Modelling Land Use / Cover Change using Markov Chains

Overview:

This exercise uses Markov chains to model land use / cover change in southeast England. The objectives of the exercise are:

- To provide an introduction to simpler methods for modelling land use / cover change
- To provide practical experience in handling land use / cover products

As part of the exercise, we will look at the Corinne land use / cover product, which is used throughout Europe. We will develop a simple Markov chain model of land use / cover change for just one broad land use / cover class – 'agricultural areas'. However, the practical could be extended so that land use / cover was modelled not just in this one class, but for many classes. This model will be based on transition probabilities for 2000 to 2006 and simulate land use / cover in 2012.

Data:

As the Corinne data sets are very large files in their raw form, they have been downloaded and pre-processed for this exercise. There are three files in total:

- **UK2000**: Land use / cover in raster format for the year 2000 for southeast England, stored in a file geodatabase
- **UK2006**: Land use / cover in raster format for the year 2006 for southeast England, stored in a file geodatabase.
- **Clc_legend.csv**: An explanation of the meaning of the codes that are stored in each raster grid cell in both the 2000 and 2006 map layers.

Data were downloaded from here:

http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-2

http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2000-raster-2

There are several different variants of the Corinne data available, e.g. with differing spatial resolutions and in both raster and vector format. In this exercise, we have used Version 16 of the raster product with a spatial resolution of 250 metres.

After downloading and unzipping the data, the '*clip raster*' tool was used to select out just the area in the southeast of England. The projection was defined using the .prj file provided on this web site via the *define projection* tool. If you have a fast broadband connection, you may wish to try out these initial steps to prepare the data yourself.

Practical Instructions

Familiarise yourself with the data

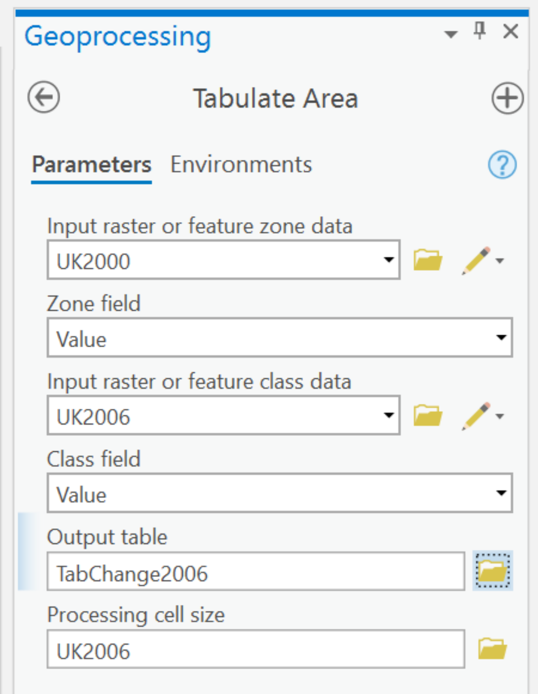
First, take a look at clc_legend.csv by opening up this file from within Excel (note: to view this file within Excel, head for *file*¸then *open* and select *files of type* **text files**). You should see that Corinne is a hierarchical classification. Various GRID_CODES are stored in each raster grid. These grid codes represent the most detailed land use / cover classes, as shown in the LABEL3 column. These detailed classes are grouped into intermediate classes (shown in the LABEL2 column), which in turn are grouped into broad classes (shown in the LABEL 1 column). Thus, the detailed class 'continuous urban fabric' is represented by a code of 1 in the raster map layer and forms part of the intermediate class 'urban fabric', which in turn is part of the broad class 'artificial surfaces'.

Now open up a new map display in ArcGIS Pro and take a look at the two raster grids that are provided for this exercise. Note how the codes that you have just viewed relate to the raster grids in ArcGIS Pro. Note in particular any areas where land use/cover has changed between 2000 and 2006 that you can identify by visually inspecting these map layers.

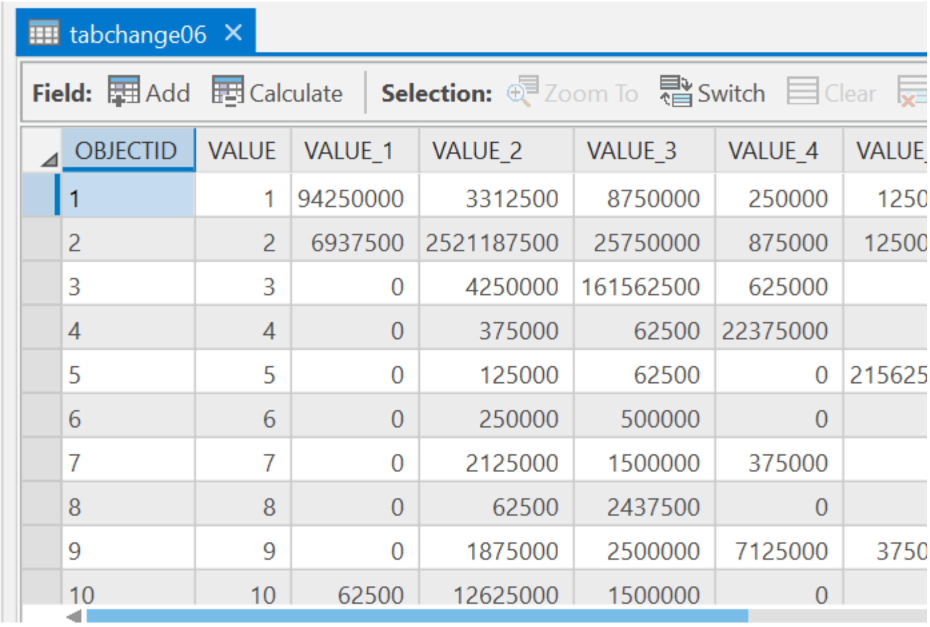
Generate a transition probability matrix

Having found our way around the data, we now need to generate a transition probability matrix, as the basis for our Markov chain simulation. The ArcGIS tool 'tabulate areas' can be used to generate such a matrix. As this is a spatial analyst tool, you will need to activate this extension (via *project* menu / *licencing / configure your licencing options*).

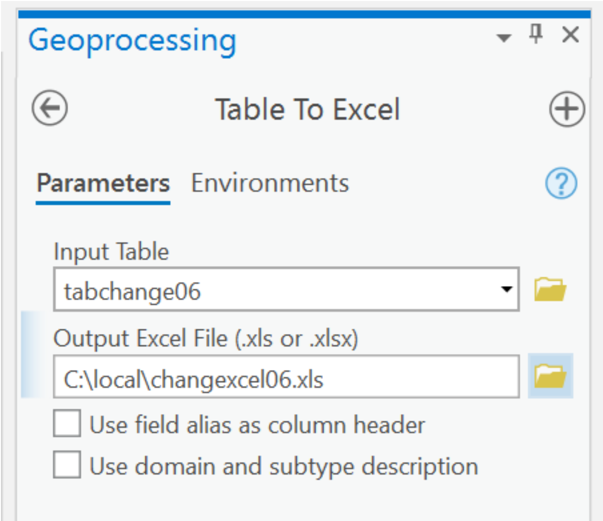
Head for the *analysis* menu, then use *tools* to bring up the geoprocessing panel. Search for and then run then 'tabulate areas'. *'Tabulate areas'* will generate an output table that has land use/cover categories in 2000 as rows and land use / cover categories in 2006 as columns. Each cell in this output table will contain the area (in metres sq) that was occupied by a given combination of land use / cover classes in 2006 and in 2000. To generate this table, use your land use / cover grids for 2000 and 2006 as inputs as shown below:



ArcGIS Pro should generate a table by way of output. Right-click on this table when it appears in the left-hand table of contents window, and then click on *open* to view the results of your calculation. You should see the land use / cover codes that we saw earlier for the year 2000 featuring in the first column. Each row of the table represents a different land use / cover class from the year 2000. In the columns are the land use / cover classes from the year 2006. To make our transition probability calculations easier, we need to export this table to Excel so that we can make use of Excel formulas.

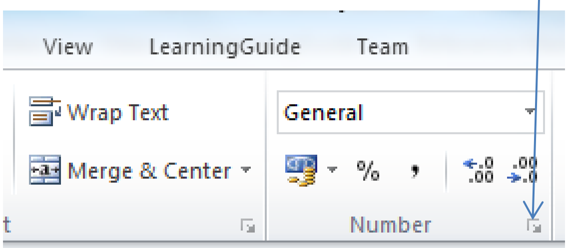


OBJECTID	VALUE	VALUE_1	VALUE_2	VALUE_3	VALUE_4	VALUE_
1	1	94250000	3312500	8750000	250000	1250
2	2	6937500	2521187500	25750000	875000	12500
3	3	0	4250000	161562500	625000	
4	4	0	375000	62500	22375000	
5	5	0	125000	62500	0	215625
6	6	0	250000	500000	0	
7	7	0	2125000	1500000	375000	
8	8	0	62500	2437500	0	
9	9	0	1875000	2500000	7125000	3750
10	10	62500	12625000	1500000	0	



Once you open up your output, you may wish to make a few changes to make the output more readable in Excel:

- Click on cell b2. Click on the View menu then select freeze panes, then freeze panes again. This will mean that the row and column headers are always visible, no matter how far you scroll down or to the right.
- The areas (in metres squared) are currently shown with a great many decimal places. You may wish to highlight the area of your spreadsheet that contains these area figures. Next, click on the *home* menu, then click here to change the *category* to **number** and *decimal places* to be **0**:



This should make your spreadsheet more readable. At this point, you may wish to select *save* as from the *file* menu and save a copy of your work in Excel (rather than dbase) format.

At the moment, the matrix we have generated is based on the most detailed land use / cover classes. To make life simpler, we will work with the broadest land use / cover classes. Take another look at clc_legend.csv and you will see that our row / column headings in our current spreadsheet relate to LEGEND3. What we need to do is to total up the figures for these detailed land use / cover classes using the information in LEGEND1.

Task 1:

1. By using the sum function in excel (i.e. typing in =sum(*range*) where *range* is a reference to a set of cells in your spreadsheet such as B3:D4), see if you can calculate the area in sq m of agricultural areas that changed to the broad classes 'artificial surfaces', remained as 'agricultural areas', changed

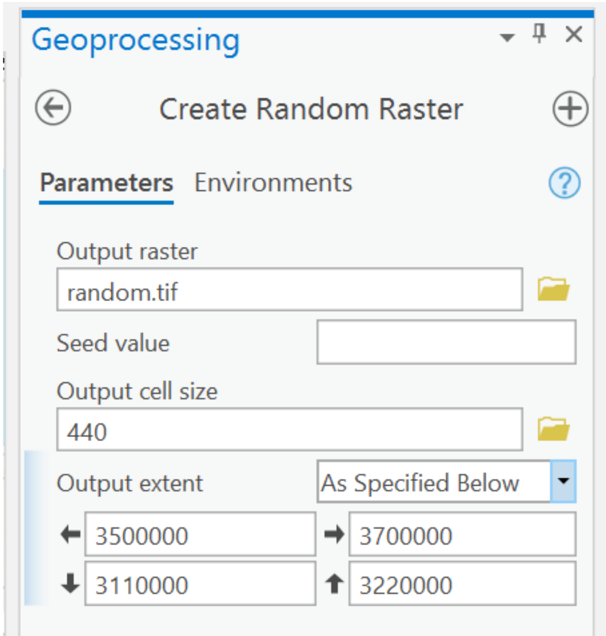
to 'forest and semi-natural area', to 'wetland' or to 'water bodies'. You will need to use both the spreadsheet you have just generated and **clc_legend.csv** to do this.

Having calculated these figures, work out what proportion of agricultural area in 2000 lay in each of these classes in 2006.

If you are not familiar with functions in Excel, you may wish to refer to '**ModelAnswer_probabilitytransition.xlsx**' in the .zip for this exercise if you are unsure or want to check your results.

Generate random numbers for your simulation

Now, we will generate a set of random numbers between 0 and 1 to help with our land use change simulation – such a simulation that makes use of random numbers is sometimes known as a Monte Carlo simulation. To do this, head for the *Analysis* menu again, choose *tools* and search for 'create random', then run *Create random raster.* Select an appropriate name for your output raster, e.g. **random** and set the *output extent* to be the same as one of your existing map layers, e.g. **UK2006**. The *output cell size* should automatically then be set to 440 (metres), so that it matches your other existing map layers also:



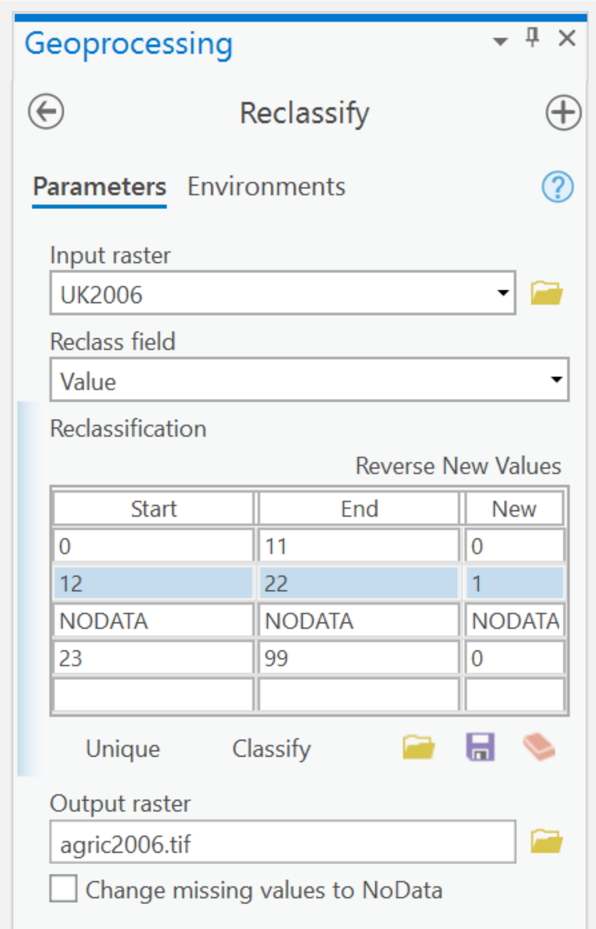
ArcGIS Pro should now generate a new raster for your study area, filled with random numbers between 0 and 1. We will use this in conjunction with our transition probability matrix to simulate land use/cover change.

Apply the transition probabilities to your random numbers

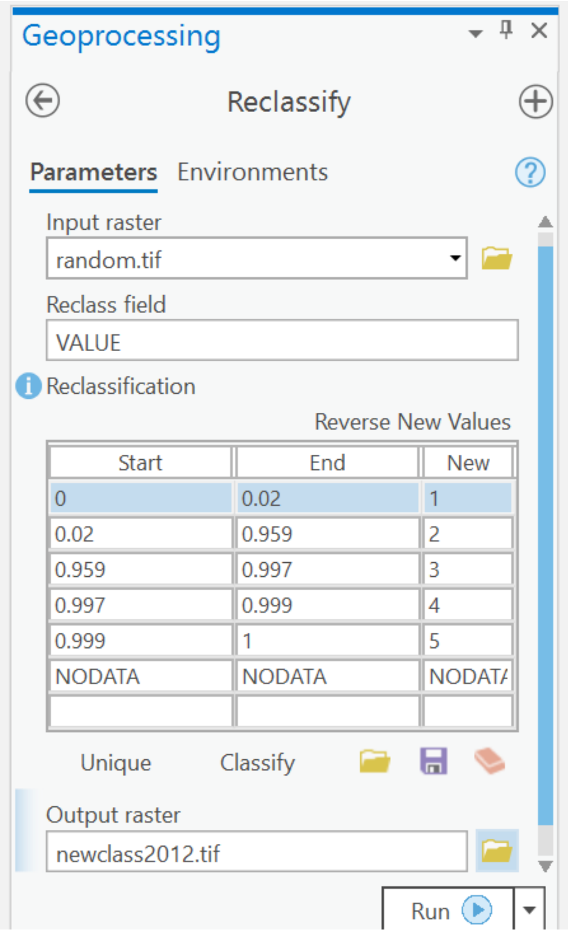
We have only calculated transition probabilities for the 'agricultural areas' class, so we need to pick out just that land use / cover class from our 2006 map layer as part of our calculations before proceeding further.

To do this, head for the geoprocessing panel again and search for 'reclassify', then run the *reclassify* tool. You will see that reclassify lets you enter individual or ranges of values in an existing raster grid and replace them with new values. There is a tabular area in the *reclassify* scheme where you can specify the old raster grid values as well as the new values that will replace them. Select **UK2006** as the *input raster* and *value* (meaning the numbers stored in each grid cell in UK2006) should automatically be selected as the *reclass field.* Choose an appropriate *output raster* name, e.g. **agric2006**.

By referring back to the **clc_legend.**csv spreadsheet, assign new values of 1 to those detailed classes that are within the 'agricultural areas' broad class (in the label1 column **clc_legend.**csv) and 0 to all other classes (see the illustration below if you are stuck). Note that pressing 'tab' at the end of a row will create a new row beneath the current row. When you are finished, press OK and ArcGIS will generated a binary raster output, where all grid cells classified as agricultural areas in 2006 are coded as 1 and everything is coded as 0.

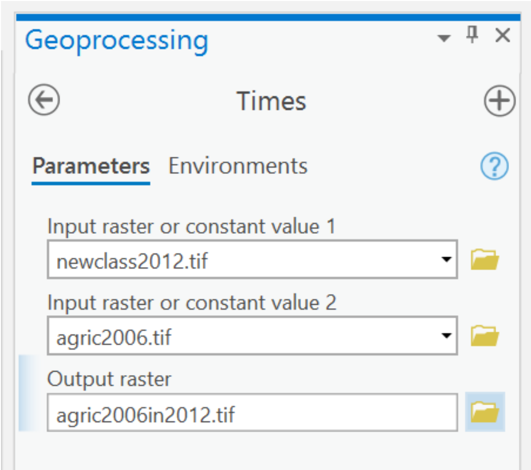


Next, we will run *reclassify* again, but this time we will use it to convert our random numbers into simulated land use / cover classes for the year 2012. Head back to *reclassify* again, and select **random** as the *input raster*. Again *value* should be automatically selected as the *reclass field*. Delete any rows that you do need in the same way as you did before when using *reclassify*. Next, enter in probability ranges that reflect the chances of agricultural area being converted to 'artificial surfaces' (coded as 1 in the screenshot below), remaining unchanged (coded as 2 below), becoming 'forest and semi-natural' (coded as 3 below), becoming 'wetland' (coded as 4 below), or becoming a 'water body' (coded as 5 below):



Click on Run and ArcGIS Pro will generate a set of simulated land use/cover categories for 2012 (with land uses coded 1 to 5 as detailed in the paragraph above), based on the transition probability matrix values for 'agricultural areas'.

Finally, to restrict our simulation results just to those areas that were classified as 'agricultural areas' in 2006, head for the geoprocessing panel again and search for and then run *times.* Multiply your simulated land use/cover classes (**newclass2012**) by the grid cells classified as 'agricultural areas' in 2006 – **agric2006**. Choose an appropriate output file name:



Click on *Run* again, and ArcGIS Pro will generate an output raster where zeros indicate non-agricultural areas in 2006 and the codes 1 to 5 are used for the five major land use / cover classes.

Task 2: By zooming in and out, explore the simulated land use / cover classes generated by this process. Are there any reasons why the land use change that we have just simulated is unrealistic? Can you think of any ways you might address these in ArcGIS Pro? [See the following page for some possible ideas on this]

Task 2: possible answers

- **PROBLEM: The transition probabilities are not very realistic and do not vary over space.** POSSIBLE SOLUTION: Instead of assuming that they are the same for a given land use / cover class regardless of where it is in the study area, we could vary the transition probabilities, depending on other map layer characteristics. For example, if a grid cell lay in a national park or nature reserve, we might calculate separate transition probabilities for such areas.
- **PROBLEM: The transition probabilities do not take account of classification accuracy, so our probability matrix will mix up real land use / cover change with misclassification of land use / cover in one or both of our study years.** POSSIBLE SOLUTION: If we had information on the accuracy of the land use / cover classifications in 2000 and 2006, we could use this information to adjust our transition probabilities downwards.
- **PROBLEM: The transition probabilities for each pixel are independent of those for neighbouring pixels**. Land use / cover change is thus much more piecemeal in our simulation than in reality and our simulation has a rather 'speckled' appearance. POSSIBLE SOLUTION: Instead of working with pixels, we could work with larger land use/cover parcels / polygons whose area matches patches of observed land use / cover change between 2000 and 2006.