TRAINING COURSE
ON

APPLICATIONS OF REMOTE SENSING AND GIS IN WATER RESOURCES MANAGEMENT

ORGANISED UNDER
HYDROLOGY PROJECT PHASE II

AT
NATIONAL INSTITUTE OF HYDROLOGY, ROORKEE

OCTOBER 20-24, 2008

LECTURE NOTE
ON

APPLICATIONS OF REMOTE SENSING AND ANN's IN LAND FEATURE IDENTIFICATION

By

ARCHANA SARKAR
NIH ROORKEE

NATIONAL INSTITUTE OF HYDROLOGY
JALVIGYAN BHAWAN
ROORKEE – 247 667 (UTTARAKHAND)

APPLICATION OF REMOTE SENSING AND ANNs IN LAND FEATURE IDENTIFICATION

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 modeling, watershed and irrigation management, water resources inventory and management etc.

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 These maps soon become outdated with passage of time, particularly in a rapidly changing environment.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. The accuracy of these techniques is subjected to various human, spatial, temporal or radiometric factors. Alternative processing techniques includes expert system, hyperspectral classification, knowledge based classification and ANN

based classification. Beside this in some cases, normalized ratio as the classification input instead of the original spectral bands gives better results.

In this lecture, criteria for land use/ land cover classification, different approaches of remote sensing and advanced techniques such as ANN for identification of land features have been discussed.

CRITERIA FOR LAND FEATURE IDENTIFICATION AND LAND USE/LAND COVER CLASSIFICATION

In order to develop a land use/land cover classification system for the land use conditions prevalent in India, it is essential to consider certain criteria and limitations of satellite data and that of the study areas. The generation of remotely sensed data/images by various types of sensors flown aboard different platforms at varying heights about the terrain and at different times of the day and the year, do not lead to a simple classification system. In fact, many researchers believe that no single classification could be used with all types of imagery and all scales. To date, the most successful attempt in developing a general-purpose classification scheme compatible with remote sensing data has been attempted by Anderson et al., which is also referred to as USGS classification scheme. This system was designed according to the following criteria:

- The minimum level of interpretation accuracy using remote sensor data should be atleast 85 percent
- The accuracy of interpretation for the several categories should be about equal.
- Repetitive results should be obtainable from one interpreter to another and from one time of sensing to other.
- The system should be applicable over extensive areas.
- The system should be suitable for use with remote sensor data obtained at different times of the year.
- Categories should be divisible into more detailed subcategories that can be obtained from large-scale imagery.
- Aggregation of categories must be possible
- Comparison with future land use and land cover data should be possible.
- Multiple uses of land should be recognized when possible.

Other classification schemes available for use with remotely sensed data are basically modifications of the above classification scheme. Four levels of classification are suggested in this scheme, each level being useful at various information/decision-making levels. Level-1 information can be used at inter-state and state-level planning; level-II, at state-level to regional level planning, level-III, at regional to local level of planning; whereas level-IV information can be used at local or micro level planning.

The National Land Use/Land Cover Classification system was designed as a reconnaissance scheme applicable in Indian environment with varying needs and perspectives. The land use/land 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 is a brief discussion on each of the categoric levels.

Level-I: The level-I classes are readily available from IRS imagery. The ground area of the minimum mapping unit would vary depending upon the method of interpretation and scale of mapping. U sing visual interpretation, the minimum mapping unit (3 mm x 3 mm) on 1: 1 million scale is equivalent to 900 ha (9 sq. km) on the ground (Table I). 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 scales. The minimum mapping unit, on this scale, represents 56.25 ha on the ground. Satellite imagery of different cropping seasons of the year are required to obtain level-II - information. It should be noted that 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. At this stage, the classification scheme is not intended to be final and likely to undergo further changes. However, it contains alL possible categories, which might be encountered in the interpretation process.

The array of information available on land use/land cover need to be grouped under a suitable classification system. National Land Use/Land Cover Classification System for India has been developed under the National Remote Sensing Agency, Department of Space in consideration of the views of the several user departments including the Planning Commission

of India. The system is fairly compatible with those followed by most of the other government departments in the country (Table 1).

Table 1: Land use/land covers classification system

Level-I	Level-II
1. Built-up land	1.1 Built-up land
2. Agriculutral land	2.1- Crop Land
	i) Kharif
	ii) Rabi
3. Forest[d]	iii) Kharif +Rabi[a]
	2.2 Fallows[b]
	2.3 Plantations[c]
	3.1 Evergreen/semi-evergreen forest
4. Wastelands	3.2 Deciduous forest
	3.3 Degraded or scrub land
	3.4 Forest blank
	3.5 Forest plantation[e]
5. Water bodies	3.6 Mangrove
6. Others	4.1 Salt affected land
	4.2 Waterlogged land
	4.3 Marshy/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
	5.1 River/stream
	5.2 Lake/reservoir/tank/canal[f]
	6.1 Shifting cultivation
	6.2 Grassland/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 without 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 notified forest boundary as shown on the Survey of India topographic maps on I :250,000 scale. Those occurring outside the notified areas are also included under forest

class, but the area estimates of the two will be shown separately.

e It includes plantations within the notified forest boundary eg., cahew, casuarina, eucalyptus, etc. Those occurring out- side the notified areas will be classified under category 2.3. The area estimates of the two will be shown separately.

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

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 electromagnatic 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.

The limitations of the conventional classifiers have compelled researchers either to modify the existing techniques or to search for new techniques. Some of the new generation techniques used to classify remotely sensed data are ratio based approaches, fuzzy techniques, knowledge-based techniques and artificial neural networks (ANN). Unlike conventional classifiers, these techniques are non-parametric in nature, as these do not make assumptions about the distribution of data. Moreover, these techniques have been found to be efficient in classifying data from different sources and sensors.

The ANN approach, in particular, has many advantages over conventional ones and has produced quality classified products from remote sensing data. These techniques can handle data at any scale of measurement. Moreover, ANN can deal with data showing inconsistencies, noise and missing data.

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 (Jensen, 1996). 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 (Avery and Barlin, 1992). Ratio may also provide unique information not available in any single band that is useful for discriminating soils and vegetation (Satterwhite, 1984). 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 (Mather, 1987). 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 (Singh, 1989).

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:

$$NDVI = \frac{(Ch2 - Ch1)}{(Ch2 + Ch1)} \qquad (3.13)$$

`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 (Tucker et al. 1982, Tarpley et al. 1984, Justice et al. 1985) 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 (Kogan 1987a, Dobrowska-Zielinskan et al. 2002).

NDWI Approach

McFeeters (1996) developed an index similar to the NDVI, which is called the NDWI. This stands for the Normalized Difference Water Index. Any instrument having a green band

and a near infrared band can apply this index. The NDWI was derived using principles similar to those that were used to derive the NDVI. The NDWI is calculated as follows:

$$NDWI = \frac{(GREEN - NIR)}{(GREEN + NIR)} \tag{2.2}$$

where GREEN is a band that encompasses reflected green light and NIR represents reflected near-infrared radiation. The selection of these wavelengths was done to: (1) maximize the typical reflectance of water features by using green light wavelengths; (2) minimize the low reflectance of NIR by water features; and (3) take advantage of the high reflectance of NIR by terrestrial vegetation and soil features. When equation (2.2) is used to process a multi-spectral satellite image that contains a reflected visible green band and an NIR band, water features have positive values; while soil and terrestrial vegetation features have zero or negative values, owing to their typically higher reflectance of NIR than green light. Image processing software can easily be configured to delete negative values. This effectively eliminates the terrestrial vegetation and soil information and retains the open water information for analysis. The range of NDWI is then from zero to one. Multiplying equation (2.2) by a scale factor (e.g. 255) enhances the resultant image for visual interpretation.

NDSI Approach

This approach is used for identification of snow cover in hilly areas. Snow exhibits high reflectance in visible band and strong absorption in SWIR band. Cloud on the other hand shows uniform reflectance due to non-selective scattering. The Normalized Difference Snow Index (NDSI) uses the above spectral characteristics of snow and is based on the concept of Normalized Difference Vegetation Index (NDVI) used in vegetation mapping from remote sensing data (Dozier, 1989: Hall et al., 1995, Gupta et al., 2005). The NDVI is defined as the difference of reflectance observed in a visible band and the short-wave infrared band divided by the sum of the two reflectance (Gupta et al., 2005). The equation is given below:

$$\text{NDSI} = \frac{\text{Visible Band} - \text{SWIR Band}}{\text{Visible Band} + \text{SWIR Band}}$$

NDSI maps were prepared from NOAA-AVHRR images. The NDSI map was further classified into two classes: (a) snow and (b) snow free area based on threshold value of 0.4 (Dozier,

1989). This type of classification provided an advantage of SCA estimation under mountain shadow condition and discrimination between snow and cloud.

EXPERT SYSTEM

An Expert System classification model can be constructed using the Knowledge Engineer module of the ERDAS Imagine image-processing software. The expert classification software provides a rules-based approach to multispectral image classification, post-classification refinement, and GIS modeling. In essence, an expert classification system is a hierarchy of rules, or a decision tree, that describes the conditions under which a set of low level constituent information gets abstracted into a set of high level informational classes. The constituent information consists of user-defined variables and includes raster imagery, vector coverages, spatial models, external programs, and simple scalars.

A rule is a conditional statement, or list of conditional statements, about the variable's data values and/or attributes that determine an informational component or hypothesis. Multiple rules and hypotheses can be linked together into a hierarchy that ultimately describes a final set of target informational classes or terminal hypotheses. Confidence values associated with each condition are also combined to provide a confidence image corresponding to the final output classified image.

The Expert Classifier is composed of two parts: the Knowledge Engineer and the Knowledge Classifier. The Knowledge Engineer provides the interface for an expert with first-hand knowledge of the data and the application to identify the variables, rules, and output classes of interest and create the hierarchical decision tree. The Knowledge Classifier provides an interface for a non-expert to apply the knowledge base and create the output classification. An Expert Classifier classifies based on ancillary data. This is done using a hierarchy of rules. The Knowledge Engineer is a part of the Expert Classifier that provides the interface for an expert with first-hand knowledge of the data and the application to identify the variables, rules, and output classes of interest and create the hierarchical decision tree (ERDAS, 1999).

Knowledge Engineer

The Knowledge Engineer is a separate application program invoked from the Classification dialog. The interface follows a document editing paradigm where the document is the decision tree. The File menu provides access to knowledge base file Open, Save, and

Print functions. The Edit menu provides additional access to functions more readily available from the toolbar. However, it provides a couple of powerful globally acting functions. The Delete All Disabled option allows removing all currently disabled branches from the knowledgebase. The Clear All Work Files... option allows removing all work files associated with the current session.

After a knowledge base is opened or created, a graphical representation of the decision tree is presented in the editing window as shown in Figure 4.1. The Rule Properties dialog allows each of the variables, or conditions of the rule to be defined. A name string identifies the rule.
A method of determining Confidence for the rule's conditions must be selected. If the compute from conditions option is selected, the confidence values for each rule are computed by the Knowledge Classifier from the confidence of each condition associated with the rule. If the **Specify** option is selected confidence value to be assigned to the rule has to be entered.
The list of conditions is presented in a cell array where each condition is defined by a Variable, a relation, a value, and a confidence.
A variable, relation, or value in the list of conditions can be changed by selecting its cell in the Cell array and choosing from the popup list. The confidence entries can be edited by clicking in the appropriate cell in the Cell array.

ANNS FOR MULTISPECTRAL LAND USE/LAND COVER CLASSIFICATION

A new mathematical model that has emerged recently, and which has made a great impact in the scientific community is the artificial neural networks (ANNs). ANN has attracted increasing attention from researchers in many fields during the last decade, resulting in studies aiming to solve a wide range of problems. ANN has been proved to be more robust compared to conventional statistical classifiers in recognising patterns from noisy and complex data and in estimating their nonlinear relationships. In short, it is known to be good at learning the internal representation of data in any form. Artificial neural networks are heuristic algorithms, in that they can learn from experience via samples and can subsequently be applied to recognise new data. These systems are intended, in an extremely simple way, to imitate the behaviour of the network of neurons in the human brain. The primary aim of the ANNs is to improve the performance of computer recognition processes by simulating the superior characteristics of the human brain. Based on the biological theory of the human brain, artificial

neural networks are models that attempt to parallel and simulate the functionality and decision-making processes of the human brain.

The power of artificial neural network techniques rests in their unique advantages that may be listed as follows:

- they are non-parametric,
- they have arbitrary decision boundary capabilities,
- it is easy to incorporate different types of data and input structures,
- they can handle data at any scale of measurement,
- they yield fuzzy output values that can enhance classification,
- they can generalise better, especially in the use of multiple images,
- they are tolerant to noise.

Of the advantages of ANN techniques, the most important one may be their nonparametric nature. In other words, there is no underlying assumption about the frequency distribution of the data. They learn the characteristics of the training data (or the internal structure of these data), typically in an iterative way, so they may be called data-dependent techniques. It is also worth noting that artificial neural networks can give considerably better results for small training datasets compared to conventional statistical classifiers.

STRUCTURE OF ANNs

A neural network may be considered as being composed of a relatively large number of simple processing units (nodes) that work in parallel to classify input data into output classes. The processing units are generally organized into layers, each unit in a layer being connected to every other unit in the following layer. This is known as a feed forward multilayer network.

The architecture of a typical multi-layer neural network is shown in the Figure 1. It consists of an input layer, a hidden layer and an output layer. The input layer is passive and merely receives the data (for example, the multi spectral remote sensing data). Consequently the units in the input layer equal the number of variables (for example, number of wavebands) to be used in the classification. Unlike the input layer, both the hidden and the output layers actively process the data. The output layer, as its name suggests, produces the neural network's results. In land-cover applications, the result denotes the various land-cover classes. Thus, the number of units in the output layer corresponds to the number of land-cover classes of interest. Hence, the number of units in the input and the output layers is typically fixed by the

application design. Introducing the hidden layer between the input and output layers increases the network's ability to model complex functions.

The input to the network is received by the neurons in the input layer. The data passing through the connections from one neuron to another are manipulated by weights, which control the strength of a passing signal. When these weights are modified, the data transferred through the network changes and the network output alters. The neurons in a layer share the same input and output connections, but do not interconnect among themselves. Each layer performs specific functions. All the nodes within a layer act synchronously, meaning at a point of time, they will be at the same stage of processing. The activation levels of the hidden nodes are transmitted across connections with the nodes in the output layer. The level of activity generated at the output node(s) is the network's solution to the problem presented at the input nodes.

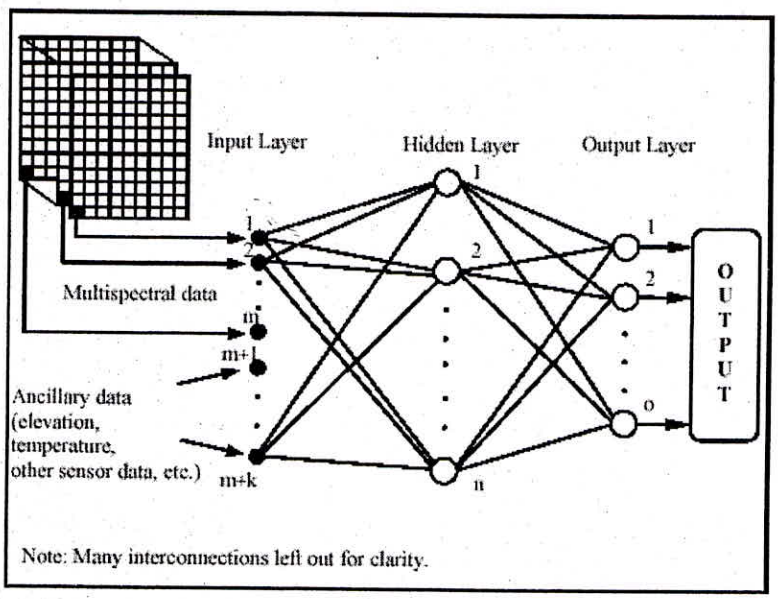

Figure 1: Image Classification using ANNs

A typical neuron is shown in Figure 2. Depending on the layer in which the neuron is located, its input may be an input casual variable or outputs of neurons of previous layer. Every neuron receives signals from each neuron in the previous layer. At each neuron in hidden and output layer, every input is multiplied by its weight; the product is summed (NET) and passed through a transfer function to produce its result.

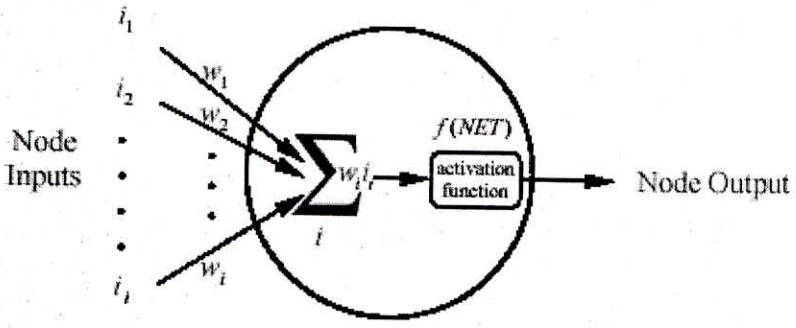

Fig. 2: A typical processing node of an ANN

The most common form of the activation function is a sigmoid function, defined as:

$$F(NET) = 1/1 - e^{-NET} \quad (1)$$

And accordingly

$$Output = f(NET) \quad (2)$$

Where NET is the sum of the weighted inputs to the processing unit and may be expressed as:

$$NET = \sum_{i=1}^{n} w_i i_i \quad (3)$$

Where i_i is the magnitude of the i^{th} input and w_i is the weight of the interconnected path.

The determination of appropriate weights is referred to as "learning" or "training" which is briefly described in subsequent sections.

DESIGNING AN ANN

The ANN design consists of finding a simple architecture, which yields the desired performance. There is no analytical solution to determine an optimal ANN architecture and a unique solution cannot be guaranteed. Since the numbers of input and output nodes are problem dependent, the designer has to determine the number of hidden layers and the number of nodes in each hidden layer. According to Hsu et al. (1995), three-layer feed forward ANNs can be used to model real world functional relationships that may be of unknown or poorly defined form and complexity. Therefore, in such networks, the problem reduces to finding the optimal number of nodes in the hidden layer. Generally, a trial-and-error approach is used. This number should be chosen carefully since the performance of a network critically depends on it- a network with too few nodes will give poor results, while it will overfit the training data if too many nodes are present.

Among the automatic algorithms, there are two major variants. The pruning algorithms, as the name suggests, begin with a large network and systematically remove the nodes whose contribution is minimal. The other variant, the growing algorithms, begin with a small network and add nodes till the improvement in performance is insignificant.

TRAINING OF ANN

The knowledge of an ANN is contained in its weights. The objective of training/learning is to determine the set of weights and thresholds so that the ANN gives desired output. The process is similar to calibration of a watershed model. In general, it is assumed that an ANN does not have any prior knowledge about the problem before it is trained. At the beginning of training, weights are initialized either with a set of random values or based on some previous experience. When the network weights are altered, the data transfer through the ANN changes and the network performance alters. The learning algorithm adjusts the weights such that for an input signal, the ANN output is close to the desired output. Several learning examples are presented to the network, each contributing to the optimization of weights. The results of an ANN keep on improving as more and more data are made available to it because it has a built-in mechanism of growing 'wiser' with 'experience'. This adjustment can be continued recursively until a weight space is found, which results in the smallest overall prediction error. At this stage when an ANN has learned enough examples, it is considered trained. The final weight matrix of a successfully trained neural network represents its knowledge about the problem. Note that the aim of learning is to get a network that generalizes the relationship between input and output rather than the one that memorizes it.

There are primarily two basic learning strategies for ANNs -supervised and unsupervised, as in a conventional statistical classification. The supervised training algorithm uses a large number of inputs and outputs patterns. The inputs are cause variables of a system and the outputs are the effect variables. This training procedure involves the iterative adjustment and optimization of connection weights and threshold values for each of nodes. The primary goal of training is to minimize the error function by searching for a set of connection strengths and threshold values that cause the ANN to produce outputs that are equal or close to targets. In contrast, an unsupervised training algorithm uses only an input data set. The ANN adapts its connection weights to cluster input patterns into classes with similar properties.

Generally, a back propagation algorithm is applied, which is a supervised algorithm that has been widely used in neural network classification applications. In supervised training, the available data set is generally partitioned into two parts: training set and validation set. The training data set should contain sufficient input and output pairs and the entire range of inputs should be included so that the network can adequately learn the underlying relationship between input and output variables. The learning algorithm iteratively minimizes an error function over the network outputs and a set of target outputs, taken from the training data set. The process continues until the error value converges to minima. Conventionally, the error function is given as:

$$E = 0.5\Sigma (T_i - O_i)^2 \qquad (4)$$

where T_i is the target output vector and O_i is the network output.

After an optimal amount of training has been achieved, the errors for the training set continue to decrease, but those associated with the cross training data set begin to rise. This is an indication that further training will likely result in the network overfitting the training data. Training is stopped at this stage and the current set of weights and thresholds are assumed to be optimal.

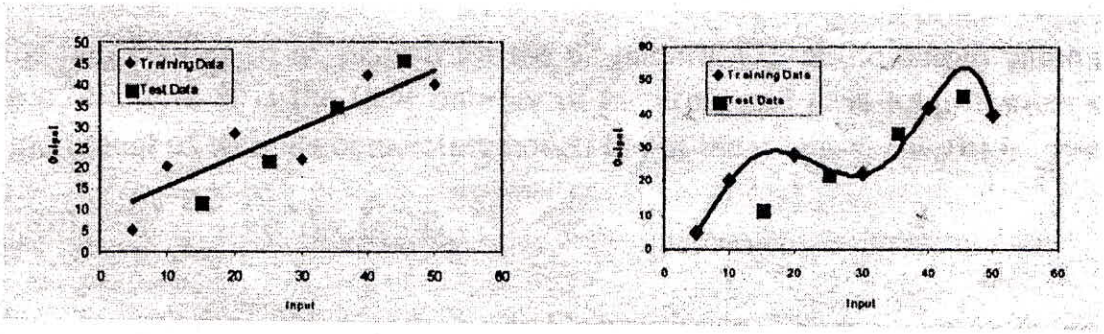

Figure 3: A well-trained ANN (left) and an overfit model (right).

After classification has been performed, it is evaluated for its accuracy. The error matrix approach may be utilized to derive a range of accuracy measures such as overall accuracy or kappa coefficient, which has been discussed in the following section.

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 (Jensen, 1996). 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.

REFERENCES

1. ASCE. Task Committee on Application of Artificial Neural Networks in Hydrology. (2000), "Artificial Neural Networks in Hydrology. II: Hydrologic Applications", Journal of Hydrologic Engineering, ASCE, Vol.5 (2), pp.124-137.
2. Avery, T.E., and Berlin, G.L. (1992). Fundamentals of Remote Sensing and Airphoto Interpretation. Macmillan, NY: 433-436
3. Foody, G. M. and Arora, M.A. (1997), "An Evaluation of some factors affecting the accuracy of classification by an artificial neural network", Int. J. Remote Sensing, vol.18(4), pp 799-810.
4. Hsu, K., Gupta, H.V., & Sorooshian, S. (1995), "Artificial Neural Network Modelling of the Rainfall-Runoff process", Water Resources Research, vol. 31(10), pp.2517-2530.

5. Jensen, J.R., (1996). Introductory Image Processing: a Remote Sensing Perspective. Second edition, Upper Saddle River, NJ: Prentice Hall :316p.

6. Kavzoglu, T and. Mather, P.M (1999), "Pruning Artificial Neural Networks: An example Using Land Cover Classification of Multi-Sensor Images", Journal of Remote Sensing, vol. 20(4), pp. 2787-2803

7. Mather, P.M. (1987), "Computer processing of remotely sensed Images", Chichester: Wiley.

8. Satterwhite, M.B. (1984). Discriminating Vegetation and Soils Using Landsat MSS and Thematic Mapper bands and Band Ratios. Proc. American society of Photogrammetry, No.2: 479-485.

9. Schowengerdt R. A. (1997), "Remote Sensing: Models and Methods for Image Processing", Academic Press, San Diego, NY.

10. Singh, A. (1989). Digital Change Detection techniques using Remotely Sensed Data. International J. of Remote Sensing, 10(6): pp989-1003.

** *** **