TRAINING COURSE
ON

HYDROLOGICAL MODELING AND GIS

(MAY 26 TO JUNE 06, 2014)

FOR

UNFAO & Ministry of Energy and Water, Afghanistan

LECTURE NOTE
ON

CREATION OF DATA BASE IN GIS: LAND USE/COVER AND DEM

By

SANJAY K. JAIN
NIH ROORKEE

NATIONAL INSTITUTE OF HYDROLOGY
AND
INDIAN ASSOCIATION OF HYDROLOGISTS

ROORKEE – 247 667 (UTTARAKHAND)

CREATION OF DATA BASE IN GIS: LAND USE/COVER AND DEM

Land is the most important natural resources, which embodies soil, water and associated flora and fauna involving the total ecosystem. Land cover is a fundamental parameter describing the Earth's surface. Land feature identification is the assessment of different land cover types over a certain geographic extent. These features can be agriculture area, water, forest area, snow cover area and built up area etc. These features are needed for the optimal utilization and management of land resources of the country. Land features control many hydrological processes in the water cycle e.g. infiltration, evapotranspiration, surface runoff etc. The land use maps are useful in hydrological modelling, watershed and irrigation management, water resources inventory and management etc.

Watershed models require physiographic information such as configuration of the channel network, location of drainage divides, channel length and slope, and subcatchment geometric properties. Traditionally, these parameters are obtained from maps or field surveys. Over the last two decades this information has been increasingly derived directly from digital representations of the topography. Topography plays an important role in the distribution and flux of water and energy within natural landscapes. The automated extraction of topographic parameters from DEMs is recognized as a viable alternative to traditional surveys and manual evaluation of topographic maps, particularly as the quality and coverage of DEM data increases.

LAND USE/COVER

There are many ways of land use mapping. Field workers can go into the field and actually take an inventory of species and draw a map showing what species occur in what areas. Conventional ground methods of land use mapping are labour intensive, time consuming and are done relatively infrequently. These maps soon become outdated with passage of time, particularly in a rapidly changing environment. Aerial photographs can be used to trace areas of certain species, or damage classes of trees. This method has been used extensively in the past. It is much more cost effective and faster than field workers, but it still has its limitations.

In recent years, satellite remote sensing techniques have been developed, which have proved to be of immense value for preparing accurate land use and cover maps and monitoring changes at regular intervals of time. Remote sensing techniques have ability to represent of land cover categories by means of classification process. A remote sensor records response, which is based on many characteristics of the land surface, including natural and artificial cover. Object identification through remote sensing applications is accomplished through identification of unique spectral response curves. An interpreter uses the elements of tone, texture, pattern, shape, size, shadow, site and association to derive information about land use activities, which is also the basic information about land cover. Land Cover classification through remote sensing applications involves some of the well accepted conventional classification techniques such as unsupervised (computer generated) interpretation), or supervised (human interpretation) classification.

Land use controls many hydrological processes in the water cycle e.g. infiltration, evapotranspiration, surface runoff etc. Surface cover provides roughness to the surface. This reduces the discharge, thereby increasing the infiltration. There are many classification systems exist for land use classification. The maps for the land use are prepared following these systems. The maps are useful in hydrological modelling, watershed and irrigation management, water resources inventory and management etc. Thus, they are very good and cost effective source for land use mapping. Land use maps are also available in hardcopy form at many scales. These are prepared using remotely sensed data. For preparing the maps from satellite data, suitable data are selected and processed using various classification techniques. Ground truth information is also required for preparing the maps.

INTERPRETATION OF REMOTELY SENSED DATA FOR LAND USE/ LAND COVER

Although initial efforts were made since mid seventies for application of different interpretation techniques in land use mapping, the major thrust for operational methodologies (visual, digital) came from the project on national land use/land cover mapping.

- Data loading, merging and georeferencing.
- Ground truth collection, training sets.
- Signature generation for classification.
- Demarcation of boundaries and transfer of administrative and cultural features.
- Extraction of statistics and final report.

The data are processed to extract information from them. To extract information from the remotely sensed digital data, multi spectral classification techniques are most often used. In a multi-spectral classification, DNs in various electromagnetic bands are processed using certain techniques to obtain useful information from them. Two approaches namely supervised and unsupervised classification is used. Unsupervised classification algorithms (such as ISODATA) cluster data according to several user-defined statistical parameters in an iterative fashion until either some percentage of pixels remain unchanged or a maximum number of iterations has been performed. This method of classification is most useful when no previous knowledge or ground truth data of an area is available.

The supervised classification approach is used where location of land use classes e.g. crops, urban, water, wetlands etc are known a priori. Traditionally, conventional (statistical) classification techniques such as maximum likelihood and minimum distance to mean classifiers have been used to classify remotely sensed data. These classifiers are based on the assumption that data follow some standard statistical distribution. For instance, the maximum likelihood classifier (MLC) makes the assumption that training data for each class are normally distributed. When data follow this assumption, they produce high classification accuracy, but when the data deviate from normal distribution, they give low accuracy. The statistical classifiers also require a large number of training data sets, as many as 10 to 30 times the number of bands. Adequate training sample extraction from remotely sensed data is a cumbersome and expensive process. Hence, statistical classifiers are not often the best choice for classification.

The current trend in digital image classification is to use data in a number of bands from different satellites and a number of sources such as topographic data, socio-economic data and others. Data obtained from different sources may possess different statistical properties and, therefore, the statistical classifier may not be able to utilize this data efficiently. Another problem is that the spatial data collected from different sources may have different accuracies and different scales of measurement. For example, the accuracy of a map depends upon several factors apart from its scale, such as how data were collected and processed. So maps utilized from other sources may have different accuracies. Similarly, the scales of measurement of the data can vary, from ordinal, nominal, interval or ratio scales. When such data sets are used, conventional classifiers may not provide any mechanism to incorporate this variability in the data sets. Besides this problem, conventional classifiers may become computationally inefficient when multidimensional data are used in classification. These situations often occur with remotely sensed data, and conventional classifiers may not classify these data with the desired accuracy.

LEVELS OF CLASSIFICATION USING SATELLITE DATA

The Natioanl land use/land cover classification system was designed as a recnnaissance scheme applicable in Indian environment with varying needs and perspectives. The land use/cover categories can be expanded or reduced to any degree and be made more responsive to the information the region needs. At this stage, the classification scheme is not intended to be final, but is so designed that it contains all possible categories, which might be encountered in the interpretation process. The following s a brief discussion on each of the categoric levels.

Level-I: The level classes are readily available from IRS imagery. The ground area of the minimum unit would vary depending upon the method of interpretation and scale of mapping. The level-I classification has been successfully applied using both the digital and the visual methods of data interpretation.

Level-II: The level II classification is readily achieved on IRS LISS-I FCC imagery of 1:250,000 scale. Satellite imagery of different cropping seasons of the year are required to obtain level-II information. It should be noted the reference level, knowledge of the area and skill of the remote sensing scientist/interpreter have a determining effect on the level of details and accuracy of mapping.

COMPREHENSIVE CLASSIFICATION SYSTEM DEVELOPED BY NRSC

The array of information available on land use/cover need to be grouped under a suitable classification system. The classification system should not only be flexible in its scope, definition, and nomenclature of its categories, but also be capable of incorporating information obtained from different sensor data and other sources. Such a land use classification system, based on the understanding that the remote sensing techniques can be used effectively to complement traditional surveys for an accurate inventory of the land use and land cover in the country has been developed under the National Remote Sensing Agency, Department of Space. The system is fairly compatible with those followed by most of the other government departments in the country (Table 1).

RATIO BASED APPROACHES

Ratioing is considered to be a relatively rapid means of identifying land use/cover features. Sometimes differences in brightness values from identical surface materials are caused by topographic slope and aspect, shadows or seasonal changes in sun light illumination angle and intensity. These conditions may hamper the ability of an interpreter or classification algorithm to identify correctly landuse/cover features in a remotely sensed image. Ratio transformation of the remotely sensed data can be applied to reduce the effects of such environmental conditions. Ratio may also provide unique information not available in any single band that is useful for discriminating soils and vegetation. Ratio technique is accomplished by dividing the data base brightness values in one spectral band by the data base brightness values in second spectral band for each spatially registered pixel pair. Rationing two spectral bands negates the effect of any extraneous multiplicative factors in remote sensor data that act equally in all wave bands of analysis. The ratio images have two important properties. First, strong differences in the intensities of the spectral response curves of different features may be emphasized in ratioed images. Second, ratios can suppress the topographic effects and normalized differences in irradiance when using multidate images.

NDVI approach

Numerous vegetation indices have been developed to estimate vegetation cover with the remotely sensed imagery. A vegetation index is a number that is generated by some combination of remote sensing bands. The most common spectral index used to evaluate vegetation cover is the Normalized Difference Vegetation Index (NDVI). The basic algebraic structure of a spectral index takes for form of a ratio between two spectral bands Red and near infrared (NIR). This index is calculated by subtracting Red reflectance from NIR reflectance, and dividing by the sum of the two. For instance, in vegetation areas, the NIR portion of the spectrum is reflected by leaf tissue, and the sensor records the reflectance.

The index is calculated as follows:

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)}$$

The value of NDVI index can range from -1 to +1. Vegetated surfaces tend to have positive values, bare soil may have near zero, and open water features have negative values. Now from the output we have to select different values as the limits pertaining to be water required in NDVI. In addition to NDVI values, we have to apply one algorithm, which is discussed below.

The NDVI is the measure of estimate of amount of radiation being absorbed by plants. Green and healthy vegetation reflects much less solar radiation in the visible (Channel 1) compared to near infrared (Channel 2). More importantly, when vegetation is under stress, Channel 1 value may increase and Channel 2 values may decrease. The NDVI is defined by Rouse et al. (1974) as follows:

Table 1 A classification system

Level I	Level II
1 Built-up land	1.1 Built-up land
2 Agricultural land	2.1 Crop land (i) Kharif (ii) Rabi (iii) Kharif+Rabi[a]
	2.2 Fallows[b] 2.3 Plantations[c]
3 Forest[d]	3.1 Evergreen/semi-evergreen forest 3.2 Deciduous forest 3.3 Degraded or scrub land 3.4 Forest blank 3.5 Forest plantation[e] 3.6 Mangrove
4 Wastelands	4.1 Salt affected land 4.2 Waterlogged land 4.3 Marsh/swampy land 4.4 Gullied/ravinous land 4.5 Land with or without scrub 4.6 Sandy area (coastal and desertic) 4.7 Barren rocky/stony waste/sheet rock area
5. Water bodies	5.1 River/stream 5.2 Lake/reservoir/tank/canal[f]
6. Others	6.1 Shifting cultivation 6.2 Grass land/grazing land 6.3 Snow covered/Glacial area

[a] It includes land under agricultural crops during Kharif, Rabi (both irrigated + unirrigated) and the area under double crop, during both the seasons

[b] It is that land which remains vacant with out crop during both the Kharif and the Rabi seasons

c It includes all agricultural plantations like tea, coffee, rubber, coconut, arecanut, citrus and other orchards

d It includes those areas which occur within the notfied forest boundary as shown on the Survey of India topographic maps on 1:250,000 scale. Those occuring outside the notified areas are also included under forest class, but the area estimates of the two will be shown seperately.

e It includes plantations within the notified forest boundary eg., cahew, casurina etc. Those occuring outside the notifid areas will be classfied under category 2.3.

f It includes inland fresh water lakes, salt lakes, coastal lakes and lagoons.

$$NDVI = \frac{(Ch2 - Ch1)}{(Ch2 + Ch1)}$$

`Where, Ch1: Radiation measured in channel 1 (Visible)
Ch2: Radiation measured in channel 2 (NIR)

The healthy and dense vegetation show a large NDVI. In contrast cloud, water, and snow have larger visible reflectance than those of NIR, thus those features yield negative index yields. Rock and bare soil covered areas have similar reflectance in the visible/near infrared band and result in vegetation indices near zero. Because of these properties NDVI has become primary tool for mapping change in vegetation cover and analysis of impact of environmental phenomenon. The NDVI can be used not only for accurate description of continental land covers, vegetation classification and vegetation phonology, but it is also effective for monitoring rainfall and drought, estimating net primary production of vegetation, crop growth condition and crop yields, detecting weather impacts and other events important for agriculture, ecology and economics.

ACCURACY ASSESSMENT

Classifications derived from remotely sensed images are subject to error and uncertainty. In classifying an image, the spectral response of a pixel, representing a fixed area on the ground defined by the resolution of the sensor, is used to assign it to one of a number of classes using various classification techniques.

Accuracy assessment is an important final step of the classification process. Accuracy is defined in terms of misclassifications, where a pixel is assigned to the wrong class. Misclassifications are usually presented in the form of a matrix, which is referred to as a confusion or error matrix. The error matrix can be used to generate various statistics that characterize the accuracy of a classification technique. The goal is to quantitatively determine how effectively pixels were groups into the correct land cover classes. The procedure is relatively simple. Pixels are randomly selected throughout the image using a specified random distribution method. Then the analyst use the original image along with ancillary information such as aerial photographs or direct field observation to determine the true land cover represented by each random pixel. This ground truth is compared with the classification map. If the ground truth and classification match, then the classification of that pixel is accurate. Given that enough random pixels are checked, the percentage of accurate pixels gives a fairly good estimate of the accuracy of the whole map. A more rigorous and complicated estimate of accuracy is given by the kappa statistics, which are obtained by a statistical formula that utilizes information in an error matrix. An error matrix is simply an array of numbers indicating how many pixels were associated with each class both in terms of the classification and the ground truth. Statistics that can be generated from the error matrix include errors of omission (producer's error) and errors of commission (user's error). These are based on individual classes, dividing the number of pixels that are incorrectly classified by either the column or row totals, respectively.

DIGITAL ELEVATION MODEL (DEM)

One of the capabilities of GIS is the description of the topography of a region. Techniques used in the computer description of topography are called as Digital Elevation Model (DEM). DEM's are arrays of numbers that represent the spatial distribution of terrain altitudes. Main data sources for DEM's are ground surveys, existing topographic maps, photogrammetric stereomodels and surveys done by radar or laser altimeters carried in aircrafts and spacecrafts.

Various data structures are in use for DEMs, each with their own merits and shortcomings. There is no structure which satisfies all requirements; much will depend on the purpose and also on the computer facilities available. The basic structures are the line model, the triangulated irregular network and the grid network. The classical form of representing topography is the contour line mapping. The contours can be represented digitally as a set of point to point paths (vectors) of a common elevation. The line model describes the elevation of terrain by contours (stored as Digital Line Graphs i.e. the x,y coordinate pairs along each contour of specified elevation). Typical GIS operations based on the line model are carried out by overlaying the contours on to thematic maps or remotely sensed classifications.

An alternative approach to producing DEM's relies upon determination of significant peaks and valley points in the terrain, which is then represented by a collection of irregularly spaced points connected by lines. The TIN model splits up the true surface into triangular elementary planes. The terrain surface is sampled by points (nodes) that are located at positions which capture the terrain characteristics. The grid based methods may involve the use of a regularly spaced triangular, square or regular angular grid. The element area is the cell bounded by three or four adjacent grid points, depending upon the method. The raster based GISs use the square grid networks. The advantage of the regular grid method is the simplicity of the data storage usually as sequential z coordinate along the x (or y) direction, with a specified starting point and grid spacing.

DEM can be put in use in variety of applications. The most common products derived from DEM relate elevation, slope, aspect, convexity/concavity of terrain etc. Considerable research has been carried out in the field of drainage network extraction and watershed boundary delineation by hydrologists and geologists. Both the slope and the aspect of land are important determinants for ecological site classification and natural resource management and use. For example soil classifications often use landscape categories such as 0-5% slope, 5 to 15% slope, and 15 to 45% slope because it influences the pattern of native vegetation. In a similar fashion land with northern aspects produces different plant communities than land with southern aspect.

The slope of the surface may be calculated in a number of different ways, and depending on which way it is done, pixels will have different values. After obtaining an interpolated raster DEM; other terrain properties can be extracted using filtering techniques. First, gradients in X and Y direction are derived, transforming the scalar into vector field. Each pixel becomes a vector with two components the partial derivatives in X and Y direction. After creating gradients say, GX and GY, the absolute gradients can be used as a slope map. Slope is a widely used topographic measurement that describes the nature of the land surface and, more importantly, influences the flow rates of water.

A drainage network can be extracted from a DEM with an arbitrary drainage density or resolution. The characteristics of the extracted network depend on the definition of channel sources on the digital land surface topography. Once the channel sources are defined, the essential topology and morphometric characteristics of the drainage network are implicitly defined because of their close dependence on channel source definition. Thus, the proper identification of channel sources is critical for extraction of a representative drainage network from DEMs.

SOI Derived DEM

The most common method of acquiring elevation data in a digital raster format is to digitize the contours from a topographic map and apply an interpolation method to transform the contour data into a DEM. Although the quality of DEM essentially depends on the topographic map used, the interpolation algorithm used for DEM resampling has also a significant influence. The interpolation algorithm used for DEM resampling may be designed according to different requirements. Many technique aim at minimising the RMS-error; this criterion is the one most commonly applied in DEM quality assessment. On the contrary, some other techniques tend to preserve the terrain texture.

CARTOSAT DEM

The CARTOSAT-1 spacecraft launched by the Indian Space Research Organisation in May 2005 is dedicated to stereo viewing for large-scale mapping and terrain modeling applications. It is configured with two panchromatic cameras, AFT (Afterward looking) and FORE (Foreword looking) with a spatial resolution of 2.5 m, which facilitates along-track stereo vision of the imaging scene. It covers a swath of ≈30 km with a base-to-height ratio of 0.62. The time difference between the acquisitions of the same scene by two cameras is about 52 sec. The principle of binocular vision is used to produce a 3-D view of the imaging scene. The two cameras are mounted on the satellite in such a way that nearsimultaneous imaging of the same scene from two different angles +26° (FORE) and −5° (AFT) along the track with respect to nadir is possible. This provides stereoscopic image pairs in the same pass and facilitates the generation of accurate 3-D maps.

SRTM DEM

The Shuttle Radar Topography Mission (SRTM) is an international project led by the U.S. National Geospatial-Intelligence Agency (NGA), U.S. National Aeronautics and Space Administration (NASA), the Italian Space Agency (ASI) and the German Aerospace Center (DLR). SRTM obtained elevation data on a near-global scale to generate the most complete high resolution digital topographic database of Earth, including three resolution products of 1 km and 90 m resolutions for the world, and a 30 m resolution for the US (USGS, 2004). The elevation data used in this study is the 90 m resolution (3-arc SRTM), which consists of a specially modified radar system that flew onboard the Space Shuttle Endeavour during an 11-day mission in February of 2000. All SRTM data are freely available at: http://seamless.usgs.gov/Website/Seamless/. The SRTM-DEM was downloaded from the USGS ftp site. These data are presently supplied free of cost for scientific study. The data were supplied in GeoTIFF format. These DEMs were exported to the format as unsigned 32-bit data in ERDAS Imagine platform. Sometimes variation in pixel intensity (digital numbers) may be caused by differing sensitivities or malfunctioning of the detectors, topographic effects, or/and atmospheric effects. To correct such variations, radiometric calibration were carried out. The SRTM-DEM was already projected in Geographic lat/long and WGS84 datum.

ASTER DEM

On 29 June 2009, the Global Digital Elevation Model (GDEM) was released to the public. A joint operation between NASA and Japan's Ministry of Economy, Trade and Industry (METI), the Global Digital Elevation Model is the most complete mapping of the earth ever made, covering 99% of its surface. The previous most comprehensive map, NASA's Shuttle Radar Topography Mission, covered approximately 80% of the Earth's surface, with a global resolution of 90 meters, and a resolution of 30 meters over the USA. The GDEM covers the planet from 83 degrees North to 83 degrees South (surpassing SRTM's coverage of 56 °S to 60 °N), becoming the first earth mapping system that provides comprehensive coverage of the Polar Regions. It was created by compiling 1.3 million VNIR images taken by ASTER using single-pass stereoscopic correlation techniques, with terrain elevation measurements taken globally at 30 meter (98 ft) intervals. The GDEM is produced with 30 meter postings, and is formatted in 1 x 1 degree tiles as GeoTIFF files. The GDEM is referenced to the WGS84/EGM96 geoid.

ASTER DEM data are freely available at: ***http://gdem.ersdac.jspacesystems.or.jp***. These data are presently supplied free of cost for scientific study. The data supplied in Geo-TIFF format. Anyone can easily use the ASTER GDEM to display a bird's-eye-view map or run a flight simulation, and this should realize visually sophisticated maps. By utilizing the ASTER GDEM as a platform, institutions specialized in disaster monitoring, hydrology, energy, environmental monitoring etc. can perform more advanced analysis.

ASTER GDEM tiles can be downloaded electronically from ERSDAC. By clicking on the above link.

****