Chapter 3

Emerging Techniques for Data Acquisition and Systems Modeling

The objectives of this chapter are:

- to introduce fundamental concepts of data collection through remote sensing and its analysis,
- to explain the concepts of a GIS and its analytical capabilities,
- to introduce the ANN theory and describe its applications, and
- to give an exposition of the expert systems.

This chapter discusses four tools which are now extensively used in analysis of hydrosystems. These tools are labeled as emerging techniques since their application in water resources area began in right earnest only during the last two decades or so but during this short period, they have occupied an important place in the tool kit of an analyst. As these tools are still being evolved, there are some unresolved issues regarding their application. The topics that have been included in this chapter are: Remote Sensing, Geographic Information Systems, Artificial Neural Networks, and Expert Systems. Since each of these topics is subject matter of a complete book in itself, this chapter will only give a broad overview of the techniques and their applications to water resources problems.

3.1 REMOTE SENSING

The land phase of the hydrologic cycle is influenced and controlled by surface and near surface features of earth which have inherent spatial variability. The inputs to this system, e.g., radiation, precipitation, also have spatial attributes. In the absence of a reliable technique to measure spatial features, point values are commonly used in the analysis. The need for measurement of areal properties was thus summarized by Klemes (1986): "It also seems that the search for new measurement methods that would yield areal distribution, or at least reliable areal totals or averages of hydrologic variables such as precipitation, evapotranspiration, and soil moisture would be a much better investment for hydrology."

Remote sensing is a technique which has the potential to realize this necessity. In fact, it has already achieved so in many instances.

The term Remote Sensing (RS) implies the acquisition of information about an object without establishing any physical contact between the object and the sensing device. Although it appears to be a new technique, this way of obtaining information about an object is quite old. Photography, which has been in use for a long time, is a technique based on the same principle. Of late, the term RS technique is chiefly used to denote the acquisition and analysis of satellite data. This is a powerful technique for exploration, mapping, and management of the earth resources. The main advantage of the RS technology is that it provides a broad perspective over a large area. One can "see" beyond visible electro-magnetic (EM) radiation band, and data of inaccessible areas can be obtained just as easily. Remote sensing techniques have extended the scope of utilization of the EM spectrum to almost its entire range. Depending on the sensor, it is also possible to infer the characteristics of a top thin layer of the earth's surface. The interaction of EM radiation with an object can reveal a tremendous amount of information about the object: What is it? Where is it? What are some of its physical properties? What are its spatial relationships with the surroundings?

Electromagnetic energy can be generated by changes in the energy levels of electrons, acceleration of electrical charges, decay of radioactive substances, and the thermal motion of atoms and molecules. Nuclear reactions within the sun produce a full spectrum of EM radiation which is transmitted through space without major changes in its character until it reaches the atmosphere. Visible light, radio waves, thermal, ultraviolet and X-rays are the familiar forms of EM radiation and propagate in accordance with the basic wave theory. The electromagnetic energy travels in a harmonic sinusoidal fashion at the velocity of light ($c = 3 \times 10^8$ m/sec). The distance from one wave peak to the next is the wavelength λ and the number of peaks passing through a fixed point in space per unit time is the wave frequency ν. The wave velocity is computed by $c = \nu\lambda$. The EM spectrum ranges from very short wavelength (10^{-7} μm) cosmic rays to very long (10^8 μm) radio waves.

In RS, the measurements of the EM spectrum are used to determine the properties of various earth features and vegetation. The basic concept is that each object, depending on its physical characteristics, reflects, emits, and absorbs varying intensities of radiation at different EM wavelength ranges. The chief source of EM radiation is the Sun. The EM spectrum can be divided into various ranges of wavelengths. Using the reflectance information from one or more wavelength ranges, it is possible to discriminate between different types of ground objects (e.g., water, dry soil, wet soil, vegetation, rocks, etc.) and map their distribution on the ground. One of the remote sensing techniques is aerial photography. Aerial photography essentially makes use of only the visible part of the EM spectrum of which it is a very small portion.

The wavelength ranges which are transmitted by the atmosphere without much loss of energy are known as *atmospheric windows*. The various ranges of the EM spectrum and the transmission of these by atmosphere are depicted in Fig. 3.1. Naturally, the wavelength ranges which are efficiently passed by the atmosphere are used in remote sensing analysis.

Useful atmospheric windows are available in the range 0.3-0.75 μm (ultra violet, UV, and visible), 0.77-0.95μm (near infra-red or NIR), 1.0-1.12μm, 1.19-1.34μm, 1.55-1.75μm, 2.05-2.4μm (shortwave IR), 3.5-4.16μm, 4.5-5.0μm, 8.0-9.2μm, 10.2-12.4μm, 17.0-22.0μm (middle and thermal IR), and microwave region.

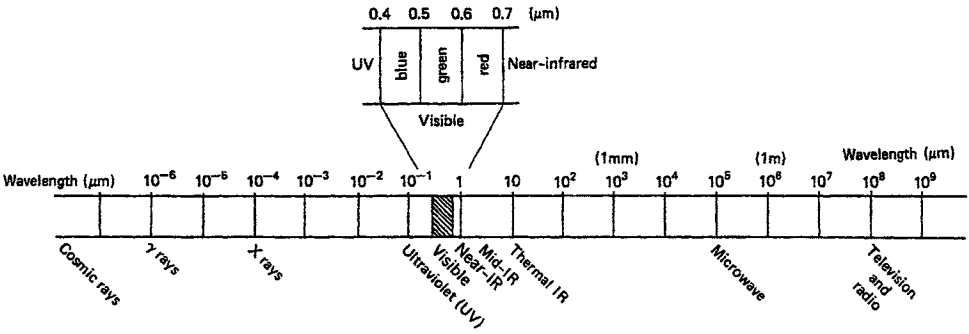


Fig. 3.1 The Electromagnetic Spectrum [Source: Lillesand and Kiefer (1994). Copyright © John Wiley & Sons, Inc. Used by permission of John Wiley & Sons, Inc.].

The EM radiation emitted or reflected from the objects of interest has to pass through atmosphere before the remote sensor detects it. Thus, the characteristics of the atmosphere significantly determine the effective use of electromagnetic spectrum for remote sensing. The absorption and re-emission processes in the atmosphere are through changes in electronic, vibrational and rotational quantization levels. The transmission and absorption of the energy by the gases present in the atmosphere depends on the wavelength ranges. The most important atmospheric constituents that influence the incident radiation are water vapour (H_2O), oxygen (O_2), ozone (O_3), carbon dioxide (CO_2), and aerosols. The spectral characteristics of energy sources, atmospheric effects, and sensing systems are shown in Fig. 3.2. Note that only a small portion of EM spectrum can be detected by human eye. This *visible* range is between 0.4 to 0.7 μm. The approximate wavelengths of three primary colours are: Blue 0.45-0.50μm, Green 0.5-0.6μm, and Red 0.6-0.7μm. When an object is illuminated by visible light, its color depends upon the reflectance properties. Chlorophyll present in healthy vegetation absorbs blue and red portion and reflects green light.

Depending on the source of EM energy, the RS systems can be classified in two categories: active and passive. The sensors that generate their own energy are called *active sensors*. These sensors, such as radar, transmit energy with certain properties and record the energy that is reflected back by the features that happen to be in the signal path. The *passive sensors* use an external source of energy. Most RS systems rely on the Sun to generate the EM energy that is needed to image objects and, therefore, are classified as passive. The design of sensors is influenced by atmospheric windows as the sensors measure energy corresponding to a particular range.

Advantages: One of the main advantages of RS techniques in natural resources management is the synoptic coverage of the earth on a periodic basis with small expenses. The advantages of RS over the conventional methods are enumerated next.

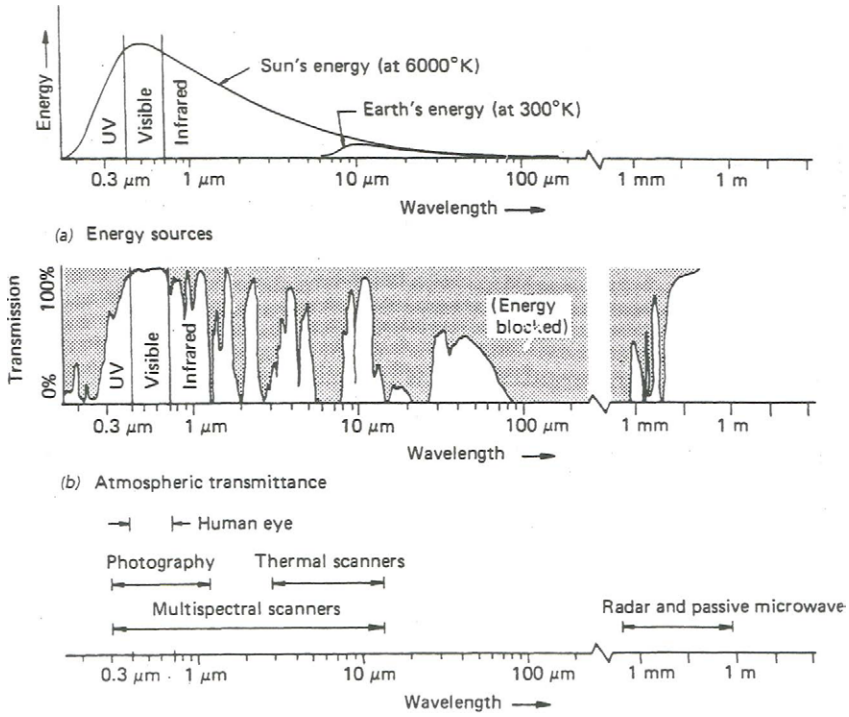


Fig. 3.2 The spectral characteristics of a) energy sources, b) atmospheric effects, and c) common sensing systems [Source: Lillesand and Kiefer (1994),Copyright © John Wiley & Sons, Inc. Used by permission of John Wiley & Sons, Inc.].

Measurement frequency: With the availability of RS data of sensors having resolution of the order of a few meters, highly accurate maps can be prepared. For water resources modeling, the accuracy is well within the desired limits.

Speed of analysis: Great savings in time can be made with the use of the RS approach, particularly the time spent in field. The ground truth can be collected by visiting selected spots in the study area. The satellite RS data collection is as easy for a remote, inaccessible area as for a nearby area.

Sampling frequency: Currently, RS data of the same area are easily available at an interval of 15 days or less. This is a very welcome feature because the time series of changes, such as land use, can be easily prepared and used. Such data was the most difficult to gather until the advent of RS.

3.1.1 Basic Components of Remote Sensing Data Collection

The basic components of a RS system are shown in Fig. 3.3. The EM energy source or illumination (A) provides the energy to sense the object. The radiation energy comes in contact with the atmosphere (B) and interacts with it as it passes through. Once the energy

reaches the target (C) through the atmosphere, it interacts with it in a manner that depends on the properties of both the target and the radiation. The radiation scattered or emitted from the target travels to the sensor and it again interacts with the atmosphere. The sensor (D) senses the energy and either records it or transmits it (often in electronic form) to a receiver on earth (E). At the receiver, the data are stored on a computer media. The data are passed on the users at (F) in the form of an image (digital and/or hard copy). These days most images are available in digital form; some data products may also be given in hard copy form. At the processing centers (F), the image is interpreted and analyzed, by visual or digital means to extract information about the target. The final product is usually produced in the form of a map (G) and is used in planning, problem solving, decision making, etc.

3.1.2 Remote Sensing Sensors

The commonly used sensors in RS are described in the following.

Gamma ray sensors: These sensors measure the difference between the natural gamma rays radiation of earth and that attenuated by soil, water or snow.

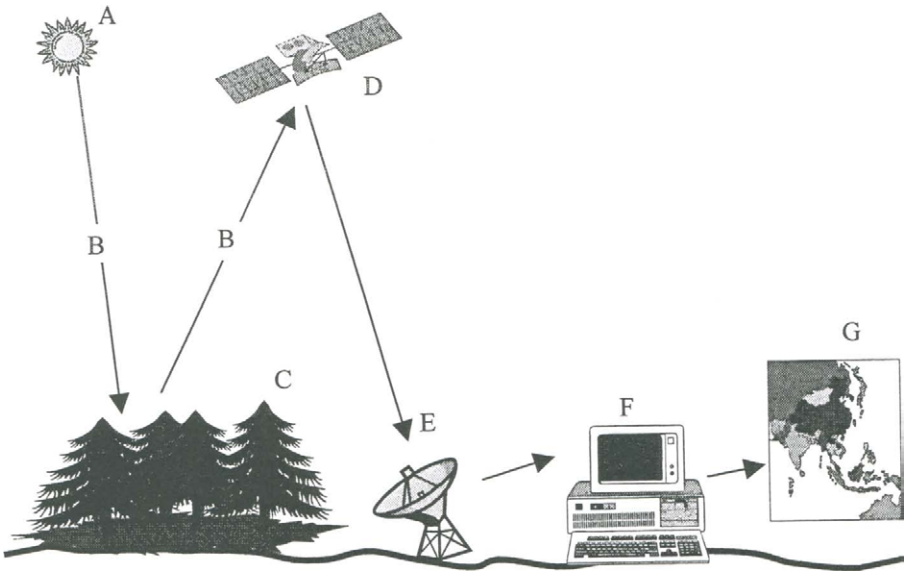


Fig. 3.3 Basic components of a remote sensing system.

Multispectral scanners: The detectors of these sensors are sensitive to specific regions of the spectrum and measure the spectral reflectance in narrow range of wavelengths or *bands*. Typical bands are: Blue (0.45 – 0.50μm), Green (0.50 – 0.60μm), Red (0.60 – 0.70μm), Panchromatic (0.50 – 0.75μm), and Thermal Infrared (TIR) (10 – 12.5μm). Note that the wavelength ranges of different bands may not be exactly the same in different satellites.

Thermal sensors: The temperature of a body depends on the incident and emitted energy

as well as its thermo-dynamic properties. The thermal sensors measure radiation in the thermal range. The thermal IR range is very useful in water resources studies.

Micro-wave sensors: These sensors measure the dielectric properties of the various features on earth surface. The micro-waves can penetrate through clouds and for this reason, such sensors can work in cloudy weather too. They are useful in measurement of soil moisture and mapping of areas during the flood season when the sky is usually overcast.

Applications of data of various spectral ranges in the field of water resources are indicated in Table 3.1.

Table 3.1 EM wavelength bands and their applications in water resources (Source: Engman 1993).

Band, μm	Applications
Blue (0.45-0.50)	Water penetration, land use, soil and vegetation characteristics, sediment
Green (0.50-0.60)	Green reflectance of healthy vegetation
Red (0.60-0.70)	Vegetation discrimination because of red chlorophyll absorption
Panchromatic (0.50-0.75)	Mapping, land use
Reflective (0.75-0.90)	Biomass, crop identification, soil-crop, land-water boundaries
Mid-infrared (1.5-1.75)	Plant turgidity, droughts, clouds- snow-ice discrimination
Mid-infrared (2.0-2.35)	Geology, rock formation
Thermal infrared (10-12.5)	Relative temperature, thermal discharges, vegetation classification, moisture studies

3.1.3 Remote Sensing Platforms

The sensors acquiring the RS data have to be mounted on platforms from which measurements can be taken. Such platforms may be (1) ground observation platforms (towers, vehicles, etc.); (2) air borne balloons, (3) aircrafts, and (4) satellites. The sensors mounted on ground based vehicles are mostly used during sensor development while verifying the design and understanding its response with respect to the characteristics of the target. If the data of a large area are to be obtained, it is necessary to take observations from a platform high up in the air. Due to this reason, satellites are the most convenient platforms. The satellites provide the coverage of all parts of earth and therefore are an ideal platform and are widely employed. The sensor may be mounted on a geostationary satellite which moves synchronously with the earth and is always positioned above the same point on the earth. Most weather or communication satellites (e.g., Meteosat: 0° longitude, 0° latitude) are of this type. They have little application in water resources since they can provide data of limited area and have poor resolution.

The polar orbiting satellites follow elliptical orbits (crossing near the poles) and can view a given area repetitively after a certain number of days. These satellites are sun synchronous which means that they image a particular area at the same local time in each revisit. The polar satellites provide a complete coverage of earth's surface. Some current satellites have sensors that can be steered sideways. This allows the cameras to also view and map off-nadir areas. By the use of this feature, the image of an area can be obtained more frequently than is normally possible but the satellite has to be pre-programmed for this purpose. This feature also permits stereo coverage of an area. Some satellite systems whose data are most commonly used include the Landsat (including Thematic Mapper, TM) of NASA, the French SPOT (Systeme Probatoire d'Observation de la Terre) satellites, the Indian Remote Sensing (IRS) satellites, the ERS and JERS satellites, and GOES, GMS, etc. The Landsat data are available since 1972, the SPOT since 1986, and the IRS since 1988. The most relevant spectral bands are visible (VIS) at various colours, infrared (IR), near infrared (NIR), thermal infrared, passive microwave and active microwave (Radar). Besides satellites, ground based RS data collection systems are also of importance in water resources. For instance, ground based weather radar is very useful for areal estimation of precipitation.

Currently, a number of satellite systems are operational. A few of these are described below.

LANDSAT Program

The Landsat program has provided the most extensively used RS data the world over. Its chief plank has been in delivering unrestricted global data of good geometric accuracy. Under the Landsat programme, six satellites (Landsat - 1, 2, 3, 4, 5 & 6) were launched till year 2000 (Landsat-6 having failed). These satellites have been placed in near-polar, near-circular, sun-synchronous orbit at an altitude of 700 to 900 km. As the satellite orbits in the north-south plane, the earth below it spins around its axis, from west to east. Thus, different parts of the globe are 'seen' by the satellite during different passes. The data are acquired in the descending node, i.e., as the satellite moves from the north pole to the south pole.

Two important sensors on-board the Landsat satellites are Multispectral Scanning System (MSS) and Thematic Mapper (TM). Both are on-line scanners and produce ground scenes of 185*185 km size. The MSS sensor has been a regular payload of the Landsat and has made this programme a tremendous success. The TM is an advanced multispectral scanner used in Landsat 4 & 5 missions. TM operates in seven wavelength bands, out of which six are in the solar reflection region and one in thermal-IR region. The Landsat satellites offer high resolution in space (30 x 30m pixels) and data of 7 spectral bands which are suitable for many applications.

The web site of Landsat (NASA) is: http://landsat.gsfc.nasa.gov/.

SPOT Program

The French satellite system SPOT system commenced its operation in 1986. These satellites have been placed in near-polar sun-synchronous 830 km high orbit with a repeat cycle of 26

days. The sensor here is called HRV (High Resolution Visible) instrument which is a CCD-line scanner. The HRV's acquire data in two modes: (a) panchromatic mode in a swath of 60 km with ground resolution of 10*10 m, and (b) multispectral mode, in three channels (green, red and infrared) with a ground resolution of 20*20 m in a swath width of 60 km. The HRV's can also be tilted to acquire data in off-nadir viewing mode for more frequent repetitive coverage and for stereoscopy.

Details of this system and data are available at http://www.spot.fr.

Indian Space Program

Under the Indian Remote Sensing Satellite (IRS) program, the Indian Space Research Organisation (ISRO) has launched a series of land observation satellites. The operational first generation RS satellites IRS-1A and IRS-1B were launched in 1988 and 1991, respectively. These were placed in near-polar, sun-synchronous orbit, with repetitive time of 22 days. The satellites had two Linear Imaging Scanning Sensors (LISS-I and LISS-II) for providing data in four spectral bands: Visible, Infra Red (IR) and Near Infra Red (NIR). Their ground resolutions were 72.5 m and 36.25 m, respectively. The second generation, operational, multi-sensor satellites IRS-1C and IRS-1D were launched in 1995 and 1997. These were placed in near-polar, sun-synchronous orbit with a repetitive time of 22 days. These satellites have three on-board cameras. The PANchromatic (PAN) camera operates in the panchromatic region of the EM spectrum and has a spatial resolution of 5.8m. It can be steered up to 26° across-track, thus enabling generation of steoroscopic imagery and improved revisit capability. The Linear Imaging Self Scanner-III (LISS-III) camera operates in four spectral bands: three in Visible/ Near Infrared (VNIR) and one in Short Wave Infrared (SWIR) region. It has a resolution of 23.5m in VNIR bands (swath 141 km) and 70m in SWIR band. The Wide Field Sensor (WiFS) is a coarse resolution camera with spatial resolution of 188.3m. A satellite for oceanographic studies, namely, IRS-P4 (OCEANSAT-1) having an ocean color monitor with 8 spectral bands and a multi-frequency scanning microwave radiometer operating in four frequencies has been recently launched to enable measurement of physical and biological ocean parameters.

In view of rapid developments taking place in space technology and increasing application of remote sensing and GIS techniques, ISRO has planned a number of future satellite IRS missions. These will provide a cadastral level information up to 1:5000 scale thematic applications, vegetation and multi-crop discrimination, species level discrimination etc. Further details of satellites and data can be obtained from http://www.nrsa.gov.in.

Besides, the National Oceanic and Atmospheric Administration (NOAA) has a system of satellites for meteorological purposes. Active microwave data are useful in study of soil moisture in areas that are covered with clouds. The Radarsat satellite carries sensors to collect microwave data.

The information on European Remote Satellites is available at www.esrin.esa.it.

3.1.4 Resolution of Remote Sensing Data

The RS data has the following four types of resolutions.

Spatial resolution. Mather (1987) lists four criteria for the definition of spatial resolution: geometrical properties of the imaging system, the ability to distinguish between point targets, the ability to measure the periodicity of repetitive targets, and the ability to measure the spectral properties of small targets. The instantaneous field of view (IFOV) is the most common measure of geometric properties. It is the area on ground that is viewed by the instrument from a particular altitude at a given time. As shown in Fig. 3.4, it can be measured either as an angle α or as the equivalent distance on the ground. Due to various reasons, IFOV is not a useful measure of resolution and, moreover, it need not be linked to pixel size. The values can be interpolated over the cells of the image to represent any ground spacing (Mather, 1987). For example, the pixel size of IRS-1C LISS III sensor is 23.5m. Nowadays, images with pixel sizes of the order of a meter are easily available.

The pixel size is chosen based on the purpose of study and the size of study area. In case of large pixel sizes, problems arise in georeferencing because ground control points with desired accuracy cannot be located.

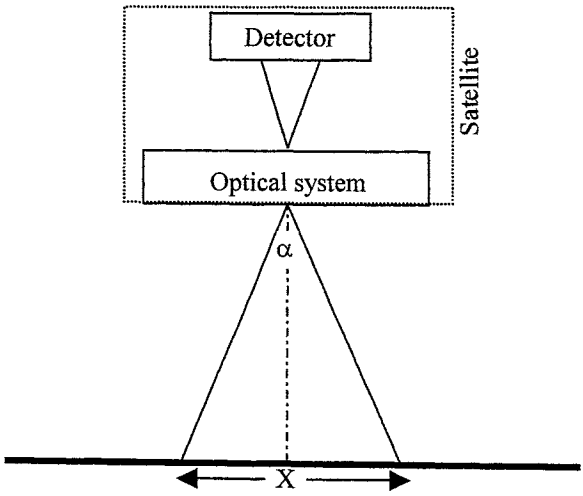


Fig. 3.4 Angular IFOV and projection X on ground. X is the diameter of a circle.

Spectral resolution indicates the number and bandwidth of specific wavelength intervals in the EM spectrum to which a sensor is sensitive. For example, the LISS-III sensor of IRS-1C satellite operates in four spectral bands.

Radiometric resolution – In RS, the reflected radiation from an object is measured by the sensor. An electrical signal is generated whose strength depends on the radiation received

by the sensor and this strength is converted and stored as digital numbers. The radiometric resolution is the number of levels which can be realized in the range of radiation levels. Typically there are 128 or 256 levels.

Temporal resolution or re-visit frequency means how often the sensor records imagery of the same area. Commonly it is of the order of two weeks. Geostationary satellites produce a very high resolution in time (say, about 30 minutes) but have a rather coarse spatial resolution, e.g., of the order of several km. Ground based weather radars produce a rather high resolution in time and space (e.g., 15 minutes in time, 1 km^2 in space).

The RS data are also available as hardcopy prints of images. However, these images differ from normal photographs. In normal photographs, an object of green color is printed in green, red color in red and so on. The resultant print is known as true color composite. The visual perception of an RS image can be altered by assigning, for example, blue, green, and red colors to observations in the green, red, and near-infrared (NIR) bands. The product so generated is known as *false color composite (FCC)*. Since vegetation is highly reflective in NIR band, it will appear as red in standard FCC. The hard copy FCC can be analyzed visually but visual analysis is gradually becoming outdated.

The digital data contains the reflectance of each pixel in various spectral ranges. A number of hardware and software systems are available for analysis of digital imagery or *image processing*. The current trend of software development is for PC or workstation based systems. The following discussion is focused on analysis techniques and applications of remote sensing to water resources.

3.1.5 Reflectance Characteristics of Earth Features

Reflectance is the ratio of energy reflected to the total energy incident on a body, expressed in percentage. An object appears green because it reflects only wavelengths corresponding to green color in the visible spectrum. Thus, blue colored objects absorb all light waves except those pertaining to blue color, and so on. The spectral radiant flux ϕ that is incident on the earth's surface is reflected, or absorbed, or transmitted. Thus,

$$\phi = \phi_r + \phi_a + \phi_t \tag{3.1}$$

where ϕ_r, ϕ_a, ϕ_t are the reflected, absorbed, and transmitted parts of the flux. The relative magnitudes of these depend on the characteristics of the target on the earth's surface. This property is used to identify various features on the earth surface through measurements on their spectral properties. The remotely sensed measured signal (ϕ_r) as a function of the wavelength is often referred to as the *spectral signature* of the target because its analysis provides useful information about the properties of target.

Spectral reflectance is the ratio of reflected energy to incident radiation (ϕ_r/ϕ) as a function of the wavelength. The spectral reflectance characteristics of most targets depend upon the illuminating region of the EM spectrum. The MSS, TM, LISS are some of the sensors to sense the radiances at distinct spectral bandwidths. Vegetation, for example, may

reflect only 10 to 15 percent in the green band of the EM spectrum but as much as 40 to 60 per cent in the NIR band. However, under certain conditions, the spectral reflectance characteristics of different objects may be the same. For example, water and wet black soil may show the same reflectance properties in a band. In such cases, it would be necessary to examine the data of other spectral bands to differentiate them. This is the basic motivation behind multispectral analysis. Fig. 3.5 shows spectral reflectance curves for water, soil, and vegetation.

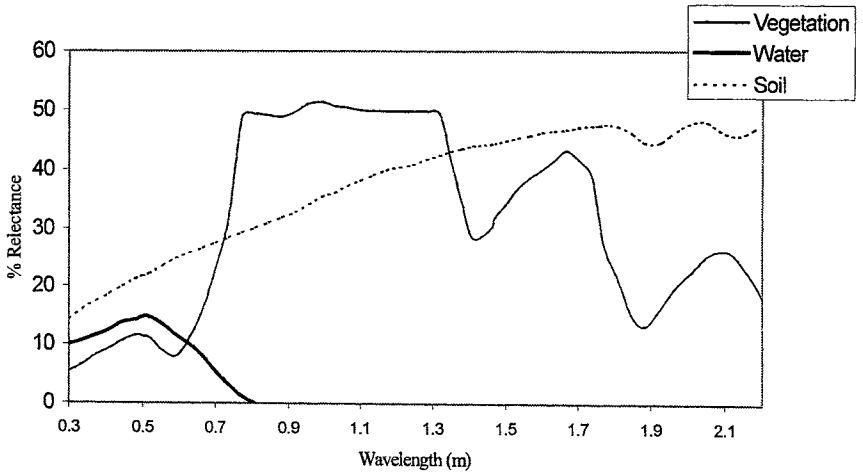


Fig. 3.5 Typical spectral reflectance curves for water, soil, and vegetation.

The mapping and identification of vegetation is an important aspect of RS. A leaf is built-up of layers of structural fibrous organic matter. In this are pigmented water filled cells and air spaces. The properties of pigment, leaf physiological structure, and water content affect its reflectance, absorptance and transmittance properties. In the visible band, the green colour of the leaf is due to peak reflectance at 0.54 μm. The presence of chlorophyll is responsible for very low reflectance (10-30%) in red and blue bands. Some middle infrared (MIR) bands are absorbed by water vapour. The reflectivity rises rapidly near 0.75μm and remains high in the NIR region in the range 0.75 to 1.35μm. The peak reflectance in MIR occurs at 1.6 and 2.2 μm. Therefore, measurements in 1.55 – 1.75 and 2.08 – 2.35 μm bands give useful information about moisture content of the plant canopy.

A leaf with low pigment content gives a higher reflectance in the red region. Stress in vegetation due to diseases, insect, and nutrient deficiency affects the reflectance characteristics. A healthy plant gives a less reflectance in the red region and a high reflectance in NIR. The stress is noticeable in the NIR imagery even in the early stage of disease. The seasonal state of maturity of a plant also influences its spectral reflectance and thus the spectral signature of a plant species usually varies during a season and in its life cycle. These spectral reflectance properties of vegetation are used to discriminate vegetated and non-vegetated areas. There can be considerable differences between the plant species and these can be used to discriminate among the species.

The reflectance of soil depends on the chemical and physical properties of its components, organic matter, texture, moisture, surface roughness, and sun angle. Soils usually have a higher reflectance than plants in the visible bands. Typical curve shows rising reflectivity with increase in wavelength. In the visible band, reflectivity is affected by organic matter and moisture. In NIR band, plants have a higher reflectance than do soils. Generally, a reduction in grain size of soil results in an increase of reflectance. In case of rocks, the reflectance properties significantly depend on their type, chemical composition, weathering, rock outcrop, etc. The influence of weathering on spectral signatures can be either way; generally, the weathered rocks have a low reflectance.

Clear water gives low reflectance in visible and almost completely absorbs the radiation in NIR regions. The signal received by the satellite mainly depends on the conditions near the surface of the water body. Fig. 3.6 shows processes acting on solar radiation in the visible part of the spectrum over an area of shallow water. A part of solar irradiance is scattered by the atmosphere and some of it reaches the sensor. Of the irradiance reaching water surface, a part is reflected. The incident solar energy that is not specularly reflected by water is refracted downward. Within the water body, radiation is either absorbed by water or dissolved substances or is backscattered by suspended particles (volume reflectance). The scatter by clear water depends on wavelength.

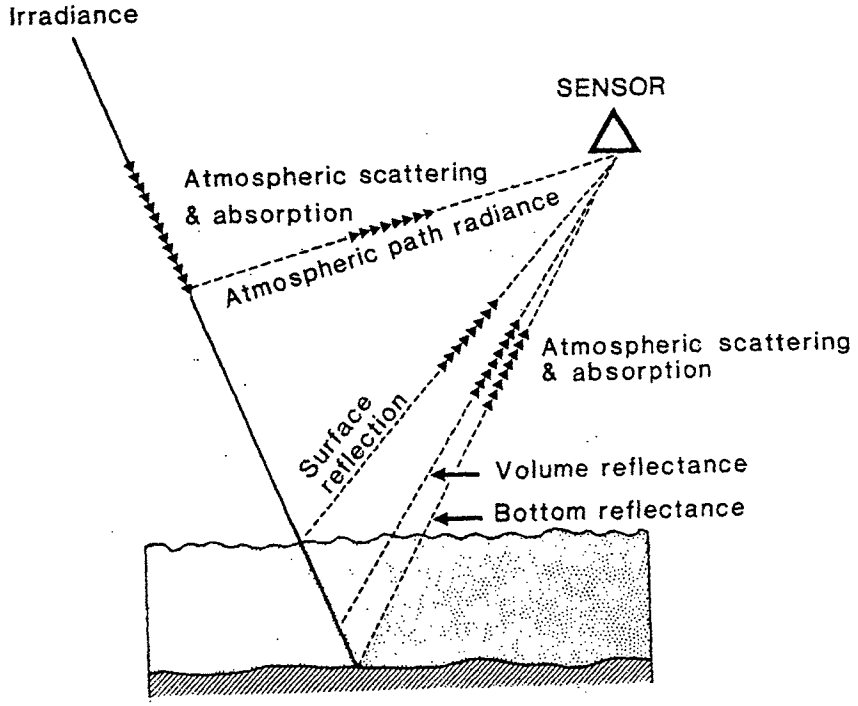


Fig. 3.6 Processes acting on solar radiation in the visible part of the spectrum over an area of shallow water [Source: Mather (1987). Copyright © John Wiley & Sons, Limited. Used by permission].

Mapping and monitoring the quality of water by RS is considerably difficult because the signal from the water body is composed of many components. A number of variables, such as suspended particles, floating materials, and the depth of water body, affect the reflectance of EM by impure water. Generally, the reflectance by turbid water is more than clear water in the visible and NIR bands; the change being dependent on the characteristics of the suspended material. Different types of materials differently affect the radiation. For a particular size and shape of particles, the reflected energy increases with the concentration of impurities and a correlation can be established between the two.

3.1.6 Remote Sensing Data Analysis

The satellite image can be analyzed using either the visual or the digital techniques. These are now discussed.

Visual Image Processing

The visual image processing is based on how the analyst views and interprets an image. The analyst uses his experience and knowledge about the area to interpret the hard copy of the image. Clearly, the results depend on his ability and vary from one analyst to another. The major characteristics that form the basis of image interpretation are tone/color, texture, size, shape, pattern, and association. On a standard FCC, water appears in black color, vegetation in red, etc. The water bodies have smooth texture, while cities have medium texture. The agricultural fields can be easily identified by their typical rectangular size and nestling; an airstrip will be long and thin. Shape wise, canals will be long and thin, drainage curvilinear. In terms of pattern, roads will be radial, streams in tree-like pattern, and lineaments thin, long, and crisscrossing. Examples of association are canals and agricultural fields, roads, cities, etc.

There are a few key limitations of visual interpretation. An imagery contains a huge amount of digital data and human beings are not the best number crunchers. Computers can carry out this job much more effectively. The eye of a normal human being can only detect the difference between 8 - 16 different shades of gray and thus a lot of useful information remains unused. It is not normally possible to get consistent results from different interpreters.

3.1.7 Digital Image Processing

A digital image is formed by a number of *pic*ture *el*ements or *pixel*s of the same size. Each pixel represents the spectral response of a small square shaped area on the ground and has certain spatial and spectral properties. The spatial property defines the ground coverage of the pixel. The size of the area depends on the resolution of the on-board camera. For example, the resolution of LISS III camera of IRS-1C satellite is 23.5 m resulting in a pixel size of 23.5m*23.5m. Smaller the pixel size, more details of the target are mapped. The spectral properties define the intensity of spectral response for a pixel in various bands. Associated with each pixel is a number known as *digital number* which is the integrated radiance response of the ground covered by that pixel.

A row of pixels represents a scan line that is recorded as the sensor moves left to right or recorded through the use of a linear array of photodetectors. An image is composed of a number of geographically ordered m scan lines, placed adjacent to one another in the direction of the y-axis. Each scan line consists of n pixels in the direction of the x-axis. If the data of only one band of the EM spectrum is recorded, an image with pixels in gray shades is obtained. Multispectral sensors detect the light reflectance in more than one band of the EM spectrum. The computer processing of an imagery allows a quantitative analysis of all spectral bands simultaneously. It is also possible to detect and interpret small differences in spectral response that a human eye cannot.

A digital image is a two-dimensional light intensity function denoted by f(i, j), where the value or amplitude of f at spatial co-ordinate (i, j) gives the intensity (brightness) of the image at that point. Since light is a form of energy, f(i, j) must be positive and finite, and we have

$$0 < f(i, j) < u \tag{3.2}$$

where u is the upper limit of response. A continuous digital image f(i, j) is approximated by equally spaced samples arranged in the form of a $n*m$ array as shown below, where each element of the array is a pixel:

$$f(i,j) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,n-1) \\ f(1,0) & f(1,1) & \dots & f(1,n-1) \\ \dots & \dots & \dots & \dots \\ f(m-1,0) & f(m-1,1) & \dots & f(m-1,n-1) \end{bmatrix} \tag{3.3}$$

When the raw digital data is displayed on a computer screen, it is difficult to discriminate the various terrain features. Digital Image Processing (DIP) involves the manipulation of digital data to improve the image qualities or to enhance the features of interest with the aid of a computer. The process helps in maximising clarity, sharpness, and details of objects of interest and leads to better information extraction. It improves the image's interpretability. Image processing operations are carried out to remove noise from the data and enhance certain features based on their spectral response. DIP is a broad subject and may involve procedures that can be simple as well as quite complex. Basically, each pixel of an image is mathematically manipulated and the operation may involve more than one image. The results of computations for each pixel are stored and form a new image. The new digital image may be subject to further manipulation, may be stored or a hard copy may be taken.

Geometric Correction and Registration

The remotely sensed images are frequently integrated with maps. In *geometric correction*, an image is transformed so that it has the same scale and projection properties as a map. When many images of the same area are to be processed, it is often helpful if these are registered. In *registration*, the coordinate system of one image (called master image) is fit to other images of the same area.

Image Enhancement

Image enhancement algorithms are commonly applied to remotely sensed data to improve the appearance of an image and a new enhanced image is produced. The enhanced image is generally easier to interpret than the original image.

RS images are collected in multispectral bands, i.e., the same scene is simultaneously scanned in several spectral bands of the EM spectrum. The radiance measured in each band is an average value over a range of wavelengths in the spectral region, termed as bandwidth. Image contrast is related to the range of gray levels (GL) in an image, larger the range, greater the contrast and vice versa. Both linear or non-linear contrast enhancement techniques are used for contrast enhancement. Contrast C may be computed in several ways, e.g.:

$$C = GL_{Max} / GL_{Min} \quad (3.4)$$

or $$C = GL_{Max} - GL_{Min} \quad (3.5)$$

where GL_{Max} and GL_{Min} are the maximum and minimum gray levels in the image, respectively. Contrast is an indicator of the visual quality and is also a measure of the signal-to-noise ratio of an image. It is desirable to utilise the entire brightness range of the display system or hard-copy photographic film.

A contrast enhancement (often referred to as a contrast stretch) expands the original input brightness values to make use of the total dynamic range or sensitivity of the output device. The linear contrast enhancement is best applied to remotely sensed images with Gaussian or near-Gaussian histograms, wherein all the brightness values generally fall within a single, relatively narrow range of the histogram and only one mode is apparent. Unfortunately, this is a rare case, especially for scenes with extensive land and water bodies.

Non-linear contrast enhancements can be applied to low-contrast imagery. One of the most useful techniques is *histogram equalisation.* The histogram of the image is determined based on the number of output gray-scale classes into which the data are to be redistributed. Now the input data are modified to assign approximately equal number of pixels under each of the output gray-scale classes. The histogram equalisation applies the greatest contrast enhancement to the most populated range of brightness values in the image. It automatically reduces the contrast in very light or dark parts of the image associated with the tail portion of a normally distributed histogram.

Principal Component Analysis

If there is significant correlation among the data of different bands, this shows that there is redundancy in the data. Fig. 3.7 contains a plot of two highly correlated variables and AB is the best-fit line. If a plot is prepared with AB as the major axis and a line CD right angle to it as the minor axis, in most cases, the new plot will reveal more information about the structures that are present in the data. In this example, although there are two variables, the dimensionality of the data is one.

The Principal Component Analysis (PCA), also referred to as factor analysis, is an important technique for the analysis of RS digital data. In this, the data of a number of bands is transformed into the same number of principal components (PC) such that the first two to three PCs contain nearly 95% of the input information. The information content of higher order PCs is not much, they mostly contain noise. The ability to reduce the dimensionality from seven to two or three bands is an important consideration, especially when the transformed data contains nearly as good information as the original data. Consider a scene of Landsat TM satellite which has data of seven bands. If the dimensionality of the data is three, than a FCC can be prepared in which the three primary colours can represent three principal components. The information content of this FCC will be more than any other band combination.

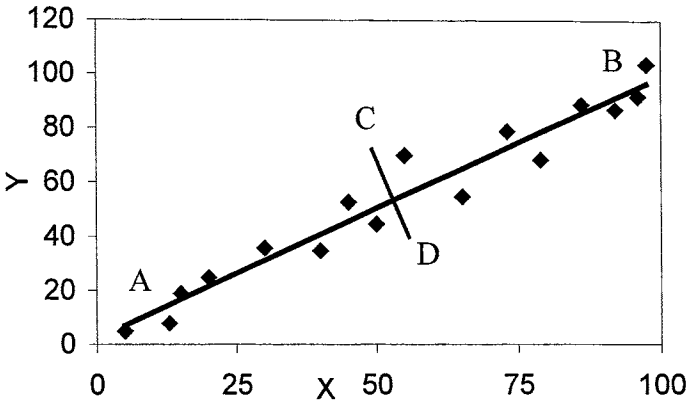


Fig. 3.7 Variability of two high positively correlated variables.

Spectral Indices

The spectral indices are calculated using the data of two or more bands and are very useful in image classification. Bastiaanssen (1998) has described many indices that are routinely used in image processing. A few important indices are given below.

Normalized Difference Vegetation Index (NDVI)

This is the most commonly used index which is an indicator of the presence of chlorophyll in the vegetation:

$$NDVI = (NIR - RED) / (NIR + RED) \tag{3.6}$$

where NIR and RED are reflectances in NIR and RED band. The red band of EM spectrum is largely absorbed by chlorophyll and reflectance is low. On the other hand, the NIR band is largely reflected by it and has low reflectance. Therefore, healthy vegetation will have high DN value in the NIR band and low in red and consequently, large NDVI. The range of NDVI is –1 to +1. NDVI is a high positive number for healthy vegetation and a small

positive number for a water stressed or deceased vegetation.

Normalized Difference Wetness Index (NDWI)

Rao and Mohankumar (1994) proposed this index for identifying irrigated crops and suggested that this index is better than NDVI for this purpose. NDWI is defined as:

$$NDWI = (SWIR - MIR) / (SWIR + MIR) \quad (3.7)$$

Soil Adjusted Vegetation Index (SAVI)

It is defined as

$$SAVI = 1.5(NIR - RED)/(NIR + RED + 0.5) \quad (3.8)$$

Many attempts have been made to relate SAVI with features of crops. These are explained in Section 3.1.8.

Classification

Image classification is a problem of pattern recognition whose aim is to find the characteristics of objects on the earth at corresponding pixels in the image. The classification algorithms make use of the spectral reflectance values at the pixel in various bands and then tries to identify the most likely class to which the pixel belongs. Thus, using the spectral reflectances, a particular pixel may be labelled as water or dense forest, etc. The aim may also be to identify all the pixels in the image that correspond to water. Two types of classification techniques are normally employed: supervised or unsupervised.

Supervised classification is carried out when the identity and location of some of the features on the image, such as urban area, wetland, and forest, are known a priori through information gathered from field visits and study of toposheets, etc. For each of these features, statistical parameters are calculated and then one can build a series of templates representing these patterns. These templates or areas are known as *training sites* and their spectral characteristics are used to guide the classification algorithm. Next, every pixel of the image is compared with each template and is assigned the class whose properties are closest to it. In other words, it is assigned to a class of which it has the highest likelihood of being a member. The maximum likelihood classifier is the most widely used criterion.

If the identities of land features that are to be classified in a scene are not generally known a priori due to the absence of the ground truth data or other supporting data, an unsupervised classification is performed. The software then groups the pixels into different classes according to some statistical criteria. The analyst, based on his experience and familiarity with the scene, labels these clusters. If there are only a few pixels in some groups, these groups may be deleted or merged with others. Similarly, if some groups are too heterogeneous, these may be split. A combination of supervised and unsupervised classification is often helpful, particularly if adequate field data are missing.

The problem of mixed pixel occurs in classification when the ground area corresponding to the pixel has mixed features, such as part crops and part water body. The spectral properties of such a pixel show mixed behavior. Such pixels can be classified using additional ground information, supporting information from other imageries, or the judgment of the analyst. Advanced techniques, such as fuzzy classification or ANN, have also been used in image classification. Image processing techniques have been discussed in detail in texts, such as Lillesand and Kiefer (1994) and Mather (1987).

Density Slicing

Density slicing is a technique of image processing in which the histogram of the image is divided in a number of intervals or slices. All the pixels that fall in a given interval are supposed to belong to a particular category. This technique has been used in a number of studies. For example, the identification of water spread area of a surface water body. However, many times, this way of classification is considered to be inadequate and liable to errors.

Change Detection

A *change detection* of the feature involves analysis of a series of images of different dates. In water resources studies, land cover is a most commonly studied phenomenon. The analyst might be interested in both short–term (e.g., flood inundation or snow cover) as well as long-term (e.g., deforestation or expansion of urban area) changes. It will be good if the data of the same sensor and the same resolution are used. The analysis also requires that all the images are accurately registered so that overlays can be prepared to visualize the change.

The simplest way to detect changes is by subtracting DN values of one image from the other. At places where changes are minor, the result will be close to zero. Another possibility is to take ratio of two images. If there is no change at a place, the ratio will be close to one; at places of change, the ratio will be some higher or lower number. The classification of different images can be performed singly or jointly. If the images are classified independently, the accuracy of change detection will depend on the accuracy of classification. The classification becomes complex if many images are classified together. Of course, the analyst has to decide a threshold above which the change is considered significant and below which it is insignificant or could be due to error in analysis.

Advantages of Digital Image Processing

The technique of digital image processing is very cost-effective for mapping large geographic areas, particularly when imageries of a number of dates are to be analyzed. It is possible to simultaneously use data of several bands for improved interpretation and modern software permit formulation of complex interpretation algorithms which can use input data from a number of imageries. The analysis results are consistent because subjectivity is not involved. With the availability of fast computers, the interpretation and analysis can be carried out in a short time.

Although the skills and efforts required for digital image processing are considerably less now due to the availability of user-friendly software, it still demands sufficient experience, particularly where other than routine interpretation is to be carried out. The analysis can be expensive for small areas or for one-time interpretations as the start-up costs may be high. A good set-up requires not only a faster computer with sufficient disk storage space but also special peripheral devices, such as a laser (preferably color) printer, a scanner, digitizer, CD-ROM writer, etc. An Internet connection is also necessary since currently a lot of data and literature can be accessed through it. Many organizations are now using Internet as the preferred medium for data supply. Although the cost of satellite data has come down recently, it is still expensive, particularly the microwave data. It is difficult to evaluate the accuracy of the results, more so when the target area lies in an inaccessible region.

3.1.8 Applications of Remote Sensing to Water Resources Problems

Currently, RS data for an area are available from a number of satellites. For applications dealing with water resources, the analyst first identifies which available RS data are suitable for his problems. Then a choice is made based on the considerations, such as the spatial and temporal resolution, the sensors and spectral bands available, and the cost of data products.

The main fields of RS application in hydrology are measurement of precipitation, evapotranspiration, irrigation water management, snow cover mapping, flood mapping, and hydrological modelling. The use of RS data in hydrology requires efficient storage and retrieval of RS raster data in a data bank coupled to a GIS. Usually the original RS data have to be stored for all spectral bands (e.g., 7 channels for Landsat) as well as the derived products, e.g., landuse classification maps, snow cover maps, vegetation index maps, etc. The GIS aids the hydrologist to produce these derived maps from the original RS data and it is again of assistance when these derived data are used in hydrological modelling. The main application areas of RS technology in the field of water resources are briefly described here.

Precipitation Estimation: Direct measurement of rainfall from satellite data has not been much successful thus far but the same can be inferred by indirect means, e.g., by deriving the spatial distribution of rain producing clouds, temperature and other cloud characteristics from RS data. The geo-stationary meteorological satellites are used for this purpose but currently their spatial resolution is very coarse. The data that are used for rainfall studies are generally in the spectral range of 0.5 – 0.7 μm (visible), 3.5-4.2 μm & 10.5-12.5 μm (infrared) and 0.81 –1.55 cm (microwave). The data from the visible region can be used to track cloud development.

The major attempts in this field are the Global Precipitation Climatology Project (GPCP) and the Tropical Rainfall Measurement Mission (TRMM, a joint mission between NASA and National Space Development Agency, NASDA, of Japan. Their web-site is: www.trmm.gsfc.nasa.gov). The thermal IR data have been used to estimate cloud-top temperature and attempts have been made to relate this to rainfall. The satellite data can be combined with meteorological observations to develop a cloud index which identifies different types of rain clouds and estimates the rainfall based on the number and duration of

clouds or their area. The temperature threshold method has been widely used to estimate monthly rainfall from satellite IR images (Richards and Arkin, 1981). The monthly rainfall over a grid cell of 2.5° latitude by 2.5° longitude is given by (Smith 1993):

$$R = (3 \text{ mm/h}) (FRAC)(HOURS) \tag{3.9}$$

where R is rainfall (mm), *FRAC* is the fractional coverage of cloud-top temperature less than -38°C for the cell, and *HOURS* is the number of hours in the observation period. In the *cloud indexing approach*, the cloud data such as cloud type and amount are related to measured rainfall at ground over certain time period by a relationship:

$$R = f\{C, i(a)\} \tag{3.10}$$

where R is total rainfall, and C, i, and a are cloud area, type, and altitude respectively.

In the concept of *Cold Cloud Duration* (CCD), Dugdale and Milford (1986) suggested that the duration above a threshold temperature can be related to the amount of rainfall that is produced by the clouds. The thermal IR data of Meteosat satellite was used by the authors. This technique has been successfully applied at several places. Hsu et al. (1997) applied Artificial Neural Networks (ANN) to estimate rainfall over a specified area using satellite IR images. The inputs were normalized values of the IR brightness temperature of the pixel for which rainfall had to be predicted, and the mean and standard deviation of the brightness temperature of pixel windows of various sizes centered at the prediction pixel. The rainfall rate over the prediction pixel was the ANN output. The case studies using data from Japan and USA showed good estimation of rainfall using this approach.

Irrigation Water Management: The attempts to identify crop types using RS began in the 1980s. Over the years, a large number of RS studies related to various aspects of irrigation water management have been conducted. Irrigation is the main consumer of freshwater and in view of the water scarcity, these studies are highly relevant and useful. The attractiveness of RS stems from its ability to provide repetitive and distributed agricultural and hydrological status of a command. The major aspects that have been covered in the studies include mapping of land use, crop types, crop water requirements, soil salinity, and waterlogged areas.

Leaf area index (LAI) is defined as the cumulative area of leaves per unit ground area at nadir orientation. It varies from 0 for bare soil to around 3 to 8 for annual crops and 15 or more for dense evergreen forests. Many attempts have been made to relate LAI with vegetation indices, such as NDVI and SAVI. The procedure and its success depends upon the properties of RS data, availability of supporting ground truth data, and the properties of the target area. Bastiaanssen (1998) reports that the crop classification accuracy up to 90% and above have been obtained in many studies. The active microwave data of satellites, such as Radarsat, have been used in soil moisture studies.

Choudhary et al. (1994) gave the following type of regression relationship between

SAVI and LAI for cotton, maize, and soybean:

$$SAVI = c_1 - c_2 \exp(-c_3 * LAI) \quad (3.11)$$

where c_1, c_2, and c_3 are regression coefficients. Bastiaanssen (1998) showed that the relationship between LAI and SAVI is fairly linear during the development of a crop, until a threshold value of LAI is reached. He noted that selection of a vegetation index to obtain biophysical parameters is to a certain extent redundant because most vegetation indices exhibit a similar spatial behavior. Therefore, SAVI should be considered the best indicator of crop biophysical parameters for irrigation management.

Snow Cover Mapping: RS is a valuable tool in snow and ice studies. This topic has been covered in depth by Hall and Martinec (1985). Fresh snow is the brightest formation on the Earth surface. The albedo or reflectance of the snow surface can be measured by RS techniques. The albedo of snow rapidly decreases with its age. The freshly fallen snow typically has an albedo of about 90%. Due to accumulation of dust, etc., it may reduce to 40% or so in a few days. However, as both snow and clouds have high reflectance in the visible bands, differentiation could be problematic. Besides, problems may also arise when snow is obscured by thick vegetation, bare rocks with reflectance similar to snow are present, and low sun angles generate shadows on north facing mountain slopes.

The RS data are helpful in mapping snow cover areas and assessing snow water equivalent. The data of visible and near-IR bands are widely used for snow cover mapping. Dozier (1989) developed a model using atmospheric, topographic, and radiation data to simulate planetary reflectance for a range of snow grain sizes and topographic conditions. He developed an automatic procedure for mapping snow cover area and used data of Landsat TM band 1, 2, and 5 to distinguish snow from other surfaces and clouds. Goel and Jain (1997) modified the Dozier algorithm for IRS-1C data and used it to map snow cover in a Himalayan catchment. Many studies, such as Rango and Martinec (1999), have used RS data in snowmelt modeling. Using Landsat MSS and TM data, Seidel et al. (2000) were able to distinguish between snow and ice and temporally monitor their areas. The microwave data provides information on snow covered areas as well as snow water equivalent and the presence of melt water in the snow pack. This technique has great potential in snow and ice studies.

Evapotranspiration: Evapotranspiration (ET) is one of the most important processes of the hydrological cycle. Depending upon the climate, it may account for 40% or more of the moisture lost from a catchment. The conventional ground-based instruments and techniques can provide information over a very limited area but one really needs areal evapotranspiration values for atmospheric general circulation models, hydrological, and agricultural studies.

The satellite data can be used to measure variables that are used in the energy and moisture balance methods of computing ET or the empirical ET relations can be applied to large areas in conjunction with such techniques. The variables that can be estimated by RS include incoming solar radiation, surface albedo, vegetation areas and vegetation properties

(NDVI), etc. Many studies, such as Bausch (1995), have attempted to estimate *crop coefficients* using RS. Crop coefficient (k_c) is the ratio of potential ET for a given crop to that of a reference crop (generally grass of alfalfa):

$$k_c = ET_{pot}/ET_{ref} \tag{3.12}$$

Attempts have been made to relate crop coefficients with vegetation indices. For example, Bausch and Neale (1987) related k_c with NDVI. Ambast et al. (2000) estimated regional ET using RS and ancillary data. They used the Landsat TM data in visible and IR bands to generate surface albedo, surface temperature, NDVI, and LAI images. Next, soil heat flux and sensible heat flux images were generated. Latent heat flux density was obtained by surface energy balance. Regional ET was obtained using evaporative fraction on instantaneous time basis with the total energy integrated over the day.

Catchment Modeling: The RS technology is effective in obtaining inputs to catchment models, e.g., area, channel network, land use, and soil cover, etc. Some future satellites are are being especially designed for cartographic purposes and the imageries from these will be useful to prepare accurate contour maps. Thus, all the spatial and land use data which are typically used by the catchment model will be available through RS. The United States Geological Survey has developed land use/land cover classification system for use with RS data (Table 3.2). A number of software are also being developed to analyze imageries and extract the data necessary for a particular model. Since the RS data are available pixel wise which is nothing but the raster format of the GIS data, RS and GIS are used many times in conjunction to exploit the strength of both these techniques.

Flood Mapping and Monitoring

The RS data can be used to map the flood-inundated area in near real-time. The data of meteorological satellites, such as NOAA/AVHRR, INSAT/VHRR, and GMS, have helped in improved rainfall forecasting. The ground conditions rapidly change during floods and it is necessary to have repetitive data. With the availability of a number of satellites, the images can now be obtained at an interval of a few days. However, during flood periods, the clouds may inhibit data collection and it may be necessary to use microwave data.

Other Applications

Drought is a creeping water related hazard that affects many parts of the world. RS is a convenient tool to predict drought. This prediction is based on vegetation mapping. Drought causes structural changes in vegetation which affect its spectral reflectance. Temporal monitoring of vegetation indices in a region and comparison with the values of a normal year will reveal the onset of drought and its areal extent. In India, agricultural drought assessment programs have been launched that use ground data along with the data of IRS satellites, NOAA AVHRR (advanced very high resolution radiometer), and SAR (synthetic aperture radar) to monitor droughts (Jayaseelan and Rao, 1997). Nagarajan (2000) monitored land use and land cover features of an area using IRS satellite data to evaluate crop water and human demand and assess the probability of a drought event. RS application for assessment of

reservoir sedimentation is discussed in Chapter 12. An array of recent RS applications in hydrology have been included in a special issue of the journal *Hydrologic Processes* (Vol. 16, No. 8, 2002).

Table 3.2 Land use/land cover classification system for use with RS data by the United States Geological Survey.

Level I	Level II	
1. Urban or Built-Up Land	11	Residential
	12	Commercial and Services
	13	Industrial
	14	Transportation, Communications, and Utilities
	15	Industrial and Commercial Complexes
	16	Mixed Urban or Built-up Land
	17	Other Urban or Built-up Land
2. Agricultural Land	21	Cropland and Pasture
	22	Orchards, Groves, Vineyards, Nurseries, and Ornamental Horticultural Areas
	23	Confined Feeding Operation
	24	Other Agricultural Land
3. Rangeland	31	Herbaceous Rangeland
	32	Shrubs and Brush Rangeland
	33	Mixed Rangeland
4. Forest Land	41	Deciduous Forest Land
	42	Evergreen Forest Land
	43	Mixed Forest Land
5. Water	51	Streams and Canals
	52	Lakes
	53	Reservoirs
	54	Bays and Estuaries
6. Wetland	61	Forested Wetland
	62	Nonforested Wetland
7. Barren Land	71	Dry Salt Flats
	72	Beaches
	73	Sandy Areas other than Beaches
	74	Bare Exposed Rock
	75	Strip Mines, Quarries, and Gravel Pits
	76	Transitional Areas
	77	Mixed Barren Land
8. Tundra	81	Shrub and Brush Tundra
	82	Herbaceous Tundra
	83	Bare Ground Tundra
	84	Wet Tundra
9. Perennial Snow or Ice	91	Perennial Snowfields
	92	Glaciers

A number of new satellites with smaller pixel size and larger number of spectral bands would be launched in the years to come. Besides many other applications, these will provide much needed useful input to water resources studies. This field is bound to see a rapid development.

3.1.9 Cost of Remote Sensing Analysis

The cost of RS data analysis can be split into three components: hardware and software cost, data cost, and staff and other costs. The hardware prices are monotonously falling all over the world. Current (2002) estimates show that a medium capability PC is available for about Rs. 50000 (US $ 1100). Prices of peripherals such as a digitizer, scanner, and printers are extra. An image analysis software under Windows operating system will cost several hundred-thousands rupees. Note that these costs are one-time only, although the rapid technological advances tempt 'upgrading' the set-up after a few years. One scene of IRS-1C satellite covering 141km x 141km with a pixel size of 23.5m can be had for Rs. 14000. Trained staff to do routine analysis can be employed at the same order of monthly remuneration.

3.2 GEOGRAPHIC INFORMATION SYSTEMS

Conventionally, mapping, map analysis, and measurements were done manually. With the advent of the computer technology, software were written to handle geographic data on computers. This has culminated in development of Geographic Information Systems (GIS). A GIS, as the name suggests, is a system for input, storage, analysis, and output of geographic data or maps. While maps have been the most common conventional form of representing topography, the advent of digital maps in GIS provides an alternate method of storing and retrieving this information. The general configuration of a GIS is given in Fig. 3.8. The movement of water is linked to processes at the earth's surface and is influenced by them. In analysis and management of water projects, maps are frequently used. Besides, many other spatial data, such as the limits of community habitation and reserved areas for wild life, are important for water resources management but are not appropriately accounted for in the absence of a mechanism to do so. A GIS can link land cover data to topographic data and to other information concerning processes and properties related to geographic location. In this way, a GIS can be effective in such studies. The effective use of GIS depends on many factors, viz., proper problem formulation, availability of data of desired accuracy, availability of software, hardware, trained manpower, etc. Some commercially available GISs are: Arc/Info, ERDAS Imagine, ILWIS (Integrated Land and Water Information System), IDRISI, GRASS (Geographic Resource Analysis Support System), etc. A comprehensive list of major GISs has been provided by Mendicino (1996). Johnston (1998) lists many sources of GIS data, including some Internet sites. Answers to Frequently Asked Questions (FAQs) are available at many sites, such as ftp.census.gov/pub/geo/gis-faq.txt. In addition, there are newsgroups and mailing lists. Since Internet is a dynamic medium where new sites are always coming up and some old sites closing, it is helpful to do a search through search engines/sites like Yahoo!, Google etc.

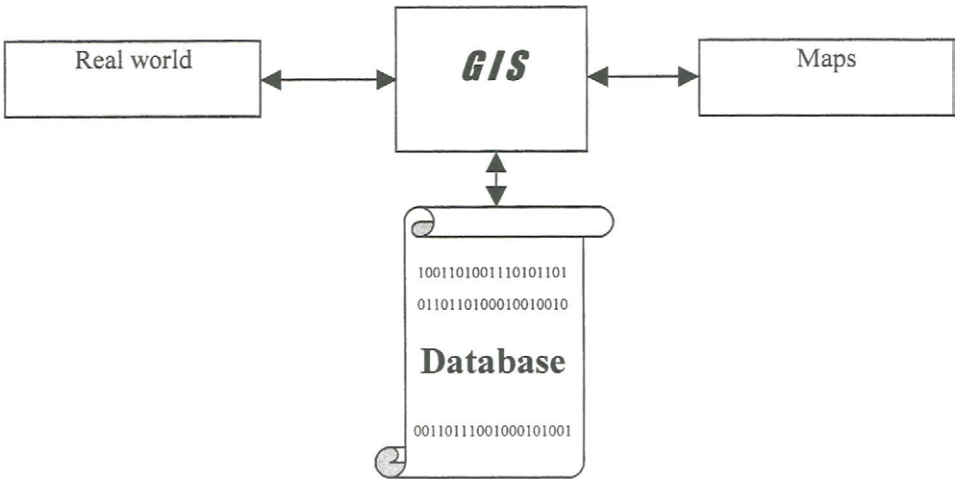


Fig. 3.8 General configuration of a GIS.

Most resource management problems have a spatial dimension. In the environmental modeling domain, this is handled by spatially distributed models that describe the phenomena in one (for example in river flow models), two (overland flow and water quality models), or three dimensions (atmosphere circulation water models). The increasing use of spatially distributed models in place of spatially lumped models is, to some extent, motivated by the wider availability of powerful and affordable computers (Loucks and Fedra, 1987). Developments in the GIS technology are important from the point of view of water resources engineering.

In GIS, the basic concept is of location, spatial distribution and relationship; the basic elements are spatial objects. In resource modeling, the basic concept is of state, expressed in terms of numbers, mass, or energy, of interaction and dynamics; the basic elements are biological species, chemicals, and environmental media, such as air, water or sediment. The overlap and relationship is apparent, and thus the integration of these two fields of research, technologies, or sets of methods, is an obvious and promising idea (Fedra, 1994).

When applied to water resources systems, nontopographic information can include description of soils, land use, ground cover, ground water conditions, as well as man-made systems and their characteristics on or below the land surface. Description of topography is called terrain modelling, and because of the tendency of surface water to flow downhill, the hydrologic importance of terrain modeling is clear. The characteristic that differentiates a GIS from general computer mapping or drawing systems is the link to the information database. Once the database is constructed, the analysis and preparation of output maps is easy. Some GISs, such as GRASS, allows treatment of large data quantities and contains a number of routines which are useful for water resources problems. The major advantage of such a GIS is that it is possible to write a routine for the analysis and integrate it with the GIS. This is a very

welcome flexibility because an existing and tested routine can be directly used.

Spatial data can be visualised as a template composed of cells and/or points and lines, to which specific information is linked. A map can be considered to be composed of several layers (Fig. 3.9), each containing the spatial structure associated with a single category, such as channel network, soils, and land cover. If one of the layers contains the boundary of the basin, spatial operators can be used to define boundaries in the remaining layers as well as to define specific relationship among them.

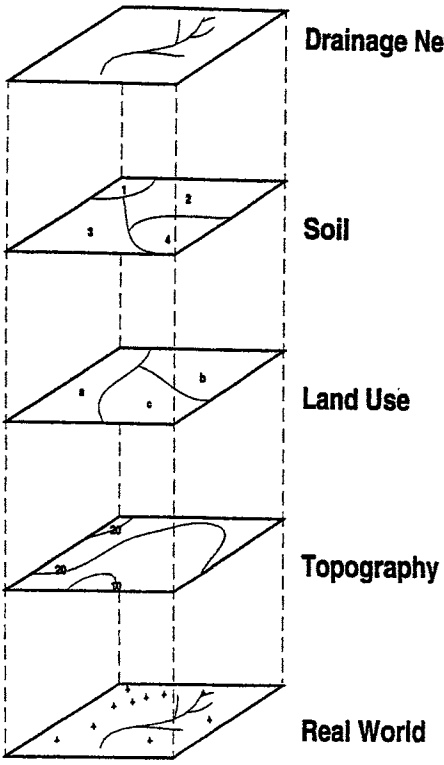


Fig. 3.9 Categorical approach for handling spatial data.

To understand GIS, it is helpful to first discuss the geographic data types, data models, etc.

3.2.1 Geographic Data Types

Geographic or spatial data have certain characteristics which make them different from other data used in water resources and other areas. The spatial objects or features have some spatial *attributes* associated with them. The main features of spatial data are: geographic location, attributes and time index. These indicate where the object is, what it is, and when it existed. The position of a terrain feature is specified with respect to a certain datum. Usually the geographic location of a point on the earth's surface is fixed by its latitude, longitude, and

altitude. In addition, a cartesian co-ordinate system or a grid system can also be used. The position of an object can also be specified with respect to a landmark in the region.

The attributes of a spatial feature define its characteristics. The attribute can be the class to which the object belongs, e.g., water body or forest; it can be an ordinal number; or it can be some measurement associated with the feature. In the last case, mathematical operations are possible on the data. The attributes of geographic data change with time and therefore it is necessary to note the time to which the attribute refers. Here, time can be specified by the calendar date, e.g., April 18, 1999; it can be some time duration, e.g., daily rainfall, or it can be a frequency, e.g., gauge data measured twice a day.

In the GIS parlance, the real-word phenomena are *entities* while their digital representations are *objects*. On a map, entities are represented as zero-, one-, or two-dimensional spatial objects. In an effort to standardise the data structure and storage, and facilitate and systematise data transfer among the various GISs, the Spatial Data Transfer Standard (NDTS) was developed by National Institute of Standards and Technology (NIST, 1992). Spatial objects are graphical elements that are used to represent spatial phenomena in a map. These objects can be aggregated into more complex spatial objects. A summary of simple spatial objects is given in Fig. 3.10.

The geographic data can be classified into four types: points, lines, polygons and surfaces.

Point data: A point is an object with zero-dimension which has some position but no size. Its position is specified by some co-ordinate system and some attributes are associated with it. A point may represent a village, a city, a stream gauging station, a rain gauge station, or a well. In a raster database, a point refers to one cell.

Line data: A line is a one-dimensional object of certain length. A line may represent a road, a river, a canal or an electrical power line. Depending on the object to which a line refers, it has attributes associated with it. A contour line has some altitudes, a river may have certain width, depth and hydraulic properties. In case of a river network, each segment may have a different order according to some system, e.g., Strahler order.

Polygons: Polygons are two-dimensional objects with certain length, width and area. These are the most common data types in a geographical database. The boundaries of a polygon may be natural features or these may be artificial. Some examples of polygons are areas of the same soil, a reservoir, a watershed, a forest or a country. The boundaries of a catchment are natural while the boundaries of an administrative unit, such as a district, are man-made. In case two polygons are adjacent to each other, they will have shared boundaries. A polygon may also be enclosed within another polygon, e.g., a small island enclosed within a large lake. The most common polygon in water studies is a catchment. The attributes associated with such a polygon are soil types, land cover, geology, and drainage network.

Continuous surface: A continuous surface represents the altitude values of a polygon. For example, the variation of rainfall or piezometer levels can be represented as a continuous

surface. A continuous surface is normally constructed using the data of point measurements. These surfaces help in better visualization of the underlying process.

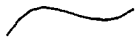
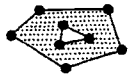
Number of Dimensions	Geometry	Geometry and Topology	Graphical Representation	Purpose
Zero	Point			Geometric location of point features
		Node		Topological junction
One	Line segment			Direct line between two points
	String			Connected nonbranching sequence of line segments
	Arc			Curve defined by a mathematical expression
		Link		Topological connection between nodes
		Chain		Sequence of nonintersecting lines and/or arcs bounded by nodes
	G-ring			Sequence of nonintersecting strings and/or arcs, with closure
		GT-ring		Ring created with chains
Two	Interior area			Area not including its boundary
	G-polygon			Area bounded by one outer G-ring
		GT-polygon		Area bounded by one or more GT-rings
	pixel			Smallest nondivisible element of a digital image
	grid cell			Smallest nondivisible element of a grid

Fig. 3.10 Summary of simple spatial objects in the SDTS model (NIST, 1992).

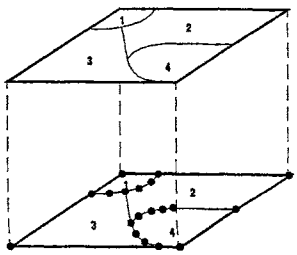
3.2.2 GIS Data Structure

The approach used to represent geographic data in GIS is defined as data model. The spatial behavior of the four data types discussed above is represented using two basic models: vector and raster. The actual method of storing the spatial data on the computer is known as data structure. There are many data structures in use for vector and raster data. These are discussed below.

Vector Data: A vector is a quantity which has a starting point, a magnitude, and a direction. In a vector model, objects are created by connecting two adjacent points by a straight line. A point is a one-dimensional object that specifies location and a node is a point that is a junction of two or more links. The curve joining two nodes is called an arc or a segment. A polygon is made up of arcs or segments. Areas can be created by enclosing a region by these lines to form a polygon. The conventional form of representing topography is contour line mapping. The contours can be digitally represented as a set of point-to-point vectors of a common elevation. When an entire map is stored in digital form, it is called a digital line graph (DLG).

Vector objects can be assigned several attributes, usually associated with an attribute table. Vector spatial data structures provide a compact but complex way to conceptualise geographic objects. As a result, data files are smaller in size. These are also very efficient in modeling topology. Three popular structures for handling vector data are topological, spaghetti, and TIN.

Spaghetti: This type has its origin in drawing software. In this structure, lines are drawn as seen in drawings, maps, etc. A polygon is defined as a sequence of (x, y) ordinates which represent a closed area. An example of this data structure is shown in Fig. 3.11. Strings are not inter-connected, even though they may appear to be so in the map. The data structure of this model is simple. However, this structure has many drawbacks and is used for simple applications only.



Original Soil Map

Data Structure

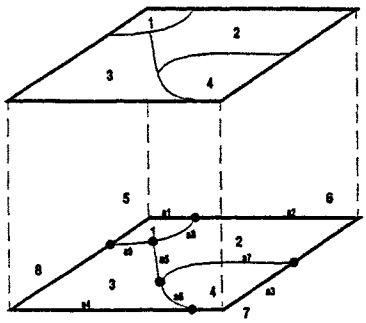
Feature	No.	Coordinate data
G-ring	1	$X_1, Y_1, X_2, Y_2, \ldots\ldots, X_1, Y_1$
G-ring	2	$X_1, Y_1, X_2, Y_2, \ldots\ldots, X_1, Y_1$
G-ring	3	$X_1, Y_1, X_2, Y_2, \ldots\ldots, X_1, Y_1$
G-ring	4	$X_1, Y_1, X_2, Y_2, \ldots\ldots, X_1, Y_1$

Data Encoding

Fig. 3.11 Spaghetti data structure.

Topologic Data Structure: Topology is defined as the transformation of geometric computations and the mathematical study of the relationships. In this structure, all lines are explicitly linked and the spatial relationships between entities are explicitly defined. The basic entity is a segment of straight line which is defined by the coordinates of its end points. An arc is a sequence of straight lines which begin and end at a node. The location value of starting and ending points of the areas are the same. The points where two lines join are called nodes.

With this technique, spatial relationships among entities, mainly connectivity, are explicitly recorded. In the general case, polygons may be formed using several connected chains or arcs, as shown in the example of Fig. 3.12. The database contains information about the points in an arc, the arcs meeting at a node, and the arcs forming a polygon. The names of the polygons on both sides of the line are also stored. In the database, spatial relationships among entities, mainly connectivity, are explicitly stored. Adjacency is defined based on the polygons that surround every chain. As with other data structures, tables can be constructed to describe coordinates, connectivity land adjacency characteristics. In Fig. 3.12, these tables are represented as summarized lists. The polygon-arc list includes the arcs associated with each polygon, even those that are common to more than one polygon. This list can be related to the arc-coordinate list to determine the coordinates of all points that constitute each polygon. The topology list defines the adjacency characteristics of all arcs.



Polygon-Arc List

Polygon	Arcs
1	a1, a8, a9
2	a2, a7, a5, a8
3	a5, a6, a4, a9
4	a7, a3, a6

Arc Topology List

Arc	Left pol.	Right pol.
a1	5	1
a2	6	2
a3	7	4
a4	8	3
a5	2	3
a6	4	3
a7	2	4
a8	1	2
a9	1	3

Arc Coordinate List

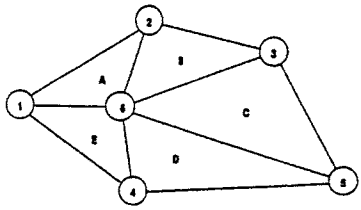
Arc	X, Y pairs
a1	0,4 0,6 1,6
a2	1,6 6,6 6,3
a3	6,3 6,0 ...
a4	5,0 0,0 ...
a5	1,5 3,2
a6	3,2 4,1...
a7	3,2 3,3 ...
a8	1,5 2,5 ...
a9	0,4 1,4 1,5

Fig. 3.12 Topologic data structure.

Topologic data structures are more sophisticated than spaghetti data structures and are included in practically all vector-based GIS. Their ability to handle topology makes

them well suited for contiguity and connectivity analysis (Aronoff, 1989), both of which have ample applications in hydrology. Contiguity capabilities, for example, can be applied to determine flood impact zones, correlations between soil types and land use, and others. Connectivity capabilities can be applied to drainage network and transportation analysis, emergency evacuation routes, and so on. Like spaghetti data structures, however, topologic data structures do not constitute the best solution for representing surfaces and transition zones.

Triangular Irregular Network (TIN): Triangular Irregular Networks (TINs) constitute a special case of topologic data structures in which nodes are interconnected using single links (Fig. 3.13), resulting in a set of triangular facets that covers the area of interest. The location of each node is given by its X, Y, and Z coordinates, which provide TINs with a powerful tool to simulate highly variable surfaces. The TIN topology is based on triangle adjacency and the definition of the nodes that compose each triangle.



Node Coordinates

Node	Coordinates
1	X_1, Y_1
2	X_2, Y_2
3	X_3, Y_3
4	X_4, Y_4
5	X_5, Y_5

Triangle Adjacency

Triangle	Adjacent Triangle
A	B, E
B	A, C
C	B, D
D	C, E
E	A, D

Triangle Nodes

Triangle	Nodes
A	1, 2, 6
B	2, 3, 6
C	3, 5, 6
D	4, 5, 6
E	1, 4, 6

Z Coordinates

Node	Value
1	Z_1
2	Z_2
3	Z_3
4	Z_4
5	Z_5

Fig. 3.13 Triangular irregular networks.

The number of possibilities for defining and locating nodes is nearly endless. As a result, many algorithms and techniques have been devised to optimize the process of TIN generation. Many GIS contain built-in routines. Customized procedures are continuously being developed for particular applications. In general, algorithms begin with a regular mesh or uniform grid of triangular elements upon which a series of point and/or linear well-

defined features are drawn. Point features include peaks and wells, among others, and are maintained in the TIN as triangle nodes. Linear features include channel networks, lake shorelines, etc., and are maintained in the TIN as additional triangle edges. The algorithm then takes the location of these additional nodes and edges, and subdivides the original grid into finer triangles, causing, a denser spatial resolution. The shape and spatial resolution of the original mesh may actually be the result of an interpolation process based on contour data or a digital elevation model.

TINs are well suited for representing surface spatial variability. Raster data structures can also be used for this purpose, although their applicability may not necessarily be overlapping. TINs usually provide a more compact representation for the same level of accuracy. On the other hand, raster data structures are more suited for the evaluation of parameters, such as slope and aspect, which work best when systematic sampling is used.

Raster Data

In the raster structure, the map area is divided into a number of cells (or pixels, in a satellite image). Each cell represents an area with uniform properties. Most raster implementations are built around square or rectangular grid cells, but other shapes, such as triangular and hexagonal, are also possible. Since grid cells are uniform, these can be referenced by their row and column numbers. Strictly speaking, raster spatial data structures are two-dimensional arrays. This means that the area that each grid cell represents can be used both to define map resolution and the number of grid cells needed to describe the spatial distribution of the attribute under study. Each grid cell may be assigned only one value, which represents the value of the attribute being mapped. If more than one attribute is to be handled, data layers or overlays, as shown in Fig. 3.14, need to be created. These layers can be visualized as stacked, one on top of the other.

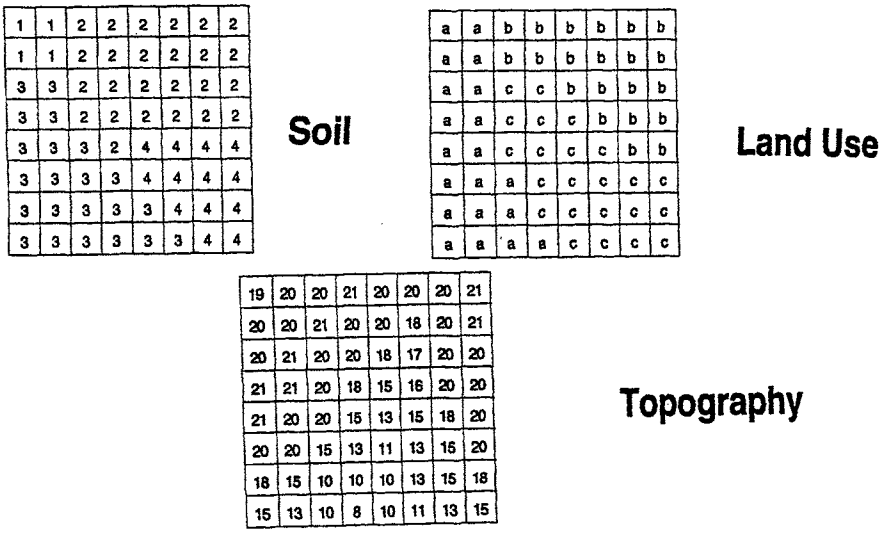


Fig. 3.14 Raster Representation of the Soil, Land Use, and topography categories shown in Fig. 3.9.

Initial applications of GIS in water resources were mostly based on raster data. The grid is made up of regularly spaced lines, and the enclosed area of each rectangle is described in terms of its center coordinates. If the terrain is a visual image with the dots having various colors and intensities similar to a computer video screen, the use of the term raster image used for grid data as well as computer screen images is easily understood. The GIS GRASS is an example of a widely used raster-based GIS developed by the U.S. Army Corps of Engineers. This is a public domain package and its details are available on Internet at http://www3.baylor.edu/grass.

The size of a GIS database can be very large and several methods have been devised to reduce the storage requirement. In the regular tessellation, the size of cells is uniform which is not an efficient way to store data when there is less variation in the features. The important raster models of arranging data are regular tessellation, nested tessellation, and irregular tessellation. In the regular tessellation, the data are arranged in arrays. The square grid is the most commonly used arrangement although triangular and hexagonal meshes are also used. In the square grid, raster data are arranged in arrays; each cell can be referred to by its row and column number. The referencing system of grid raster data is very convenient and its interfacing with hardware is also simpler. In triangular tessellation, all the cells do not have the same orientation. Due to this, the comparison of data is complex but this structure is very efficient in representing topography. The size of the database can be reduced if the size of a cell depends on the variation in the surface features -- the cell size can be small where the variation is more and vice versa. This concept is used in a nested tessellation model in which the cells are recursively subdivided into smaller cells of the same shape and orientation.

Raster spatial data structures are simple to conceptualize and use for overlay analysis. Further, they can be efficiently used to model high spatial variability, which makes them appealing for the manipulation of remotely sensed data and digital images. On the negative side, data files tend to be huge, although data compression techniques can reduce this burden. The size of the cell denotes the resolution of the data, bigger is the size, coarser is the resolution. Note that the size of the data files increases exponentially when higher resolution is used. The number of grid cells needed is inversely proportional to the square of the resolution and, consequently, the need to develop methods to minimize storage requirements is crucial. Most data compression methods are based on the fact that adjacent grid cells often have the same attribute value and, as a result, it is better to group them under the same category. Two such methods, quadtrees and run-length encoding, are described below.

Quadtrees. Quadtrees are perhaps the most common structure used to represent raster data. In general, the technique is based on a successive division of the map or image into quadrants until every subdivision can be assumed to be spatially homogeneous, i.e., one single attribute value can be associated with it. In practical terms, grid cells can adopt variable sizes, depending on the number and spatial distribution of the original grid cells having the same attribute value. However, the analysis of this type of data is complex.

Fig. 3.15 shows a sample region and its quadtree representation. For easy

visualization, a node with children is represented by a circle; a node without children and no attribute values is represented by a blank square; and a node with an attributed value is represented by a hashed square. NN, NE, SW and SE are used to designate quadrants even though a numerical coding sequence such as 1, 2, 3, is more convenient.

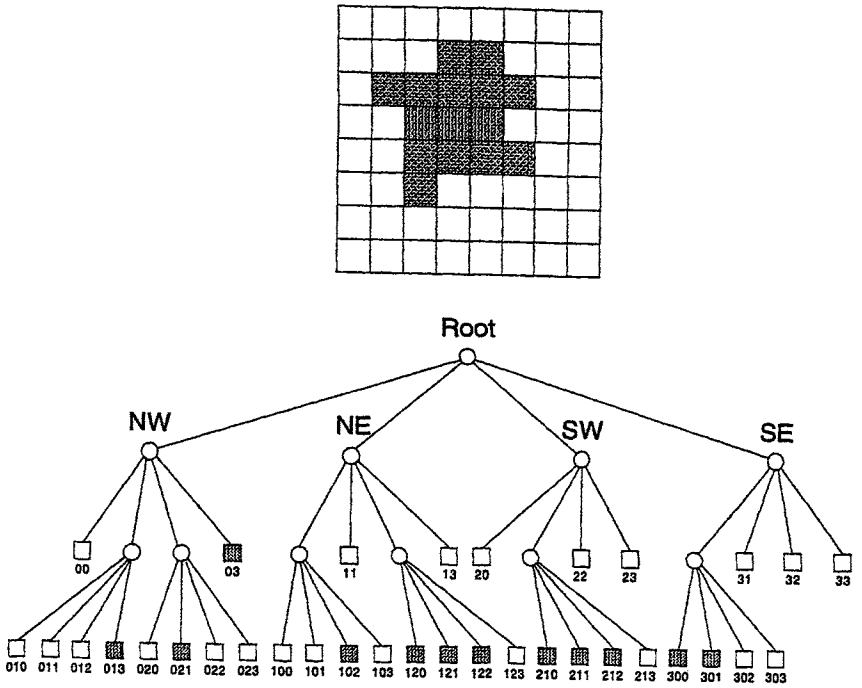


Fig. 3.15 Sample region and its quadtree representation.

Run-Length Encoding. With this technique, grid cells having the same attribute value are organized into blocks or runs. A table can be formed to describe, for each row, the number of consecutive grid cells that are associated with the same attribute value. Table 3.2 shows, as an example, the run-length encoding that would correspond to category 'soils' in Fig. 3.14.

Table 3.2 Run-Length Encoding of the Raster Spatial Data Structure of Fig. 3.14.

Row	Value	Length	Row	Value	Length	Row	Value	Length
1	1	2	4	3	2	6	4	4
1	2	6	4	2	6	7	3	5
2	1	2	5	3	3	7	4	3
2	2	6	5	2	1	8	3	6
3	3	2	5	4	4	8	4	2
3	2	6	6	3	4			

Comparison of Raster and Vector Models

Most commercially available GISs have the ability to transform between DLGs, grid DEMs, and TIN DEMs, but contour-based methods require an order of magnitude more data storage, so that the transformation is typically from DLGs to the other forms. Both vector and raster models have some advantages as well as disadvantages. The vector data are compact, accurate and are widely used to describe polygons. The lines and boundaries are represented in a better manner using a vector data model. The vector models are more suited for network analysis. However, the topographic analysis is difficult because the data structure is complex and each polygon has a different topological form. Further, overlaying of these maps is difficult. The structure of a raster model is simple, the analysis is also simple because each cell has the same size and shape. There is loss of information when data are changed to a coarse resolution. Also, when raster is converted to vector, the lines and boundaries appear as staircase. The integration of this data with satellite data is easy. Its major disadvantage is the requirement of large computer storage. However, with development of high capacity and cheaper storage devices, this no longer appears to be a major limitation.

It is necessary that GIS for hydrological purposes have the capability of dealing with raster as well as vector data. The remote sensing analysis involved large quantities of multi-temporal raster data. The vector data are also analysed in hydrological modelling, e.g., delineation of drainage basin boundaries, automatic generation of drainage map in a catchment, etc. Most current GIS packages have modules to convert raster data into vector data or vice versa.

Time in GIS

Most geographic databases are built assuming static geographic feature conditions. However, geographic features evolve through time. Many of these change as a result of natural processes, but they also change as a result of human intervention. Land cover, land use, land slope, and stream network geometry, including alignment, slope and cross section, are just some examples of landscape characteristics that may change over time and that may have a direct impact on watershed hydrologic modeling.

Unfortunately, current GIS are normally atemporal in that they describe only one data state. Because most GIS in existence are based on the categorical approach, they are not capable of providing explicit linkages for objects through time. However, a reasonable goal for GIS would be to be able to respond to queries, such as what, where, when, how fast, and how often changes have taken place. Many applications in engineering, in general, and in water resources, in particular, would surely benefit from having a system capable of answering such types of queries.

3.2.3 Geographic Coordinate Systems

The position of any point on the earth's surface is commonly expressed in terms of its latitude and longitude. The latitude of a point is the angle measured at the centre of the

earth between the plane of the equator and the line connecting the point with the centre of the earth. If the point lies in the northern hemisphere, the letter N is suffixed to the angle and the letter S is suffixed if the point lies in the southern hemisphere. The latitudes vary between 0° (equator) and 90° (poles). The meridian passing through Greenwich (U.K.) is the reference meridian for longitude and is termed the prime meridian. The horizontal angle between the prime meridian and the line connecting a point with the earth's centre is the longitude of that point. By convention, E is suffixed for the points lying to the east of the prime meridian and W for those lying to the west. The longitudes vary between 0 and 180°. Both latitudes and longitudes are expressed in degrees, minutes, and seconds. For example, the longitude of a point may be 79° 30' 45" E and latitude 29° 19' 55" N.

Map Projection

Although the shape of the earth is considered to be spherical, actually the equatorial axis of the earth is longer than its polar axis or the earth is somewhat flattened near the poles. This shape of the earth is close to an ellipsoid or spheroid which is considered to be a better representation of the earth's shape. A spherical representation of the earth, for example, a globe is not convenient to use and work; it cannot be included as part of a book or a report. Therefore, it is necessary to use a map projection to project the 3-dimensional surface of the earth on a 2-dimensional flat surface. Conceptually, the earth's surface is typically projected on cylinders and cones. These surfaces are 'opened' and stretched to conform to the shape of a plain paper.

A number of transformations are employed in map projection which can be mathematically represented as:

$$x = f(\phi,\theta) \tag{3.13}$$
$$y = g(\phi,\theta) \tag{3.14}$$

where ϕ and θ are the latitude and longitude of the place.

There are three basic classes of map projections: the cylindrical projection, the conical projection, and the azimuthal projection. In cylindrical projection, it is visualized that a cylindrical shaped paper is wrapped around the globe. The globe is opened up and stretched so as to conform to the cylindrical shape. If this cylinder is opened, the map of the world on a rectangular sheet will be obtained. In case of conical projections, the globe is visualized to be wrapped by a cone and projection is taken. When the cone is opened, the fan-shaped map of the world will be obtained. In azimuthal projection, the spherical earth is projected on a plane tangent to the globe.

A map projection is named according to its class and property as well as the name of the originator and the nature of the modification of earth's surface. Some well-known projections are Mercator projection, Normal Cylindrical Equal Area (Lambert projection), Polyconic projection, etc. A few commonly used map projections are described here.

Polyconic

This projection system was developed by Ferdinand Hassler in 1820. It is suitable for north-south oriented maps. It uses an infinite number of cones as developing surfaces. The parameters needed for the projection are central meridian, false easting, and northing. False northing is assigned a value of zero. The false easting is assigned a value of 500,000 m. Central meridian is generally a meridian passing through the center of the study area. The parallels are circular arcs and the meridians are radial lines. This projection is used by the Survey of India for topographic maps at the scales of 1:250,000 and 1:50,000.

Universal Transverse Mercator (UTM)

The Universal Transverse Mercator (UTM) projection, developed by the U.S. Army, is widely used in topographic maps. This projection is recommended for areas lying between 84°N to 80°S. In UTM, the earth surface is divided in 60 zones, each 6° wide in the longitudal direction resulting in rectangular graticule mesh (Fig. 3.16). These are numbered sequentially from west to east. The western edge of the first zone touches the 180° W meridian and the eastern edge of the 60th zone touches the 180° E meridian. In the latitude direction, each zone covers an area of 8° (except the northern most zone that covers 12°). The bottom-most zone (80°S to 72°S) is assigned letter C and the topmost letter X. The origin of each zone is located at a point at the equator where it is intersected by the central meridian of the zone. The eastings of the origin of each zone is 500,000 m. Regarding northings, for the northern hemisphere it is 0 at the equator and for southern hemisphere, it is 1,000,000 m at the equator.

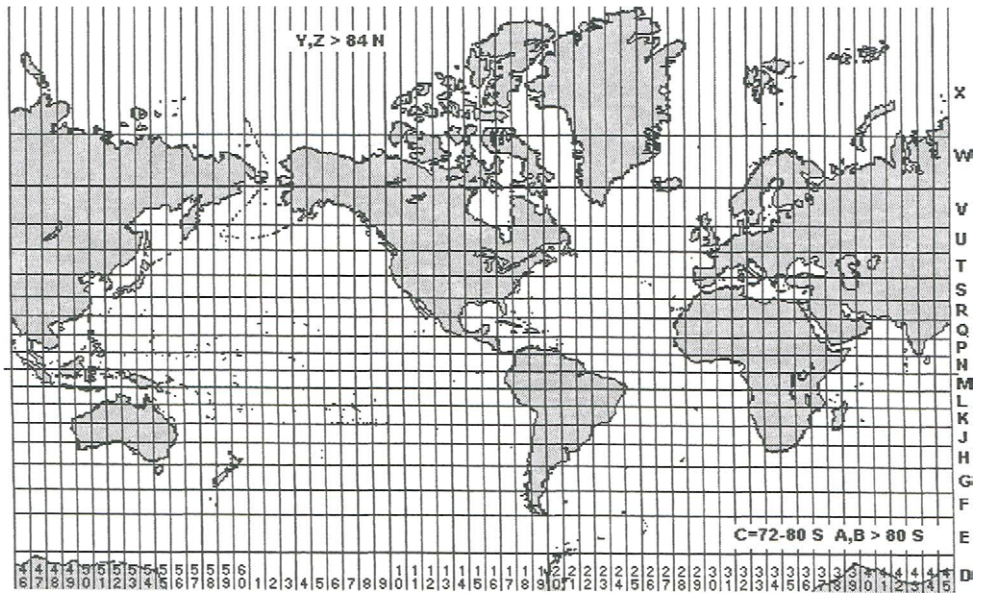


Fig. 3.16 The UTM map projection.

Albers Equal Area Conic

This projection is used for small-scale maps. It requires two standard parallels.

Note that all maps have some kind of distortion and while choosing a map projection, the aim is to select the projection in which the distortions are within acceptable limits. The following general guidelines are followed while choosing a projection:

a) For the countries lying in tropics, a cylindrical projection is to be used.
b) If the country lies in temperate zone (between latitudes 23.5° to 66°), a conical projection should be used.
c) If the region is between 66° latitude and pole, an azimuthal projection should be used.

The above is a very simplified treatment of a complex topic. For further details, a good book on geography and coordinate systems, for example, by Maling (1992), may be referred to.

3.2.4 GIS-User Interface

A user interface is the way a user interacts with a GIS software to perform certain tasks. An interface is a link between computer and the user. The interface is an important part of a GIS because a GIS is used by people with diverse backgrounds and many of them have little computer experience. Therefore, the interface should be such that it is easy and convenient for a typical user to use the system and exploit all its strengths. There are many standard interfaces, viz., menu, icon, command line, script etc.

In a menu-based interface, the user works with cascading menu. A branch structure exists wherein from the main menu, submenus open. The actual function or operation resides at the end of the tree. In a menu, the operations are grouped based on their closeness in operations. For example, the functions dealing with a file operation (opening, creating, copying, etc) many be grouped in one menu group. Icons are interface methods in which graphical elements are used to pictorially depict the operations that they will launch. For example  indicates file opening under the Windows operating system. Under the command lines interface, the user types in a command and it is executed on pressing the return key. The optional parameters can be supplied in the command line itself. In the scripts or model mode, many commands can be written and executed as a batch. Most of the packages provide a macro language which is the programming language for that particular package and can be executed from within the package.

3.2.5 Steps of GIS Based Analysis

The major steps of any analysis using a GIS are discussed below.

Data Preparation

Data preparation requires identification of data and its source, scale, geographic extent of

data, and locating an area of interest, etc. Data identification involves deciding what data are needed and it depends on the problem. Source and scale of data will be dependent on many factors, e.g., availability, storage space, processing time, computer, areal extent, etc. The geographic extent is known a priori and the boundaries are drawn on a map. It is helpful to gather requisite information for the focus area prior to the analysis in a GIS.

Data required for a GIS can be collected from ground surveys, digitizing existing maps, digitally recorded aerial photography, satellite imaging data, or combinations of thereof. A problem of the scale of accuracy arises when these data are used in combination.

In a typical GIS application, the availability of spatial database is the first problem, particularly in developing countries. In many cases, one may have to start from scratch. Note that the acquisition and compilation of the information is hardly a trivial exercise. Generally, the requisite data are available only as hardcopy map. Even after important developments in digitizing hardware and software, the process is labour intensive and requires skilled manpower, time, and patience.

Data Input

Data are input to GIS in many ways, namely, using digitizers, scanners, computer files, etc. Digitizers are computer devices that use a tablet and cursor to convert a paper map to a digital one. A tablet contains closely spaced mesh of wires below a flat surface over which the map is placed. The cursor is moved over the map and by pressing buttons on the cursor, coordinates of current location are sent to the computer. This is the most common method of inputting a map to a GIS but requires a lot of time and patience. While digitizing a map, many errors can occur, common amongst them are:

1. Dangle or dead end in segment: A polygon in a GIS should be a closed area. Error occurs when it is not so and this may be due to presence of a dangle or dead end in a segment.
2. Intersection without node: At all intersections of lines, nodes should be formed, otherwise error will be generated.
3. Self-overlap: This error occurs by the erroneous movement of cursor causing a line to intersect itself or formation of loops.
4. Double digitization: If a line is digitized more than once by mistake, this causes multiple intersections without node.

The digitization modules have routines for minimizing errors, their checking and correction. The technique of auto snapping is used to avoid dangles. When two lines start or end closer than a specified distance, these are automatically joined and a node is created. This function can also be turned off. To delete the intermediate redundant points from a line segment, the *tunneling* function is used. In this, a tunnel of specified width is drawn between the first and the last points of consecutive sets of three points. The middle point is retained only if it lies outside the tunnel.

Scanners are similar to digitizers, except that the information is automatically input

to computer. A scanner can have either a moving sensor or a fixed sensor. In the first case, a map is placed on a flat bed and sensor moves over the map. In the latter case, the map is moved over the sensor. The light reflected by the map is picked up by the sensor to form an image of the map which can be interpreted by software routines to extract the desired information. After a map is scanned, it needs to be edited to remove errors before it can be used in a GIS.

Output

The final step of GIS analysis is to present the results in the form of maps for the purpose of report or a presentation. In the output stage, various maps and information can be directed to a computer monitor or a printer. The output creation involves design of layout, contents, legend, scale, etc. The map can also include scale, graticules or grids, direction arrow, nameplate (author, date and other information), texts, symbols, etc. Note that the design of maps, scale, contents, etc. is dependent on the objective of the problem. The required information can also be displayed on maps by various cartographic methods, such as pie charts, histograms, proportional circles, etc. Blending is an output method in which two raster maps can be superimposed to show both maps partially. This method can be used to overlay hazard maps on watershed maps.

The choice of colors in a map can be natural colors, e.g., water can be shown in blue color, or an artificial color scheme can be used. Sometimes, the options, such as fonts, colors, and annotation procedure, may be limited by the hardware and software features. The gray shades and symbols can be selected depending upon the variation in the output data.

3.2.6 Analysis of Geographic Data using a GIS

In data analysis, one or more geographic data layers are manipulated to extract useful information. Many manipulations and measurements can be done on the geographic data. The GIS analytical capabilities can be broadly divided into three general groups: operations that may be performed on spatial or non-spatial data, operations performed on individual spatial data layers, and operations performed on multiple spatial data layers. Despite the differences in data structure between and most of these operations can be performed with either raster data or vector structure. The important operations are described here.

Data Editing and Management: The properties of geographic features change with time due to natural and human activities – new features might be added, the existing modified or deleted. The scope and purpose of a study may also change with time. Thus, there is a frequent need to edit the existing GIS database. Typical edit operations are: add, delete, extend, and modify. The names of the objects may have to be changed. Since a user has to frequently use the edit module, it should be easy to use and powerful.

A file management function carries tasks, such as rename, delete, copy, etc. These can also be done through the operating system also but are relatively safe in the GIS environment. In general, it is better to do these tasks from the GIS itself. If a certain item is

to be removed, the GIS will also remove all the associated files and clean up memory. The delete operation is to be used with care, otherwise useful data can be lost forever.

Classification: It is a method of information retrieval from the data. This is somewhat similar to RS analysis discussed in Section 3.1.8. In classification, the value groups are created and the group numbers are assigned to each value in the output map. Examples of the classification operation are creating a contour map from a DEM, an isohyetal map from rainfall raster map, etc. Reclassification is a GIS operation in which another map is created by assigning new values to a map. For example, from a soil class map, soil-pH map can be prepared; from a 2nd order land use map, 1st order map can be prepared, etc. The operation is also called 'recoding'.

Data selection and query: A GIS can be used to select data that meet certain criteria. Selection of database records may be done: i) interactively; ii) by specifying numerical thresholds; or iii) using logical operators. Interactively, data can be selected from tables or on screen using a pointing device. Quantitative data can be selected by either specifying numeral thresholds or ranges. Boolean logic uses the following four operators (Fig. 3.17) to perform on two or more data sets: AND (intersection), OR (union), NOT (negative) and XOR (exclusionary or).

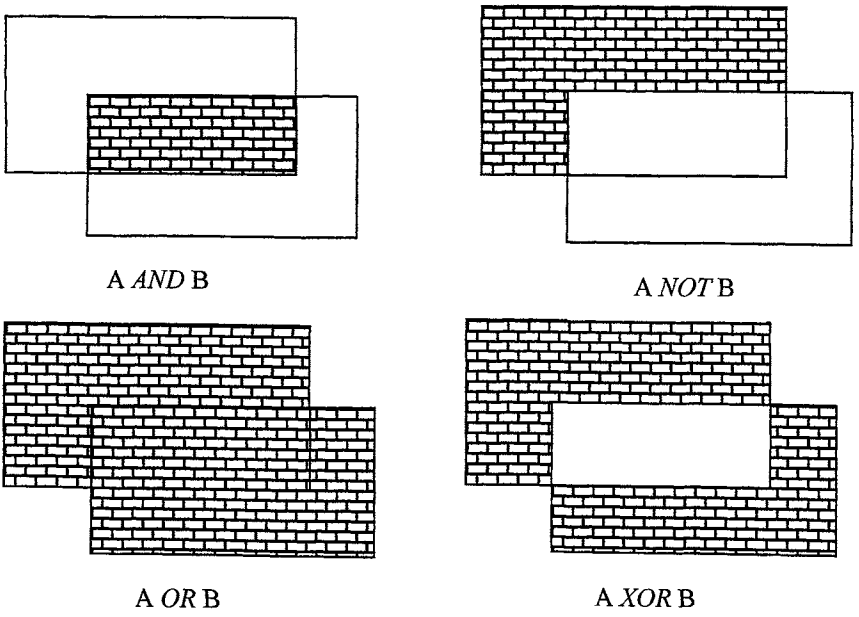


Fig. 3.17 Venn diagram showing Boolean operations.

Another important GIS operation is query. Using query, the location of a certain geographic object can be retrieved, or a certain geographic object can be displayed, etc. For example, in a river network, river with a particular name or all rivers with particular order can be highlighted. The flooded areas or particular soil or land use areas can be displayed.

Two or more maps can also be overlaid, e.g., flooded cropland may be displayed. The query results can also be saved as a map or different geographic data. Thus, information retrieval is also combined with query operation.

Intersection operation is used for overlay of two different types of geographic data, e.g., point and area or line and area. Examples of this operation are extracting ground water wells for a basin, extracting village names for the flood affected area, extracting drainage lines in a basin, etc. The output of a union operation shows the details of both input maps.

Measurements: GIS provide a range of capabilities for measurements on spatial data depending on the type of geographic data. For quantitative attributes, the statistic, like mean pixel attribute and their standard deviations can be computed. Area calculation is straight for both vector and raster databases. In a raster database, multiplying the number of target pixels by the area of one pixel gives the total area. Distance (d) between two points whose coordinates are (x_1, y_1), and (x_2, y_2) is:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{3.15}$$

Other similar operations are length of string and perimeter of an area. It is also possible to compute volume between two surfaces.

Spatial aggregation and generalization: Spatial aggregation is used to reduce excess details in a raster data layer. The user defines the numbers of cells to be aggregated in each *nxn* window, or *kernel*. This kernel should have the same orientation as the input layer and the size should be integer multiple of cells. A GIS systematically examines all *nxn* non-overlapping blocks of the input data layer, aggregates them and assigns a summary value to it (Fig. 3.18). This summary value could be the majority value (the most frequent value), average, maximum, etc. for the kernel. In case of tie, a mixed class may be created. Spatial aggregation reduces database size and may also be used to visualize patterns at a coarser grid than the original file.

A somewhat similar technique is data resampling. Here, the kernel need not be an integer number of cells and may have a different orientation and size than the input cells. Resampling is performed to convert a raster data layer to another coordinate systems or projection, or to match raster data layers having different cell sizes prior to overlaying them. Methods for determining which data value should be assigned to the new cell include nearest neighbor, bilinear interpolation, and cubic convolution. The nearest neighbor method assigns to new cell the value of its closest neighbor cell in the original data layer. The bilinear interpolator computes the new cell value from the values of the four cells surrounding the new cell.

A useful GIS operation is clumping. It is used to aggregate patches of contiguous cells with the same attribute value is a raster database. Various groups of cells that touch at a corner or edge and have the same value are identified and merged or assigned same value.

F	F	F	A	A	A	A	A
F	F	F	F	F	F	A	A
A	F	F	F	F	F	F	A
A	A	A	F	A	F	F	F
A	A	A	A	A	A	F	F
A	A	A	A	A	A	A	A
F	F	F	A	A	F	F	A
F	A	A	A	F	F	F	A

(a)

F	F	A	A
A	F	F	F
A	A	A	A
F	A	F	A

(b)

F	F	M	A
A	F	M	F
A	A	A	M
F	A	F	M

(c)

A – Agriculture, F – Forest, M – Mixed.

Fig. 3. 18. Spatial aggregation of data of two categories in (a). To resolve tie, majority rule is followed in (b) and a mixed class is created in (c).

Buffer zones: An often used capability of a GIS is to generate a buffer zone around spatial objects, such as a flood plain zone around a river, right-of-way for road/rail, etc. The inputs are the spatial entity for which the buffer is desired and the distance or width of the buffer zone. In raster systems, buffering is done by a spreading process which computes the distance between the feature of interest and every other cell within a defined limit, resulting in halos of cells with incrementally larger distances from the central feature. In some raster GISs, additional criteria can be applied to the spread command to control the direction (uphill, downhill), resistance, and barriers to spreading (Johnston 1998). These criteria may be helpful in studies, such as routing of overland flow.

Geometric transformation: A geometric transformation manipulates a data layer to correctly overlay it on another layer of the same area. The need for this arises because maps can have different coordinate systems and/or projections and these must be reconciled before they can be overlaid. Note that this operation involves at least two data layers. There can be mismatches even when both data layers have the same coordinate systems and projections. Such mismatches can be adjusted by techniques such as registration by absolute position, registration by relative position, and edge matching. Most GISs can transform data among a number of projections.

Registration by absolute position: Each data layer is transformed independently using ground coordinates, such as those obtained from topographic maps. The reference points should be carefully chosen features, such as road intersections, that can be clearly identified on the digital maps. Since each layer is independently registered, errors are not propagated from one data layer to another. Even then, small errors in various data layers may cause alignment problems.

Registration by relative position: An accurately georeferenced reference layer is chosen and the other layers are transformed based upon this layer. A prerequisite here is that the reference layer should be very accurate with clear reference points which could be easily and accurately located on both data layers. A disadvantage is that the errors in the reference layer are propagated to transformed layers.

Edge matching: The idea here is that lines and polygons that cross map boundaries should match. This mismatch is quite common in the edges of adjacent maps and data layers. Most GISs have routines for (semi) automatic edge matching.

Overlay operations: This is a very useful GIS function that allows to digitally place multiple data layers on top of each other. It is also possible to overlay a vector layer over a raster data layer. Thus, one can digitize the catchment boundary from a topographic map and overlay it on top of a satellite imagery. Similarly, on a map of catchment boundary, different layers such as soil, land cover, slope, and channel network, can be laid. This is useful in studies, such as catchment modeling, soil erosion, etc. The maps can be manipulated using mathematical and logical functions.

Aspect: Aspect is the direction which a surface (typically hillslope) faces. It is usually expressed in degrees ($0\text{-}360^0$) or as compass directions such as N, S, NE, NW etc. In a TIN database, aspect is the direction of maximum slope of the triangle plane. On a hillslope, aspect controls input of solar energy which controls snow melt, evapotranspiration, etc.

Line-of-sight maps: A GIS can be used to generate maps identifying the intervisibility of landscape features. A *viewshed* shows the land area that can be seen from a given point. A *visual impact* map shows all the area from which a tall object can be seen. Input data layers include elevation, height and location of the observation point (viewshed map) or target (visual impact map). The height and location of features projecting above the land surface that could block the view (e.g., buildings, trees) are also input (Fig. 3.19). Such analysis is useful in planning communication towers.

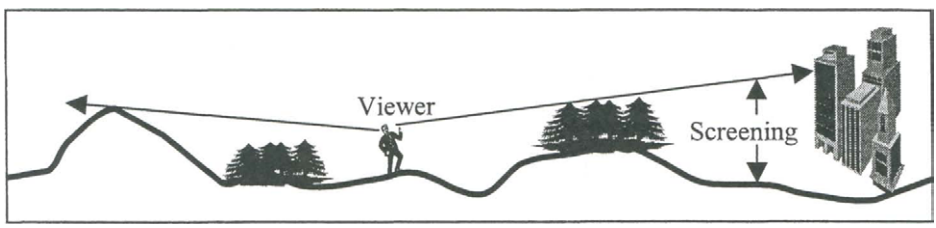


Fig. 3.19 Pictorial presentation of viewshed concept.

3.2.7 Digital Terrain Model (DTM)

The primary source of topographic information prior to the 1980s was contour maps. In these maps, elevations are represented by points and contours. Advances in digital mapping have offered essential tools to closely represent the three-dimensional nature of natural landscapes. A capability of a GIS that is most important to water resources applications is the description of the topography of a region. One such tool is the digital terrain model (DTMs). DTMs are 'ordered arrays of numbers that represent the spatial distribution of terrain attributes' (Vieux, 2001). These attributes include elevation, slope, slope steepness, and soil depth. A subset of DTMs which describes elevations above some arbitrary datum is the digital elevation model (DEM). Topographic variables, such as basin geometry, stream networks, slope, aspect, flow direction, can be extracted from DEMs. Three schemes for structuring elevation data for DEMs are: triangulated irregular networks (TIN), grid networks, and vector or contour-based networks (Moore and Grayson, 1991). Some spatial information is not directly described by elevation, and can be described as topologic data. Topologic data define how the various pieces of the region are connected. Topology is the spatial distribution of terrain attributes.

DEMs are useful to determine physical variables, such as slope and aspects that are useful inputs in various hydrological computations, e.g., overland and channel flow velocity, travel time, soil erosion, slope stability, etc. While topographic data fit within the general classification of topologic data, there are significant hydrologic attributes not related to land surface elevation. The more obvious of these are catchment areas, flow lengths, land slope, surface roughness, soil types, and land cover. These attributes help describe the ability of a region to store and transmit water.

The input data for DTM usually comes from topographic maps. Other sources could be surveys, photogrammetry, and satellite images. Special attention should be given to surface discontinuities and pits, peaks, ridges, and streams. After the data are entered in a GIS, a model is build to represent surface behaviour. The two models that are common are regular grids and TINs. In regular grids, elevation values are computed at equally spaced grid points (Fig. 3.20). In TIN, a surface is represented as a network of adjacent triangles whose vertices are the sample points.

Elevation data will rarely be available at regular grid points and interpolation will have to be carried out. DEMs are created from point and contour data using an interpolation function. The commonly used interpolation techniques include linear/bilinear interpolation, inverse weighted distance method, and kriging. The weighted average method consists of taking the weighted average of point values within a specified radius from the interpolation point. The weighted surface method fits an n-degree surface using the points within a specified radius from the point of interpolation. Kriging is also a popular interpolation technique. These have been described in Chapter 2.

The information that is useful in hydrologic analysis and can be derived from DTM includes slope, aspect, catchment boundary, and channel network. The slope can be calculated as:

$$\tan S = [(dz/dx)^2 + (dz/dy)^2]^{0.5} \tag{3.16}$$

and the aspect is defined as (Burrough, 1986):

$$\tan A = -(dz/dy)/(dz/dx) \quad -\pi < A < \pi \tag{3.17}$$

where x, and y are coordinates, and z is altitude.

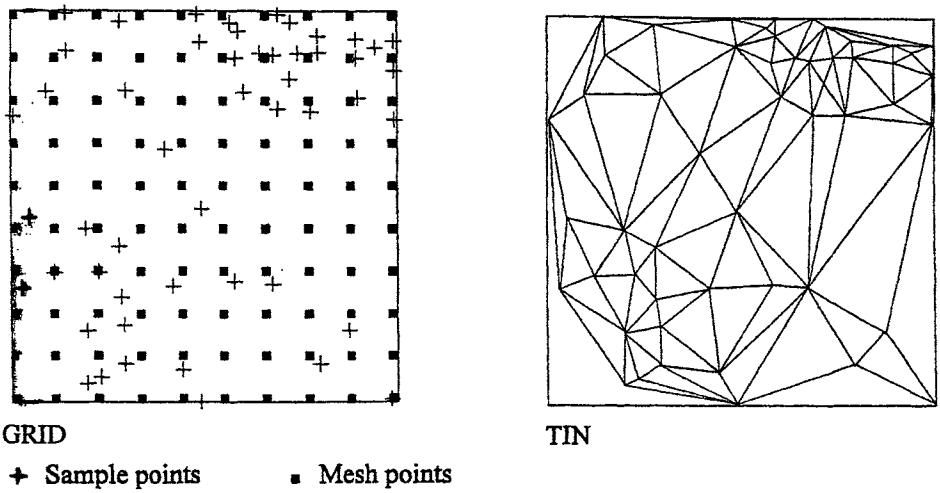


Fig 3.20 Regular grid and TIN models for DTM.

An inherent problem in hydrologic modeling with grid DEM data is the production of non-physical depressions due to the noise in the elevation data, coarse contour interval, or sparse elevation data. These affect interpolation schemes resulting in an unwanted termination of drainage paths in pits. The problem is particularly acute for relatively flat areas. O'Callaghan and Mark (1984) and Jenson and Domingue(1988) have described techniques to locate and remove depressions in gridded DEM data. The situation may, however, be complicated by the existence of naturally pitted topography, sometimes called pothole regions. Sometimes the user has to manually alter the elevations to remove pits. The pits are filled by raising their levels iteratively to nearest higher elevation points

The algorithms to demarcate catchment boundary and channel network mainly make use of slope and aspect maps and the outlet point is given by the user. Sole and Valanzano (1996) have reviewed and described several such algorithms. BASINS (Endreny 2002) is a toolkit for tabular and spatial data management with several modeling tools. It can be accessed at http://www.epa.gov/ost/BASINS.

DEM Visualization

Several techniques are used to display a DEM, e.g., contouring, hill shading, perspective view, fly through, perspective cartography, etc. Contouring is the conventional technique of elevation representation. In hill shading, gray shades are assigned to a map based on its

orientation with respect to the source of illumination. The location of the source of illumination can be specified. A perspective view is created by joining adjacent relief profiles. This is also called 'wire net' technique. Profiles may be used in single directions or in two perpendicular directions. The technique is complex and uses computations of hidden points. The location of an observer can be varied to create many views.

Fly-through is a special 'perspective view' technique. In this technique, many perspective views are used in an animation to create an impression as if one is flying over the terrain. The direction and the speed of the flight, height, etc. can be defined interactively. In perspective cartography and texture mapping, an image or cartographic objects, e.g., roads, drainage, etc. can be overlaid on the perspective view.

3.2.8 GIS Applications in Water Resources

There are innumerable applications of GIS in water resources and the list is fast growing. This technique has been widely applied in conjunction with remote sensing. GIS are particularly helpful in distributed modeling where they are used in managing input data and display of output. For example, long-term soil erosion from agricultural lands are estimated using Universal Soil Loss Equation (USLE). The erosion values are given for bare standard plots under different slopes and these are modified by certain factors. Geographic data of different factors and soils are overlaid using a GIS to obtain the mean-annual soil erosion rates. The main inputs influencing soil erosion are geology, physiography, soils, drainage density, and land use. Overlay of these maps can be used to delineate the areas that are vulnerable to soil erosion.

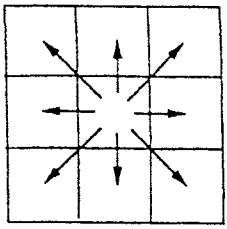
The SCS curve number technique is frequently used to estimate direct runoff from a watershed. Using a GIS, land use and soil maps are overlaid to get a composite map and curve numbers are assigned based on it. An average curve number can be determined or a distributed model can be applied. The SCS curve number method with GIS was used by Muzik (1996) to derive the distributed unit hydrograph.

Stream ordering is a method of assigning a numeric order to chains in a stream network based upon the number and arrangement of tributaries. Stream order has been demonstrated to be related to numerous characteristics and processes of river ecosystems. In stream ordering, headwater streams (those that receive water only from overland flow) are assigned to order number one, and chains in a downstream direction are incremented based on two different numbering methods, described below.

Surface water flow

A GIS is also a powerful tool to model hydrological processes in catchments and command areas. Since surface water flows in the downslope direction, the overland and channel flow direction can be determined using elevation data. Note that terrain slope and flow direction is a local property and GIS permits a realistic representation of these in the model. Grid-based DEMs generally provide the most efficient structures for estimating terrain attributes.

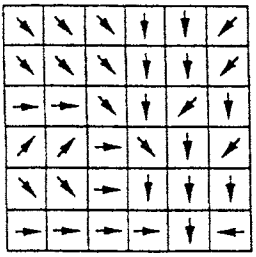
A comparable flow path procedure can be done in a vector GIS, except that the topologies of the network chains (i.e., the beginning and ending nodes specified as part of the data structure) are used to determine the directionality of flow from the "start" location to the outlet. Fig. 3.22 shows watershed terrain analysis using a grid model. The map of flow direction for each grid is prepared using a GIS and the equivalent network showing flow accumulation is also shown.



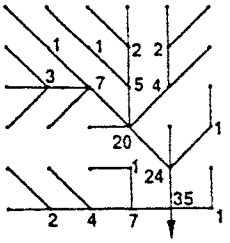
(a)

78	72	69	71	58	49
74	67	56	49	46	50
69	53	44	37	38	48
64	58	55	22	31	24
68	61	47	21	16	19
74	53	34	12	11	12

(b)



(c)



(d)

Fig. 3.22 Watershed terrain analysis using Grids; (a) the eight-direction pour-point model; (b) a grid of terrain elevations; (c) the corresponding grid of flow directions, (d) the equivalent network showing flow accumulation.

Catchments are defined at a point on a stream or are surrounding contributing areas for lakes. Before the advent of GIS, catchment boundaries were manually delineated. The same process can now be done automatically and quite accurately in a GIS. Water flow data layers can be used to determine the path from a known start location in the watershed to the outlet. This capability is useful in tracking the path of a pollutant from a point or non-point source through the drainage network. To distinguish diffuse from channelized flow, thresholds can be applied to the flow accumulation data layer such that only cells which receive runoff from a sufficiently large drainage area are identified as channels. A network is a series of interconnected lines. Networks may be dendritic or rectilinear, as exemplified by stream and road networks. Stream networks branch hierarchically, have unidirectional flow, and are composed of network chains. Several algorithms for delineating watersheds have been covered by Vieux (2001) who has discussed distributed watershed modeling using GIS at length.

The 'WaterWare' decision support system for river basin planning developed by Jamieson and Fedra (1996) uses a GIS to manage spatial data. The most widely used distributed parameter catchment models include SHE (DHI 1998), and TOPMODEL (Quinn et al. 1995). Kite et al. (1996) describe application of SLURP watershed model. Many more applications of GIS and remote sensing to water resources have been described in recent books and papers such as DeVantier and Feldman (1993), Engman (1993), Shih (1996), Singh and Fiorentino (1996), and Islam and Sado (2002).

3.2.9 Conclusions

The amount of digital data required to accurately describe the topography of even small geographic regions can be huge. This makes GIS a memory intensive and computationally intensive system. A good computer system for GIS requires large CPU memory, high capacity disk drives, 21" monitor, and peripherals like digitizer, scanner, colour printer, etc. Even with sharp reduction in prices of hardware in recent times, the total cost of hardware and software can be quite large. Of course, all the set-up would be futile if adequate skilled manpower is not employed.

From the above discussion, it emerges that GISs are being tried in a variety of applications. However, the time and cost of using a GIS can be significant, more so if the database has to be created from the scratch. The application of GIS becomes very convenient when the database exists or when it can be shared for several related purposes. Ideally, a GIS database should be created at the time of planning a project and it can be shared later on in all the studies related to the project. This may not work in practice because the organizations involved in various activities are distinctly different, they may have different hardware and software, and the human element may prohibit such cooperation.

In future, the use of GIS in water resources is likely to grow rapidly. The main limitation will not be the availability of computing power but the innovative ideas and trained man-power. Further technical advancements are likely to result in improved tools for data collection, database creation, and numerical modeling. Education and awareness will also play a key role in the success of the methodology. It will require efforts of many more years before the technology is sufficiently propagated and percolated and it is viewed as an alternative analysis tool.

3.3 ARTIFICIAL NEURAL NETWORKS

Many attempts have been made to develop a technique that does not require algorithm or rule development and thus reduce the quality and complexity of the software. One such technique is known as "neurocomputing". The origin of this technique can be traced to the functioning of the human brain which contains billions of neurons and their interconnections. Due to the structure in which the neurons are arranged and operate, humans are able to quickly recognize patterns, process data, and learn from past experiences. An interesting idea that emerged in the 1940s was the possibility of emulating the processing mechanism of the brain. Although the biological unit still out-performs any

man-made tool in terms of recognition, analysis, prediction and especially learning, the alluring success from the brain-simulation models has provided enough motivation for extended research. *Artificial Neural Networks* (ANNs) refer to computing systems whose central theme is borrowed from the analogy of biological neural networks. ANNs represent highly simplified mathematical models of our understanding of the biological neural networks. They include the ability to learn and generalize from examples, to produce meaningful solutions to problems even when input data contains error or are incomplete, and to adapt solutions over time to compensate for changing circumstances and to process information rapidly. Artificial neural networks are also referred to as "neural nets," "artificial neural systems," and "parallel distributed processing systems".

The ANN approach is faster compared to its conventional compatriots, robust in noisy environments, flexible in the range of problems it can solve and highly adaptive to the newer environments. An ANN has the ability to learn from examples, to recognize a pattern in the data, to adapt solutions over time, and process information rapidly. Due to these established advantages, by now ANNs have numerous real world applications, such as image processing, speech processing, performing general mapping from input pattern (space) to output pattern (space), grouping similar patterns, solving constrained optimization problems, robotics, and stock market predictions. Mathematically, an ANN is often viewed as a universal approximator (ASCE, 2000a). The applications of ANNs to water resources problems are rapidly gaining popularity due to their immense power and potential in mapping of non-linear system data.

Some of the reasons why the ANNs have become an attractive modeling tool are:

1. They are able to 'learn' the relation between the input and output variables even when the underlying physical laws are unknown or not precisely known.
2. The ANNs are a useful tool in modeling complex processes.
3. The mathematics is simple and one need not solve complex partial differential equations with attendant problems like instability of algorithm.
4. They work well even when the training sets are incomplete or contain noise.
5. They are able to adapt to solutions over time.
6. Once a network is trained, it is easy to use.

A water resources system may be nonlinear and multi-variate, and the variables involved may have complex inter-relationships. Often, the problems are ill-defined and solutions are difficult to come by using physically-based methods. Such problems can be efficiently solved using ANNs. In many cases, the existing knowledge is far from perfect and, therefore, empirical models are used. The ANNs are handy in such cases too. Because of their built-in mechanism of growing 'wiser' with 'experience', the ANNs are capable of adapting their complexity and their accuracy increases as more and more input data are made available to them. They are capable of extracting the relation between inputs and outputs of a process without any knowledge of the underlying principles. The processes involving several parameters are easily amenable to neuro-computing. Because of the generalizing capabilities of the activation function, one need not make any assumption about the relationship (linear, non-linear etc.) between input and output as in case of

regression analysis. All these properties make ANNs an attractive tool for water resources practitioners.

Definition

The origin of ANNs lies in the quest to mimic the functioning of human brain, the definition and jargon can hardly defy this connection. There are many definitions of an ANN. The definition proposed by Govindraju and Rao (2000) is one of them: "A neural network is a massively parallel distributed processor that has a natural propensity for storing experimental knowledge and making it available for use. It resembles the brain in two respects:

1. Knowledge is acquired by the network through a learning process.
2. Interneuron connection strengths, known as synaptic weights, are used to store the knowledge."

3.3.1 Structure and Classification of ANNs

An ANN is a network of parallel, distributed information processing system that relates an input vector to an output vector. It consists of a number of information processing elements called neurons or nodes, which are interconnected via unidirectional, weighted signal channels called connections. The networks with a large number of neurons are frequently used for practical problems. The way these neurons are interconnected determines how computations proceed.

The most common way of classifying ANNs is based on the number of layers: single layer, bilayer and multi-layer. Another classification is based on the direction of data flow through the network. The networks where information passes one way (forward) are known as feed-forward networks. The information is received by the input layer nodes which process and pass it on to the next (hidden) layer. The hidden layer(s) nodes also process it and pass to the next layer till the final output layer. In a recurrent ANN, the information flows through the nodes in both directions. To achieve this, the previous network outputs are recycled as current inputs. In fully connected networks, each node is connected to every other node.

The most widely used network structures in water resources area are the multi-layer, feed-forward networks. The remaining discussion is focused only on such networks.

3.3.2 Feed-forward ANNs

A feed-forward ANN has an input layer, an output layer and one or more hidden layers in between the input and output layers. Each of the neuron in one layer is connected to all the neurons of the next layer and the neurons in one layer are only connected to the neurons of the immediate next layer. The strength of the signal passing from one neuron to the other depends on weights of the inter-connections. The intermediate layers enhance the network's ability to model complex functions. The optimal architecture of an ANN is the one that

yields the best performance in terms of error minimization while retaining a simple and compact structure (ASCE, 2000a).

A three-layer feed forward ANN is shown in Fig. 3.23. 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 any 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.

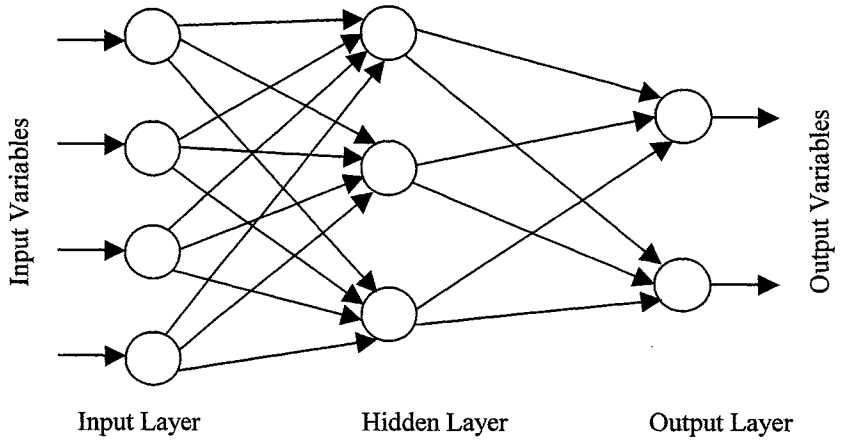


Fig. 3.23 Schematic representation of a three-layer feed forward ANN.

A typical neuron is shown in Figure 3.24. Depending on the layer in which the neuron is located, its input may be an input causal 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 and passed through a transfer function to produce its result. The weights leading to the j^{th} neuron in a layer form a weight vector $W_j = (w_{1j},\ldots w_{ij},\ldots,w_{nj})$, where w_{ij} represents the weight of the connection from the i^{th} neuron in the previous layer to the current neuron.

The most commonly used transfer function is a steadily increasing S-shaped curve, called a *sigmoid or logistic function.* The basic characteristics of the sigmoid function are that it is continuous, differentiable everywhere, and is monotonically increasing and these make it suitable for use with an ANN.

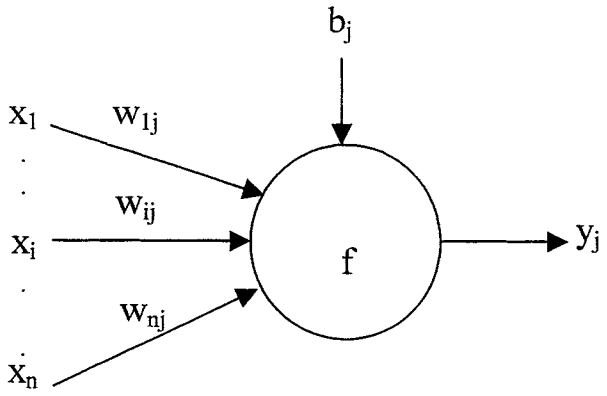


Fig. 3.24 A typical processing node of an ANN.

The sigmoid function is shown in Fig. 3.25. The input to the function can vary between $\pm\infty$ and output y_j is always bounded between 0 and 1. The attenuation at the upper and lower limbs of the "S" constrains the raw sums smoothly within fixed limits. The transfer function also introduces a non-linearity that further enhances the network's ability to model complex functions.

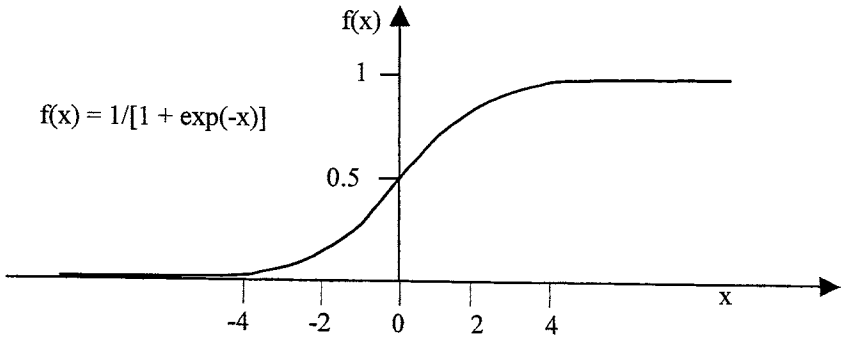


Fig. 3.25 The Sigmoid function.

To obtain output of a node *j* (y_j), the weighted sum of input signals is taken. This should exceed the threshold or bias of the neuron (b_j) before the neuron can fire. The resulting value is passed through the activation function to get the output. The sigmoid function is:

$$y_j = 1 / [1 + \exp(-z)] \tag{3.18}$$

and the *tan h* function is

$$y_i = (e^z - e^{-z})/(e^z + e^{-z}) \tag{3.19}$$

where $z = \Sigma\, w_{ij}\, x_i$.

The potential of feed-forward neural networks can be attributed to three main factors (Kothari and Agyepong 1996): (1) multilayered feedforward neural networks do not need an explicit mathematical equation relating inputs and outputs; (2) a feed-forward network with a single hidden layer with an arbitrary number of sigmoidal hidden nodes can approximate any continuous function; and (3) a feedforward network with a single hidden layer of m sigmoidal nodes achieves an integrated squared error of O(1/m) while a linear combination of a set of m fixed basis functions achieves an integrated squared error of $O(1/m^{2/d})$, where d is the dimension of the input (Barron 1993).

3.3.3 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. Maren et al. (1990) recommend the geometric mean of the numbers of neurons in input and output layers as a good starting guess.

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.

3.3.4 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. 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. Supervised training is most common in water resources 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.

Caution is to be exercised in supervised learning process so that it does not end up in *overtraining a network* or *overfitting*. This happens when the network "learns" a data set too well, i.e., while trying to capture the underlying principles, it also tries to fit the noise that is present in the data set (Fig. 3.26). Note that the network is supposed to learn the trends in the data set and not remember the individual patterns. In this eventuality, the ANN results will be very good for the training data set but poor for others. ASCE (2000a) has recommended a cross training procedure to overcome the overfitting problem. The goal of this procedure is to stop training when the network begins to overtrain. A portion of the data set is reserved for cross training purpose. After the adjustment of network parameters with each epoch, the network is used to find the error for this data set. Initially, errors for both the training and cross training data sets go down. 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.

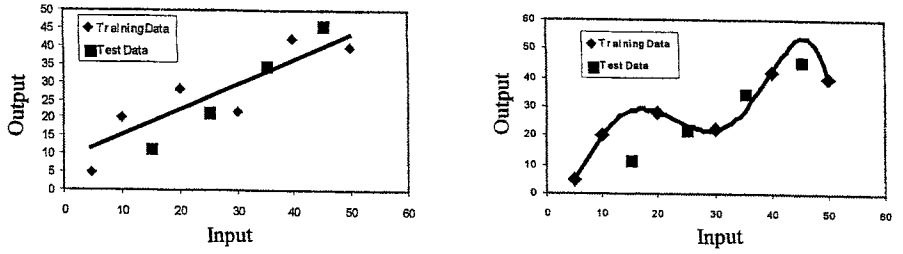


Fig. 3.26 A well-trained ANN (left) and an overfit model (right).

The validation of a trained ANN is performed by subjecting it to patterns that it has not seen during training. Some error criterion, such as MSE or a plot of ANN output versus desired response, can form a basis to assess its performance.

There are many algorithms to train a network. Hsu et al. (1995), proposed a Linear Least Squares SIMplex (LLSSIM) algorithm for training a three-layer feed forward network and demonstrated its application. Error back propagation algorithm is perhaps the most popular algorithm to train multi-layer feed-forward network and is discussed next.

Data Standardization

Before applying ANN, the input data need to be standardized so as to fall in the range [0,1]. A typical variable, say discharge (Q), which can vary between zero to some maximum value Q_{max} can be standardized by the following formula:

$$Q_s = Q / Q_{max} \tag{3.20}$$

where Q_s is the standardized discharge. A different formula will be more suitable for a variable that varies within a certain range. Minns and Hall (1996) have rightly emphasized the importance of the correct standardization. There is, however, some danger of losing information in standardization.

3.3.5 Error Back Propagation Algorithm

The error back propagation (BP) algorithm based on the generalized delta rule was proposed by Rumelhart et al. (1986) to adjust the inter-connection weights during training. In this algorithm, a set of inputs and outputs is selected from the training set and the network calculates the output based on the inputs. The actual output is subtracted from the target output to find the output-layer errors. The weights of all neurons are adjusted by an amount that is proportional to the strength of the signal in the connection and the total measure of the error. The total error at the output layer is then reduced by redistributing this error backwards through the hidden layers until the input layer is reached. This backward propagation of errors gives the algorithm its name. This process continues for a number of prescribed sweeps or until a prescribed error tolerance is reached. The mean square error (MSE) over the training samples is a typical objective function. If all possible sets of weights are plotted against the corresponding sum-of-squares of errors, the result is an error surface shaped like a bowl. Its bottom marks the set of weights with the smallest sum-of-squared error. The goal during the training is to find the bottom of the bowl or the best set of weights. A typical error function can be given as

$$E = \sum_{p=1}^{N} \sum_{n=1}^{m} (T_{pn} - O_{pn})^2 \tag{3.21}$$

where T_{pn} is the target value of n^{th} neuron for the p^{th} pattern, O_{pn} is the output value of the n^{th} neuron for the p^{th} pattern, N is the total number of patterns, and m is total number of output neurons. The increment of weights connecting node i to j at the n^{th} pass, $\Delta w_{ij}(n)$ is

given by

$$\Delta w_{ij}(n) = -\varepsilon^* \frac{\partial E}{\partial w_{ij}} + \alpha^* \Delta w_{ij}(n-1) \tag{3.22}$$

where α and ε are known as momentum factor and learning rate. The momentum factor controls the speed of training and helps prevent the oscillations in weights. The algorithm can be trapped in a local minima and learning rate can be adjusted to increase the chance of avoiding the same.

The BP training algorithm involves two steps. The first step is a forward pass, in which the effect of the input is passed forward through the network to reach the output layer. This is compared with the desired output and the error is computed. In the second step, the error is propagated back towards the input layer and the weights are modified according to eq. (3.4). The BP algorithm is based on the steepest descent method. The problem of local minima is faced in most non-linear optimization problems. It can be addressed to some extent by adjusting the step size and choosing different starting points. Besides, the closer are the initial guesses to the optimum point, the faster is the training but there is no definite way of making a good initial guess of the weights.

The computations begin with initialization of the weights. The steps of algorithm, following ASCE (2000a), are as follows:

- Step 1. Do Steps 2-9 till the stopping condition is met.
- Step 2. For each training pair of data set, perform Steps 3-8.

Feed-forward:

- Step 3. Each input neuron (X_i, i = 1, 2, … , n) receives input signal x_i and sends it to all units in the next (hidden) layer.

- Step 4. Each neuron in the hidden layer (Z_j, j = 1, 2, … , p) sums its weighted input signals

$$Zin_j = v_{oj} + \sum x_i v_{ij} \quad \text{for } i=1,2, \ldots ,n \tag{3.23}$$

where v_{ij} is the connection weight and v_{oj} is the bias value. The activation function is applied its to compute the output signal:

$$Z_j = f(Zin_j) \tag{3.24}$$

This signal is sent to all units in the output layer. Typically, "*f*" is the sigmoidal nonlinear function, defined in eq. (3.18) or tanh function defined in eq. (3.19).

- Step 5. Each output neuron (Y_k, k = 1, 2, …, m) sums its weighted input signals

$$Yin_k = w_{ok} + \sum z_i w_{ij} \quad \text{for } j=1,2,\ldots, p \tag{3.25}$$

Again, the activation function is applied to compute the output signal:

$$Y_k = f(Yin_k) \tag{3.26}$$

Back-propagation of error:

- Step 6. Each output neuron (Y_k, k = 1, 2, …, m), computes its error using the target pattern corresponding to the input training pattern

$$\delta_k = (t_k - y_k)f'(Yin_k) \tag{3.27}$$

calculates its weight correction term (to update w_{jk})

$$\Delta w_{jk} = \partial\delta_k Z_j \tag{3.28}$$

calculates its bias correction term (to update w_{ok} later)

$$\Delta w_{ok} = \alpha\delta_k \tag{3.29}$$

and sends δ_k to nodes in the previous layer.

- Step 7. Each hidden unit (Z_j, j = 1, 2, …, p) sums its delta inputs (from units in the next layer)

$$\delta in_j = \sum \delta_k w_{jk} \text{ for } k=1,2,\ldots..,m \tag{3.30}$$

multiplied by the derivative of its activation function to calculate its error information term

$$\delta_j = \delta in_j\, f'(Zin_j) \tag{3.31}$$

calculates its weight correction term (used to update v_{ij} later):

$$\Delta v_{ij} = \partial\delta_j x_i \tag{3.32}$$

and calculates its bias correction term (used to update v_{oj} later):

$$\Delta v_{oj} = \alpha\delta_j \tag{3.33}$$

Update weights and biases:

- Step 8. Each output neuron, Y_k, k = 1, 2, …, m, updates its bias and weights (j = 0, 1, …, p):

$$w_{jk}(new) = w_{jk}(old) + \Delta w_{jk} \tag{3.34}$$

Each hidden node (Z_j, j = 1, 2,…, p) updates its bias and weights (i = 0, 1,…, n):

$$v_{ij}(new) = v_{ij}(old) + \Delta v_{ij} \tag{3.35}$$

- Step 9. Test if the stopping condition is satisfied.

Although BP is a powerful algorithm, it has several drawbacks that can lead to problems during training. The most common problems are long and slow training process, network paralysis, and local minima.

The training process usually starts with a long step size which is gradually reduced till the desired convergence is achieved. There are no definite guidelines to determine the step size. Too small a step size will increase the training time abnormally while a large step size may skip the optima. Jain et al. (1999) found that learning rates between 0.4 to 0.8 give good results. The problem of moving target arises as the weights have to continuously adjust their values from one output to another for successive patterns. The change in a weight during one training pass may be nullified in the next pass because of a different pattern and this may also slow down training.

When there are large adjustments of weights in the initial epochs, this may lead to *network paralysis*. When all the nodes produce large outputs, the derivative of the activation function can become small and this will slow down the learning process (change in weights) because the change in weights is proportional to the derivative of the activation function. This problem is usually avoided by reducing the step size or learning rate.

Example 3.1: Train the following 3-layer network by BP method: Assume the learning rate to be 0.25. The input pattern is (0, 1) and the output 0.

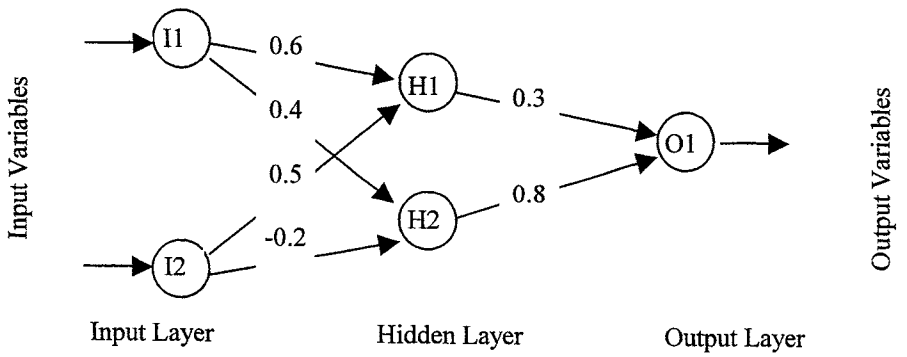


Fig. 3.27 Example 3-layer ANN.

Solution: To begin with, the weight values are set to random values: 0.6, 0.4, 0.5, -0.2 for weight matrix 1, and 0.3 and 0.8 for weight matrix 2. The input signal set to the neurons I1 and I2 of the input layer which just pass the signal to the hidden layer. Now consider the working of the hidden layer.

Input of hidden neuron H1:	$0 * 0.6 + 1 * 0.5 = 0.5$
Output of hidden neuron H1:	$1 / [1 + \exp(-0.5)] = 0.6225$
Input of hidden neuron H2:	$0 * 0.4 + 1 * (-0.2) = -0.2$
Output of hidden neuron H2:	$1 / [1 + \exp(+0.2)] = 0.4502$

The signal now reaches the output layer.

Input of output neuron O1: $0.6225* 0.3 + 0.4502* 0.8 = 0.5469$

Output of output neuron O1: $1 / [1 + \exp(-0.5469)] = 0.6334$

Since the target output is 0,

Error at the output neuron = $0 - 0.6334 = -0.6334$

To modify the weights, we first calculate

$\delta = \text{Error}*(\partial\text{Out}/\partial x) = (\text{Target} - \text{Out})*\text{Out}*(1.0 - \text{Out})$

Change in weight

$\Delta w_{pq,k} = \eta \delta_{q,k} \text{Out}_{p,j}$

and $w_{pq,k}(n+1) = w_{pq,k}(n) + \Delta w_{pq,k}$

First, change the weights in weight matrix 2:

$\delta = (-0.6334) * 0.6334 * (1-0.6334) = -0.1470.$

$\Delta w_{11,3} = 0.25*(-0.1470)*0.6225 = -0.0229$

$\Delta w_{21,3} = 0.25*(-0.1470)*0.4502 = -0.0165$

New value of weight 1: $0.3 + (-0.0229) = 0.2771$

New value of weight 2: $0.8 + (-0.0165) = 0.7835.$

Now consider the weights in weight matrix 1:

Change in weight 1: $0.25 * (-0.6334) * 0* 0.6225 * (1-0.6225) = 0$

Change in weight 2: $0.25 * (-0.6334) * 0* 0.4502 * (1-0.4502) = 0$

Change in weight 3: $0.25 * (-0.6334) * 1* 0.6225 * (1-0.6225) = -0.0372$

Change in weight 4: $0.25 * (-0.6334) * 1* 0.4502 * (1-0.4502) = -0.0392$

Hence,

New value of weight 1: $0.6 + 0 = 0.6$ (not changed)

New value of weight 2: $0.4 + 0 = 0.4$ (not changed)

New value of weight 3: $0.5 + (-0.0372) = 0.4628$

New value of weight 4: $-0.2 + (-0.0392) = -0.2392.$

3.3.6 Cascade Correlation Algorithm

This algorithm starts training with a minimal network, i.e., without any node in the hidden layer. During the training, the network grows by adding new hidden units one by one, maximizing the impact of the new node on the network error, creating a multilayer structure. If the output error is greater than the desired value, a node is added to the hidden layer. Once a new hidden node has been added to the network, its input-side weights are frozen. The hidden nodes are trained so as to maximize the correlation between output of the nodes and output error. A training cycle is divided into two phases. First, the output nodes are trained to minimize the total output error. Then, a new node is inserted and connected to every output note and all previous hidden nodes. The new node is trained to correlate with the output error. The addition of new hidden nodes is continued until the

maximum correlation between the hidden nodes and error is attained.

The training or weight updating is done for any two layers at a time, and the weights are optimized with the help of a gradient ascent method. In this method, the correlation between the hidden unit output and residual network error is maximized. A noteworthy property of this algorithm is that the determination of architecture is part of training. The steps for this algorithm are summarized as follows (Thirumaliah and Deo, 1998):

1. Consider only the input and output nodes.
2. Train directly the input-output weights over the entire training set by the delta rule.
3. Add one new hidden node. Connect it to all input nodes as well as to all other existing hidden nodes. Take all training sets one by one and adjust the input weights of this new hidden node after each training set so as to maximize the overall correlation S between the new hidden node's value and the residual error defined as:

$$S = \sum_{o} \left| \sum_{p} (V_p - \overline{V})(E_{p,o} - \overline{E}) \right| \tag{3.36}$$

where V_p is the output of the new hidden node for training pattern p; E_{po} is the network output error for output node o on pattern p; and bars denote the averages of the respective quantities. Pass the training data set one by one and adjust input weights of the new node after each training set unit S does not change appreciably.

4. Once the training of the new node is done, freeze its input weights and the output side weights are trained once again using the delta rule.

5. Go to step 3, and repeat the procedure until the specified minimum error is reached or a specified maximum number of iterations is over.

Thirumalaiah and Deo (2000) have reported that this algorithm significantly reduced the training time. However, in some instances, it is possible to get better results (in terms of a higher coefficient of correlation and efficiency) with the help of a different training algorithm.

3.3.7 Modular Neural Network

In applications in which the training data are fragmented or a discontinuous representation of the process having a significant variation over the range of the inputs, the normal three-layer ANNs may not give good performance. For example, the relationships between rainfall and runoff are likely to be quite different for low and high flow events because different catchment properties govern the flow in these ranges. To a conventional ANN, this data may appear to be inconsistent and noisy, and if this is the case, the ANN will not be able to properly learn the data behavior. A Modular Neural Network (MNN) can adequately handle such problems. An MNN is based on the assumption that a complex nonlinear problem can be divided into several sub-problems.

A schematic diagram of an MNN is shown in Fig. 3.28. It consists of several expert feed forward networks, each of which receives the input vector. The input is also received by a gating network which produces scalar outputs that are partitions of unity at each point in the input space. Corresponding to the input vector, each expert network produces certain response and the weighted sum of the responses is the network output. Here, the weights are assigned by the gating network. According to Zhang and Govindaraju (2000), the gating network output can be regarded as the probability that an input vector is attributed to a particular expert. The number of expert modules is equal to the number of partitions in the input and is dependent on the problem. The data of three medium sized watersheds was used for monthly runoff predictions based on rainfall and temperature data. In this study, the input was partitioned in three subsets representing low, medium, and high flows and three expert networks were used. The authors showed that the performance of the MNN was better or comparable to the performance of a three layer ANN.

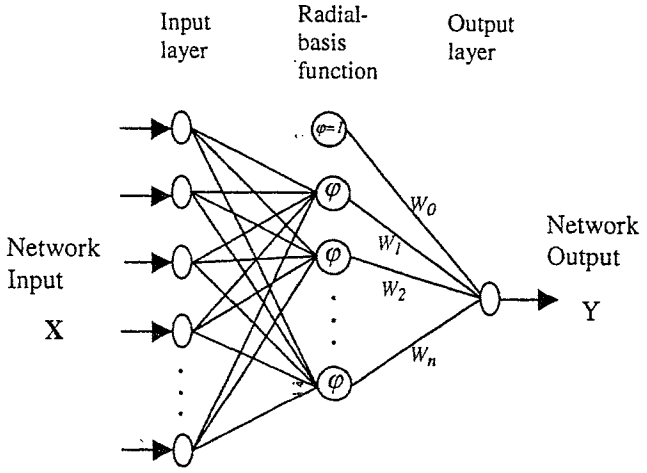


Fig. 3.28 Schematic diagram of a Modular Neural Network (MNN) [Source: Govindaraju and Rao (2000)].

3.3.8 Radial-Basis Function Networks

The back-propagation algorithm of a multi-layer feed-forward ANN is a gradient descent algorithm that may terminate at a local optimum. This problem is overcome in Radial-Basis Function (RBF) networks by incorporating the non-linearity in the transfer functions of the nodes of the hidden layer and thus the parameter optimization becomes a linear search. The theory and structure of the RBF networks is based on the theory of interpolation in multi-dimensional spaces. However, in case of exact interpolation of every data point, a very large number of nodes in the hidden layer will be required. As most water resources data contain a noise element too, the network would also try to mimic the noise. To overcome these problems, the number of radial-basis functions (M) employed in the RBF networks is much smaller than the number of training patterns (N). The number of nodes in the hidden layer M is determined during training.

The RBF networks consist of three layers, viz., a transparent input layer, a hidden layer, and an output layer. The input layer passes the variables to each node in the hidden layer. A transformation function at each node in the hidden layer transforms the incoming values. The transfer function in RBF networks is a radially symmetric basis function and hence this name of the networks. A RBF has a center μ where the function value is the highest and a spread σ that indicates the radial distance from the center within which the function value is significantly different from zero. There is a wide choice of such functions; the Gaussian RBF is the most commonly used function. The responses from the hidden layer are multiplied by the weights of links connecting hidden and output layers and then summed up to yield the network output. The diagram of a typical network is shown in Fig. 3.29.

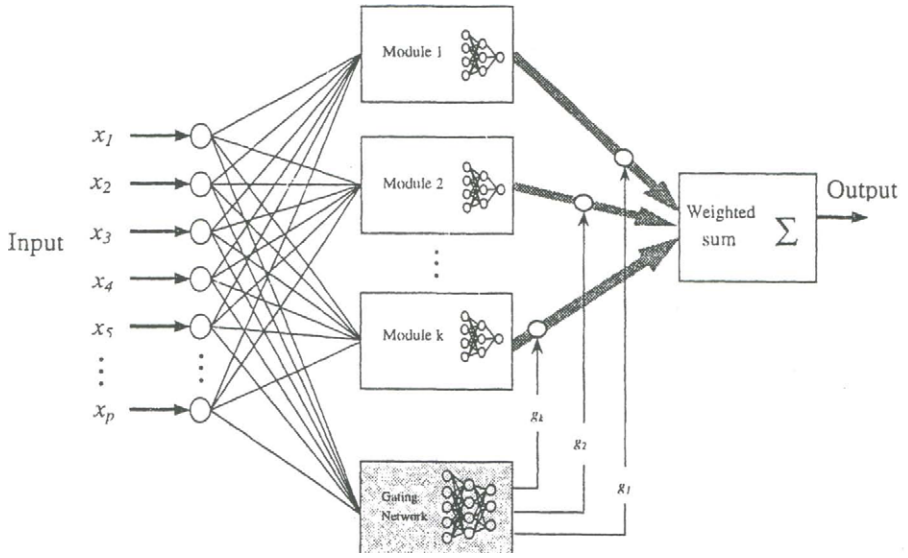


Fig. 3.29 A schematic diagram of RBF network [after Govindaraju and Zhang (2000)]

The response y_j of the j^{th} node of the hidden layer for the p^{th} input pattern is given by

$$y_j = f\left(\frac{\|X^p - \mu_i\|}{2\sigma^2}\right) \tag{3.37}$$

where $\|\ .\ \|$ denotes Euclidean distance. The output of the k^{th} node of the output layer becomes

$$z_{p,k} = \sum_{j=1}^{M} y_j w_{k,j} \tag{3.38}$$

where w_{kj} is the weight of the link from the j^{th} node in the hidden layer to k^{th} node in the output layer.

The training of RBF networks is a two-step process. In the first step, the input patterns are clustered in groups and the center and spread of each group are determined. In the next step, the interconnection weights are obtained. Govindaraju and Zhang (2000) describe several training strategies of these networks.

The RBF network and BP mainly differ in terms of handling of non-linearities. In BP networks, non-linearity is handled through the activation (e.g., sigmoid) function, in the RBF network, the Euclidean distance between the center and input is computed by a non-linear function. The results are combined at the output layer in a linear summation fashion.

3.3.9 Applications

An ANN is a network that relates the inputs and outputs of a system and these networks have been successfully used to map non-linear input and output relationships in a wide range of areas. The immense success with which ANNs have been used to model the non-linear system behavior indicates that this approach can be useful in the field of water resources. ANNs have been used for flow predictions, flow/pollution simulation, parameter identification and to model complex non-linear input-output time series.

A brief review of selected applications follows. A set of two papers published by the ASCE task committee on application of ANNs in hydrology, (ASCE, 2000a, 2000b) contains an exhaustive review of theory and applications of ANN in water resources.

French et al. (1992) developed a three-layer feed forward ANN to forecast rainfall intensity in space and time and compared the results with two other methods of short term forecasting. Since an ANN relates the pattern of inputs to the pattern of outputs, volume continuity is not a constraint. However, care must be taken to avoid the presentation of contradictory information to the ANN. Chang and Tsang (1992) compared the multiple regression and ANN approaches to model snow water equivalent from multichannel brightness temperatures and reported that the results of the ANN approach were better.

Hsu et al. (1995) have shown that the ANN model approach provides a better representation of the rainfall-runoff relationship of a medium sized basin than the ARMAX approach or the Sacramento soil moisture accounting model. Raman and Sunilkumar (1995) investigated the use of ANNs for synthetic inflow generation and compared the model performance with that of a multi variate time-series (ARMA) model. Smith and Eli (1995) used a neural network model for simulating runoff using data of a synthetic watershed. They trained a back propagation network to predict the peak discharge and the time to peak. Minns and Hall (1996) have reported a series of numerical experiments in connection with the application of ANN to rainfall-runoff modeling and concluded that the ANNs are capable of identifying usable relationships between discharges and antecedent rainfalls.

Carriere et al. (1996) designed a virtual runoff hydrograph system based on ANN by training a recurrent back propagation neural network. They obtained good correlation between observed and predicted data. An advantage of using ANN for rainfall-runoff modeling is that parameters relating to the catchment can be avoided in the input and

virtually no model parameter needs to be manually calibrated. Raman and Chandramauli (1996) derived reservoir operating policies for a dam in India by two approaches: i) a DP and neural network procedure (DPN model) and ii) DP and a multiple linear regression procedure (DPR model). Based on a comparison of the performance, it was concluded that the DPN model provided better performance than other models. Dawson and Wilby (1998), while using an ANN for river flow forecasting, have given an overview of ANNs, their training and data standardization. Based on the results of an application study, they have highlighted the ability of ANN to cope with missing data and to learn from the event currently being forecast in real-time. They have also emphasized the need for a thorough investigation into the relationship between the training period length and the hydrological realism of the ANN forecast.

Fernando and Jayawardena (1998) applied RBF networks for runoff forecasting. In general, the performance of RBF networks was found to be as good or better than multi-layer feed-forward networks. They also mention that the efforts needed in case of a RBF network with OLS algorithm are considerably less. For reservoir inflow prediction, time-series analysis and ANNs were used by Jain et al. (1999). It was found that high flows were modeled better through the ANN. For reservoir operation, the ANN model performance was found to be the best as compared to the other models. Birikundavyi et al. (2002) also found that a simple ANN can achieve accuracy superior with that of ARMAX and deterministic models for 7-days ahead forecasting.

A very useful application of ANNs is to stage-discharge relationship at a gauging site. Jain and Chalisgaonkar (2000) applied ANN for this purpose and demonstrated that an ANN can also successfully model a loop rating curve (hysterisis effect). This is not possible using the conventional technique which can only fit an average or steady-state curve. Besides streamflow, the phenomenon of hysteresis is noticed in other branches of hydrology too, e.g., soil moisture retention curve and ANNs can be conveniently used in these areas also. Jain (2001) further extended this work by applying the ANN concept to establish the relation between river stage, discharge, and sediment discharge. The input to the ANN consisted of the river stage at the current and previous time periods, and the water discharge and sediment concentration at previous time periods. Such an ANN has two output nodes, one corresponding to the water discharge and the other corresponding to the sediment discharge. It was found that for ANNs, the SSE is about an order of magnitude smaller and the correlation coefficient is very high than is for the conventional method.

Minasny and McBratney (2002) found ANN to be a significantly improved tool to model parametric pedotransfer functions of soils. Kumar et al. (2002) concluded that the ANN can predict reference crop ET for an area better than the Penman-Monteith method. ANNs have also been applied to groundwater remediation problems, identification of pollution sources, and infilling streamflow data. Some researchers have a feeling that the ANN could perhaps be regarded as the ultimate black-box model.

3.3.10 Issues in ANN Applications

By now, ANNs are firmly established as a viable black-box modeling tool. However, along

with numerous advantages, ANNs have some disadvantages too. The first and foremost is the requirement of adequate data of desirable quality and quantity. To be fair to ANNs, this is a crucial requirement with all modeling techniques and ANN cannot be an exception. Presently clear guidelines are not available except that the entire range of likely inputs should be covered. Guidelines to select network architecture for a given type of problems are also badly missing. In this context, ASCE (2000b) has raised the following very pertinent questions which need to be resolved.

a) Can ANNs be made to reveal any physics? The application of ANNs can get a boost if some physical explanation of their functioning is available. This will also help in selecting the appropriate type of network and learning algorithm for a given problem. Some neural networks can provide statistical interpretations in terms of conditional probabilities. For instance, a feed-forward network can learn the posterior probability of a classification. This problem is receiving attention of many researchers.

b) Can an optimal training set be identified? ANNs cannot learn without data – they are data intensive and poor training data will result in poor learning. An optimal data set should fully represent the modeling domain, have minimum required data points and there should not be repetition of data. So far, there are no guidelines about these.

c) Can ANNs improve time series analysis? The time-series models are based on extracting the correlation and dependence structure of the data. While ANNs have been shown to work better than a time series model for river stage and discharge prediction, they have not given any insight into the process. It would be welcome if ANNs can bring out the relationships among the variables and highlight those features of input data that are not revealed by other techniques.

d) Can training of ANNs be made adaptive? The most time-consuming part of ANN application is training. As new data become available, the previously trained ANN has to be re-trained. Since the catchment properties change with time, it is important to incorporate the new information in the model. The ANN applications will be immensely benefited if the training can be made adaptive, i.e., the new information is incorporated into the models without the necessity of complete re-training.

e) Are ANNs good extrapolators? Many studies have shown that ANNs work well in the range of input data that were used for training and their performance deteriorates if during application, the input data are out of this range. This aspect is receiving attention of hydrologists. Imrie et al. (2002) have presented a methodology for training ANNs to produce models that generalize well on new data and can extrapolate beyond the range of values included in the calibration range. They claimed good results with the data from the catchment of the River Trent and a modified cascade-correlation algorithm.

A thinking and solution to the above issues will certainly help in a better understanding of the Artificial Neural Networks. Of course ANNs cannot be considered as a panacea for all types of problems of water resources or an alternative to other modeling approaches. Gupta et al. (2000) have aptly commented: "*We do not advocate that the ANN approach be generally used in place of the conceptual modeling approach, because the conceptual methodology provides the strengths that the ANN approach does not – in particular, the conceptual approach has the potential to be applied to ungaged watersheds*

or to simulate the potential behavior of a watershed under land use changes... Further, implementation of the ANN approach does not require the considerable amount of expertise and data required to calibrate a conceptual watershed model." Logically then, ANN should be viewed as alternative to conventional computing techniques.

3.4 EXPERT SYSTEMS

Ever since the computers have become an integral part of the emerging technology, efforts are on to use their capabilities in decision making. One of the outcomes of these efforts is the emergence of 'Artificial Intelligence' (AI). Artificial Intelligence is concerned with efforts in making computers think and do things intelligently. According to Barr and Feigenbaum (1981), '*Artificial intelligence is that part of computer science which is concerned with designing intelligent computer systems, that is, systems that exhibit the characteristics we associate with intelligence in human behavior'*. Based on the AI concepts, a wide spectrum of areas are being developed, viz., robotics, natural language processing, speech recognition, ANN, computer aided intelligent instructions and knowledge based expert systems. In earlier periods, AI was considered only as a research topic, but later on it found more applications in specialised programs.

Intelligence requires software to be knowledgeable not only about its own possibilities and constraints, but also about the application domain and about the user, i.e., the context of its use. Defaults and predefined options in a menu, sensitivity to context and history of use, built-in estimation methods, or alternative ways of problem solving that depend on the user can all be achieved by the integration of expert systems technology in the user interface and in the system itself.

The development of expert systems (ES) involves separation of problem solving techniques and domain dependent knowledge so that existing problem solving frameworks may be used in new domains with appropriate knowledge base. This led to knowledge engineering which is the process of constructing an ES. Basically, the expert knowledge is stored in the computer in an organized manner. This knowledge-base is used to provide advice and, if necessary, logic behind it (Tuber, 1995).

ES follow the reasoning process that a human decision maker would go through to arrive at a decision. The most prominent problem-solving assets of an expert, particularly in the domain of water resources, are theoretical and practical experience, archived data, and operational expertise. The expert must have sound theoretical knowledge that has been acquired through studies, thinking, and discussions. An expert gathers experience by solving a range of practical problems and in this process, he also builds a large database. His involvement in operational tasks further helps in refining his tools and techniques.

At the core of an ES is knowledge about a specific problem domain. An ES is a combined human-computer system designed to solve problems that normally require logical consideration of both facts and heuristics, or rules-of-thumb, to arrive at a decision. An ES can be defined as '*interactive computer program that incorporates experience, judgment, rules of thumb, intuition and other expertise to provide knowledgeable advice about a*

variety of tasks'. Simonovic (1991) defined a water resources ES as 'a computer application that assists in solving complicated water resources problems by incorporating engineering knowledge, principles of systems analysis, and experience, to provide aid in making engineering judgments and including intuition in the solution procedure'. Professional engineers using computer aided analysis and design have incorporated expertise in the program in the form of constraints and limitations, assumptions and approximations, interactive inputs and outputs including graphics. Some people consider the results more as advice rather than answers.

The important characteristics and advantages of an ES are:

a) When experts retire, the valuable knowledge that they have acquired over the years goes with them; ES helps in preserving such a perishable knowledge.
b) ES can also help an expert either as an assistant, or a partner.
c) ES provides justification and explanation to the conclusions that it has arrived at.
d) ES can be designed to interact with humans in a suitable way.
e) ES incorporates knowledge associated with humans which is otherwise scarce and mobile.
f) ES can help in reduction of the expenses due to mediocrity.
g) ES can provide an efficient solution for complex decision problems requiring large amount of information and many possible outcomes.

A rule-based approach can also be a substitute for a numerical model, even in the socio-economic domain. An example is environmental impact assessment based on a checklist of problems, which can be understood as a diagnostic or classification task. A qualitative label is assigned to potential problems based on the available data on environment and planned action, and a set of generic rules assessing and grading the likely consequences. One of the famous early ES was a software named as MYCIN which was developed by the Stanford Research Institute in 1976 to help a physician in his diagnosis and prescription. It successfully used domain dependent information. The first attempt to develop an ES for water resources problems was by Gasching et al. (1981) who developed a consultation system called HYDRO. This system was intended to provide advice comparable to that of an expert hydrologist in selecting parameter values characteristic of the watershed under consideration.

ESs are attractive because they offer an excellent way of organizing rules. However, note that ESs provide no new intrinsic validity to the use of rules or uncertainty in the process of forming a decision. While some problem domains are very complex but deterministic in nature, most experientially derived rules do not perfectly capture the relationships within a problem domain. The expert systems are helpful because they deal with two types of uncertainties: those of imperfect rules and imperfect information. Imperfect rules may result from a poor understanding or intentional simplification of the system. Rather than saying that "If A then B", one might express to what degree he believes that B follows from A. The answer might be: "If A then probability of B = 0.8". Here probability represents the chance of premise B following premise A. Such a rule is termed an uncertain rule because of its imperfect association. Imperfect information may

either be incomplete or it may have noise.

An effective ES must provide the following which are of direct concern to a system manager (Palmer and Holmes 1988): a) Consider only operating rules that are feasible and that follow existing management procedures. b) Consider operating rules that are appropriate for a wide range of conditions. c) Assume that managers consider the same criteria for each decision. d) Provide quantitative information (such as the probability of yield and failure for a wide variety of initial conditions and overall historic flow regimes) that is not readily available otherwise. e) Operate on a computer available in a typical office. f) Serve as a training tool for future managers.

3.4.1 Expert Systems Architecture

A typical ES consists of (Fig. 3.30): i) a knowledge base; ii) a working memory; iii) an inference engine; iv) system analysis, graphics, and other software; and v) a communication module including interfaces to the developer and the user. The knowledge base consists of declarative knowledge which are facts about the domain, and procedural knowledge which are scientific, analytical or heuristic rules from the domain. The rules in turn a have trigger part which, based on facts, triggers the body part of the rule which are actions for processing instructions on the knowledge base, input-output instructions or control instructions. They may include metarules which are rules about rules. The working memory is the current active set of the knowledge base and may include a knowledge management module. The problem solving module is the inference engine which may also provide justification for the advice from ES. Graphic, numeric analysis, statistical analysis, system analysis and simulation software may also be available. Data communication module provides an exchange of information among the various modules and may also provide interfaces to the developer and user.

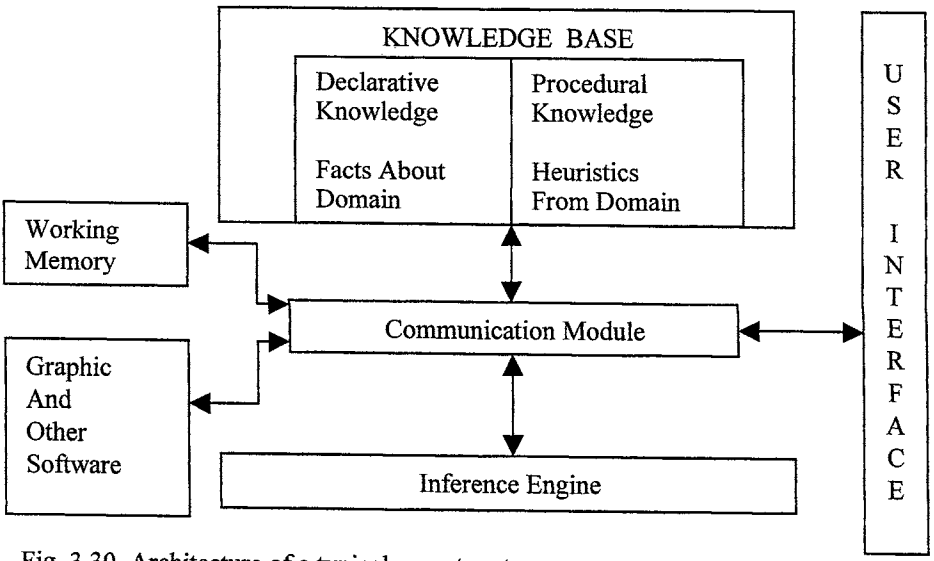


Fig. 3.30 Architecture of a typical expert system.

ES differ from traditional computer programs in the following respects:

a) ES are knowledge intensive programs;
b) ES mimic the decision making and reasoning process of human experts. They can provide advice, answer questions and justify their conclusions;
c) In ES, expert knowledge is usually divided into a number of separate rules;
d) The rules forming expert knowledge base are separated from the methods (inference mechanism, reasoning mechanism or rule interpreter) for applying the knowledge to the current problem;
e) ES are highly interactive; and
f) ES have a friendly/intelligent user interface.

Knowledge is driven by a mechanism called inference engine, which needs data in the context space. An explanation module is sometimes provided and the user interface, as the name suggests, assists in interaction between the user and context space. A knowledge acquisition facility is provided to manage (add or edit) knowledge base.

3.4.2 ES Development

Five sequential stages in the development of ES were enumerated by Hayes-Roth et al. (1983):

i) Identification - determining characteristics of the problem,
ii) Conceptualisation - finding concepts to represent the knowledge,
iii) Formalisation – designing structures to organise knowledge,
iv) Implementation – formulating rules embodying the knowledge, and
v) Testing - validating the rules.

While building an expert system, the gathering and structuring of rules involve significant efforts and resources. Therefore, it is necessary to approach in an organized manner. The first step is to precisely define the problem domain. The next step is to document the expert decision-making process, most commonly by watching or asking the expert. Rouhani and Kangari (1987) have recommended the following steps:

a) Observe the expert without interruption.
b) Informally discuss representative problems with him.
c) Ask the expert to describe the basis for his judgment and empirical rules. Organize knowledge base.
d) Ask the expert to solve a series of problems. Focus on the decision process.
e) Solve a range of problems with the rule base.
f) Verify the accuracy of judgments. Iterate.
g) Ask other experts.

Another good approach is the "thinking-aloud" experiment wherein the expert is asked to think through the problem while speaking out loudly and a verbatim transcript is taken. This is then followed with a "cross-examination" period where the expert is asked

in-depth questions about specific topics. The transcripts of those two sessions are then broken into short meaningful phrases used in the expert's analysis. From those phrases, premises and their relations are identified and formed into rules. Another source of rules is simulation models. While these models can be integrated into the system model, it may pose a computational burden to do so. In such cases, heuristics that generalize the results of a number of simulation runs are developed. Cuena (1983) describes how rules for a flood-control problem are generated in this manner.

To take full advantage of the ES features, a good coordination between the knowledge engineer and the expert is necessary. At times, this requires patience and understanding because an expert will be a busy person and may not always be willing to share his expertise. One individual expert should not be given too much importance because this will give a strong personal bias. A review and critical examination by other experts will help to overcome this bias.

3.4.3 ES Techniques

The order of execution of the rules and/or procedures in an ES is governed by the inference engine in terms of the problem solving strategy used. Maher (1986) considers two approaches:

The derivation approach: A solution is derived that is most appropriate for the problem from a list of predefined solutions stored in the knowledge base of ES. It includes forward chaining (which works from an initial state of known facts to the goal state), backward chaining (which works from a hypothetical goal state to the facts in terms of sub-goals which are preconditions for the goal stated. If the hypothesis is not supported by facts, it tests for another goal state and so on in a predefined order of goals), and the mixed initiative which combines forward and backward chaining strategies.

The formation approach: A solution is formed from eligible solution components stored in the knowledge base. It includes problem reduction (factoring the problem into sub-problems), plan-generate-test (which generates all possible solutions, prunes inconsistent solutions and tests the remaining solutions), and agenda control (assigning a priority rating to each task in the agenda and perform tasks with higher priority before those of lower priority). These may be combined with other techniques for hierarchical planning and least commitment, backtracking and constraint handling.

3.4.4 ES Tools

A wide variety of development tools and environments are available for ES. They include programming languages like PASCAL, C++, and Java. ES shells provide a framework and tools to build a system. They contain all the modules required for an ES except that the knowledge base is hollow. If this base is filled with knowledge in specified syntax, then it becomes an ES. Although the criteria for selection of a tool solely depend on type of problems, a few general principles can guide a knowledge engineer in selecting an approximate shell. For an engineering problem, the prototype of ES must be able to perform

hybrid chaining, i.e., forward and backward. They must be able to handle uncertainty, and provide interface to other softwares. The hardware requirements must be such that the shell is highly portable and can be used on a PC. Expert system shells make the development of these systems easier than previously possible. Many such shells are commercially available. The use of a good tool helps in lessening the effort to develop an ES and the developer can pay more attention on acquiring knowledge and refining the rule base.

The ES also have a module for editing the knowledge base, i.e., adding, deleting or modifying knowledge. The user interface acts as intermediary between the end user and inference engine and conceals the complexity of the system from end user by helping him in a friendly manner. ES must be friendly with a national language interface, should be able to support user-defined functions, and must not be too complicated.

Some ES also an have explanation facility. This module has the knowledge to explain how the system has arrived at particular answers. The explanation includes displaying the inference chains and explaining the rationale behind the use of each rule in chains. The ability to examine the reasoning process and explain their operation are truly innovative and important qualities of expert systems. This facility becomes very handy in debugging the system when its performance is not as expected.

Every ES shell may not have all these facilities. Depending upon the type of the problem, the knowledge engineer has to choose the shell that best suits the requirement. For example, a backward chaining rule is suitable with an ES in medical field while for an engineering design, a system with forward chaining is better.

3.4.5 Knowledge Base

A knowledge base is that module which contains domain specific knowledge, which must be of high quality as the ES behavior critically depends on this. There are mainly three ways of representing knowledge, viz., Rule, Texonomy (semantic nets) and Frames. Representation of knowledge in the form of rules is the most popular formalism with 'IF-condition-THEN-action' statements. A set of rules specifies the program to react and is useful when the knowledge is in the form of condition action. It has its inherent simplicity, understandability and ease of modification. The rule base contains facts and heuristics that are specific to the problem being addressed. An example of a rule for management of a storage reservoir in water scarce region in high flow season is:

```
IF    Demand          >     Average
AND   Storage >=      Capacity/2
AND   Date            >=    August 15
THEN  Curtail irrigation releases to half.
```

Typically, such a rule will be one of many in a rule base relating system conditions and management action.

Semantics (implying meaning) nets representation is used when knowledge is a

subset of some other bigger set. A semantic network consists of points (nodes) connected by links (arcs) that describe the relation between nodes. The biggest advantage of this is that it is possible to represent hierarchical information. Frames refer to a special way of representing common concepts and situations. A frame is a network of nodes and relations organised in a hierarchy, where topmost nodes represent general concepts and lower ones more specific instances. A node is a collection of attributes called slots and their values are stored in the respective slots. Frames are useful in giving specification details.

Knowledge Acquisition

Knowledge in any specialty is of two types: public and private. The body of information that is widely shared and agreed upon is public knowledge. Pump efficiencies, pipe roughness coefficients, and channel roughness coefficients are some examples of public information in a water resource setting. Private knowledge consists of rules of thumb or heuristics developed by human experts through experience. These heuristics are subjective rules of good judgment that allow human experts to make educated guesses and deal with incomplete data. Perceived public resistance to water restrictions, estimated political liabilities, and the risk avoidance characteristics of an operator are examples of private information. This characterization between public and private knowledge is of particular value in water resources planning. Often modelers devote significant effort in capturing public information but devote little attention to private information.

Knowledge acquisition is obtaining of problem-solving skills from an individual and incorporating these in a computer program. These techniques establish information as specialized facts, procedures, and judgmental rules about a narrow domain area (Hayes-Roth et al. 1983). Knowledge acquisition is extremely important in expert systems development and is often a significant obstacle. The success of an ES depends upon the quality of the knowledge gathered and its effective assimilation into a rule base. Unfortunately, knowledge acquisition is an iterative process and many revisions of the system rule base may be necessary. Knowledge acquisition is difficult for several reasons. Domain experts often are unaccustomed, or unable to formulate and articulate their problem-solving methods. They may forget portions of information, introduce inconsistencies, or not express their problem-solving algorithms explicitly. They may fear losing their jobs to computers or exposing their techniques to the scrutiny of coworkers and the public. Finally, they may simply mistrust or feel uncomfortable with the process.

3.4.6 Inference Engine

An inference engine is computer software that examines the knowledge base and answers the questions posed by the user. Simple logic schemes based upon the status of system variables and on the rule base accomplish this task. The inference engine is the most crucial component of the ES since it makes and manipulates the database for problem solution. It is a mechanism that derives the knowledge, i.e., a sophisticated system guiding the selection of a proper response to a specific situation. This is known as pruning. Three formal approaches used in this case are production rules, structured objects and predicate logic. Production rules consist of a rule set, a rule interpreter which specifies when and how to

apply the rules and a working memory that holds data, goals or intermediate results. Structured objects use vector representation of essential and accidental properties. Predicate logic uses propositional and predicate calculi. Context is the work space for the problem constituted by the inference mechanism from information provided by the user and meta rules in the knowledge base. The inference engine can work in two ways, viz., forward chaining and backward chaining.

Forward chaining starts with known initial state and proceeds in the forward direction until the goal state is arrived at. From the given information, the inference engine searches the knowledge base for rules whose precedence matches the given current state and fires those rules adding more information to the working memory. The basic steps are:

1) The system is presented with one or more conditions.
2) For each condition, the system searches the rules in the knowledge base for those rules that correspond to the condition in IF part.
3) Each rule can in turn generate new conditions from the conclusions of the invoked THEN part, which are added to the existing ones.
4) If any condition is added to the system in step (3), it will be again processed from step (2). If there are no new conditions, the session ends.

Backward chaining engines (goal-driven engines) involve reasoning in backward direction. The system selects a goal state and reasons in the backward direction and establishes the initial state conditions necessary for that goal state to be true. If the given initial state conditions match with the arrived initial state conditions, then that goal is the solution. Otherwise, the system selects another goal and repeats the process. This process is very similar to the backward chaining associated with dynamic programming. The steps are:

1) Select a goal state and rules whose THEN portion has that goal state as conclusion.
2) Using IF portion of the selected rules, establish sub goals to be satisfied for the goal state to be true.
3) Using steps (1) and (2), establish initial conditions necessary to satisfy all sub-goals.
4) If the given initial state corresponds with required initial state, then the selected goal is one solution. If not, select another goal state.

3.4.7 Applications in Water Resources

A large amount of domain dependent expertise is available in planning, design, construction and integrated operation of water resources systems and they are often heuristic in nature. Hence, the water resources engineering has good potential for ES applications. The role and context of application of an ES in water systems management is shown in Fig. 3.31. The following examples demonstrate the variety of issues in water resources management that need and can be tackled effectively through judicial use of ES.

Some of the application areas of ES in water resources are detailed below.

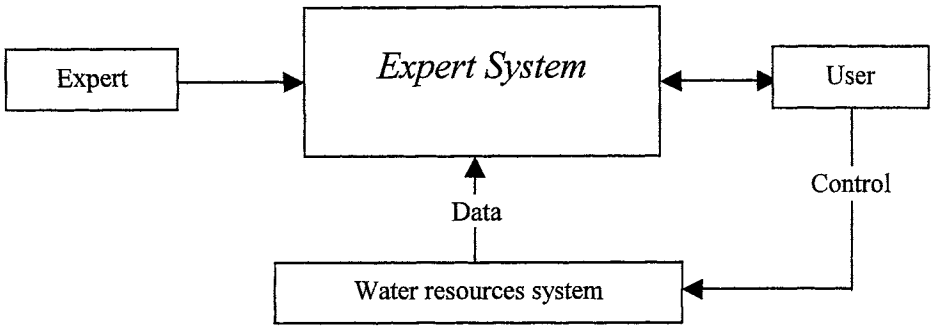


Fig. 3.31 Role and context of an ES in water management.

System modeling and analysis: Mathematical modeling and analysis of water resources systems is an area that is well suited for ES application. ES can be useful in data analysis (hazard evaluation; Wilson, 1986); hydrologic modeling and parameter estimation (Engman et al, 1986; Delleur, 1988) and in formulation of the management problem, choice of model to be used, preparation of input data etc., say, for a reservoir system (Savic and Simonovic, 1989); water-supply system operations (Shepherd and Ortolano, 1996), and for an urban storm sewer system (Lindberg and Nielsen, 1986). An ES for calibration of SWMM model was developed by Baffaut and Delleur (1990).

Reservoir Operation: Integrated management of a system of multipurpose reservoirs and conjunctive use of surface and groundwater are complex problems requiring multi-disciplinary knowledge and experience and, therefore, are suitable for ES applications. An ES for dam operations would use available data on current reservoir levels and inflows, downstream channel conditions, storage capacities, precipitation forecasts, data on demands, etc., and suggest the best operational strategies to maximize overall benefits. In a multi-reservoir system, such data on all the structures can be jointly analysed to formulate the best policy for integrated operation. Some applications of ES for this purpose include ES for real-time reservoir operations by Armijos et al. (1990), ES for real-time operation of a multi-purpose reservoir system by Fischer and Schultz (1991), ES for short-term reservoir operation by Simonovic et al. (1992).

Flood Management System: ES can be very effective in overall flood management planning and flood forecasting. A flood management planning system would contain knowledge of the basin like topography, channel network, soil properties, flood damage centers, meteorological characteristics of the region etc. These data could be used in preparing a blueprint for flood management using structural and non-structural measures. Some relevant applications are: ES for dam site selection by Engel and Beasley (1991); ES to select methods to calculate design flood flows by Varas and Von Chrismar (1995), for flood frequency analysis by Chow and Watt (1990).

A river stage or discharge forecasting system could use the data, such as stages at various locations in the reach, lateral flow data, current and forecasted precipitation data, cross-section data, longitudinal-section data, characteristics of flood plains, and roughness

properties, and use these data in a mathematical model. The results along with the knowledge base can be used to formulate improved flood forecasts. In a similar way, a system for drought forecasting can also be developed. Raman et al. (1992) report an ES for crop planning during droughts.

Water Quality Monitoring: A water quality monitoring system would take data readings and determine possible causes of poor water quality, aid in the location of pollution sources, and suggest remedial actions to improve water quality. Wishart et al. (1990) developed expert systems for the interpretation of river quality data. Sponemann and Fahs (1989) describe control of an urban storm sewer system using ES.

The MEXSES system developed by Fedra et al. (1991) for the Lower Mekong basin is a rule-based ES, using hierarchical checklists to perform screening level environmental impact assessment (EIA). The system is geared for assessment of water resources development projects, such as dams and reservoirs, hydropower and irrigation schemes, flood control, navigation, aquaculture, etc. The indicators used to assess a given project are based on checklists of items specific to a project type, covering environmental as well as selected socio-economic topics. Each indicator is rated on a qualitative scale, from 'not significant' to 'major'. A system of hierarchical checklists was used with a rule-based deduction process including a recursive explanation function and a knowledge base browser. MEXSES uses a knowledge representation, combining an object-oriented design for the descriptors, the basic elements in the inference procedure, with near natural language rules.

Jamieson and Fedra (1996a) described a decision support system called 'WaterWare' that has, among other modules, an expert system. The capabilities of this product were demonstrated through two applications to Thames basin in England and Rio Lerma in Mexico by Jamieson and Fedra (1996b).

3.5 CLOSURE

Knowledge-Based Expert Systems aim at bringing together the expertise of individuals gathered over long periods of professional practice and the power of the digital computers, especially the personal computers. This can be placed at the disposal of the scientists and field engineers for more effective management of the scarce water resources particularly in extreme events like floods and droughts.

The current ES simulate the thinking of an expert only in a gross manner. Still much work is needed to include a number of important aspects of human thinking process like perceiving significance, reaching intuitive conclusions, examining a single issue from different perspectives, understanding basic principles, generalization and breadth of knowledge. Hence, compared to human experts, they appear narrow and shallow. However, in the current form, they have the potential to at least partly relieve an expert of a difficult task and give timely advice to beginners and non-experts. As ES applications increase, it is expected that the vast body of heuristic knowledge and expertise available with water resources professional will be successfully compiled, organized, formalized and

codified into ES packages for efficient management of scarce resources. There are reasons to believe that ES will have commonplace applications in water resources in near future and these will be an integral part of most decision support systems.

An advantage of expert systems over traditional modeling approaches is its ability to incorporate human intuition and experience into the modeling process. This argument rests on the ability of expert systems to capture human decision-making expertise and represent this expertise as a series of rules and facts. Another major advantage is in bridging gaps between scientific developments and their practical application or reducing the time gap between 'lab-to-land' transition.

3.6 REFERENCES (RS & GIS)

Ambast, S.K., Keshari, A.K., and Gosain, A.K. (2000). Estimation of regional evapotranspiration for a part of WYC command using remote sensing technique. Proceedings of International Conference on Integrated Water Resources Management for Sustainable Development, New Delhi.

Arnoff, S. (1989). Geographical Information Systems: A management Perspective. WDL publications, Ottawa, Canada.

Bastiaanssen, W.G.M. (1998). Remote sensing in water resources management: The state of the art. International Water Management Institute, Colombo, Sri Lanka.

Bausch, W.C., and Neale, C.M.U. (1987). Crop coefficients derived from reflected canopy radiation: A concept. Transactions of American Society of Agricultural Engineering. 30(3), 703-709.

Bausch, W.C. (1995). Remote sensing of crop coefficients for improving irrigation scheduling of corn. Agricultural Water Management, 27(1), 55-68.

Burrough, P.A. (1986). Principles of GIS for land resource assessment. Oxford Clarendon Press, London.

Choudhury, B.J., Ahmed, N.U., Idso, S.B., Reginato, R.J., and Daughtry, C.S.T. (1994). Relations between evaporation coefficients and vegetation indices studied by model simulations. Remote Sensing of Environment. Vol. 50, 1-17.

Demers, M.N. (1997). Fundamentals of Geographic Information Systems. John Wiley & Sons, New York.

DeVantier, B.A., and Feldman, A.D. (1993). Review of GIS applications in hydrologic modelling. Journal of Water Resources Planning and Management, ASCE, 119(2), 246-261.

DHI (1998). MIKE-SHE Users Guide and Technical Reference Manual. Danish Hydraulic Institute, Denmark.

Dozier, J. (1989). Spectral signature of Alpine snow cover from the Landsat thematic mapper. Journal of Remote Sensing of Environment, 28, 9-22.

Dugdale, G., and Milford,J.R. (1986). Rainfall estimation over the Sahel using Meteosat thermal infrared data. In ISLCP. Parametrisation of land-surface characteristics: Use of satellite data in climate studies. Edited by Rolfe,E., and Battrick,B. European Space Agency, Paris.

Endreny, T.A. (2002). BASINS toolkit for hydrological monitoring, modelling, and assessment. Hydrological Processes, 16, 1331-1335.

Engman, E.T. and R.J. Gurney (1991). Remote Sensing in Hydrology, Chapman and Hall, London.

Engman, E.T. (1993). Remote Sensing. Chapter 24 in Handbook of Hydrology, Edited by D.R. Maidment, McGraw-Hill, New York.

Goel, M.K., and Jain,S.K. (1997). Automated mapping of snow cover using IRS-1C data. Report No. TR(BR)-6/96-97. National Institute of Hydrology, Roorkee.

Hall, D.K., and Martinec,J. (1985). Remote Sensing of Ice and Snow. Chapman and Hall, London.

Hsu, K., Gao,X., Sorooshian,S., and Gupta,H.V. (1997). Precipitation estimation from remotely sensed information using artificial Neural Networks. Journal of Applied Meteorology, 36(9), 1176-1190.

Islam, M.M., and Sado, K. (2002). Development priority map for flood countermeasures by remote sensing data with geographical information system. Journal of Hydrologic Engineering, ASCE, 7(5), 346-355.

Jamieson, D.G., and Fedra,K. (1996). The 'WaterWare' decision-support system for river basin planning. 1. Conceptual design. Journal of Hydrology, Vol. 177, 163-175.

Jayaseelan, A.T., and Rao,D.P. (1997). Satellite contribution to natural disaster management in India. Conference on Geomatics in Disaster Management. Dehradun, India.

Jenson, S.K., and Domingue,J.O. (1988). Extractring topographic structure from digital elevation data for geographic information system analysis. Photogrammetric Engineering and Remote Sensing. Vol. 54.

Johnston, C.A. (1998). Geographic Information Systems in Ecology. Blackwell Science Ltd., London.

Kite, G.W., Ellehoj,E., and Dalton,A. (1996). GIS for large-scale watershed modeling. in Singh,V.P., and Fiorentino,M. (ed.). Geographical Information Systems in Hydrology. Kluwer Academic Publishers, Dordrecht.

Klemes, V. (1986). Dilettantism in hydrology: Transition or destiny? Water Resources Research, 22(9), 177S-188S.

Lillesand, T.M. and Kiefer, R.W. (1994). Remote Sensing and Image Interpretation. Wiley, New York.

Loucks, D.P. and Fedra, K. (1987). Impact of Changing Computer Technology on Hydrologic and Water Resource Modeling. Review of Geophysics, 25(2), March.

Maling, D.H. (1992). Coordinate Systems and Map Projections. Pergamon Press, Oxford.

Mather, P.M. (1987). Computer Processing of Remotely Sensed Images, An Introduction. John Wiley and Sons, Chicester.

Mendicino, G. (1996). Design of GIS for hydrological applications. in Singh,V.P., and Fiorentino, M. (ed). Geographical Information Systems in Hydrology. Kluwer Academic publishers, Dordrecht.

Moore, I. D. and Grayson, R. B. (1991). Terrain-based catchment partitioning and runoff prediction using vector elevation data. Water Resources Research, Vol. 27, No. 6, pp. 1177-1191.

Muzik, I. (1996). Flood modeling with GIS-derived distributed unit hydrographs. Hydrological Processes. Vol. 10, 1401-1409.

Nagarajan, R. (2000). Temporal orbital satellite images in the drought assessment and early warning decisions. Proceedings of International Conference on Integrated Water Resources Management for Sustainable Development, New Delhi.

NIST (1992). Spatial data transfer standards (SDTS), FIPS Pub. 173, US Department of Commerce, Computer Systems Laboratory, Maryland. SuDocs C13.52:173.

O'Callaghan, J.F., and Mark, D.M. (1984). The extraction of drainage networks from digital elevation data. Computer Vision, Graphics and Image Processing. Vol. 28, 232-344.

Pauwels, V.R.N., Hoeben, R., Verhoest,N.E.C., Troch,F.P.De, and Troch, P.A. (2002). Improvement of TOPLATS-based discharge prediction through assimilation of ERS-based remotely sensed soil moisture values. Hydrologic Processes, 16(5), 995-1013.

Quinn, P.F., Beven,K.J., and Lamb,R. (1995). The ln(a/tanβ) index: how to calculate it and how to use it in the TOPMODEL framework. Hydrological processes. Vol. 9, 161-182.

Rao, P.P.N., and Mohankumar,A. (1994). Cropland inventory in the command area of Krishnarajsagar project using satellite data. International Journal of Remote Sensing, 15(6), 1295-1305.

Rango, A., and Martinec,J. (1999). Modeling snow cover and runoff response to global warming for varying water years. World Resource Review, 11(1), 76-91.

Richards, F., and Arkin,P. (1981). On the relationship between satellite-observed cloud cover and precipitation. Monthly Weather Review, Vol. 109, 1081-1093.

Sabins, F.F. (1987). Remote Sensing Principle and Interpretation, W.H. Freeman & Co., San Francisco.

Schultz, G.A. (1993). Application of GIS and remote sensing in hydrology. HydroGIS 93: Application of Geographic Information Systems in Hydrology and Water Resources. IAHS Pub. No. 211, 127-140.

Seidel, K., Schaper, J., and Martinec, J. (2000). Mapping of snow cover and glaciers with high resolution remote sensing data for improved runoff modeling. Proceedings of International Conference on Integrated Water Resources Management for Sustainable Development, New Delhi.

Shih, S.F. (1996). Integration of remote sensing and GIS for hydrological studies. in Singh, V.P., and Fiorentino,M. (ed). Geographical Information Systems in Hydrology. Kluwer Academic Publishers, Dordrecht.

Sole, A., and Valanzano,A. (1996). Digital Terrain Modeling, in Singh,V.P., and Fiorentino, M. (ed). Geographical Information Systems in Hydrology. Kluwer Academic Publishers, Dordrecht.

Singh, V.P., and Fiorentino,M. (1996). Geographical Information Systems in Hydrology. Kluwer Academic Publishers, Dordrecht.

Smith, J.A. (1993). Precipitation. Chapter 3 in Handbook of Hydrology, Edited by D.R. Maidment, McGraw-Hill, New York.

Vieux, B.E. (2001). Distributed Hydrologic Modeling Using GIS. Kluwer Academic Publishers, Dordrecht.

REFERENCES (ANN)

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000a). Artificial Neural Networks in hydrology, I: Preliminary Concepts. Journal of Hydrologic Engineering, ASCE, 5(2), 115-123.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000b). Artificial Neural Networks in hydrology, II: Hydrological Applications. Journal of Hydrologic Engineering, ASCE, 5(2), 124-137.

Birikundavyi, S., Labib, R., Trung, H.T., and Roisselle, J. (2002). Performance of neural networks in daily streamflow forecasting. Journal of Hydrologic Engineering, ASCE, 7(5), 392-398.

Carriere, P., Mohaghegh, S., and Gaskari, R. (1996). Performance of a virtual runoff hydrograph system. Journal of Water Resources Planning and Management, ASCE, 121(6), 421-427.

Chang, A.T.C., and Tsang,I. (1992). A neural network approach to inversion of snow water equivalent from passive microwave measurements. Nordic Hydrology, 23, 173-182.

Dawson, C.W., and Wilby,R. (1998). An artificial neural network approach to rainfall-runoff modelling. Hydrological Sciences Journal, 43(1), 47-66.

Fernando, D.A.K., and A.W. Jayawardena (2000). Runoff forecasting using RBF networks with OLS algorithm. Journal of Hydrologic Engineering, ASCE, 393), 203-209.

French, M. N., Krajewski, W. F., and Cuykendall, R. R. (1992). Rainfall forecasting in space and time using a neural network. Journal of Hydrology, 137(1-4), 1-31.

Govindaraju, R.S., and A.R. Rao (2000). Artificial Neural Networks in Hydrology, Kluwer Academic Publishers, Dordrecht.

Govindaraju, R.S., and B. Zhang (2000). Radial-basis function networks. In Artificial Neural Networks in Hydrology, edited by R.S. Govindaraju and A.R. Rao, Kluwer Academic Publishers, Dordrecht.

Hsu, K-L., Gupta, H.V., and Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process. Water Resources Research, 31(10), 2517-2530.

Gupta, H.V., Hsu, K-L., and Sorooshian, S. (2000). 'Effective and efficient modeling for streamflow forecasting' in Artificial Neural Networks in Hydrology, edited by R.S. Govindaraju and A.R. Rao, Kluwer Academic Publishers, Dordrecht.

Imrie, C.E. , Durucan, S., and Korre, A. (2000). Riverflow prediction using artificial neural networks: generalisation beyond the calibration range. Journal of Hydrology, 233,138-153.

Jain, S.K., Das, A., and Srivastava, D.K. (1999). Application of ANN for reservoir inflow prediction and operation. Journal of Water Resources Planning and management, ASCE, 125(5), 263-271.

Jain, S.K., and Chalisgaonkar, D. (2000). Setting stage-discharge relations using ANN. Journal of Hydrologic Engineering, ASCE, 594), 428-433.

Jain, S.K. (2001). Development of Integrated Sediment Rating Curves using Artificial Neural Networks. Journal of Hydraulic Engineering, ASCE, 127(1), 30-37.

Karunanithi, N., Grenney, W.J., Whitley, D., and Bovee, K. (1994). Neural networks for river flow prediction. J. Comp. in Civ. Engrg., ASCE, 8(2), 210-220.

Kothari, R., and Agyepong, K. (1996). On lateral connections in feed-forward neural networks. Proc. IEEE International Conference on Neural Networks, 13-18. Institute of Electrical and Electronics Engineers, New York.

Kumar, M., Raghuwanshi, N.S., Singh, R., Wallender, W.W., and Pruitt, W.O. (2002). Estimating evapotranspiration using artificial neural network. Journal of Irrigation and Drainage Engineering, ASCE, 128(4), 224-233.

Maren, A.J., Hartson, C.T., and Pap, R.M. (1990). Handbook of Neural Computing Applications. Academic Press, San Diego.

Minasny, B., and McBratney, A.B. (2002). The neuro-m method for fitting neural networks parametric pedotransfer functions. Journal of Soil Society of America, 66, 353-361.

Minns, A.W., and Hall, M.J. (1996). Artificial Neural Networks as rainfall runoff models. Hydrological Sciences Journal, 41(3), 399-418.

Raman, H., and Sunilkumar, N. (1995). Multivariate modeling of Water Resources time-series using Artificial Neural Networks. Hydrological Sciences Journal. 40(2), 145-163.

Raman, H., and Chandramauli, V. (1996). Deriving a general operating policy for reservoir using Neural Network. Journal of Water Resources Planning and Management, 122(5), 342-347.

Rumelhart, D.E., McLelland, J.L. and the PDP Research Group, (1986). Parallel Distributed Processing, Explorations in the Micro Structure of Cognition, Vol. I: Foundations, The MIT Press, Cambridge.

Smith, J., and Eli, R. N. (1995). Neural Network models of rainfall-runoff process. Journal of Water Resources Planning and Management, 121(6), 499-508.

Thirumalaiah, K. and Deo, M.C. (2000). Hydrological forecasting using neural networks. Journal of Hydrologic Engineering, ASCE, 5(2), 180-189.

Tokar, A.S., and Markus,M. (2000). Precipitation-runoff modeling using artificial neural networks and conceptual models. Journal of Hydrologic Engineering, ASCE, 5(2), 156-161.

Zhang, B., and Govindaraju,R.S., (2000). Modular neural networks for watershed runoff. In Artificial Neural Networks in Hydrology, edited by R.S. Govindaraju and A.R. Rao, Kluwer Academic Publishers, Dordrecht.

REFERENCES (ES)

Armijos, A., Wright, J. R., and Houck, M. H. (1990). Bayesian inferencing applied to real-time reservoir operations. Journal of Water Resources Planning and Management, ASCE, 116(1): 38-51.

Barr, A. and Feigenbaum, E.A. (eds.) (1981). The Handbook of Artificial Intelligence, Vol.1, Morgan Kaufman.

Baffaut, C. and Delleur, J. W. (1990). Calibration of SWMM runoff quality model with expert system. ASCE Journal of Water Resources Planning and Management, 116(2): 247-261.

Chow, K. C. A., and Watt, W. E. (1990). A knowledge-based expert system for flood frequency analysis. Canadian Journal of Civil Engineering, 17(4): 597-609.

Cuena, I. (1983). The use of simulation models and human advice to build an expert system for the defense and control of river floods. Proceedings of eighth IJCAI, 246-249.

Delleur, J.W. (1988). A front end expert system for the calibration of SWMM runoff block, Annual Water Res. Conf., WRP&M Div., ASCE, Norfolk, Va.

Engel, B. A. and Beasley, D. B. (1991). DSS: Dam site selector expert system for education. ASCE Journal of Irrigation and Drainage Engineering, 117(5): 774-782.

Engman, E.T., et al. (1986). An expert system for snowmelt runoff modeling and forecasting. Proceedings of Water Forum '86, ASCE, pp.174-180.

Fedra, K., Winkelbauer, L., and Pantulu, V.R. (1991). Expert Systems for Environmental Screening. An Application in the Lower Mekong Basin. RR-91-19. International Institute for Applied Systems Analysis. A-2361 Laxenburg, Austria.

Fischer, H., and Schultz, G.A. (1991). An expert system for real-time operation of a multi-

purpose multi-unit reservoir system. Proceedings of the Symposium on Hydrology of Natural and Manmade Lakes. IAHS Pub. No. 206, 151-157.

Gasching, J., Reboh, R., and Reiter, J. (1981). Development of knowledge-based expert system for water resources problems. SRI Project 1619. SRI International, USA.

Hayes-Rolth, F.D., Waterman, D.A., and Lenat, D.B. (1983). Building Expert Systems, Addison-Wesley, Reading, Mass.

Jamieson, D. G. and Fedra, K. (1996a). The 'WaterWare' decision-support system for river-basin planning. 1: Conceptual design. Journal of Hydrology, 177(3-4): 163-175.

Jamieson, D.G. and Fedra, K. (1996b). The 'Water Ware' decision-support system for river-basin planning 3. Example applications. Journal of Hydrology, 177(3-4), 199-211.

Lindberg, S. and Nielsen, J.B. (1986). Modeling of urban storm sewer systems in 31, pp.687-696.

Maher, M.L. (1986). Problem solving using expert system techniques, in Proceedings of Symposium Expert Systems in Civil Engineering. Kostem, C.N., and Maher, M.L. (Editors), Seattle, USA.

Maksimovic, C. and Radojkovic, M. (1989). Computational Modelling and Experimental Methods in Hydraulics, Proc. HYDROCOMP '89, Elsevier.

Palmer, R. N. and Holmes, K. J. (1988). Operational guidance during droughts: Expert system approach. ASCE Journal of Water Resources Planning and Management, 114(6): 647-666.

Raman, H., Mohan, S., and Rangacharya, N. C. V. (1992). Decision support for crop planning during droughts. Journal of Irrigation and Drainage Engineering, ASCE, 118(2): 229-241.

Rouhani, S., and Kangari, R. (1987). Expert systems in water resources. In Water for the future: hydrology in perspective. Proceedings of Rome Symposium. IAHS Publication No. 164.

Savic, D.A., and Simonovic, S.P. (1989). Introduction of fuzzy reasoning in intelligent decision support for reservoir analysis, in Maksimovic, C. and M. Radojkovic, 1989. Computational Modelling and Experimental Methods in Hydraulics, Proceedings of HYDROCOMP '89, pp. 404-414, Elsevier.

Sell, P.S. (1985). Expert system – A Practical Introduction, Macmillan, London.

Shepherd, A. and Ortolano, L. (1996). Water-supply system operations: Critiquing expert-system approach. ASCE Journal of Water Resources Planning and Management, 122(5): 348-355.

Simonovic, S.P. (1989). Expert system design for intelligent decision support for surface water quality data management, pp. 383-393, in Maksimovic, C. and M. Radojkovic, 1989. Computational Modeling and Experimental Methods in Hydraulics, Proceedings of HYDROCOMP '89, Elsevier, Dordrecht.

Simonovic, S. P. (1990). An expert system for the selection of a suitable method for flow measurement in open channels. Journal of Hydrology, 112(3-4): 237-256.

Simonovic, S. P. (1991). Knowledge-based systems and operational hydrology. Canadian Journal of Civil Engineering, 18(1): 1-11.

Simonovic, S. P., Venema, H. D., and Burn, D. H. (1992). Risk-based parameter selection for short-term reservoir operation. Journal of Hydrology, 131(1-4): 269-291.

Sponemann, P. and Fuchs, L. (1986). Use of artificial Intelligence for real time control in Sriram, D. and R. Adey (eds.). Applications of AI in Engineering Problems, Vol. II,

pp. 415-422. Springer Verlag.

Turban, E. (1995). Decision support system and expert system: Management support system. Prentice-Hall Inc., Englewood Cliffs, New Jersey.

Varas, E. A. and Von Chrismar, M. (1995). Expert system for the selection of methods to calculate design flood flows. Hydrological Sciences Journal, 40(6): 739-749.

Wilson, J.L. et. al (1986). GEOTOX: A knowledge based system for hazard site evaluation in Sriram, D. and R. Adey (eds.). Applications of AI in Engineering Problems, Vol. II, pp. 661-671, Springer Verlag.

Wishart, S. J., Lumbers, J. P., and Griffiths, I. M. (1990). Expert systems for the interpretation of river quality data. Journal of Ins. of Water and Environment Management, 4(2), 194-202.