Modelling Land Use and Land Cover Change using StocModLCC

Isabel Rosa, Robert Ewers

Before you start

First you need to go to Blackboard and download the PDF with the practical handout, the .zip file (SouthAfricaInputData.zip) containing the dataset and StocModLCC itself (an R script called stocmodlccSA.r).

These data come from a subset of a coastal South African landscape studied by Pieter Olivier:

> Olivier, P. I., R. J. van Aarde, and A. T. Lombard. 2013. The use of habitat suitability models and species–area relationships to predict extinction debts in coastal forests, South Africa. Diversity and Distributions **19**:1353-1365.

Normally, before you start running models you'd need to do a couple of things, but we've already done that for you today:

- Ensure that all layers of data have exactly the same cell or pixel size (e.g. 100 m) and the same spatial extent (i.e. same number of columns and rows).
- Ensure that all three forest cover maps (2005, 2008 and 2011) only have two classes (1- forest, 0- deforested). The other input data can be either numeric or categorical (e.g. 1- protected/0- unprotected).

What you need to do to get underway is:

1) Unzip the data files into your own folder – make sure you create a new folder for this data and place all of the files in the same folder. Copy the stocmodlccSA.r script to a different folder.

2) Inside the folder you just created, you need to create a new empty folder and name it "outputs". This is where all the outputs from StocModLCC will be placed automatically.

3) Take a couple of minutes to quickly look at the basic data. Open the forest cover data layers in QGIS, focusing on the differences between forest cover in 2005, 2008 and 2011 (filenames 'for05b.asc', 'for08b.asc' and 'for11.asc' respectively).

Running StocModLCC

1) To start the graphical user interface (GUI) of StocModLCC you need to open a new R session, simply by double-clicking its icon, and then type the following line of code into the command line of your R console:

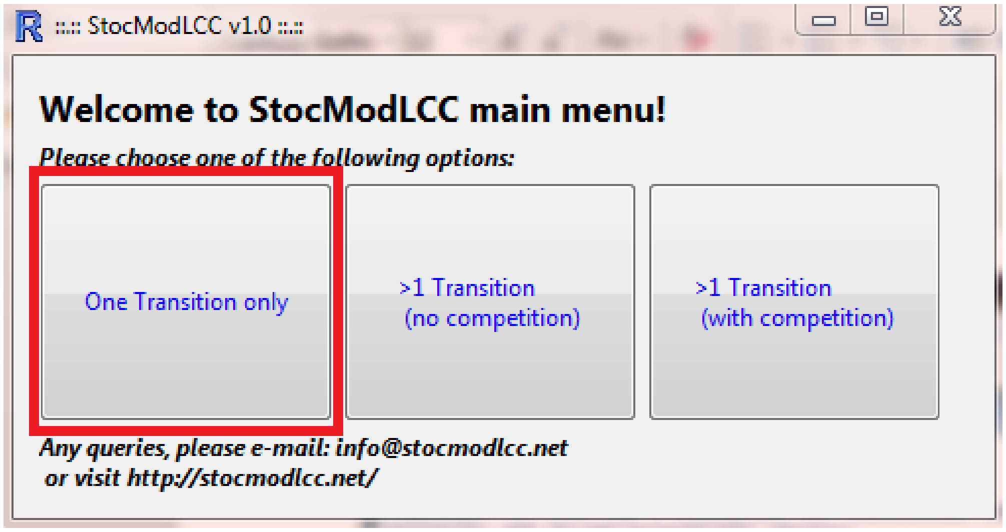
   ➢ source("C:/stocmodlccSA.R")

   - NB: make sure to write the correct path to the file, depending on where you placed the StocModLCC R script

   When you call StocModLCC for the first time it takes a little bit longer to load because it will install (when needed) and read in all the packages it depends on.

   You will almost certainly get an error message as the first thing you see. The first error message will tell you there's a missing .dll file; click 'Okay'. The second error message will say something about needing to install something called GTK+; select 'Install' and click OK.

2) Once all packages are loaded successfully you will see a window like the one being shown below. This is the main menu of StocModLCC. You should select 'One Transition Only'.



One Transition Model

In any land cover/use change modelling exercise you always need (at least) two land cover/use maps, from two different points in time. This will allow you to calibrate your model, i.e., determine which variables (that represent drivers and proximate causes of change) better explain the variations observed between those two years. The relationships found between variables and change in land cover will then be used to simulate land cover change rates and patterns. StocModLCC let you forecast and back-cast changes in land cover based on the data you input data. In this class we will learn how to forecast deforestation. The logic followed here for these particular land cover transitions is the same for any other type of land cover change you might be interested in modelling.

Modeling one land cover transition means that you only have two land cover classes in your initial land cover maps. For instance, the dataset "SouthAfrica" you downloaded from Blackboard contains three land cover maps (for05.asc, for08.asc and for11.asc), which only show forested and deforested areas in coastal South Africa (cell size = 100 m). In particular, **forest is coded as 1** and the areas **deforested are coded as 0**.

The goal of this modeling exercise is to **determine the probability of forest (1) changing to deforested areas (0) until 2020**, based on the input variables (or a selection of them). To do so you will first calibrate a model using the 2005-2008 data.

1) First, we need to indicate where our data (in .asc format) are and import it into R. To do so, you need to define your **working directory** by clicking the browse button and opening the **folder** where you placed your data.

   And don't forget that inside this folder you should have an empty folder named **"outputs"** where StocModLCC will place all of its outputs.

2) Once you selected the data folder, you just need to click "Import All Ascii Files". You'll get a popup telling you when this is complete.

3) After you get this message you can see in the R console the file paths of each file imported as well as the dimensions (number of rows and columns) of each data layer. This is a quick way for you to check if they all match (they must!).

4) If you want, you can now click on "Visualise data" to see the layers that were imported. It's slow though and could be more frustrating than useful. If you do take a look and get bored with waiting for it to load map after map, then use the escape button to get out.

   This is a good way to check if your data has been correctly imported. Once you see the first map you can either press "Enter" or use the mouse's left-click to see all maps imported (one at a time). If any of your maps looks 'funny' or if it does not look the way you expected then probably there is something wrong with the ascii file, and it is better for you to fix that before proceeding.

5) If all maps look good then click "Start StocModLCC" to proceed to Model Calibration.

6) Now you're ready to calibrate the model. First, we need to set up a couple of things. Importantly, we need to tell StocModLCC which of the data layers imported represent the land cover map in time 1 and time 2 (highlighted in red and orange below). This is extremely important! When forecasting you should have a first date (t1) that is from any year before the second date (t2). In our example this should be for05.asc (t1) and for08.asc (t2).

7) Once you indicate the two land cover maps you must click on “Compute Transitions Matrix” to determine the area that has changed between the two years. This information will be printed on the R console, and again you will see a popup message once it is done. Under “Land cover in time 1” and “Land cover in time 2” you see the number of pixels coded as 1 and 0. Under “Land Cover Change From time 1 to time 2” you see the number of pixel that did not change or persisted (coded as 0) and the number of pixels that changed from 1 to 0 (coded as 1).

8) Now that you have prepared the transitions matrix, you can plot this information on a chart. To do so, click “Plot Transitions”. What you will get is a graph that shows you how many pixels from each land cover in time 1, in this case 1 (forest) and 0 (deforested), have changed or persisted in the same class in time 2.

   In this example, the great majority of the pixels of forest persisted from time 1 to time 2 and there was no change in the deforested pixels, because these can never be forest again in this particular example. The only change that occurred from time 1 to time 2 was forest (1) changing to deforested areas (0), which is the transition we are interested in modelling.

9) Now you need to select which variables, of the data layers that you imported to R, you will be using to calibrate the models, which you do using the checkboxes in the “Variables to consider” section.

   Choose for yourself what variables you want to model, but try to base it on an expectation of what you think should be driving deforestation. You can choose as few or as many variables as you wish. Here, however, you just need to indicate whether you will be using all the data you imported or leaving some of it behind.

   You’ve only been given a small number of variables to choose from to ensure the simulations in this session run rapidly. They are:

   - droads2: Distance to roads
   - dtowns2: Distance to towns

- pop2: Estimated persons per grid square (1 ha). AfriPop (www.afripop.org) project data downloaded from website as tif for South Africa, reprojected scaled to 100 m x 100 m and clipped. Population count map for 2010 version 4 (v4).
- pareas2: The WDPA v10 database downloaded 2014 for Africa - Extract boundaries of PAs for South Africa, reproject and convert to raster as protected / not protected. Here using all PAs.

10) Next to the “Variables to consider” box is another titled “Categorical variables”. This is where you indicate if any of the variables are non-numeric. The implication of this choice is that further ahead you will have to choose a reference level for the variables indicated here. For instance, in our example, we only have protected areas as a categorical variable (two levels of protection: unprotected (0), protected (1)). If you chose to model protected areas, you would choose unprotected (0) as the reference level because we are interested in investigating whether protection is preventing deforestation effectively.

11) There is one more step before we start calibrating our model. StocModLCC is based on the idea that land cover change is a contagious process, meaning that changes in a particular land cover tend to happen closer to where it has happened before. Therefore, one of the variables that you must include in your model is the proportion of neighbors that belong to the land cover you are interested in modelling. This will be the only dynamic variable in the model. This mean that at each time step of the simulations the neighborhood will be updated automatically (using a moving window you define), whereas all other variables will remain the same.

    In this example, we are modeling deforestation; therefore, we are interested in determining the proportion of deforested neighbors surrounding any given forest pixel.

    To do this, click on “Calculate Neighborhood”. A popup box will ask you to enter the code of the land cover we want, which in this case is “0” (aka, deforested pixels).

    You will then get another popup asking you to define a neighbourhood size. This determines the width of the window in which you consider nearby pixels to be neighbours. Setting this value to 3 means only pixels that touch you are considered neighbours; a value of 5 means pixels located one-pixel distance away are considered neighbours; and so on. Choose a value that seems right to you based on what you think is happening in the landscape; i.e. if you think a man with a chainsaw is doing the damage, how far can he walk? If you think it's a man with a bulldozer, how fast can he drive?

The window size might affect the rate of change, so you might want to use a different size to that of the person beside you.

Yet another popup will popup to tell you when the neighbourhood has been calculated.

12) Now it's time to "Choose calibration method" through the dropdown menu. In this version of StocModLCC there is only the one method available for calibration, and that is Logistic Regression. Logistic regression is probably the most widely used statistical method for calibrating land use change models.

13) After you select "Logistic Regression" you can choose whether you want to tick the "Perform Automatic Stepwise" checkbox that appears at the bottom of the window. Leaving it unticked will mean all variables are left in the final model, whereas ticking it will run an automatic stepwise regression to reduce the number of variables in the final model. You will be asked to choose whether you wish to perform a forward or a backward stepwise regression. In addition, you will have the opportunity to manually update your model (remove an uninteresting variable) later on.

14) The final decision to make is to select the proportion of samples for training and testing. By default, 50/50 is selected according to the approached followed by Rosa et al. (2013) but you can choose to put more of your data in training if you wish.

15) Now go for it – click "Calibrate your model"

16) If you've put any categorical variables into your model, you'll instantly get hit with a popup asking you to input the reference level for it. In our example, we only have protected areas as a categorical variable, and as described above we chose 0 (aka, not protected) to be our reference level. 16)

17) If you asked for the automatic stepwise procedure to be performed, StocModLCC will ask whether you wish to perform a forward or backward stepwise regression.

    If it all goes well, the output of the model will show up in the R console, a .txt file will be added to the "outputs" folder with the summary of the fitted model.

18) Open the "best_model.txt" file that StocModLCC exported to the "outputs" folder and interpret the model calibration.

    a. How does each variable impact the probability of change?

    b. Which variables have a positive/negative impact?

19) You can compute the confidence interval around each parameter by clicking on the "Get GLM C.I." button – this button will have become visible once you finish the first round of model calibration.

    The confidence intervals are particularly important because StocModLCC simulations of future and past land cover change rely on randomly selecting parameter values from these intervals.

20) You can manually simplify your model if you're not happy with the automated output by using the "Update model" button. Be warned, though, that if you remove a variable and want to put it back in the model then you will have to restart the whole calibration. It's easy enough to do – just click "Calibrate model" again.

21) It's worth evaluating the statistical fit of your model using the "Evaluate your model" dropdown menu. This is not the same as validating a model, but is an important component of determining the reliability of your land use change predictions.

    There are two methods which you can use to perform this evaluation: the Receiver operating characteristic curve (ROC), including the area under that same curve (AUC); and pseudo-R2. Both methods are well explained in Hosmer et al. (2013), and you can search online for more information about these two metrics.

22) Choose one or multiple methods of model evaluation and set it running. A popup will tell you when it's complete.

a. When you choose to evaluate your model using the ROC curve, you will get a graph added to the "outputs" folder. The number on the top of the graph is the AUC value, and values that are close to 1 are what you're after. Values of 0.5 mean your model is random, and realistically you should be in the AUC > 0.9 to consider your model decent.

b. If you used pseudo-R2, the result will display in the R console. Contrarily to linear regression, in logistic regression (or GLM with binomial errors), pseudo-R2 values between 0.20 and 0.40 are considered good (Hosmer et al. 2013).

23) Play around with your model to get one that works. Choose different set of variables and calibrate it again, and use ROC and pseudo-R2 as your guides to model performance. This model you choose will be what you use to simulate future land use change.

24) Now it's time to simulate the future patterns of deforestation. To do so, you need to tell StocModLCC how many iterations and time steps you will be simulating.

a. StocModLCC is based on random procedures; therefore, each model run (or iteration) will produce slightly different outputs, based on sampling parameters from the confidence interval determined above. As a result, you are advised to run multiple iterations (100 is the default). However, large datasets take a long time to run, so in this exercise choose a small number (e.g. 2-5). If each iteration is taking too long, press Escape to crash out and reset with a reduced number of iterations.

b. The time steps correspond to the number of future points in time that you want to project. If you were using an annual calibration period (e.g. 2001-2002), you would be projecting annual changes in land cover (first year of predictions would be 2003). However, in this example, you're using a multi-year calibration period (2005-2008), so the model will output projections in three-year time steps (first year of projections will be 2011, second 2014, etc.).

   To project deforestation until 2020, using the calibrated 2005-2008 model, make sure to fill in the Time Steps box with the number 5 (time step 0 will be 2008, time step 1 will be 2011, time step 2 will be 2014, time step 3 will be 2017 and time step 4 will be 2020).

c. StocModLCC will export a number of files into the output folder:

- the change between time steps, exported as the file 'predicted_rates.csv'. This contains information on the **rate of land cover change**. In particular, it has a column with the number of pixels coded as 1 (forest), a column with the number of pixels coded as 0 (non-forest), and another column with the number of pixels that changed from 1 to 0 (deforested), at each iteration and time step.

- the 'annual' and the cumulative probability of change through time, exported as .asc mapfiles that you can open, see and manipulate in any GIS software. These are named in three ways: "predchange_i0_0yr.asc" (where the number after "i" corresponds to the iteration and the number before "yr" corresponds to the time step being simulated), "probyear0.asc", "cumprobyr4.asc" (where in both cases the number corresponds to the time step – StocModLCC averages all iterations in these two files!), respectively. The cumulative probability is given by the sum of the probability of change on each time step. The probability of change at each time step is the sum of the number of times a pixel was selected to change on that particular time point, divided by the total number of iterations.

- the model **parameters** used in each iteration of the model, exported as the file 'coefficients.csv'

25) Open the cumulative probability map for 2020 in QGIS and investigate where changes from forest to deforested areas, are most likely to happen in the future.

- Do the locations of the changes match your expectations based on the parameters you fed into the model?

- Any surprise locations showing up as deforestation hotspots?

- Any locations you thought should get deforested but haven't? What might explain that?

26) StocModLCC will give you some canned options for looking at the predicted changes in land cover using the 'Rates (barplot)' button. Alternatively, you can open the predicted_rates.csv to get the exact amounts.

- How much deforestation did your model predict for the 2008-2011 transition?
- How much deforestation actually occurred during that time?

Extension work: model validation

You have the data on observed forest cover in 2011, so if you want you can take the extra step to validate your model predictions over that first timestep. You can do this the simple way or the hard way – it's up to you.

The simple way is to use the canned 'Model Output Validation' options StocModLCC now provides on its console window and follow through the popup windows. If you need more precise instructions, visit the manuals on http://stocmodlcc.net/. Keep in mind these validations are relatively simplistic and ignore many of the issues we discussed in the lecture.

The hard way is to get online to pick up the R code for SimiVal from here: http://www.luflondon.co.uk/lulcc-models.html. This is the more complete validation approach I presented in the lecture, and it's described more fully in this paper:

> Bradley, A. V., I. M. D. Rosa, R. G. Pontius Jr, S. E. Ahmed, M. B. Araújo, D. G. Brown, A. Brandão Jr, G. Câmara, T. G. S. Carnerio, A. J. Hartley, M. J. Smith, and R. M. Ewers. 2016. SimiVal, a multi-criteria map comparison tool for land-change model projections. Environmental Modelling & Software **82**:229-240.

If you tackle this validation, this to think about are:

- What do you think it could be done to improve model's performance?
- Which variables could we be missing here?
- Do different model formulations give better results? Why?