READER CT5401, Lecture 4, Introduction remote sensing

This reader provides most of the theoretical and practical background for the exercise that comes with Lecture 4. This reader consists of relevant chapters from the book *Remote Sensing Digital Image Analysis, An Introduction*, by John A. Richards, ISBN0-387-16007-8, and a specific chapter on mapping small reservoirs (=exercise) from the M.Sc. thesis by Jens Liebe. It is recommended to also briefly look at the slides of the lecture that will be provided on Blackboard.

For any further questions, please contact me (n.c.vandegiesen@tudelft.nl).

Content:

I *Chapter 1 Richards*: Pay special attention to:
1.1, 1.3 (no details needed but look especially at the Landsat info) and 1.6 (Rest just background info, not essential)

II *Chapter 3 Richards*: Pay special attention to:
3.4, 3.5, 3.6 (Rest just background info, not essential)

III *Sections 8.1 and 8.2 Richards*: Read all

IV *Appendix E Richards*: Mathematical background for maximum likelihood classifier, mainly meant as background info)

V *Chapter 3.2 M.Sc. Thesis Jens Liebe*
(see: http://www.glowa-volta.de/publications/printed/thesis_liebe.pdf for complete thesis). This is essential reading and describes the techniques for the exercise in most detail. Perhaps start with this part and read other chapters as needed.

Chapter 1

Sources and Characteristics of Remote Sensing Image Data

1.1 Introduction to Data Sources

1.1.1 Characteristics of Digital Image Data

Remote sensing image data of the earth's surface acquired from either aircraft or spacecraft platforms is readily available in digital format; spatially the data is composed of discrete picture elements, or *pixels*, and radiometrically it is quantised into discrete brightness levels. Even data that is not recorded in digital form initially can be converted into discrete data by use of digitising equipment such as scanning microdensitometers.

The great advantage of having data available digitally is that it can be processed by computer either for machine assisted information extraction or for embellishment before an image product is formed. The latter is used to assist the role of photointerpretation.

A major characteristic of an image in remote sensing is the wavelength band it represents. Some images are measurements of the spatial disposition of reflected solar radiation in the ultraviolet, visible and near-to-middle infrared range of wavelengths. Others are measurements of the spatial distribution of energy emitted by the earth itself (dominant in the so-called *thermal* infrared wavelength range); yet others, particularly in the microwave band of wavelengths, measure the relative return from the earth's surface of energy actually transmitted from the vehicle itself. Systems of this last type are referred to as *active* since the energy source is provided by the remote sensing platform; by comparison remote sensing measurements that depend upon an external energy source, such as the sun, are called *passive*.

From a data handling and analysis point of view the properties of image data of significance are the number and location of the spectral measurements (or spectral bands) provided by a particular sensor, the spatial resolution as described by the pixel size, in equivalent ground metres, and the radiometric resolution. The last describes the range and discernable number of discrete brightness values and is sometimes referred to alternatively as dynamic range or signal to noise ratio. Frequently the radiometric resolution is expressed in terms of the number of binary digits, or bits, necessary to represent the range of available brightness values. Thus data with 8 bit radiometric resolution has 256 levels of brightness (see Appendix B).

Together with the frame size of an image, in equivalent ground kilometres, the number of spectral bands, radiometric resolution and spatial resolution determine the data volume provided by a particular sensor and thus establish the amount of data to be processed, at least in principle. As an illustration consider the Landsat Thematic Mapper instrument. It has 7 wavelength bands with 8 bit radiometric resolution, six of

which have 30 m spatial resolution and one of which has a spatial resolution of 120 m (the thermal band, for which the wavelength is so long that a larger aperture is necessary to collect sufficient signal energy to maintain the radiometric resolution) An image frame of 185 km × 185 km therefore represents 2.37 million pixels in the thermal band and 38 million pixels in the other six bands. At 8 bits per pixel a complete 7 band image is composed of 1.848×10^9 bits or 1 848 Gbit; alternatively and more commonly the data value would be expressed as 231 Mbytes, where 1 byte is equivalent to 8 bits.

The later sections of this chapter provide an overview of common remote sensing missions and their sensors in terms of the data-related properties just described This is necessary to provide orders of magnitude and other parameters necessary in determining time requirements and other figures of merit used in assessing the usefulness of image analysis procedures; it will also place the analytical material in context with the data gathering phase of remote sensing Before proceeding to such an overview of typical data sources it is of value to examine the spectral dimension in some detail since the choice of the spectral bands for a particular sensor significantly determines the information that can be extracted from the data for a particular application.

1.1.2 Spectral Ranges Commonly Used in Remote Sensing

In principle, remote sensing systems could measure energy emanating from the earth's surface in any sensible range of wavelengths However technological considerations, the selective opacity of the earth's atmosphere, scattering from atmospheric particulates and the significance of the data provided exclude certain wavelengths. The major ranges utilized for earth resources sensing are between about 0.4 and 12 μm (referred to below as the visible/infrared range) and between about 30 to 300 mm (referred to below as the microwave range) At microwave wavelengths it is often more common to use frequency rather than wavelength to describe ranges of importance Thus the microwave range of 30 to 300 mm corresponds to frequencies between 1 GHz and 10 GHz For atmospheric remote sensing frequencies in the range 20 GHz to 60 GHz are encountered

The significance of these different ranges lies in the interaction mechanism between the electromagnetic radiation and the materials being interrogated In the visible/infrared range the reflected energy measured by a sensor depends upon properties such as the pigmentation, moisture content and cellular structure of vegetation, the mineral and moisture contents of soils and the level of sedimentation of water. At the thermal end of the infrared range it is heat capacity and other thermal properties of the surface and near subsurface that control the strength of radiation detected In the microwave range, using active imaging systems based upon radar techniques, the roughness of the cover type being detected and its electrical properties, expressed in terms of complex permittivity (which in turn is strongly influenced by moisture content) determine the magnitude of the reflected signal. In the range 20 to 60 GHz, atmospheric oxygen and water vapour have a strong effect on transmission and thus can be inferred by measurements in that range. Thus each range of wavelength has its own strengths in terms of the information it can contribute to the remote sensing process. Consequently we find systems available that are optimised for and operate in particular spectral ranges, and provide data that complements that from other sensors.

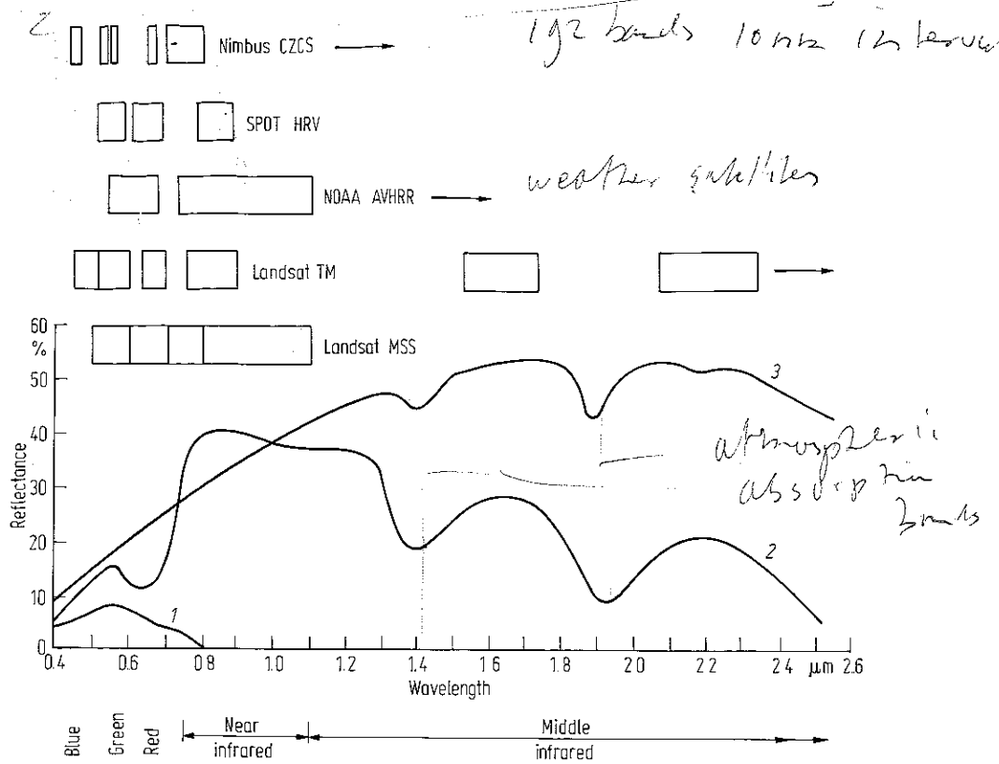


Fig. 1.1. Spectral reflectance characteristics of common earth surface materials in the visible and near-to-mid infrared range *1* Water, *2* vegetation, *3* soil The positions of spectral bands for common remote sensing satellites are indicated These are discussed in the following sections

Figure 1 1 depicts how the three dominant earth surface materials of soil, vegetation and water reflect the sun's energy in the visible/infrared range of wavelengths It is seen that water reflects about 10% or less in the blue-green range, a smaller percentage in the red and certainly no energy in the infrared range. Should the water contain suspended sediments or should a clear water body be shallow enough to allow reflection from the bottom then an increase in apparent water reflection will occur, including a small but significant amount of energy in the near infrared range. This is a result of reflection from the suspension or bottom material

As seen in the figure soils have a reflection that increases approximately monotonically with wavelength, however with dips centred at about 1.4 μm, 1 9 μm and 2.7 μm owing to moisture content. These water absorption bands are almost unnoticeable in very dry soils and sands In addition to these bands clay soils also have hydroxyl absorption bands at 1.4 μm and 2.2 μm.

The vegetation curve is considerably more complex than the other two. In the middle infrared range it is dominated by the water absorption bands at 1 4 μm, 1 9 μm and 2 7 μm The plateau between about 0 7 μm and 1 3 μm is dominated by plant cell

structure while in the visible range of wavelengths it is plant pigmentation that is the major determinant The curve sketched in Fig 1.1 is for healthy green vegetation This has chlorophyll absorption bands in the blue and red regions leaving only green reflection of any significance This is why we see chlorophyll pigmented plants as green.

An excellent review and discussion of the spectral reflectance characteristics of vegetation, soils, water, snow and clouds can be found in Hoffer (1978) This includes a consideration of the physical and biological factors that influence the shapes of the curves, and an indication of how the various cover types appear in images recorded in different wavelength ranges

In wavelength ranges between about 3 and 14 μm the level of solar energy actually irradiating the earth's surface is small owing to both the small amount of energy leaving the sun in this range by comparison to the higher levels in the visible and near infrared range (see Fig 1.2), and the presence of strong atmospheric absorption bands between 2.6 and 3.0 μm, 4 2 and 4.4 μm, and 5 and 8 μm (Chahine, 1983) Consequently much remote sensing in these bands is of energy being emitted from the earth's surface or objects on the ground rather than of reflected solar radiation.

Figure 1 2 shows the relative amount of energy radiated from perfect black bodies of different temperatures As seen the sun at 6000 K radiates maximally in the visible and near infrared regime but by comparison generates little radiation in the range around 10 μm Incidentally, the figure shown does not take any account of how the level of solar radiation is dispersed through the inverse square law process in its travel from the sun to the earth Consequently if it is desired to compare that curve to others corresponding to black bodies on the earth's surface then it should be considerably reduced overall.

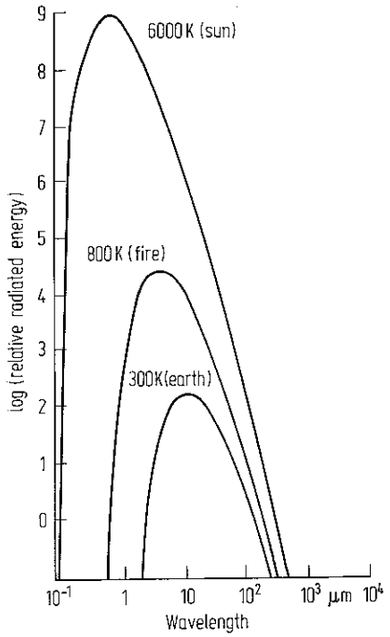


Fig. 1.2 Energy from perfect radiators (black bodies) as a function of wavelength

The earth, at a temperature of about 300 K has its maximum emission around 10 to 12 μm Thus a sensor with sensitivity in this range will measure the amount of heat being radiated from the earth itself Hot bodies on the earth's surface, such as bushfires, at around 800 K have a maximum emission in the range of about 3 to 5 μm Consequently to map fires, a sensor operating in that range would be used.

Real objects do not behave as perfect black body radiators but rather emit energy at a lower level than that shown in Fig 1.2 The degree to which an object radiates by comparison to a black body is referred to as its emittance Thermal remote sensing is sensitive therefore to a combination of an object's temperature and emittance, the last being wavelength dependent.

Microwave remote sensing image data is gathered by measuring the strength of energy scattered back to the satellite or aircraft in response to energy transmitted The degree of reflection is characterized by the scattering coefficient for the surface being imaged As noted earlier this is a function of the electrical complex permittivity of the surface material and the roughness of the surface in comparison to a wavelength of the radiation used Roughness is assessed according to the Rayleigh criterion which, for radiation with a wavelength λ incident at an angle θ to the normal of a surface with a local vertical variation of h, describes a surface as smooth if $h < \frac{\lambda}{8 \cos \theta}$ and rough otherwise.

Smooth surfaces act as so-called specular reflectors (i.e. mirror-like) in that the direction of scattering is predominantly away from the incident direction as shown in Fig 1 3. Consequently they appear dark to black in image data. Rough surfaces act as diffuse reflectors; they scatter the incident energy in all directions as depicted in Fig 1.3, including back towards the remote sensing platform As a result they appear light in image data. A third type of scattering mechanism is often encountered in microwave image data, particularly associated with man-made features such as buildings This is a corner reflector effect, as seen in Fig 1.3, resulting from the right angle formed between

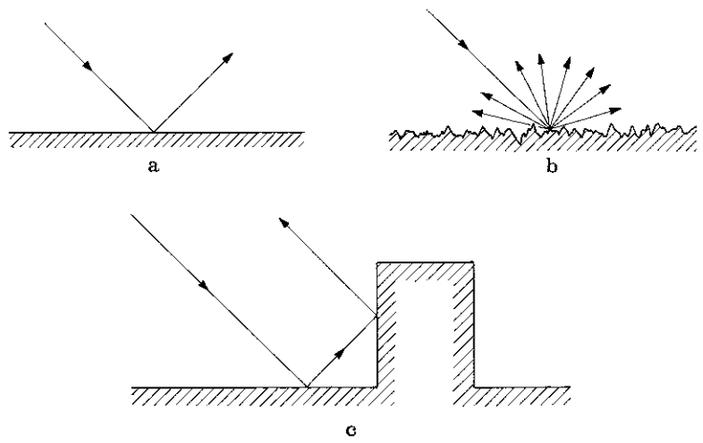


Fig. 1.3 **a** Specular, **b** diffuse and **c** corner reflector behaviour, encountered in the formation of microwave image data

a vertical structure such as a fence, building or ship and a horizontal plane such as the surface of the earth or sea

In interpreting image data acquired in the microwave region of the electromagnetic spectrum it is important to recognise that the three reflection mechanisms of Fig. 1.3 are present and modify substantially the tonal differences resulting from surface complex permittivity variations. By comparison, imaging in the visible/infrared range in which the sun is the energy source, results almost always in diffuse reflection, allowing the interpreter to concentrate on tonal variations resulting from factors such as those described in association with Fig. 1.1.

An outstanding treatment of the essential principles of microwave remote sensing will be found in the three volume series by Ulaby, Moore and Fung (1981, 1982, 1985).

1.1.3 Concluding Remarks

The purpose of acquiring remote sensing image data is to be able to identify and assess, by some means, either surface materials or their spatial properties. Inspection of Fig. 1.1 reveals that cover type identification should be possible if the sensor gathers data at several wavelengths. For example, if for each pixel measurements of reflection at 0.65 μm and 1.0 μm were available (i.e. we had a two band imaging system) then it should be a relatively easy matter to discriminate between the three fundamental cover types based on the relative values in the two bands. For example, vegetation would be bright at 1.0 μm and very dark at 0.65 μm whereas soil would be bright in both ranges. Water on the other hand would be black at 1.0 μm and dull at 0.65 μm. Clearly if more than two measurement wavelengths were used more precise discrimination should be possible, even with cover types spectrally similar to each other. Consequently remote sensing imaging systems are designed with wavebands that take several samples of the spectral reflectance curves of Fig. 1.1. For each pixel the set of samples can be analysed, either by photointerpretation, or by the automated techniques to be found in Chap. 8 and 9, to provide a label that associates the pixel with a particular earth surface material.

A similar situation applies when using microwave image data; viz. several different transmission wavelengths can be used to assist in identification of cover types by reason of their different scattering cross sections with wavelength. However a further data dimension is available with microwave imaging owing to the coherent nature of the radiation used. That relates to the polarizations of the transmitted and scattered radiation. The polarization of an electromagnetic wave refers to the orientation of the electric field during propagation. Often for spaceborne radar systems this is chosen to be parallel to the earth's surface on transmission, a situation referred to as *horizontal polarization*. On scattering some polarization changes can occur and energy can be received as horizontally polarized and/or vertically polarized. The degree of polarization rotation that occurs can also be a useful indicator of surface material.

Another consequence of using coherent radiation in radar remote sensing systems, of significance to the interpretation process, is that images exhibit a degree of "speckle". This is a result of constructive and destructive interference of the reflections from surfaces that have random spatial variations of the order of one half a wavelength, or so. Noting that the wavelengths commonly employed in radar remote sensing are between about 30 mm and 220 mm it is usual to find images of most common cover types showing a considerably speckled appearance. Within a homogeneous region for example, such as a crop field, this causes adjacent radar image pixels to have large differences in brightness, a factor which complicates machine-assisted interpretation.

1.2 Weather Satellite Sensors

Weather satellites and those used for earth resources sensing operate in much the same bands of wavelength. Perhaps the major distinctions in the image data they provide lies in the spatial resolutions available. Whereas data acquired for earth resources purposes generally has a pixel size less than 100 m, that used for meteorological applications usually has a much coarser pixel – often of the order of 1 km × 1 km. This is the distinction used herein in order to separate the two types of sensor. Having made that distinction however it is important to note that because of the similarity in wavebands, meteorological satellite data such as that from the NOAA Advanced Very High Resolution Radiometer (AVHRR) does find application in remote sensing when large synoptic views are required.

1.2.1 Polar Orbiting and Geosynchronous Satellites

Two broad types of weather satellite are in common use. One is of the polar orbiting, or more generally low earth orbit, variety whereas the other is at geosynchronus altitudes. The former typically have orbits at altitudes of about 700 to 1500 km whereas the geostationary altitude is approximately 36,000 km (see Appendix A). Typical of the low orbit satellites are the current NOAA series, and their forerunners the TIROS, TOS and ITOS satellites. The principal sensor of interest from this book's viewpoint is the NOAA AVHRR mentioned above. This is described in Sect. 1.2.2 following.

The Nimbus satellites, while strictly test bed vehicles for a range of meteorological and remote sensing sensors, also orbit at altitudes of around 1000 km. Nimbus sensors of interest include the Coastal Zone Colour Scanner (CZCS) and the Scanning Multichannel Microwave Radiometer (SMMR). Only the former is treated below.

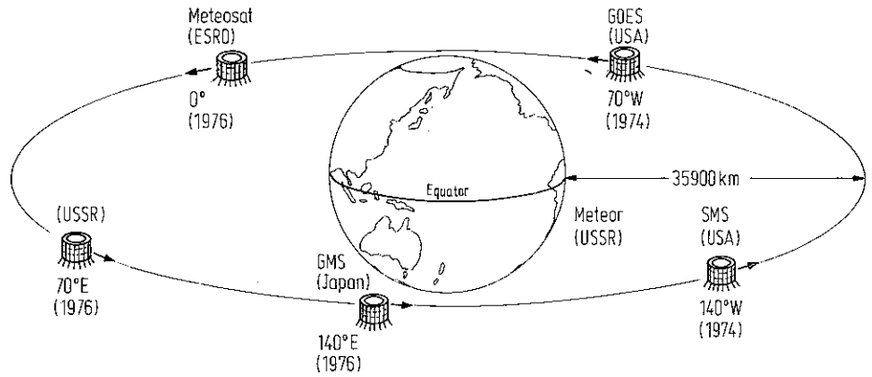


Fig. 1.4. Positions of the five geosynchronous meteorological satellites that provide global weather watch capabilities

Geostationary meteorological satellites have been launched by the United States, USSR, ESA and Japan to form a global atmospheric monitoring system. These are placed in equatorial geosynchronous orbits at the longitudes shown in Fig 1.4. The primary sensor carried on these satellites is the Visible and Infrared Spin Scan Radiometer (VISSR).

1.2.2 The NOAA AVHRR (Advanced Very High Resolution Radiometer)

The AVHHR has been designed to provide information for hydrologic, oceanographic and meteorologic studies, although data provided by the sensor does find application also to solid earth monitoring. An earlier version of the AVHRR contained four wavelength bands. Table 1.1 however lists the bands available on the current generation of instrument (NOAA 7).

Table 1.1. NOAA advanced very high resolution radiometer

Ground resolution	1.1 km at nadir
Dynamic range	10 bit
Ground swath	2700 km
Spectral bands:	
band 1	0.58 – 0.68 μm
band 2	0.725 – 1.1 μm
band 3	3.55 – 3.93 μm
band 4	10.3 – 11.3 μm
band 5	11.5 – 12.5 μm

1.2.3 The Nimbus CZCS (Coastal Zone Colour Scanner)

The CZCS is a mirror scanning system, carried on Nimbus 7 designed to measure chlorophyll concentration, sediment distribution and general ocean dynamics including sea surface temperature. Its characteristics are summarised in Table 1.2.

1.2.4 GMS VISSR (Visible and Infrared Spin Scan Radiometer)

The geostationary meteorological satellites such as GMS (Japan), SMS, GOES (USA) are spin stabilized with their spin axis oriented almost north-south. The primary sensor on these, the VISSR, scans the earth's surface by making use of the satellite spin to acquire one line of image data (as compared with an oscillating mirror in the case AVHRR, CZCS, MSS and TM sensors), and by utilizing a stepping motor to adjust the angle of view on each spin to acquire successive lines of data (on orbiting satellites it is the motion of the vehicle relative to the earth that displaces the sensor between successive scan lines). The characteristics of the VISSR are summarised in Table 1.3.

An improved version of the VISSR referred to as the VAS-VISSR Atmospheric Sounder (although VAS as an acronym has more widespread applications) has been in use since late 1980 on the US GOES series of meteorological satellite (Allison and Schnapf, 1983).

Table 1.2. Nimbus coastal zone colour scanner

Ground resolution	825 m at nadir
Dynamic range	8 bit
Ground swath	1566 km
Spectral bands:	
channel 1	0.433 – 0.453 μm
channel 2	0.510 – 0.530 μm
channel 3	0.540 – 0.560 μm
channel 4	0.660 – 0.680 μm
channel 5	0.700 – 0.800 μm
channel 6	10.5 – 12.5 μm

Table 1.3. GMS visible and infrared spin scan radiometer

Band		Spatial resolution	Dynamic range
Visible	0.55 – 0.75 μm	1.25 km at nadir	6 bit
Thermal infrared	10.5 – 12.5 μm	5.0 km	8 bit

1.3 Earth Resource Satellite Sensors in the Visible and Infrared Regions

1.3.1 The Landsat System

The Landsat earth resources satellite system was the first designed to provide near global coverage of the earth's surface on a regular and predictable basis. Consequently it has been the "workhorse" image data acquisition system for remote sensing purposes and has become the benchmark against which newer systems are judged.

The first three Landsats have had identical orbit characteristics, as summarised in Table 1.4. The orbits are near polar and are sun synchronous – i.e., the orbital plane precesses about the earth at the same rate that the sun appears to move across the face of the earth. In this manner data is acquired at about the same local time on every pass.

All satellites acquire image data nominally at 9:30 a.m. local time on a descending (north to south) path; in addition Landsat 3 obtained thermal data on a night-time ascending orbit for the few months that its thermal sensor was operational. Fourteen complete orbits are covered each day, and the fifteenth, at the start of the next day, is 159 km advanced from orbit 1, thus giving a second day coverage contiguous with that of the first day. This advance in daily coverage continues for 18 days and then repeats. Consequently complete coverage of the earth's surface is given, with 251 revolutions in 18 days.

Table 1.4. Landsat 1, 2, 3 orbit characteristics

Orbit:	Sun synchronous, near polar; nominal 9:30 am descending equatorial crossing; inclined at about 99° to the equator
Altitude:	920 km (570 mi)
Period:	103 min
Repeat cycle:	14 orbits per day over 18 days (251 revolutions)

The orbital characteristics of the second generation Landsats, commencing with Landsats 4 and 5 are different form those of their predecessors. Again image data is acquired nominally at 9:30 a m local time in a near polar, sun synchronous orbit; however the spacecraft are at the lower altitude of 705 km, both to assist in achieving higher resolution and to aid shuttle recovery for refurbishing. This lower orbit gives a repeat cycle of 16 days at 14 56 orbits per day This correspond to a total of 233 revolutions every cycle. Table 1 5 summarises the Landsat 4, 5 orbit characteristics Unlike the orbital pattern for the first generation Landsats, the day 2 ground pattern for Landsats 4 and 5 is not adjacent and immediately to the west of the day 1 orbital pattern. Rather it is displaced the equivalent of 7 swath centres to the west. Over 16 days this leads to the repeat cycle.

Table 1.5. Orbit parameters for Landsats 4, 5

Orbit:	Near polar, sun synchronous; nominal 9:30 am descending equatorial crossing.
Altitude:	705 km
Period:	98.9 min
Repeat Cycle:	14 56 orbits per day over 16 days (total of 233 revolutions)

It is intended that Landsats 4 and 5 be retrievable by Space Shuttle, and that ultimately they make use of GPS navigational satellite data for positioning information (U S. Geological Survey 1982) Also, whereas Landsats 1, 2 and 3 contained on-board tape recorders for temporary storage of image data whilst the satellites were out-of-view of earth stations, Landsats 4 and 5 do not, and depend on transmission either to earth stations directly or via the geosynchronous communication satellite TDRS (Tracking and Data Relay Satellite). TDRS is a high capacity communication satellite, leased to NASA, that is used to relay data from a number of missions, including the Space Shuttle. Its ground receiving station is in White Sands, New Mexico from which data is relayed via domestic communication satellites.

1.3.2 The Landsat Instrument Complement

Three Imaging instruments have been used with the Landsat satellites. These are the Return Beam Vidicon (RBV), the Multispectral Scanner (MSS) and the Thematic Mapper (TM). Table 1.6 shows the actual imaging payload for each satellite along with historical data on launch and out-of-service dates Two different RBV's were used: a multispectral RBV package was incorporated on the first two satellites, whilst a panchromatic instrument with a higher spatial resolution was used on Landsat 3. The MSS on Landsat 3 also contained a thermal band, as discussed in Sect. 1 3.4; however this operated only for a few months

It is intended that the MSS will not be used after Landsat 5. With the launches of Landsats 6 and 7 (currently scheduled for 1988 and 1991 respectively) an improved TM will be carried along with newer devices such as a multispectral linear array scanner.

The following sections provide an overview of the three Landsat instruments, especially from a data characteristic point-of-view. More details on mechanical considerations, design issues, sampling strategies etc will be found in Freden and Gordon (1983), Silva (1978) and U.S. Geological Survey (1979).

Table 1.6. Landsat payloads, launch and out of service dates

Satellite	Imaging Instruments			Launched	Out-of-service
Landsat 1	RBV^{m}	MSS		23 Jul 1972	6 Jan 1978
Landsat 2	RBV^{m}	MSS		22 Jan 1975	27 Jul 1983
Landsat 3	RBV^{p}	MSS^{t}		5 Mar 1978	7 Sept 1983
Landsat 4		MSS	TM	16 Jul 1982	–
Landsat 5		MSS	TM	1 Mar 1984	–

m – multispectral RBV
p – panchromatic RBV
t – MSS with thermal band

1.3.3 The Return Beam Vidicon (RBV)

As the name suggests the RBV's were essentially television camera-like instruments that took "snapshot" images of the earth's surface along the ground track of the satellite. Image frames of 185 km × 185 km were acquired with each shot, repeated at 25 s intervals to give contiguous frames in the along track direction at the equivalent ground speed of the satellite

Three RBV cameras were used on Landsats 1 and 2, distinguished by different transmission filters that allowed three spectral bands of data to be recorded as shown in Table 1.7. In Landsat 3 two RBV cameras were used; however both operated panchromatically and were focussed to record data swaths of 98 km, overlapped to give a total swath of about 185 km. By so doing a higher spatial resolution of 40 m was possible, by comparison to 80 m for the earlier RBV system.

Historically the spectral ranges recorded by the RBV's on Landsats 1 and 2 were referred to as bands1, 2 and 3. The MSS bands (see following) in the first generation of Landsats were numbered to follow on in this sequence.

1.3.4 The Multispectral Scanner (MSS)

The Multispectral Scanner has been the principal sensor on Landsats 1, 2 and 3 and was the same on each spacecraft with the exception of an additional band on Landsat 3 The MSS is a mechanical scanning device that acquires data by scanning the earth's surface in strips normal to the satellite motion. Six lines are swept simultaneously by an oscillating mirror and the reflected solar radiation so monitored is detected in four wavelength bands for Landsats 1 and 2, and five bands for Landsat 3, as shown in Table 1.7 A schematic illustration of the six line scanning pattern used by the MSS is shown in Fig. 1 5 It is seen that the sweep pattern gives rise to an MSS swath width of 185 km thereby corresponding to the image width of the RBV The width of each scan line corresponds to 79 m on the earth's surface so that the six lines simultaneously correspond to 474 m. Approximately 390 complete six-line scans are collected to provide an effective image that is also 185 km in the along track direction. For Landsats 1 and 2, 24 signal detectors are required to provide four spectral bands from each of the six scan lines. A further two are added from the thermal band data of Landsat 3. These detectors are illuminated by radiation reflected from the oscillating scanning mirror in the MSS, and produce a continuously varying electrical signal corresponding to the energy received along the 79 m wide associated scan line. The

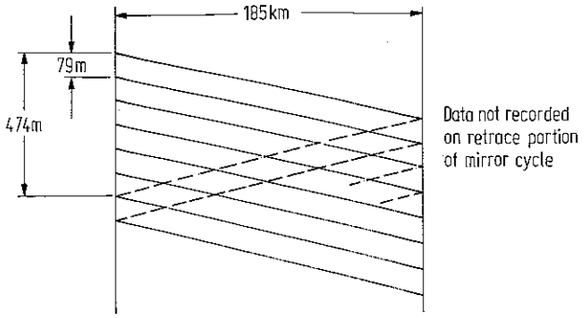


Fig. 1.5. The six line scanning pattern used by the Landsat multispectral scanner. Dimensions are in equivalent measurements on the ground. This scanning pattern is the same in each of bands 4 to 7. The same six line pattern is used on Landsats 4 and 5 except that the strip width is 81.5 m and 82.5 m respectively

optical aperture of the MSS and its detectors for bands 4 to 7 is such that at any instant of time each detector sees a pixel that is 79 m in size also along the scan line. Consequently the effective pixel size (or instantaneous field of view IFOV) of the detectors is 79 m × 79 m. At a given instant the output from a detector is the integrated response from all cover types present in a 79 m × 79 m region of the earth's surface. Without any further processing the signal from the detector would appear to be varying continuously with time. However it is sampled in time to produce discrete measurements across a scan line. The sampling rate corresponds to pixel centres of 56 m giving a 23 m overlap of the 79 m × 79 m pixels, as depicted in Fig. 1.6. The thermal infrared band on Landsat 3, band 8, has an IFOV of 239 m × 239 m. As a result there are only two band 8 scan lines corresponding to the six for bands 4 to 7, as indicated above.

The IFOV's of the multispectral scanners on Landsats 4 and 5 have been modified to 81.5 m and 82.5 m respectively although the pixel centre spacing of 56 m has been retained. In addition the bands have been renamed as bands 1, 2, 3 and 4, corresponding to bands 4, 5, 6 and 7 from the earlier missions.

After being spatially sampled, the data from the detectors is digitised in amplitude into 6 bit words. Before so encoding, the data for bands 4, 5 and 6 is compressed allowing decompression into effective 7 bit words upon reception at a ground station.

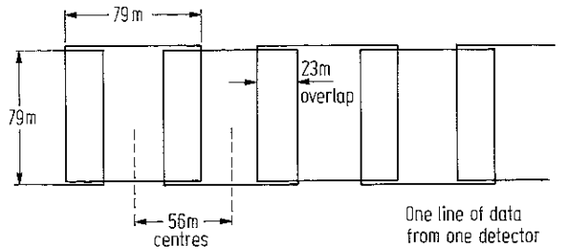


Fig. 1.6. The relationship between instantaneous field of view and pixel overlap for Landsat MSS pixels

Table 1.7. Characteristics of the Landsat imaging devices

Instrument	Spectral bands (μm)		IFOV (m)	Dynamic range (bits)
RBV[m]	1	0.475– 0.575 (blue)	79 × 79	
	2	0.580– 0.680 (red)	79 × 79	
	3	0.698– 0.830 (near IR)	79 × 79	
RBV[p]		0.505– 0.750 (panchromatic)	40 × 40	
MSS	4[a]	0.5 – 0.6 (green)	79 × 79	7
	5	0.6 – 0.7 (red)	79 × 79	7
	6	0.7 – 0.8 (near IR)	79 × 79	7
	7	0.8 – 1.1 (near IR)	79 × 79	6
	8[b]	10.4 –12.6 (thermal)	237 × 237	
TM	1	0.45 – 0.52 (blue)	30 × 30	8
	2	0.52 – 0.60 (green)	30 × 30	8
	3	0.63 – 0.69 (red)	30 × 30	8
	4	0.76 – 0.90 (near IR)	30 × 30	8
	5	1.55 – 1.75 (mid IR)	30 × 30	8
	7[c]	2.08 – 2.35 (mid IR)	30 × 30	8
	6	10.4 –12.5 (thermal)	120 × 120	8

[a] MSS bands 4 to 7 have been renumbered MSS bands 1 to 4 from Landsat 4 onwards. IFOV = 81.5, 82.5 m for Landsats 4, 5

[b] MSS band 8 was used only on Landsat 3

[c] TM band 7 is out of sequence since it was added last in the design after the previous six bands had been firmly established. It was incorporated at the request of the geological community owing to the importance of the 2 μm region in assessing hydrothermal alteration

1.3.5 The Thematic Mapper (TM)

The Thematic Mapper is a mechanical scanning device as for the MSS, but has improved spectral, spatial and radiometric characteristics. Seven wavelength bands are used, with coverage as shown in Table 1.7. Note that band 7 is out of place in the progression of wavelengths, it having been added, after the initial planning phase, at the request of the geological community.

Whereas the MSS of all Landsats scans and obtains data in one direction only, the TM acquires data in both scan directions again with a swath width of 185 km. Sixteen scan lines are swept simultaneously giving a 480 m strip across the satellite path, as illustrated in Fig. 1.7. This permits a lower mirror scan rate compared with the MSS and thus gives a higher effective dwell time for a given spot on the ground, making possible the higher spatial resolution and improved dynamic range.

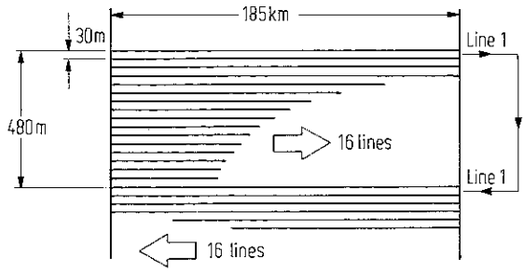


Fig. 1.7. Scanning characteristics of the Landsat Thematic Mapper

1.3.6 The SPOT High Resolution Visible (HRV) Imaging Instrument

The French satellite SPOT (Systeme Probatoire d'Observation de la Terre), planned for launch in 1986 carries two imaging devices referred to as HRV's These instruments utilize a different technology for image acquisition from that employed in the Landsat MSS and TM devices Rather than using oscillating mirrors to provide cross-track scanning during the forward motion of the space platform, the SPOT HRV instruments consist of a linear array of charge coupled device (CCD) detectors These form what is commonly referred as a "push broom" scanner Each detector in the array scans a strip in the along track direction – by having several thousand such detectors a wide swath can be imaged without the need for mechanical scanning Moreover, owing to the long effective dwell time this allows for each pixel, a higher spatial resolution is possible. A trade-off however is that charge coupled device technology currently is not available for wavelengths into the middle infrared range. Consequently the spectral bands provided by the HRV are not unlike those of the Landsat MSS

The HRV covers a ground swath width of 60 km; two instruments are mounted side by side in the spacecraft to give a total swath width of 117 km, there being a 3 km overlap of the individual swaths

Two imaging modes are possible One is a multispectral mode and the other panchromatic. The imaging characteristics of these are summarised, along with the satellite orbital properties, in Table 1.8.

Table 1.8. Spot satellite and HRV sensor characteristics

SPOT:		Altitude	832 km
		Orbit	sun synchronous, 98.7° inclination 10:30 am nominal equator crossing
		Repeat cycle	26 days
HRV:		Multispectral mode	Panchromatic Mode
	bands	0.50–0.59 μm 0.61–0.68 μm 0.79–0.89 μm	0.57–0.73 μm
	IFOV	20 m × 20 m	10 m × 10 m
	Dynamic range	8 bit	8 bit
	Total swath	117 km	117 km

An interesting property of the HRV is that it incorporates a steerable mirror to allow imaging to either side of nadir. This allows daily coverage for a short period along with a stereoscopic viewing capability.

1.3.7 The Skylab S192 Multispectral Scanner

Skylab was launched in May 1973 and taken out of service in February 1975 It carried a number of instrument packages, one of which was known as EREP (Earth Resources Experiment Package). It was put into an orbit inclined at 50° to the equator, at an altitude of 435 km. From this orbit it was able to view 75% of the Earth's surface.

The EREP package contained six sensor types, one of which was a multispectral scanner As with the Landsat MSS this scanner constructs images of the earth's surface, in discrete spectral or wavelength bands, by scanning or sweeping the earth's surface through an aperture, although in a conical fashion rather than a linear raster normal to the spacecraft's motion (Silva 1978). The spectral bands in the Skylab EREP multispectral scanner are described in Table 1.9

Table 1.9. Imaging characteristics of the Skylab S192 MSS

channel 1 0.41–0.46 μm	channel 8 0.98– 1.08 μm
2 0.46–0.51 μm	9 1.09– 1.19 μm
3 0.52–0.56 μm	10 1.20– 1.30 μm
4 0.56–0.61 μm	11 1.55– 1.75 μm
5 0.62–0.67 μm	12 2.10– 2.35 μm
6 0.68–0.76 μm	13 10.2 –12.5 μm
7 0.78–0.88 μm	
	spatial resolution = 79 m
	swath width = 72.4 km
	dynamic range = 8 bit

1.3.8 The Heat Capacity Mapping Radiometer (HCMR)

The Heat Capacity Mapping Mission (HCMM) satellite was the first of the planned "applications explorer missions" involving placement of small dedicated spacecraft in special orbits to satisfy mission-unique data acquisition requirements. It was launched in April 1978 into a sun synchronous orbit with an ascending equatorial crossing at 2:00 p.m. and a descending crossing at 2:00 a.m. Its altitude was 620 km with a repeat cycle of 16 days. It was taken out of service in September 1980

HCMM carried a heat capacity mapping radiometer (HCMR) with two spectral bands. One is in the thermal infrared region of 10.5 to 12.5 μm and the other in the reflective range 0.5 to 1.1 μm

The purposes of HCMM were to measure reflected solar energy, determine the heat capacity of rock types and to monitor soil moisture, snow cover, plant canopy temperature and thermal effluents. A summary of the HCMR characteristics is shown in Table 1.10, taken from the HCMM Data User's Handbook (Price, 1978).

Table 1.10. Summary of heat capacity mapping radiometer properties

Swath width	716 km
Repeat cycle	16 days
Bands:	0.5– 1.1 μm (visible and near IR)
	10.5–12.5 μm (thermal IR)
Resolution	600 m × 600 m at nadir (infrared)
	500 m × 500 m at nadir (visible)

1.4 Aircraft Scanners in the Visible and Infrared Regions

1.4.1 General Considerations

Multispectral line scanners, similar in principle to the Landsat MSS and TM instruments, have been available for use in civil aircraft since the late 1960's and early 1970's. As with satellite image acquisition it is the forward motion of the aircraft that provides along track scanning whereas a rotating mirror or a linear detector array provides sensing in the across track direction.

There are several operational features that distinguish the data provided by aircraft scanners from that produced by satellite-borne devices. These are of significance to the image processing task. First, the data volume can be substantially higher. This is a result of having (i) a large number of spectral bands or channels available – typically as high as twelve and (ii) a large number of pixels produced per mission, owing to high spatial resolution available. Frequently up to 1000 pixels may be recorded across the swath, with many thousands of scan lines making up a flight line; each pixel is normally encoded to 8 bits. Recording of the data during a mission usually requires therefore use of a high density digital tape recorder or a robust disc unit carried on board the aircraft.

A second feature of importance relates to field of view (FOV) – that is the scan angle either side of nadir over which data is recorded. This is depicted in Fig. 1.8. In the case of aircraft scanning the FOV, 2γ, is typically about 70 to 90°. Such a large angle is necessary to acquire an acceptable swath of data from aircraft altitudes. By comparison the FOV for the Landsats 1 to 3 is 11.56° while that for the Landsats 4 and 5 is slightly larger at about 15°. The consequence of the larger FOV with aircraft scanning is that significant distortions in image geometry can occur at the edges of the scan. Often these have to be corrected by digital processing.

Finally, the attitude stability of an aircraft as a remote sensing platform is much poorer than the stability of a satellite in orbit, particularly the Landsat 4 generation for which the pointing accuracy is 0.01° with a stability of 10^{-6} degrees per second. Because of atmospheric turbulence, variations in aircraft attitude described by pitch, roll and yaw can lead to excessive image distortion. Sometimes the aircraft scanner is mounted on a three axis stabilized platform to minimise these variations. It is more common however to have the scanner fixed with respect to the aircraft body and utilize a variable sampling window on the data stream to compensate for aircraft roll.

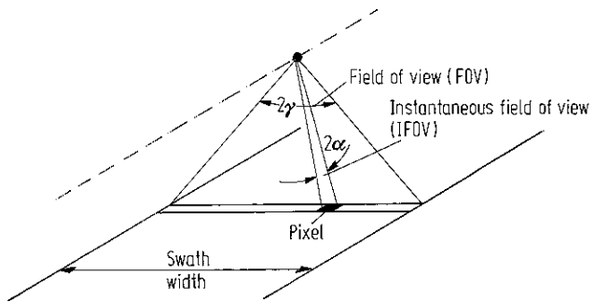


Fig. 1.8. The concept of field of view (FOV) and instantaneous field of view (IFOV)

Use of airborne multispectral scanners offers a number of benefits. Often the user can select the wavebands of interest in a particular application, and small bandwidths can be used. Also, the mission can be flown to specific user requirements concerning time of day, bearing angle and spatial resolution, the last being established by the aircraft height above ground level. As against these however, data acquisition from aircraft platforms is expensive by comparison with satellite recording since aircraft missions are generally flown for a single user and do not benefit from the volume market and synoptic view available to satellite data.

In the following sections the characteristics of a number of different commercial aircraft scanners are reviewed. Those chosen are simply illustrative of the range of device available.

1.4.2 The Daedalus AADS1240/1260 Multispectral Line Scanner

As an illustration of the properties of aircraft multispectral scanners and their data, the Daedalus AADS 1240/1260 is discussed here in a little detail. This is an example of an instrument with interchangeable detectors to permit data acquisition in a variety of waveband configurations. Image scanning is by means of a rotating mirror from which reflected light is passed via a series of mirrors and a dichroic lens to two sensor ports as shown in the diagram in Fig. 1.9. These ports can be fitted with two detectors in the thermal infrared range (ADDS 1240) or one infrared detector and a 10 channel spectrometer (AADS 1260). Alternatively a detector operating in the ultraviolet regime can be fitted to gone of the ports.

The wavebands, or channels, possible with this arrangement are summarised in Table 1.11. All data are recorded as 8 bit words; spatial resolution and swath width depend upon aircraft altitude. As an illustration, at 1000 m above ground level, the sub-nadir resolution on the ground is 2.5 m if the angular IFOV of the detector is 2.5 mrad. At this altitude a typical swath width, with an angular FOV of 86°, is 1.87 km.

Table 1.11. Spectral channels available with Daedalus visible and infrared multispectral line scanners

Ultraviolet	0	0.32 – 0.38 μm
Visible/Reflective IR	1	0.38 – 0.42 μm
	2	0.42 – 0.45 μm
	3	0.45 – 0.50 μm
	4	0.50 – 0.55 μm
	5	0.55 – 0.60 μm
	6	0.60 – 0.65 μm
	7	0.65 – 0.69 μm
	8	0.70 – 0.79 μm
	9	0.80 – 0.89 μm
	10	0.92 – 1.10 μm
Thermal	11	3.0 – 5.0 μm
	12	8.0 – 14.0 μm

FOV = 86°
IFOV = 2.5 mrad
Dynamic range = 8 bit

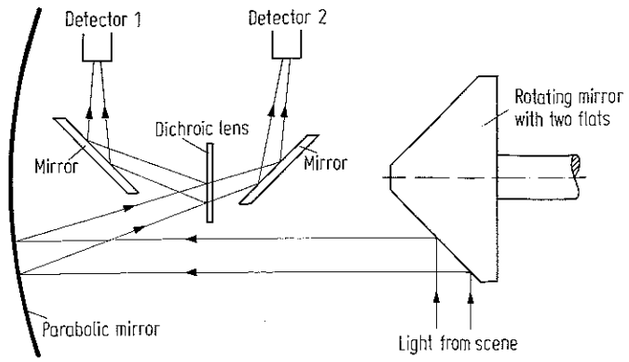


Fig. 1.9. Schematic diagram of the Daedalus infrared/visible aircraft multispectral scanner

1.4.3 The Airborne Thematic Mapper (ATM)

Because of the versatility possible with aircraft scanner band positioning it is a relatively straightforward matter to choose channels that correspond to spaceborne sensors. For example inspection of Table 1.11 shows that the spectral bands of the Landsat MSS can be approximately simulated by recording the sum of channels 4 and 5 (for MSS band 4), the sum of channels 6 and 7 (MSS band 5), channel 8 (MSS band 6) and the sum of channels 9 and 10 (for MSS band 7)

Because of the intense interest in thematic mapper simulation studies prior to the launch of Landsat 4, an aircraft scanner with bands that approximately matched those of the TM was designed. Referred to as an Airborne Thematic Mapper (ATM), this has the bands shown in Table 1.12. Again the spatial resolution achievable is a function of the altitude at which the instrument is flown. At an altitude of 12.5 km with an IFOV of 2.5 mrad this produces the same equivalent ground pixel as the Landsat TM of 30 m.

Table 1.12. Airborne thematic mapper spectral characteristics

Channel	Landsat TM equivalent band	
1 0.42 – 0.45 μm		
2 0.45 – 0.52 μm	1	
3 0.52 – 0.60 μm	2	
4 0.605– 0.625 μm		
5 0.63 – 0.69 μm	3	
6 0.695– 0.75 μm		FOV = 86°
7 0.76 – 0.90 μm	4	IFOV = 2.5 mrad
8 0.91 – 1.05 μm		Dynamic range = 8 bit
9 1.55 – 1.75 μm	5	
10 2.08 – 2.35 μm	7	
11 8.5 –13.0[a] μm	6	

[a] Thermal band is broader to permit same aircraft IFOV as preceding channels

1.4.4. The Thermal Infrared Multispectral Scanner (TIMS)

The thermal infrared range of wavelength, typically between 8 and 12 μm, contains spectral emission features that can be used to diagnose silicate rocks. Owing to stretching vibrations of the silicon-oxygen bond in the silicate crystal lattice, a broad mimimum in emissivity occurs between 8 and 11 μm. The depth and position of this band is related to the crystal structure of the constituent minerals, varying with changing quartz content. Clay silicates additionally have aluminium-oxygen-hydrogen bending modes, while carbonates exhibit no distinguishing spectral variability in the thermal infrared range (Kahle and Goetz 1983)

To exploit these vibrational features, as means for identifying rock and soil types using multispectral scanner data, it is necessary to have finer spectral resolution in the thermal regime than is possible with scanners such as those discussed previously. However, by using a dispersive grating, a multichannel thermal scanner has been developed by Daedalus Enterprises Inc for use by NASA/Jet Propulsion Laboratory for remote sensing of non-renewable resources. A summary of the characteristics of this instrument referred to as TIMS (Thermal Infrared Multispectral Scanner) is given in Table 1.13.

Table 1.13. Thermal infrared multispectral scanner characteristics

Channels	
1 8.2– 8.6 μm	
2 8.6– 9.0 μm	FOV = 76°
3 9.0– 9.4 μm	IFOV = 2.5 mrad
4 9.5–10.2 μm	Dynamic range = 8 bit
5 10.2–11.2 μm	
6 11.2–12.2 μm	

1.4.5 The MDA MEIS-II Linear Array Aircraft Scanner

The Canada Centre for Remote Sensing has developed an airborne scanner which employs linear array technology for across track sensing and recording, such as with the HRV sensor on the French SPOT satellite (Till et al., 1983) This permits a greater effective dwell time on a pixel thereby allowing a narrow spectral band width to be used if required for each spectral region of interest, without sacrificing IFOV or dynamic range (i.e. radiometric resolution) Known as MEIS-II (for Multispectral Electro-optical Imaging Scanner – II) the instrument is marketed by the Canadian firm of MacDonald, Dettwiler and Associates

Table 1.14. MEIS-II linear array scanner characteristics

Spectral range:	0.4–1.0 μm bands selected by front mounted filters
FOV:	39.66°[a]
IFOV:	0.7 mrad[a]
Dynamic range:	8 bit

[a] After real time resampling to remove geometric distortion and to achieve channel to channel pixel registration with high accuracy

Unlike the aircraft scanners treated previously, which make use of dichroic lenses and diffraction gratings to achieve spectral discrimination the MEIS-II instrument consists of eight separate charge coupled device detectors that have their own lenses and are geometrically registered. The spectral band for each detector is determined by interchangeable spectral filters placed in front of the lens.

Characteristics of the MEIS-II instrument are summarised in Table 1.14

1.5 Image Data Sources in the Microwave Region

1.5.1 Side Looking Airborne Radar and Synthetic Aperture Radar

Remote sensing image data in the microwave range of wavelengths is generally gathered using the technique of side-looking radar, as illustrated in Fig. 1.10. When used with aircraft platforms it is more commonly called SLAR (side looking airborne radar), a technique that requires some modification when used from spacecraft altitudes, as discussed in the following.

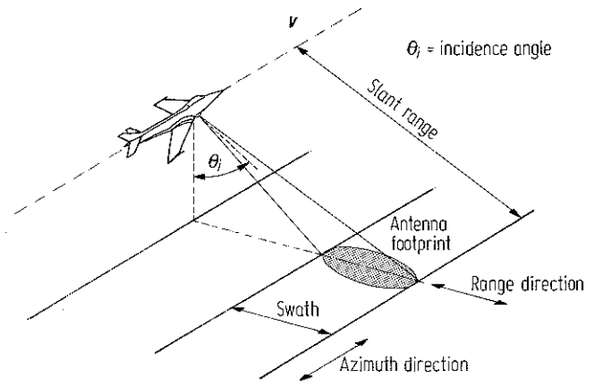


Fig. 1.10. Principle of side looking radar

In SLAR a pulse of electrical energy at the microwave frequency (or wavelength) of interest is radiated to the side of the aircraft at an incidence angle of θ_i. By the same principle as radars used for air navigation and shipping, some of this transmitted energy is scattered from the ground and returned to the receiver on the aircraft. The time delay between transmission and reflection identifies the slant distance to the "target" from the aircraft, while the strength of the return contains information on the so-called scattering cross-section of the target region of the earth's surface. The actual received signal from a single transmitted pulse consists of a continuum of reflections from the complete region of ground actually illuminated by the radar antenna. In Fig. 1.10 this can be identified as the range beamwidth of the antenna. This is chosen at design to give a relation between swath width and altitude, and tends to be rather broad. By comparison the along-track, or so-called azimuth, beamwidth is chosen as small as possible so that the reflections from a single transmitted pulse can be regarded

as having come from a narrow strip of terrain broadside to the aircraft. The forward velocity of the aircraft is then arranged so that the next transmitted pulse illuminates the next strip of terrain along the swath. In this manner the azimuth beamwidth of the antenna defines the spatial resolution in the azimuth direction whereas the time resolution possible between echos from two adjacent targets in the range direction defines the spatial resolution in the slant direction.

From an image product viewpoint the slant range resolution is not of interest. Rather it is the projection of this onto the horizontal plane as ground range resolution that is of value to the user. A little thought reveals that the ground range resolution is better at larger incidence angles and thus on the far side of the swath; indeed it can be shown that the ground range size of a resolution element (pixel) is given by

$$r_g = c\tau/2 \sin \theta_i$$

where τ is the length of the transmitted pulse and c is the velocity of light. (Often a simple pulse is not used. Instead a so-called linear chirped waveform is transmitted and signal processing on reception is used to compress this into a narrow pulse. For the present discussion however it is sufficient to consider the transmitted waveform to be a simple pulse or burst of the frequency of interest.)

The azimuth size of a resolution element is related to the length (or aperture) of the transmitting antenna in the azimuth direction, l, the wavelength λ and the range R_0 between the aircraft and the target, and is given by

$$r_a = R_0 \lambda / l$$

This expression shows that a 10 m antenna will yield an azimuth resolution of 20 m at a slant range of 1 km for radiation with a wavelength of 20 cm. However if the slant range is increased to say 100 km – i.e. at low spacecraft altitudes – then a 20 m azimuth resolution would require an antenna of 1 km length, which clearly is impracticable. Therefore when radar image data is to be acquired from spacecraft a modification of SLAR, referred to as synthetic aperture radar (SAR) is used. Essentially this utilizes the motion of the space vehicle, during transmission of the ranging pulses, to give an effectively long antenna, or a so-called synthetic aperture. This principle is illustrated in Fig. 1.11, wherein it is seen that an intentionally large azimuth beamwidth is employed

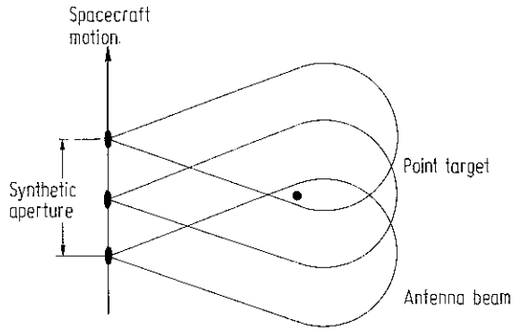


Fig. 1.11. The concept of synthesizing a large antenna by utilizing spacecraft motion along its orbital path. Here a view from above is shown, illustrating that a small real antenna is used to ensure a large real beamwidth in azimuth. As a consequence a point on the ground is illuminated by the full synthetic aperture

to ensure that a particular spot on the ground is illuminated and thus provides reflections over a length of spacecraft travel equivalent to the synthetic aperture required.

A discussion of the details of the synthetic aperture concept and the signal processing required to produce a high azimuth resolution is beyond the scope of this treatment. The matter is pursued further in Ulaby, Moore and Fung (1982), Elachi et al (1982), Tomiyasu (1978) and Cumming and Bennett (1979).

1.5.2 The Seasat SAR

The first earth observational space mission to carry a synthetic aperture imaging radar was the Seasat satellite launched in June 1978. Although only short lived it recorded about 126 million square kilometres of image data, including multiple coverage of many regions. Several other remote sensing instruments were also carried, including a radar altimeter, a scatterometer, a microwave radiometer and a visible and infrared imaging radiometer. Relevant characteristics of the satellite and its SAR are summarised in Table 1.15. Polarization referred to in this table relates to the orientation of the electric field vector in the transmitted and received waveforms. Free space propagation of electromagnetic energy, such as that used for radar, takes place as a wave with electric and magnetic field vectors normal to each other and also normal to the direction of propagation. Should the electric field vector be parallel to the earth's surface, the wave is said to be horizontally polarized. Should it be vertical then the wave is said to be vertically polarized. A wavefront with a combination of the two will be either elliptically or circularly polarized. Even though one particular polarization might be adopted for transmission, some rotation can occur when the energy is reflected from the ground. Consequently at the receiver often both vertically and horizontally polarized components are available, each having its own diagnostic properties concerning the earth cover type being sensed. Whether one or the other, or both, are received depends upon the antenna used with the radar. In the case of Seasat, horizontally polarized radiation was transmitted (***H***) and horizontally polarized returns were received (***H***).

Table 1.15. Characteristics of Seasat and its synthetic aperture radar

Altitude	800 km
Wavelength	0.235 m
Polarization	***HH***
Incidence angle	20°
Swath width	100 km
Range resolution	25 m
Azimuth resolution	25 m
Dynamic range	5 bit

Further details on the Seasat SAR will be found in Elachi et al. (1982).

1.5.3 Shuttle Imaging Radar-A (SIR-A)

A modified version of the Seasat SAR was flown as the SIR-A sensor on the second flight of Space Shuttle in November of 1981. Although the mission was shortened to three days, image data of about 10 million square kilometres was recorded. In contrast to Seasat however, in which the final image data was available digitally, the data in SIR-A was recorded and processed optically and thus is available only in film format. For digital processing therefore it is necessary to have areas of interest digitized from film using a device such as a scanning microdensitometer. A summary of SIR-A characteristics is given in Table 1.16, wherein it will be seen that the incidence angle was chosen quite different from that for the Seasat SAR. Interesting features of landform can be brought out by processing the two together.

Table 1.16. Characteristics of Shuttle Imaging Radar-A (SIR-A)

Altitude	245 km
Wavelength	0.235 m
Polarization	***HH***
Incidence angle	47°
Swath width	50 km
Range resolution	40 m
Azimuth resolution	40 m

More details on SIR-A will be found in Elachi et al. (1982) and Elachi (1983).

1.5.4 Shuttle Imaging Radar-B (SIR-B)

SIR-B, the second instrument in the NASA shuttle imaging radar program was carried on Space Shuttle mission 41 G in October 1984. Again the instrument was essentially the same as that used on Seasat and SIR-A, however the antenna was made mechanically steerable so that the incidence angle could be varied during the mission. Also about half the data was recorded digitally with the remainder being optically recorded. Details of the SIR-B mission are summarised in Table 1.17; NASA (1984) contains further information on the instrument and experiments planned for the mission. Because of the variable incidence angle both the range resolution and swath width also varied accordingly.

Table 1.17. Characteristics of Shuttle Imaging Radar-B (SIR-B)

Altitude	225 km–235 km
Wavelength	0.235 m
Polarization	***HH***
Incidence angle	15°–57°
Swath width	20–50 km
Range resolution	58–17 m
Azimuth resolution	25 m

A feature of the later SIR-C and SIR-D instruments will be the use of cross polarized returns (***VH*** and ***HV***) and the use of multiple wavelengths (L, band, X band and C band).

1.5.5 Aircraft Imaging Radar Systems

Airborne imaging radar systems in SLAR and SAR technologies are available for commercial project data acquisition. As with airborne multispectral scanners these

offer a number of advantages over equivalent satellite based systems including flexibility in establishing mission parameters (bearing, incidence angle, spatial resolution etc.) and proprietary rights to data However the cost of data acquisition is also high.

Table 1.18 summarises the characteristics of three commercial aircraft imaging radars, chosen to illustrate the operating parameter of these devices by comparison to satellite based systems

Table 1.18. Characteristics of aircraft imaging radars

	MARS LTD AN/APS-94D	MDA LTD IRIS	GOODYEAR/ GEMS
Technology	SLAR	SAR	SAR
On-board quick look processing	yes	yes	yes
Wavelength	X band	X, C, L band	X band
Range resolution	30 m	5 m to 30 m	~10 m
Azimuth resolution	IFOV = 7.8 mrad	6 m, 10 m	~10 m
Incidence angles	1° to 45°		
Swath width	25 km to 100 km		

1.6 Spatial Data Sources in General

1.6.1 Types of Spatial Data

The foregoing sections have addressed sources of multispectral digital image data of the earth's surface; each image considered has represented the spatial distribution of energy coming from the earth in one or several wavelength ranges in the electromagnetic spectrum Other sources of spatially distributed data are also often available for regions of interest. These include simple maps that show topography, land ownship, roads and the like, through to more specialised sources of spatial data such as maps of geophysical measurements of the area Frequently these other spatial data sources contain information not available in multispectral imagery and often judicious combinations of multispectral and other map-like data allow inferences to be drawn about regions on the earth's surface not possible when using a single source on its own. Consequently the image analyst ought to be aware of the range of spatial data available for a region and select that subset likely to assist in the information extraction process.

Table 1.19 is an illustration of the range of spatial data one might expect could be available for a given region. This differentiates the data into three types according as to whether it represents point information, line information of area information Irrespective of type however, for a spatial data set to be manipulated using the techniques of digital image processing it must share two characteristics with multispectral data of the types discussed in the previous sections. First it must be available in discrete form spatially, and in value. In other words it must consist of, or be able to be converted to, pixels with each pixel describing the properties of a given (small) area on the ground: the value ascribed to each pixel must be expressible in digital form. Secondly it must be in correct geographic relation to a multispectral image data set if the two are to be manipulated together. In situations where multispectral

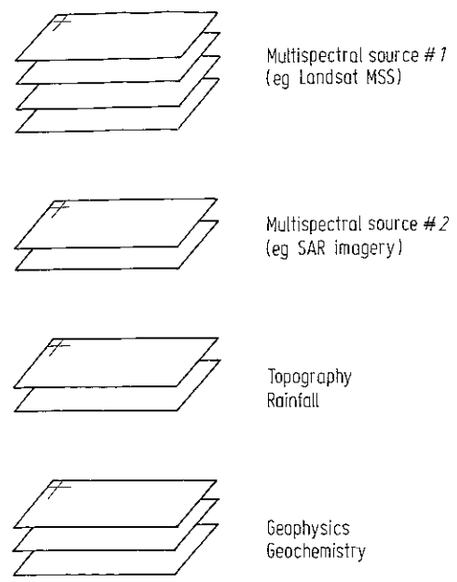


Fig. 1.12. An integrated spatial data source database

data is not used, the pixels in the spatial data source would normally be arranged to be referenced to a map grid It is usual however, in digital spatial data handling systems, to have *all* entries in the data set relating to a particular geographical region, mutually registered and referenced to a map base such as the UTM grid system. When available in this manner the data is said to be geocoded. Means by which different data sets can be registered are treated in Sect 2.5 Such a database is depicted in Fig 1.12

Table 1.19. Sources of spatial data

Point	Line	Area
Multispectral data	road maps	land ownership
Topography	powerline grids	town plans
Magnetic measurements	pipeline networks	geological maps
Gravity measurements		land use licenses
Radiometric measurements		land use maps
Rainfall		land cover maps
Geochemistry (in ppm)		soil type maps

1.6.2 Data Formats

Not all sources of spatial data are available originally in the pixel oriented digital format depicted in Fig. 1.12 Indeed often the data will be available as analog maps that require digitisation before entry into a digital data base That is particularly the case with line and area data types, in which case also consideration has to be given to the "value" that will be ascribed to a particular pixel. In line spatial data sources the pixels could be called zero if they were not part of a line and coded to some other number if

they formed part of a line of a given type. For a road map, for example, pixels that fall on highways might be given a value of 1 whereas those on secondary roads could be given a value of 2, and so on. On display using a colour output device the different numbers could be interpreted and output as different colours. In a similar manner numbers can be assigned to different regions when digitizing area spatial data sources.

Conceptually the digitization process may not be straightforward. Consider the case for example of needing to create a digital topographic map from its analog contour map counterpart. Figure 1.13 illustrates this process. First it is necessary to convert the contours on the paper map to records contained in a computer. This is done by using an input device such as a stylus or cross-hair cursor to mark a series of points on each contour between which the contour is regarded by the computer to be a straight line. Information on a contour at this stage is stored in the computer's memory as a file of points. This format is referred to as *vector format* owing to the vectors that can be drawn from point to point (in principle) to reconstruct a contour on a display or plotter. Some spatial data handling computer systems operate in vector format entirely. However to be able to exploit the techniques of digital image processing the vector formated data has to be turned into a set of pixels arranged on rectangular grid centres. This is referred to as *raster format* (or sometimes grid format); the elevation values for each pixel in the raster form are obtained by a process of interpolation over the points recorded on the contours. The operation is referred to as *vector to raster conversion* and is an essential step in entering map data into a digital spatial data base.

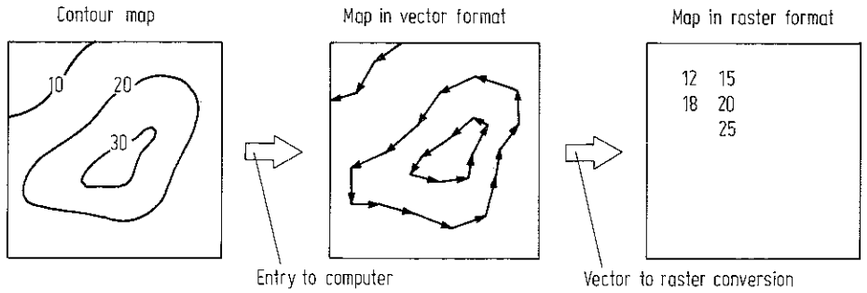


Fig. 1.13. Definition of vector and raster format using the illustration of digitising contour data

Raster format is a natural one for the representation of multispectral image data since data of that type is generated by digitising scanners, is transmitted digitally and is recorded digitally. Moreover many image forming devices such as filmwriters and television monitors operate on a raster display basis, compatible with digital data acquisition and storage. Raster format however is also appealing from a processing point of view since the logical records for the data are the pixel values (irrespective of whether the data is of the point, line or area type) and neighbourhood relationships are easy to establish by means of the pixel addresses. This is important for processing operations that involve near neighbouring groups of pixels. In contrast, vector format does not offer this feature. However an advantage of vector format, often exploited in high quality graphics display devices, is that resolution is not limited by pixel size.

1.6.3 Geographic Information Systems (GIS)

The amount of data to be handled in a database that contains spatial sources such as satellite and aircraft imagery along with maps, as listed in Table 1.19, is enormous, particularly if the data covers a large geographical region. Quite clearly therefore thought has to be given to efficient means by which the data types can be stored and retrieved, manipulated, analysed and displayed. This is the role of the geographic information system (GIS). Like its commercial counterpart, the management information system (MIS), the GIS is designed to carry out operations on the data stored in its database, according to a set of user specifications, without the user needing to be knowledgeable about how the data is stored and what data handling and processing procedures are utilized to retrieve and present the information required. Unfortunately because of the nature and volume of data involved in a GIS many of the MIS concepts developed for data base management systems (DBMS) cannot be transferred directly to GIS design although they do provide guidelines. Instead new design concepts have been needed, incorporating the sorts of operation normally carried out with spatial data, and attention has had to be given to efficient coding techniques to facilitate searching through the large numbers of maps and images often involved.

To understand the sorts of spatial data manipulation operations of importance in GIS one must take the view of the resource manager rather than the data analyst. Whereas the latter is concerned with image reconstruction, filtering, transformation and classification, the manager is interested in operations such as those listed in Table 1.20. These provide information from which management strategies and the like can be inferred. Certainly, to be able to implement many, if not most, of these a substantial amount of image processing may be required. However as GIS technology progresses it is expected that the actual image processing being performed would be transparent to the resource manager; the role of the data analyst will then be in part of the GIS design. A good summary of the essential issues to be faced in the establishment of a GIS, and a discussion of some systems implemented for particular purposes will be found in Marble and Peuquet (1983).

A problem which can arise in image data bases of the type encountered in a GIS is the need to identify one image by reason of its similarity to another. In principle, this could be done by comparing the images pixel-by-pixel; however the computational demand in so doing would be enormous for images of any practical size. Instead effort has been

Table 1.20. Some GIS data manipulation operations

Intersection and overlay of data sets (masking)
Intersection and overlay of polygons (grid cells, etc.) with spatial data
Identification of shapes
Identification of points in polygons
Area determination
Distance determination
Thematic mapping
Proximity calculations (shortest route, etc.)
Search by data
Search by location
Search by user-defined attribute
Similarity searching (e.g. of images)

directed to developing codes or signatures for complete images that will allow efficient similarity searching. For example an image histogram could be used (see Sect. 4.2); however as geometric detail is not preserved in a histogram this is rarely a suitable code for an image on its own. One effective possibility that has been explored is the use of image pyramids. A pyramid is created by combining groups of pixels in a neighbourhood to produce a new composite pixel of reduced resolution, and thus a low resolution image with fewer pixels. This process is repeated on the processed image to form a new image of lower resolution (and fewer pixels) still. Ultimately the image could be reduced to one single pixel that is a global measure of the image's brightness. Since pixels are combined in neighbourhood groups spatial detail is propagated up through the pyramid, albeit at decreasing resolution. Figure 1.14 illustrates how an image pyramid is constructed by simple averaging of non-overlapping sets of 2×2 pixels. It is a relatively easy matter (see Problem 1.6) to show that the additional memory required to store a complete pyramid, constructed as in the figure, is only 33% more than that required to store just the image itself.

Having developed an image pyramid, signatures that can be used to undertake similarity searching include the histograms computed over rows and colums in the uppermost levels of the pyramid (see Problem 1.7). A little thought shows that this allows an enormous number of images to be addressed, particularly if each pixel is represented by an 8 bit brightness value. As a result very fast searching can be carried out on these reduced representations of images.

Image pyramids are discussed by Rosenfeld (1982) and have been considered in the light of image similarity searching by Chien (1980).

There is sometimes an image processing advantage to be obtained when using a pyramid representation of an image. In edge detection, for example, it is possible to localise edges quickly, without having to search every pixel of an image, by finding apparent edges (regions) in the upper levels of the pyramid. The succeeding lower pixel groupings are then searched to localise the edges better.

Finally the pyramid representation of an image is felt to have some relation to human perception of images. The upper levels contain global features and are therefore not unlike the picture we have when first looking at a scene – generally we take the scene in

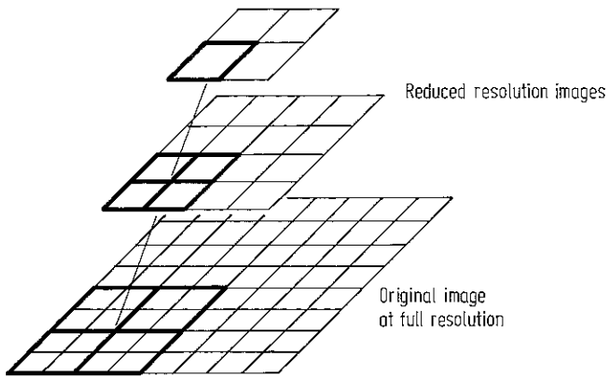


Fig. 1.14. Construction of an image pyramid by successively averaging groups of 2×2 pixels

initially "as a whole" and either miss or ignore detail. Then we focus on regions of interest for which we pay attention to detail because of the information it provides us with.

1.6.4 The Challenge to Image Processing and Analysis

Much of the experience gained with digital image processing and analysis in remote sensing has been with multispectral image data. In principle however any spatial data type in digital format can be processed using the techniques and procedures presented in this book. Information extraction from geophysical data could be facilitated, for example, if a degree of sharpening is applied prior to photointerpretation, while colour density slicing could assist the interpretation of topography. However the real challenge to the image analyst arises when data of mixed types are to be processed together. Several issues warrant comment.

The first relates to differences in resolution, an issue that arises also when treating multi-source satellite data such as Landsat MSS and NOAA AVHRR. The analyst must decide, for example what common pixel size will be used when co-registering the data, since either resolution or coverage will normally be sacrificed. Clearly this decision will be based on the needs of a particular application and is a challenge more to the analyst than the algorithms.

The more important consideration however is in relation to techniques for machine assisted interpretation. There is little doubt that combined multispectral and, say, topographic or land ownership maps can yield more precise thematic (i.e. category of land cover, etc.) information for a particular region than the multispectral data on its own. Indeed the combination of these sources is often employed in photointerpretive studies. However digitally, acceptable algorithms have yet to be devised that will permit integrated but diverse data types to be analysed automatically as effectively, and with the theoretical foundations, as have satellite and aircraft spectral data. Some attempts show promise and indicate the benefits of treating combined data (Strahler et al., 1980 and Richards et al., 1982) but fall short of providing an acceptable and general theoretical basis for machine-assisted analysis of mixed data. Such a basis will be required if geographic information systems are to incorporate a significant degree of digital processing and analysis between data retrieval and display.

The issue is complicated further when it is recalled that much of the non-spectral, spatial data available is not in enumerable point form but rather is in nominal area or line format. With these it seems clear that image analysis algorithms developed algebraically will not be suitable. Rather some degree of logical processing of labels combined with algebraic processing of arithmetic values (such as pixel brightnesses) will be necessary.

1.7 A Comparison of Scales in Digital Image Data

Because of IFOV differences the digital images provided by various remote sensing sensors will find application at different scales. As a guide Table 1.21 relates scale to spatial resolution; this has been derived somewhat simplistically by considering an image pixel to be too coarse if it approaches 0.1 mm in size on a photographic product

at a given scale Thus Landsat MSS data is suggested as being suitable for scales smaller than about 1:500,000 whereas NOAA AVHRR data is suitable for scales below 1:10,000,000 Detailed discussions of image quality in relation to scale will be found in Welch (1982) and Forster (1985).

Table 1 21. Suggested maximum scales of photographic products as a function of effective ground pixel size (Based on 0.1 mm printed pixel)

Scale	Pixel Size (m)	Sensor (nominal)
1: 50,000	5	aircraft MSS
1: 250,000	25	Spot HRV, Landsat TM
1: 500,000	50	Landsat MSS
1: 5,000,000	500	HCMM
1: 10,000,000	1000	NOAA AVHRR
1: 50,000,000	5000	GMS thermal IR band

References for Chapter 1

More details on the satellite programs treated in this chapter, along with information on sensors and data characteristics can be found in the *Manual of Remote Sensing* (1983). Many other missions are also covered, along with typical applications Other important sources that can be consulted include the *Landsat Data User's Handbook* (U.S. Geological Survey, 1979) and the *Landsat Data Users Notes*, published quarterly by the US National Oceanic and Atmospheric Administration Information on new satellite and Shuttle missions is often found in the magazine *Aviation Week and Space Technology* published by McGraw-Hill every two weeks. In March each year this contains a table of current remote sensing and other satellite missions

L J. Allison and A Schnapf, 1983: Meteorological Satellites. In Manual of Remote Sensing, loc cit

M. T. Chahine, 1983: Interaction Mechanisms within the Atmosphere. In Manual of Remote Sensing, loc cit

Y T. Chien, 1980: Hierarchical Data Structures for Picture Storage, Retrieval and Classification. In Pictorial Informations Systems, S. K. Chang and K. S. Fu (Eds.), Springer-Verlag, Berlin

I G. Cumming and J R Bennett, 1979: Digital Processing of Seasat SAR Data. Proc Int Conf. on Acoustics, Speech and Signal Processing, Washington D.C 2–4 April.

C Elachi (Chairman), 1983: Spaceborne Imaging Radar Symposium, Jet Propulsion Laboratory, January 17–20. JPL Publication 83–11

C Elachi, T Bicknell, R.L. Jordan & C. Wu, 1982: Spaceborne Synthetic Aperture Imaging Radars Applications, Techniques and Technology Proc. IEEE, 70, 1174–1209.

S Freden and F. Gordon, Jr 1983: Landsat Satellites In Manual of Remote Sensing, loc cit.

B C Forster, 1985: Mapping Potential of Future Spaceborne Remote Sensing Systems Institution of Surveyors (Australia) Annual Congress, Alice Springs

R M Hoffer, 1978: Biological and Physical Considerations in Applying Computer-Aided Analysis Techniques to Remote Sensor Data. In P H Swain and S M Davis, Eds., Remote Sensing: The Quantitative Approach, N.Y., McGraw-Hill

A B Kahle & A F H. Goetz, 1983: Mineralogic Information from a New Airborne Thermal Infrared Multispectral Scanner, Science, 222, 24–27.

Manual of Remote Sensing, 1983: R N Colwell (Ed.) 2e American Society of Photogrammetry, Falls Church, Va

D F Marble and D. J Pequet, 1983: Geographic Information Systems and Remote Sensing. In Manual of Remote Sensing, loc. cit

NASA, 1984: The SIR-B Science Investigations Plan, Jet Propulsion Laboratory Publication 84–3.

J C. Price, 1978: Heat Capacity Mapping Mission Users Guide, Goddard Space Flight Center, Dec.

J A. Richards, D.A Landgrebe and P H Swain, 1982: A Means for Utilizing Ancillary Information in Multispectral Classification. Remote Sensing of Environment, 12, 463–477.

A. Rosenfeld, 1982: Quadtrees and Pyramids: Hierarchical Representation of Images, Report TR-1171, Computer Vision Laboratory, University of Maryland

L F Silva, 1978: Radiation and Instrumentation in Remote Sensing. In P H Swain & S M Davis, Eds., Remote Sensing: The Quantitative Approach, N.Y., McGraw-Hill.

A H Strahler, J.E Estes, P.F. Maynard, F C. Mertz and D A Stow, 1980: Incorporating Collateral Data in Landsat Classification and Modeling Procedures Proc 14th Int Symp on Remote Sensing of Environment, Costa Rica.

S M Till, W D. McColl and R.A Neville, 1983: Development, Field Performance and Evaluation of the MEIS-II Multidetector Electro-Optical Imaging Scanner Proc. 17th Int. Symp. on Remote Sensing of Environment, Ann Arbor, Michigan, May 9–13.

K Tomiyasu, 1978: Tutorial Review of Synthetic-Aperture Radar (SAR) with Applications to Imaging of the Ocean Surface. Proc IEEE, 66, 563–583

R Welch, 1982: Image Quality Requirements for Mapping from Satellite Data. Proc. Int Soc Photogrammetry and Remote Sensing, Commission 1. Primary Data Acquisition, Canberra.

F T Ulaby, R K. Moore and A K Fung, 1981, 1982, 1985: Microwave Remote Sensing, Active and Passive Vols 1, 2, 3 Reading Mass Addison-Wesley.

U S Geological Survey, 1982: Landsat Data Users Notes, No 23, July.

U S Geological Survey, 1979: Landsat Data Users Handbook, Virginia

is no aliasing From sampling theory it is well known that the original data can be recovered by low pass filtering the spectrum, using the ideal filter as indicated in the figure Multiplication of the spectrum by this ideal filter is equivalent to convolving the original line samples by the inverse Fourier transform of the filter function From the theory of the Fourier transform, the inverse of the filter function is

$s(x)=\frac{2d}{\pi}\frac{\sin x}{x}$ with $x=\zeta/2d$ in which ζ is a spatial variable along lines of data, and d is the inter-pixel spacing This is known generally as an interpolating function Determine some cubic polynomial approximations to this function These could be determined from a simple Taylor series expansion or could be derived from cubic splines For a set of examples see Shlien (1979)

Chapter 3
The Interpretation of Digital Image Data

3.1 Two Approaches to Interpretation

When image data is available in digital form, spatially quantised into pixels and radiometrically quantised into discrete brightness levels, there are two approaches that may be adopted in endeavouring to extract information One involves the use of a computer to examine each pixel in the image individually with a view to making judgements about pixels specifically based upon their attributes This is referred to as *quantitative analysis* since pixels with like attributes are often counted to give area estimates Means for doing this are described in Sect 3 4 The other approach involves a human analyst/interpreter extracting information by visual inspection of an image composed from the image data In this he notes generally large scale features and is often unaware of the spatial and radiometric digitisations of the data This is referred to as *photointerpretation* or sometimes *image interpretation*; its success depends upon the analyst exploiting effectively the spatial, spectral and temporal elements present in the composed image product Information spatially, for example, is present in the qualities of shape, size, orientation and texture Roads, coastlines and river systems, fracture patterns, and lineaments generally, are usually readily identified by their spatial disposition Temporal data, such as the change in a particular object or cover type in an image from one date to another can often be used by the photointerpreter as, for example, in discriminating deciduous or ephemeral vegetation from perennial types Spectral clues are utilised in photointerpretation based upon the analyst's fore-knowledge of, and experience with, the spectral reflectance characteristics of typical ground cover types, and how those characteristics are sampled by the sensor on the satellite or aircraft used to acquire the image data

The two approaches to image interpretation have their own roles and often these are complementary Photointerpretation is aided substantially if a degree of digital image processing is applied to the image data beforehand, whilst quantitative analysis depends for its success on information provided at key stages by an analyst This information very often is drawn from photointerpretation

A comparison of the attributes of photointerpretation and quantitative analysis is given in Table 3.1 From this it can be concluded that photointerpretation, involving direct human interaction and therefore high level decisions, is good for spatial assessment but poor in quantitative accuracy Area estimates by photointerpretation, for instance, would involve planimetric measurement of regions identified visually; in this, boundary definition errors will prejudice area accuracy By contrast, quantitative analysis, requiring little human interaction and low level software decisions, has poor spatial ability but high quantitative accuracy Its high accuracy comes from the ability

of a computer, if required, to process every pixel in a given image and to take account of the full range of spectral, spatial and radiometric detail present. Its poor spatial properties come from the difficulty with which decisions about shape, size, orientation and texture can be made using standard sequential computing techniques. Developments in parallel computing and multiresolution image analysis may obviate this in due course.

Table 3.1. A comparison of the two approaches

Photointerpretation (by a human analyst/interpreter)	Quantitative analysis (by computer)
On a scale large relative to pixel size	At individual pixel level
Inaccurate area estimates	Accurate area estimates possible
Only limited multispectral analysis	Can perform true multispectral (multidimensional) analysis
Can assimilate only a limited number of distinct brightness levels (say 16 levels in each feature)	Can make use quantitatively of all available brightness levels in all features (e.g. 64, 128, 256)
Shape determination is easy	Shape determination involves complex software decisions
Spatial information is easy to use in a qualitative sense	Limited techniques available for making use of spatial data

3.2 Forms of Imagery for Photointerpretation

Multispectral or single dimensional image data can be procured either in photographic form or as digital data on magnetic tape. The latter is usually more expensive although more flexible since first, photographic products can be obtained via suitable apparatus from the digital data and secondly, the data can be processed digitally for enhancement before photographs are produced.

There are two fundamental types of photographic product. The first is a black and white reproduction of each band in the image data independently. If produced from the raw digital data then black will correspond to a digital brightness value of 0 whereas white will correspond to the highest digital value. This is usually 63, 127 or 255 (for 6 bit, 7 bit and 8 bit data respectively).

The second photographic product is a colour composite in which selected features or bands in multispectral data are chosen to be associated with the three additive colour primaries in the display device which produces the colour product. When the data consists of more than three features a judgement has to be made as to how to discard all but three, or alternatively, a mapping has to be invented that will allow all the features to be combined suitably into the three primaries. One possible mapping is the principal components transformation developed in Chap. 6. Usually this approach is not adopted since the three new features are synthetic and the analyst is therefore not able to call upon his experience of spectral reflectance characteristics. Instead a subset of original bands is chosen to form the colour composite. When the data available consist of a large number of bands (such as those produced by aircraft scanners or by the

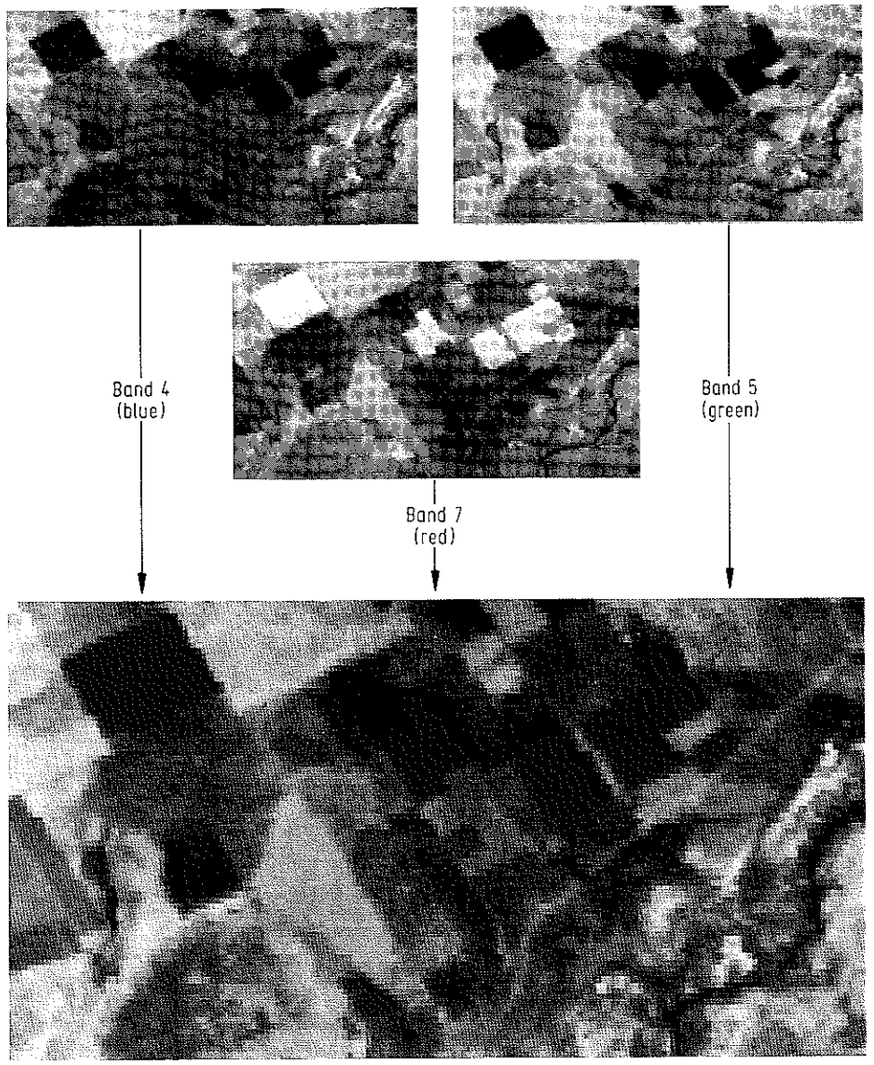


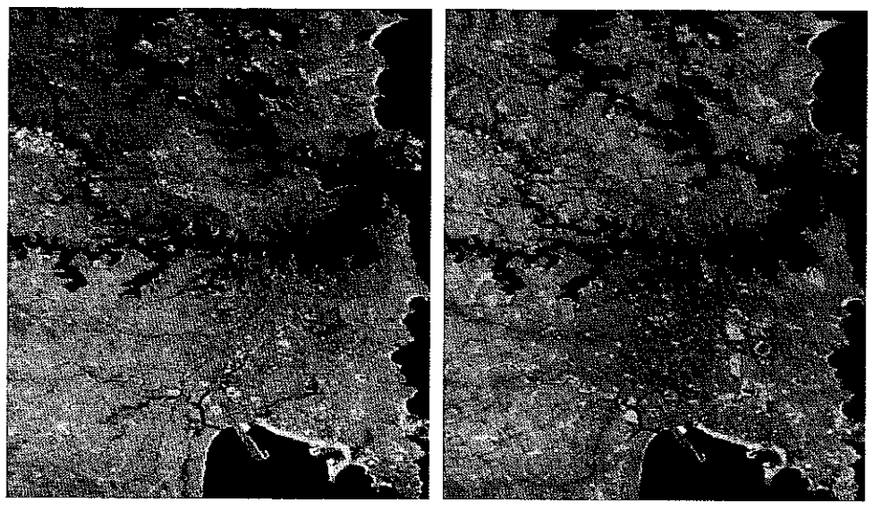
Fig. 3.1. Formation of a Landsat multispectral scanner false colour composite by displaying band 7 as red, band 5 as green and band 4 as blue

Landsat thematic mapper) only experience, and the area of application, tell which three bands should be combined into a colour product. For data with limited spectral bands however the choice is more straightforward. An example of this is Landsat multispectral scanner data. Of the available four bands, frequently band 6 is simply discarded since it is highly correlated with band 7 for most cover types and also is more highly

correlated with bands 4 and 5 than is band 7 Bands 4, 5 and 7 are then associated with the colour primaries in the order of increasing wavelength:

Landsat MSS Bands		Colour primaries
4 (green)		blue
5 (red)		green
7 (infrared)		red
	increasing wavelength	

An example of the colour product obtained by such a procedure is seen in Fig 3 1 This is often referred to as a false colour composite or sometimes, by association with colour infrared film, a colour infrared image. In this, vegetation shows as variations in red (owing to the high infrared response associated with vegetation), soils show as blues, greens and sometimes yellow and water as black or deep blue. These colour associations are easily determined by reference to the spectral reflectance characteristics of earth surface cover types in Sect 1.1; it is important also to take notice of the spectral imbalance created by computer enhancement of the brightness range in each wavelength band as discussed in Sect. 3.3 following



a b

Fig. 3.2. Standard Landsat multispectral scanner false colour composite **a** compared with a product in which band 7 has been displayed as green, band 5 as red and band 4 as blue **b**. Finer detail is more apparent in the second product owing to the sensitivity of the human vision system to yellow-green hues. The image segment shown is Sydney, the capital of New South Wales, Australia, acquired on December 14, 1980

It is of interest to note that the correlation matrix for the image of Fig. 3 1 is

	Band 4	Band 5	Band 6	Band 7
Band 4	1.00			
Band 5	0 85	1 00		
Band 6	0 31	0.39	1.00	
Band 7	−0 09	−0.07	0.86	1.00

wherein the redundancy present in band 6 can be seen.

In many ways the choice of colour assignments for the Landsat multispectral scanner bands is an unfortunate one since this yields, for most scenes, an image product substantially composed of reds and blues. These are the hues in which the human visual system is least sensitive to detail. Instead it would have been better to form an image in which yellows and greens predominate since then many fine details become more apparent. An illustration of this is given in Fig 3.2

3.3 Computer Processing for Photointerpretation

When image data is available in digital form it can be processed before an actual image or photograph is written in order to ensure that the clues used for photointerpretation are enhanced. Little can be done about temporal clues, but judicious processing makes spectral and spatial data more meaningful This processing is of two types One deals with the radiometric (or brightness value) character of the image and is termed radiometric enhancement. The other has to do with the image's perceived spatial or geometric character and is referred to as geometric enhancement. The latter normally involves such operations as smoothing noise present in the data, enhancing and highlighting edges, and detecting and enhancing lines. Radiometric enhancement is concerned with altering the contrast range occupied by the pixels in an image. From the point of view of computation radiometric enhancement procedures involve determining a new brightness value for a pixel (by some specified algorithm) from its existing brightness value. They are often referred to therefore as point operations and can be effectively implemented using look up tables. These are two column tables (either in software or hardware) that associate a set of new brightness values with the set of old brightnesses Specific radiometric enhancement techniques are treated in Chap. 4

Geometric enhancement procedures involve establishing a new brightness value for a pixel by using the existing brightnesses of pixels over a specified neighbourhood of pixels. These cannot be implemented in look up table form and are usually time-consuming to evaluate The range of geometric enhancement techniques commonly used in the treatment of remote sensing image data is given in Chap. 5 Both radiometric and geometric enhancement can be of value in highlighting spatial information. It is generally only radiometric or contrast enhancement, however, that amplifies an image's spectral character. A word of caution however is in order here. When contrast enhancement is utilised, each feature in the data is generally treated independently. This can lead to a loss of feature to feature relativity and thus, in the case of colour composites, can lead to loss of colour relativity. The reason for this is depicted in Fig. 3 3, and the effect is illustrated in Fig 3 4.

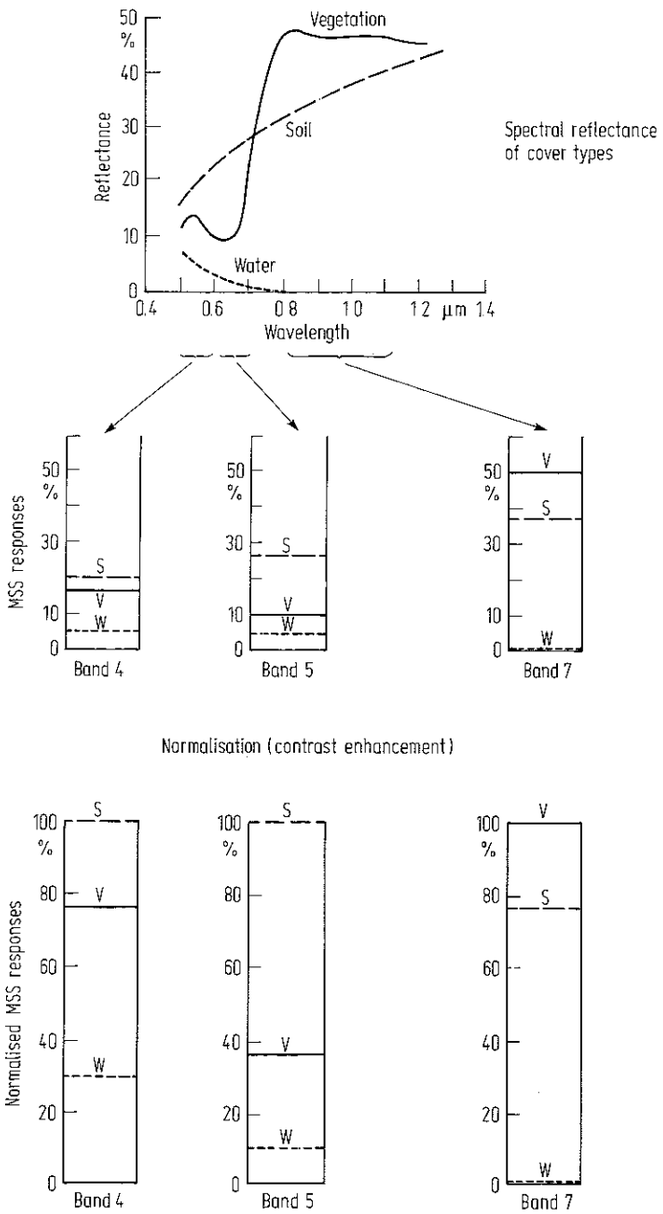
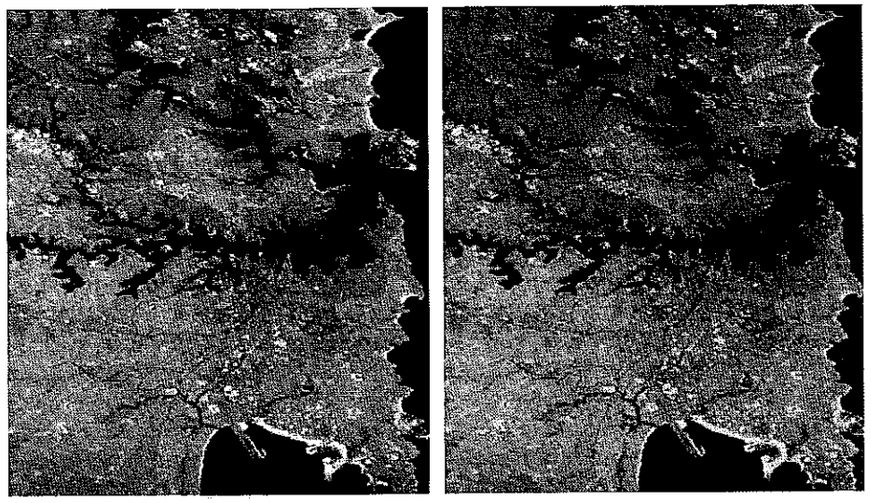


Fig. 3.3. Indication of how contrast enhancement (normalisation) can distort the feature to feature or band to band relativity (and thus colour relativity) in an image Without contrast enhancement both soil and vegetation cover types would have a reddish appearance, whereas after enhancement the soil takes on its characteristic bluish tones



a b

Fig. 3.4. Image in which each band has identical contrast enhancement before colour composition **a** compared to that in which each band has been enhanced independently to cover its full brightness range **b**. This causes a loss of band to band relativity and thus gives a different range of hues

3.4 An Introduction to Quantitative Analysis — Classification

Identification of features in remote sensing imagery by photointerpretation is effective for global assessment of geometric characteristics and general appraisal of ground cover types. It is, however, impracticable to apply at the pixel level unless only a handful of pixels is of interest. As a result it is of little value for determining accurate estimates of the area in an image corresponding to a particular ground cover type, such as the hectarage of a crop Moreover as noted in Sect 3 2, since photointerpretation is based upon the ability of the human analyst-interpreter to assimilate the available data, only three or so of the complete set of spectral components of an image can be used readily Yet there are four bands available in Landsat multispectral scanner imagery, seven for Landsat thematic mapper data and probably twelve or so for aircraft scanner data It is not that all of these would necessarily need to be used in the identification of a pixel; rather, should they all require consideration or evaluation, then the photointerpretive approach is clearly limited Furthermore the human analyst is unable to discriminate to the limit of the radiometric resolution available in scanner and other forms of imagery By comparison if a computer can be used for analysis, it could conceivably do so at the pixel level and could examine and identify as many pixels as required. In addition, it should be possible for computer analysis of remote sensing image data to take full account of the multidimensional aspect of the data, and its radiometric resolution

Computer interpretation of remote sensing image data is referred to as quantitative analysis because of its ability to identify pixels based upon their numerical properties

and owing to its ability for counting pixels for area estimates Systems available for implementing quantitative analysis techniques range from software packages on commercial computer networks to fully interactive dedicated image processing computers. In any of these, the procedures for quantitative analysis are classification based Classification is a method by which labels may be attached to pixels in view of their spectral character. This labelling is implemented by a computer by having trained it beforehand to recognise pixels with spectral similarities

Clearly the image data for quantitative analysis must be available in digital form. This is an advantage with more recent image types, such as Landsat multispectral scanner and thematic mapper imagery, as against more traditional aerial photographs. The latter require digitisation before quantitative analysis can be performed.

Detailed procedures and algorithms for quantitative analysis are the subject of Chap 8, 9 and 10; Chap. 11 is used to show how these are developed into classification methodologies for effective quantitative analysis. The remainder of this chapter however is used to provide an outline of the essential concepts in classification. As a start it is necessary to devise a model with which to represent remote sensing multispectral image data in a form amenable to the development of analytical procedures

3.5 Multispectral Space and Spectral Classes

The most effective means by which multispectral data can be represented in order to formulate algorithms for quantitative analysis is to plot them in a pattern space, or multispectral vector space, with as many dimensions as there are spectral components. In this space, each pixel of an image plots as a point with co-ordinates given by the brightness value of the pixel in each component This is illustrated in Fig 3 5 for a simple two dimensional band 5, band 7 Landsat MSS space Provided the spectral bands have been designed to provide good discrimination it is expected that pixels would form groups in multispectral space corresponding to various ground cover types, the sizes and shapes of the groups being dependent upon varieties of cover type, systematic noise and topographic effects. The groups or clusters of pixel points are referred to as *information classes* since they are the actual classes of data which a computer will need to be able to recognise

In practice the information class groupings may not be single clusters as depicted in Fig 3 5 Instead it is not unusual to find several clusters for the same region of soil, for the same apparent type of vegetation and so on for other cover types in a scene These are not only as a result of specific differences in types of cover but also a result of differences in moisture content, soil types underlying vegetation and topographic influences. Consequently, a multispectral space is more likely to appear as shown in Fig 3.6 wherein each information class is seen to be composed of several *spectral classes*

In many cases the information classes of interest do not form distinct clusters or groups of clusters but rather are part of a continuum of data in the multispectral space. This happens for example when, in a land systems exercise, there is a gradation of canopy clusure with position so that satellite or aircraft sensors might see a gradual variation in the mixture of canopy and understory. The information classes here might correspond to nominated percentage mixtures rather than to sets of well defined sub-

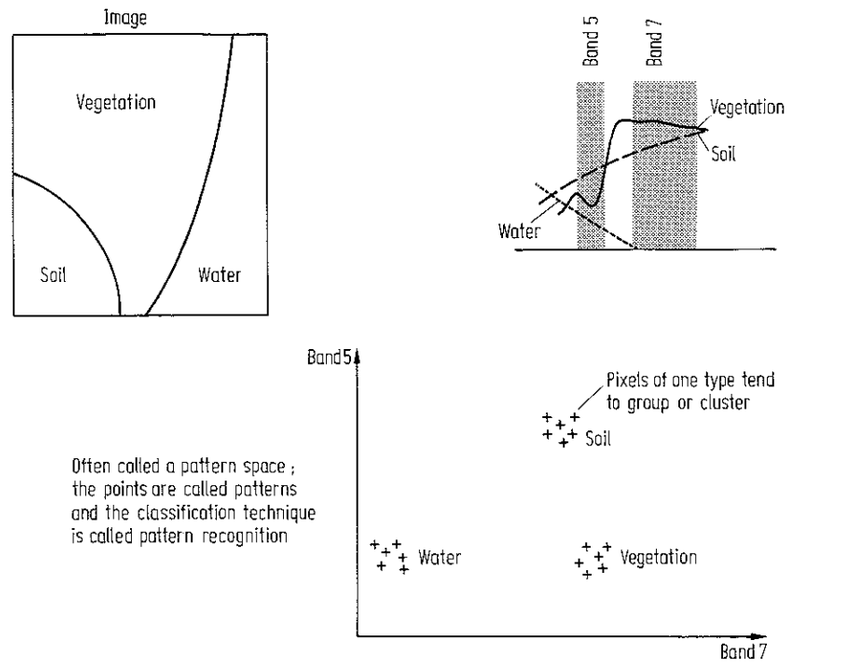


Fig. 3.5. Illustration of a two dimensional multispectral space showing its relation to the spectral reflectance characteristics of ground cover types

classes as depicted in Fig 3.6. It is necessary in situations such as this to determine appropriate sets of spectral classes that represent the information classes effectively. This is demonstrated in the exercises chosen in Chap 11

In quantitative analysis it is the spectral classes that a computer will be asked to work with since they are the "natural" groupings or clusters in the data After quantitative analysis is complete the analyst simply associates all the relevant spectral classes with

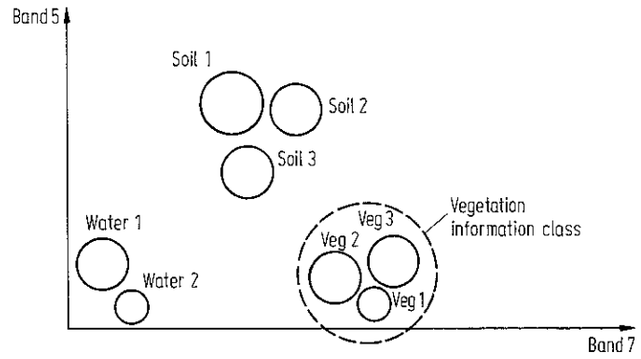


Fig. 3.6. Representation of information classes by sets of spectral classes

the one appropriate information class. Shortly, spectral classes will be seen to be unimodal probability distributions and information classes as possible multimodal distributions. The latter need to be resolved into sets of single modes for convenience in analysis.

3.6 Quantitative Analysis by Pattern Recognition

3.6.1 Pixel Vectors and Labelling

Recognition that image data exists in sets of spectral classes, and identification of those classes as corresponding to specific ground cover types, is carried out using the techniques of mathematical pattern recognition or pattern classification. The patterns are the pixels themselves, or strictly the mathematical *pixel vectors* that contain the sets of brightness values for the pixels arranged in column form:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

where x_1 to x_n are the brightnesses of the pixel x in bands 1 to n respectively. It is simply a mathematical convention that these are arranged in a column and enclosed in an extended square bracket. A summary of essential results from the algebra used for describing and manipulating these vectors is given in Appendix C.

Classification involves labelling the pixels as belonging to particular spectral (and thus information) classes using the spectral data available. This is depicted as a mapping in Fig. 3.7. In the terminology of statistics this is more properly referred to as allocation rather than classification. However throughout this book, classification, categorization, allocation and labelling are generally used synonomously.

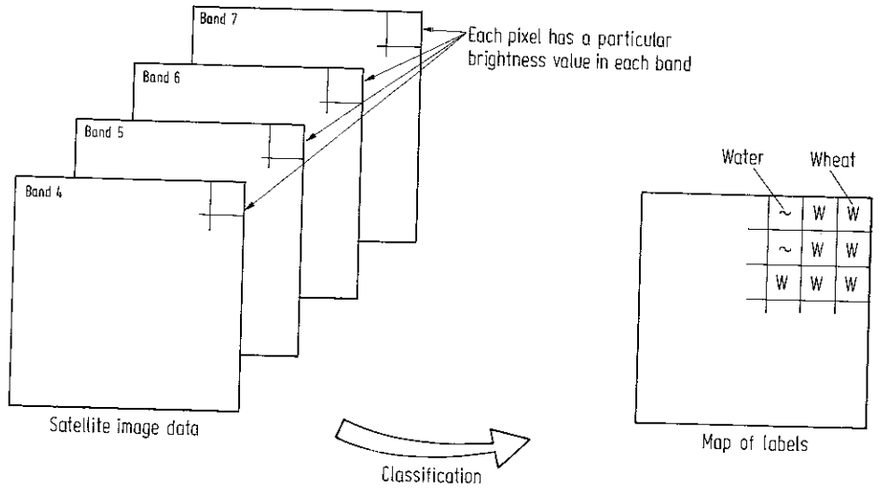


Fig. 3.7. The role of classification in labelling pixels in remote sensing image data. Here an illustration from Landsat multispectral scanner data is depicted

There are two broad classes of classification procedure and each finds application in the analysis of remote sensing image data. One is referred to as supervised classification and the other unsupervised classification. These can be used as alternative approaches but are often combined into hybrid methodologies as demonstrated in Chap. 11.

3.6.2 Unsupervised Classification

Unsupervised classification is a means by which pixels in an image are assigned to spectral classes without the user having foreknowledge of the existence or names of those classes. It is performed most often using clustering methods. These procedures can be used to determine the number and location of the spectral classes into which the data falls and to determine the spectral class of each pixel. The analyst then identifies those classes *a posteriori*, by associating a sample of pixels in each class with available data, which could include maps and information from ground visits. Clustering procedures are generally computationally expensive yet they are central to the analysis of remote sensing imagery. Whilst the information classes for a particular exercise are known, the analyst is usually totally unaware of the spectral classes, or sub-classes as they are sometimes called. Unsupervised classification is therefore useful for determining the spectral class composition of the data prior to detailed analysis by the methods of supervised classification.

The range of clustering algorithms frequently used for determination of spectral classes and for unsupervised classification is treated in Chap. 9.

3.6.3 Supervised Classification

Supervised classification procedures are the essential analytical tools used for the extraction of quantitative information from remotely sensed image data. There are a number of specific techniques and these are treated in detail in Chap. 8. It is the role of this section to introduce the framework of the approach.

An important assumption in supervised classification usually adopted in remote sensing is that each spectral class can be described by a probability distribution in multispectral space: this will be a multivariable distribution with as many variables as dimensions of the space. Such a distribution describes the chance of finding a pixel belonging to that class at any given location in multispectral space. This is not unreasonable since it would be imagined that most pixels in a distinct cluster or spectral class would lie towards the centre and would decrease in density for positions away from the class centre, thereby resembling a probability distribution. The distribution found to be of most value is the normal or Gaussian distribution. It gives rise to tractable mathematical descriptions of the supervised classification process, and is robust in the sense that classification accuracy is not overly sensitive to violations of the assumptions that the classes are normal. A two dimensional multispectral space with the spectral classes so modelled is depicted in Fig. 3.8. The *decision boundaries* shown in the figure represent those points in multispectral space where a pixel has equal chance

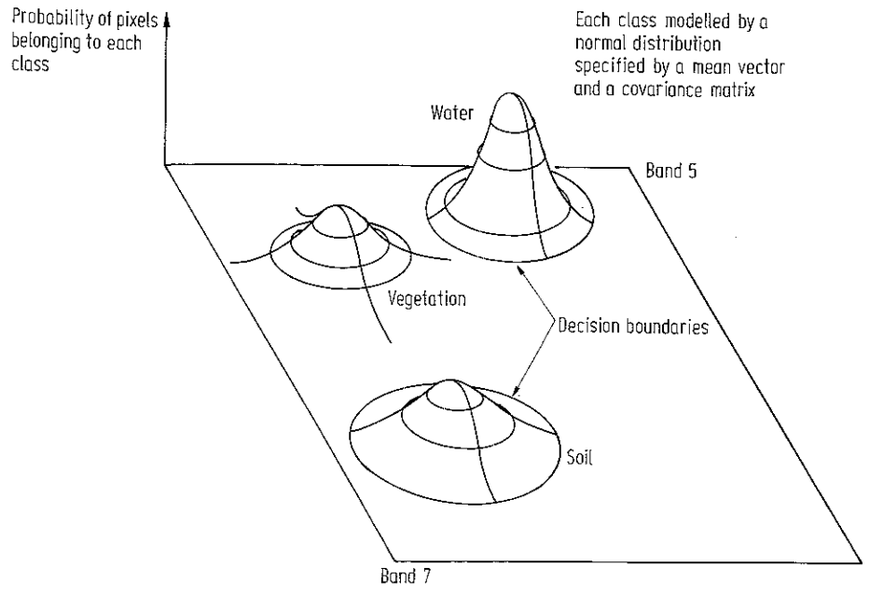


Fig. 3.8. Two dimensional multispectral space with the spectral classes represented by Gaussian probability distributions Here the space is defined in terms of the Landsat multispectral scanner bands 5 and 7

of belonging to two classes The boundaries therefore partition the space into regions associated with each class; this is developed further in Sect 8.2 4

A multidimensional normal distribution is described as a function of a vector location in multispectral space by:

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}\,|\Sigma|^{1/2}} \exp\left\{-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^t\,\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{m})\right\}$$

where $\boldsymbol{x}$ is a vector location in the N dimensional pixel space: $\boldsymbol{m}$ is the mean position of the spectral class – i e. the position $\boldsymbol{x}$ at which a pixel from the class is most likely to be found, and Σ is the covariance matrix of the distribution, which describes its spread directionally in the pixel space Equation (6.2) shows how this matrix is defined; Appendix D summarises some of the important properties of this distribution

The multidimensional normal distribution is specified completely by its mean vector and its covariance matrix Consequently, if the mean vectors and covariance matrices are known for each spectral class then it is possible to compute the set of probabilities that describe the relative likelihoods of a pattern at a particular location belonging to each of those classes. It can then be considered as belonging to the class which indicates the highest probability Therefore if $\boldsymbol{m}$ and Σ are known for every spectral class in an image, every pixel in the image can be examined and labelled corresponding to the most likely class on the basis of the probabilities computed for the particular location for a pixel. Before that classification can be performed however $\boldsymbol{m}$ and Σ are estimated for

each class from a representative set of pixels, commonly called *a training set.* These are pixels which the analyst knows as coming from a particular (spectral) class. Estimation of $\boldsymbol{m}$ and Σ from training sets is referred to as supervised learning Supervised classification consists therefore of three broad steps. First a set of training pixels is selected for each spectral class This may be done using information from ground surveys, aerial photography, topographic maps or any other source of reference data. The second step is to determine $\boldsymbol{m}$ and Σ for each class from the training data This completes the learning phase The third step is the classification phase, in which the relative likelihoods for each pixel in the image are computed and the pixel labelled according to the highest likelihood.

The view of supervised classification adopted here has been based upon an assumption that the classes can be modelled by probability distributions and, as a consequence, are described by the parameters of those distributions Other supervised techniques also exist, in which neither distribution models nor parameters are relevant. These are referred to as non-parametric methods and are discussed extensively in Nilsson (1965) While they are sometimes used in remote sensing they are encountered considerably less frequently than the approach outlined above

References for Chapter 3

An overview of the essential issues in image interpretation, including the role of computer processing for improving the impact of imagery and for quantitative analysis, will be found in Estes et al (1983) A good summary, with extensive references, of the spectral reflectance characteristics of common earth surface cover types has been given by Hoffer (1978). Material of this type is important in photointerpretation. Landgrebe (1981) and Hoffer (1979) have provided good general discussions on computer classification of remote sensing image data

J E Estes, E J. Hajic & L R Tinney, 1983: Fundamentals of Image Analysis: Analysis of Visible and Thermal Infrared Data in D S Simonett (ed.), Manual of Remote Sensing, Vol. 1, 2e, American Society of Photogrammetry, Falls Church, Virginia

R. M Hoffer, 1978: Biological and Physical Considerations in Applying Computer-Aided Analysis Techniques to Remote Sensing Data, in P H. Swain & S M. Davis, Eds.: Remote Sensing: The Quantitative Approach, McGraw-Hill, N Y.

R M Hoffer, 1979: Computer Aided Analysis Techniques for Mapping Earth Surface Features, Technical Report 020179, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.

D A Landgrebe, 1981: Analysis Technology for Land Remote Sensing, Proc IEEE, 69, 628–642

N J Nilsson, 1965: Learning Machines, McGraw-Hill, N Y

Problems

7.1 Compute the discrete Fourier transform of the square wave shown in Fig. 7.3 using $K = 2$, 4 and 8 samples per period of the waveform respectively. You can use the flow chart of Fig. 7.10 to help in this.
7.2 Compute the discrete Fourier transform of the unit pulse shown in Fig. 7.4. Use respectively $K = 2$, 4 and 8 samples over a time interval equal to 8a, where 2a is the width of the pulse as shown in the Figure. Compare the results with those obtained in problem 7.1.
7.3 (a) A common technique for smoothing an image is to compute averages over square or rectangular windows as discussed in Sect. 5.5. Consider a 3×1 smoothing template used to smooth a single line of image data in the manner of Fig. 5.4. Determine the corresponding filter function in the spatial frequency domain by finding the discrete Fourier transform of the template. You may find the material of Fig. 7.4 to be of value.
(b) Imagine an ideal low pass filter function in the spatial frequency domain that could be used to smooth just the lines of an image. Determine the corresponding function in the image domain by computing the inverse Fourier transform of the ideal filter. Taking into account the discrete pixel nature of the image, approximate the inverse transform by an appropriate one dimensional template.
7.4 Verify the results in Sect. 7.5.2 graphically.
7.5 (a) The periodic sequence of impulses of (7.11) is an idealised sampling function. In practice it is not possible to take infinitessimally short samples of a function; rather the samples will have a finite, albeit small duration. This could be modelled mathematically by replacing $\Delta(t)$ in (7.12) by a periodic pulse waveform. This periodic sequence of pulses can be represented by the convolution of a single pulse with the periodic sequence of impulses in (7.11). With this in mind describe what modifications are needed to Fig. 7.6 to account for samples of finite duration.
(b) Suppose the total period of sampling is equivalent to ten sample intervals. Describe the effect this has on Fig. 7.6.
7.6 In Fig. 7.6a suppose the function $f(t)$ is a sinewave of frequency B Hz. Its frequency spectrum will consist of two impulses, one at $+B$ Hz and the other at $-B$ Hz. Produce the spectrum of the sampled sinusoid if only three samples are taken every two periods. Suppose the waveform is then reconstructed by feeding the samples through a low pass filter that will pass all frequency components unattenuated, up to $1/2\,T$ Hz, where T is the sampling interval, and will exclude all components with frequencies in excess of $1/2\,T$ Hz. Describe the shape of the reconstructed signal; this will give an appreciation of aliasing distortion.

Chapter 8

Supervised Classification Techniques

8.1 Steps in Supervised Classification

Supervised classification is the procedure most often used for quantitative analysis of remote sensing image data. It rests upon using suitable algorithms to label the pixels in an image as representing particular ground cover types, or classes. A variety of algorithms is available for this, ranging from those based upon probability distribution models for the classes of interest (such as outlined in Chap. 3) to those in which the multispectral space is partitioned into class-specific regions using optimally located surfaces. Irrespective of the particular method chosen, the essential practical steps are:

1. Decide the set of ground cover types into which the image is to be segmented. These are the information classes and could, for example, be water, urban regions, croplands, rangelands, etc.
2. Choose representative or prototype pixels from each of the desired set of classes. These pixels are said to form training data. Training sets for each class can be established using site visits, maps, air photographs or even photointerpretation of a colour composite product formed from the image data (either in hardcopy form or on the colour display of an image analysis system). Often the training pixels for a given class will lie in a common region enclosed in a border. That region is then often called a training field.
3. Use the training data to estimate the parameters of the particular classifier algorithm to be used; these parameters will be the properties of the probability model used or will be equations that define partitions in the multispectral space. The set of parameters for a given class is sometimes called the *signature* of that class.
4. Using the trained classifier, label or classify every pixel in the image into one of the desired ground cover types (information classes). Here the whole image segment of interest is typically classified. Whereas training in Step 2 may have required the user to identify perhaps 1% of the image pixels by other means, the computer will label the rest by classification.
5. Produce tabular summaries or thematic (class) maps which summarise the results of the classification.

It is the role of this chapter to present the range of algorithms that could be used in 3 and 4. In so doing it will be assumed that the information classes each consists of only one spectral class, so that the two names will be used synonomously. (See Chap. 3 for a discussion of the two class types.) By making this assumption, problems with establishing sub-classes will not distract from the algorithm development to be given. Handling sub-classes is taken care of explicitly in Chaps. 9 and 11.

In the following sections it is assumed that the reader is familiar at least with the sections on quantitative analysis contained in Chap 3. This relates particularly to definitions and terminology.

8.2 Maximum Likelihood Classification

Maximum likelihood classification is the most common supervised classification method used with remote sensing image data. This is developed in the following in a statistically acceptable manner; it can be derived however in a more general and rigorous manner and this is presented for completeness in Appendix E. The present approach is sufficient though for most remote sensing exercises.

8.2.1 Baye's Classification

Let the spectral classes for an image be represented by

$$\omega_i,\ i=1,\ \ldots\ M$$

where M is the total number of classes. In trying to determine the class or category to which a pixel at a location x belongs it is strictly the conditional probabilities

$$p(\omega_i|\boldsymbol{x}),\ i=1,\ \ldots\ M$$

that are of interest. The position vector $\boldsymbol{x}$ is a column vector of brightness values for the pixel. It describes the pixel as a point in multispectral space with co-ordinates defined by the brightnesses, as shown in the simple two-dimensional example of Fig. 3.5. The probability $p(\omega_i|\boldsymbol{x})$ gives the likelihood that the correct class is ω_i for a pixel at position $\boldsymbol{x}$. Classification is performed according to

$$\boldsymbol{x}\in\omega_i \quad \text{if} \quad p(\omega_i|\boldsymbol{x})>p(\omega_j|\boldsymbol{x}) \quad \text{for all} \quad j\neq i \tag{8.1}$$

i.e., the pixel at $\boldsymbol{x}$ belongs to class ω_i if $p(\omega_i|\boldsymbol{x})$ is the largest. This intuitive *decision rule* is a special case of a more general rule in which the decisions can be biased according to different degrees of significance being attached to different incorrect classifications. The general approach is called Baye's classification and is the subject of the treatment in Appendix E.

8.2.2 The Maximum Likelihood Decision Rule

Despite its simplicity, the $p(\omega_i|\boldsymbol{x})$ in (8.1) are unknown. Suppose however that sufficient training data is available for each ground cover type. This can be used to estimate a probability distribution for a cover type that describes the chance of finding a pixel from class ω_i, say, at the position $\boldsymbol{x}$. Later the form of this distribution function will be made more specific. For the moment however it will be retained in general terms and represented by the symbol $p(\boldsymbol{x}|\omega_i)$. There will be as many $p(\boldsymbol{x}|\omega_i)$ as there are ground cover classes. In other words, for a pixel at a position $\boldsymbol{x}$ in multispectral space a set of

probabilities can be computed that give the relative likelihoods that the pixel belongs to each available class.

The desired $p(\omega_i|\boldsymbol{x})$ in (8.1) and the available $p(\boldsymbol{x}|\omega_i)$ – estimated from training data – are related by Baye's theorem (Freund, 1962):

$$p(\omega_i|\boldsymbol{x})=p(\boldsymbol{x}|\omega_i)\,p(\omega_i)/p(\boldsymbol{x}) \tag{8.2}$$

where $p(\omega_i)$ is the probability that class ω_i occurs in the image. If, for example, 15% of the pixels of an image happen to belong to spectral class ω_i then $p(\omega_i)=0.15$; $p(\boldsymbol{x})$ in (8.2) is the probability of finding a pixel from *any* class at location $\boldsymbol{x}$. It is of interest to note in passing that

$$p(\boldsymbol{x})=\sum_{i=1}^{M} p(\boldsymbol{x}|\omega_i)\,p(\omega_i),$$

although $p(\boldsymbol{x})$ itself is not important in the following. The $p(\omega_i)$ are called *a priori* or prior probabilities, since they are the probabilities with which class membership of a pixel could be guessed before classification. By comparison the $p(\omega_i|\boldsymbol{x})$ are posterior probabilities. Using (8.2) it can be seen that the classification rule of (8.1) is:

$$\boldsymbol{x}\in\omega_i \quad \text{if} \quad p(\boldsymbol{x}|\omega_i)\,p(\omega_i)>p(\boldsymbol{x}|\omega_j)\,p(\omega_j) \quad \text{for all} \quad j\neq i \tag{8.3}$$

where $p(\boldsymbol{x})$ has been removed as a common factor. The rule of (8.3) is more acceptable than that of (8.1) since the $p(\boldsymbol{x}|\omega_i)$ are known from training data, and it is conceivable that the $p(\omega_i)$ are also known or can be estimated from the analyst's knowledge of the image. Mathematical convenience results if in (8.3) the definition

$$\begin{aligned} g_i(\boldsymbol{x}) &= \ln\{p(\boldsymbol{x}|\omega_i)\,p(\omega_i)\} \\ &= \ln p(\boldsymbol{x}|\omega_i)+\ln p(\omega_i) \end{aligned} \tag{8.4}$$

is used, where ln is the natural logarithm, so that (8.3) is restated as

$$\boldsymbol{x}\in\omega_i \quad \text{if} \quad g_i(\boldsymbol{x})>g_j(\boldsymbol{x}) \quad \text{for all} \quad j\neq i \tag{8.5}$$

This is, with one modification to follow, the decision rule used in maximum likelihood classification; the $g_i(\boldsymbol{x})$ are referred to as *discriminant functions*.

8.2.3 Multivariate Normal Class Models

At this stage it is assumed that the probability distributions for each class are of the form of multivariate normal models. This is an assumption, rather than a demonstrable property of natural spectral or information classes; however it leads to mathematical simplifications in the following. Moreover it is one distribution for which properties of the multivariate form are well-known.

In (8.4) therefore, it is now assumed for N bands that (see Appendix D)

$$p(\boldsymbol{x}|\omega_i)=(2\pi)^{-N/2}\,|\Sigma_i|^{-1/2}\exp\{-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_i)^t\,\Sigma_i^{-1}(\boldsymbol{x}-\boldsymbol{m}_i)\} \tag{8.6}$$

where $\boldsymbol{m}_i$ and Σ_i are the mean vector and covariance matrix of the data in class ω_i. The resulting term $-N/2 \ln (2\pi)$ is common to all $g_i(\boldsymbol{x})$ and does not aid discrimination. Consequently it is ignored and the final form of the discriminant function for maximum likelihood classification, based upon the assumption of normal statistics, is:

$$g_i(\boldsymbol{x}) = \ln p(\omega_i) - \tfrac{1}{2} \ln |\Sigma_i| - \tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{m}_i)^t \Sigma_i^{-1} (\boldsymbol{x} - \boldsymbol{m}_i) \tag{8.7}$$

Often the analyst has no useful information about the $p(\omega_i)$, in which case a situation of equal prior probabilities is assumed; as a result $\ln p(\omega_i)$ can be removed from (8.7) since it is the same for all i. In that case the $1/2$ common factor can also be removed leaving, as the discriminant function:

$$g_i(\boldsymbol{x}) = -\ln |\Sigma_i| - (\boldsymbol{x} - \boldsymbol{m}_i)^t \Sigma_i^{-1} (\boldsymbol{x} - \boldsymbol{m}_i) \tag{8.8}$$

Implementation of the maximum likelihood decision rule involves using either (8.7) or (8.8) in (8.5). There is a further consideration however concerned with whether any of the available labels or classes is appropriate. This relates to the use of thresholds as discussed in Sect. 8.2.5 following.

8.2.4 Decision Surfaces

As a means for assessing the capabilities of the maximum likelihood decision rule it is of value to determine the essential shapes of the surfaces that separate one class from another in the multispectral domain. These surfaces, albeit implicit, can be devised in the following manner.

Spectral classes are defined by those regions in multispectral space where their discriminant functions are the largest. Clearly these regions are separated by surfaces where the discriminant functions for adjoining spectral classes are equal. The ith and jth spectral classes are separated therefore by the surface

$$g_i(\boldsymbol{x}) - g_j(\boldsymbol{x}) = 0.$$

This is referred to as a *decision surface* since, if all the surfaces separating spectral classes are known, decisions about class membership of an image pixel can be made on the basis of its position relative to the complete set of surfaces.

The construction $(\boldsymbol{x} - \boldsymbol{m}_i)^t \Sigma_i^{-1} (\boldsymbol{x} - \boldsymbol{m}_i)$ in (8.7) and (8.8) is a quadratic function of $\boldsymbol{x}$. Consequently the decision surfaces implemented by maximum likelihood classification are quadratic and thus take the form of parabolas, circles and ellipses. Some indication of this can be seen in Fig. 3.8.

8.2.5 Thresholds

It is implicit in the foregoing development that pixels at every point in multispectral space will be classified into one of the available classes ω_i, irrespective of how small the actual probabilities of class membership are. This is illustrated for one dimensional data in Fig. 8.1a. Poor classification can result as indicated. Such situations can arise if spectral classes (between 1 and 2 or beyond 3) have been overlooked or, if knowing

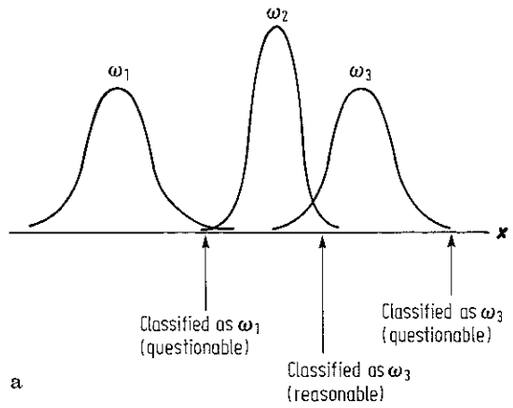


Fig. 8.1. **a** Illustration of poor classification for patterns lying near the tails of the distribution functions of all spectral classes; **b** Use of a threshold to remove poor classification

other classes existed, enough training data was not available to estimate the parameters of their distributions with any degree of accuracy (see Sect. 8.2.6 following). In situations such as these it is sensible to apply thresholds to the decision process in the manner depicted in Fig. 8.1b. Patterns which have probabilities for *all* classes below the threshold are not classified.

In practice, thresholds are applied to the discriminant functions and not the probability distributions, since the latter are never actually computed. With the incorporation of a threshold therefore, the decision rule of (8.5) becomes

$$\boldsymbol{x} \in \omega_i \quad \text{if} \quad g_i(\boldsymbol{x}) > g_j(\boldsymbol{x}) \quad \text{for all} \quad j \neq i \tag{8.9a}$$

and

$$g_i(\boldsymbol{x}) > T_i \tag{8.9b}$$

where T_i is the threshold seen to be significant for spectral class ω_i. It is now necessary to consider how T_i can be estimated. From (8.7) and (8.9b) a classification is acceptable if

$$\ln p(\omega_i) - \tfrac{1}{2}\ln|\Sigma_i| - \tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_i)^t \Sigma_i^{-1}(\boldsymbol{x}-\boldsymbol{m}_i) > T_i$$

i.e.

$$(\boldsymbol{x}-\boldsymbol{m}_i)^t \Sigma_i^{-1}(\boldsymbol{x}-\boldsymbol{m}_i) < -2T_i - \ln|\Sigma_i| + 2\ln p(\omega_i) \tag{8.10}$$

The left hand side of (8.10) has a χ^2 distribution with N degrees of freedom, if $\boldsymbol{x}$ is (assumed to be) distributed normally (Swain and Davis 1978). N is the dimensionality of the multispectral space. As a result χ^2 tables can be consulted to determine that value of $(\boldsymbol{x}-\boldsymbol{m}_i)^t \Sigma_i^{-1}(\boldsymbol{x}-\boldsymbol{m}_i)$ below which a desired percentage of pixels will exist (noting that larger values of that quadratic form correspond to pixels lying further out in the tails of the normal probability distribution). This is depicted in Fig. 8.2.

As an example of how this is used consider the need to choose a threshold for Landsat multispectral scanner data such that 95% of all pixels in a class will be classified (i.e. such that the 5% least likely pixels for each spectral class will be rejected). χ^2 tables show that 95% of all pixels have χ^2 values (in Fig. 8.2) less than 9.488. Thus, from (8.10)

$$T_i = -4.744 - \frac{1}{2}\ln|\Sigma_i| + \ln p(\omega_i)$$

which thus can be calculated from a knowledge of the prior probability and covariance matrix of the ith spectral class.

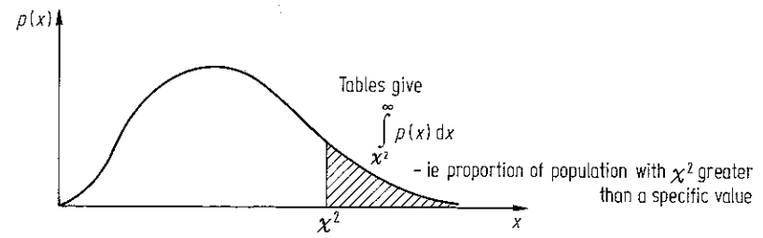


Fig. 8.2 Use of the χ^2 distribution for obtaining classifier thresholds

8.2.6 Number of Training Pixels Required for Each Class

Sufficient training samples for each spectral class must be available to allow reasonable estimates of the elements of the mean vector and the covariance matrix to be determined. For an N dimensional multispectral space at least $N+1$ samples are required to avoid the covariance matrix being singular. Should that happen its inverse in discriminant function expressions cannot be found. Apart from this consideration it is clearly important to have as many training pixels as possible, particularly as the dimensionality of the pixel vector space increases, since in higher dimensional spaces there is an increased chance of having some individual dimensions poorly represented. Swain and Davis (1978) recommend as a practical minimum that $10N$ samples per spectral class be obtained for training, with $100N$ as being highly desirable if it can be attained.

8.2.7 A Simple Illustration

As an example of the use of maximum likelihood classification, the segment of Landsat multispectral scanner image shown in Fig. 8.3 is chosen. This is a 256×276 pixel array of image data in which four broad ground cover types are evident. These are water, fire burn, vegetation and "developed" land. Suppose one wished to produce a thematic map of these four cover types in order to enable the area and extent of the fire burn to be evaluated.

The first step is to choose training data. For such a broad classification, suitable sets of training pixels for each of the four classes are easily identified visually in the image.

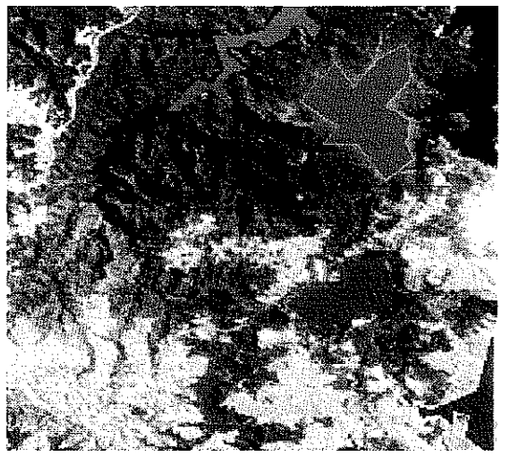


Fig. 8.3. Image segment to be classified, consisting of a mixture of natural vegetation, waterways, urban development and vegetation damaged by fire. Four training regions are identified in solid colour. These are water (violet), vegetation (green), fire burn (red) and urban (dark blue). Pixels from these were used to generate the signatures in Table 8.1

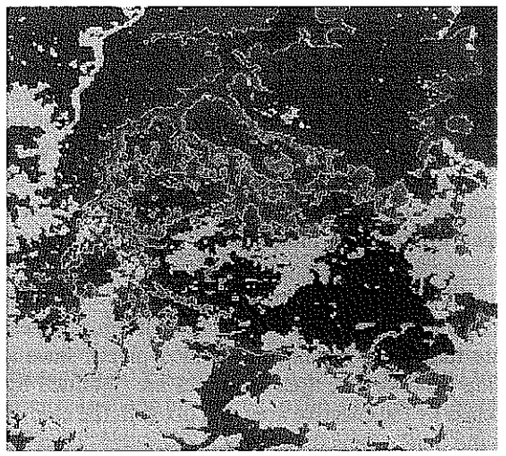


Fig. 8.4. Thematic map produced by maximum likelihood classification. Blue represents water, red is fire damaged vegetation, green is natural vegetation and yellow is urban development

Table 8.1. Class signatures for the training areas in Fig. 8.3. Numbers are on a scale of 0 to 255 (8 bit)

Class	Mean vector	Covariance matrix			
Water	44.27	14.36	9.55	4.49	1.19
	28.82	9.55	10.51	3.71	1.11
	22.77	4.49	3.71	6.95	4.05
	13.89	1.19	1.11	4.05	7.65
Fire burn	42.85	9.38	10.51	12.30	11.00
	35.02	10.51	20.29	22.10	20.62
	35.96	12.30	22.10	32.68	27.78
	29.04	11.00	20.62	27.78	30.23
Vegetation	40.46	5.56	3.91	2.04	1.43
	30.92	3.91	7.46	1.96	0.56
	57.50	2.04	1.96	19.75	19.71
	57.68	1.43	0.56	19.71	29.27
Developed (urban)	63.14	43.58	46.42	7.99	−14.86
	60.44	46.42	60.57	17.38	−9.09
	81.84	7.99	17.38	67.41	67.57
	72.25	−14.86	−9.09	67.57	94.27

Table 8.2. Tabular summary of the thematic map of Fig. 8.4

Class	No. of pixels	Area (ha)
Water	4830	2137
Fireburn	14182	6274
Vegetation	28853	12765
Developed (urban)	22791	10083

data. Fig. 8.3 also shows the locations of four training fields used for this purpose. Sometimes, to obtain a good estimate of class statistics it may be necessary to choose several training fields for the one cover type, located in different regions of the image.

The four band signatures for each of the four classes, as obtained from the training fields are given in Table 8.1. The mean vectors can be seen to agree generally with known spectral reflectance characteristics of the cover types. Also the class variances (diagonal elements in the covariance matrices) are small for water as might be expected but on the large side for the developed/urban class, indicative of its heterogeneous nature.

Using these signatures in a maximum likelihood algorithm, to classify the four bands of the image in Fig. 8.3, the thematic map shown in Fig. 8.4 is obtained. The four classes, by area, are given in Table 8.2. Note that there are no unclassified pixels, since a threshold was not used in the labelling process. The area estimates are obtained by multiplying the number of pixels per class by the effective area of a pixel. In the case of the Landsat 2 multispectral scanner the pixel size is 0.4424 hectares.

8.3 Minimum Distance Classification

8.3.1 The Case of Limited Training Data

The effectiveness of maximum likelihood classification depends upon reasonably accurate estimation of the mean vector $\boldsymbol{m}$ and the covariance matrix Σ for each spectral class. This in turn is dependent upon having a sufficient number of training pixels for each of those classes. In cases where this is not so, inaccurate estimates of the elements of Σ result, leading to poor classification. When the number of training samples per class is limited it can be more effective to resort to a classifier that does not make use of covariance information but instead depends only upon the mean positions of the spectral classes, noting that for a given number of samples these can be more accurately estimated than covariances. The so-called minimum distance classifier, or more precisely, minimum distance to class means classifier, is such an approach. With this classifier, training data is used only to determine class means; classification is then performed by placing a pixel in the class of the nearest mean.

The minimum distance algorithm is also attractive since it is a faster technique than maximum likelihood classification, as will be seen in Sect. 8.5. However because it does not use covariance data it is not as flexible as the latter. In maximum likelihood classification each class is modelled by a multivariate normal class model that can account for spreads of data in particular spectral directions. Since covariance data is not used in the minimum distance technique class models are symmetric in the spectral domain. Elongated classes therefore will not be well modelled. Instead several spectral classes may need to be used with this algorithm where one might be suitable for maximum likelihood classification. This point is developed further in the case studies of Chap. 11.

8.3.2 The Discriminant Function

The discriminant function for the minimum distance classifier is developed as follows:

Suppose $\boldsymbol{m}_i, i=1, \quad M$ are the means of the M classes determined from training data, and $\boldsymbol{x}$ is the position of the pixel to be classified. Compute the set of squared Euclidean distances, defined in vector form as

$$\begin{aligned} \mathrm{d}(\boldsymbol{x}, \boldsymbol{m}_i)^2 &= (\boldsymbol{x}-\boldsymbol{m}_i)^{\mathrm{t}}\,(\boldsymbol{x}-\boldsymbol{m}_i) \\ &= (\boldsymbol{x}-\boldsymbol{m}_i)\ \ (\boldsymbol{x}-\boldsymbol{m}_i)\ \ i=1, \quad M \end{aligned}$$

Expanding the product gives

$$\mathrm{d}(\boldsymbol{x}, \boldsymbol{m}_i)^2 = \boldsymbol{x}\ \ \boldsymbol{x} - 2\,\boldsymbol{m}_i\ \ \boldsymbol{x} + \boldsymbol{m}_i\ \ \boldsymbol{m}_i$$

Classification is performed on the basis of

$$\boldsymbol{x} \in \omega_i \quad \text{if} \quad \mathrm{d}(\boldsymbol{x}, \boldsymbol{m}_i)^2 < \mathrm{d}(\boldsymbol{x}, \boldsymbol{m}_j)^2 \quad \text{for all} \quad j \neq i$$

Note that $\boldsymbol{x}\ \ \boldsymbol{x}$ is common to all $\mathrm{d}(\boldsymbol{x}, \boldsymbol{m}_j)^2$ and thus can be removed. Moreover, rather than classifying according to the smallest of the remaining expressions, the signs can be reversed and classification performed on the basis of

Appendix E

Penalty Function Derivation of the Maximum Likelihood Decision Rule

E.1 Loss Functions and Conditional Average Loss

The derivation of maximum likelihood classification in Sect 8.2 is generally acceptable for remote sensing applications and is used widely However it is based implicity on the understanding that misclassifying any particular pixel is no more significant than misclassifying any other pixel in an image The more general approach presented in the following allows the user to specify the importance of making certain labelling errors compared with others. For example, for crop classification involving two sub-classes of wheat it would probably be less of a problem if a particular wheat pixel was erroneously classified into the other sub-class than it would if it were classified as water.

To develop the general method we introduce the penalty function, or loss function

$$\lambda(i|k)\ i, k = 1, \ldots M \tag{E-1}$$

This is a measure of the loss or penalty incurred when an algorithm erroneously labels a pixel as belonging to class ω_i when in reality the pixel is from class ω_k. It is reasonable to expect that $\lambda(i|i) = 0$ for all i: this implies there is no penalty for a correct classification In principle, there are M^2 distinct values of $\lambda(i|k)$.

The penalty incurred by erroneously labelling a pixel at position $\boldsymbol{x}$ in multispectral space into class ω_i is

$$\lambda(i|k)\ p(\omega_k|\boldsymbol{x})$$

where the pixel comes correctly from class ω_k and $p(\omega_k|\boldsymbol{x})$ is the posterior probability that ω_k is the correct class for pixels at $\boldsymbol{x}$ Averaging this over all possible ω_k we have the average loss, correctly referred to as the *conditional average loss*, associated with labelling a pixel as belonging to class ω_i. This is given by

$$L_{\boldsymbol{x}}(\omega_i) = \sum_{k=1}^{M} \lambda(i|k)\ p(\omega_k|\boldsymbol{x}) \tag{E-2}$$

and is a measure of the accumulated penalty incurred given the pixel could have belonged to any of the available classes and that we have available the penalty functions relating all the classes to class ω_i. Clearly, a useful decision rule for assigning a label to a pixel is to choose that class for which the conditional average loss is the smallest, viz

$$\boldsymbol{x} \in \omega_i \quad \text{if} \quad L_{\boldsymbol{x}}(\omega_i) < L_{\boldsymbol{x}}(\omega_j) \quad \text{for all} \quad j \neq i \tag{E-3}$$

An algorithm that implements (E-3) is often referred to as a Baye's optimum algorithm.

Even if the $\lambda(i|k)$ were known, the $p(\omega_k|\boldsymbol{x})$ usually are not. Therefore, as in Sect. (8.2.2) we adopt Baye's theorem which allows the posterior probabilities to be expressed in terms of the class probability distribution functions $p(\boldsymbol{x}|\omega_k)$; viz

$$p(\omega_k|\boldsymbol{x}) = p(\boldsymbol{x}|\omega_k)\ p(\omega_k)/p(\boldsymbol{x})$$

where $p(\omega_k)$ is the class prior probability. Using this in (E-2) gives

$$L_{\boldsymbol{x}}(\omega_i) = \frac{1}{p(\boldsymbol{x})}\ l_{\boldsymbol{x}}(\omega_i)$$

with

$$l_{\boldsymbol{x}}(\omega_i) = \sum_{k=1}^{M} \lambda(i|k)\ p(\boldsymbol{x}|\omega_k)\ p(\omega_k) \tag{E-4}$$

Since $p(\boldsymbol{x})$ is common to all classes it is sufficient to decide class membership on the basis of the $l_{\boldsymbol{x}}(\omega_i)$.

E.2 A Particular Loss Function

Suppose $\lambda(i|k) = 1 - \Phi_{ik}$ with $\Phi_{ii} = 1$ and $\Phi_{ik} (k \neq i)$ to be defined. Then (E-4) can be expressed

$$\begin{aligned} l_{\boldsymbol{x}}(\omega_i) &= \sum_{k=1}^{M} p(\boldsymbol{x}|\omega_k)\ p(\omega_k) - \sum_{k=1}^{M} \Phi_{ik} p(\boldsymbol{x}|\omega_k)\ p(\omega_k) \\ &= p(\boldsymbol{x}) - g_i(\boldsymbol{x}) \end{aligned}$$

with

$$g_i(\boldsymbol{x}) = \sum_{k=1}^{M} \Phi_{ik} p(\boldsymbol{x}|\omega_k)\ p(\omega_k) \tag{E-5}$$

Again since $p(\boldsymbol{x})$ is common to all classes it does not aid discrimination and thus can be removed from the conditional average loss expression, leaving just $l_{\boldsymbol{x}}(\omega_i) = -g_i(\boldsymbol{x})$. Because of the minus sign in this expression we can then decide the "least cost" labelling of a pattern at position $\boldsymbol{x}$ in multispectral space according to maximisation of the *discriminant function* $g_i(\boldsymbol{x})$, viz

$$\boldsymbol{x} \in \omega_i \quad \text{if} \quad g_i(\boldsymbol{x}) > g_j(\boldsymbol{x}) \quad \text{for all} \quad j \neq i \tag{E-6}$$

It is of interest at this stage to put

$$\Phi_{ik} = \delta_{ik}, \text{ the Kroneker delta function,}$$

defined by

$$\begin{aligned} \delta_{ik} &= 1 \quad \text{for} \quad i = k \\ &= 0 \quad \text{for} \quad i \neq k. \end{aligned}$$

Equation (E-5) then becomes

$$g_i(\boldsymbol{x}) = p(\boldsymbol{x}|\omega_i)\, p(\omega_i)$$

so that the decision rule in (E-6) is

$$\boldsymbol{x} \in \omega_i \quad \text{if} \quad p(\boldsymbol{x}|\omega_i)\, p(\omega_i) > p(\boldsymbol{x}|\omega_j)\, p(\omega_j) \quad \text{for all} \quad j \neq i$$

which is the classification rule adopted in (8.3) in Chap. 8. Frequently this is referred to as the unconditional maximum likelihood decision rule.

References for Appendix E

Nilsson (1965) gives an excellent account of the derivation of the maximum likelihood decision rule based upon penalty functions in the manner just derived. Other treatments that can be consulted include Duda and Hart (1973) and Andrews (1972).

H. C. Andrews, 1972: Introduction to Mathematical Techniques in Pattern Recognition, N. Y., Wiley.
R. O. Duda & P. E. Hart, 1973: Pattern Classification and Scene Analysis N. Y., Wiley.
N. J. Nilsson, 1965: Learning Machines, N. Y., McGraw-Hill.

FROM:
Jens Liebe, Storage Estimation of Water Storage Capacity and Evaporation Losses of Small Reservoirs in the Upper East Region of Ghana.
Diploma thesis, University of Bonn, September 2002

3.2 The Landsat 7 ETM+ System and Remote Sensing of Surface Water Bodies

3.2.1 Characteristics of Landsat 7 ETM+ - a brief overview

Landsat 7 was launched in April 1999 and is currently the latest platform of the Landsat program[12] that started in 1972. It carries the "Enhanced Thematic Mapper Plus" (ETM+) sensor, a passive, optical across-track-scanner, which records the reflected portion of the electromagnetic radiation, and the earth's emitted radiation in the thermal band, respectively. With an inclination to the equator of 98.2° and an equatorial crossing time at 10:00am +/- 15 min (descending node), Landsat 7 belongs to the near-polar, sun-synchronous orbiters. Orbiting at an altitude of 705km, it has a period of revolution of 99 minutes and repeatedly covers an area every 16 days (LANDSAT 7 SCIENCE DATA USERS HANDBOOK).

Table 2: Spectral and Spatial Resolution of the Landsat 7 ETM+ Sensor

Band	Sensitivity (µm)	Spatial Resolution (m)
1	0.45-0.52	30
2	0.52-0.60	30
3	0.63-0.69	30
4	0.76-0.90	30
5	1.55-1.75	30
6 (thermal)	10.4-12.5	60
7	2.08-2.35	30
8 (panchromatic)	0.50-0.90	15

(Adapted from: LILLESAND/KIEFER 2000, p. 379)

The spectral and spatial resolutions of Landsat 7 are denoted in Table 2. For the benefit of a continuous comparable dataset, the sensitivities – or bandwidths – of the bands one through seven abut on those of ETM+'s predecessor Thematic Mapper (TM), which was mounted on Landsat 4 and 5. While the resolution of the visible and infrared bands remained unchanged at 30 m, the resolution of the thermal band six was increased from 120 to 60 m. Additionally, a panchromatic eighth band with a resolution of 15 m was introduced to the ETM+ sensor.

[12] An extensive description of the Landsat program can be found in LILLESAND/KIEFER 2000, pp. 373-416.

The approximate dimension of a Landsat scene is 183 x 170 kilometers (EROS DATA CENTER). Indexed with path and row numbers according to the Worldwide Reference System (WRS), the scenes and thus the areas they cover are uniquely identifiable. The full coverage of the Upper East Region of Ghana requires four Landsat Scenes (Appendix 5).

The vast part is covered by the image strip consisting of the rows 052 and 053 in path 194, while, taking into account the overlap between neighboring paths, only a small fragment in the very western part is solely covered by the respective scenes in path 195. The scenes mainly used in this study were captured on the 7th November 1999 (194/052 and 194/053). An image captured on the 13th October 1999 (195/052), and a last one that was acquired on the 2nd February 2000 (195/053) laterally supplemented them. Their scene parameters are documented in Appendix 6.

3.2.2 Influence of date selection on classification results

In order to determine the maximum dimensions of reservoirs by means of remotely sensed data, which are represented by their largest area extent, the time of acquisition should be as close as possible to the end of the rainy season, when reservoirs are filled to their maximum capacity, and losses (draft, seepage, evaporation) are still negligible.

A look at Figure 2 showing the climatology of Navrongo reveals the end of the rainy season to be in September/October. Cloud coverage of up to 100% during the rainy season make optical remote sensing of ground features impossible. With the rains subsiding from October to November and a decline in cloud cover, the month of November turns out to meet the specified demands best. Another constraint for remote sensing, however, then arises through the emergence of the Harmattan, the dust laden NE wind, which causes considerable atmospheric scattering and reduces the contrast in the satellite images.

Having found the month of November to be most suitable with respect to seasonal variability, the year 1999 proves to have been a wet year (Figure 13, compare with Figure 2 and Appendix 7). It can thus be expected that the reservoirs absolute maximum storage volumes were reached in that year. Hence, the Landsat imagery from the 7th November 1999 is expected to optimally show the greatest extent of surface waters, both in number and area. With respect to the difficult atmospheric conditions, it has to be mentioned that the availability of useful satellite imagery is very limited. The good quality of the image strip from the 7th November that covers the greatest part of the Upper East Region is a fortunate incidence. The scenes from the 13th October 1999 and, especially, from the 2nd February 2000 are suboptimal, nonetheless they were used in that they only cover a small edge in the very west of the Upper East Region.

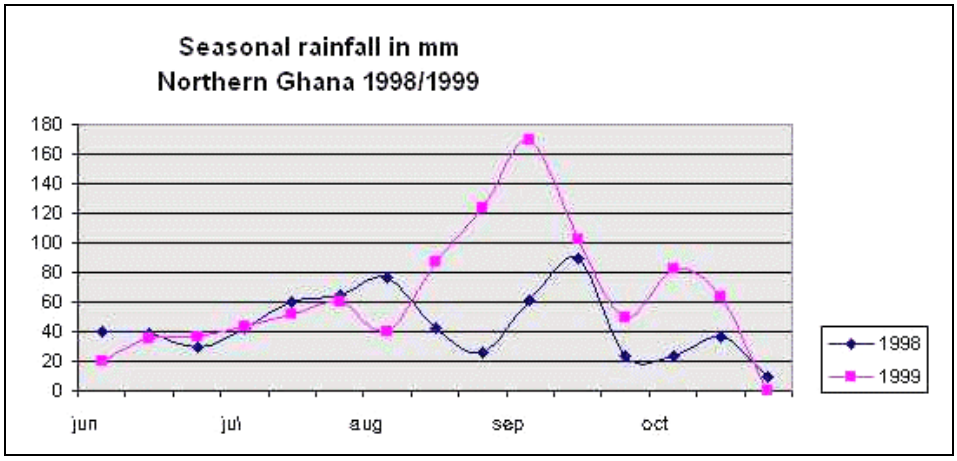


Figure 13: Seasonal Rainfall in Northern Ghana for 1998 and 1999 (in mm).
(Source: LUPO/REGINSTER 2000)

A second important date is the time of the field survey. This will be addressed together with the description of the fieldwork later in Chapter 5.1.1.

3.2.3 Water Body Detection with LANDSAT ETM+ and the reflectance properties of surface water bodies

The detection of water bodies is a concern that has been pursued since the first LANDSAT images became available in 1972. Since then, a large number of studies regarding the techniques of water detection and their quality have been published. The techniques range from visual interpretation by density slicing methods and band-ratio-approaches (NDWI[13]) to different supervised and unsupervised classification methods.

Using optical systems like LANDSAT, the detection and delineation of open surface water bodies is best done with imagery from the infrared and visible part of the spectrum. "The characteristic spectral reflectance curve for water shows a general reduction in reflectance with increasing wavelength, so that in the near-infrared the reflectance of deep, clear water is virtually zero" (MATHER 1999, p. 22). The reflectances of soils and, especially healthy vegetation are higher in these spectral bands, and therefore stand in distinct contrast to water bodies. The visible bands (VIS) enhance the contrast between water and soils (compare Figure 14).

[13] Normalized Difference Water Index (GAO 1996, McFEETERS 1996, VONDER/CLEVERS 1998)

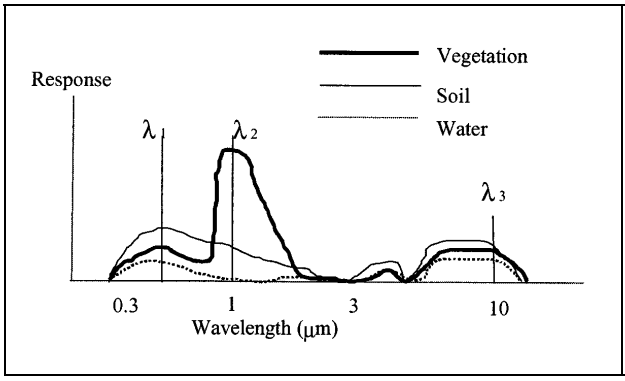


Figure 14: Reflectance vs. Wavelength for soil, vegetation and water.
(Source: KITE/PIETRONIRO 2000, p.218.)

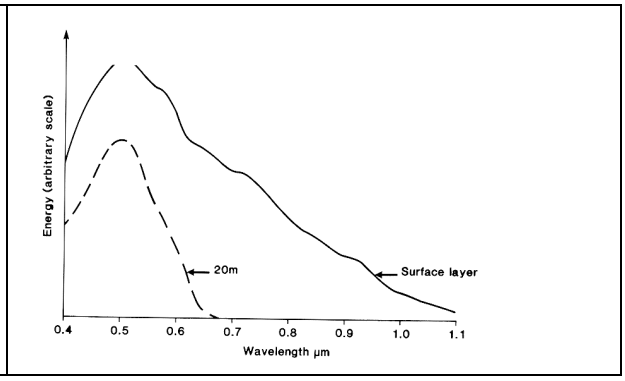


Figure 15: Spectral reflectance curve for surface water layer (solid line) and 20m (dashed).
(Source: MATHER 1999, p.24.)

This very general description of the surface water reflectance pattern, however, does not embrace the full spectrum of the diversity of surface waters as they appear in nature. Open "surface water"[14] in terms of a land cover type can have a wide range of different reflectance patterns and should hence be seen as an umbrella term for various sub-classes. The spectral reflectance patterns from surface waters are composed of three influencing parameters[15]: Surface reflection, volume reflectance and bottom reflectance [Figure 16], meaning that "it is affected by the presence and concentration of dissolved and suspended organic and inorganic material, and by the depth of the water body [Figure 15]" (MATHER 1999, p.22).

The *surface reflection* is influenced by the sun-sensor-constellation and the roughness of the water surface created by wind. Under calm conditions, the reflection might be specular, while surface waves and certain sun-to-sensor constellation cause sunglint (MATHER 1999, p. 22). "Because of strong absorption in the middle infrared, the magnitude of the scattered component is

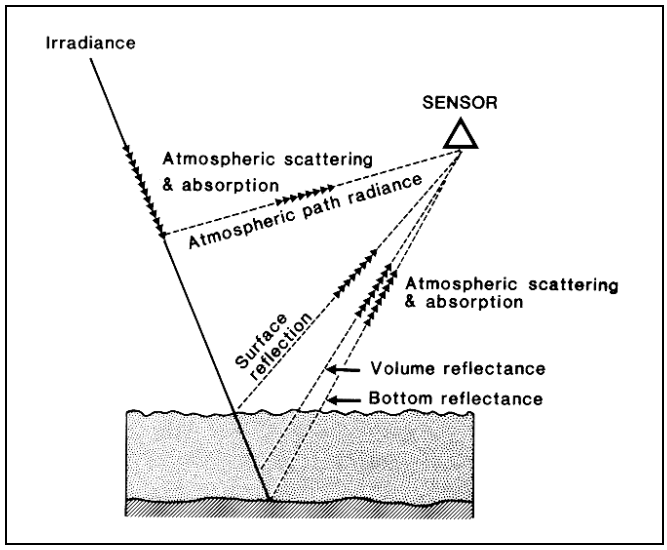


Figure 16: Processes acting on solar radiant energy in the visible part of the spectrum over an area of shallow water.
(Source: MATHER 1999, p. 24.)

[14] Refers to liquid water. Snow and ice add further reflectance patterns.
[15] An extensive description of influencing parameters on water return signals in optical remote sensing can be found in KONDRATYEV/FILATOV (1999, pp. 169-268).

chiefly derived from the shorter wavelengths, particularly those in the visible range" (MEIJERINK et al. 1994, p.64). Standing and floating vegetation alters the reflectance properties of a water body by adding a "red edge" (MATHER 1999, p. 20) to the return signal, the steep ascent in reflectance from the visible to the infrared part of the spectrum, which is characteristic for healthy vegetation (Figure 14).

The *volume reflection* influences the signal due to turbidity, dissolved matter, the trophic status and algae content of the reservoir water. Yet, the degree to which the volume reflection contributes to the total reflection signal of a water body depends on the penetration depth of light, which varies with the wavelengths (Figure 15). "The depth to which light can penetrate depends on the wavelength, about 10 m in the 0.5-0.6 µm waveband and less than 10 cm in the 0.8–1.1 µm range" (MEIJERINK et al. 1994, p. 64).

In shallow water bodies, *bottom reflection* occurs as a third influencing component. Similar to the volume reflection, the major influence of bottom reflection comes from the shorter wavelengths, as their penetration depth is greater than that of the higher wavelengths. Because bottom materials and the suspension freight often constitute of the same material, their influence on the reflectance can hardly be distinguished. "Therefore, it is difficult to use the penetration properties for the determination of water depths […]" (MEIJERINK et al. 1994, p.65).

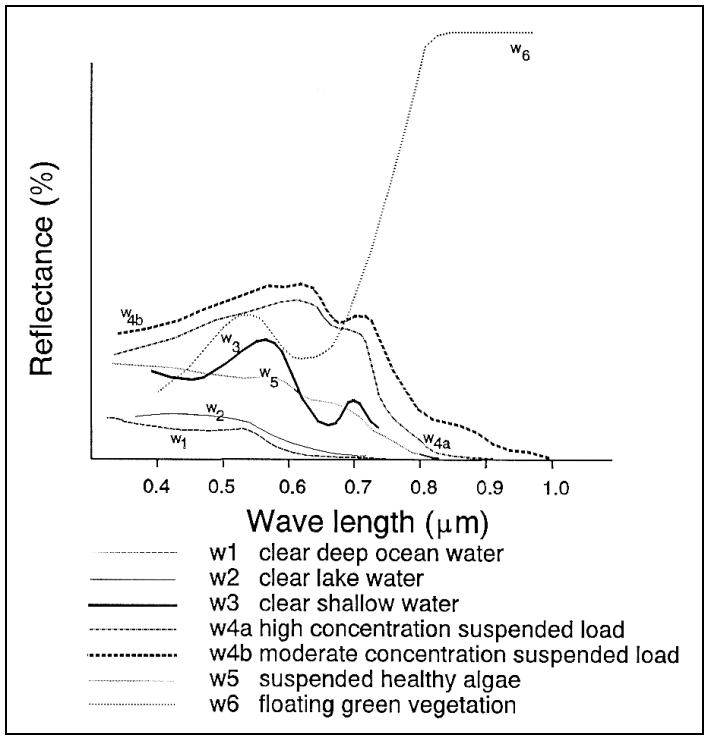


Figure 17: Possible reflectance curves of water bodies (relative reflectances).
(Source: MEIJERINK et al. 1994, p. 64.)

Resuming the influencing factors on the return signal from open water bodies, the following conclusions can be drawn:

1. *Surface water bodies can have a wide range of different reflectance patterns (Figure 17)*
2. *Water body delineation using RS data is best done with IR but also VIS bands*
3. *Bathymetry is difficult due to penetration properties of light and the often found uniformity of suspended and bottom materials*

3.2.4 Pre-Processing, classification and data extraction

The classification of the Landsat ETM+ scenes based on the above-mentioned principles and previous experience with the classification of water bodies in this area. The images 194-052 and 194-053 were geo-registered and, after pixel resizing to 30x30 m, merged into one image strip as they were recorded on the same orbit. Likewise, the images 195-052 and 195-053 were both geo-registered to each other and to the image strip 194, but not merged due to different surface water characteristics and extents, phonological states and atmospheric conditions.

Due to the lack of meteorological data on the constitution of the atmosphere at the times of acquisition, complex atmospheric corrections are hardly possible. Instead, a simple but effective atmospheric correction algorithm, the COST-Model (CHAVEZ 1996; 1989), was then applied to the scenes. It expands the dark object subtraction (DOS) method (JENSEN 1996, pp. 116-122; MEIJERINK et al. 1994, pp. 66-67; CHAVEZ 1996), which corrects the additive scattering effect and is used for the reduction of haze effects. The COST method additionally corrects for multiplicative transmittance effects of absorption by atmospheric gases and Rayleigh scattering (LILLESAND/KIEFER 1999, pp. 9f, MATHER 1999, pp. 16-17) by using the cosine of the solar zenith angle. "The corrections generated by the entirely image-based COST model are as accurate as those generated by the models that used in-situ atmospheric field measurements […]" (CHAVEZ 1996, p. 1025, accentuation changed).

After creating false color composites with the channel combination RGB 453, the classification of reservoirs was performed using a maximum likelihood classifier. Using the channel combination RGB 453, water bodies are depicted in colors ranging from bright blue to almost black. In order to recognize reservoirs on a satellite image, the human mind makes use of several visual characteristics, such as shape, contiguity and the association with the drainage network. Image processing software, such as IDRISI, operates differently. It recognizes reservoirs as it makes use of their spectral reflectance patterns. Therefore, the image classifier needs to be "trained" with spectral information, which comes from reservoirs that were identified by visual interpretation. Lead by the approach of thematic extraction of these surface water bodies, the classifier was trained with various water sub-classes to make use of all the facets of the 'land cover category' water. Large numbers of training sets were therefore defined for different water classes by on-screen digitizing based on visual differences in tone. These different shades are associated with the manifold compositions and combination of the influencing factors and by this means take into account different depths as well as different algae contents and turbidity.

A clear benefit of using several sub-classes instead of one water class is that of achieving spectrally more compact classes (Figure 18). Taking into account the functioning of the maximum likelihood classifier, the advantage of this procedure becomes clearer. Whether or not a pixel is assigned to one class or the other is herein based on equi-probability contours around the statistical averages of each class, which normally show elliptical shapes, whereby each class's statistical average is calculated from its spectral training set (Figure 19). Furthermore, a threshold applied to the distance, defines the number of pixels that remain unclassified (MEIJERINK et al. 1994, pp. 91-92; MATHER 1999, pp. 181-185).

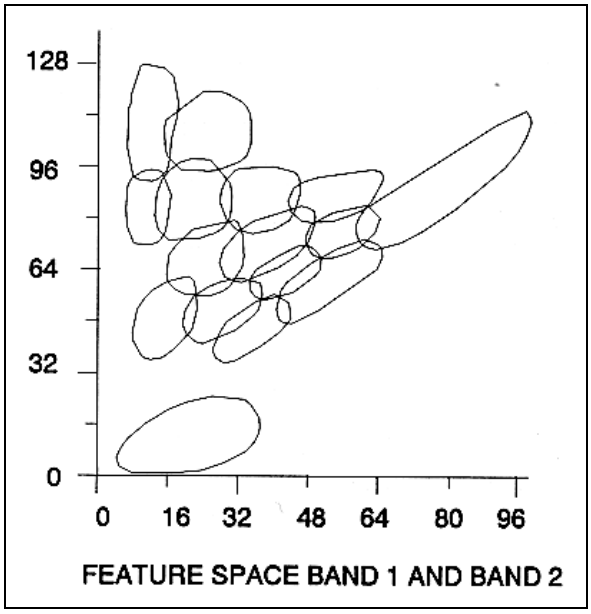


Figure 18: Arrangement of compact sub-classes in the Feature Space.
Exemplary arrangement of a large number of compact sub-classes in the feature space of band 1 plotted against band 2.
(Source: MEIJERINK et al. 1994, p. 64.)

In addition to the water classes, 'dump-classes' were introduced into the classification, which are primarily not of interest, but improve the classification results significantly. As it was done for the different water sub-classes, the classifier was thus also trained with the spectral patterns of burned[16] areas and parts of cloud shadows. These have a spectral overlap with what we summarize under the umbrella term "water", and lead to misclassifications. In combination with the various but compact water sub-classes, it is then possible to separate these different features. A pixel that represents a burnt surface is then assigned to the category "burnt" due to the shorter equi-probability distance to the mean of this class, rather than to a water class that is also close and would have

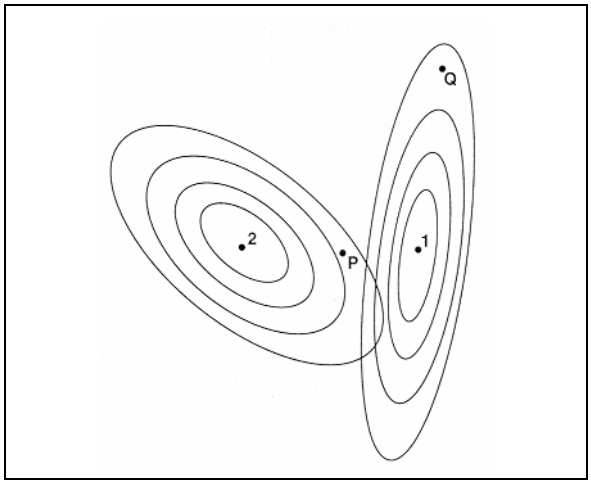


Figure 19: Equi-probability contours as used in Maximum Likelihood classification
The depicted ellipses represent equal probability contours around the classes 1 and 2. Point P, which is closer to the center of class 1, is assigned to class 2 due to the higher probability of its affiliation to class 2. Q is a member of class 1.
(Source: MATHER 1999, p. 181.)

[16] A different approach to map burned land is described by (KOUTSIAS/KARTERIS/CHUVIECO 2000). They distinguish burned areas upon color space transformation (IHS). Their method was tested on a subset of the Upper East's Landsat images, which also worked well, but finally the above-described approach was chosen.

taken-up that pixel if there was no "burnt" category introduced in the classification.

The difficulties arising from the spectral overlap between water and burned areas become much clearer in the comparison of the two Landsat scenes depicted in Appendix 8 that indicate the degree to which a landscape can be affected by bushfires. In the already familiar image from the 7th November 1999, reservoirs can easily be identified, such as the two large reservoirs Tono and Vea (see also Figure 1), but also the small dark spots in the bright patches that are less vegetated due to higher population density. Mainly at the edge of these bright patches, but also in the vegetated areas (green shades) there are less contiguous dark spots that represent burnt surfaces. The second image that was recorded only one month later (9th December 1999) shows a dramatic increase in the extent of burnt areas. As it is more difficult for the human eye to spot reservoirs in this image, also the classifier underperforms when not including the above described "dump classes". At a much lower level, the same accounts for the 7th November image.

After the classification was completed and the "dump classes" were discarded, a Boolean water image was created by regrouping the water sub-classes into one class. It was then filtered with a 3x3 median filter (JENSEN 1996, pp. 156-157; MATHER 1999, pp. 152-153) which both removes single pixels and fills 'single-pixel-holes' in reservoirs. Finally, based on visual interpretation, 'misclassified' river and riverbed pixels, oxbow lakes and the very few remaining burnt spots and cloud shadows were manually deleted.

The resulting reservoir maps were then mosaiked and georeferenced (quadratic mode) using 27 reference points derived from 1:50,000 maps that were available in digital format. The reference points were intersections, bridges and river mouths, however it was difficult to find large numbers of adequate reference points well spread throughout the image. Many times, intersections are the only points that qualify for geo-registration, but they are often located in a village or cluster of houses which appear as bright spots on the satellite image. This problem is explained by the blending of the dust roads with their mostly sparsely vegetated, bare or urban surroundings. Particularly in rural areas, houses are often built out of similar materials that lie bare in the surroundings. The same accounts for roads. Also at intersections outside villages, this problem occurs when they are surrounded by bare surfaces[17]. By also reverting to more variable features, such as river mouths, and bridges intersecting with rivers, lower accuracy was taken into account, resulting in an RMS of 30.17. An effect of combining the scenes of path 195 with the image strip of path 194 is best observed with the map depicted in Appendix 5. Among other things, it shows the Landsat ETM coverage of the Upper East Region and the point distribution of the classified reservoirs. Those parts of the Upper East Region that are solely covered by the

[17] Bare ground at intersections is a result of behavior of pedestrians and cattle. They do not walk to the center of the intersection to turn but take shortcuts as they approach the intersection and thereby tread down vegetation. Rectangular crossroads thus often shows roughly diamond shaped bare areas.

images of path 195, turned out to contribute no additional reservoir surfaces[18]. Only in the overlap with the image strip of path 194, reservoirs were classified. As the image strip of path 194 from the 7th November 1999 represents larger surface extents than the scenes that were acquired later in the season (scene 195/052 from the 13th October 1999 and scene 195/053 from the 2nd February 2000), their classification results were overlain by the larger ones from the 7th November image strip. Therefore, further references to the results of the satellite image classification refer to the image strip of path 194 (rows 052 and 053) from the 7th November 1999.

For the data extraction, the Upper East Region was masked with a Boolean image of its extent, leaving only the data that correspond to this region. The contiguous water pixels were grouped to clusters that resemble reservoirs, and given unique identifiers, upon which the area calculation was performed and each reservoir could be associated with its surface area. For later use in ArcView, a raster-to-vector conversion of the reservoirs was performed.

With the previously described technique, a total of 504 reservoirs was mapped adding up to a total acreage of 3,408 ha. Reservoirs classified with an area of less than 1 ha were deleted. Also, the two largest reservoirs, Tono (1,894 ha) and Vea (435 ha), were omitted in the study. Thereafter, a total of 154 reservoirs was left, summing up to an acreage of 999.54 ha with reservoirs that range from 1 to 35 ha in size. The 348 discarded reservoirs smaller than 1 ha only account for 80 ha of surface areas. Their hydrological impact is minimal and the classification uncertainty is getting high for such small features as each pixel is already 0.09 ha.

The selection of sites to visit was based on the evaluation of the data regarding the distribution of reservoir surface areas in the Upper East Region. No significant breaks were found (Figure 20), however it was not clear whether splitting the data in categories was necessary in further steps. For the benefit of a good distribution of samples over the full range of the different reservoir sizes, the dataset was split into three approximately equal categories. Category 1 consists of 51 reservoirs and ranges from 1 to 2.79 ha. Category 2 is the largest comprising of 53 reservoirs from 2.88 to 6.93 ha, and finally, category 3 consists of 50 reservoirs and spans from 7.02 to 32.49 ha.

[18] An earlier study on reservoir distribution that partially covers the same area confirms this absence of reservoirs in the area that is solely covered by scene 195/052. Further west, reservoirs are found again in the sparsely populated Upper West Region (see VAN DE GIESEN et al. 2002, p. 167).

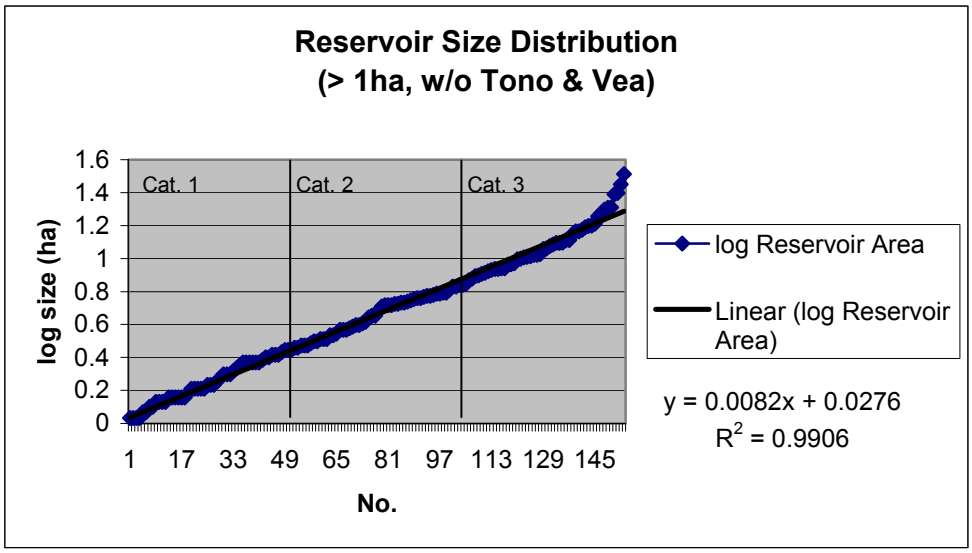


Figure 20: Size distribution of reservoirs and categorization.
Plotted against the number of reservoirs is the distribution of logarithmic reservoir sizes. The total no. of reservoirs involved is 154, which make up a total area of 999.54 ha. The vertical lines indicate the segmentation into three equally large categories. Reservoirs smaller than 1 ha, as well as the two large reservoirs Tono and Vea were discarded.

Problematic was the decision on the number of reservoirs to be measured. There is no statistical means to determine the number of reservoirs that needs to be measured in order to achieve a representative sample. A plausible solution is to have an extensive sample set to practically ensure to have a sufficient coverage. By measuring 60 out of the 154, or 40% of the total population, the variance that occurs is expected to be adequately represented.

Thereupon, 25 reservoirs were randomly selected within each category. The first 20 drawn reservoirs of each category were selected for field evaluation, leaving five spare samples per category. Following the order they were drawn, they were supposed to be subsequently used as substitutes in case of misclassification, dam failure, or other unforeseen events making the evaluation impossible (Appendix 8).

An extensive description and evaluation of the reservoirs follows in Chapter 3.3.2 and 5.