9.2 Remote Sensing

We have seen in the previous section that raster data can be acquired by interpolation of point samples. In this section we look at how remote sensing can be used to acquire huge amounts of spatial information which is, in the first instance, best represented in the field model and stored in the raster data structure.

The concept of remote sensing goes back a long way (our own eyes are the most familiar remote sensing tools) but only recently, in the digital age, has the full power of remote sensing been realised. The essence of remote sensing is to view, from a remote location, the object of interest. We are probably most familiar with this in the context of photography, where we use a photosensitive film to capture an image of a remote object. Digital cameras are now common-place, and these replace the film with an array of charge-coupled device (CCD) based sensors which measure the intensity of the projected light at a certain wavelength over an array of points in the image. This produces an array of pixels, whose value represents the intensity of light emanating from the corresponding point in the image.

In remote sensing the same technology is used. In the past optical cameras with film were used, but now the sensors are almost all electronic – even on the 'spy-satellites'. In general we think of remote sensing as referring to satellite borne instruments, although a large portion of data acquired by remote sensing is still from aircraft mounted systems. The great benefit of remote sensing over other methods of data acquisition is the ability to cover large regions of the Earth's surface at relatively low cost (although remote sensing is still often not economically viable without government support). These large regions can also be assessed synoptically (at the same time) and repeatedly, to build up a picture of change. If you think remote sensing is largely concerned with sensing at visible wavelengths you are wrong. Increasingly, other regions of the electro-magnetic spectrum are utilised to provide very useful information.

9.2.1 The electro-magnetic spectrum

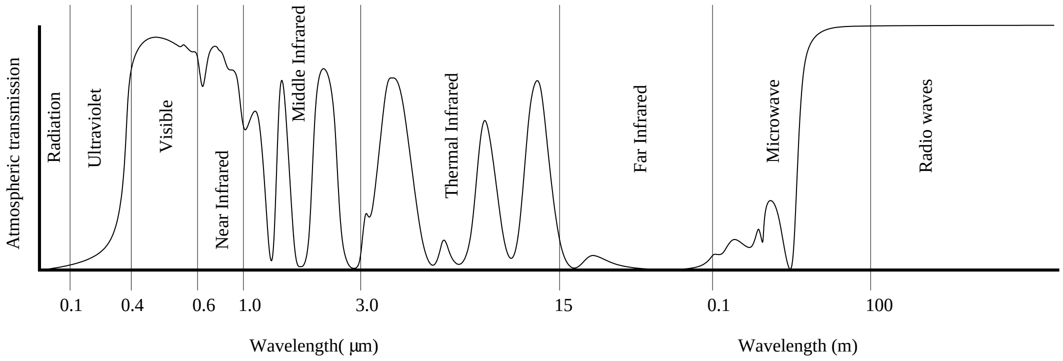

Figure 9.11: A *sketch* of the atmospheric transmission of EMR for the full region commonly used in remote sensing, based on (Mather, 1999).

Electro-magnetic radiation (EMR) is emitted by all objects which have energy. You and I are currently giving off large amounts of electro-magnetic radiation in the thermal infrared region of the spectrum. We are doing this because our bodies are hot – indeed the wavelength of the radiation emitted is a function of the temperature of the emitting body. The useful EMR spectrum is shown in Figure 9.11, together with the amount of radiation at each wavelength absorbed by the atmosphere. Roughly speaking we have:

- 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.

When deciding on what wavelength to use for remote sensing there are several factors to consider. The most important question is: what is the object or property that is to be observed? Here we will assume that it is something at the Earth's surface, so the first requirement is that the atmosphere, through which the radiation will generally have to pass twice, is transparent to EMR of that wavelength. Regions of the EMR spectrum for which this is true are shown in Figure 9.11. In general there are certain ranges of wavelengths in which the cloud free atmosphere is approximately transparent, these being called window regions.

It is also important to consider what source of radiation is to be used. The sun provides good illumination (on one half of the globe at a time) in the ultraviolet, visible and near infrared regions of the spectrum. Thus in these regions we can use passive sensors which measure the amount of sunlight reflected or scattered by the surface in window regions of these wavelengths. The Earth itself emits radiation in the thermal infrared, regardless of solar illumination, thus passive sensors can be used in this region of the spectrum. The final region commonly used, is the microwave region – here there are no natural emitters, thus an active instrument, which creates its own radiation source and bounces this off the surface of the Earth must be used. Radar based ideas operate in this region of the spectrum, but are complicated to use in satellites because of the large amount of power needed to generate the radar beam.

Typically the instrument is composed of an array of CCD sensors which respond in the desired wavelength only, with some optics in front to focus the image onto the sensors. In some systems the number of sensors is reduced by using mirrors to scan across a linear array of sensors, or even a single sensor, however this means that mechanical parts are needed in the instrument – a potential disadvantage in space.

9.2.2 The platform

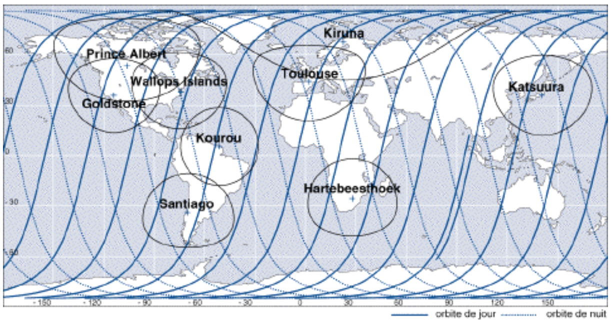

Figure 9.12: The orbit of the SPOT series of satellites.

Another issue that must be considered is the platform. As a general rule the further the platform is from the surface of the Earth, the bigger the region that can be seen, but the worse the resolution of the instrument becomes. Considering satellites there are two options:

- **Geostationary** orbits mean that the satellite remains over the same location on the surface of the planet at all times. For this orbit to be stable the satellite must be a long way from the Earth (about 36000 km – to minimise the effect of its gravity). Thus geostationary satellites typically have poor ground resolution (of the order of 1–10 km). There is, however, the benefit of being able to see almost half the globe at the same time, which means very good spatial coverage and time resolution can be provided. Thus this sort of system is largely used for weather monitoring / forecasting.
- **Polar** orbits place the satellites much closer to the Earth's surface, but rather than staying in a fixed place they orbit the Earth, typically almost over the poles. Thus they observed a swathe of the Earth's surface (Figure 9.12) on each orbit, this swathe moving across the surface of the Earth both as the satellite orbits and the Earth rotates. This means that for most instruments the same point on the Earth's surface is sampled once every 1 to 30 days. The benefit of this lower orbit (about 650 km typically) is that the images acquired have a very fine resolution, typically from 1 km down to 1 m. The finer the resolution, the smaller the swathe in general.

Constraints on the resolution of the instruments include not only the optics, but also how to transmit the information down to Earth quickly enough, and what to do with all the data when it arrives.

Once the image has been acquired and transmitted to Earth we need to be able to correct several image characteristics. There are at least three corrections which are typically applied:

- photo-rectification – remove any distortion caused by the optics of the imaging system,
- geo-rectification (also called geo-referencing) – map the image into the ground coordinate system,
- ortho-rectification – remove any distortion cause by the topography.

Each of these processes is a highly mathematical operation which we will not cover, but it is important to understand that remotely sensed data imported into a GIS have already undergone a number of manipulations which themselves may have introduced additional errors into the data. Thus although remote sensing does provide pretty much the only means for obtaining primary, synoptic, spatial data the 'primary' data has been heavily processed.

9.2.3 Some satellite systems

This section briefly outlines some of the more commonly used remote sensing (satellite) systems. Possibly the most well known is the Landsat series of satellites. Like any other machine, satellites have a limited life thus there are often many satellites all launched for the same mission (although often with upgraded or additional sensors). The Landsat series of satellites dates back from 1972. These are polar orbiting satellites, with sensors in 7 bands on the thematic mapper instrument:

- 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.

All bands other than six have a ground resolution of approximately 30 m, with band 6 having a resolution of 120 m.

The primary mission for Landsat, as the name suggests, was the monitoring of land-use, land-use change, crop growth and mineral exploration. Since many of the sensors are passive and in the solar range, the satellite has a sun synchronous orbit, meaning that each orbit is half in full (near midday) sunlight. The main problem for Landsat is cloud cover, which can mean that some locations are only very rarely sampled.

Landsat has been used in many applications, but largely in the field of land-cover mapping, particularly in less well surveyed countries. This type of information is often very useful in a GIS and has many applications, from agricultural planning to site selection to mineral exploitation. The relatively regular temporal coverage means that the satellite has been used to monitor crop growth and climate change (e.g. in the Sahel).

Another well known system is the SPOT series of satellites. These French satellites are polar orbiters with sensors in the green, red and near infrared parts of the spectrum (at 20 m ground resolution) and a panchromatic sensor (sensing at all visible wavelengths) with a 10 m ground resolution. The SPOT satellites can fulfil some of the roles that Landsat does (although the fewer spectral classes mean less accurate classification of land-cover) but the better spatial resolution means that features such as roads and buildings can be resolved on SPOT scenes. Thus SPOT data can be used to automatically update and improve existing GIS databases.

The final satellites to be considered are the polar orbiting ERS series of satellites. These include an on board Synthetic Aperture Radar (SAR) sensor. Unlike the other satellites considered so far this is an active instrument, which bounces a pulse of C-band radar (6 cm wavelength) off the surface at an oblique angle. The amount of backscattered energy is measured, which gives information on the surface roughness, and electrical properties over the 25 m footprint. SAR images are not like those we are used to seeing with our eyes, but have proved very useful for land-cover mapping, mineral prospecting and DEM creation. SAR is particularly effective since it is not really affected by cloud.

9.2.4 Applications

Once we have the remotely sensed images, they will generally need further processing to extract the information desired. At their most simple remote sensing images are used as backdrops for other information in the GIS, e.g. showing the route of a new power line in context. This is the most elementary use of remote sensing data but still requires rectification of the images.

More commonly we will use a wide variety of methods (some of which are very powerful and mathematically advanced) to extract the information we need from the remotely sensed images. This might involve:

- ➢ 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.

One of the fastest developing areas of science is the extraction of information from huge data sets (data mining), and remote sensing is one area where these methods are very relevant.

9.2.5 Into the future

The use of remote sensing is set to increase in the future, with higher resolution (spectral and ground footprint) instruments and ever more satellites. Current satellites can give ground resolutions of 1 m in pan-chromatic bands (e.g. IKONOS), for small regions of the Earth's surface. At this resolution it is possible to detect cars and other small objects. This is a relatively cheap way of surveying very large areas accurately although good rectification is important. Future trends look set to provide still higher resolution and in smaller (and more) spectral bands.

The Millennium mapping project is another interesting remote sensing project, which is aiming to produce a complete digital aerial photograph coverage of Great Britain at very high resolutions. This can be found on the web, from where you can probably obtain an image of your home.

An interesting recent innovation in remote sensing is to use lidar (laser-radar) fired from a plane at a known position directly down to the surface to infer a very accurate DEM of the surveyed area. This can produce 1 m resolution DEMs which are accurate to within 10 cm. This sort of information might be very useful when assessing the risk of flooding of a given area. Insurance companies are particularly interested in this.

But the real challenge for the future, in my opinion, will be the extraction of useful information from the accurate, high resolution data. This is a challenge that requires involved mathematical and statistical models, which we cannot cover here.

For more details on remote sensing the reader is referred to Mather (1999).

Mather, P. M., 1999. *Computer Processing of Remotely-Sensed Images.* Chichester: Wiley.

9.3 Scanning

An alternative to remote sensing and possibly the easiest way to acquire new raster data layers is to scan in existing maps. This is commonly done to provide a backdrop to show other analyses, or provide simple map display systems. It is not easy to directly analyse these maps, thus further processing is often required before they can be used in a GIS. It is also necessary to overcome problems caused by errors in the original maps and the distortion of the paper copies in storage.

9.4 Conversion between fields and objects

We do not have time to cover this in any depth, but there are a series of very interesting algorithms which we can use to convert from raster to vector. This is most easily accomplished for scanned images, where the scanned maps contain only the features of interest. In remotely sensed images the problem is extremely difficult, since the images tend to contain a lot of information and detail which is not relevant to your data requirements, and this must be filtered in some way. Details of the Zang-Suen erosion algorithm, which can be used to extract vector data from scanned maps, can be found in Worboys (1995, p. 230).

10 Sources of Vector Spatial Data

There are several sources of vector data which can be accessed. The most commonly used method for acquiring vector data is digitising. This involves a user defining the vertices of the objects being digitised, often from an existing paper map or aerial photograph using a digitising tablet. Attributes are also often added at this point. The main drawback here is that the digitised data is only as reliable as the original map from which it is digitised, and that extra operator errors are often added during digitising. It is also expensive, since it is very labour intensive.

Scanning can also be used, generally in a semi-automated way, to vectorise existing maps. We could also use remote sensing images as the source. Methods to extract vector information from these raster data structures require complicated mathematics, or a great deal of operator intervention.

The global positioning system (GPS) can be used to obtain vector data from field based studies. GPS based surveys are probably the most accurate source of geo-referenced data we have available. GPS uses a constellation of satellites (which are controlled by the US military) to triangulate the receivers ground position using a number of 'visible' satellites. Using repeated samples it is possible to estimate position to around 25 m, however if an additional fixed base unit, which is sited at a known position, is used in addition to the field instrument, many of the atmospheric effects which cause most of the errors on the position measurement can be eliminated. Thus using differential GPS, as the method is called, positional accuracies to tens of centimetres can be achieved. To obtain greater accuracy, traditional surveying methods, using laser enhanced equipment is necessary. This is labour and skill intensive.

11 Errors in Spatial Data

One of the major problems in a GIS is the keep a handle on the analysis errors. This is particularly important for two reasons. First, people tend to believe that results presented as digital maps are exact. Secondly, errors in the base data tend to build up (propagate) during many analyses. Thus the results produced by a GIS should be regarded as skeptically as those produced by any other model.

We have already seen an example, related to precision, where the intersection point of two lines could not be represented in the computer system with sufficient precision. Precision is a measure of the ability of the digital number (which by its very nature represents a discretisation of the real number line) to represent the location of an object.

Say we were interested in storing the location of buildings to a precision of 1 meter, across the whole of Great Britain. The domain has a size of $O(1000\ km)$ or $10^6\ m$. With a 16 bit numbers we can represent the spatial location to a precision of about 16 m. Thus we will need to use at least 32 bit numbers to represent the coordinates, and if we increase the precision required and the areas considered we may have to go up to 64 bit reals (doubles).

Of course precision is just one part of the story because we can have very precise measurements which are very inaccurate, such as the distance between London and Birmingham is 24.3432545434512 km – very precise, but this is false precision. However if we assume that one day our data will be accurate, precision may be a significant source of error. Analyses such as point in polygon and line intersection algorithms are potentially sensitive to precision errors.

Of more concern is the propagation of errors. Consider trying to pot a ball on a pool table. Even if you are not that good the chances are you can pot the ball if you hit it directly with the cue. But if you have to hit the white first and play this onto the ball you want to pot it gets harder. A small error on hitting the white might produce a big error on the red. If you have to do a plant (hit the white into a red which hits the other red you want to pot) it becomes even harder because the errors multiply. This is common in some GIS analyses, although others are more conservative and have better error propagation properties. We can think of several types of error:

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

We do not have time to fully cover all these issues, so they are raised for awareness. In general all other data in databases of any kind will also be susceptible to most of the errors described above. Let us consider those errors which are especially prevalent in spatial data. We have already discussed accuracy briefly when looking at sources of GIS data. The geometric (positional) accuracy that is required will depend upon the application and the scale of the phenomena. If the application is in utility network management in a city it will be important that the positions of the cables and pipes is known to say $\pm 0.1\ m$. If the aim is to study the global distribution of iron deposits, then an accuracy of $\pm 1\ km$ might be acceptable. Of course the more geometrically accurate the data needs to be the more it costs to collect it.

There is also a question of consistency between the different data sources. If one of the 'layers' of information being used is derived from a 1:250,000 map and another from a 10 m SPOT image, it is important to ensure that the two layers are registered together. A good example of this can be seen in the GRASS tutorials – try plotting the roads coverage on top of the image coverage, and look at the motorway – which is correct? In general when combining multiple maps we should try and ensure that they are co-registered – that is the relative errors between them are as small as possible. It is often very difficult to obtain absolute ground truth – differential GPS is probably best here – but we can use the well surveyed triangulation points in the UK as known 'anchors' to which the maps can be attached.

Scale is an important issue in spatial data, because different scales of information are required for different applications. We have discussed this above, to some extent. Maps at scales of 1:2,500 should have a positional accuracy of 2 m while maps at 1:250,000 scales can be expected to have accuracies of about 200 m. We have also seen how scale might affect the way we represent curves – e.g. the Douglas-Peucker algorithm.

Undertaking modelling using the geometrically imprecise data can produce unpredictable effects. An obvious example is that if our boundary data is only accurate to $\pm 200\ m$ and we are looking for individual houses in a given region we might get the answer quite wrong, which may be critical if the boundary is an electoral district. If our modelling is more complex the positional errors could accumulate, although using most geometric algorithms the growth of the errors will not be too severe. Tracing the error propagation through geometrically based modelling is complex, and very rarely done in practice. Most effort is put into producing as error free data sets as possible. Often the errors relating to attribute data, or the model itself are more gross than the geometrical errors in a high quality spatial data set.

Attribute errors can have many sources. These may be as simple as operator error, or as complex as the errors introduced into a dataset due to the use of an interpolation method to estimate a variable at an unsampled location. Where the data is stored in a raster data structure this type of error can be quantified (when using appropriate, probabilistic models) and propagated through the analysis analytically. For instance Heuvelink

(1998) is a book that deals almost exclusively with this issue. We cannot cover the details here, but it is worth noting that operations such as the computation of slope and aspect (i.e. neighbourhood based operations) from a DEM produce very complicated error distributions, given simple attribute errors on the DEM.

Since many of the models we use will be non-linear in the attributes, the distribution of the errors at the end of the analysis will often not be of a recognised form such as a Gaussian or normal distribution. If this is the case, we must either make approximations, or use Monte Carlo based methods to estimate the effects of the errors. Monte Carlo methods work by having multiple attribute fields, which are samples from the true distribution (we hope) but include the errors present in the data. It is rather like assuming there are many worlds, and calculating the outcome of the modelling in each of these different worlds. When we put this together we can see how the errors added on to the raw data in the many worlds has affected the outcome. This will give us a measure of uncertainty. The problem here is that if our model is complex and big, we need to run it many thousands, possibly millions of times to accurately account for the errors.

We can do similar things but changing only certain key bits of data. This is called sensitivity analysis, because we will only change one thing at a time – for instance we might ask how sensitive is the analysis undertaken in the second piece of coursework to positional errors in the location of the existing supermarkets. This is often undertaken when it is suspected that a few key errors may have a significant effect on the final conclusions.

Another common problem is that of data currency. Since houses, roads, cable networks and many other objects are constantly being built, moved and removed, the data in a GIS needs to be kept current. This requires constant effort in data acquisition, which is usually handled by the mapping agencies such as the Ordnance Survey in Great Britain. Thus when you buy data from the Ordnance Survey it is generally relatively up to date, but in five years time (and a lot less in some applications) the data will be out of date. When updating spatial data, all the problems of database transactions occur and some form of version control and locking may be necessary.

For spatial data we also need to consider completeness. If you think about the second piece of coursework the data was incomplete from two (maybe more) aspects. First, the data only covered a small region – you had no idea what went on outside that small region. If you were undertaking that analysis in reality a much larger region would be needed. Secondly, the data was incomplete because not all the information you required was available. This is often the case – it might have been nice to know which supermarkets people currently shop at (if any). But this data was not available, so maybe you used a surrogate, such as assuming that people will shop at the closest supermarket. This ability to use surrogate data is a strength and a weakness of GIS. It means if we are inventive we can use existing information to infer something useful, **but** this will introduce additional errors into our model.

In general we will hope that we can trace the lineage of the data using meta-data. Meta-data should describe the creation, intended accuracy, measures of accuracy derived from other field surveys of the data, any changes made to the data. This is vital if the data is to be used in any analysis, and is something that is often overlooked in standard data structures. Another area that is often included in discussion of errors is the issue of data availability. If you are working near a military installation it may be difficult to get high resolution data. There are also issues of data format, copyright and cost. We do not treat these here.

One of the big changes that I think will have to happen to GIS in the future is that there will be built in error propagation modelling, although this raises a number of issues concerning error quantification for existing datasets and who will do this (do we trust the optimistic data providers??), how we can represent the uncertainty and how we then use this in decision making. Many people are much happier with a numerical answer, as opposed to a probability distribution. Probabilistic modelling makes decision making a more complicated process, but should produce better decisions.

Heuvelink, G. B. M., 1998. *Error propagation in environmental modelling with GIS.* London: Taylor and Francis.

12 Coordinate Systems and Map Projections

As far as our everyday experience is concerned the Earth is flat. This is because visibility is generally less than 30 *km* and the Earth is so big that we cannot really detect the fact it is a sphere. However **it is a sphere**, and when we are using data over large areas (e.g. the United States) we need to be aware that representing points as if they existed on a flat plane will produce some distortion. Most GIS assume that the coordinates of the objects within them are 2D, that is exist on the surface of the Earth.

The most general coordinate system to use would be spherical polar coordinates - that is two angles and a distance from the centre of the Earth. You will be familiar with these coordinates as latitude and longitude – giving the angle from the centre of the Earth, in terms of the south–north angle across the equator (latitude), and west–east angle from Greenwich (longitude). Typically one assumes a certain geoid (shape for the Earth) and these are sufficient to represent any point on the Earth's surface.

The Earth is not a perfect sphere, the shape being distorted due to the rotation of the Earth and the locations of the continents and to a small extent the pull of the moon. The exact measurement of the shape of the Earth is difficult, although recent advances in satellite technology mean that there are several versions of the ellipsoid which can be used. These define the amount by which the Earth is 'squashed' at the poles, as well as the radius of the Earth. The most commonly used reference ellipsoid is the WGS-84 datum, which gives the equatorial radius as 6378.137 *km* and the polar radius as 6356.752 *km*.

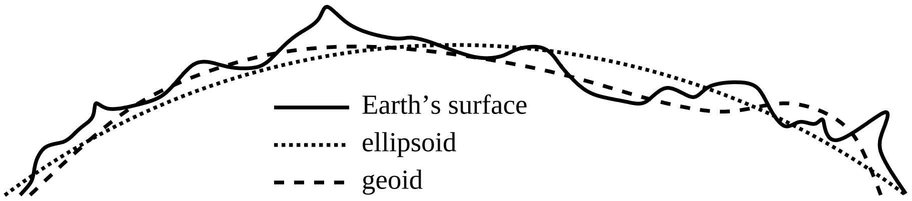

Figure 12.1: The relation between the ellipsoid, the geoid and the Earth's surface.

There is also a reference geoid – this includes the distortions caused by the continents and is illustrated in Figure 12.1. Geoids have been acquired using satellite borne altimeters and other methods. If we are going to use two different data sets then we must take care to ensure that a common geoid model is used to define the projections of both data sets. If we do not account for the uses of different geoids serious errors can result.

When we consider projections there are three properties which we might be concerned with:

- conformal – preserves angles locally;
- equal area – preserves area;
- equal distance – preserves distance.

There are no 2D projections which can preserve all these properties thus the choice of projection will depend upon the use that is to be made of the data.

12.1 Planar projections

If we project the objects on the surface of the Earth directly onto a flat plane, we get a so called azimuthal projection of half the surface of the Earth. These are so called because the azimuth (direction) from the centre point of the projections is correct, although it is wrong elsewhere, the distortion getting worse the nearer the edge one goes. Several examples are shown in Figure 12.2.

The orthographic projection is often used for showing what the Earth might look like from space. It preserves no useful properties but 'looks good'. The stereographic projection is frequently used for areas near the polar regions and is conformal, but not equal area or length. The other two projections are area and distance preserving respectively but neither are conformal. In general azimuthal projections are not widely used.

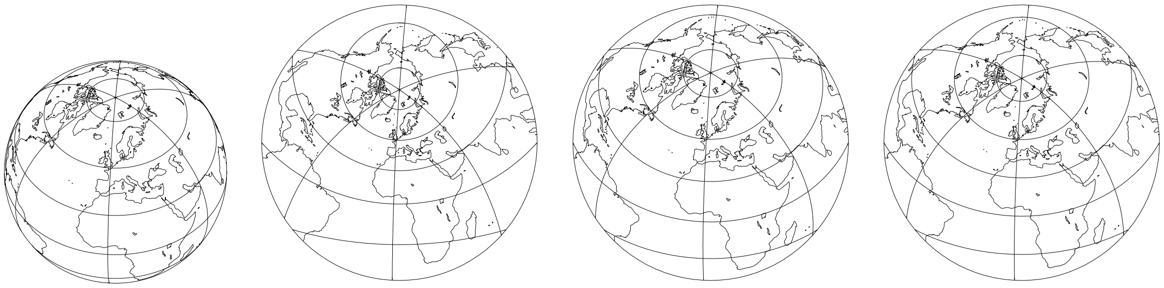

Figure 12.2: Four azimuthal projections: orthographic, stereographic, equal area and equal distance, from left to right.

12.2 Cylindrical projections

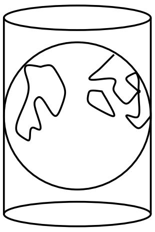

Figure 12.3: Cylindrical projection.

Most of the projections used in GIS are cylindrical. In these projections the surface of the Earth is projected onto a cylinder which surrounds the globe as shown in Figure 12.3. Typically the cylinder will only touch the sphere on a great circle (one of the longest circles around the globe from any point). The most common cylindrical projection is to allow the cylinder to wrap around the equator. This produces the well known Mercator projection, which is shown in Figure 12.4. The Mercator projection is conformal, and lines give constant bearings, so the map is good for navigation, but it is not area or length preserving.

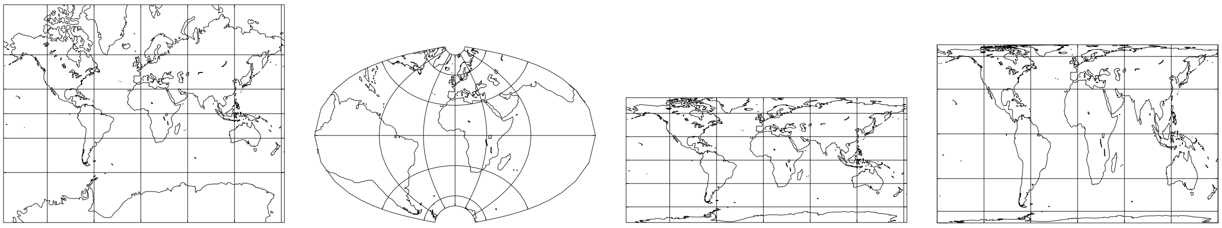

Figure 12.4: Four cylindrical projections: Mercator, transverse Mercator, equal distance and Peters equal area, from left to right.

If we take the great circle to be a line of constant longitude (i.e. the cylinder is on its side) then a transverse Mercator projection is achieved. It is conformal, but again not area or length preserving. This projection can be used to show fairly large areas of the world with big north–south extents.

If we take a Mercator projection, but ensure that distances are equal we produce the equidistant projection, which is not conformal or area preserving. We can however directly measure distances from this.

The final global projection we consider in the Peters or equal area projection which as the name suggests preserves areas, but is not conformal or length preserving.

If we are looking at smaller regions of the Earth's surface then the projection chosen becomes less noticeable (but the geoid becomes more important). Four projections are shown for smaller regions in Figure 12.5. These are all Mercator projections, with the cone touching the sphere along different great circles. The middle left figure shows an oblique Mercator, where the great circle is at an angle across the region. The right hand figure show the universal transverse Mercator (UTM) projection, which is the most commonly used projection in GIS. There are several UTM zones defined across the world (which means this is a local coordinate system) and in each of these zones the great circle that defines the projection changes.

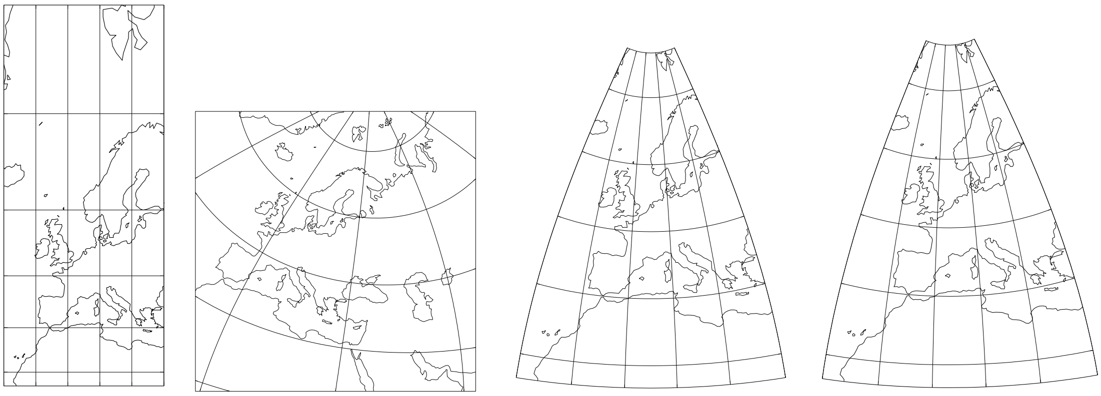

Figure 12.5: Four cylindrical projections for a smaller area: Mercator, oblique Mercator, transverse Mercator and universal transverse Mercator from left to right.

The British national grid system is a transverse Mercator projection, based on the Airy Ellipsoid. This is the most commonly used coordinate system in the UK, which gives the distance east and north of the origin to the south and west of the Scilly Isles. The data in the GRASS tutorial is displayed under this projection.

12.3 Conic projections

The other projection commonly used is to project the surface of the Earth onto a cone surrounding the Earth. This works well in specific hemispheres, with the cones generally placed so that the apex (tip) is over one of the poles.

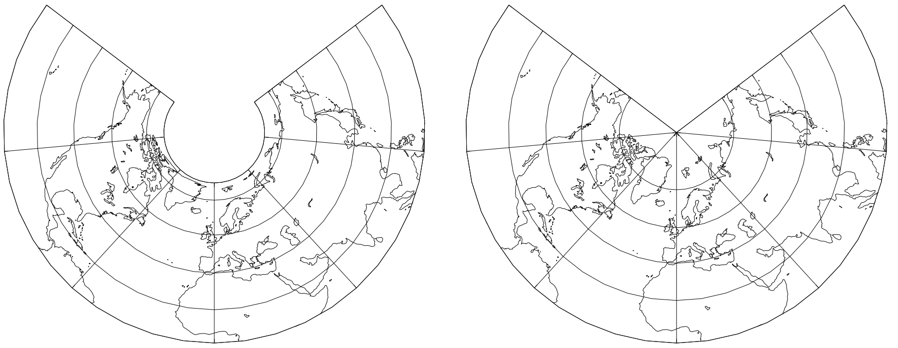

Figure 12.6: Conic projections for the Northern Hemisphere: equal area to the left and equal distance to the right.

Examples of two conic projections are shown in Figure 12.6. The Albers equal area conic projection is not conformal, while the Lambert equal distance is conformal but not equal area. Conic projections are most useful for displaying large longitudinal extents in mid-latitudes for instance.

12.4 Which projection to use?

The decision on which of the available projections to use will depend upon the region which is to be represented. In general vector data is best stored with the coordinates specified in latitude, longitude form, since this can then be projected to any desired coordinate system. However, authorities such as Birmingham City Council who only require GIS for the local area might fix their coordinate system once and for all and keep all data in the British national grid projection for example.

For raster data the grid system (projection) that is used is more important, because changing projections between different coordinate systems is computational expensive and likely to introduce additional interpolation errors.

Thus most raster GIS use UTM like system to define their grids. In the UK this is almost always based on the British national grid projection.

In the United States a slightly different approach is adopted due to the increased size of the area, thus there are local projections which change for different regions to provide local reference systems. This can make inter-boundary GIS rather complicated, thus the vector data is generally stored internally in latitude, longitude form and projected for display.

13 The Future of GIS?

In this section **I hypothesise** about the future role that GIS might have. I will use a practical example of the application of GIS in the insurance industry to illustrate the impact these advances are likely to have. In recent years all insurance companies have realised that GIS can give them a competitive edge by adding the spatial dimension to their insurance activities. In the past this was based on post-code and previous claims, tomorrow it will be based on predictive, spatial modelling of threats.

13.1 Role in business

Increasingly businesses are seeking competitive advantage through the use of information (the information economy). GIS deal specifically with spatial data and extracting spatial information from that data. Current trends, such as the spatial extensions added to the Oracle database and the extension of SQL to support some spatial and temporal queries, suggest that GIS as independent software might not last long into the future. The distinction between database and GIS is already very blurred, thus the addition of spatial analysis capabilities to standard databases could mean the need for GIS is reduced. The other side of the coin is that increasing use of spatial data will require more experts (i.e. computer scientists and geographers) to acquire, maintain and process the spatial data but possibly not within a GIS. The advanced spatial modelling capabilities of GIS seem unlikely to be provided in more standard databases, thus for more specialised modelling it seems likely that GIS will survive, possibly as extensions to standard databases.

There is also a question over whether the current trend to desktop based GIS will continue. The growth of the internet and the potential of the fat host / thin client model of computing (which is somewhat at odds with Java) suggest that a client-server approach to GIS may be the most likely model. If the server is a powerful, multi-processor machine, the client merely requires map display capabilities which will probably be provided by XML. Many of the big GIS vendors are selling internet based mapping packages. Another area that seems likely to strongly interact with GIS is that of Virtual Reality (VR) – planning a new wind turbine facility is helped by a VR simulation of the visual effect this might have. This links GIS to advances in computer graphics.

The future will, I believe, see the adoption of GIS like systems into a central role in many businesses decision support systems, and thus they will become a core part of the business information system.

In insurance companies this is already happening, although it can be seen more widely in re-insurance, where huge companies buy some of the risk under-written by insurers. It is clear that buildings and home insurance risks have a spatial dimension. The likelihood of subsidence damage to a house is a function of soil type, which varies in space. Thus if you are going to try and predict the incidence of subsidence you will need a GIS which contains information on soil types and properties as well as the spatial distribution of rainfall and evaporation. The model used will need to include details on the house and when it was built, and other factors. If you can model these down to a road, or even house scale then you can set your premiums to reflect the risk – although there are dangers here that doing this will destroy the basis for insurance.

However all the big re-insurance companies now have GIS to explore and assess their exposure to risk – and to try and minimise this – or spread the risk. There is a general trend is toward doing more things for ourselves – so it seems likely that the use of GIS will spread downwards .... possibly to home users in assessing their own risk?

13.2 Data models

Object based models seem set to take over the world (as embodied in Java). This will not be a revolution, rather an evolution. In the example of the Oracle database, which historically is based around the ER model, with tabular data structures, the spatial extensions are based on hybrid object – ER approaches. Although we have used field based approaches in the practical work, there is no reason why these fields cannot be integrated into the object approach. Everything we have seen can be translated into an object oriented approach, and I suspect that this will be the way forward.

Object oriented approaches are more natural for spatial data than layer based vector models, because the hierarchies they define correspond more closely to the way we understand space e.g. a house is on a street which is in a town which is in a country. One problem with an object oriented approach is the natural inertia of a database industry which is still largely dominated by ER approaches.

We will also see the blooming of 3D GIS, since for many applications (e.g. visibility analysis, utility management, mining, ...) information is required in at least 2.5D (this typically means a 2D data structure used with a digital elevation model) and more often in 3D (that is an additional height coordinate).

The next step will be to integrate the time dimension into GIS. Change over time is something that we are increasingly interested in modelling, whether it be for the assessment of the impact of climate change, or the changes in the demographics of a city for marketing or site selection. Integrating the temporal dimension seems to be more easily realised in the object based approach, since for the most part things stay the same, it is only small objects that change over time.

13.3 Data

This may be the area where some of the most significant changes occur. The focus of NASA in its Mission to Planet Earth, and of many other national space agencies such as the European Space Agency, has become very Earth centred. This is starting to supply vast amounts (many terabytes) of data on the surface (and other) properties of the Earth. GIS provides the framework within which this data can be stored and analysed.

It seems likely the the cost and accuracy of spatial data will come down as more becomes available and more people start using it. However the volume of data is still a very significant issue, and thus data structures and access methods remain crucial, despite faster computers. We will not only improve the accuracy of the data we might also start to do something sensible about error propagation.

In the insurance application the increased availability of data might have a very significant impact. Lidar based DEMs will allow assessment of flood potential, microwave radars might allow assessment of soil moisture to predict subsidence. The higher resolution of visible wavelength remote sensing might enable accurate characterisation of house type and location, together with the proximity of trees (which may affect subsidence).

The problem with having vast amounts of data is the extraction of information from the data becomes more computational and conceptually difficult. This will require better models and techniques to turn the data into information (although really the two are the same thing).

13.4 Models

With the increased data available we will be able to implement more sophisticated models which embody more of the factors we believe to be important in the phenomena we are trying to address. For instance when deciding on the potential sites for a new supermarket we might have some sophisticated models of competition between supermarkets to allow us to assess the effect of a promotion or a competitor placing a store nearby (what if scenarios). Not only will the models tend to become more sophisticated, I believe they will also become more principled and statistical in nature. This will allow a more careful error analysis – a measure of the uncertainty will be as important as the answer.

This is particularly so for the insurance industry where the whole thing is based on probability. Most of the events that could occur are very rare, but occasionally we get very bad floods, or storms, or dry summers. These can cause big losses, so it is important that the companies know both the mean (expected number of occurrences or expected losses) and the variance (or spread of the losses), otherwise they may be badly hit by the exceptional year.

Another area that will see growth in the coming years is that of data mining. Data mining is all about finding useful relationships, or other information in huge databases. Spatial databases often have this sort of hidden information (maybe relating expenditure on chocolate to the aspect of the persons house??!!) which might be usefully exploited by certain organisations. This may not just be for business benefit, but might be useful for social services or environmental improvement.

13.5 Summary

Overall the future of GIS is sure to be interesting if nothing else. Like much technology the evolution is occuring very rapidly, and on a global scale. This makes the exact future rather difficult to predict, however a change to a more object based view seems rather likely. The main impact, however, is likely to come from the increased data availability and the improved modelling that will result. This suggests GIS and associated technology will be a key (and probably integral) part of most information systems in the future in business and the public sector.

References

Bowyer, A. and J. Woodwark 1983. *A programmer's geometry.* London: Butterworths.

Burrough, P. A. and R. A. McDonnell 1998. *Principles of Geographic Information Systems.* Oxford: Oxford University Press.

Date, C. J. 1995. *An Introduction to Database Systems* (6th ed.). Reading, MA: Addison-Wesley.

Environmental Systems Research Institute, Inc. 1993. *Understanding GIS: The Arc/Info Method.* Harlow, Essex: Longman Scientific and Technical.

Foley, J., A. van Dam, S. K. Feiner, J. F. Hughes, and R. L. Phillips 1993. *Introduction to Computer Graphics.* Reading, Massachusetts: Addison-Wesley.

Heuvelink, G. B. M. 1998. *Error propagation in environmental modelling with GIS.* London: Taylor and Francis.

Laszlo, M. J. 1996. *Computational Geometry and Computer Graphics in C++.* London: Prentice Hall.

Mather, P. M. 1999. *Computer Processing of Remotely-Sensed Images.* Chichester: Wiley.

Rogerson, P. A. and A. S. Fotheringham 1994. GIS and spatial analysis: introduction and overview. In A. S. Fotheringham and P. A. Rogerson (Eds.), *Spatial Analysis and GIS*, pp. 1–10. London: Taylor and Francis.

Worboys, M. F. 1995. *GIS: A Computing Persepctive.* London: Taylor and Francis.