Creating a cost surface to estimate pedestrian journey time to health facilities

The purpose of this exercise is to gain familiarity with some of the conceptual issues and GIS techniques involved in creating a cost surface that can be used to estimate geographical access to health care. The exercise is split into two sections, with each section using data from neighbouring regions in Mozambique. Some elements of the data are based on real-world data sets, whilst others are fictional and have been created for this purposes of this exercise. The first section can be thought of as a 'dry run', with fairly detailed instructions taking you through some of the basic GIS steps in creating a cost surface. The second section asks you to create a more complex cost surface for a different region.

About cost surfaces

Cost surfaces are used in GIS to estimate the 'cost' of a journey between two points across a landscape represented by a raster grid. This journey 'cost' can be measured in any units the user wishes. In this exercise we will be quantifying journey cost in terms of time, but other measures such as financial expense, or number of calories burnt could be used in a similar way. Cost surface functions in GIS generally require two basic pieces of information. Firstly, the location of one or more origins and destinations is required. In this exercise, the starting points of the journeys we are interested in are a series of villages and the destination is a health facility. Secondly, information is needed on the journey cost, or *impedance*, associated with travelling over different parts of the landscape. Since we are using a raster grid to represent the landscape, each grid cell can potentially take on a unique value of cost depending on a range of factors. These factors might include, for example, whether the grid cell represents road or non-road, whether the cell represents a barrier to travel such as a river, or whether the cell is flat or steeply sloping.

If a journey between two points requires passage through a known series of grid cells, and the time required to cross each of those cells is also known, the total journey time can be calculated by simply summing the impedance values of all the cells along the route. This is effectively the approach used by GIS cost surface functions, but there is one complication: in general, the exact route people use to travel between two points is not known in advance. The solution offered by cost surface functions is to work out the 'least-cost' route between every cell and the destination cell, and assume this is route people will use: a reasonable assumption.

Tip: A clear technical account of the *Cost* Distance function in ArcGIS is provided in the ArcGIS help. This can be accessed by clicking on Help > ArcGIS Desktop Help, selecting the Index tab and typing 'Cost Distance', and selecting the option located in the Spatial Analyst Toolbox. The link entitled 'Learn more about how Cost functions work' is particularly useful.

Section 1: Creating a basic cost surface

1.1 Data

Folder 'Section1data' contains the GIS layers required for the first part of this exercise. These layers are:

Two shapefiles:

- **Villages**: The location of two villages
- **Facility**: The location of a health facility

Three raster grids:

- **Impedance1**: A uniform raster grid containing impedance values – more on this shortly!
- **Roads**: Grid showing location of the road network
- **Lake**: Grid showing location of a lake

The three raster grids are aligned with each other and have a cell size of 0.000834 decimal degrees (approx 91m).

It is particularly important to understand how the impedance1 grid is set up. Each cell contains an impedance value. In this particular case, every cell contains the same value, so we are saying that the landscape is entirely uniform, with speed of travel equal in all grid cells. When this impedance grid is input into the Cost Distance function in ArcGIS, a calculation is performed to work out the cost of crossing each cell. The required input impedance values are <u>not</u> the cost (time) required to cross each cell, but the cost <u>per unit distance</u>. The data for this exercise are georeferenced in decimal degrees (DDs) so the appropriate impedance value for each cell is the time taken to walk one DD. The Cost Distance function can then use this value, along with knowledge of cell size, to work out the cost of crossing each cell.

The impedance value of each cell of the *impedance1* grid is 1320 minutes. Let's take a moment to understand how this value was calculated. The underlying assumption is that people walk at a steady speed of 5km/h. Remember we need to state the time taken in minutes to walk one DD. The actual size of a DD varies greatly according to where you are on the Earth's surface, but we will assume in this exercise that one DD is equal to 110km. Walking 110km at 5km/h takes 22 hours, or 1320 minutes, so this is our 'cost' per unit distance.

<u>Tip</u>: An alternative means of figuring out impedance values on this grid would be to reproject all the map layers to a coordinate system in metres. We could do this by selecting *data management tools / projections and transformations*, then using *project* (for vector layers) or *raster / project raster* (for raster layers). Under *output coordinate system,* a projection such as *projected coordinate systems / UTM / WGS 1984 / Southern Hemisphere / WGS1984 UTM Zone 36S* would be an appropriate choice for Mozambique. We would then work out impedance values per <u>metre and not per decimal degree</u>, since these would be the distance units. For UTM coordinates, if we assumed a walking speed of 5km/hour, this is equivalent to 83.33metres/minute (calculated by dividing 5000 metres by 60 minutes), so that would mean it would take 1/83.33 = 0.012 minutes to walk 1 metre.

1.2 Creating a simple 'cost surface'

We will now create a simple cost surface using the uniform impedance grid. This will give us an estimate of the pedestrian journey time between the health facility and the two villages.

- Open the Cost Distance dialogue box by selecting: Spatial Analyst Tools > Distance > Cost Distance in ArcToolbox.

- Firstly, we need to check some of the generic raster analysis settings are configured correctly. Go to the 'Environments…' tab at bottom of the Cost Distance dialogue box. Click on 'General settings' and ensure that both the 'Output Coordinate System' and 'Processing Extent' are set as 'same as Layer "impedance1" '. Now click on 'Raster analysis settings' and ensure that 'Cell size' also reads 'same as Layer "impedance1" '. Click on OK to return to the Cost Distance dialogue.

- Now enter the 'Input raster or feature data source' as the *Facility* shapefile. This will tell the function where the journey destination is that we are interested in.

- The 'Input cost raster' is the *impedance1* grid, telling the function the cost (per unit distance) associated with crossing each cell.

- Call the 'Output distance raster' *Costsurface1*. The other two boxes can be left blank.

Take a moment to examine the output, *Costsurface1*. Do the values 'make sense'?

1.3 Creating more realistic cost surfaces

The impedance grid used in the above exercise was uniform, implicitly assuming that people can walk at the same speed regardless of where they are in the landscape. We will now make some more sophisticated impedance grids to use with the Cost Distance function that incorporate two different features of the landscape, a lake and the road network, that will affect people's walking speed and the route they take between villages and the health facility.

We know that people will not be able to walk across the lake, and so will have to walk around it. We can modify our impedance grid to reflect this by masking out the lake grid cells so the Cost Distance function will ignore them.

- Firstly, we need to reclassify the *Lake* grid to set all 'lake' cells to a value of 'NoData' which means they will be masked out. Do this using the Reclassify tool in ArcToolbox (Spatial Analyst Tools > Reclass > Reclassify) and call your output *impLake*.

- Next, we want to multiply the impedance grid *impedance1* by the values of *impLake* in order to create a new impedance grid with the lake cells masked out. So this using the Times tool in ArcToolbox (Spatial Analyst Tools > Math > Times) and call your output *impedance2*.

- Now rerun the Cost Distance function as before, but this time using the new *impedance2* impedance grid, and naming your output *Costsurface2*.

Examine your new output cost surface, *Costsurface2*, and compare it to *Costsurface1*. Try and interpret the effect that the incorporation of the lake has made.

- Next, we will incorporate the road network. It is a reasonable assumption that people will choose to walk along roads, tracks, or footpaths, rather than across the landscape in an uninterrupted straight line between origin and destination. We will also assume that our original walking speed value of 5km/h applies only to the road, whilst the rest of the landscape allows walking at only half that speed, 10km/h. We can implement these assumptions in the following steps:

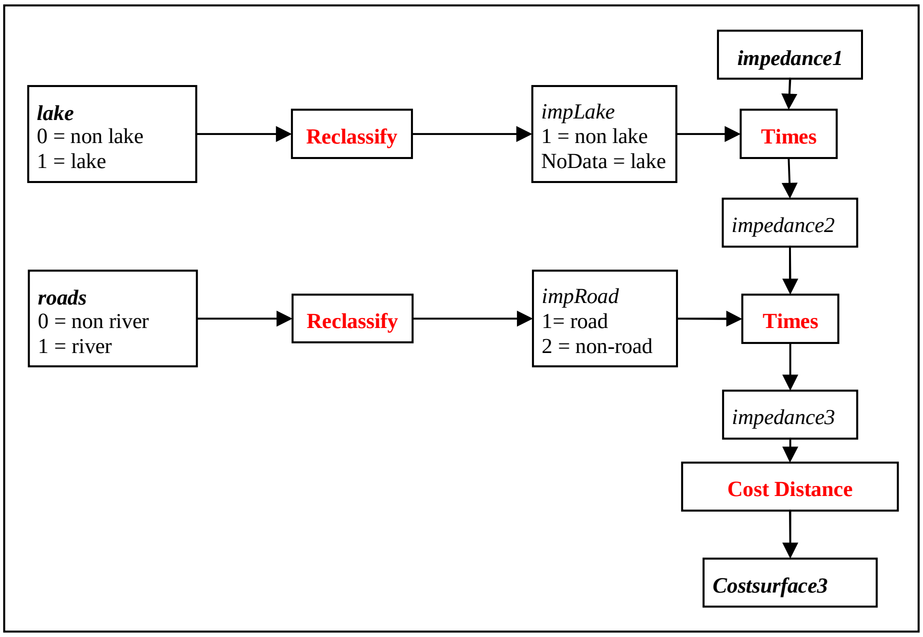
- Firstly, reclassify cells in the *Road* grid such that cells of road have a value of one, whilst cells of nonroad have a value of two. Call your new output *impRoad*.

- Now multiply the impedance grid *impedance2* by *impRoad* and call your output *impedance3*. The impedance values of non-road have now doubled, whilst road cells have remained the same. Cells masked out by the lake remain masked.

Finally, rerun the cost distance as before but using the *impedance3* impedance grid. Think about how it compares to the previous cost surfaces and what effect the incorporation of the road network has made.

1.4 Flowcharts to summarise analysis steps

The above procedures are fairly laborious to describe in words. A more effective way of representing the various stages is with a flowchart like the one below, showing the input data layers, each analysis step, intermediate data layers, and the final output.



Section 2: Creating a complex cost surface

Having outlined some of the basic stages in the production of a simple cost surface in the above 'dry-run', your task is to now create your own cost surface that incorporates more complex data about the landscape. You have been given a second set of data for a different area just to the South of that considered in Section 1.

2.1. Data

Folder 'Section2data' contains the GIS layers required for this exercise. These layers are:

Two shapefiles:

- **villages**: The locations of villages in the area. You will find fields with the total male and female populations in each village, as well as a field called **outpatient** that gives the total number of annual outpatient visits to the health facility during the course of that year.
- **HealthFacility**: The location of a health facility

Five raster grids:

- **impedance**: A uniform raster grid containing impedance values (1320 minutes per DD).
- **roads**: Grid showing location of the road network
- **lake**: Grid showing location of a lake
- **rivers**: Grid showing location of rivers
- **slope_degrees**: Gradient of terrain (in degrees), derived from a digital elevation model.

As before, the raster grids are aligned with each other and have a cell size of 0.000834 decimal degrees (approx 91m).

2.2 Activity

Your task is to create, for this new area, a cost surface to estimate the journey time from each village to the health facility, taking into account all the landscape features for which you have data. You will have to make decisions about how certain features affect walking speeds and routes. You will also have to think about how the different landscape features might interact when they occur at the same grid cell. You should also assess the extent to which geographic access to healthcare affects utilisation of the health facility.

You should produce a report of no more than 1500 words (excluding references and appendices) that includes:

- An account of how you produced your cost surface. Where you have made assumptions, these should be stated and justified. You should also produce a flowchart similar to the one presented earlier that summarises all the analysis steps that contributed to your final output.
- Final results. These may include appropriate maps and tables.
- Reflection on which features had the most substantial effect on the final results.
- Reflection on any weaknesses or limitations you perceive in the approach (these might be methodological, related to the data, conceptual, etc), and the implications of these limitations for the interpretation of the results.