Introduction

The Pearl River Delta, and particularly Shenzhen City, in Southern China has been going through a dramatic urbanization process since 1978 due to decentralization policies and market- oriented reforms.

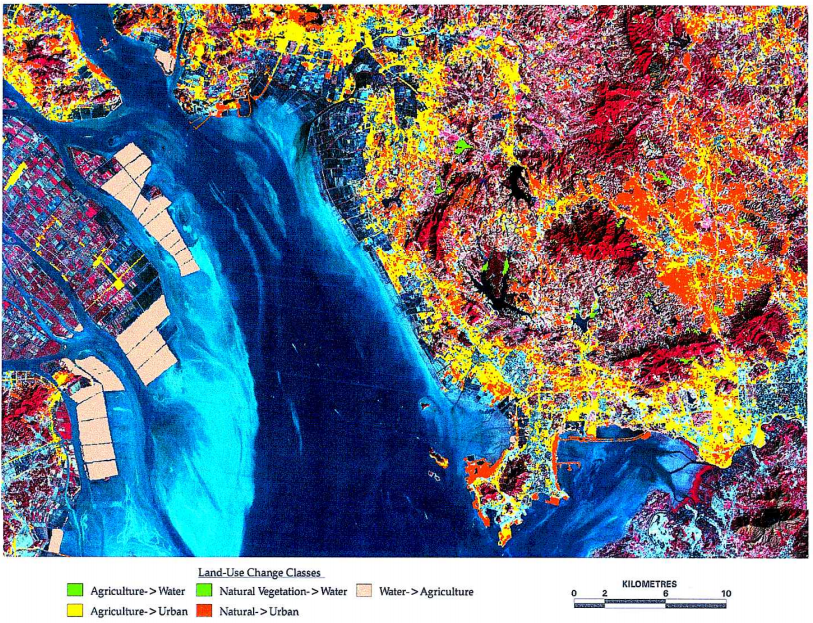
 

In this project, we aim to quantify the change of land use due to urbanization, and attempt to explain such change by a number of important socioeconomical drivers. Nigh yearly Landsat TM images are given from 1988 to 1997. See sample images below (1988 on the left and 1996 on the right) and more from NASA at <http://svs.gsfc.nasa.gov/stories/Landsat/pearl_river.html>.

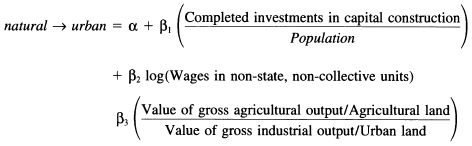
 

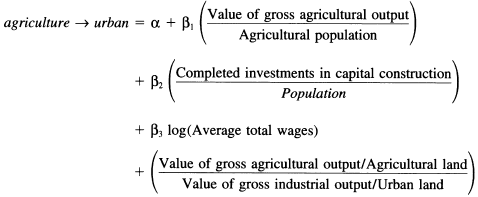
 

As a first step, we need to quantify at least three types (urban, agricultural, and natural) of land uses for each given year. A number of classification methods have been taught in this module, and Maximum Likelihood is a good starting point for multispectral TM imagery. There are also other image processing methods could help to identify land use classes (e.g. vegetation indices, filters, segmentation, etc.). Then, we can identify where changes have happened and how much land area has been transformed into urban built. See example land use change map below from Seto et al. 2002 ([Link](http://www.unc.edu/courses/2008spring/geog/577/001/www/Seto02-IJRS.pdf) to article).



Secondly, we can establish an empirical model to link the changes in land use to a number of key socioeconomical factors (e.g. capital investment, land productivity, population, wage rates, etc). A simplified model is adopted from Seto & Kaufmann (2003, [link](http://le.uwpress.org/content/79/1/106.short) to article). This second part of the project involves running this model and calibrating it with the derived land use data. Two equations are given here, one for modelling the percentage of *natural vegetation* land converted into *urban* area, and the other for the percentage of *agricultural* land converted into *urban* area:





References:

Seto, K. C., C. E. Woodcock, C. Song, X. Huang, J. Lu, and R. K. Kaufmann. (2002). Monitoring Land-Use Change in the Pearl River Delta Using Landsat TM. *International Journal of Remote Sensing*, 23 (10): 1985-2004.

Seto, K. C., and Kaufmann, R. K. (2003). Modeling the drivers of urban land use change in the Pearl River Delta, China: integrating remote sensing with socioeconomic data. *Land Economics*, *79* (1): 106-121.

1. [The first part of the project involves extracting land cover and land use area extent from Landsat TM imagery.](http://www.geog.ucl.ac.uk/~plewis/geog2021/project/EO.html)
2. [The second part of the project involves running a statistical model and calibrating it with the derived classification data](http://www.geog.ucl.ac.uk/~plewis/geog2021/project/Model.html)
3. [Project write-up](http://www.geog.ucl.ac.uk/~plewis/geog2021/project/write.html)

Part 1: Extraction of land use extend

1. **Introduction**

Your task in this section is to calculate the area of land cover & land use (LULC) extent (in m2) for each year of a series of Landsat imagery, along with an associated characterization of uncertainty in each of these areas.

1. **Download and examine the data**

The main image datasets you will be using to determine LULC extend over time are extracted sub-scenes of Landsat World Reference System path 122, row 44 from July 1988 to January 1996. The original data can be downloaded from USGS Global Visualization Viewer: <http://glovis.usgs.gov/> (you may wish to register for an account).

Make a directory to work in (e.g. ~/DATA/project) and download the image data to your working directory:

machine% mkdir ~/DATA/Project

* + It’s highly recommended to download the data by yourself. However, if you really have a hard time downloading the data, they can also be copied from here:

machine% cp ~plewis/public\_html/geog2021/project/PearlDelta19\* ~/DATA/Project

The seven files to download are: LT51220441988185BJC01 (03-Jul-88), LT41220441990358XXX05 (24-Dec-90), LT51220441991033XXX04 (02-Feb-91), LT51220441993278BJC00 (05-Oct-93), LT51220441994297BJC00 (24-Oct-94), LT51220441995364CLT00 (30-Dec-95), and LT51220441996063CLT00 (03-Mar-96).

**LT51220442000042BJC00 (11-Feb-00), LT51220442000250BJC00 (06-Sep-00),** LT51220442001060BJC00 (01-Mar-01), LT51220442001364BJC00 (30-Dec-01),

While you are downloading these files, think about why they are chosen over the other available files? What does “CC” mean among the “scene information”?

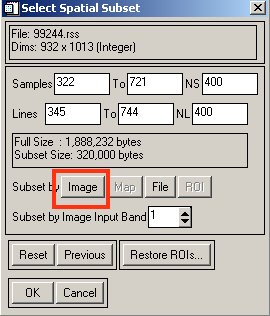
Alternatively, you can download the same files form USGS Earth Explorer links in the following format: <https://earthexplorer.usgs.gov/download/options/3372/LE71220442002007SGS00?node=GV> (login may be required)

Most of these data have been registered and calibrated for you, except the one taken on 1995364. This image needs to be re-registrered. You can use the “Image to Image Registration” tool in ENVI. Each compressed file consists of TIFF images from 7 spectral bands, as the standard of the LandSat TM data. You can unzip these TIFF files into separate folders.

Once you are able to open any of these images, you should be able to recognize features in the images, such as the Pearl River Delta and Shenzhen directly east of it. The Landsat data are at a spatial resolution of 30 meters. To reduce the amount of work, we need to “cut out” the area of less interest to us, and only keep the Shenzhen area for the rest of the study. To do so, we can use the “Spatial Subset” tool in ENVI.

Enter the beginning and ending Columns of 3731 and 5750 in the **from** and **to** fields. Enter the beginning and ending Rows of 4596 and 5925 in the **from** and **to** fields.

<FIGURE: screenshot of SubSet tool>



Now give yourself some time to explore this subset in ENVI and examine what types of land use and land cover can be observed from each of the given dates. Also, can you see any changes of land use over time?

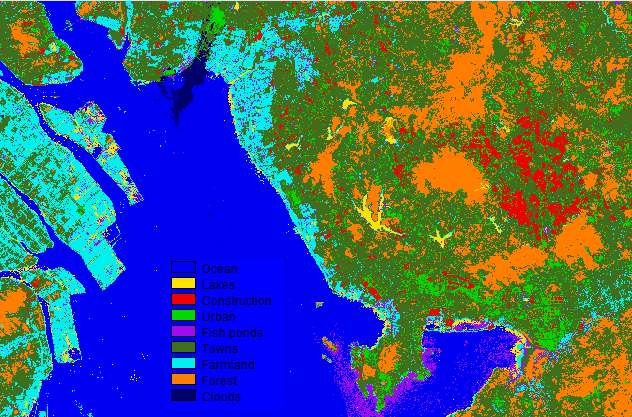
<FIGURE OF SAMPLE IMAGE of 1995>



(If using newer ETM images, students will have to deal with Destripe. Steps are *Basic Tools*→*Preprocessing*→*General Purpose Utilities*→ *Landsat ETM+ Destripe*.)

1. **Identifying land use classes**

Once you are familiar with the given scene, you can start marking some training sites (ROIs) with the tools available in ENVI. We will need to classify these seven image sets into meaningful land use categories.



You can utilize any method we have learnt from this module to classify the given images. However, Maximum Likelihood is thought to be a safe choice to perform this task. The classification goal here includes the identification of three basic “land use” classes (i.e. urban, agricultural and natural vegetation) that we will be using in building the socio-economic model. However, it is suggested to classify the image into a number of spectral distinctive “land cover” types, as opposed to simply classify it into the three basic “land use” classes. (What’s the difference between “land cover” and “land use”?)

Some possible ROI classes can be: water, clouds, fire smoke, et al. You can start building ROIs and run classification with a larger number of spectral classes, and then, after classification, combine multiple spectral classes into general land use classes.

* + You may notice some cloud contamination in the given images. The location of clouds changes over time of course. Would this affect your classification results? If so, how do you avoid it?
  + While you are classifying the images, some degree of sensitivity has to be taken into account.
  + Sample LU-LC classes:
    1. Water ( #1 ocean harbor, #2 lakes)
    2. Urban (#3 construction, #4 urban built, roads, #6 towns & villages)
    3. Agricultural (#5 fish ponds, #7 crop fields)
    4. Natural vegetation (#8 forest)
    5. Other (#9 clouds)
  + Your initial classification results probably look quite noisy. Is there any procedure you can undertake to eliminate the noise? E.g. majority filter, or post-classification segmentation?

1. **Calculate the land use area extent (and sensitivity/uncertainty)**

Having achieved satisfactory classification results, you can quantify the area of changes (in the unit of m2), and extract the proportions of changed area for **each pair** of land use type. Take screenshots of your classification outputs, note all the numbers you have arrived at, and it may even worth to make an excel sheet of your estimates.

If ETM data: Use the ENVI software program to open the two Landsat ETM images you have selected above. If using ENV 4.x or the Classic interface, from the main menu bar, select Basic Tools 🡪 Preprocessing 🡪 Data-Specific Utilities 🡪 Landsat TM 🡪 Landsat Gapfill. If you are using the new ENVI 5.x interface look under Toolbox  Extensions. This will open the Select input file(s) and processing type dialog, Figure 2. You should only use the option Two band gap-fill (Local histogram matching), the other options do not perform as well. More detailed instructions can be found here: <http://www.yale.edu/ceo/Documentation/Landsat_ETM_Gap_Fill.pdf>.

1. **Generate a time series of land change data**

Next, we will model the “growth” of urban land. It is essentially modeling the areas that have “changed” into urban area. So now we need to calculate the amount of land area changed from “natural vegetation” to “urban”, and land area changed from “agriculture” to “urban” respectively, for each time step. Save them into a .dat file, and name the file something like “LUC\_UrbanGrowth.dat” - we will need this file later.

Part 2: Model the growth of urban land

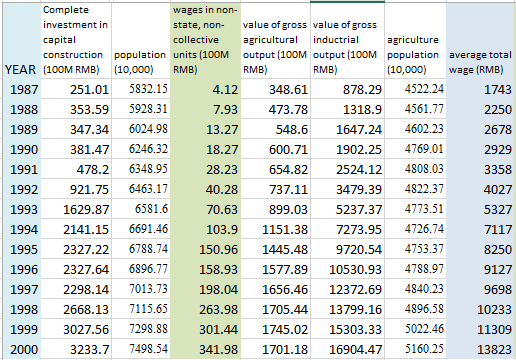
1. **Understand the urban-growth model**

There are two parts in the original growth model, first being modeling the urban land transformed from natural vegetation (“Nat-Urban model” hereafter), and second being urban land transformed from agricultural land (“Agr-Urban model” hereafter). To simplify this process, we only ask you to calculate the total amount of urban area gained in each time period, and a combined model would look like:

In this model, presumably changes are driven by five factors: per capita investment, per capita agricultural output, wages in private sectors, average wages, and per unit productivity of land. The per unit productivity of land, , takes into account the agricultural and urban land area from the previous time step. This model assumes the amount of land changed into urban built area is a linear combination of these five factors, and we have collected or calculated data for these factors by now. What we don’t know yet is how much each factor contributes in the model, or, in another word, the coefficient of each factor (i.e. α, β1, β2, β3, β4 and β5). In order to build this model for the use of future prediction, we need to figure out the best value of the coefficients. This is essentially a model “inversion” problem.

1. **Gathering socio-economic data**

There are many freely available online database for downloading important socio-economic data. Some examples are the Center for International Earth Science Information Network (CIESIN) at Columbia University and the World Bank. The data we will be using here are from Guangdong Statistical Yearbook: <http://www.gdstats.gov.cn/tjnj/2006/ml_e.htm>. A spreadsheet with relevant data has been compiled and can be downloaded here: “Guangdong.dat”



1. **Building the model (with Numpy? Or in Excel)**

We have written a little python module (copy it: Guangdong\_Urban\_Growth.py) for you to inversely derive the coefficients based on your calculation of land use change and the socio-economic data. Before you run the module, make sure your land use change data have been saved in to the “LUC\_UrbanGrowth.dat” file. Run this python module “Guangdong\_Urban\_Growtt.py” from the terminal with the following command line:

% python Guangdong\_Urban\_Growth.py LUC\_UrbanGrowth.dat

This module reads your input file “LUC\_UrbanGrowth.dat” into an array, and then calls a number of functions to solve the model inversion. The output coefficients and a figure will be listed on your terminal once the calculation is finished.

Take notes of your output coefficients for the model, and save your model figures. In your final write-up, you will need to explain this model. What is the final form of this model? What is the RMSE? How well do you think this model fits into the given data?

**Uncertainty?**

Part 3: Write-up