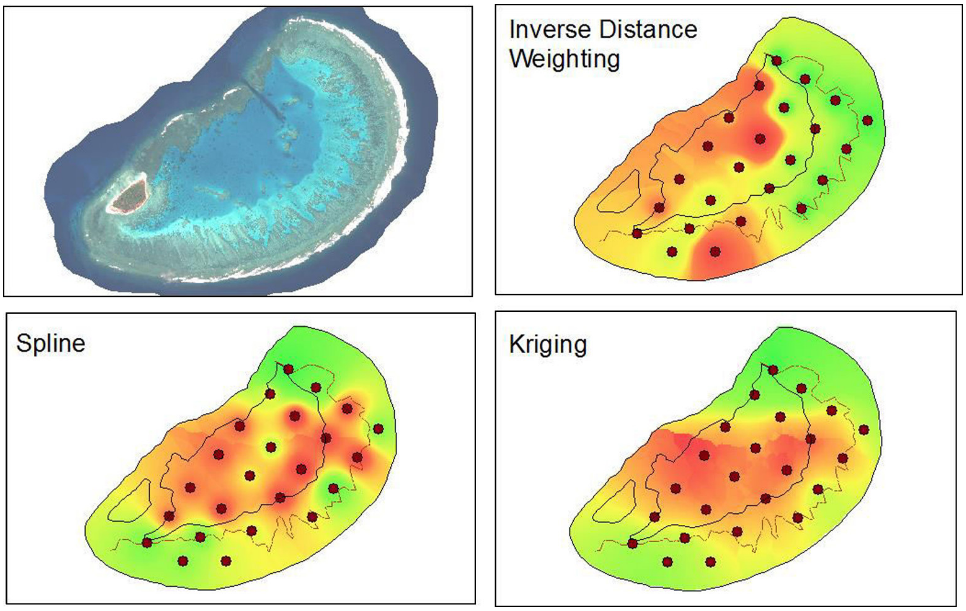


Figure 1 Upper left: Lady Musgrave reef, showing the island on the west side, the blue lagoon in the centre and the reef flat around the periphery. Illustrative examples of the interpolated raster layers using IDW, spline and Kriging methods.

The Z value field (i.e. the value within the dataset that you wish to interpolate) is the proportion of sand that was measured in each sample, labelled "Sand2". You may need to adjust the processing extent within the environments tab at the bottom of the tool so that the extent window says "same as layer Reef_flat". This means that your interpolated surface layer covers the entire reef platform.

Part 2: Adjust the Visualisation Properties of the Interpolated Layers

Clip the Interpolated Layer to the Extent of Lady Musgrave Reef:

One you have produced the three interpolated raster layers, you will need to clip them to the extent of the Musgrave reef flat (*Data Management, Raster, Raster Processing, Clip*). The clip function acts like a cookie-cutter to reduce the layer to the specific geographic area that you are interested in.

You will need to run the clip tool three times, each time setting the input raster to the interpolated raster layer you have produced, setting the output extent to that of the *Reef_flat shapefile* and putting a tick in the box that says *"Use input features for Clipping Geometry"*.

Adjust the Symbology of the Clipped Raster Layer:

Within the properties window of each clipped and interpolated raster dataset, apply the following symbology settings: Show = Stretched, select the color ramp that progresses from red to green (15^{th} option from the bottom in the drop down menu), within the stretch options, select histogram equalize and then add a tick in the invert box.

Part 3: Compare your Output Layers

Use the questions below as a basis for comparing the interpolated outputs of the sediments dataset:

- What are the upper and lower values for the inverse distance weighting, spline and kriging interpolated layers? How do these values compare to the range of the input dataset?
- Which areas appear to be associated with the largest concentrations of sand?
- Adjust the symbology settings (try changing the colours or the stretch type applied to the layer). How does this affect the way you interpret the output dataset?

Predicting Live Coral Reef Occurrence Around The Lizard Island Group, Great Barrier Reef

Background Information and Objectives

This exercise introduces quantitative methodologies for hypothesis testing using mapped variables. It tests whether there is a statistically significant relationship between physical variables (e.g. those related to bathymetry) and a biological variable (e.g. the presence of live coral). It covers classical regression (the ordinary least squares model) and spatial regression (spatially lagged autocorrelation). The objectives of Part 1 are to statistically test the relationship between the presence (or absence) of coral reef and physical parameters (in this case, we will be using a combination of bathymetry, rugosity, BPI and variety- all information layers that you generated in the last practical). In Part 2, the relationship defined in Part 1 will be used to predictively map the presence of coral reef around Lizard Island.

Data

Filename = Ground_ref.shp This dataset was collected from 365 snapshots of underwater video footage during an expedition to Lizard Island (December 7–17th, 2011). It contains coral reef benthic information combined with relevant physical information at the same point locations. It has been generated by applying the "Add surface information" to the underwater video shapefile to extract physical information into a series of new fields (in each case, the new "Z" field generated was copied across to a new field). Also supplied is a vector file indicating the shoreline around Lizard Island (Lizard_land_outline). Additional data supplied for part 2 include a model of water depth around Lizard Island (see Hamylton et al. 2015 for more information on how this was derived), associated terrain variables of variety (a measure of localised variability in water depth) and slope. Both variability and slope are derived using the Spatial Analyst tools in ArcGIS. Finally, wavebands 1 to 4 of a WorldView2 satellite image of Lizard Island are provided.

Data Sources: Field data for the Ground_ref shapefile were collected by the author, the water depth model and associated terrain variables were produced by the author from a combination of satellite imagery and field data. The shoreline file (Lizard_land_outline) was digitised by the author at a resolution of 1:500 from the satellite image.

Hamylton, Sarah M., John D. Hedley, and Robin J. Beaman. "Derivation of high-resolution bathymetry from multispectral satellite imagery: a comparison of

empirical and optimisation methods through geographical error analysis." *Remote Sensing* 7.12 (2015): 16257–16273.

Software: GeoDa (GeoDa is free spatial software that can be downloaded at http://geodacenter.asu.edu/).

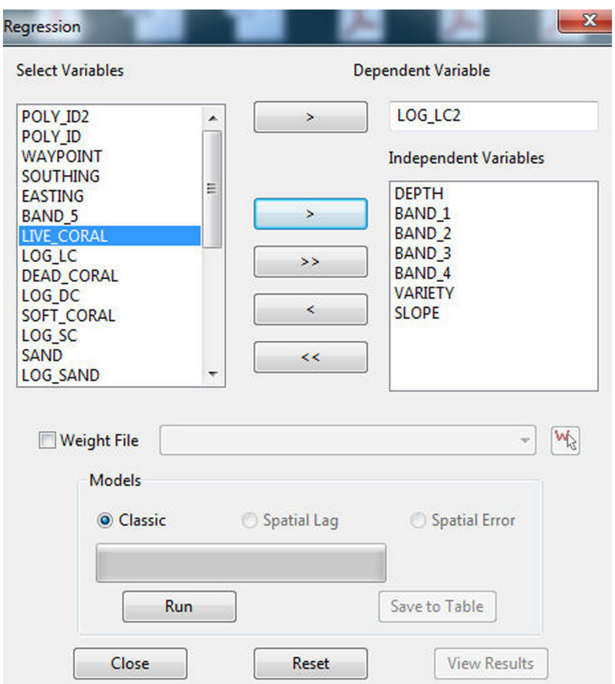
Instructions

Part 1: Modelling Coral Reef Occurrence

1. Open the point file
   File>open shapefile, or left click the icon to input the shapefile. Browse to where you have stored the Ground_ref shapefile and open.

2. Check the attribute table
   Left click the button to open the table and check that it contains the variables described above.

3. Perform a classical (ordinary least squares) regression
   Methods > Regress. As with ArcGIS, you need to type in the file name and location. It is also necessary to type in a report title (if you want to run several regressions, use a naming convention that will remind you which iteration you have run). Hit okay. Add your Dependent variable, which denotes the presence of live coral cover (in this case we use the logarithm of live coral cover because it is necessary to) and your independent variables, which denote physical aspects of the environment (bathy, rugosity, variety and BPI). Click Run to execute the model.



4. When the model has run, click "view results" to examine the diagnostics associated with the regression analysis.

```
Regression Report

Regression
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set                :   Ground_ref
Dependent Variable      :      LOG_LC2  Number of Observations:  365
Mean dependent var      :     0.637331  Number of Variables   :    8
S.D. dependent var      :      1.27101  Degrees of Freedom    :  357



R-squared               :     0.670007  F-statistic           :      103.549
Adjusted R-squared      :     0.663537  Prob(F-statistic)     :            0
Sum squared residual:          194.578  Log likelihood        :     -403.109
Sigma-square            :     0.545037  Akaike info criterion :      822.217
S.E. of regression      :     0.738266  Schwarz criterion     :      853.416
Sigma-square ML         :     0.533091
S.E of regression ML:         0.730131


-----------------------------------------------------------------------
     Variable     Coefficient      Std.Error     t-Statistic   Probability
-----------------------------------------------------------------------
    CONSTANT        1.287918        1.714219      0.7513145     0.4529546
       DEPTH     0.007821502      0.01326378      0.5896889     0.5557700
      BAND_1   -0.0008144387      0.00767862     -0.1060658      0.9155639
      BAND_2    -0.001849191     0.008971383     -0.206121      0.8368294
      BAND_3     0.001708558     0.003983848      0.4288713     0.6682822
      BAND_4    -0.005203245     0.002018282     -2.578056      0.0103359
     VARIETY       0.3196013       0.1132603       2.821831     0.0050422
       SLOPE       0.1347068      0.01309244        10.2889     0.0000000
-----------------------------------------------------------------------


REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   316.956746
TEST ON NORMALITY OF ERRORS
TEST                      DF           VALUE           PROB
Jarque-Bera                2           261.4114        0.0000000
```

5. What is the R-squared value and its associated p-value?

R^2= p <

R^2 is a measure of the proportion of variation in live coral occurrence that is being explained by the combination of reflectance (bands 1–4), variety and slope. Is the relationship between coral reef occurrence and the physical variables statistically significant (probability < 0.05)?

____

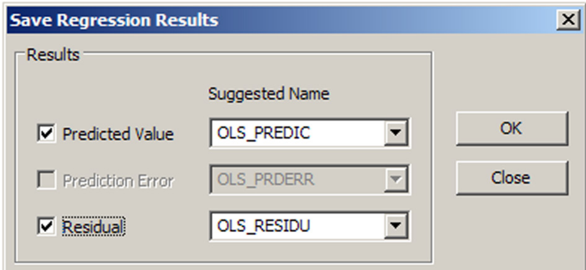
Note down the T-statistic associated with each of the independent variables (this is a measure of how powerful these are independently as predictors of coral reef occurrence),

____

____

____

6. Map the residuals of the model.
Save the predicted values and residuals of the regression to the Ground_ref shapefile table of attributes > save to table > tick "predicted value" and "residual" and click okay.



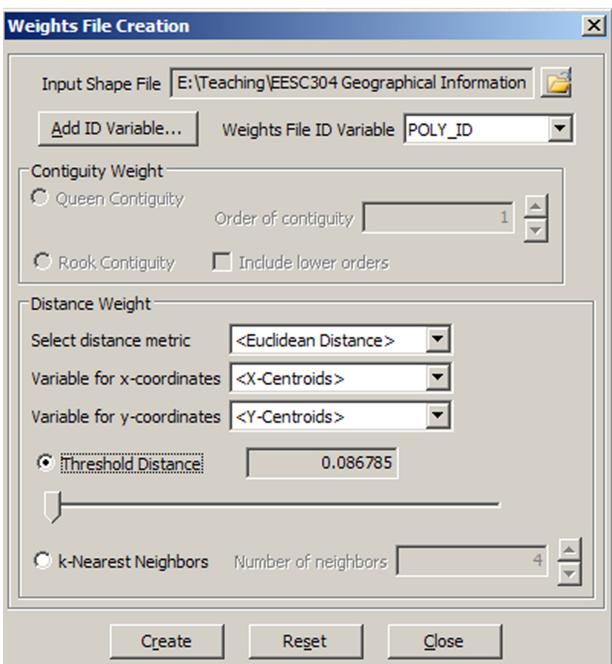
These should now appear as new columns in the shapefile table of attributes. With the map window open, select Map>Standard Deviation.
Does the map display any spatial pattern of the residuals? Based on the regression model used, explain what positive and negative residuals indicate:

___

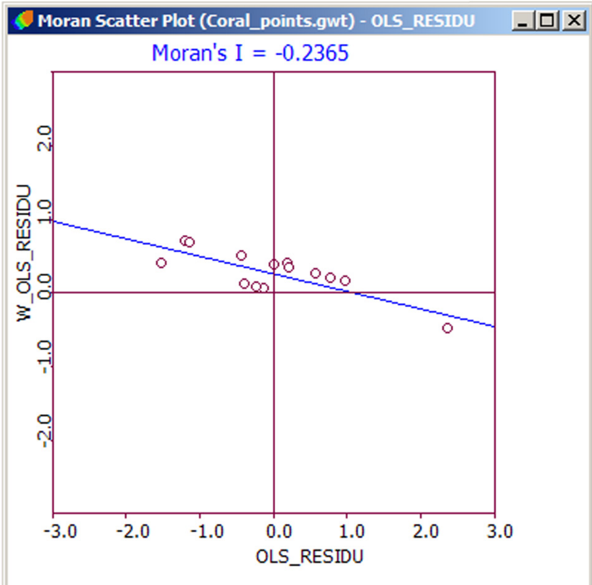
___

___

7. Quantify the spatial structure of the residuals, if any, using Moran's I
Create a spatial weight matrix Tools > weights > Create. Browse and select the input file (Ground_ref) from which the weight matrix is to be created. Select POLY_ID (the default) as the ID variable and highlight the Threshold Distance box for the Distance Weight. Hit Create and save the.gwt file.

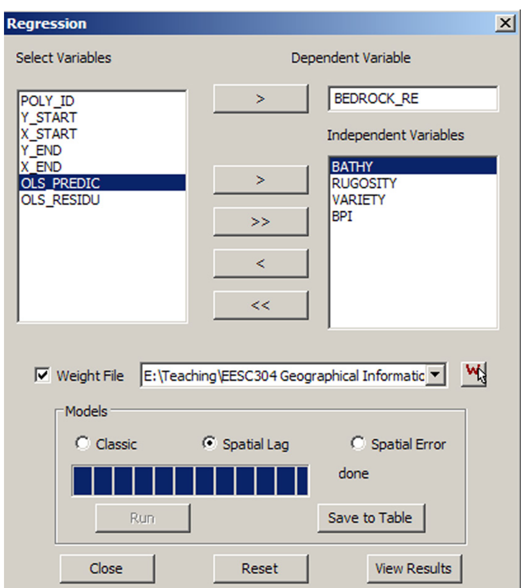


Create a Moran's *I* scatter plot of the residuals using the variable "OLS_RESIDU" that you have already created. This can be done by selecting Space > Univariate Moran's I and then selecting the variable "OLS_RESIDU".

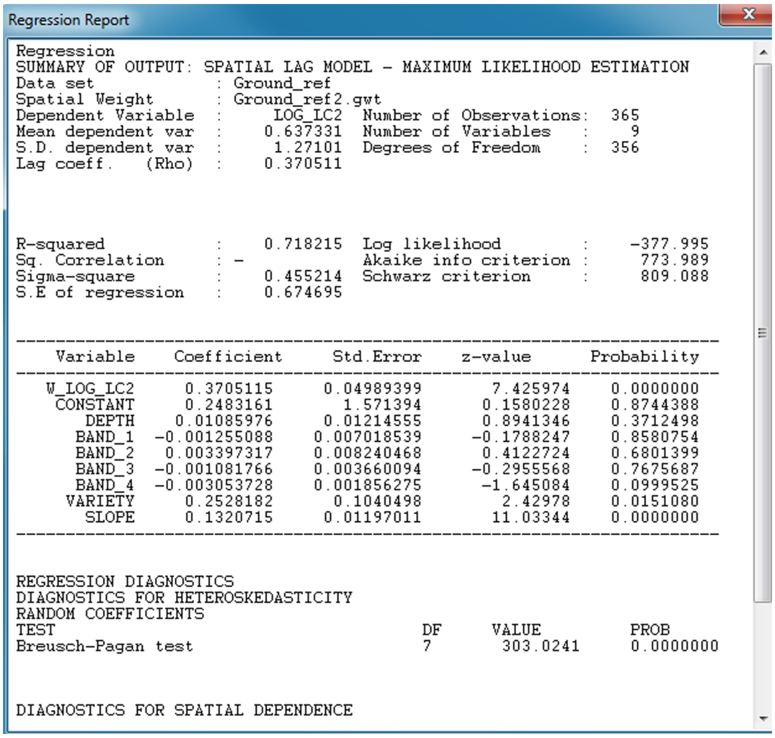


Does it suggest from the value of the Moran's I that the residuals are spatially autocorrelated?

8. Re-run the classical regression as a spatially lagged autoregression.
   Menu > Regress. Specify a name for the output file (e.g. Spatial_regression). Tick the predicted value and residual box. Run the model as before, but this time as a "Spatial lag" model with the weights matrix specified.



When done, click “Save to Table” and save the predicted values and residuals of the regression result to the attribute table. Then click “View Results”. The results of the regression analysis for the spatial lag model should appear.



```
Regression Report
Regression
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : Ground_ref
Spatial Weight      : Ground_ref2.gwt
Dependent Variable  :     LOG_LC2  Number of Observations:  365
Mean dependent var  :    0.637331  Number of Variables   :    9
S.D. dependent var  :     1.27101  Degrees of Freedom    :  356
Lag coeff.    (Rho) :    0.370511

R-squared           :    0.718215  Log likelihood        :     -377.995
Sq. Correlation     : -            Akaike info criterion :      773.989
Sigma-square        :    0.455214  Schwarz criterion     :      809.088
S.E of regression   :    0.674695

-----------------------------------------------------------------------
    Variable    Coefficient      Std.Error     z-value      Probability
-----------------------------------------------------------------------
   W_LOG_LC2      0.3705115     0.04989399      7.425974     0.0000000
    CONSTANT      0.2483161       1.571394     0.1580228     0.8744388
       DEPTH     0.01085976     0.01214555     0.8941346     0.3712498
      BAND_1  -0.001255088    0.007018539    -0.1788247     0.8580754
      BAND_2   0.003397317    0.008240468     0.4122724     0.6801399
      BAND_3  -0.001081766    0.003660094    -0.2955568     0.7675687
      BAND_4  -0.003053728    0.001856275     -1.645084     0.0999525
     VARIETY      0.2528182      0.1040498       2.42978     0.0151080
       SLOPE      0.1320715     0.01197011      11.03344     0.0000000
-----------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                       DF      VALUE         PROB
Breusch-Pagan test                         7        303.0241     0.0000000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
```

Does the spatial lag model perform any better for explaining the spatial distribution of the coral reef?

R^2 ________________

Re-run the Moran’s *I* scatter plot for the residuals of the spatially lagged regression (you will need to re-add the new residuals data column to the Table of attributes and specify the weights file for this plot).

What is the new value of the Moran’s *I* statistic? What does this mean?

____

____

____

Part 2: Predict the Presence of Live Coral Around the Lizard Island Group

Finally, it is possible to use the coefficients from the regression equations you have just generated to predict the presence of live coral around the Lizard Island Group (as a spatially continuous layer), given the physical information you already have (reflectance bands 1–4, variety, slope).

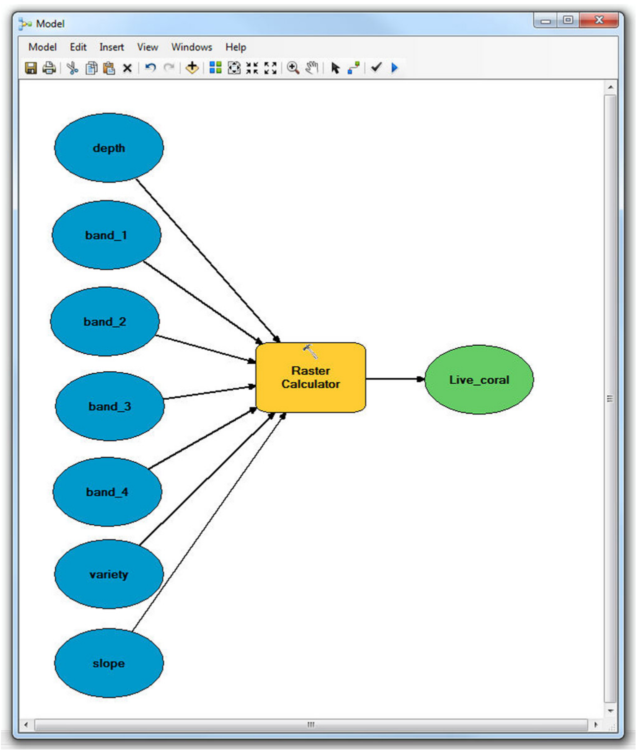
The form of the classic ordinary least squares regression is as follows:

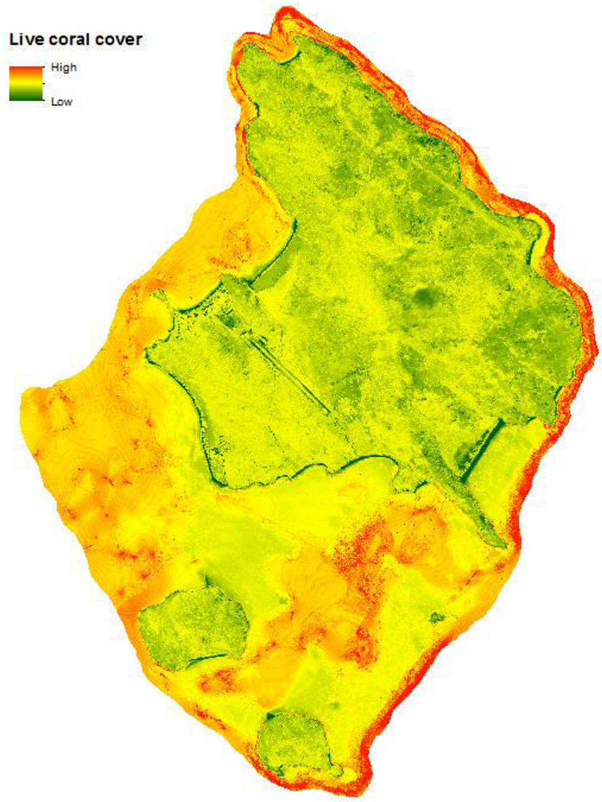
$$Y_{(i)} = \beta_0 + \beta_1 X_{1(i)} + \beta_2 X_{2(i)} + \beta_3 X_{3(i)} + \beta_4 X_{4(i)} + \beta_5 X_{5(i)} + \beta_6 X_{6(i)}$$

Where Y(i) is a predicted measure of the dependent variable for location i (in this case, the cover of live coral on the reef) and X1 through to X6 are the independent physical variables (reflectance bands 1–4, variety, slope), multiplied by their associated regression coefficients. It is necessary to insert the constant regression coefficient β_0 to complete the expression in the map calculator, which should read as follows on the basis of the results from the classic ordinary least squares model:

1.287918 + (0.007821 * “%depth%”) + (–0.000814 * “%band_1%”) + (–0.001849* “%band_2%”) + (–0.001709* “%band_3%”) + (–0.005203* “%band_4%”) + (0.319601* “%variety%”) + (0.134707* “%slope%”)

Build and run this model using the model builder in ArcGIS. The output should look like a feasible distribution for coral reef around Lizard Island. This is a simple predictive habitat model for coral reef. (NB It may be necessary to rescale the output values so that they fall between 0 and 1).




Live coral cover
High
Low

How Long is the Length of Britain's Coastline?

Background Information and Objectives

This exercise poses the question *"How long is the length of Britain's coastline?"* to demonstrate that the uncertainty associated with any measurement or estimation is a function of the spatial scale at which that measurement or estimation was generated.

Data

The imagery basemap supplied by ArcGIS can used as a back drop against which the coastline of Britain can be digitised across at a series of different geographical scales in order to measure, or estimate, its length.

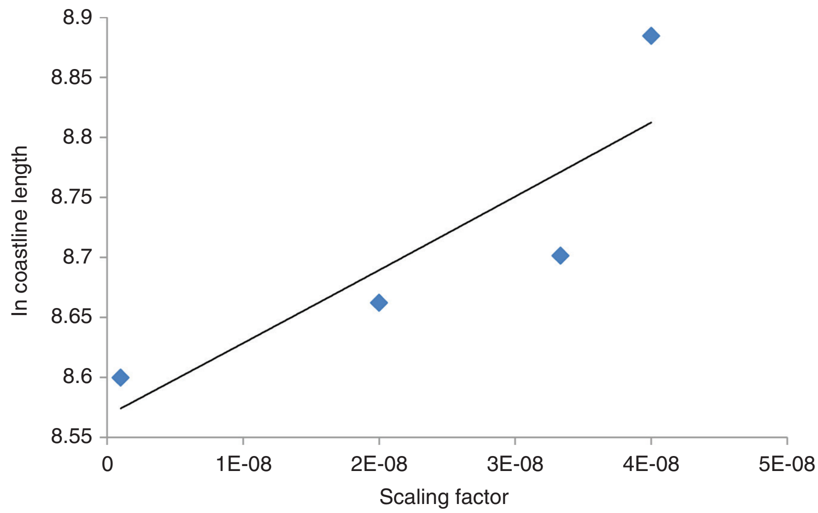
Instructions

Using the ESRI Imagery Basemap as a guide, create a polyline shapefile and edit this to digitise the length of the coastline of Britain at a given scale. Then generate a new field in the table of attributes for the file and use the calculate geometry function to estimate the length of the coastline in kilometres. Repeat this procedure to create several different shapefiles for which digitisation can be carried out while the screen is zoomed in on the image at a different scale. Compare your results to those presented here using the Table and plot provided.

A table is supplied below to illustrate estimates of the coastline length, alongside the scale at which the coastline was digitised in order to generate the overall length estimate.

Scale	Scale factor	Coastline length/ km	Ln Coastline length
1:25000000	0.00000004	7219	8.884472
1:30000000	3.33333E-08	6011	8.701346
1:50000000	0.00000002	5779	8.661986
1:100000000	0.000000001	5430	8.599694

The natural logs of the perimeter length estimations for the UK coastline are plotted against the scaling factor (area on the map divided by the areas on the ground) for a series of estimation values derived across a range of geographical scales to explore the influence of scale on measurement.



This question can be answered by digitising the coastline as a vector dataset composed of many lines. The overall length can then be calculated by adding together the lengths of all the individual lines. If several digitisations are carried out across a range of scales, it becomes apparent that the greater the scale at which the coastline is viewed during the digitization procedure (i.e. the more 'zoomed in' the analyst is to the coastline), the longer the overall length that is estimated. This is because the more you zoom in on the coastline, the more detail is apparent. Thus, a line digitised at a larger scale more closely follows the intricate meandering shape of the coastline. As a consequence, the estimated length increased as a function of scale.

Figure 8.2 illustrates the lines digitised to estimate the length of the mainland UK coastline at a scale of 1:1,000,000 (the largest scale), 1:10,000,000 and 1:30,000,000 (the smallest scale). It can be seen that a greater degree of error is associated with the line digitised at the smallest scale. This is because it is not possible to see the detailed meanderings of the coastline. Indeed, the lines digitised at this smaller scale often miss the coastline altogether. Such an error can be quantified through repeat digitisations of the coastline and statistical analysis of the distribution spread associated with the resulting coordinates belonging to the vertices of the digitised lines (see Chapter 8 of textbook, section 8.5.1). To get a more accurate measure, it is necessary to increase the scale. The estimate of length will continue to increase as the scale is increased. Even if the length is measured of every boulder, rock, pebble or grain of sand, this estimate will continue to increase as the scale or precision increases, approaching infinity. This phenomenon makes it very difficult to provide a definitive answer to the question of how long the coast of Britain is!

Evaluating Uncertainty in the Vulnerability of Manhattan, New York to a Tsunami

Background Information and Objectives

This case study assesses the land area that would be inundated in the event of a 6m height tsunami wave hitting the borough of Manhattan in New York City. The exercise illustrates how the uncertainty associated with a dataset can translate into a substantially different result for any analysis that is carried out on this dataset. This can have considerable implications for the practical objectives of the analysis. This inundation assessment is a common spatial analysis exercise used for coastal vulnerability assessment.

Due to the relevance of the results for infrastructure planning, it is important to explicitly report the main limitations and uncertainty of the analysis. To do this, different scenarios will be presented to explore the uncertainty inherent in the data. This exercise consists of two steps:

1. Calculating the 95% confidence interval associated with a DEM of the Manhattan area, and
2. Mapping the areas of Manhattan inundated by a 6 m tsunami, accounting for the vertical accuracy of the DEM dataset

Data

Two datasets are used for this case study: a two metre resolution digital elevation model (DEM) that indicates the height above sea level for each pixel (filename: Manhattan_2mDEM) and an independent set of 100 point elevations extracted from a LiDAR survey of Manhattan (filename: NYC_spot_heights).

Data Source: Both datasets were sourced from the Department of Environmental Protection and accessed through the NYC Open Data portal.

Calculating the 95% Confidence Interval of the Digital Elevation Model

In the first step, values of the LiDAR point elevations are compared to the DEM in order to calculate the root mean square error and associated 95% confidence intervals for the DEM. The 95% confidence interval allows the analyst to state a range of values within which they are 95% confident that their data point lies. This is a conventional way of expressing the uncertainty associated with DEMs. Vertical accuracy of the DEM is essential for this assessment. Vertical accuracy is an expression of the overall quality of the elevations contained in the DEM in comparison to "true" (i.e. more accurate) ground elevations. Accuracy standards and guidelines exist in general for geospatial data and specifically for elevation data. A common way to report the vertical accuracy of a DEM is by using the root mean square error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(z_i - \tilde{z}_i\right)^2}{n}}$$

Where n = number of data points

z_i = estimated value at location i (extracted from the DEM)
$\tilde{z}$ = true or expected value at location i (provided from the LiDAR survey)

For a standard normal distribution (or Gaussian distribution), the central limit theorem states that 95% of the area under the curve lies within plus or minus 1.96 standard deviations of the mean (Figure 1). For a given dataset, it is therefore possible to calculate the mean, standard deviation and a 95% confidence interval for any value within the dataset. A corresponding vertical linear error (LE) with confidence levels can be established by multiplying the RMSE by 1.96 (see spreadsheet provided).

To calculate the DEM's vertical accuracy based on the Lidar point dataset using the root mean square error (RMSE), it is necessary to extract values from the 2m DEM that correspond to the locations for which independent Lidar height points are known. Both sets of elevations can then be exported to a spreadsheet to perform the required calculations. Calculating the DEM root mean square error and corresponding vertical accuracy proceeds as follows:

- *Load both the DEM dataset (Manhattan_2mDEM) and the spot heights dataset (NYC_spot_heights) into ArcGIS*
- *Make sure the 3D Analyst extension is activated (Customize – Extensions).*

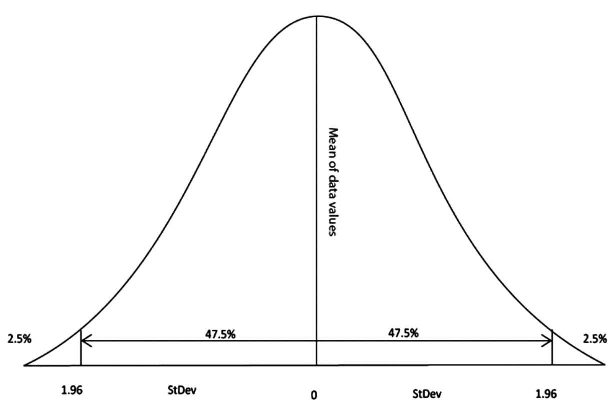


Figure 1 A standard normal distribution, or Gaussian curve. The total area under the curve is 100% and the central limit theorem states that 95% of this area lies within plus or minus 1.96 standard deviations of the mean.

- *Open the Surface Information tool from 3D Analyst Tools – Functional Surface – Add Surface information*
- *Use NYC_spot_heights as your input Feature Class and Manhattan_2mDEM as the Input Surface. Place a tick in the z box in output property. Press OK and run the model.*
- *Open the NYC_spot_heights attribute table. You should have a new field called 'z' with the 2m DEM values.*
- *Export the table to dbf format. From the attribute table, select Table Option/ Export.*
- *Open the dbf table using Excel. You can now calculate the RMSE in Excel. check the Excel file 'RMSE.xlsx'.*

What is the confidence interval for 6m if the linear error at 95% confidence level is 7.91 m?

A. 6m [m, m]

Mapping the Areas of Manhattan Inundated by a 6m Tsunami, Accounting for the Vertical Accuracy of the DEM Dataset

In the second step, the land area of Manhattan that falls below an elevation of 6m is calculated. Following application of the Bruun Rule, this corresponds to the area that would be inundated by a 6m tsunami (Bruun, 1988, Bruun, 1983). Note that the Bruun rule has been widely criticised as an over simplification of complex inundation dynamics, but is still widely used as a guide for coastal vulnerability modeling. The area mapped must therefore be adjusted to incorporate the uncertainty associated with the DEM into the analysis. This is achieved by adding the vertical error corresponding to the upper 95% confidence limit to the height of the tsunami and re-evaluating the area inundated by an adjusted, larger tsunami height. Finally, a comparison of the land areas inundated in the first and second models runs is undertaken to illustrate the implications of uncertainty associated with DEM for the overall results of the coastal vulnerability analysis (see Figure 1).

Calculating the Area of Manhattan Inundated by a 6m Tsunami

- Use the conditional tool (*Spatial Analyst Tools – Conditional – Con*). Double click 'Con' to convert elevations less than 6 m to a value of **1** (this is the conventional value assigned to raster pixels if the condition is true). Select 'dem_2m' in Input conditional raster. Create the following SQL condition '**Value < 6'.** Type 1 for the Input true raster or constant value. Leave blank for the Input false raster or constant value (optional)**.**
- Save the *output raster* as 'tsunami_6m' and hit okay to run the model

- You have defined areas prone to tsunami flooding as a new raster dataset.
- To calculate the area (sq.km) that is prone to flooding, it may be easier to work in vector data format so that the associated area geometry of the polygon shapefiles can be calculated. Convert the raster dataset "tsunami_6m' to vector using the *Raster to Polygon* tool (*Conversion tools – From Raster – Raster to Polygon*)
- Add the vector layer representing 'tsunami_6m' and open the attributes table.
  - Add a new field (float) and name it 'flooded'.
  - Right-click on the field and select *Calculate Geometry*…
  - Calculate *Area* in *Square Kilometers [sq km].* To get the total area covered by the vector dataset, right click on the header of the column that contains the area values, hit statistics and read off the 'sum' value. (Note: this area could also be calculated by multiplying the number of raster cells coded one by the area of a raster pixel, i.e. 4sqm, there may be a small offset between the asnwers reached from these different approaches because of the differing precisions of the raster and vector datasets)
- Repeat the process to include the maximum interval defined by the 95% confidence interval (you will need to adjust the height of the tsunami to 13.91 m for this step, which corresponds to adding the additional 7.91m height onto the initial 6m tsunami)

Solution

Comparison of the LiDAR point elevations with the DEM yielded a root mean square error of approximately 4.04 m for the DEM. The central limit theorem could be used to calculate the associated 95% confidence interval limit as 1.96 * RMSE under the assumption that the error was random (i.e. equally likely to an over or underestimation of the actual elevation) and normally distributed. This yielded a vertical linear error of 7.91 m (see spreadsheet provided). Thus, the analyst can make the statement for the elevation value associated with every pixel in the DEM that they can be 95% confident that the accepted elevation in real life of the ground area corresponding to that particular pixel falls within the range of the value specified by the DEM grid, plus or minus 7.91 m.

In the second step of the analysis, the area of the DEM that falls beneath an elevation of 6 m was calculated to be 25.36 sq.km. In the third step of the analysis, the uncertainty associated with the DEM was incorporated into the assessment of the area inundated. For this coastal vulnerability exercise, which involves projecting the height of a hypothetical tsunami onto a DEM, this uncertainty range can be transferred to the height of the tsunami because vulnerability is assessed as a function of the vertical relationship between the tsunami height and the DEM height. The confidence limit of 7.91 m was therefore added to the tsunami, to project a tsunami of 13.91 m in height. This step was taken because it is possible, within the 95% confidence limit that the DEM could actually lie 7.91 m lower than indicated by the values quoted. This would have the same effect on the area inundated as raising the height of the tsunami. The area of the DEM falling beneath an elevation of 13.91 m was found

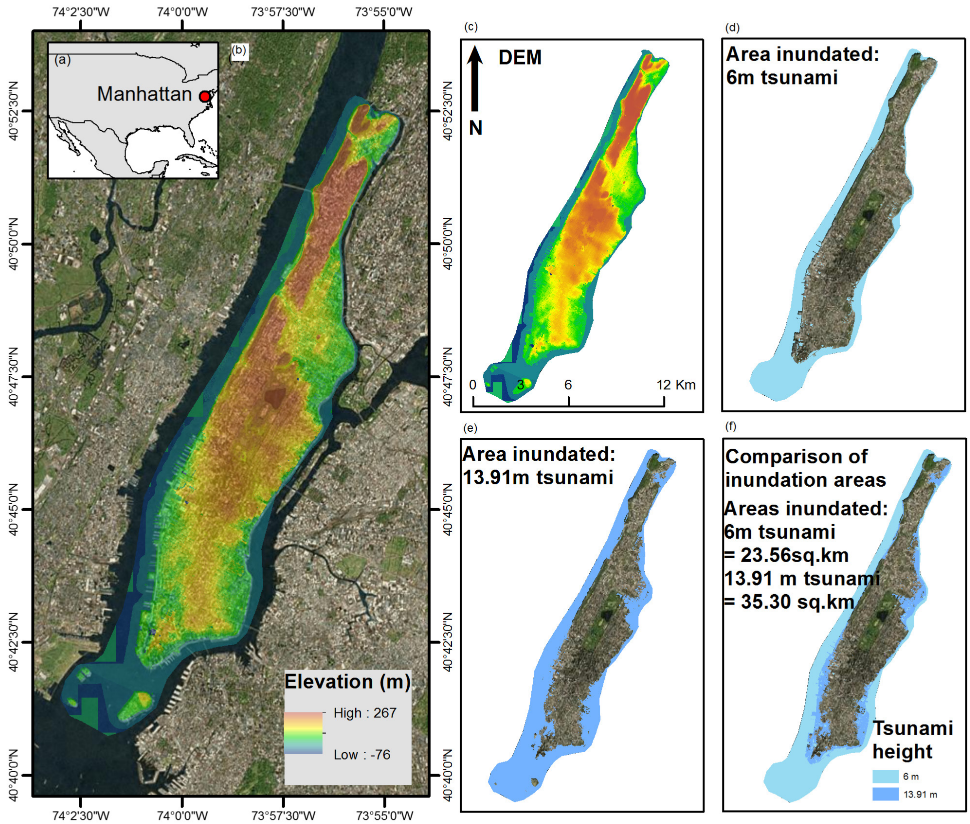


Figure 2 Uncertainty associated with an assessment of the vulnerability of Manhattan, New York to a 6 m tsunami (a) Location of Manhattan on the North American continent, (b) Digital elevation model (DEM) overlaid onto a satellite image of Manhattan (c) DEM of Manhattan, (d) Area inundated by a 6 m tsunami indicated in light blue, (e) Area inundated by a 13.91 m tsunami indicated in dark blue (the upper 95% confidence limit) (f) Area inundated by a 6m tsunami (light blue) overlaid onto the are inundated by a 13.91 m tsunami for comparison of assessment outputs, allowing for uncertainty associated with the DEM.

to be 35.30 sq.km. When the uncertainty of the DEM dataset is accounted for, this therefore equated to an additional land area of 9.94 sq.km (Figure 2).

In practical terms, this difference in area related to some of the world's highest density, most expensive land. At real estate prices of $1363 per square foot (http://www.dnainfo.com/new-york/20140401/upper-east-side/manhattans-real-estate-prices-reach-record-1363-foot), the difference in value of the land inundated equates to US $1400 Billion. This figure assumed single storey properties, which is clearly not the case in the high-rise suburb of Manhattan. Accounting for sky scrapers would likely make this number much higher. It would therefore be of interest to an insurance company calculating the vulnerability of Manhattan properties to incorporate uncertainty into their calculations!