How to acquire GIS data, remote sensing and errors in GIS

Lecturer: Dr Dan Cornford

d.cornford@aston.ac.uk

http://wiki.aston.ac.uk/DanCornford

CS3210, Geographic Information Systems, Aston University, Birmingham, UK

http://wiki.aston.ac.uk/CS3210

March 10, 2009
How do we get raster and vector data in the first place?
What is remote sensing?
The electro-magnetic spectrum.
Some common remote sensing platforms.
Errors in GIS – what are they and do they matter?
Assume we are trying to **acquire** raster data to be stored in an array structure that represents square cells.

Several possible methods:

- remote sensing;
- sampling and (interpolation);
- scanning existing sources;
- conversion of existing data.

Different methods needed for different variables.

Assume that the value in a cell represents the mean value within the cell?
Acquiring vector data

- Most common is **digitising**:
  - expensive to do since it is manual;
  - uses existing paper maps (propagates their errors).
- Can scan data – this is OK but **automatic feature extraction** is unreliable – requires raster to vector conversion.
- Alternative is to go into the field and record data directly:
  - surveying – using theodolites and poles is slow, expensive, but very accurate;
  - GPS (**global positioning system**) – uses satellite constellation and triangulation – quicker but slightly less accurate.
What is remote sensing?

- The act of **sensing remotely**!
- Examples: our eyes, cameras, digital cameras ....
- Digital methods use arrays of **charge-coupled devices (CCDs)**.
- Project the image onto the CCDs using some optics.
- Not just using visible light, the whole **electromagnetic spectrum** is used.
Electro-magnetic radiation

- Emitted by objects with energy .... wavelength of the radiation determines its properties:

- Different spectral regions have different properties.
Electro-magnetic radiation

The spectrum has the following characteristics:

- **X-ray** – high energy and not very useful since absorbed in the atmosphere,
- **ultraviolet** – high energy, largely absorbed (ozone),
- **visible** – from the sun, not heavily absorbed,
- **near infrared**,
- **thermal infrared** – emitted by objects on the surface, partially absorbed,
- **far infrared**,
- **microwave** – increasingly used, but not intuitive,
- **radio-wave** – used for communications.
Spectral regions of **low atmospheric absorption** are called window regions.

Clouds cause problems in the visible and IR regions.

In general **passive instruments** are used.

For spectral regions with no large natural emitters (e.g. microwave) satellites must provide their own source (**active instruments**).

There are a number of instrument geometries - arrays of CCDs, linear CCD arrays, and scanning instruments which use mirrors and small numbers of CCDs.
• **Geostationary** orbits mean that the satellite remains over the same location. Best for *synoptic* data acquisition - e.g. weather observations.

• **Polar** orbits place the satellites in orbit around the Earth. Orbit means that any one spot on the surface of the Earth is observed once every 1-30 days. Polar orbits are about 650 km (c.f. Geostationary 36000 km), thus the ground resolution is much better - down to 1 m.

• Having acquired the image it requires rectification: photo-; geo- and ortho-.
Launched 1972, polar orbiter, thematic mapper has 7 bands:

- Band 1 – blue green (visible),
- Band 2 – green (visible),
- Band 3 – red (visible),
- Band 4 – near infrared,
- Band 5 – middle infrared,
- Band 6 – thermal infrared,
- Band 7 – additional middle infrared,

with a 30, 120 and 60 m ground resolution.
Examples - Landsat

Landsat bands, and Amsterdam, 1987
Examples - SPOT

Chesapeake Bay and flooding in El Salvador

- Polar orbiters, four bands:
  - 3 bands in the green, red and near infrared parts of the spectrum – 20 m ground resolution.
  - 1 pan-chromatic (all visible) band – 10 m ground resolution.
- Not as good for land-cover, but often used to create / update / correct GIS databases.
- Good in urban areas – 10 m can resolve many buildings.
Examples - SAR

Another polar orbiter, but this time the sensor is Synthetic Aperture Radar (SAR).

Works in the microwave region of the spectrum (6 cm wavelength) – day and night, cloudy or clear.

Gives information on the surface roughness, and electrical properties over the 25 m footprint.

Can be used to construct DEMs, land-cover mapping and mineral exploitation.
Remote sensing and GIS strongly linked. Use of GIS will drive the market for remotely sensed data.

Most simple application is to use the images as a backdrop – rather common.

Other uses may require a lot of processing:
- classification of land-cover,
- updating / correction of existing data (often semi-manual),
- creation of DEMs using two images taken from different positions,
- monitoring of crop growth, pests or irrigation needs.

Future lies in data mining.
Remote sensing use likely to increase.

Better resolution (ground and spectral), more bands, more frequent.

IKONOS satellite has a 1 m pan-chromatic band, e.g. Rio (store the whole of UK: \(\sim 600 \text{ Gb} \) at 8 bit resolution).

LIDAR - creation of high resolution DEMs – flood insurance.
Increasing use of aircraft acquired results - e.g. thermal mapping of cities.

Millennium mapping project – 12.5 cm resolution aerial photographs of the whole of the UK.
Errors in spatial data

- Errors are especially important in GIS because:
  - people are inclined to absolutely believe digital maps?;
  - any modelling that uses base data with errors can produce unpredictable results.
- First consider error propagation: a good example is pool.
- The implication is that if our models behave as a pool table our data must be very accurate (and precise).
- Fortunately not many models are this sensitive.
Types of errors in spatial data

- Precision;
- Accuracy:
  - geometric,
  - scale,
  - attribute,
  - topological;
- Consistency:
  - scale,
  - source,
  - classifications;
- Currency;
- Completeness;
- Model.
Dealing with errors

- How to represent errors?
- Depends on the data model.
- Using a raster data structure we might store the mean and variance, to denote attribute accuracy.
- In an object model we could do the same to store geometric accuracy.
- These are the main sources of error.

Realisations of slope from a DEM with errors.
Dealing with errors

- Recall models can produce extreme errors on the outputs given small errors on the inputs.
- One way to deal with this is to create multiple realisations of the data which cover the range of errors on the data, and run the model using these inputs.
- This is Monte Carlo analysis – powerful tool and does not need a parametric (equation based) error model.
- In sensitivity analysis we carefully choose those variables we think are most important, perturb these, and then run the model.
- This is very computationally expensive with big models or data sets.
The key points were:

- Raster data acquired using remote sensing or interpolation, vector data using a variety of methods.
- Remote sensing uses a range of spectral bands to determine information on the Earth’s surface characteristics.
- Remote sensing is a cheap way of getting near global coverage.
- Much processing is required to convert the raw data into information useful in a GIS.
- Errors in data are important and should be, but are generally not, treated carefully when using GIS.

A real challenge for the future is the accurate extraction of useful information from the high resolution remote sensing data.