Fundamentals of Remote Sensing

Edited and written by Noam Levin

November 1999

1st Hydrographic Data Management course, IMO - International Maritime Academy, Trieste, Italy

Remote Sensing Laboratory, Geography Department, Tel Aviv University, Israel

GIS unit, the Society for the Protection of Nature in Israel

noam71levin@hotmail.com

1

Table of contents

Thanks:

First, I would like to express my thanks to the following people, for making this work possible:

- The staff of IMO-IMA, for supplying me with the appropriate working conditions (free access to the Internet, the academy's library) and helping whenever needed,
- the staff of the Earth Sciences library of the Trieste university, for their help in lending books and making photocopies of articles,
- all the people and institutions that maintain internet sites, and publish on-line articles, tutorials, technical documentations, etc (references mentioned along the text, the principal sources at the end of the text),
- Dr. Eyal Ben Dor from the geography department of Tel Aviv's university for introducing me to the field of Remote Sensing, and
- to my family and friends.

Preface:

- *Craig J.M. (1998), The application of satellite imagery in support of nautical charting; past experience and future possibilities - a practical view, International Hydrographic Review, LXXV(1), no. 142, pp. 95-105*

The field of Remote Sensing is very wide, both in the data acquisition methods, data processing procedures and techniques and the applications it is used for; it is also a fast developing field, in all the above themes. This text is therefore intended to give only a general overview about several subjects, yet I hope, an extensive one, covering all the important topics regarding Remote Sensing of the surface of the Earth. The text also attempts to give the reader an understanding of the capabilities and limitations of Remote Sensing. Very few equations and formulas will be given in the text, as the focus will be on understanding the basic ideas. The main subjects covered are:

- The role Remote Sensing plays in our understanding of the Earth and the natural and human processes affecting it,
- The radiometric and geometric principles of Remote Sensing,
- The principal sensors used, and their characteristics, in passive and active, imaging and non-imaging Remote Sensing, on airborne or on satellite platforms, from monochromatic to hyperspectral,
- The pre processing phase of data: radiometric, atmospheric, geometric and noises corrections, and
- The image processing phase of data: visualisation, enhancement and classification.

Due to the wide scope covered, the subjects could not be covered in details and the interested reader should turn to the relevant literature.

Detailed examples of Remote Sensing applications will be given in areas which have direct importance for either hydrography or oceanography. The advantages of acquiring information by Remote Sensing apply, irrespective of platform or sensor, also for hydrography:

1. It is cheaper than conventional surveying;
2. It is safer than hydrographic surveying in shoal areas such as coral reefs;
3. It is capable of change detection in rapidly developing ports and regular monitoring of mobile areas such as deltas and sandbanks;
4. World-wide coverage is comercially available, without security, political, or copyright restrictions, enabling dataa acquisition from remote areas;
5. The inherent geometry and therefore the relative positioning of features within a single scene is generally very good.

Through out the work all the references from which I have taken material are given, though not in the formal academic way.

1. Introduction:

1.1 Definition

- *http://www.vtt.fi/aut/rs/virtual/defin.html*
1. The Definition of **Remote Sensing** In the broadest sense, the measurement or acquisition of information of some property of an object or phenomenon, by a recording device that is not in physical or intimate contact with the object or phenomenon under study; e.g., the utilization at a distance (as from aircraft, spacecraft, or ship) of any device and its attendant display for gathering information pertinent to the environment, such as measurements of force fields, electromagnetic radiation, or acoustic energy. The technique employs such devices as the camera, lasers, and radio frequency receivers, radar systems, sonar, seismographs, gravimeters, magnetometers, and scintillation counters.
2. The practice of data collection in the wavelengths from ultraviolet to radio regions. This restricted sense is the practical outgrowth from airborne photography. Sense (1) is preferred and thus includes regions of the EM spectrum as well as techniques traditionally considered as belonging to conventional geophysics.

- *http://www.ciesin.org/TG/RS/satremot.html*

As humans, we are intimately familiar with remote sensing in that we rely on visual perception to provide us with much of the information about our surroundings. As sensors, however, our eyes are greatly limited by 1) sensitivity to only the visible range of electromagnetic energy; 2) viewing perspectives dictated by the location of our bodies; and 3) the inability to form a lasting record of what we view. Because of these limitations, humans have continuously sought to develop the technological means to increase our ability to see and record the physical properties of our environment.

Beginning with the early use of aerial photography, remote sensing has been recognized as a valuable tool for viewing, analyzing, characterizing, and making decisions about our environment. In the past few decades, remote sensing technology has advanced on three fronts: 1) from predominantly military uses to a variety of environmental analysis applications that relate to land, ocean, and atmosphere issues; 2) from (analog) photographic systems to sensors that convert energy from many parts of the electromagnetic spectrum to electronic signals; and 3) from aircraft to satellite platforms.

> Today, we define satellite remote sensing as the use of satellite-borne sensors to observe, measure, and record the electromagnetic radiation reflected or emitted by the Earth and its environment for subsequent analysis and extraction of information.

1.2 Comparison to maps, GIS, aerial photography / Photogrammetry, SONAR

Here will be given the main points of similarity and difference between the field of Remote Sensing (analysis and images) and the fields/products mentioned above.

1.2.1 Satellite Images *Vs* Maps

- *Anson R.W. and Ormeling F.J., (1993), Basic Cartography – for students and technicians, Vol 1., Elsevier*
- *Portugaly Y. (1996), "The maps hidden from the eye", Mishkafayim, 27, pp. 44-47 (in Hebrew)*

According to the International Cartographic Union, a map is "a conventionalised image representing selected features or characteristics of geographical reality, designed for use when spatial relationships are of primary importance". This definition does declare that in every map there's a process of selection present (and in addition - symbolization, abstraction and generalization), but also keeps the aura of scientific accuracy of a map. But, we should remember, that "a map shows us the world as we know it, and what we know, is a very complex subject, that is comprised of:

- The limits of matter, technology and our measurement tools,
- what we believe that exists,
- what we think to be important,
- and what we want and aspire to"

Thus, a map is a subjective, for we always decide what to put on it, and how to represent it. A Remote Sensing image in contrast, is an objective recording of the Electromagnetic reaching the sensor.

Another important difference, is that a map is a projection of the earth on paper, without any relief displacements, while in a Remote Sensing image both relief displacements and geometrical distortions.

1.2.2 Remote Sensing *Vs* GIS

GIS (Geographic Information System) is a kind of software that enables:

- The collection of spatial data from different sources (Remote Sensing being one of them).
- Relating spatial and tabular data.
- Performing tabular and spatial analysis.
- Symbolize and design the layout of a map.

A GIS software can handle both vector and raster data (some handle only one of them). Remote Sensing data belongs to the raster type, and usually requires special data manipulation procedures that regular GIS does not offer. However, after a Remote Sensing analysis has been done, its results are usually combined within a GIS or into database of an area, for further analysis (overlaying with other layers, etc). In the last years, more and more vector capabilities are being added to Remote Sensing softwares, and some Remote Sensing functions are inserted into GIS modules.

1.2.3 Remote Sensing *Vs* Aerial Photography / Photogrammetry

Both systems gather data about the upper surface of the Earth, by measuring the Electromagnetic radiation, from airborne systems. The following major differences can be given:

- Aerial photos are taken by an analog instrument: a film of a (photogrammetric) camera, then scanned to be transformed to digital media. Remote Sensing data is usually gathered by a digital CCD camera.

- The advantage of a film is its high resolution (granularity), while the advantage of the CCD is that we measure quantitatively the radiation reaching the sensor (radiance values, instead of a gray-value scale bar). Thus, Remote Sensing data can be integrated into physical equations of energy-balance for example.
- An Aerial photograph is a central projection, with the whole picture taken at one instance. A Remote Sensing image is created line after line; therefor, the geometrical correction is much more complex, with each line (or even pixel) needing to be treated as a central projection.
- Aerial photos usually gather data only in the visible spectrum (there are also special films sensitive to near infrared radiation), while Remote Sensing sensors can be designed to measure radiation all along the Electromagnetic spectrum.
- Aerial photos are usually taken from planes, Remote Sensing images also from satellites.
- Both systems are affected by atmospheric disturbances. Aerial photos mainly from haze (that is, the scattering of light – the process which makes the sky blue), Remote Sensing images also from processes of absorption. Atmospheric corrections to Aerial photos can be made while taking the picture (using a filter), or in post-processing, as in done Remote Sensing. Thermal Remote Sensing sensors can operate also at nighttime, and Radar data is almost weather independent.
- In Photogrammetry the main efforts are dedicated for the accurate creation of a 3d model, in order to plot with high accuracy the location and boundaries of objects, and to create a Digital Elevation Model, by applying sophisticated geometric corrections. In Remote Sensing the main efforts are dedicated for the analysis of the incoming Electromagnetic spectrum, using atmospheric corrections, sophisticated statistical methods for classification of the pixels to different categories, and analysing the data according to known physical processes that affect the light as it moves in space and interacts with objects.
- Remote Sensing images are very useful for tracking phenomena on regional, continental and even global scale, using the fact that satellites cover in each image a wide area, and taking images all the time (whether fixed above a certain point, or "revisiting" the same place every 15 days (for example).
- Remote Sensing images are available since the early 1970's. Aerial photos, provide a longer time span for landscape change detection (the regular coverage of Israel by Aerial photos started in 1944/5, for example, with many Aerial photos taken also during World War 1).
- Remote Sensing images are more difficult to process, and require trained personnel, while aerial photographs can be interpreted more easily.

1.2.4 Remote Sensing *Vs* SONAR

The SONAR can also be considered as Remote Sensing – that is, studying the surfaces of the sea (bathymetry and sea bed features) from a distance. The SONAR is an active type of Remote Sensing (like Radar; Not depending on an external source of waves, measuring the time between the transmission and reception of waves produced by our instruments, and their intensity), but using sound waves, and not Electromagnetic radiation.

Both systems transmit waves through an interfering medium (water, air), that adds noise to the data we are looking for, and there for corrections must be applied to the

raw data collected. In Remote Sensing however, Radar is considered to be almost weather independent, and atmospheric disturbances affect mainly passive Remote Sensing). To make these necessary corrections, both systems depend on callibration from field data (be it salinity, temperature and pressure measured by the ship while surveying, or measurements of the atmospheric profile parameters by a meteorological radiosonde for example).

Sonar's are mainly used to produce the bathymetry of the sea, while Remote Sensing techniques are focusing more on identification of the material's properties than on its height.

Echo-sounders (single or multi-beam) can be compared to Airborne Laser Scanning – both of them create point (vector) data containing X,Y,Z, that needs to be further post processed in order to remove noise (spikes). An added complexity when dealing with bathymetry (as opposed to topography) is the need for tide corrections.

Side Scan SONAR can be compared to Side Looking Aperture RADAR, both of them creating images (raster) analyzing the surface.

Another major difference is that in Remote Sensing the results of the analysis can be compared easily to the field (aerial photos, maps, field measurments), while in SONAR the underlying bottom of the sea is hidden from us, and we depend totally on the data gathered.

1.3 Applications in general

▪ *http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/indexe.html*

As will be learned in the section on sensors, each one is designed with a specific purpose. With optical sensors, the design focuses on the spectral bands to be collected. With radar imaging, the incidence angle and microwave band used plays an important role in defining which applications the sensor is best suited for.

Each application itself has specific demands, for spectral resolution, spatial resolution, and temporal resolution. There can be many applications for Remote Sensing, in different fields, as described below. In the body of this tutorial of Remote Sensing, some applications relevant for hydrography and oceanography will be given in more detail.

1.3.1 Agriculture

Agriculture plays a dominant role in economies of both developed and undeveloped countries. Satellite and airborne images are used as mapping tools to classify crops, examine their health and viability, and monitor farming practices. Agricultural applications of remote sensing include the following:

- crop type classification
- crop condition assessment
- crop yield estimation
- mapping of soil characteristics
- mapping of soil management practices
- compliance monitoring (farming practices)

1.3.2 Forestry

Forests are a valuable resource providing food, shelter, wildlife habitat, fuel, and daily supplies such as medicinal ingredients and paper. Forests play an important role in balancing the Earth's CO2 supply and exchange, acting as a key link between the atmosphere, geosphere, and hydrosphere.

Forestry applications of remote sensing include the following:

- reconnaissance mapping:
  Objectives to be met by national forest/environment agencies include forest cover updating, depletion monitoring, and measuring biophysical properties of forest stands.
- Commercial forestry:
  Of importance to commercial forestry companies and to resource management agencies are inventory and mapping applications: collecting harvest information, updating of inventory information for timber supply, broad forest type, vegetation density, and biomass measurements.
- Environmental monitoring:
  Conservation authorities are concerned with monitoring the quantity, health, and diversity of the Earth's forests.

1.3.3 Geology

Geology involves the study of landforms, structures, and the subsurface, to understand physical processes creating and modifying the earth's crust. It is most commonly understood as the exploration and exploitation of mineral and hydrocarbon resources, generally to improve the conditions and standard of living in society.

Geological applications of remote sensing include the following:

- surficial deposit / bedrock mapping
- lithological mapping
- structural mapping
- sand and gravel (aggregate) exploration/ exploitation
- mineral exploration
- hydrocarbon exploration
- environmental geology
- geobotany
- baseline infrastructure
- sedimentation mapping and monitoring
- event mapping and monitoring
- geo-hazard mapping
- planetary mapping

1.3.4 Hydrology

Hydrology is the study of water on the Earth's surface, whether flowing above ground, frozen in ice or snow, or retained by soil

Examples of hydrological applications include:

- wetlands mapping and monitoring,
- soil moisture estimation,
- snow pack monitoring / delineation of extent,
- measuring snow thickness,
- determining snow-water equivalent,
- river and lake ice monitoring,
- flood mapping and monitoring,
- glacier dynamics monitoring (surges, ablation)
- river /delta change detection
- drainage basin mapping and watershed modelling

- irrigation canal leakage detection
- irrigation scheduling

1.3.5 Sea Ice

Ice covers a substantial part of the Earth's surface and is a major factor in commercial shipping and fishing industries, Coast Guard and construction operations, and global climate change studies.

Examples of sea ice information and applications include:
- ice concentration
- ice type / age /motion
- iceberg detection and tracking
- surface topography
- tactical identification of leads: navigation: safe shipping routes/rescue
- ice condition (state of decay)
- historical ice and iceberg conditions and dynamics for planning purposes
- wildlife habitat
- pollution monitoring
- meteorological / global change research

1.3.6 Land Cover & Land Use

Although the terms land cover and land use are often used interchangeably, their actual meanings are quite distinct. Land cover refers to the surface cover on the ground, while Land use refers to the purpose the land serves. The properties measured with remote sensing techniques relate to land cover, from which land use can be inferred, particularly with ancillary data or a priori knowledge.

Land use applications of remote sensing include the following:
- natural resource management
- wildlife habitat protection
- baseline mapping for GIS input
- urban expansion / encroachment
- routing and logistics planning for seismic / exploration / resource extraction activities
- damage delineation (tornadoes, flooding, volcanic, seismic, fire)
- legal boundaries for tax and property evaluation
- target detection - identification of landing strips, roads, clearings, bridges, land/water interface

1.3.7 Mapping

Mapping constitutes an integral component of the process of managing land resources, and mapped information is the common product of analysis of remotely sensed data.

Mapping applications of remote sensing include the following:
- Planimetry:
  Land surveying techniques accompanied by the use of a GPS can be used to meet high accuracy requirements, but limitations include cost effectiveness, and difficulties in attempting to map large, or remote areas. Remote sensing provides a means of identifying and presenting planimetric data in convenient media and efficient manner. Imagery is available in varying scales to meet the requirements of many different users. Defence applications typify the scope of planimetry

applications - extracting transportation route information, building and facilities locations, urban infrastructure, and general land cover.

- digital elevation models (DEM's):
  Generating DEMs from remotely sensed data can be cost effective and efficient. A variety of sensors and methodologies to generate such models are available and proven for mapping applications. Two primary methods if generating elevation data are 1. Stereogrammetry techniques using airphotos (photogrammetry), VIR imagery, or radar data (radargrammetry), and 2. Radar interferometry.

- Baseline thematic mapping / topographic mapping:
  As a base map, imagery provides ancillary information to the extracted planimetric or thematic detail. Sensitivity to surface expression makes radar a useful tool for creating base maps and providing reconnaissance abilities for hydrocarbon and mineralogical companies involved in exploration activities. This is particularly true in remote northern regions, where vegetation cover does not mask the microtopography and generally, information may be sparse. Multispectral imagery is excellent for providing ancillary land cover information, such as forest cover. Supplementing the optical data with the topographic relief and textural nuance inherent in radar imagery can create an extremely useful image composite product for interpretation.

1.3.8 Oceans & Coastal Monitoring

The oceans not only provide valuable food and biophysical resources, they also serve as transportation routes, are crucially important in weather system formation and CO_2 storage, and are an important link in the earth's hydrological balance. Coastlines are environmentally sensitive interfaces between the ocean and land and respond to changes brought about by economic development and changing land-use patterns. Often coastlines are also biologically diverse inter-tidal zones, and can also be highly urbanized .

Ocean applications of remote sensing include the following:

- Ocean pattern identification:
  - currents, regional circulation patterns, shears
  - frontal zones, internal waves, gravity waves, eddies, upwelling zones, shallow water bathymetry ,
- Storm forecasting
  - wind and wave retrieval
- Fish stock and marine mammal assessment
  - water temperature monitoring
  - water quality
  - ocean productivity, phytoplankton concentration and drift
  - aquaculture inventory and monitoring
- Oil spill
  - mapping and predicting oilspill extent and drift
  - strategic support for oil spill emergency response decisions
  - identification of natural oil seepage areas for exploration
- Shipping
  - navigation routing
  - traffic density studies
  - operational fisheries surveillance
  - near-shore bathymetry mapping

- Intertidal zone
  - tidal and storm effects
  - delineation of the land /water interface
  - mapping shoreline features / beach dynamics
  - coastal vegetation mapping
  - human activity / impact

2. Electromagnetic radiation:

Sabins Floyd F. (1976), Remote Sensing – Principles and Interpretation, Freeman
▪ *http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/indexe.html*

2.1 Electromagnetic energy

Electromagnetic energy refers to all energy that moves with the velocity of light in a harmonic wave pattern. The word *harmonic* implies that the component waves are equally and repetitively spaced in time. The wave concept explains the propagation of Electromagnetic energy, but this energy is detectable only in terms of its interaction with matter. In this interaction, Electromagnetic energy behaves as though it consists of many individual bodies called *photons* that have such particle-like properties as energy and momentum.

Electromagnetic waves can be described in terms of their:

- Velocity: The speed of light, $c=3*10^8$ m*sec^{-1}).
- Wavelength: l, the distance from any position in a cycle to the same position in the next cycle, measured in the standard metric system. Two units are usually used: the micrometer (mm, 10^{-6}m) and the nanometer (nm, 10^{-9}m).

- Frequency: n, the number of wave crests passing a given point in specific unit of time, with one hertz being the unit for a frequency of one cycle per second.

Wavelength and frequency are related by the following formula:

$$c = l *\ n$$

Electro-Magnetic radiation consists of an electrical field (E) which varies in magnitude in a direction perpendicular to the direction in which the radiation is traveling, and a magnetic field (M) oriented at right angles to the electrical field. Both these fields travel at the speed of light (c).
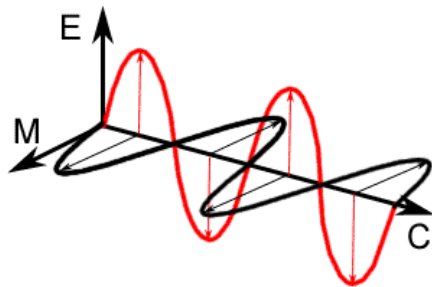
Figure: Electro-Magnetic radiation   Figure: wavelength and frequency
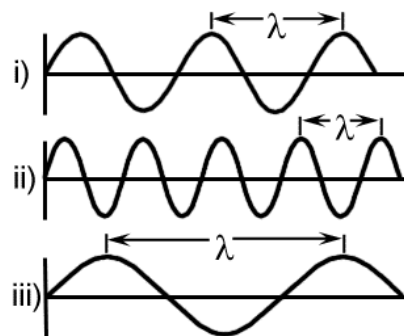
2.2 Interaction mechanisms

A number of interactions are possible when Electromagnetic energy encounters matter, whether solid, liquid or gas. The interactions that take place at the surface of a substance are called *surface phenomena*. Penetration of Electromagnetic radiation beneath the surface of a substance results in interactions called *volume phenomena*. The surface and volume interactions with matter can produce a number of changes in the incident Electromagnetic radiation; primarily changes of magnitude, direction, wavelength, polarization and phase. The science of Remote Sensing detects and records these changes. The resulting images and data are interpreted to identify remotely the characteristics of the matter that produced the changes in the recorded Electromagnetic radiation.

The following interactions may occur:

- Radiation may be *transmitted*, that is, passed through the substance. The velocity of Electromagnetic radiation changes as it is transmitted from air, or a vacuum, into other substances.
- Radiation may be *absorbed* by a substance and give up its energy largely to heating the substance.
- Radiation may be *emitted* by a substance as a function of its structure and temperature. All matter at temperatures above absolute zero, 0°K, emits energy.
- Radiation may be *scattered*, that is, deflected in all directions and lost ultimately to absorption or further scattering (as light is scattered in the atmosphere).
- Radiation may be *reflected*. If it is returned unchanged from the surface of a substance with the angle equal and opposite to the angle of incidence, it is termed *specular* reflectance (as in a mirror). If radiation is reflected equally in all directions, it is termed *diffuse*. Real materials lie somewhere in between.
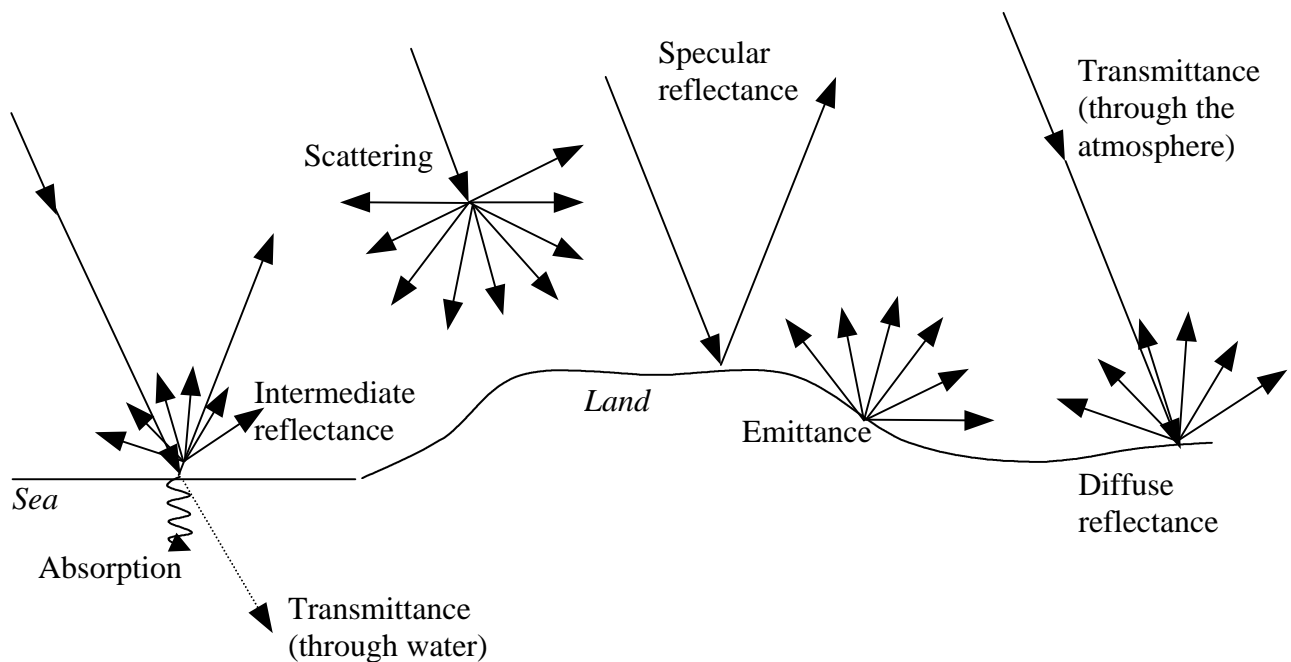
The interactions with any particular form of matter are selective with regard to the Electromagnetic radiation and are specific for that form of matter, depending primarily upon its surface properties and its atomic and molecular structure.

2.3 Laws regarding the amount of energy radiated from an object

2.3.1 Planck Radiation Law
▪ *http://csep10.phys.utk.edu/astr162/lect/light/radiation.html*

The primary law governing blackbody radiation is the *Planck Radiation Law*, which governs the intensity of radiation emitted by unit surface area into a fixed direction (solid angle) from the blackbody as a function of wavelength for a fixed temperature. The Planck Law can be expressed through the following equation.

$$E(\lambda, T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1}$$

$$h = 6.625 \times 10^{-27} \text{ erg-sec} \quad \text{(Planck Constant)}$$

$$k = 1.38 \times 10^{-16} \text{ erg/K} \quad \text{(Boltzmann Constant)}$$

$$c = 3 \times 10^{10} \text{ cm/sec} \quad \text{(Speed of Light)}$$

Every object with a temperature above the absolute zero radiates energy. The relationship between wavelength and the amount of energy radiated at different wavelengths, is shown in the following figure, and formulated above.

Emitted Radiance of Sun and Earth
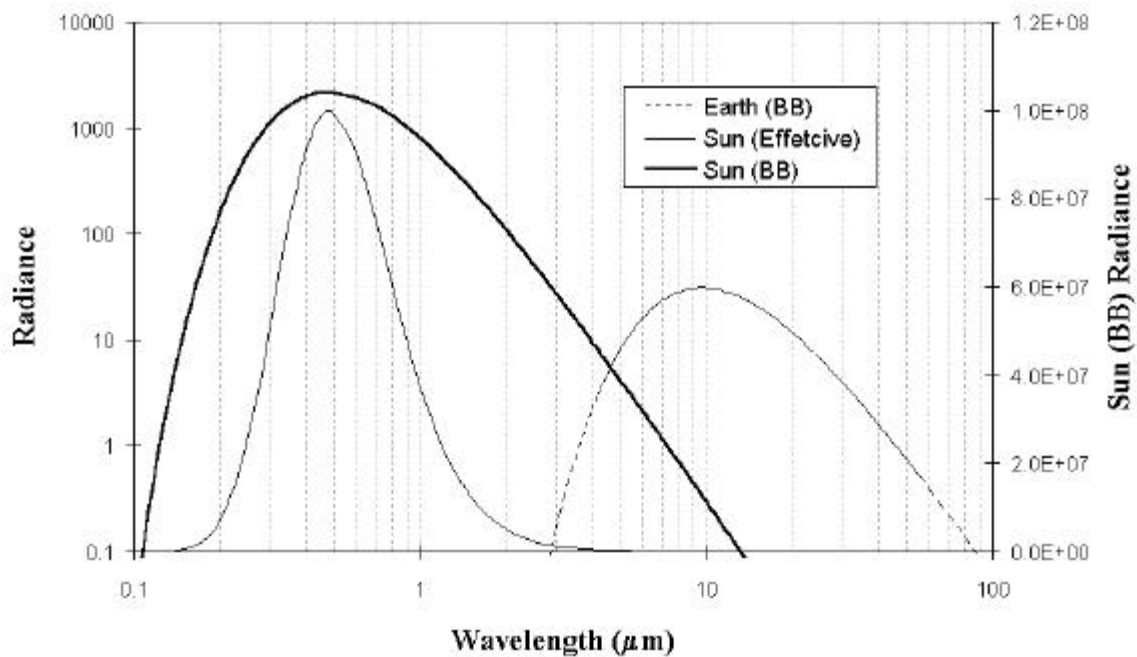
The figure shows the emitted radiance of the Earth and the Sun as Black Bodies (each given on a different Y scale), according to Plank Law, and also the effective radiance of the sun reaching the Earth. It can be seen, that from about 2-3 μm the radiance emitted from the Earth is greater than reaching us from the Sun (both of them presented on the same scale).

2.3.2 Wien's displacement law

For an object at a constant temperature the radiant power peak refers to the wavelength at which the maximum amount of energy is radiated, which is expressed as l_{max}. The sun, with a surface temperature of almost 6000°K, has its peak at 0.48μm (wavelength of yellow). The average surface temperature of the earth is 290°K (17°C), which is also called the *ambient temperature*; the peak concentration of energy emitted from the earth is at 9.7μm.

This shift to longer wavelengths with decreasing temperature is described by Wien's displacement law, which states:

$$l_{max} = 2,897\mu m°K / T_{rad}°K$$

19

2.3.3 Black body concept, Emissivity and Radiant Temperature

Temperature is a measure of the concentration of heat. The concentration of kinetic heat of a body of material may be called the *kinetic temperature*, T_{kin} and is measured by a thermometer placed in direct contact with the material.

By definition a *black body* is a material that absorbs all the radiant energy that strikes it. A black body also radiates the maximum amount of energy, which is dependent on the kinetic temperature. According to the Stefan-Boltzman law the radiant flux of a black body, F_b, at a kinetic temperature, T_{kin}, is $\boxed{F_b = s * T_{kin}^4}$ where s is the Stefan-Boltzman constant, $5.67 * 10^{-12}$ $W * cm^{-2} * {}^\circ K^{-4}$.

A black body is a physical abstraction, for no material has an absolute absorptivity, and no material radiates the full amount of energy as defined in the equation above. For real materials a property called *emissivity*, e, has been defined, as $\boxed{e = F_r/F_b}$, where F_r is radiant flux from a real material. For a black body $e=1$, but for all real materials $e<1$. Emissivity is wavelength dependent, which means that the emissivity of a material is different when is measured at different wavelengths of radiant energy (each material has both a reflectance spectrum and an emissivity spectrum – see later). In the next table are given the average emissivities of various materials in the 8 to 12mm wavelength region (thermal), which is used in Remote Sensing.

Material	Emissivity, e
Polished metal surface	0.006
Granite	0.815
Quartz sand, large grains	0.914
Dolomite, polished	0.929
Basalt, rough	0.934
Asphalt paving	0.959
Concrete walkway	0.966
A coat of flat black paint	0.970
Water, with a thin film of petroleum	0.972
Water, pure	0.993

Thus, the radiant flux of a real material may be expressed as $\boxed{F_r = e * s * T_{kin}^4}$. Emissivity is a measure of the ability of a material to both radiate and absorb energy. Materials with a high emissivity absorb and radiate large proportions of incident and kinetic energy, respectively (and vice-versa). The result, is that two surfaces, with the same kinetic temperature but with a different emissivity will have a different radiant temperature. As Remote Sensing devices measure the radiant temperature, in order to derive the kinetic temperature of the object, we need to know its emissivity, that is, w need to be able to identify the material (see later, in the chapter about atmospheric corrections). Then we can apply the following equation: $\boxed{T_{rad} = e^{1/4} T_{kin}}$.

2.4 Electromagnetic Spectrum

The Electromagnetic spectrum is the continuum of energy ranging from kilometers to nanometers in wavelength. This continuum is commonly divided into the following ranges, called spectral bands, the boundaries between them being gradational.
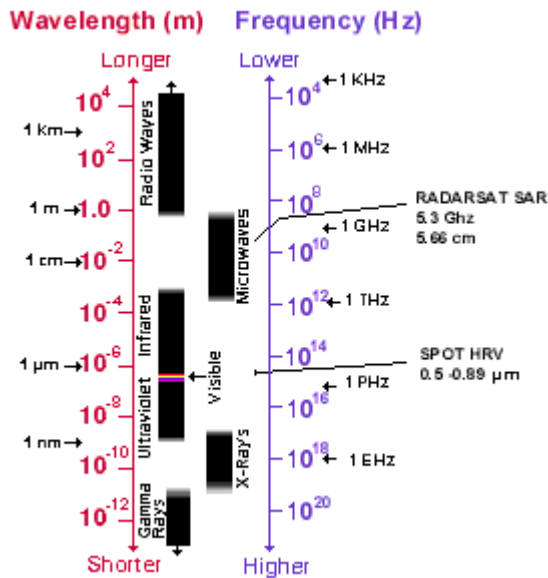
A *passive* Remote Sensing system records the energy naturally radiated or reflected from an object. An *active* Remote Sensing system supplies its own source of energy, which is directed at the object in order to measure the returned energy. Flash photography is active Remote Sensing in contrast to available light photography, which is passive. An other common form of active Remote Sensing is radar, which provides its own source of Electromagnetic energy in the microwave region. Airborne laser scanning is a relatively new form of active Remote Sensing, operating the in the visible and Near Infra Red wavelength bands.
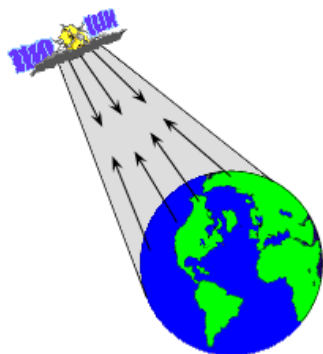
Active system
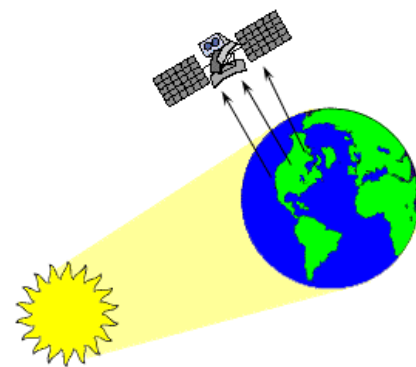
Passive system

2.4.1 Wavelength bands

Band	Wavelength	Remarks
Gamma ray	<0.03 nm	Incoming radiation from the sun is completely absorbed by the upper atmosphere, and is not available for Remote Sensing. Gamma radiation from radioactive minerals is detected by low-flying aircraft as a prospecting method.
X-ray	0.03 to 0.3 nm	Incoming radiation is completely absorbed by atmosphere. Not employed in Remote Sensing.
Ultraviolet, UV	3 nm to 0.4 µm	Incoming UV radiation atmosphere wavelengths <0.3 µm is completely absorbed by ozone in the upper atmosphere.
Photographic UV	0.3 to 0.4 µm	Transmitted through the atmosphere. Detectable with film and photodetectors, but atmospheric scattering is severe.
Visible	0.4 to 0.7 µm	Detected with film and photodetectors. Includes earth reflectance peak at about 0.5 µm.
Infrared, IR	0.7 to 300 µm	Interaction with matter varies with wavelength. Atmospheric transmission windows are separated by absorption bands.
Reflected IR	0.7 to 3 µm	This is primarily reflected solar radiation and contains no information about thermal properties of materials. Commonly divided into the following regions: • Near Infra Red (NIR) between 0.7 to 1.1 µm. • Middle Infra Red (MIR) between 1.3 to 1.6 µm. • Short Wave Infra Red (SWIR) between 2 to 2.5 µm. Radiation from 0.7 to 0.9 µm is detectable with film and is called photographic IR radiation.
Thermal IR	3 to 5 µm 8 to 14 µm	These are the principal atmospheric windows in the thermal region. Imagery at these wavelengths is acquired through the use of optical-mechanical scanners, not by film.
Microwave	0.3 to 300 cm	These longer wavelengths can penetrate clouds and fog. Imagery may be acquired in the active or passive mode.
Radar	0.3 to 300 cm	Active mode of microwave Remote Sensing.

2.4.2 Atmosphere effects

▪ *http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/indexe.html*

Our eyes inform us that the atmosphere is essentially transparent to light, and we tend to assume that this condition exists for all Electromagnetic radiation. In fact, however, the gases of the atmosphere selectively scatter light of different wavelengths. The gases also absorb Electromagnetic energy at specific wavelength intervals called *absorption bands*. The intervening regions of high energy transmittance are called *atmospheric transmission bands,* or *windows*. The transmission and absorption bands are shown in the following figure, together with the gases responsible for the absorption bands.

Particles and gases in the atmosphere can affect the incoming light and radiation. These effects are caused by the mechanisms of **scattering** and **absorption**.

2.4.2.1 Scattering

Scattering occurs when particles or large gas molecules present in the atmosphere interact with and cause the electromagnetic radiation to be redirected from its original path. How much scattering takes place depends on several factors including the wavelength of the radiation, the abundance of particles or gases, and the distance the radiation travels through the atmosphere. There are three (3) types of scattering which take place.

Rayleigh scattering occurs when particles are very small compared to the wavelength of the radiation. These could be particles such as small specks of dust or nitrogen and oxygen molecules. Rayleigh scattering causes shorter wavelengths of energy to be scattered much more than longer wavelengths. Rayleigh scattering is the dominant scattering mechanism in the upper atmosphere. The fact that the sky appears "blue" during the day is because of this phenomenon. As sunlight passes through the atmosphere, the shorter wavelengths (i.e. blue) of the visible spectrum are scattered more than the other (longer) visible wavelengths. At sunrise and sunset the light has to travel farther through the atmosphere than at midday and the scattering of the shorter wavelengths is more complete; this leaves a greater proportion of the longer wavelengths to penetrate the atmosphere (thus the sky is "painted" in red).
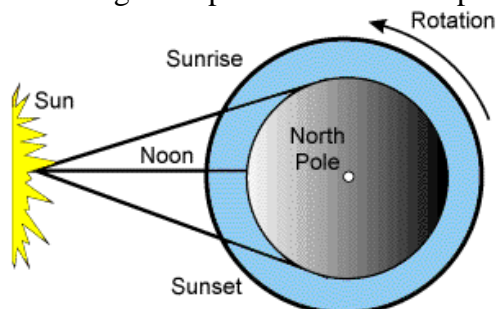
Mie scattering occurs when the particles are just about the same size as the wavelength of the radiation. Dust, pollen, smoke and water vapour are common causes of Mie scattering which tends to affect longer wavelengths than those affected by Rayleigh scattering. Mie scattering occurs mostly in the lower portions of the atmosphere where larger particles are more abundant, and dominates when cloud conditions are overcast.

The final scattering mechanism of importance is called **nonselective scattering**. This occurs when the particles are much larger than the wavelength of the radiation. Water droplets and large dust particles can cause this type of scattering. Nonselective scattering gets its name from the fact that all wavelengths are scattered about equally. This type of scattering causes fog and clouds to appear white to our eyes because blue, green, and red light are all scattered in approximately equal quantities (blue+green+red light = white light).

2.4.2.2 Absorption

▪ *http://speclab.cr.usgs.gov/PAPERS.refl-mrs/refl4.html*

Absorption is the other main mechanism at work when electromagnetic radiation interacts with the atmosphere. In contrast to scattering, this phenomenon causes molecules in the atmosphere to absorb energy at various wavelengths. Ozone, carbon dioxide, and water vapour are the three main atmospheric constituents which absorb radiation.

Any effort to measure the spectral properties of a material through a planetary atmosphere, must consider where the atmosphere absorbs. For example, the Earth's atmospheric transmittance is shown in next Figure. The drop toward the ultraviolet is due to scattering and strong ozone absorption at wavelengths short of 0.35 μm. Ozone also displays an absorption at 9.6 μm. Oxygen absorbs at 0.76 μm in a narrow feature. CO_2 absorbs at 2.01, 2.06, and a weak doublet near 1.6 μm. Water causes most of the rest of the absorption throughout the spectrum and hides additional (weaker) absorptions from other gases. The mid-IR spectrum in following Figure shows the effect of doubling CO_2, which in this case is small compared to the absorption due to water. While we will see that the spectral region near 1.4 and 3 μm can be diagnostic

of OH-bearing minerals, we can't usually use these wavelengths when remotely measuring spectra through the Earth's atmosphere (it has been done from high elevation observatories during dry weather conditions). Those areas of the spectrum which are not severely influenced by atmospheric absorption and thus, are useful to remote sensors, are called **atmospheric windows**.

However, these spectral regions can be used in the laboratory where the atmospheric path lengths are thousands of times smaller, or when measuring spectra of other planets from orbiting spacecraft.
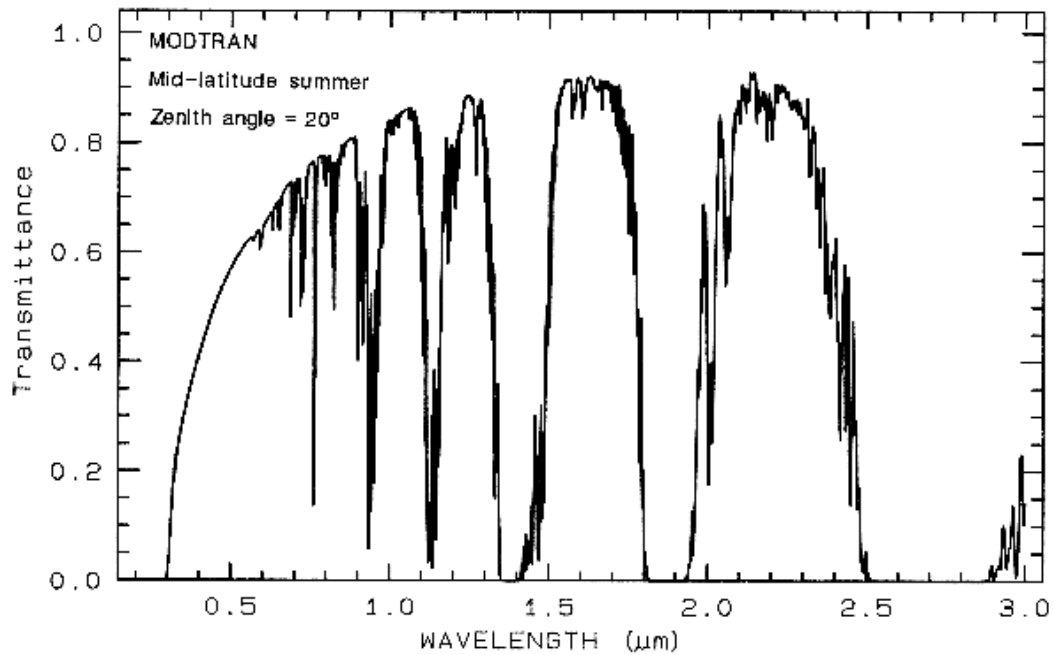
Figure: Modtran (Berk *et al.*, 1989) modeled atmospheric transmittance, visible to near-infrared. Most of the absorptions are due to water. Oxygen occurs at 0.76 μm, carbon dioxide at 2.0 and 2.06 μm.
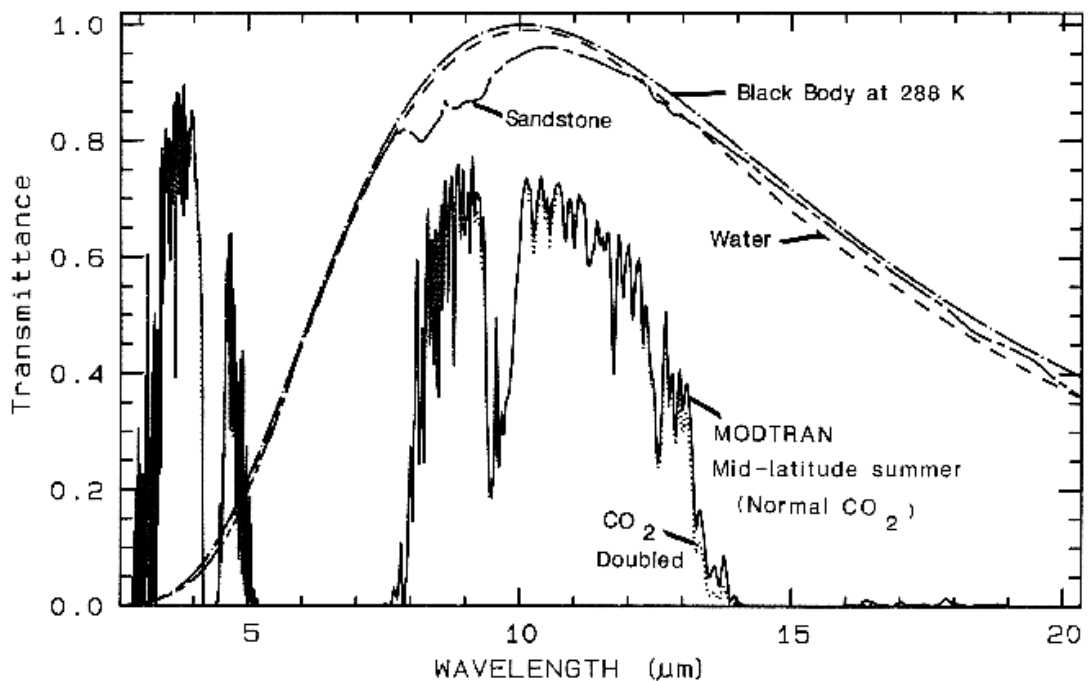
Figure: Atmospheric transmittance, mid-infrared is compared to scaled grey-body spectra. Most of the absorption is due to water. Carbon dioxide has a strong 15-μm band, and the dotted line shows the increased absorption due to doubling CO_2. Also shown is the black-body emission at 288 K and the grey-body emission from water and a sandstone scaled to fit on this transmittance scale. The water and sandstone curves were computed from reflectance data using: 1 - reflectance times a black-body at 288 Kelvin.
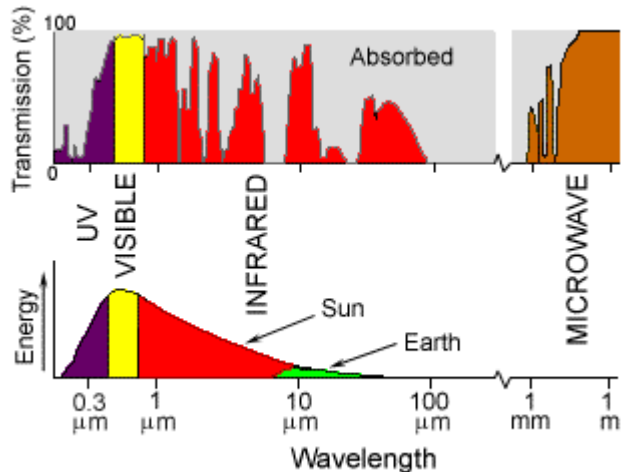
Figure: wavelengths we can use most effectively

2.4.3 Reflectance spectra

Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy by Roger N. Clark
- *http://speclab.cr.usgs.gov/PAPERS.refl-mrs/refl4.html#section4.3*

- *http://priede.bf.lu.lv/GIS/Descriptions/Remote_Sensing/An_Online_Handbook/Glossary/glossary.s html*

Albedo, is defined as the ratio of the amount of electromagnetic energy reflected by a surface to the amount of energy incident upon it. It differs from *spectral reflectance*, since usually *albedo* is averaged over the visible range of the Electro-Magnetic spectrum, while the term reflectance relates to a specific wavelength (or a specific band of a satellite).

The complex interaction of light with matter involves reflection and refraction from index of refraction boundaries, a process we call scattering, and absorption by the medium as light passes through the medium. The amount of scattering versus absorption controls the amount of photons we receive from a surface. As each material has its own unique interaction with light, we can find the unique reflectance spectra of each material. Absorption bands in the spectra of materials are caused by two general processes: electronic and vibrational (however, these processes will not be explained here). In addition, the reflectance spectra of a material is also dependent upon the mixture ratio of the material with other, and the material's grain size.

As an example for the above dicussion, are given below the reflectance spectra of some organic materials

25
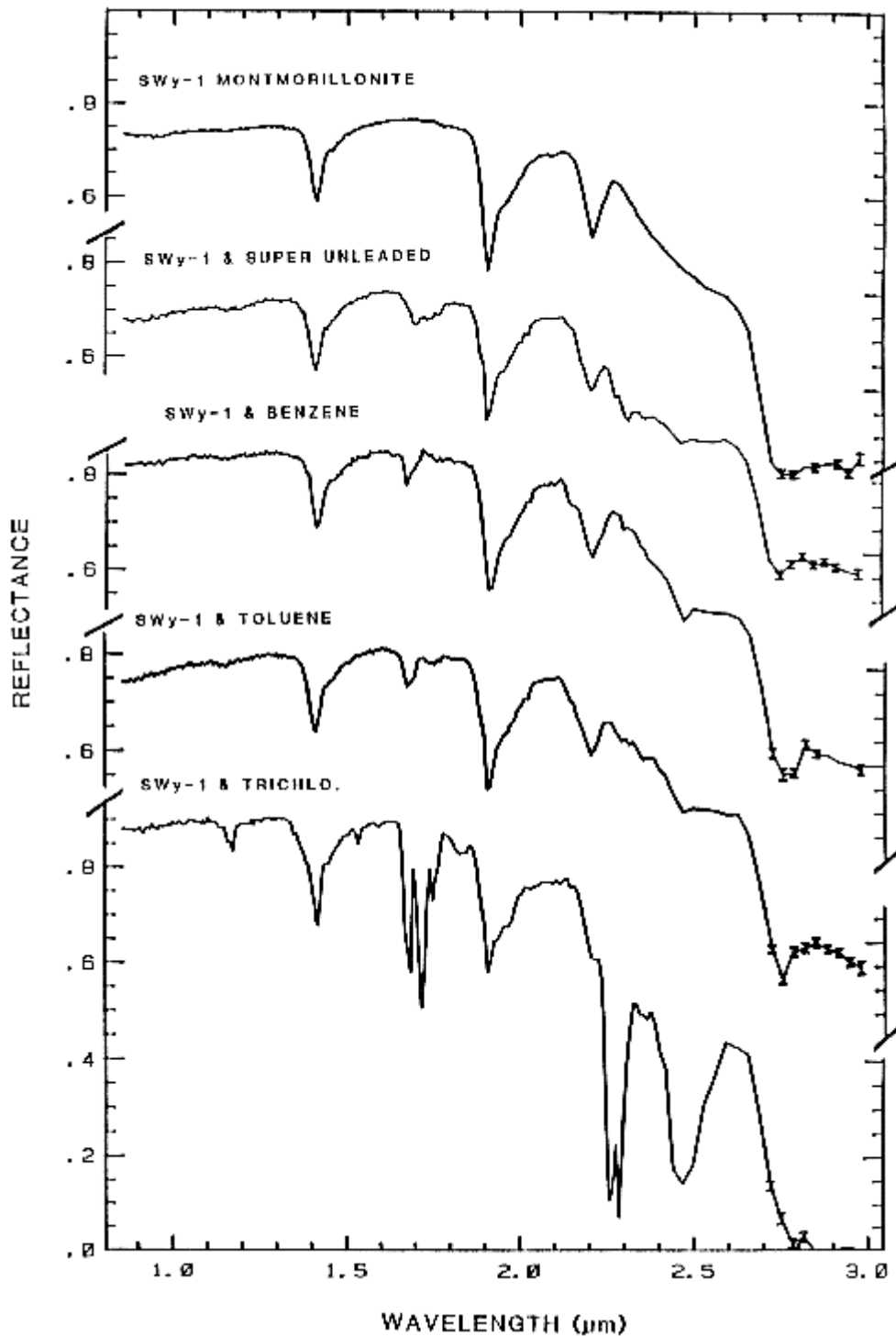
Figure: Reflectance spectra of montmorillonite, and montmorillonite mixed with super unleaded gasoline, benzene, toluene, and trichlorethylene. Montmorillonite has an absorption feature at 2.2 μm, whereas the organics have a CH combination band near 2.3 μm.

Spectra of vegetation come in two general forms: green and wet (photosynthetic), and dry non-photosynthetic but there is a seemingly continuous range between these two end members. The spectra of these two forms are compared to a soil spectrum in the next Figure. Because all plants are made of the same basic components, their spectra appear generally similar. However, in the spectral analysis section we will see methods for distinguishing subtle spectral details. The near-infrared spectra of green

26

vegetation are dominated by liquid water vibrational absorptions. The water bands are shifted to slightly shorter wavelengths than in liquid water, due to hydrogen bonding. The absorption in the visible is due to chlorophyll, and is discussed in more detail in Chapter XX of Manual of Remote Sensing. The dry non-photosynthetic vegetation spectrum shows absorptions due to cellulose, lignin, and nitrogen. Some of these absorptions can be confused with mineral absorptions, unless a careful spectral analysis is done.
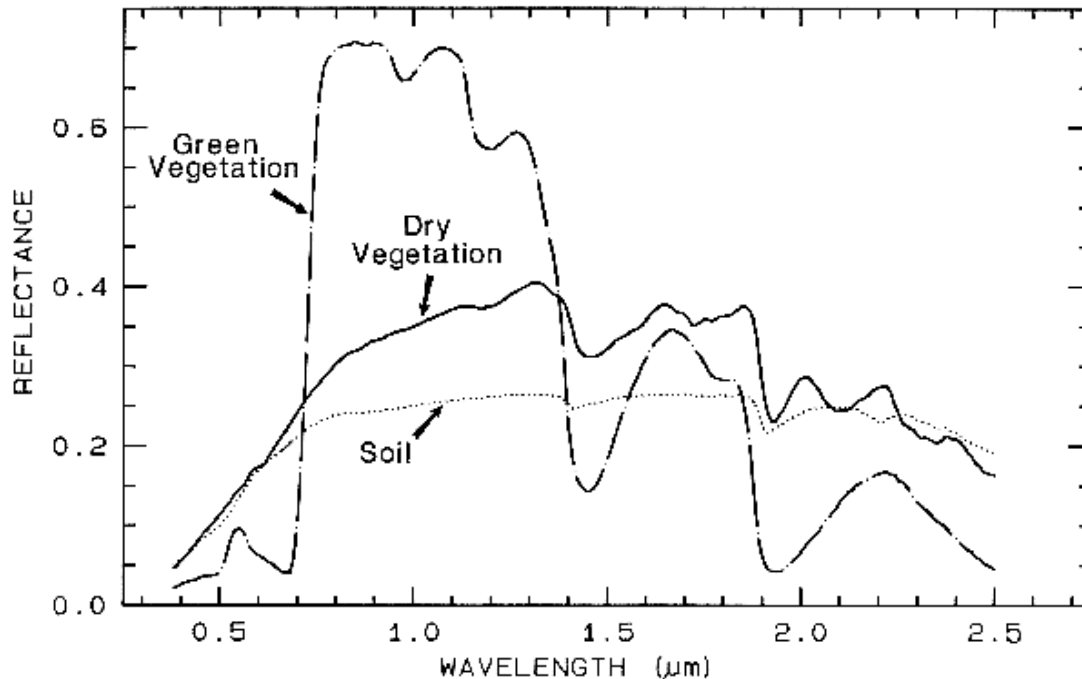
Figure : Reflectance spectra of photosynthetic (green) vegetation, non-photosynthetic (dry) vegetation, and a soil. The green vegetation has absorptions short of 1 μm due to chlorophyll. Those at wavelengths greater than 0.9 μm are dominated by liquid water. The dry vegetation shows absorptions dominated by cellulose, but also lignin and nitrogen. These absorptions must also be present in the green vegetation, but can be detected only weakly in the presence the stronger water bands. The soil spectrum shows a weak signature at 2.2 μm due to montmorillonite.

2.4.3.1 Mixtures

The real world (and for that matter, the universe) is a complex mixture of materials, at just about any scale we view it. In general, there are 4 types of mixtures:

1) *Linear Mixture*. The materials in the field of view are optically separated so there is no multiple scattering between components. The combined signal is simply the sum of the fractional area times the spectrum of each component. This is also called areal mixture.

2) *Intimate Mixture*. An intimate mixture occurs when different materials are in intimate contact in a scattering surface, such as the mineral grains in a soil or rock. Depending on the optical properties of each component, the resulting signal is a highly non-linear combination of the end-member spectra.

3) *Coatings.* Coatings occur when one material coats another. Each coating is a scattering/transmitting layer whose optical thickness varies with material properties and wavelength.

4) *Molecular Mixtures*. Molecular mixtures occur on a molecular level, such as two liquids, or a liquid and a solid mixed together. Examples: water adsorbed onto a mineral; gasoline spilled onto a soil. The close contact of the mixture components can

cause band shifts in the adsorbate, such as the interlayer water in montmorillonite, or the water in plants.

An example mixture comparison is shown in the next Figure for alunite and jarosite. Note in the intimate mixture how the jarosite dominates and the 0.4 to 1.3-μm region. The reason is because in an intimate mixture, the darker material dominates because photons are absorbed when they encounter a dark grain. In the areal mixture, the brighter material dominates.
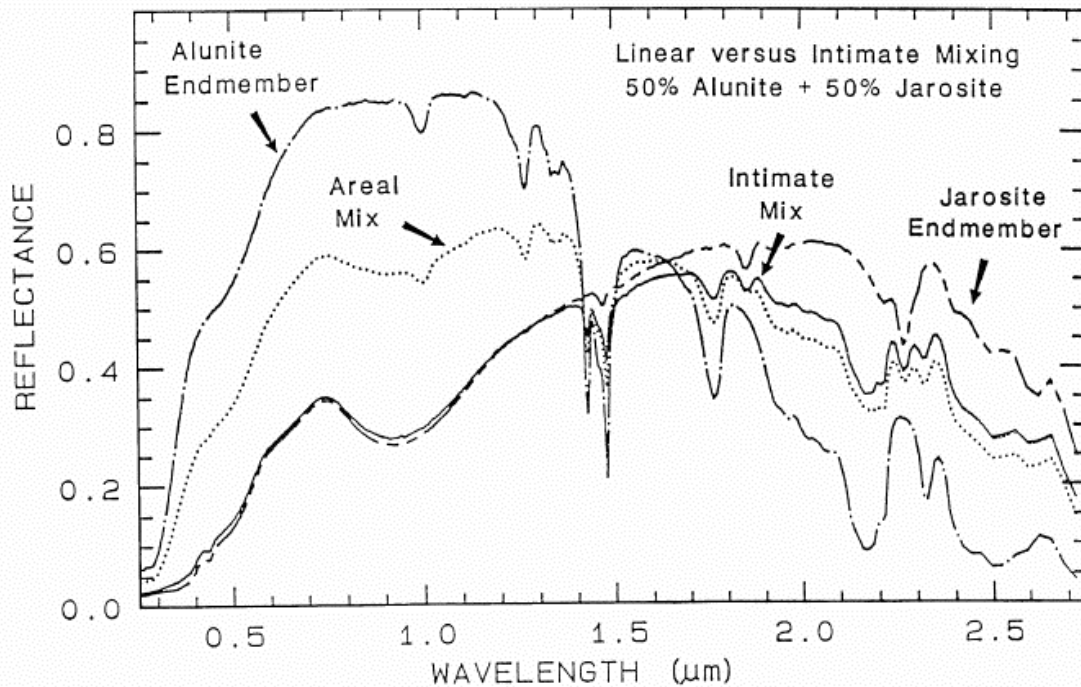
Figure. Reflectance spectra of alunite, jarosite and mixtures of the two. Two mixture types are shown: intimate and areal. In the intimate mixture, the darker of the two spectral components tends to dominate, and in and areal mixture, the brighter component dominates. The areal mixture is a strictly linear combination and was computed from the end-members, whereas the intimate mixture is non-linear and the spectrum of the physical mixture was measured in the laboratory. The jarosite dominates the 0.3-1.4 μm wavelength region in the intimate mixture because of the strong absorption in jarosite at those wavelengths and because the jarosite is finer grained than the alunite and tends to coat the larger alunite grains.

2.4.3.2 Grain Size Effects

The amount of light scattered and absorbed by a grain is dependent on grain (see figures of ice spectra for example). A larger grain has a greater internal path where photons may be absorbed according to Beers Law. It is the reflection from the surfaces and internal imperfections that control scattering. In a smaller grain there are proportionally more surface reflections compared to internal photon path lengths, or in other words, the surface-to-volume ratio is a function of grain size. If multiple scattering dominates, as is usually the case in the visible and near- infrared, the reflectance decreases as the grain size increases. However, in the mid-infrared, where absorption coefficients are much higher and the index of refraction varies strongly at the Christensen frequencies, first surface reflection is a larger or even dominant component of the scattered signal. In these cases, the grain size effects are much more complex, even reversing trends commonly seen at shorter wavelengths.
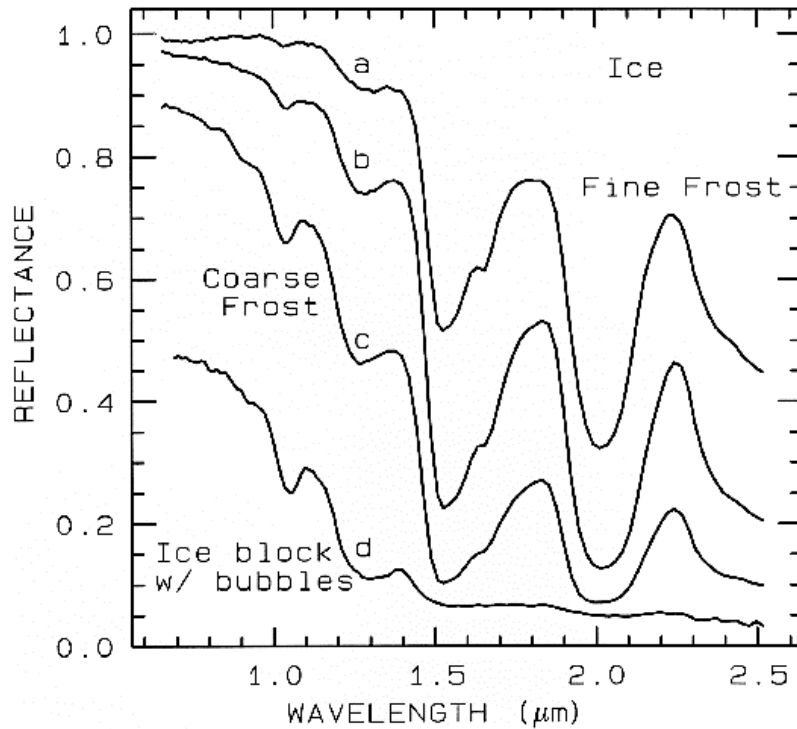
Figure 22a. The near-infrared spectral reflectance of A) a fine grained (~50 μm) water frost, B) medium grained (~200 μm) frost, C) coarse grained (400-2000 μm) frost and D) an ice block containing abundant microbubbles. The larger the effective grain size, the greater the mean photon path that photons travel in the ice, and the deeper the absorptions become. Curve D is very low in reflectance because of the large path length in ice. The ice temperatures for these spectra are 112-140 K. From Clark *et al*. (1986).
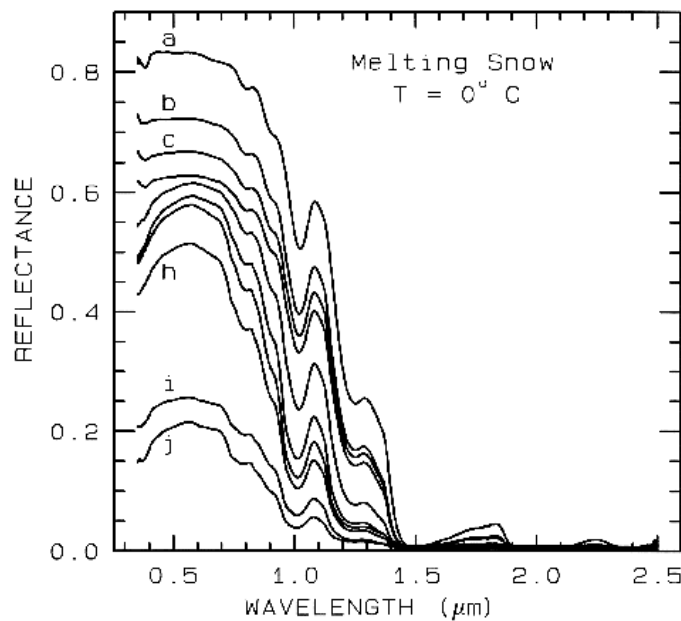
Figure 22b. A series of reflectance spectra of melting snow. The top curve (a) is at 0° C and has only a small amount of liquid water, whereas the lowest spectrum (j) is of a puddle of about 3 cm of water on top of the snow. Note in the top spectrum, there is no 1.65-μm band as in the ice spectra in figure 22a because of the higher temperature.. The 1.65-μm feature is temperature dependent and decreases in strength with increasing temperature (see Clark, 1981a and references therein). Note the increasing absorption at about 0.75 μm and in the short side of the 1-μm ice band, as more liquid water forms. The liquid water becomes spectrally detectable at about spectrum e, when the UV absorption increases. Spectra from Clark, King and Swayze, in preparation.

2.4.3.3 The Continuum and Band Depth.

Absorptions in a spectrum have two components: continuum and individual features. The continuum is the "background absorption" onto which other absorption features are superimposed (e.g. see Clark and Roush, 1984). It may be due to the wing of a larger absorption feature. For example, in the pyroxene spectra in Figure 21a, the weak feature at 2.3 μm is due to a trace amount of tremolite in the sample and the absorption is superimposed on the broader 2-μm pyroxene band. The broader pyroxene absorption is the continuum to the narrow 2.3-μm feature. The pyroxene 1.0-μm band is superimposed on the wing of a stronger absorption centered in the ultraviolet.

The depth of an absorption band, D, is usually defined relative to the continuum, R_c:

$$D = 1 - R_b / R_c \quad (2)$$

where R_b is the reflectance at the band bottom, and R_c is the reflectance of the continuum at the same wavelength as R_b (Clark and Roush, 1984).

The depth of an absorption is related to the abundance of the absorber and the grain size of the material. Consider a particulate surface with two minerals, one whose spectrum has an absorption band. As the abundance of the second mineral is increased, the band depth, D, of the absorption in the first mineral will decrease. Next consider the visual and near-infrared reflectance spectrum of a pure powdered mineral. As the grain size is increased from a small value, the absorption-band depth, D, will first increase, reach a maximum, and then decrease. This can be seen with the pyroxene spectra in Figure 21a and more so in the ice spectra in Figure 22. If the particle size were made larger and larger, the reflectance spectrum would eventually consist only of first surface reflection, like at most wavelengths beyond 1.45 μm in the ice spectra in Figure 22. The reflectance can never go to zero because of this reflection, unless the index of refraction of the material is 1.0. These concepts, called "band saturation" are explored further in Clark and Lucey (1984) and Lucey and Clark (1985).

A sloping continuum causes an apparent shift in the reflectance minimum, as shown in the above Figure. Continuaa can be thought of as an additive effect of optical constants, but in reflectance spectra, scattering and Beers Law make the effects non-linearly multiplicative (see Clark and Roush, 1984 for more details). So the continuum should be removed by division, whether you are working in reflectance or emittance. The continuum should be removed by subtraction only when working with absorption coefficients. In a spectrum with a sloping continuum, correction removes the effect of shifts in the local reflectance minimum. Note your perception of spectrum E versus A in the Figure. The spectral features do not appear to be the same, but if you remove the continuum, it is obvious they are the same (Figure, top).
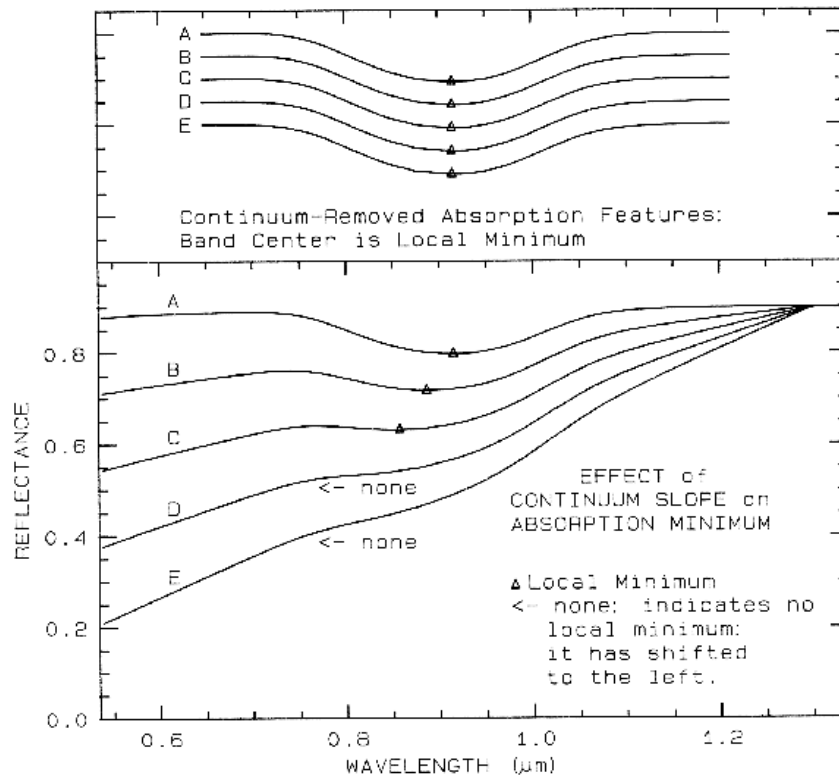
Figure: Illustration of continua and continuum removal. In the lower set of curves, the local minimum in the curve shifts to shorter wavelengths with increasing slope. removal of the continuum by division isolates the spectral features so they may be compared (top). The top set of curves are offset for clarity. In the continuum-removed spectra, we can see there is no real shift in the absorption-band center.

2.4.3.4 Continuum-Removed Spectral Feature Comparison.

The continuum-removal process isolates spectral features and puts them on a level playing field so they may be intercompared. Continuum removal and feature comparison, is in this author's opinion, the key to successful spectral identification. For example, compare the spectra of calcite ($CaCO_3$) and dolomite ($CaMg(CO_3)_2$). If we isolate the spectral features, remove the continuum, and scale the band depth (or band area) to be equal, we can see subtle band shifts and shapes:
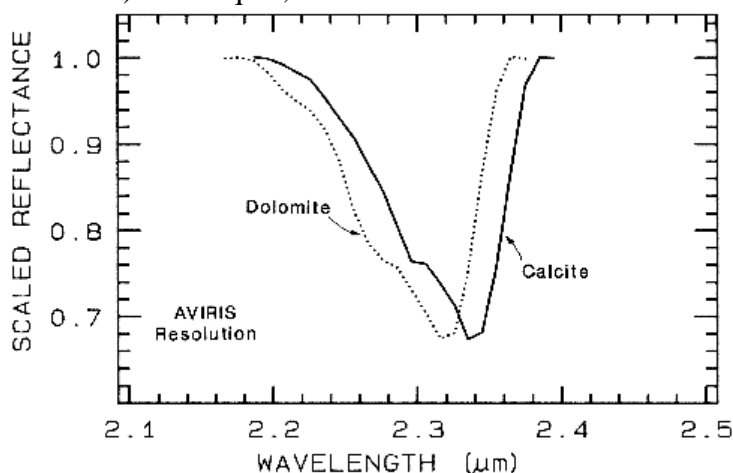
Figure 24a. Comparison of calcite and dolomite continuum-removed features. The dolomite absorption occurs at a shorter wavelength than the calcite absorption.

One of the most challenging spectral features to distinguish between are those in spectra of various plant species. Figure a shows four plant spectra (the spectra are

offset for clarity). The overall shapes are quite similar. If we remove the continuum according, we see the detailed chlorophyll absorption spectral variations for these as well as other plants in Figure b. Shape matching algorithms, like that presented in Clark *et al*. (1990b), can distinguish between these spectra and accurately compare them to spectral libraries.
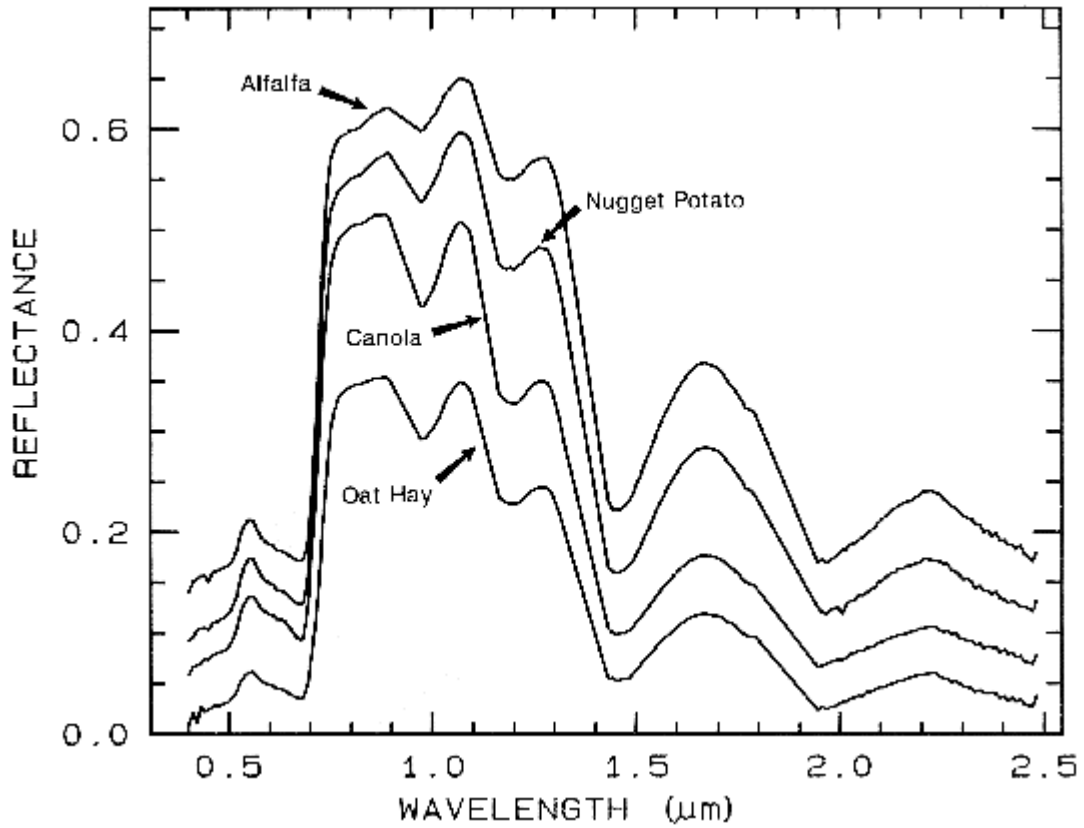
Figure a. Reflectance spectra of four types of vegetation. Each curve is offset by 0.05 reflectance unit from the one below. From Clark *et al*. (1995, 1997).
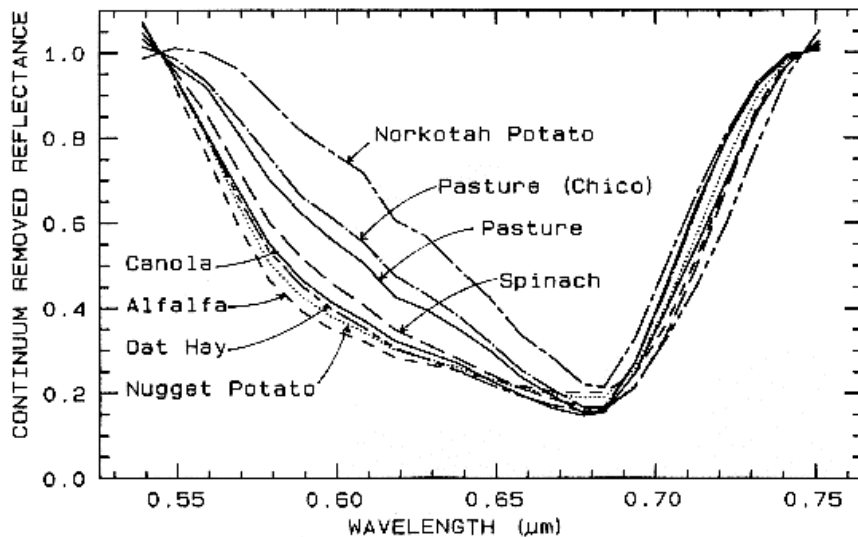
Figure b. Continuum-removed chlorophyll absorptions for 8 vegetation types (including the 4 from Figure a) showing that the continuum removed features can show subtle spectral differences. From Clark *et al*. (1995, 1997).

32

2.4.3.5 Viewing Geometry

We have seen tremendous variation in the spectral properties of minerals and materials in general, due to composition, grain size, and mixture types. So far viewing geometry has not been discussed. Viewing geometry, including the angle of incidence, angle of reflection, and the phase angle: the angle between the incident light and observer (the angle of reflection), all affect the intensity of light received. Varying the viewing geometry results in changes in shadowing and the proportions of first surface to multiple scattering (e.g. Hapke, 1993; Nelson, 1986; Mustard and Pieters, 1989), which can affect band depths a small amount except in rare cases (like extreme specular reflection off a mirror or lake surface). While measuring precise light levels are important for things like radiation balance studies, they are of lesser importance in spectral analysis. The following illustrates why.

First, your eye is a crude spectrometer, able to distinguish the spectral properties of materials in a limited wavelength range by the way we interpret color. Pick up any colored object around you. Change the orientation of the local normal on the surface of the object with respect to the angle of incident light, the angle at which you observe it (called the emission or scattering angle), and the angle between the incident and scattered light (the phase angle). As you do this, note any color changes. Unless you chose an unusual material (like a diffraction grating or very shiny object), you will see no significant color change. Plant leaves appear green from any angle, a pile of hematite appears red from any angle. This tells you that the spectral features do not change much with viewing geometry. Your eye/brain combination normalizes intensity variations so that you see the same color, regardless of the actual brightness of the object (the amount of light falling on the surface). The continuum removal does a similar but more sophisticated normalization. The band depth, shape, and position are basically constant with viewing geometry. Band depth will only change with the proportion of specular reflection added to the reflected light. For surfaces (and at wavelengths) where multiple scattering dominates, that change in band depth is minimized.

3. Sensors:

In order for a sensor to collect and record energy reflected or emitted from a target or surface, it must reside on a stable **platform** removed from the target or surface being observed. Platforms for remote sensors may be situated on the ground, on an aircraft or balloon (or some other platform within the Earth's atmosphere), or on a spacecraft or satellite outside of the Earth's atmosphere.

Ground-based sensors are often used to record detailed information about the surface which is compared with information collected from aircraft or satellite sensors. In some cases, this can be used to better characterize the target which is being imaged by these other sensors, making it possible to better understand the information in the imagery. Sensors may be placed on a ladder, scaffolding, tall building, cherry-picker, crane, etc. Aerial platforms are primarily stable wing aircraft, although helicopters are occasionally used. Aircraft are often used to collect very detailed images and facilitate the collection of data over virtually any portion of the Earth's surface at any time. In space, remote sensing is sometimes conducted from the space shuttle or, more commonly, from satellites. Satellites are objects which revolve around another object - in this case, the Earth. For example, the moon is a natural satellite, whereas man-made satellites include those platforms launched for remote sensing, communication, and telemetry (location and navigation) purposes. Because of their orbits, satellites permit repetitive coverage of the Earth's surface on a continuing basis. Cost is often a significant factor in choosing among the various platform options.

3.1 History

Since the early 1960s, numerous satellite sensors have been launched into orbit to observe and monitor the Earth and its environment. Most early satellite sensors acquired data for meteorological purposes. The advent of earth resources satellite sensors (those with a primary objective of mapping and monitoring land cover) occurred when the first Landsat satellite was launched in July 1972. Currently, more than a dozen orbiting satellites of various types provide data crucial to improving our knowledge of the Earth's atmosphere, oceans, ice and snow, and land.

3.2 Satellite Characteristics: Orbits and Swaths

We learned in the previous section that remote sensing instruments can be placed on a variety of platforms to view and image targets. Although ground-based and aircraft platforms may be used, satellites provide a great deal of the remote sensing imagery commonly used today. Satellites have several unique characteristics which make them particularly useful for remote sensing of the Earth's surface.

We can calculate the height of a satellite above the Earth, using the following laws:

Gravitation Force $F = GmM/R^2$
Centrifugal Force $F = v^2 * m / R$

Since the gravitation force is equal to the centrifugal force in an orbiting satellite, the gravitational force or mass of the planet is proportional to the velocity squared of any of it's orbiting satellites.

$$\text{Centrifugal Force} = \text{Gravitation Force}$$
$$v^2 m / R = GmM/R^2$$

to simplify the equation, we can take the orbit of the satellite to be on a circle, thus:

$$v = 2\rho R/T$$

with the following,

m   Mass of the satellite (is of no importance for the calculation)
M   Mass of the Earth, 5.976e+24 kg
R   Distance between the center of the Earth and the satellite (consider the equatorial radius of the Earth to be about 6,378.14 km)
v   Velocity of the satellite
T   Period of the orbit of the satellite around the Earth
G   Newtonian constant of gravitation , 6.6725985e-11 m^3 kg^{-1} s^{-2}

the distance of a satellite from the face of the Earth can be easily calculated:

$$R^3 = GMT^2/4\rho^2$$

The path followed by a satellite is referred to as its **orbit**. Satellite orbits are matched to the capability and objective of the sensor(s) they carry. Orbit selection can vary in terms of altitude (their height above the Earth's surface) and their orientation and rotation relative to the Earth. Satellites at very high altitudes, which view the same portion of the Earth's surface at all times have geostationary orbits. These geostationary satellites, at altitudes of approximately 36,000 kilometres, revolve at speeds which match the rotation of the Earth so they seem stationary, relative to the Earth's surface. This allows the satellites to observe and collect information continuously over specific areas. Weather and communications satellites commonly have these types of orbits. Due to their high altitude, some geostationary weather satellites can monitor weather and cloud patterns covering an entire hemisphere of the Earth.

Figure: Geostationary orbit   Near-polar orbit
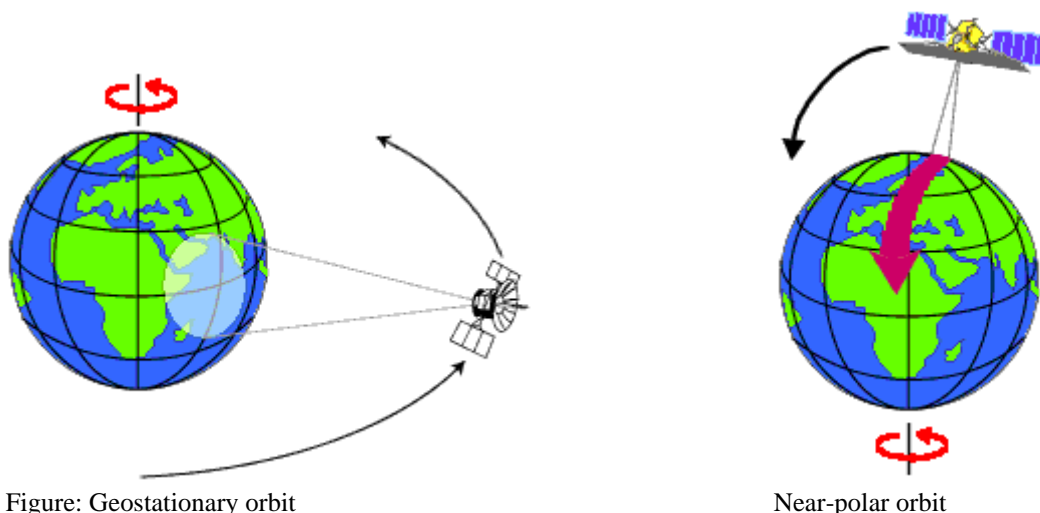
Many remote sensing platforms are designed to follow an orbit (basically north-south) which, in conjunction with the Earth's rotation (west-east), allows them to cover most of the Earth's surface over a certain period of time. These are near-polar orbits, so named for the inclination of the orbit relative to a line running between the North and South poles. Many of these satellite orbits are also **sun-synchronous** such that they cover each area of the world at a constant local time of day called **local sun time**. At any given latitude, the position of the sun in the sky as the satellite passes overhead will be the same within the same season. This ensures consistent illumination conditions when acquiring images in a specific season over successive years, or over a particular area over a series of days. This is an important factor for monitoring changes between images or for mosaicking adjacent images together, as they do not have to be corrected for different illumination conditions.

Most of the remote sensing satellite platforms today are in near-polar orbits, which means that the satellite travels northwards on one side of the Earth and then toward the southern pole on the second half of its orbit. These are called ascending and descending passes, respectively. If the orbit is also sun-synchronous, the ascending pass is most likely on the shadowed side of the Earth while the descending pass is on the sunlit side. Sensors recording reflected solar energy only image the surface on a descending pass, when solar illumination is available. Active sensors which provide their own illumination or passive sensors that record emitted (e.g. thermal) radiation can also image the surface on ascending passes.

As a satellite revolves around the Earth, the sensor "sees" a certain portion of the Earth's surface. The area imaged on the surface, is referred to as the swath. Imaging swaths for spaceborne sensors generally vary between tens and hundreds of kilometres wide. As the satellite orbits the Earth from pole to pole, its east-west position wouldn't change if the Earth didn't rotate. However, as seen from the Earth, it seems that the satellite is shifting westward because the Earth is rotating (from west to east) beneath it. This apparent movement allows the satellite swath to cover a new area with each consecutive pass. The satellite's orbit and the rotation of the Earth work together to allow complete coverage of the Earth's surface, after it has completed one complete cycle of orbits.
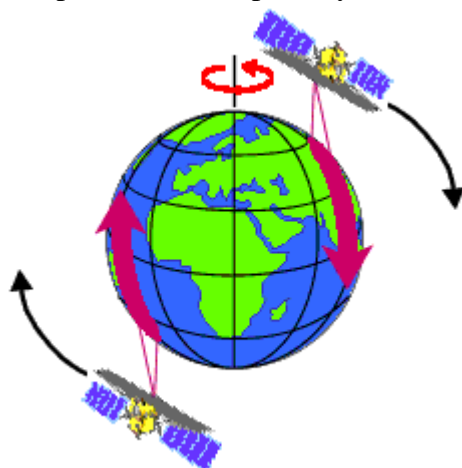
Figure: Ascending and descending passes   Swath
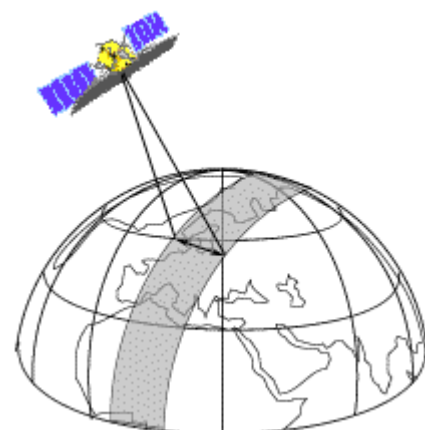
If we start with any randomly selected pass in a satellite's orbit, an orbit cycle will be completed when the satellite retraces its path, passing over the same point on the Earth's surface directly below the satellite (called the **nadir** point) for a second time. The exact length of time of the orbital cycle will vary with each satellite. The interval

of time required for the satellite to complete its orbit cycle is not the same as the "**revisit period**". Using steerable sensors, an satellite-borne instrument can view an area (off-nadir) before and after the orbit passes over a target, thus making the 'revisit' time less than the orbit cycle time. The revisit period is an important consideration for a number of monitoring applications, especially when frequent imaging is required (for example, to monitor the spread of an oil spill, or the extent of flooding). In near-polar orbits, areas at high latitudes will be imaged more frequently than the equatorial zone due to the increasing overlap in adjacent swaths as the orbit paths come closer together near the poles.

3.3 Scanner Sensor Systems

- *http://www.amesremote.com/section2.htm*

Electro-optical and spectral imaging scanners produce digital images with the use of detectors that measure the brightness of reflected electromagnetic energy. Scanners consist of one or more sensor detectors depending on type of sensor system used.

One type of scanner is called a whiskbroom scanner also referred to as across-track scanners (e.g. on LANDSAT). It uses rotating mirrors to scan the landscape below from side to side perpendicular to the direction of the sensor platform, like a whiskbroom. The width of the sweep is referred to as the sensor swath. The rotating mirrors redirect the reflected light to a point where a single or just a few sensor detectors are grouped together. Whiskbroom scanners with their moving mirrors tend to be large and complex to build. The moving mirrors create spatial distortions that must be corrected with preprocessing by the data provider before image data is delivered to the user. An advantage of whiskbroom scanners is that they have fewer sensor detectors to keep calibrated as compared to other types of sensors.

Another type of scanner, which does not use rotating mirrors, is the pushbroom scanner also referred to as an along-track scanner (e.g. on SPOT). The sensor detectors in a pushbroom scanner are lined up in a row called a linear array. Instead of sweeping from side to side as the sensor system moves forward, the one dimensional sensor array captures the entire scan line at once like a pushbroom would. Some recent scanners referred to as step stare scanners contain two-dimensional arrays in rows and columns for each band. Pushbroom scanners are lighter, smaller and less complex because of fewer moving parts than whiskbroom scanners. Also they have better radiometric and spatial resolution. A major disadvantage of pushbroom scanners is the calibration required for a large number of detectors that make up the sensor system.

(a) MULTI-SPECTRAL SCANNER (e.g., ATLAS) — WHISKBROOM SCANNER
(b) PUSHBROOM SCANNER — SWATH A IS FIRST IMAGED IN BAND A', AND LATER AS THE PLATFORM MOVES FORWARD IN BAND B'. NO SCANNING REQUIRED (e.g., SPOT)
(c) WHISKBROOM SCANNER — SINGLE PIXEL IS DISPERSED, SCANNING IS REQUIRED (e.g., AVIRIS)
(d) PUSHBROOM SCANNER — EACH CROSS-TRACK PIXEL IS DISPERSED, NO SCANNING REQUIRED (e.g., HSI)
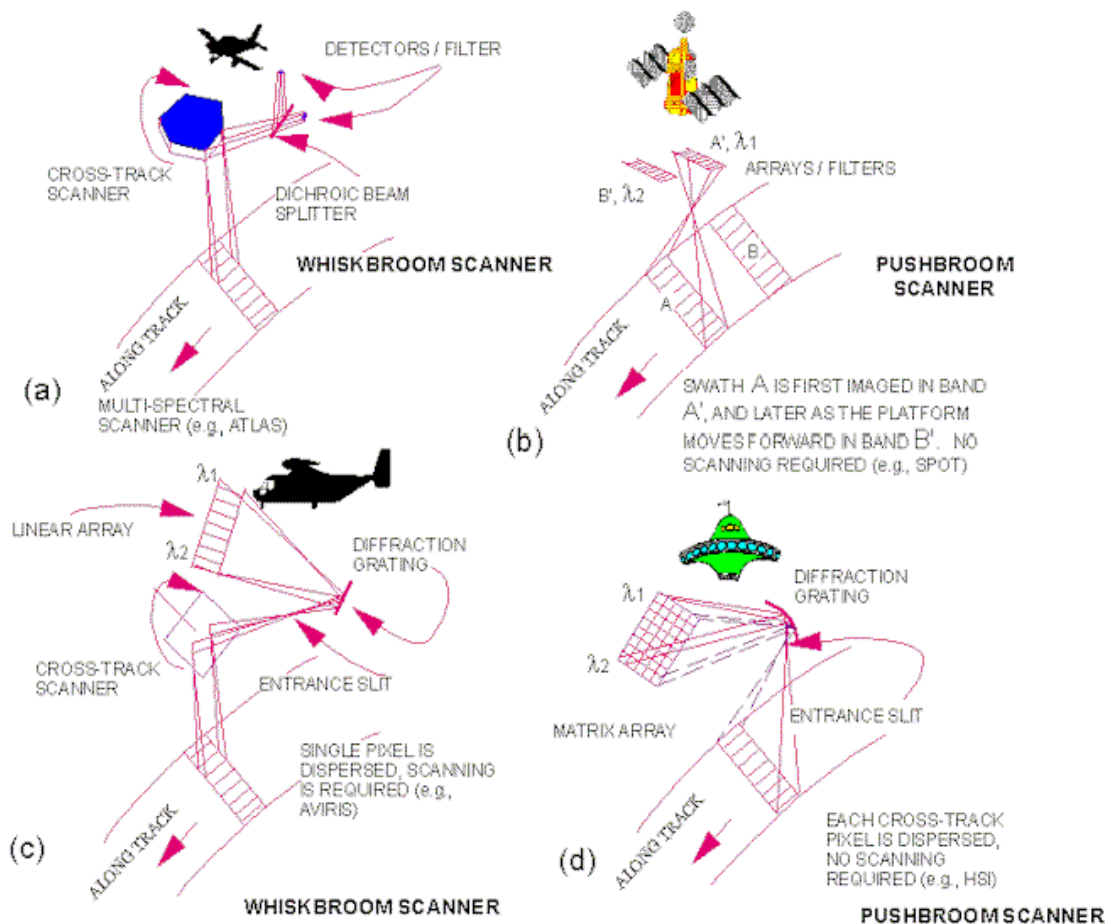
3.4 Spatial Resolution, Pixel Size, and Scale

- *http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/indexe.html*
- *http://satellite.rsat.com/rsat/*

For some remote sensing instruments, the distance between the target being imaged and the platform, plays a large role in determining the detail of information obtained and the total area imaged by the sensor. Sensors onboard platforms far away from their targets, typically view a larger area, but cannot provide great detail. Compare what an astronaut onboard the space shuttle sees of the Earth to what you can see from an airplane. The astronaut might see your whole province or country in one glance, but couldn't distinguish individual houses. Flying over a city or town, you would be able to see individual buildings and cars, but you would be viewing a much smaller area than the astronaut. There is a similar difference between satellite images and airphotos.

The detail discernible in an image is dependent on the **spatial resolution** of the sensor and refers to the size of the smallest possible feature that can be detected. Spatial resolution of passive sensors (we will look at the special case of active microwave sensors later) depends primarily on their Instantaneous Field of View (IFOV). The IFOV is the angular cone of visibility of the sensor (A) and determines the area on the Earth's surface which is "seen" from a given altitude at one particular moment in time (B). The size of the area viewed is determined by multiplying the IFOV by the distance from the ground to the sensor (C). This area on the ground is called the **resolution cell** and determines a sensor's maximum spatial resolution. For a homogeneous feature to be detected, its size generally has to be equal to or larger than

the resolution cell. If the feature is smaller than this, it may not be detectable as the average brightness of all features in that resolution cell will be recorded. However, smaller features may sometimes be detectable if their reflectance dominates within a articular resolution cell allowing sub-pixel or resolution cell detection.
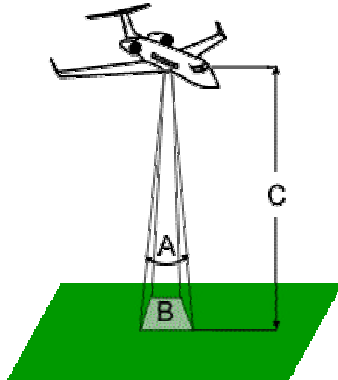
Figure: Instanteneous Field Of View

Most remote sensing images are composed of a matrix of picture elements, or **pixels**, which are the smallest units of an image. Image pixels are normally square and represent a certain area on an image. It is important to distinguish between pixel size and spatial resolution - they are not interchangeable. If a sensor has a spatial resolution of 20 metres and an image from that sensor is displayed at full resolution, each pixel represents an area of 20m x 20m on the ground. In this case the pixel size and resolution are the same. However, it is possible to display an image with a pixel size different than the resolution. Many posters of satellite images of the Earth have their pixels averaged to represent larger areas, although the original spatial resolution of the sensor that collected the imagery remains the same.

A photograph can be represented and displayed in a **digital** format by subdividing the image into small equal-sized and shaped areas, called picture elements or pixels, and representing the brightness of each area with a numeric value or **digital number**.
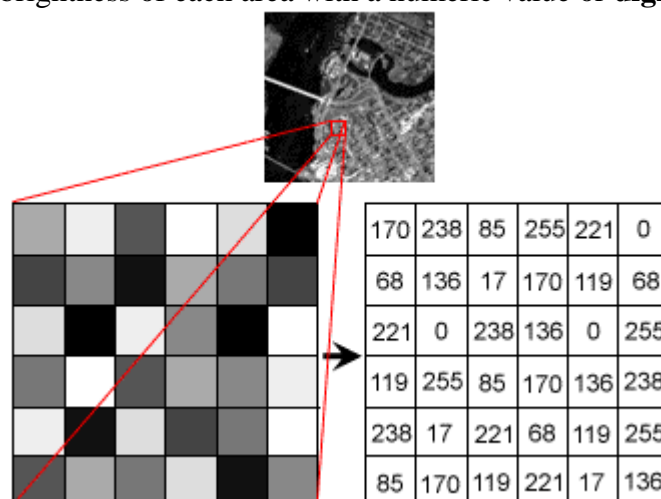
SATELLITE IMAGERY
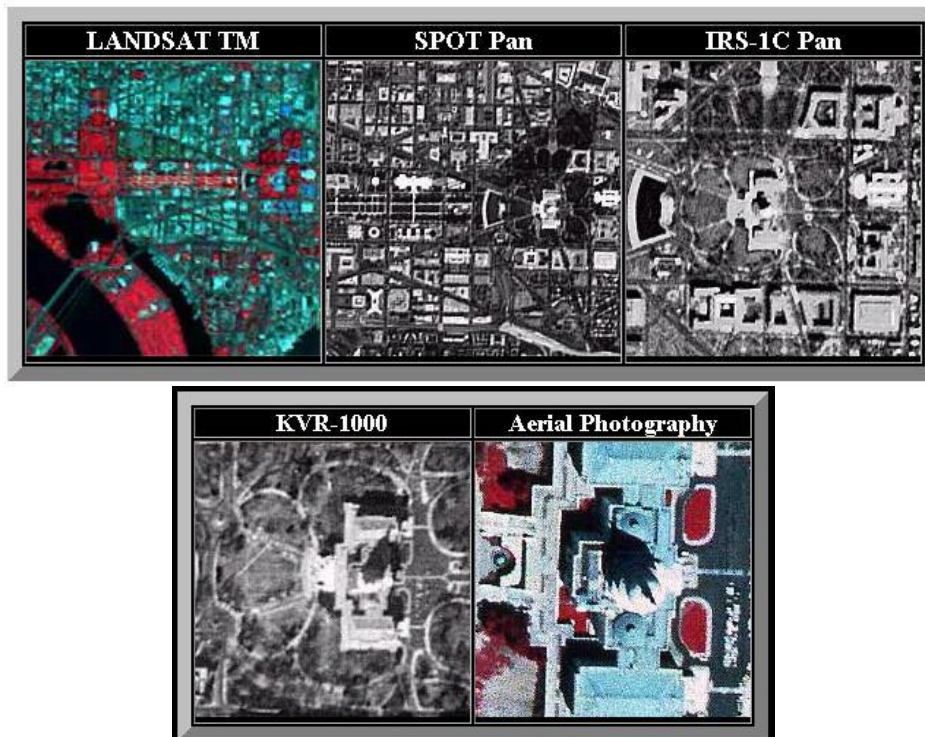
Images where only large features are visible are said to have <u>coarse or low resolution.</u> In <u>fine or high resolution</u> images, small objects can be detected. Military sensors for example, are designed to view as much detail as possible, and therefore have very fine resolution. Commercial satellites provide imagery with resolutions varying from a few metres to several kilometres. Generally speaking, the finer the resolution, the less total ground area can be seen.

The ratio of distance on an image or map, to actual ground distance is referred to as scale. If you had a map with a scale of 1:100,000, an object of 1cm length on the map would actually be an object 100,000cm (1km) long on the ground. Maps or images with small "map-to-ground ratios" are referred to as small scale (e.g. 1:100,000), and those with larger ratios (e.g. 1:5,000) are called large scale.

3.5 Spectral / Radiometric resolution

- *http://speclab.cr.usgs.gov/PAPERS.refl-mrs/refl4.html*
- *http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/indexe.html*
- *Mosher David C. and Simpkin Peter G. (), Status and Trends of Marin e High-Resolution Seismic Reflection Profiling: Data Acquisition*

3.5.1 Spectral charaecteristics

While the arrangement of pixels describes the spatial structure of an image, the radiometric characteristics describe the actual information content in an image. Every time an image is acquired on film or by a sensor, its sensitivity to the magnitude of the electromagnetic energy determines the **radiometric resolution**. The radiometric resolution of an imaging system describes its ability to discriminate very slight differences in energy The finer the radiometric resolution of a sensor, the more sensitive it is to detecting small differences in reflected or emitted energy.

Digital resolution is the number of bits comprising each digital sample. Imagery data are represented by positive digital numbers which vary from 0 to (one less than) a selected power of 2. This range corresponds to the number of bits used for coding numbers in binary format. Each bit records an exponent of power 2 (e.g. 1 bit = 2^1 = 2). The maximum number of brightness levels available depends on the number of bits used in representing the energy recorded. Thus, if a sensor used 8 bits to record the data, there would be 2^8 = 256 digital values available, ranging from 0 to 255 – termed also as the *dynamic range* of the system. However, if only 4 bits were used, then only 2^4 = 16 values ranging from 0 to 15 would be available. Thus, the radiometric resolution would be much less. Image data are generally displayed in a range of grey tones, with black representing a digital number of 0 and white representing the maximum value (for example, 255 in 8-bit data). By comparing a 2-bit image with an 8-bit image, we can see that there is a large difference in the level of detail discernible depending on their radiometric resolutions.

The range of energy values expected from a system must "fit" within the range of values possible of the data format type, and yet the value must represent accurately the energy value of the signal relative to others. The cost of more bits per data point is longer acquisition times, the requirement of larger storage capacity and longer processing time. Any signal outside the range is "clipped" and thus unrecoverable. On the other hand, if the dynamic range of the signal is widened too much as to allow the recording of extremely high or low energy values, the true variability within the signal will be lost.

Many remote sensing systems record energy over several separate wavelength ranges at various spectral resolutions. These are referred to as **multi-spectral sensors** and will be described in some detail in following sections. Advanced multi-spectral sensors called **hyperspectral** sensors, detect hundreds of very narrow spectral bands throughout the visible, near-infrared, and mid-infrared portions of the electromagnetic spectrum. Their very high spectral resolution facilitates fine discrimination between different targets based on their spectral response in each of the narrow bands.

There are 4 general parameters that describe the capability of a spectrometer: 1) spectral range, 2) spectral bandwidth, 3) spectral sampling, and 4) signal-to-noise ratio (S/N).

3.5.1.1 Spectral range

Spectral range is important to cover enough diagnostic spectral absorptions to solve a desired problem. There are general spectral ranges that are in common use, each to first order controlled by detector technology: a) ultraviolet (UV): 0.001 to 0.4 µm, b) visible: 0.4 to 0.7 µm, c) near-infrared (NIR): 0.7 to 3.0 µm, d) the mid-infrared (MIR): 3.0 to 30 µm, and d) the far infrared (FIR): 30 µm to 1 mm (e.g. see The Photonics Design and Applications Handbook, 1996 and The Handbook of Chemistry and Physics, any recent year). The ~0.4 to 1.0-µm wavelength range is sometimes referred to in the remote sensing literature as the VNIR (visible-near-infrared) and the 1.0 to 2.5-µm range is sometimes referred to as the SWIR (short-wave infrared). It should be noted that these terms are not recognized standard terms in other fields except remote sensing, and because the NIR in VNIR conflicts with the accepted NIR range, the VNIR and SWIR terms probably should be avoided. The mid-infrared

covers thermally emitted energy, which for the Earth starts at about 2.5 to 3 μm, peaking near 10 μm, decreasing beyond the peak, with a shape controlled by grey-body emission.

3.5.1.2 Spectral bandwidth

Spectral bandwidth is the width of an individual spectral channel in the spectrometer. The narrower the spectral bandwidth, the narrower the absorption feature the spectrometer will accurately measure, if enough adjacent spectral samples are obtained. Some systems have a few broad channels, not contiguously spaced and, thus, are not considered spectrometers (Figure 1a). Examples include the Landsat Thematic Mapper (TM) system and the MODerate Resolution Imaging Spectroradiometer (MODIS), which can't resolve narrow absorption features. Others, like the NASA JPL Airborne Visual and Infra-Red Imaging Spectrometer (AVIRIS) system have many narrow bandwidths, contiguously spaced (Figure 1b). Figure 1 shows spectra for the mineral alunite that could be obtained by some example broadband and spectrometer systems. Note the loss in subtle spectral detail in the lower resolution systems compared to the laboratory spectrum. Bandwidths and sampling greater than 25 nm rapidly lose the ability to resolve important mineral absorption features. All the spectra in Figure 1b are sampled at half Nyquist (critical sampling) except the Near Infrared Mapping Spectrometer (NIMS), which is at Nyquist sampling (named after H. Nyquist, that in his work published in 1928, stated that there must be at least to samplings per wavelength of the highest frequency, in order to appropriately sample the waveform). Note, however, that the fine details of the absorption features are lost at the ~25 nm bandpass of NIMS. For example, the shoulder in the 2.2-μm absorption band is lost at 25-nm bandpass. The Visual and Infrared Mapping Spectrometer (VIMS) and NIMS systems measure out to 5 μm, thus can see absorption bands not obtainable by the other systems.

Alunite as seen by three systems

REFLECTANCE + OFFSET

~65-250 nm — TM

10-50 nm — MODIS

5-10 nm
FWHM — LAB

WAVELENGTH (μm)

Alunite as seen by various imaging spectrometers

REFLECTANCE + OFFSET

(N) 25 nm

12 nm

19-12 nm — NIMS

10 nm — VIMS

5-10 nm — HYDICE

FWHM — AVIRIS

LAB

WAVELENGTH (μm)

43
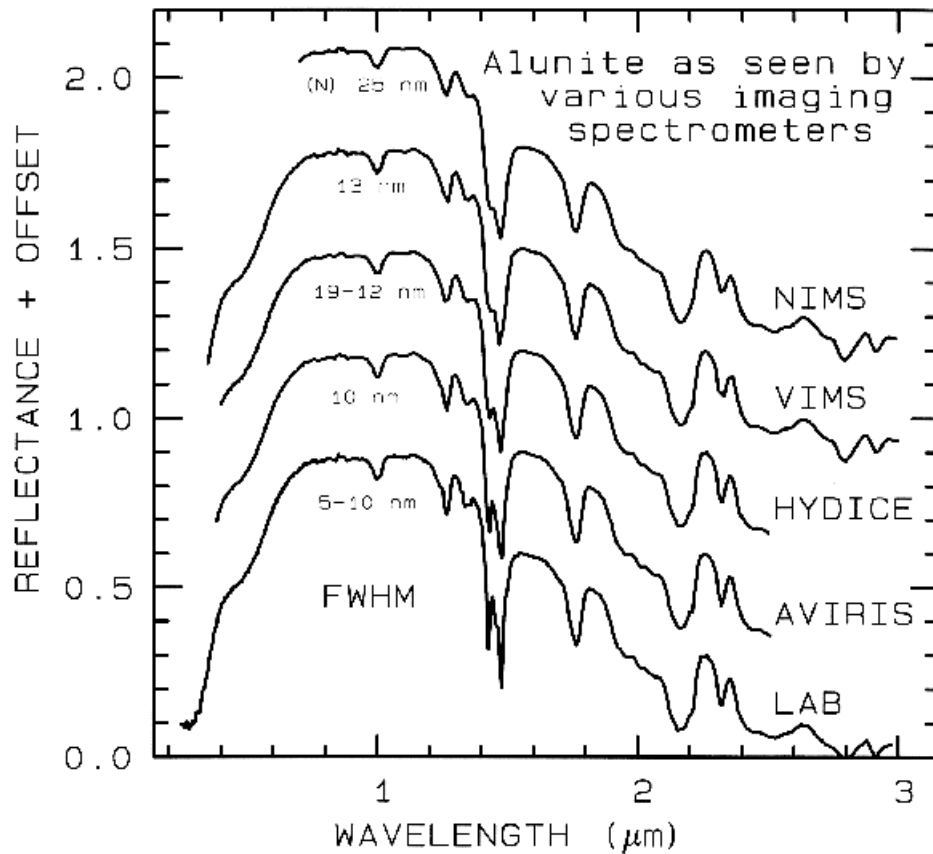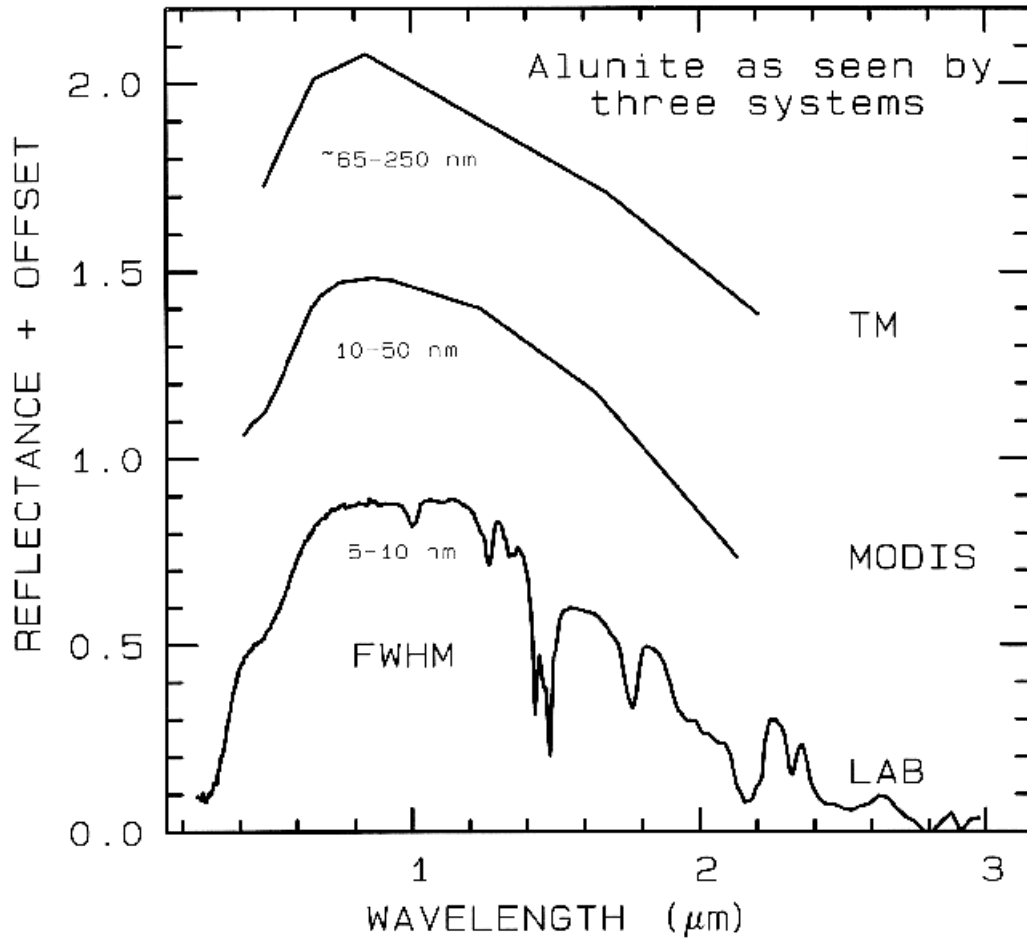
The shape of the bandpass profile is also important. Ideally each spectrometer channel rejects all light except that from within a given narrow wavelength range, but occasionally, due to optical effects too complex to discuss in detail here, light may leak in from out of the bandpass (e.g. scattering within the optical system, or inadequate blocking filters). The most common bandpass in spectrometers is a Gaussian profile. While specific spectrometer designs may have well-defined theoretical bandpass profiles, aberrations in the optical system usually smears the profile closer to a Gaussian shape. The width of the bandpass is usually defined as the width in wavelength at the 50% response level of the function, as shown in the next Figure, called the Full Width at Half Maximum (FWHM).
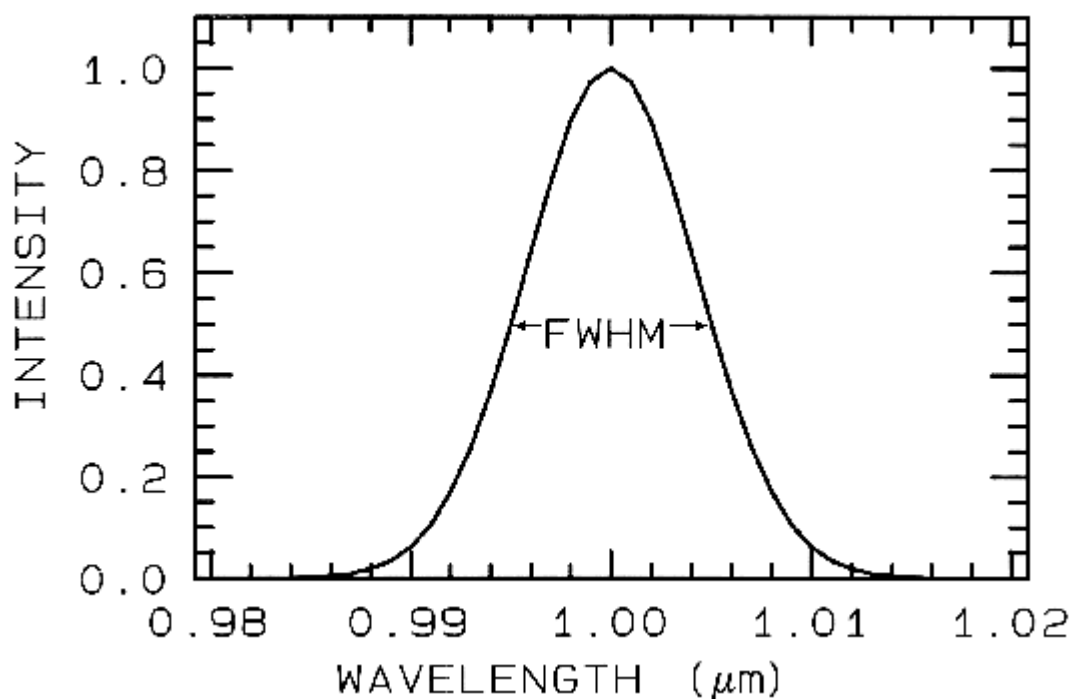
Figure: A Gaussian profile with a Full Width at Half Maximum (FWHM) of 10 nm is shown. This profile is typical of spectrometers such as AVIRIS which has 224 such profiles spaced at about 10 nm.

3.5.1.3 Spectral sampling

Spectral sampling is the distance in wavelength between the spectral bandpass profiles for each channel in the spectrometer as a function of wavelength. Spectral sampling is often confused with bandpass, with the two lumped together and called resolution. Information theory tells us that to resolve a two spectral features, we must have two samples. Further, in order to not introduce sampling bias, the samples must be close enough together to measure the peak and valley locations. The Nyquist theorem states that the maximum information is obtained by sampling at one-half the FWHM. Spectrometer design, however, sometimes dictates a different sampling, and many modern spectrometers in use (e.g. AVIRIS, VIMS) sample at half-Nyquist: sampling interval approximately equal to the FWHM. Note that the AVIRIS system has a bandpass ~0.01 μm (10 nm), a sampling of ~0.01 μm, and thus has a spectral resolution of ~0.02 μm (20 nm). The NIMS system in the Figure above can sample at Nyquist (shown), half-Nyquist, and lower.

3.5.1.4 Signal-to-noise ratio

Finally, a spectrometer must measure the spectrum with enough precision to record details in the spectrum. The signal-to-noise ratio (S/N) required to solve a particular problem will depend on the strength of the spectral features under study. The S/N is dependant on the detector sensitivity, the spectral bandwidth, and intensity of the light reflected or emitted from the surface being measured. A few spectral features are quite strong and a signal to noise of only about 10 will be adequate to identify them, while others are weak, and a S/N of several hundred (and higher) are often needed (Swayze *et al.*, 1997).

3.6 Temporal Resolution

In addition to spatial, spectral, and radiometric resolution, the concept of **temporal resolution** is also important to consider in a remote sensing system. The revisit period of a satellite sensor is usually several days. Therefore the absolute temporal resolution of a remote sensing system to image the exact same area at the same viewing angle a second time is equal to this period. However, the actual temporal resolution of a sensor depends on a variety of factors, including the satellite/sensor capabilities, the swath overlap, and latitude.

The ability to collect imagery of the same area of the Earth's surface at different periods of time is one of the most important elements for applying remote sensing data. Spectral characteristics of features may change over time and these changes can be detected by collecting and comparing **multi-temporal** imagery. For example, during the growing season, most species of vegetation are in a continual state of change and our ability to monitor those subtle changes using remote sensing is dependent on when and how frequently we collect imagery. By imaging on a continuing basis at different times we are able to monitor the changes that take place on the Earth's surface, whether they are naturally occurring (such as changes in natural vegetation cover or flooding) or induced by humans (such as urban development or deforestation). The time factor in imaging is important when:

- persistent clouds offer limited clear views of the Earth's surface (often in the tropics)
- short-lived phenomena (floods, oil slicks, etc.) need to be imaged
- multi-temporal comparisons are required (e.g. the spread of a forest disease from one year to the next)
- the changing appearance of a feature over time can be used to distinguish it from near-similar features (wheat / maize)

3.7 Overview of different sensors – satellites and airborne

3.7.1 Comparison Table

- *http://edcwww.cr.usgs.gov/glis/hyper/guide/*
- *http://ltpwww.gsfc.nasa.gov/LANDSAT/CAMPAIGN_DOCS/PROJECT/Comparison.html*
- *http://www.tbs-satellite.com/tse/online/sat_irs_1c.html*
- *http://www.spaceimaging.com/aboutus/satellites/IKONOS/ikonos.html*
- *http://www.spotimage.fr/home/system/introsat/seltec/welcome.htm*

Satellite Sensor	Bands and wavelength (µm)	Spatial Resolution	Swath width	Repeat coverage	Orbit altitude (km)
NOAA	1 (0.58-0.68)	1.1 km	2399	daily	833
	2 (0.725-1.10)	1.1 km	2399	daily	833
	3 (3.55-3.93)	1.1 km	2399	daily	833
	4 (10.3-11.3)	1.1 km	2399	daily	833
	5 (11.5-12.5)	1.1 km	2399	daily	833
MSS 4-5	1 (0.5-0.6)	79/82m	185	16 days	705
	2 (0.6-0.7)	79/82m	185	16 days	705
	3 (0.7-0.8)	79/82m	185	16 days	705
	4 (0.8-1.1)	79/82m	185	16 days	705
TM 4-5	1 (0.45-0.52)	30m	185	16 days	705
	2 (0.52-0.60)	30m	185	16 days	705
	3 (0.63-0.69)	30m	185	16 days	705
	4 (0.76-0.90)	30m	185	16 days	705
	5 (1.55-1.75)	30m	185	16 days	705
	6 (10.40-12.50)	120m	185	16 days	705
	7 (2.08-2.35)	30m	185	16 days	705
ETM 7	1 (0.45-0.515)	30m	183*170	16 days	705
	2 (0.525-0.605)	30m	183*170	16 days	705
	3 (0.63-0.69)	30m	183*170	16 days	705
	4 (0.75-0.90)	30m	183*170	16 days	705
	5 (1.55-1.75)	30m	183*170	16 days	705
	6 (10.40-12.5)	60m	183*170	16 days	705
	7 (2.09-2.35)	30m	183*170	16 days	705
	PAN (0.52-0.90)	15m	183*170	16 days	705
SPOT 4	XS 1 (0.50-0.59)	20m	60 (oblique scene at max 60 by 81)	26 days	822
	XS 2 (0.61-0.68)	20m	60 (oblique scene at max 60 by 81)	26 days	822
	XS 3 (0.79-0.89)	20m	60 (oblique scene at max 60 by 81)	26 days	822
	XS 4 (1.58-1.75)	20m	60 (oblique scene at max 60 by 81)	26 days	822
	Monospectral red (0.61-0.68)	10m	60 (oblique scene at max 60 by 81)	26 days	822
IRS 1C	LISS 1 (0.52-0.59)	23.6m	142	24 days	818
	LISS 2 (0.62-0.68)	23.6m	142	24 days	818
	LISS 3 (0.77-0.86)	23.6m	142	24 days	818
	LISS 4 (1.55-1.70)	70.8m	148	24 days	818
	WIFS 1 (0.62-0.68)	189m	810	24 days	818
	WIFS 2 (0.77-0.86)	189m	810	24 days	818
	PAN (0.5-0.75)	5.8m	70	24 days	818

Ikonos	Multispectral (0.45-0.52)	4m	13 at nadir	2.9 days 1m resolution	681
	Multispectral (0.52-0.60)	4m	13 at nadir	2.9 days 1m resolution	681
	Multispectral (0.63-0.69)	4m	13 at nadir	2.9 days 1m resolution	681
	Multispectral (0.76-0.90)	4m	13 at nadir	2.9 days 1m resolution	681
	Panchromatic (0.45-0.90)	1m	13 at nadir	2.9 days 1m resolution	681

3.7.1.1 Weather Satellites/Sensors

▪ *http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/indexe.html*

Weather monitoring and forecasting was one of the first civilian (as opposed to military) applications of satellite remote sensing, dating back to the first true weather satellite, TIROS-1 (Television and Infrared Observation Satellite - 1), launched in 1960 by the United States. Several other weather satellites were launched over the next five years, in near-polar orbits, providing repetitive coverage of global weather patterns. In 1966, NASA (the U.S. National Aeronautics and Space Administration) launched the geostationary Applications Technology Satellite (ATS-1) which provided hemispheric images of the Earth's surface and cloud cover every half hour. For the first time, the development and movement of weather systems could be routinely monitored. Today, several countries operate weather, or meteorological satellites to monitor weather conditions around the globe. Generally speaking, these satellites use sensors which have fairly coarse spatial resolution (when compared to systems for observing land) and provide large areal coverage. Their temporal resolutions are generally quite high, providing frequent observations of the Earth's surface, atmospheric moisture, and cloud cover, which allows for near-continuous monitoring of global weather conditions, and hence - forecasting. Here we review a few of the representative satellites/sensors used for meteorological applications.

3.7.1.1.1 GOES

The **GOES** (Geostationary Operational Environmental Satellite) System is the follow-up to the ATS series. They were designed by NASA for the National Oceanic and Atmospheric Administration (NOAA) to provide the United States National Weather Service with frequent, small-scale imaging of the Earth's surface and cloud cover. The GOES series of satellites have been used extensively by meteorologists for weather monitoring and forecasting for over 20 years. These satellites are part of a global network of meteorological satellites spaced at approximately 70° longitude intervals around the Earth in order to provide near-global coverage. Two GOES satellites, placed in **geostationary** orbits 36000 km above the equator, each view approximately one-third of the Earth. One is situated at 75°W longitude and monitors North and South America and most of the Atlantic Ocean. The other is situated at 135°W longitude and monitors North America and the Pacific Ocean basin. Together they cover from 20°W to 165°E longitude.

Two generations of GOES satellites have been launched, each measuring emitted and reflected radiation from which atmospheric temperature, winds, moisture, and cloud cover can be derived. The first generation of satellites consisted of GOES-1 (launched 1975) through GOES-7 (launched 1992). Due to their design, these satellites were capable of viewing the Earth only a small percentage of the time (approximately five per cent). The second generation of satellites began with GOES-8 (launched 1994) and has numerous technological improvements over the first series. They provide

near-continuous observation of the Earth allowing more frequent imaging (as often as every 15 minutes). This increase in temporal resolution coupled with improvements in the spatial and radiometric resolution of the sensors provides timelier information and improved data quality for forecasting meteorological conditions.

GOES-8 and the other second generation GOES satellites have separate **imaging** and **sounding** instruments. The **imager** has five channels sensing visible (1 km resolution) and infrared reflected and emitted solar radiation (4 km resolution). The infrared capability allows for day and night imaging. Sensor pointing and scan selection capability enable imaging of an entire hemisphere, or small-scale imaging of selected areas. The latter allows meteorologists to monitor specific weather trouble spots to assist in improved short-term forecasting. The imager data are 10-bit radiometric resolution, and can be transmitted directly to local user terminals on the Earth's surface.

The 19 channel **sounder** measures emitted radiation in 18 thermal infrared bands and reflected radiation in one visible band. These data have a spatial resolution of 8 km and 13-bit radiometric resolution. Sounder data are used for surface and cloud-top temperatures, multi-level moisture profiling in the atmosphere, and ozone distribution analysis.

3.7.1.1.2 National Oceanic and Atmospheric Administration's Advanced Very High Resolution Radiometer

▪ *http://www.ciesin.org/TG/RS/noaaavhr.html*

The Advanced Very High Resolution Radiometer (AVHRR) sensor is carried on-board the National Oceanic and Atmospheric Administration's (NOAA) series of Polar-Orbiting Operational Environmental Satellites (POES). A prototype AVHRR sensor as developed and first launched in October 1978 to acquire meteorological data, including day and night cloud mapping, surface water delineation, and sea surface temperatures. The next AVHRR sensor, launched on the NOAA-6 satellite in June 1979, included a redefined visible spectral band. The Figure of Observation Characteristics illustrates that the redefined band eliminates an overlap with a near-infrared band and corresponds more closely with the red absorption region of vegetation. This development heralded the additional use of the AVHRR as a vegetation mapping and analysis tool by enabling the acquisition of data suitable for use in a computed Normalized Difference Vegetation Index (NDVI).

AVHRR data are acquired with a wide-field scanning system that enables global coverage on a daily basis with a ground resolution of 1.1 km at nadir (directly beneath the satellite). The sensor also provides a data stream of nominal 4-km resolution that is achieved by sampling and averaging the full resolution 1.1-km data on-board the satellite.

3.7.1.2 Land Observation Satellites/Sensors

3.7.1.2.1 Landsat Multispectral Scanner

▪ *http://www.ciesin.org/TG/RS/landmss.html*

The Multispectral Scanner (MSS), launched on-board Landsat 1 in July 1972, was the world's first Earth observation satellite sensor. The MSS provided four spectral bands of Earth surface reflectance in the visible and near-infrared regions of the electromagnetic spectrum at a nominal 80-m spatial resolution, as illustrated by the Figure of Observation Characteristics. The U.S. Geological Survey's Global Land Information System (GLIS) provides information on the sensor, and acquisition and availability of Multispectral Scanner Landsat Data.

Landsat MSS data, collected by Landsats 1 through 5 from 1972 to 1993, provide the longest and most extensive archive of satellite image data for monitoring the global land surface. Efforts are ongoing to reconfigure selected portions of the historical global Landsat data archive into the Landsat Pathfinder Data Set to make it more readily useful for change detection.

In February 1993, the Earth Observation Satellite Company (EOSAT), commercial operator of Landsats 4 and 5, suspended acquisition of Landsat MSS data. Landsat MSS data are now superseded by data from the Landsat Thematic Mapper (TM).

3.7.1.2.2 Landsat Thematic Mapper

- *http://www.ciesin.org/TG/RS/landtm.html*
- *http://landsat.gsfc.nasa.gov/*

Landsats 4 and 5, launched in 1982 and 1984, respectively, were augmented with an advanced version of an Earth observation sensor known as the Thematic Mapper (TM). The TM provides a significant increase in data acquisition capability over the MSS in a number of ways, as shown in the Figure of Observation Characteristics. The TM sensor has seven spectral bands: Six acquire Earth reflectance data, and one acquires Earth temperature data. The spatial resolution of bands in the visible and reflective infrared regions is 30 m, some 2 1/2 times better than the Multispectral Scanner (MSS). The TM sensor also has greater overall radiometric sensitivity than the MSS.

To date, the Landsat TM sensor represents the most sophisticated satellite sensor to provide Earth observation data. The sensor's complement of seven spectral bands offers the most comprehensive set of multispectral measurements for land and water surface mapping, monitoring, and analysis. Data must be ordered from the Earth Observation Satellite Company (EOSAT), who holds the commercial rights to all TM data less than 10 years old.

Landsat 7 was successfully launched from Vandenberg Air Force Base on April 15, 1999 at 11:32 am PDT. The earth observing instrument on Landsat 7, the Enhanced Thematic Mapper Plus (ETM+), replicates the capabilities of the highly successful Thematic Mapper instruments on Landsats 4 and 5*. The ETM+ also includes new features that make it a more versatile and efficient instrument for global change studies, land cover monitoring and assessment, and large area mapping than its design forebears.

The primary new features on Landsat 7 are:
- a panchromatic band with 15m spatial resolution
- on board, full aperture, 5% absolute radiometric calibration
- a thermal IR channel with 60m spatial resolution

3.7.1.2.3 Systeme Probatoire d'Observation de la Terra (SPOT) High Resolution Visible Sensor

- *http://www.ciesin.org/TG/RS/spothrv.html*

The French *Systeme Probatoire d'Observation de la Terra (SPOT)* Earth observing satellite system carries two High Resolution Visible (HRV) imaging sensors. First launched in 1986, the HRV provides high spatial resolution, with three spectral bands of visible and near-infrared data acquired at 20-m resolution, and a panchromatic band at 10-m resolution, as shown in the the table above. The sensors can be pointed to either side of the orbital track (plus or minus 27 degrees), allowing the acquisition of stereo and repeat coverage in as short a period one or four days.

The high spatial resolution of the SPOT sensors has proven to be very useful for applications requiring highly detailed information. For situations where additional spectral information is desired at SPOT-equivalent resolution, the 10-m spatial resolution of the SPOT panchromatic data can be fused with Landsat Thematic Mapper (TM) data to bring the advantages of both types of data into the same image.

3.7.1.2.4 Indian Remote Sensing

▪ *http://www.tbs-satellite.com/tse/online/prog_irs.html*
Indian program to develop an indigenous capability to image Earth, particulary India. Mission: ground water exploration, land use, forest and flood mapping, inventory of surface water.

▪ *http://www.itc.nl/~bakker/earsel/9806b.html*
Having been the seventh nation to achieve orbital capability in July 1980, India is pressing ahead with an impressive national programme aimed at developing launchers as well as nationally produced communications, meteorological and Earth resources satellites. Prof U.R. Rao, who became the Chairman of the ISRO (Indian Space Research Organization) in October 1984, said that space technology had given India the opportunity to convert backwardness into an asset; developing countries could bypass the intermediate technology stage and leapfrog into the high technology area. Like France, India has benefited from simultaneous co-operation with the CIS/USSR, the US and ESA.
Currently ISRO has more than a half-dozen remote sensing satellites under development, intended for launch in the next decade. The entire fleet will be a mix of niche-role and multi-role satellites which will join India's existing constellation of five polar orbiters: IRS-1A, IRS-1B, IRS-1C, IRS-1D, IRS-P2, IRS-P3. While development of the US Landsat remote sensing satellites series has stumbled along, Space Imaging EOSAT plans to market imagery and data from the 14 satellites that India expects to fly by the year 2005. Over 1988 to 2000 there will have been 12 satellites launched towards polar orbits. They are not only placing India in a strong position to satisfy its own needs, but also dominate a significant portion of the global commercial market. The IRS series in now moving from land applications to environmental monitoring, with a first emphasis on oceanography with the IRS-P4.
The IRS-1C and 1D offer improved spatial and spectral resolution, on-board recording, stereo viewing capability and more frequent revisits over the orbital repeat cycle of 24 days. The IRS-1C is paving the way for future satellites in the 1990s. It carries three separate imaging sensor that fulfill the traditional data needs of existing customers, while also attracting new users with high-resolution data acquisition capabilities. Its Wide Field Sensor (WiFS) provides regional imagery acquiring data with 800-kilometer swaths at a coarse 188-meter resolution in two spectral bands, visible (620-680nm) and near-infrared (770-860nm), and will mainly be used for vegetation index mapping. The WiFS offers a rapid revisit time of 3 days. The Linear Imaging Self-Scanning Sensor 3 (LISS-3) serves the needs of multi-spectral imagery clients, possibly the largest of all current data user groups. LISS-3 acquires four bands (520-590, 620-680, 770-860, and 1550-1750 nm) with a 23.7 meter spatial resolution, which makes it an ideal complement to data form the aging Landsat 5 Thematic Mapper sensor. Of course, the most talked about of the three sensors is the panchromatic camera with a resolution of 5.8 meter, giving the IRS-1C (and 1D) the highest resolution of any civilian remote sensing satellite currently in orbit. With its 5.8-meter resolution the IRS-1C is carving a market niche for itself and its twin the

IRS-1D, among GIS users whose applications require spatial detail and scene size between the 10-meter SPOT satellites and the upcoming 1-meter systems. The PAN sensor is steerable up to plus or minus 26 degrees and thus offers stereo capabilities and a possible frequent revisit of about 5 days, depending on the latitude. Working together, the IRS-1C and 1D will also cater to users who need a rapid revisiting rate.

The combination of the Wide Field Sensor with the high-resolution LISS has proven to be very useful by offering an overview of regional phenomena, a capability traditionally offered by sensors like the AVHRR, and at the same time offering the possibility to zoom in to specific areas to pinpoint very local phenomena. The spectral coverage of the two WiFS bands are equal to the bands 2 and 3 of the LISS, which makes comparison on different scales particularly easy. And, let's face it, the 188 meter resolution is far sufficient for some applications. Like, for instance, in forestry over large areas, where the old cliche about seeing the forest for the trees really applies. The 188 meter resolution is sufficient for foresters to do an inventory of large land holdings, taking note of general species variations, and diseases and mortality in the tree stands. Some remote sensing experts even credit the IRS-1C WiFS with generating a renewed enthusiasm over low-resolution imagery.

For collecting data outside the visibility region of the Indian earth station, the IRS-1C and 1D satellites have an on-board tape recorder (OTBR) which is able to record and store data collected for a duration of 24 minutes. This data will be downlinked once the satellite comes into the visibility region of the Indian data receiving station. Data recording on OBTR is based on user request; the programming of the OBTR occurs at NDC. The recorded data is played back over Shadnagar during night passes.

The IRS-1D was launched on 28 September 1997, and has similar capabilities to the IRS-1C satellite. Unfortunately, initially, it entered a wrong highly elliptical orbit due to a problem with the Indian PSLV rocket. Fuel for precise orbit control had to be used for raising the orbit's perigee. Satellite life because of fuel consumption will be in the 3-5 year range. It could have been more if not as much fuel had been use to adjust the orbit. Apart from this, the spacecraft seems to be in good health, and data quality of the images seem to be very good. At SI-EOSAT the IRS images are the preferred data sets, especially when delivered in USGS Digital Ortho Quad format.

3.7.1.2.5 IKONOS

IKONOS, the world's first one-meter resolution, commercial imaging satellite, was launched in September 1999. The company, Space Imaging, was founded in late 1994, at Denver, Colorado, U.S.A. The satellite has four mutlispectral bands in the visible and Near Infrared at a spatial resolution of 4 meters, and a Panchromatic 1 meter band covering the same spectral range.

The two following images were taken be IKONOS. Notice the high spatial resolution of the images.

The red square in Beijing, China, 22.10.1999, four meter resolution colour image
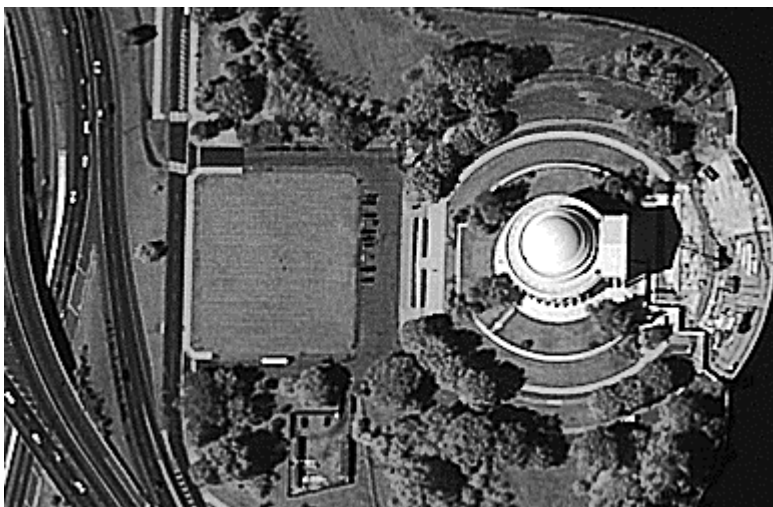
Jefferson Memorial, Washington D.C., 30.9.1999, first public image, one meter resolution B&W image

3.7.1.3 Marine Observation Satellites/Sensors

http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/indexe.html

The Earth's oceans cover more than two-thirds of the Earth's surface and play an important role in the global climate system. They also contain an abundance of living organisms and natural resources which are susceptible to pollution and other man-induced hazards. The meteorological and land observations satellites/sensors we discussed in the previous two sections can be used for monitoring the oceans of the planet, but there are other satellite/sensor systems which have been designed specifically for this purpose.

These ocean-observing satellite systems are important for global and regional scale monitoring of ocean pollution and health, and assist scientists in understanding the influence and impact of the oceans on the global climate system.

3.7.1.3.1 Coastal Zone Colour Scanner (CZCS)

The Nimbus-7 satellite, launched in 1978, carried the first sensor, the **Coastal Zone Colour Scanner (CZCS),** specifically intended for monitoring the Earth's oceans and water bodies. The primary objective of this sensor was to observe ocean colour and temperature, particularly in coastal zones, with sufficient spatial and spectral resolution to detect pollutants in the upper levels of the ocean and to determine the

nature of materials suspended in the water column. The Nimbus satellite was placed in a sun-synchronous, near-polar orbit at an altitude of 955 km. Equator crossing times were local noon for ascending passes and local midnight for descending passes. The repeat cycle of the satellite allowed for global coverage every six days, or every 83 orbits. The CZCS sensor consisted of six spectral bands in the visible, near-IR, and thermal portions of the spectrum each collecting data at a spatial resolution of 825 m at nadir over a 1566 km swath width.

3.7.1.3.2 MOS

The first Marine Observation Satellite (MOS-1) was launched by Japan in February, 1987 and was followed by its successor, MOS-1b, in February of 1990. These satellites carry three different sensors: a four-channel Multispectral Electronic Self-Scanning Radiometer (MESSR – 50m resolution), a four-channel Visible and Thermal Infrared Radiometer (VTIR – 900 to 2700m resolution), and a two-channel Microwave Scanning Radiometer (MSR), in the microwave portion of the spectrum. The MESSR bands are quite similar in spectral range to the Landsat MSS sensor and are thus useful for land applications in addition to observations of marine environments. The MOS systems orbit at altitudes around 900 km and have revisit periods of 17 days.

3.7.1.3.3 SeaWiFS

The SeaWiFS (Sea-viewing Wide-Field-of View Sensor) on board the SeaStar spacecraft is an advanced sensor designed for ocean monitoring. It consists of eight spectral bands of very narrow wavelength ranges (see accompanying table) tailored for very specific detection and monitoring of various ocean phenomena including: ocean primary production and phytoplankton processes, ocean influences on climate processes (heat storage and aerosol formation), and monitoring of the cycles of carbon, sulfur, and nitrogen. The orbit altitude is 705 km with a local equatorial crossing time of 12 PM. Two combinations of spatial resolution and swath width are available for each band: a higher resolution mode of 1.1 km (at nadir) over a swath of 2800 km, and a lower resolution mode of 4.5 km (at nadir) over a swath of 1500 km.

3.7.1.3.4 Laser fluorosensor - another kind of sensor

Some targets fluoresce, or emit energy, upon receiving incident energy. This is not a simple reflection of the incident radiation, but rather an absorption of the initial energy, excitation of the molecular components of the target materials, and emission of longer wavelength radiation which is then measured by the sensor. Laser fluorosensors illuminate the target with a specific wavelength of radiation and are capable of detecting multiple wavelengths of fluoresced radiation. This technology has been proven for ocean applications, such as chlorophyll mapping, and pollutant detection, particularly for naturally occurring and accidental oil slicks.

3.7.1.4 Hyperspectral sensors

3.7.1.4.1 Compact Airborne Spectrographic Imager CASI

CASI, the **Compact Airborne Spectrographic Imager**, is a leader in airborne imaging, being the first commercial imaging spectrometer. This hyperspectral sensor detects a vast array of narrow spectral bands in the visible and infrared wavelengths, using along-track scanning. The spectral range covered by the 288 channels is between 0.4 and 0.9 mm. Each band covers a wavelength range of 0.018 mm. While

spatial resolution depends on the altitude of the aircraft, the spectral bands measured and the bandwidths used are all programmable to meet the user's specifications and requirements. Hyperspectral sensors such as this can be important sources of diagnostic information about specific targets' absorption and reflection characteristics, in effect providing a spectral 'fingerprint'. Experimentation with CASI and other airborne imaging spectrometers has helped guide the development of hyperspectral sensor systems for advanced satellite systems.

3.7.1.4.2 *Digital Airborne Imaging Spectrometer DAIS 7915*

The **European Union**http://europa.eu.int/http://europa.eu.int/ and **DLR** are funding a new 79-channel Digital Airborne Imaging Spectrometer (DAIS 7915), which was built by the Geophysical Environmental Research corp. (GER). This new sensor covers the spectral range from the visible to the thermal infrared wavelengths at variable spatial resolution from 3 to 20 m depending on the carrier aircraft flight altitude. The DAIS 7915 is used since spring 1995 for remote sensing applications such as environmental monitoring of land and marine ecosystems, vegetation status and stress investigations, agriculture and forestry resource mapping, geological mapping, mineral exploration as well as for the supply of data for geographic information systems (GIS).

Six spectral channels in the 8000 - 12000 nm region could be used for the retrieval of temperature and emissivity of land surface objects. These and 72 narrow band channels in the atmospheric windows between 450 and 2450 nm allow to investigate land surface processes with a special emphasis on vegetation / soil interactions. Based on the requirements for on-ground calibration of the DAIS 7915 the Laboratory Calibration Facility (LCF) has been developed at DLR's Institute of Optoelectronics. The MCF covers the spectral range from 400 to 14500 nm. The DAIS 7915 has been flown several times on a Do 228 at DLR Oberpfaffenhofen since autumn 1994.

Spectrometer Characteristics
(Wavelength range: 400nm - 12.6μm, 4 Spectrometers, 79 bands)
1) 400 - 1000 nm : 32 Bands, Bandwidth = 15-30 nm Detector: Si
2) 1500 - 1800 nm : 8 Bands, Bandwidth = 45   nm Detector: InSb
3) 2000 - 2500 nm : 32 Bands, Bandwidth = 20   nm Detector: InSb
  3000 - 5000 nm : 1 Band , Bandwidth = 2.0   μm Detector: InSb
4) 8000 -12600 nm : 6 Bands, Bandwidth = 0.9   μm Detector: MCT

3.7.1.4.3 AVIRIS *Airborne Visible InfraRed Imaging Spectrometer*

AVIRIS is a world class instrument in the realm of Earth Remote Sensing. It is a unique optical sensor that delivers calibrated images of the upwelling spectral radiance in 224 contiguous spectral channels (also called bands) with wavelengths from 400 to 2500 nanometers (nm). The instrument flies aboard a NASA ER-2 airplane (a U2 plane modified for increased performance) at approximately 20 km above sea level, at about 730 km/hr. AVIRIS has flown all across the US, plus Canada and Europe.

The AVIRIS instrument contains 224 different detectors, each with a wavelength sensitive range (also known as spectral bandwidth) of approximately 10 nanometers (nm), allowing it to cover the entire range between 380 nm and 2500 nm. When the data from each detector is plotted on a graph, it yields a spectrum. Comparing the

resulting spectrum with those of known substances reveals information about the composition of the area being viewed by the instrument.

AVIRIS uses a scanning mirror to sweep back and forth ("whisk broom" fashion), producing 614 pixels for the 224 detectors each scan. Each pixel produced by the instrument covers an approximately 20 meter square area on the ground (with some overlap between pixels), thus yielding a ground swath about 11 kilometers wide.

3.7.1.5 Synthetic Aperture Radar Sensors

▪ *http://www.ciesin.org/TG/RS/sarsens.html*

Synthetic Aperture Radar (SAR) image data provide information different from that of optical sensors operating in the visible and infrared regions of the electromagnetic spectrum. SAR data consist of high-resolution reflected returns of radar-frequency energy from terrain that has been illuminated by a directed beam of pulses generated by the sensor. The radar returns from the terrain are mainly determined by the physical characteristics of the surface features (such as surface roughness, geometric structure, and orientation), the electrical characteristics (dielectric constant, moisture content, and conductivity), and the radar frequency of the sensor. By supplying its own source of illumination, the SAR sensor can acquire data day or night without regard to cloud cover. Elachi (1988) provides a technical overview of radar wave-surface interactions and their applications to land, water, and ice phenomena in Chapter 2 of *Spaceborne Radar Remote Sensing*. Most other remote sensing textbooks also provide introductory material on SAR system properties and image data applications.

Synthetic aperture radar (SAR) satellite systems currently in operation include the European Space Agency's (ESA) European Remote Sensing Satellite 1 (ERS-1), launched July 1991, and the Japanese Earth Resources satellite (JERS-1), launched February 1992. Contacts are provided for ERS-1 data and JERS-1 data. The ERS-1 sensor operates in the C-band frequency (approx. 5.6 cm wavelength) and JERS-1 operates in the L-band frequency (approx. 23 cm wavelength). Both sensors have a nominal spatial resolution of approximately 30 m. The Canadian Space Agency plans to launch its RADARSAT in 1995.

The SAR systems are now beginning to provide SAR image data on a long-term, sustained basis. The ERS-1 satellite, with a projected lifespan of three years, will be followed by an ERS-2 satellite planned to continue SAR data acquisition into the late 1990s, when advanced SAR sensors are expected to become operational as part of the Earth Observing System (EOS).

The current level of experience in operational use of SAR data is very limited compared to the use of visible and infrared data acquired by the multispectral satellite sensors. Several major characteristics of SAR data taken together, however, may promote more extensive evaluation and use of SAR data for land-use and land-cover information. These characteristics include 1) the unique information of surface roughness, physical structure, and electrical conduction properties; 2) the high spatial resolution; 3) the 24-hour, all-weather data-acquisition capability; and 4) the now-realizable long-term continuity of the data that enables repetitive (seasonal) coverage of major global land regions.

4. Corrections:

4.1 Radiometric calibration

Pixel values in commercially available satellite imagery represent the radiance of the surface in the form of Digital Numbers (DN) which are calibrated to fit a certain range of values. Sometimes the DN are referred to as the brightness values. Conversion of DN to absolute radiance values is a necessary procedure for comparative analysis of several images taken by different sensors (for example, Landsat-2 versus Landsat-5). Since each sensor has its own calibration parameters used in recording the DN values, the same DN values in two images taken by two different sensors may represent two different radiance values.

4.1.1 Main elements of sensor calibration

4.1.1.1 Absolute Radiometric Calibration – from radiance to DN and back

The following figure depicts a hypothetical response curve to a known calibration signal. There are several common curve characteristics to note. The first is that the response curve does not pass through the origin, i.e. the sensor does not give zero output even in the dark. This so-called dark signal is caused primarily by electronic noise in the sensor. The optimum operating region is the linear region indicated in the figure. Remote sensing systems are designed to measure the radiometric characteristics of the targets of primary interest in this linear region, where the output signal is linearly proportional to the input signal.
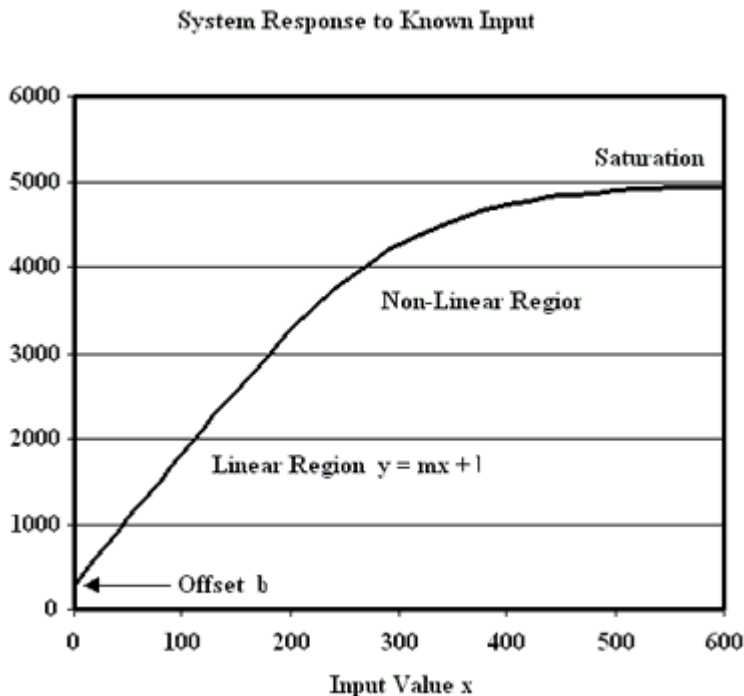
Hypothetical calibration response curve, with linear region of slope m and intercept b

This linear function is described by three parameters: the range of DN values in the image, and the lowest (Lmin) and highest (Lmax) radiances measured by a detector over the spectral bandwidth of the channel. Most commonly, the data are distributed in 8-bit format corresponding to 256 DN levels. Lmin is the spectral radiance

corresponding to the minimum DN value (usually, a value of 0). Lmax is the radiance corresponding to the maximum DN (usually, the value of 255).

For each spectral band, the output of the sensor, the grey level or DN, is related to the input signal, the radiance L in most cases, by the following equation:

$$DN = G \cdot L + D$$

The quantity G is the gain of the system and D is the dark current. All of the quantities in the equation are a function of wavelength but, for simplicity, this dependence is not indicated in the equations presented here. The image DN recorded by the sensor system can be converted to the physical quantity L (expressed in $Wm^{-2} sr^{-1}$) by inverting equation:

$$L = (DN - D) / G = (Lmax-Lmin)/255*DN + Lmin$$

The information about sensor calibration parameters (Lmin and Lmax) is usually supplied with the data or is available elsewhere (e.g., Landsat Data User's Handbook).

If the input signal exceeds the amount for which the sensor was designed, the system response will become non-linear or reach the saturation level. This is a common occurrence in land remote sensing systems when they image bright clouds and/or snow cover, for example.

4.1.1.2 Uniformity Calibration

Many remote sensing sensors in use today produce images. During the calibration process it is important to determine how the system responds to signals coming from different parts of the scene. The optics of a sensor system will be a determining factor. A phenomenon know as vignetting will reduce the light reaching the detector system towards the edges of the lens. Similarly, many sensor systems use arrays of detectors to image scenes more quickly. The detectors involved always have some differences in response to the same input signal level, which can lead to image striping. Thus, to achieve uniformity in response across the field of view of the sensor, relative radiometric calibration is also necessary.

4.1.1.3 Spectral Calibration

To determine a sensor's response to light at different wavelengths, an input signal of a known wavelength and intensity is scanned across the wavelength range and the response as a function of wavelength measured and characterized for each spectral band prior to launch (see the figure). If operational applications are not to be compromised, it is important to know the position and width of each spectral band, as well as to understand out-of-band contributions from other spectral bands.

Relative Spectral Response Profil
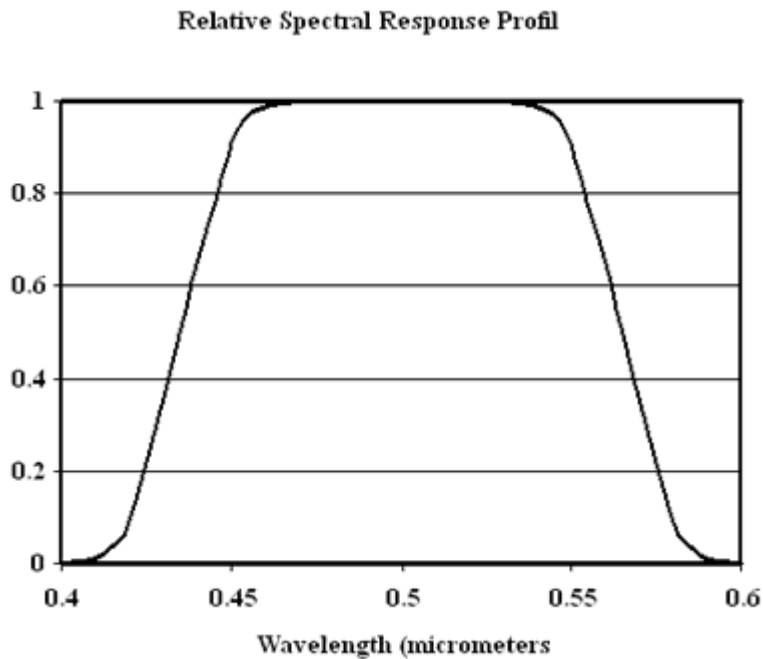
Relative spectral response profile for a broad-band sensor

4.1.1.4 Geometric Calibration

Due to optical aberrations or misalignment of discrete multiple detectors, images from different spectral bands or detectors can be misregistered. In a framing sensor, straight lines may appear curved. In this context the known input would be a test pattern of lines and geometrical shapes. The resulting output image would be compared to the input image and the appropriate geometric calibration equations implemented.

4.1.2 Calibration approaches

4.1.2.1 Prelaunch Calibration

Much of the time spent on calibration will take place before the sensor is launched into orbit or used operationally. Under controlled conditions in the laboratory, sensors can be characterized with respect to their radiometric, spectral, polarimetric and geometric properties. The equipment, time and effort devoted to such an effort are significant, but they provide a sound understanding of a given sensor's performance and facilitate stable operations on orbit over the lifetime of the mission.

Nevertheless, a satellite sensor may be calibrated months or years before being put into space. By the time the sensor is in orbit, the characteristics of the sensor's detectors and filters have often changed and thus the preflight calibrations will no longer be optimum. Despite thermal vacuum testing in the laboratory, the actual launch places the sensor in a new environment which, coupled with the effects of aging before and after launch, lead to degradations in sensor responsivity over time for all Earth observation sensor systems.

The following table shows as an example the change in the sensor gain coefficients as a function of time for the popular Advanced Very High Resolution Radiometer (AVHRR) sensors carried on board the National Oceanographic and Atmospheric Administration (NOAA) series of satellites.

Degradation of AVHRR sensor gain coefficients

Platform	Degradation in % per year	
	Channel 1	Channel 2
NOAA-7	3.6	4.3
NOAA-9	5.9	3.5
NOAA-11	1.2	2.0

4.1.2.2 Onboard Calibration

It would be gratifying to conduct as vigorous a calibration onboard the spacecraft as was done preflight. For many reasons this is never possible. However, various methods have been developed to obtain some radiometric calibration measurements onboard the spacecraft. Few if any sensor systems have provisions for post-launch checks of spectral, polarimetric and geometric sensor characteristics. Radiometric reference targets that are sometimes used include onboard standard lamps and/or solar diffuser reflectance panels, deep space, the Sun and the Moon.

4.1.2.3 Vicarious Calibration

Vicarious calibration refers to techniques that make use of natural or artificial sites on the surface of the Earth for the post-launch calibration of sensors. These targets are imaged in near-coincident fashion by the sensor to be calibrated and by one or more well-calibrated sensors from satellite or aircraft platforms or on the ground.

For example, dry lake beds or playas present bright homogeneous targets of significant size that lend themselves well to vicarious calibration. They can be characterized on the ground or from another remote sensing platform using well-calibrated sensors. With allowance for atmospheric and directional reflectance effects, as well as any differences in spectral band characteristics, a comparison between data of the same site from the reference sensor and the sensor under study can be used to update the calibration gain coefficients of the sensor under study.

4.2 Atmospheric correction - from radiance to reflectance or to temperatute\emissivity

- *http://loasys.univ-lille1.fr/informatique/sixs_gb.html*
- *http://wully.wustl.edu/epsc407/notes/notes_980217.html*
- *http://nit.colorado.edu/remsens/payne/index.html*
- *http://www.geo.unizh.ch/~dschlapf/*
- *http://www.ontar.com/*
- *http://www.ice.mtu.edu/online_docs/envi31/tut8.htm*
- *http://www.cmis.csiro.au/rsm/research/calibration/calibrate.html*
- *Erich Hernandez-Baquero (1999), Proposal of Atmospheric Compensation for Surface Temperature and Emissivity Separation, A dissertation proposal submitted in partial ful_llment of the requirements for the degree of Doctor of Philosophy in the Chester F. Carlson Center for Imaging Science of the College of Science Rochester Institute of Technology*
- *ATmosphere REMoval Program (ATREM), User's Guide, Version 3.1, Center for the Study of Earth From Space (CSES), Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder*
- *Hook Simon J. (1992), A comparison of techniques for extracting emissivity information from thermal infrared data for geologic studies, Remote Sensing of the Environment, 42; 123-135*
- *Richter Rudolf (1991), Derivation of temperature and emittance for airborne multispectral thermal infrared scanner data, Infrared Physical Technology, Vol. 35, No. 6, 817-826*
- *Schowengerdt Robert A., Remote Sensing and Methods for Image Processing, 2nd ed., Academic Press, San Diego, 520p*

After preforming the radiometric correction, image values are expressed in radiance values. As materials can be characterized by their reflectance and/or emissivity spectrum, or by their thermal properties, one needs to apply some corrections to the data, so that it will be expressed in reflectance values (for VIS-NIR passive Remote Sensing), or in emissivity and temperature values (for thermal passive Remote Sensing). These corrections are needed in order to:

- comapre images from different sensors and/or dates.
- preform environmental, ecological, or geophysical analyses.
- improve classification results

In order to do these corrections, one should consider the complex interactions the electromagnetic radiance arriving at the sensor goes through:

1. The solar radiance reaching the upper atmosphere depends on the:
   - longitude and latitude
   - time (date and hour) of the year
2. Going through the atmosphere (up and down), the Electromagnetic radiance is affected by processes of:
   - (multiple) scattering
   - absorption
   - these depend on the spatial and temporal distribution of the different gases, molecules and aerosols in the atmosphere.
3. The amount of incident radiance upon the surface, depends on the:
   - optical depth (height of the surface)
   - slope and aspect of the surface (defining the effective angle between the incident radiance and the surface)

4. Upon hitting the ground, the radiation is partially transmitted, absorbed and reflected, depending on the physical and chemical properties of the surface, and on its roughness.
5. The radiance reaching the sensor from the surface is also affected by the:
   - adjacency effect; atmopsheric scattering increases the autocorrelation between adjacent pixels, the signal from a certain pixel partly mixed with signals from its surrounding pixels.
   - optical length between the sensor and the surface, and the viewing geometry; the more a pixel is located at the edge of an image and far from the nadir, the more the signal leaving it and reaching the sensor is attenuated by the atmosphere.

The higher the spectral resolution of the sensor, the more sophisticated will be the correction algorithms, as specific absorption of gases appear (not averaged and smoothed away by wide bands), which should not be mistaken with absorption features of the target materials one wants to identify.

As to date, there is no way to absolutely correct all these effects; different methods are employed, depending on the purpose of the work, the radiometric and atmospheric data at hand, the hardware and the software. The more sophisticated ones recognize the fact that the atmospheric attenuation is not uniform over the whole image (varying topography, water vapor content, etc), and try to deal with it.

The different methods can be classified into the following, and they can be combined or applied in steps (some of these methods are further explained below):

- Disregarding the atmospheric and topographic effects, analysing the DN values.
- Calibrating images from different dates to Like-Values; this technique states that it is sufficient to convert the raw digital counts (DN's) of one image to be consistent with the counts for a chosen reference image, when comparing images to detect change.
- Topographic correction; correcting for the effects of slope and aspect (shading), and optical path length (distance between sensor and surface) using a Digital Elevation Model.
- Image based corrections:
  - Normalizing radiance values, in order to obtain values of relative reflectance. These methods include the IARR (Internal Average Relative Reflectance) and Flat-Field.
  - Extracting atmospheric correction from the image itself, e.g. water vapor and the aerosol content (One of the greatest challenges in dealing with water vapor is that its concentration changes with time, location, and altitude).
    The estimation of the aerosol content is not yet operational. Statistical approaches use the contrast reducing effect of strong scattering to estimate the aerosol content. Histogram matching allows one to even obtain the spatial distribution of the haze in the image - as long as the image data is statistically homogeneous. Simpler approaches use dark tar-gets or a series of known spectrally homogeneous areas for an estimate of the atmospherically scattered radiance within the image.
- Multiple Frequency; Absorption of electromagnetic radiation by the atmosphere is frequency dependent. By measuring the radiance at different wavelengths a correction can be calculated for the atmospheric absorption.

- Multiple View; The path length of atmosphere traveled by radiation affects the distance over which the radiation will be attenuated. If measurements are made of the same location with different scan angles of known distance, a correction can be made based upon this differential absorption.
- Path Length Correction; A correction can be included which attempts to correct for varying scan angles. This generally applies a correction which is factored by the path length of the line segment from the surface to the sensor.
- Empirical line correction; Empirical corrections fully rely on the knowledge about the spectrum of a group of pixels within the image (these are preferably gathered using a field spectrometer at the time of the flight of the sensor). They are fast and allow a pragmatic processing of the image at low costs. Their disadvantage is a low reliability in mountaineous terrain and for changing meteorologic conditions within one image.
- Adjacency correction; Each pixel has to be corrected with respect to the average reflectance of the adjacent areas. This can be done by the definition of a spatial convolution function which takes a distance-weighted average of the adjacent area in the image to calculate an adjacency weighting factor.
- Atmospheric modelling; constructing an approximate model of the atmosphere, based on world wide measurements of many atmopsheric parameters, gethered at different areas, time of the year, and atmospheric layers. Combining the above data, the software looks up through the tables and models the atmosphere as it was at the time of the flight. Atmospheric and meteorological data gathered at the time of the flight, at the surface, or using a radiosonde, can improve the atmospheric model.

Analysing the results of an atmospheric correction, one should compare the absolute values of reflectance and the shape of the reflectance spectrum, to those of the reflectance spectrum measured in the field or in a laboratory.

Below is given a more detailed description of some of the common methods used for atmospheric correction of images.

4.2.1 Calibrating images from different dates to Like-Values

Invariant targets are being used to calibrate the digital counts of coincident or overlapping scenes to the digital counts of a reference image. The digital counts of these targets are extracted from both images and robust regression techniques are used to estimate the gains and offsets.

Guidelines for the Calibration of a Sequence of Images:

1. *Selection of reference image*
   The reference image is the scene to which the other scenes are related. It is important that it is cloud free and contains no white outs (i.e. values >= 255).
2. *Selection of Invariant Targets*
   Invariant targets are features which have constant reflectance over time. The data values are used to define linear functions to transform each overpass image to the reference image by assuming these targets should have the same digital count values in each image. Targets must be selected for a range of bright, mid-range, and dark values and you must have a balanced number of bright and dark targets. Possible targets are listed below.

Dark:	Ocean	Lakes
Mid-range:	Rock outcrops	Airfields

	Quarries, gravel pits and mines	Dams
Bright:	Catchments	Sand

Targets can be located by viewing the reference image on the screen and using large-scale topographic maps. Vegetated targets should usually be avoided as they show seasonal trends. When obtaining targets, it is important to locate targets over a uniform section of the feature. . When obtaining targets, it is important to record information about the target (Target type, size, line and pixel location, geographic location, map sheet).

3. *Calculation of coefficiants*
   The next step is to calculate the regression coefficients which relate the overpass images to the reference image. Coefficients for several methods, such as least squares, s-estimation and weighted least squares estimation procedures can be calculated.

4. *Coefficient Examination*
   Once the calibration functions have been calculated, the next step is to examine and select the best calibration line for each band. This is done by plotting the lines and data. The y-axis is used for the reference image data and the x-axis is used for the overpass image data.
   Once the plots have been examined, the line of best fit is then chosen.

4.2.2 Internal Average Relative Reflectance (IARR)

The IARR calibration technique is used to normalize images to a scene average spectrum. This is particularly effective for reducing imaging spectrometer data to "relative reflectance" in an area where no ground measurements exist and little is known about the scene. It works best for arid areas with no vegetation. The IARR calibration is performed by calculating an average spectrum for the entire image and using this as the reference spectrum. Apparent reflectance is calculated for each pixel of the image by dividing the reference spectrum into the spectrum for each pixel.

4.2.3 Flat Field

The "Flat Field Calibration" technique is used to normalize images to an area of known "flat" reflectance. The method requires that you locate a large, spectrally flat, spectrally uniform area in the image data, usually defined as a Region of Interest (ROI). The radiance spectrum from this area is assumed to be composed of primarily atmospheric effects and the solar spectrum (that is, the spectrum of this ROI, which is supposed to be flat, isn't so, because of the atmospheric interference). Relative reflectance is calculated for each pixel of the image by dividing the reference spectrum into the spectrum of each pixel.

4.2.4 Empirical line

The Empirical Line calibration technique is used to force image data to match selected field reflectance. This method requires ground measurements and/or knowledge. Two or more ground targets are identified and reflectance is measured in the field. Usually the targets consist of at least one light and one dark area. The same two targets are identified in the image and average spectra are extracted for Regions of Interest. A linear regression is calculated between the field reflectance spectra and the image radiance spectra (for each band seperately) to determine a linear transform from radiance to reflectance for each band of the data set. Gains and offsets calculated in the regression are applied to the radiance spectra for each pixel to produce apparent reflectance on a pixel-by-pixel basis.
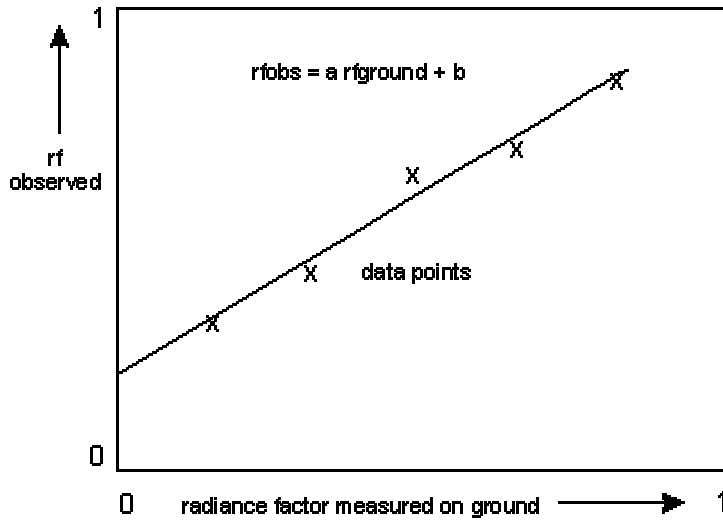
Empirical line scatter plot chart of a selected band, each point representing a Region of Interest

This technique assumes that the atmospheric effects are constant across the scene, assume that skylight scattered from the surface to the observer can be ignored, and it is basically a single scattering solution for the atmosphere.

When using only two ROI's, there is no redundancy, and therefore no computation of the RMS of the empirical line model is possible. When more than two targets are involved, by analysing the residuals on the scatter plot, it is possible to identify which of the ROI's serve well. Sometimes, different sets of ROI's may be utilized in order to preform the correction over different ranges of the spectrum. The method is also applied as a further correction, after using an atmospheric model.

4.2.5 Atmospheric modelling

The state of a static atmosphere is described in terms of the parameters of the ideal gas law:

$$P = nkT$$

where P is the atmospheric pressure, n is the number density, k is Boltzmann's constant, and T is the temperature. The atmosphere is commonly represented as a stack of layers. Layers are defined in terms of temperature, composition, mixing, and ionization distribution with altitude.

The mathematical description of the radiation propagation and emission processes that lead to an observed radiance by a remote sensing platform is often referred to as the "forward" model. The basic premise is that the observed radiance is a function of the scene in view and the composition and thermodynamic state of the intervening atmosphere. The inverse problem (otherwise known as retrieval theory) is the one in which the parameters affecting radiation are inferred from the observed radiance.

The radiometric formulation of radiative transfer is fairly standard across the literature with the exception of the notation used. The radiance reaching a sensor is the sum of the contributions from different propagation paths as shown in the following figure:

64

Radiation propagation paths for reflective and thermal regions
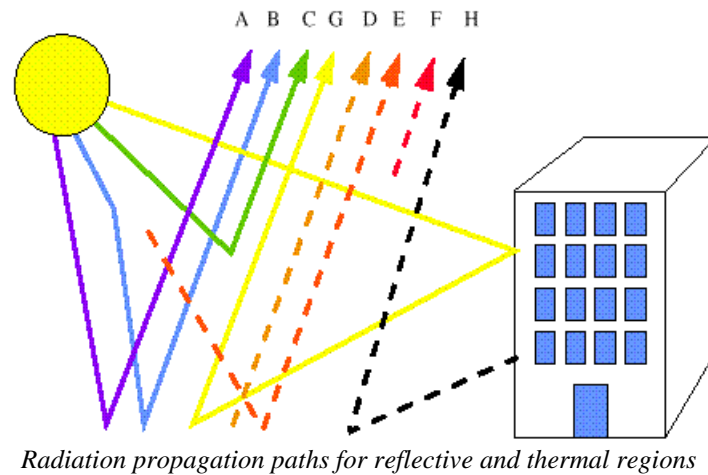
The propagation paths depend on the region of the electro-magnetic spectrum of the radiation (see the figure in the chapter describing the electromagnetic spectrum, showing the respective emitted radiance of the Earth and the Sun).

In the reflective region, the dominant paths are:
A) direct sunlight hits a target and reflects,
B) sunlight scatters in the atmosphere and reaches a target and is then reflected,
C) sunlight scatters in the atmosphere and reaches the sensor, and
G) sunlight reflects off the background, reaches the target, and is then reflected.

In the thermal region, the dominant paths are:
D) thermal photons emitted by a target reach the sensor,
E) thermal radiation from the atmosphere reaches the target and is reflected,
F) thermal photons from the atmosphere reach the sensor, and
H) thermal photons from the background reach a target and are reflected.

Atmopsheric models need to account for:
- the complex, time and space varying composition of the atmosphere, and
- the wavelength dependent interactions experienced by the radiance reaching the sensor.

A generalized form of the radiative transfer equation in the reflective part of the electromagnetic spectrum, is given below (this equation is wavelength dependent):

$$L_{sensor} = E_0 * \cos(Q) * t * r / p + E_d * t * r / p + L_{path}$$

Ground reflected   Path scattered

where,

L_{sensor} = Radiance received at the sensor
E_0 = Solar radiance at the top of the atmosphere
Q = Incidence angle of solar radiance on the surface (0 for vertical, 90 for horizontal)
t = Transmittance factor
r = Reflectance factor
E_d = Scattered background radiation (of the sky)
L_{path} = Path scattered radiation reaching the sensor

The radiative transfer equation is often implemented to some degree in computer models that use databases containing molecular and particulate absorption and scattering characteristics. Various models exist with different degrees of spectral resolution, number of atmospheric constituents, cloud models, etc. Computer models can be broken down into two major classes: band transmittance and line-by-line transmittance models.

4.2.5.1 Band Transmittance Computer Models

These models are designed to lower the computational cost of computing the radiative transfer through a inhomogeneous path in the atmosphere. This is done by fitting a band model (such as the Lorentz model) to line spectra obtained through measurements in the field or the laboratory. The model then simply uses parameters such as the line width and strength to compute the overall absorption through a given path. The most widely used model is the MODTRAN model developed by the United States Air Force Research Laboratory (AFRL), Space Vehicles Directorate, Battlespace Environment Division (originally a group within the Air Force Geophysics Laboratory). The current version of MODTRAN is version 4.

Band models are useful when an estimate of the atmospheric conditions for a particular image collection is needed. In many instances, the approach involves using a standard atmosphere that matches the particular conditions during the image collection (i.e. choosing the "mid-latitude summer profile" in MODTRAN for an image collected over Oklahoma on a summer day). Other options involve using radiosonde data as an input to the model. The spectral resolution of the MODTRAN model is limited to contiguous 2 cm^{-1} bands at 1 cm^{-1} intervals. This resolution is better than many multispectral sensors and reasonable for a higher resolution hyperspectral sensor. However, as the channel widths of these new sensors decrease, the band transmittance model starts to become inadequate due to errors arising from spectral calibration. In other words, the sensor's band centers may not line up with the band centers of the model. In that case, a line-by-line model may be needed.

4.2.5.2 Line-by-line Models

Line-by-line models are very high resolution radiative transfer models that use a large database of molecular absorption and scattering measurements. The most widely used database, also generated by the United States Air Force, is the HITRAN96 database. The model FASCODE taps into this database to generate high spectral resolution transmittance calculations. The radiation propagation is then performed on a line-by-line basis so that the Beer-Lambert Law holds. Because of this, the computation times are larger and often less practical. The University of South Florida has released the latest version 2.51 of HITRAN-PC containing extensive updates, including the ability to read the 1996 HITRAN Database available on CD-ROM. The 1996 HITRAN Database includes 999,363 absorption lines including data on new molecules.

Other "fast" forward models can be built, such as PLOD and OPTRAN, which are derivatives of the GENLN2 model. These are used in atmospheric research applications and is being currently used by the Atmospheric Infrared Sounder (AIRS) team in support of the NASA EOS program.

As Band Transmittance models are the ones more used for Remote Sensing applications, a short description of some of these will be given.

4.2.5.3 MODTRAN

MODTRAN performs accurate and speedy calculation from the UV through the visible, infrared and microwave spectrum (0 to 50,000 cm^{-1}). Calculations in the visible and ultraviolet spectral regions are performed at lower spectral resolution (20 cm^{-1}), while those in the IR and longer wavelengths are done at 2 cm^{-1} resolution. PCModWin (commercial Windows versions of the Air Force's Phillips Laboratory's MODTRAN model) accommodates standard (WMO) atmospheric profiles, numerous aerosol models (e.g. fogs, dust, maritime, etc.), water and ice cloud models, and totally arbitrary geometric paths from sea level to 100 Kms. The molecular absorption properties are based on the internationally recognized HITRAN atlas of spectroscopic parameters.

MODTRAN itself does not preform an atmospheric correction over an image (ATREM and ATCOR, described later, receive as input an image, and their output is the same image atmospherically corrected). MODTRAN receives as input a laboratory spectrum of some object (it can be also measured in the field, and be free of atmospheric interferences), and its output are the spectrum of the same object as it would be obtained on the sensor, and in addition many parameters of the atmospheric profile, by wavelength and height above ground.

The workflow of using MODTRAN for preforming an atmospheric correction of an image, goes on as follows:

1. Preparing the reference background file containing the laboratory spectrum.
2. Giving the software the input data (choosing from the available options) needed for it to build the vertical atmospheric profile and calculate the parameters of the radiative transfer model:
   - Average atmosphere model,
   - surface spectrum file and the object's temperature,
   - aerosol model,
   - geometric parameters of the optical path (height of the sensor above ground and its oblique viewing angle),

- spectral resolution to preform the calculations, and
- temporal (day of year, and the time the image was acquired) and locational (longitude and latitude) data of the image

3. Comparing the resulting output of the modelled "total radiance spectrum as seen by the sensor", with the actual radiance of the object as it is in the image.
   If the modelled spectrum does not answer to out needs, we can change the input parameters given to MODTRAN, in order to approximate more closely the prevailing atmospheric conditions during the time the image was taken.
   If the modelled spedctrum is good then we go and:
4. Extracting the atmospheric parameters from the output file, and using them with the radiative transfer model equation, on the image, in order to correct it.

When using the output parameters for the radiative transfer equation, one should remember the following points:

- The greater the reflectance r of an object, the value of L_{path} (path scatered radiation) will be higher. One can define a linear relation between the values of r and L_{sensor}, by running MODTRAN several times, each time with a theoretical spectrally flat object (e.g., 0%, 10%, 20%, 30%, 60%), and thus bypass the different values of L_{path} obtained when using different reference spectra.
- This alone is not enough, as one should also take in account the sensor's oblique viewing angle (*Initial Zenith Angle*, in MODTRAN terms). For Landsat, this effect is negligible and it can be assumed the entire scene is taken in the *nadir*. However, for SPOT (which can be directed to "look" to the side), and for airborne scanners (such as AVIRIS) that are closer to the ground and have a larger Field Of View, it is important.
   The far is the angle from the nadir, the longer is the optical path, there will be more scattering on the one hand, and more absorption on the other hand. These two contrasting effects, create a difference between dark and bright objects.
   In the far angle, over dark objects, more radiance will reach the sensor (due to more scattering; in an angle of 30° the difference can reach more than 20%), while over bright objects, less radiance will reach the sensor (scattering processes having a smaller ratio of the total radiance at the sensor relative to the direct reflectance from the object, therefore, in the far angle, there will be more absorption over a bright object; in angle of 30° the difference can be between 5-10% in specific wavelengths where there absorption features of gases in the atmosphere).
   To correct this, one can divide the image into several longitudinal stripes (5 to 10 degrees wide), computing in each one of them the regression between between the values of r and L_{sensor}. This method, is implemented in ATCOR.

As MODTRAN is "building" an atmosphere over a reference spectrum, some of the atmospheric interferences can not be taken in account:
- the adjacency effect,
- the topographic situation (shading, slope and aspect of the object's location in the real world), and
- the atmospheric interferences above the scene are assumed to be uniform.

Ancillary data can help with the characterization of the atmosphere. This is often done

through radiosonde measurements in which a sensor is mounted on a balloon and launched at the same time (or as near as possible to the same time) as the time of image acquisition. These radiosondes then measure vertical temperature, pressure, wind speed, and humidity profiles of the atmosphere. These profiles are then entered into a radiative transfer model. to estimate the atmospheric transmittance and radiation.

The problem with radiosondes is that they are susceptible to drift during their ascent and may not accurately represent the actual composition of the atmosphere for a given column of air. Furthermore, the logistics of successfully launching a coincident radiosonde for every remote sensing acquisition over the planet is impractical. Thus, it is preferable to develop techniques that can accomplish atmospheric compensation using only the in-scene data inherent in an image.

This task has not been possible until the advent of *hyperspectral*, or imaging spectrometer, sensor technology. Here, a hyperspectral sensor is any remote sensor with high spectral *and* spatial resolution. The spatial resolution aids in target definition and measurement of small-scale atmospheric events. By increasing the number of spectral measurements of the Earth-leaving radiance, the atmospheric structure can be inferred directly from the measurement.

The ATmospheric REMoval Program (ATREM), described below, is an example to a model that tries to do just that, for water vapor.

As it is very difficult to reconstruct with an atmospheric model the exact conditions pevailing at the time of the flight, an *emprical line correction* preformed after the modelling of the atmosphere, might be helpful.

MODTRAN can be downloaded from the U.S.A. Air Force Research Laboratories, at the following site:
http://www-vsbm.plh.af.mil/soft/instruct.html

The company marketing MODTRAN is ONTAR corporation (a demo version of the software can be downloaded there):
http://www.ontar.com/
http://www.ontar.com/additional.html#modwin

4.2.5.4 2nd Simulation of Satellite Signal in the Solar Spectrum - 6S code

This code predicts the satellite signal from 0.25 to 4.0 micrometers assuming a cloud-free atmosphere. The main atmospheric effects (gaseous absorption by water vapor, carbon dioxyde, oxygen and ozone, scattering by molecules and aerosols) are taken into account. Non-uniform surfaces can be considered, as well as a bidirectional reflectance as boundary conditions.

The following input parameters are needed:
- Geometrical conditions
- Atmospheric model for gaseous components
- Aerosol model (type and concentration)
- Spectral condition
- Ground reflectance (type and spectral variation)

At each step, you can either select some standard conditions (for example, spectral bands of satellite for spectral conditions) or define your own conditions.

The authors of this package are:

6S code: E. Vermote(3), D. Tanre(1), J.L. Deuze(1), M. Herman(1), J.J. Morcrette(2). Motif code: L. Gonzalez(1).

From:
(1)Laboratoire d'Optique Atmospherique
Universite des Sciences et Technologies de Lille
59655 Villeneuve d'Ascq Cedex - France
(2)E.C.M.W.F.
Reading - England
(3)Code 923 / GIMMS group:
GSFC/NASA
Greenbelt, MD 20771 - USA

4.2.5.5 ATmospheric REMoval Program (ATREM)

The ATmospheric REMoval Program (ATREM) is a radiative transfer model-based technique for deriving scaled surface reflectance from AVIRIS data without a priori knowledge of surface characteristics. The atmospheric scattering is modeled using the 6S code.

The spatial and temporal variations of atmospheric water vapor pose difficulties in removing water vapor absorption features in hyperspectral data. In this algorithm, the amount of water vapor is derived on a pixel-by-pixel basis from the data using the 0.94- and the 1.14-μm water vapor bands and a three-channel ratioing technique. The derived water vapor values are then used for modeling water vapor absorption effects in the entire 0.4-2.5 μm region. Together with the solar irradiance curve above the atmosphere, and transmittance spectra for each of the atmospheric gases CO2, O3, N2O, CO, CH4, and O2, a "scaled surface reflectance" is retrieved. This can be converted to real surface reflectance if surface topography is known.

For six gases (CO_2, O_3, N_2O, CO, CH_4, and O_2), the algorithm assumes that the amounts of the gases are uniform across the scene. Only one transmittance spectrum is calculated for each of these gases.
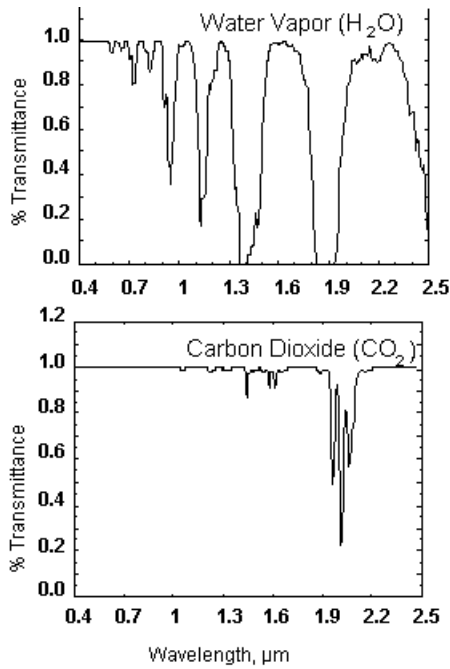
Fig. 3. Atmosphere water vapor transmittance spectrum (upper plot) and carbon dioxide transmittance spectrum (lower plot). The spectra, at a resolution of approximately 10 nm, were calculated for an observer above the atmosphere looking straight down and for a solar zenith angle of 45 degrees using the tropical model of LOWTRAN7 (Kneizys et al., 1988).
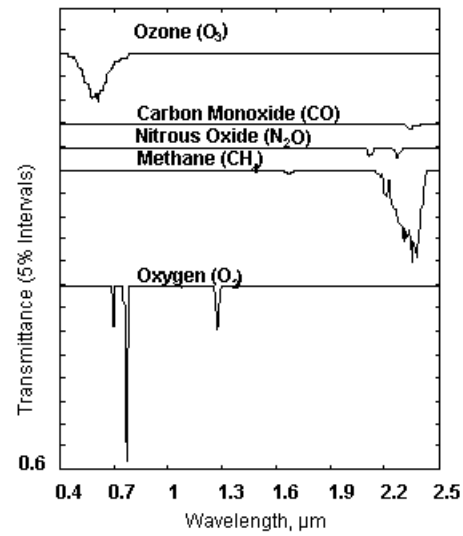
Fig. 4. Transmittance spectra of atmospheric ozone, carbon monoxide, nitrous oxide, methane, and oxygen. The spectral resolution, the geometry, and the model atmosphere used in the calculations are the same as those used in calculating the spectra in Fig. 3.

The algorithm treats the atmospheric water vapor differently. The atmospheric water vapor concentrations vary significantly with time and altitude. The derivation of water vapor values from imaging spectrometer data is mainly based on two facts. One is that the surface reflectance curves vary nearly linearly with wavelength in the 0.94- and the 1.14-μm water vapor band absorption regions for common soils and rocks. The other is that under typical atmospheric conditions, the transmittances of the two water vapor bands are sensitive to changes in the amount of atmospheric water vapor.
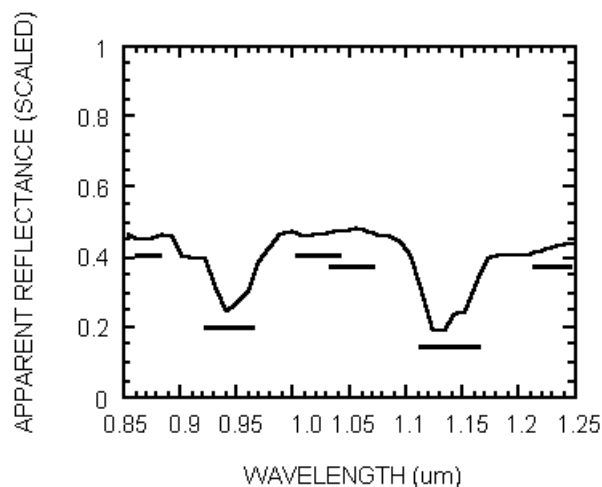
An apparent reflectance spectrum with relevant positions and widths of spectral regions used in the three channel ratioing illustrated.

A three-channel ratioing technique is used in the ATREM derivation of the water vapor value from a radiance spectrum. The mean apparent reflectance at the water vapor center is divided by one half of the sum of the mean apparent reflectances at the two window regions. The ratio effectively removes the linear surface reflectance effect and gives a mean observed transmittance for the 0.94-μm (and for the 1.14-μm) water vapor band. By matching the mean observed transmittance with theoretically calculated mean transmittances using atmospheric and spectral models, the amount of water vapor in the Sun-surface-sensor path is obtained, and an atmospheric gaseous transmittance spectrum corresponding to the pixel is then obtained.

The following systematic errors exist in ATREM:
- The sensor is assumed to be looking only in nadir.
- The atmospheric adjacency effect is not modeled.
- Over mountainous terrain, one part of the terrain may be illuminated by scattering from another part of the terrain. This introduces another kind of adjacency effect. This effect can be refered as the "topographic adjacency effect". This kind of adjacency effect is also not modeled in our program.
- Some of the surfaces, such as vegetation, snow, ice, and iron-rich soils and minerals, do not have linear reflectances in the 0.94- and the 1.14- μm water vapor band absorption regions. The three channel ratios calculated from the data over these surfaces contain the surface reflectance effects. Systematic errors are therefore introduced in the derived water vapor values. In order to decrease the errors in the derived water vapor amounts, the center positions and widths of the window and water vapor absorption channels are all allowed to vary in the actual implementation of the three-channel ratio technique.

ATREM can be obtained via anonymous ftp from *cses.colorado.edu* or by contacting the Center for the Study of Earth from Space at 303-492-5086.

4.2.5.6 ATCOR

The ATCOR (ATmospheric CORrection) algorithm, as it is implemented in the Remote Sensing software of ERDAS-Imagine, is designed for high spatial resolution satellite sensors like Landsat Thematic Mapper (TM). The algorithm works with a catalogue of atmospheric correction functions stored in look-up tables. The catalogue consists of a broad range of atmospheric conditions (different altitude profiles of pressure, air temperature, and humidity; several aerosol types; ground elevations from 0 to 1.5 km above sea level; solar zenith angles ranging from 0° to 70°). The catalogue covers visibilities (surface meteorological range) from 5 km to 80 km, values can be extrapolated down to 4 km and up to 120 km. The 1996 edition of the catalogue was compiled using the MODTRAN-2 and the SENSAT-5 codes.

The algorithm consists of an interactive and an automatic part. The interactive phase serves for the definition of a reference target (dense dark vegetation or water) as well as haze and cloud. The reflectance of the reference target in a single spectral band (dark vegetation : TM band 3, water TM band 4) has to be specified. Additionally, the image can be partitioned into subimages, called sectors. This phase also selects one of the atmospheres available in the catalogue, i.e. the altitude profile of pressure, temperature and humidity as well as the aerosol type (e.g. rural) are fixed.

The automatic phase first calculates the visibility in those regions containing reference pixels. The visibility is obtained by matching the measured signal (i.e. the digital number DN converted to a radiance using the sensor calibration) to the model-derived signal in the spectral channel of known target reflectance. The sector-average visibility over the reference regions is assigned to the non-reference areas.

The second step is the haze removal performed by histogram-matching the statistics of the haze regions to the statistics of the clear part of the scene for each sector and each channel. The last step is the calculation of the ground reflectance image including the adjacency correction, and the computation of the ground brightness temperature image (TM band 6).

Model ATCOR2 is restricted to satellite sensors with a small swath angle, like Landsat TM, MSS, Resurs-01 MSU-E and SPOT. The model assumes a terrain consisting of horizontal surfaces of Lambertian reflectances.

It should be noted however, that the full version of ATCOR, outside of ERDAS, also take in account the different viewing angles from the nadir and their influence on the optical path, and can also handle hyperspectral image data.

Algorithm Developed by:
Dr. Rudolf Richter
DLR - IB 552-03/96
Institute for Optoelectronics
D-82234 Wessling
GERMANY

World-Wide Distribution and Support:
GEOSYSTEMS GmbH
Riesstr. 10
82110 Germering
GERMANY

4.2.6 Temperature calibration of images

▪ *http://userservice.nrct.go.th/userservice/Faq/tmband6.html*
▪ *Sabins Floyd F. (1976), Remote Sensing – Principles and Interpretation, Freeman*

In order for thermal images to be quantitative, temperature calibration must be provided for the scanners. This is done in the following way: electrically heated temperature calibration sources (Black Bodies) are mounted in the scanner, in either side of the angular field of view. The scanner records the radiant temperature of the first calibration source, then sweeps the terrain, and finally records the radiant temperature of the second source. The two sources are set at two different temperatures, thus defining a scale for determining the temperature at any point along the magnetic tape record of the scan line.

4.2.7 Thermal properties of materials

▪ *http://priede.bf.lu.lv/GIS/Descriptions/Remote_Sensing/An_Online_Handbook/Sect9/nicktutor_9-3.shtml*
▪ *http://qsilver.queensu.ca/~gphy342/presentation.html*

A thermal sensor picks up radiant emitted energy from a surface target heated through radiation (solar insolation and sky radiance), convection (atmospheric circulation) and conduction (through the ground). Thus, most sensed heat from surfaces has its origin in solar illumination, that varies with both diurnal and seasonal changes as well as

cloud cover, but there is also a small, nearly constant contribution from internal heat flux from the Earth's interior (much of this is due to thermal inputs from radioactive decay). Heat is transferred into and out of near surface layers owing to external heating by the thermal processes of conduction, convection, and radiation.

A primary objective of temperature measurements and related thermal responses is to infer something about the nature of the composition and other physical attributes of materials at the Earth's surface (and, in its atmosphere). For any given material, certain characteristic internal properties play important roles in governing the temperature of a body at equilibrium with its surroundings.

Heat Capacity (C): The measure of the increase in thermal energy content (Q) per degree of temperature rise. It is given in cgs units of calories per cubic cm. per degree Centigrade, and its denotes the capacity of a material to store heat (recall from physics that a calorie [cal] is the quantity of heat needed to raise one gram of water by one degree Centigrade). Heat capacity is calculated as the ratio of the amount of heat energy, in calories, required to raise a given volume of a material by one degree Centigrade (at a standard temperature of 15° Centigrade.) to the amount needed to raise the same volume of water by one degree Centigrade. A related quantity, specific heat (c), is defined as C=c/r - units: calories per gram per degree Centigrade) where, r = density ; this associates Heat Capacity to the thermal energy required to raise a mass of 1 g(ram) of water by 1 degree Centigrade.

Thermal Conductivity (K): The rate at which heat will pass through a specific thickness of a substance, measured as the calories delivered in 1 second across a 1 centimeter square area through a thickness of 1 cm at a temperature gradient of 1 degree Centigrade (units: calories per centimeter per second per degree Centigrade).

Thermal Inertia (P): The resistance of a material to temperature change, indicated by the time dependent variations in temperature during a full heating/cooling cycle (a 24-hour day for the Earth); defined as:
$$P = (Kcr)^{1/2} = cr(k)^{1/2}$$
(k is a term, related to conductivity K, known as thermal diffusivity, in units of calories per centimeter squared per square root of degree Centigrade seconds). P is a measure of the heat transfer rate across a boundary between two materials. e.g., air/soil. Because materials with high P possess a strong inertial resistance to temperature fluctuations at a surface boundary, they show less temperature variation per heating/cooling cycle than those with lower thermal inertia.

Some characteristic values of these intrinsic thermal properties:

	Water	Sandy soil	Basalt	Stainless steel
K	0.0014	0.0014	0.0050	0.030
c	1.0	0.24	0.20	0.12
r	1.0	1.82	2.80	7.83
P	0.038	0.024	0.053	0.168

The interpretation of thermal data and images depicting temperature distribution over an area is not a simple matter. In many instances, efforts must be confined to looking

for patterns of relative temperature differences rather than the absolute values because of the many complex factors that make quantitative determinations difficult, such as:

- Number and distribution of different material classes in the instantaneous field of view
- Variations in the angle of thermal insolation relative to sensor position
- Dependency of thermal response on composition, density and texture of the materials
- Emissivities of the surface materials
- Contributions from geothermal (internal) heat flux; usually small and local
- Topographic irregularities including elevation, slope angle, and aspect (surface direction relative to Sun's position)
- Rainfall history, soil-moisture content, and evaporative cooling effects near surface
- Vegetation canopy characteristics, including height, leaf geometry, plant shape
- Leaf temperatures as a function of evapotranspiration and plant stress
- Near surface (1 to 3 meters) air temperature; relative humidity; wind effects
- Temperature history of atmosphere above surface zone
- Cloud-cover history (during heating/cooling cycle)
- Absorption and re-emission of thermal radiation by aerosols, water vapor, air gases

Some factors have fixed or constant effects; others vary with each sensor overpass. It may be possible to correct for the influence of some of the variable factors but this is difficult to do routinely. Measurements made at isolated individual points in a scene and extrapolated to the general scene have limited validity.

Unlike remote sensing of reflected light from surfaces in which only the topmost layers a few molecular layers thick are involved, thermal remote sensing is affected by energy variations extending to varying shallow depths below the ground surface. The most critical consideration in analyzing and interpreting thermal data and imagery is that of knowing the physical and temporal conditions under which the near surface layers are heated. Over the seasons, minor shifts in mean temperature in bedrock can occur to depths of 10 meters (33 ft) or more. Materials at and immediately below the surface are heated significantly during the day by incoming solar radiation and heat transfer from the air. Temperatures usually drop at night primarily by radiative cooling (with maximum radiative cooling under cloudless conditions), accompanied by some conduction and, for fluids, convection. During a single daily (diurnal) cycle, the near surface layers (commonly, unconsolidated soils) experience alternate heating and cooling to depths typically between 50 and 100 centimeters (20-40 inches). The daily mean surface temperature is commonly near the mean air temperature. Observed temperature changes are induced mainly by changes during the diurnal heating cycle but seasonal differences (averages and range) in temperature and local meteorological conditions also affect the cycle response from day to day.
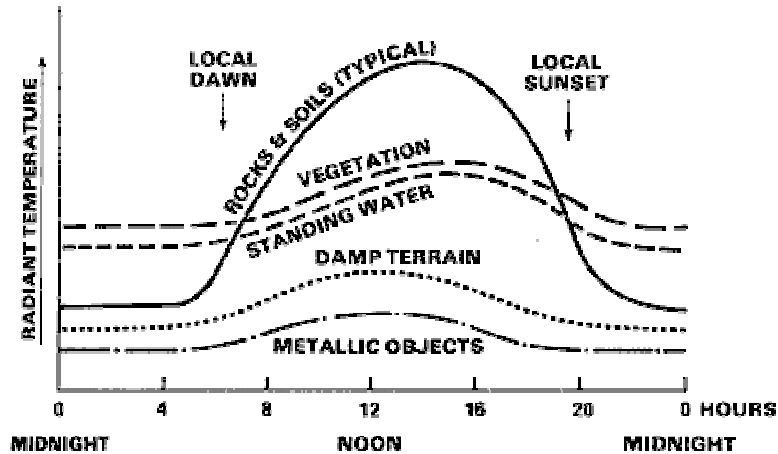
Changes in radiant temperatures of five surface-cover types during a 24-hour thermal cycle. From F.F. Sabins, Jr., Remote Sensing: Principles and Interpretation. 2nd Ed., © 1987.

The curves shown here summarize the qualitative changes in radiant temperature during a 24-hr cycle, beginning and ending at local midnight, for five general classes of materials found at the surface. From these curves one may estimate the relative gray levels that would be recorded in a thermal image, as a function of the material and the time of day. Given two thermal images of the same local, taken 12 hours apart, about noon and about midnight, one might determine the identities of co-registered pixels based on their temperatures and their thermal inertia.

Thermal inertia cannot be measured remotely as each of the three parameters must be measured in contact. Instead, Apparent Thermal Inertia (ATI) may be used as a substitute, derived from the following equation:

ATI = (1 – albedo) / (T radmax - Trad min)

where max and min radiant temps. Can be obtained from day and night thermal images.

Thermal crossover Radiant temperature curves of different materials may cross each other due to widely differing responses of each by day and by night. Occurs at or about local sunrise and sunset. Must be taken into account in mission planning in order to maximise the probability of separating surface features by their temperature characteristics.

Below is given an example of a thermal image of the lakes of Erie and Ontario, obtained from TM6. For experimental reasons, TM Band 6 on Landsat is occasionally turned on at night to obtain thermal images. One example, a full scene acquired at 9:32 P.M. on August 22, 1982 shows the familiar east half of Lake Erie and the western part of Lake Ontario. The land appears moderately cool (darker tones), with little detail, although the cities of Buffalo (east tip of Lake Erie), Toronto (top center) and Hamilton, Ontario (west end of Lake Ontario; locally hot because of steel mill effluents) may be discernible on your monitor from street patterns and slightly lighter (warmer) tones. A mottled pattern of variably warmer bands characterizes Lake Ontario; these are related to thermal overturning effects (thermoclines) possible in this deeper (237 m [782 ft]) lake. Lake Erie is uniformly "hot" because its shallowness

(less than 67 m [220 ft]) inhibits this type of circulation. Warm rivers, such as the Niagara connecting the two lakes, stand in contrast to the land.

4.2.8 Retrieval of temperature and emissivity from radiance in thermal images

The surface leaving radiance is the product of the temperature and emissivity effects. Therefore, for a given radiance measurement, there is an infinite number of solutions for temperature and emissivity. The accuracy of the temperature measurement is therefore directly related to the accuracy of the estimate of the target emissivity. Research has shown that emissivities must be known to accuracies of 0.01 or less in order to obtain adequate estimates of temperature.

When measuring temperature over the water, the task is simplified by the fact that the emissivity is well known and spectrally at in the infrared. Thus, it is possible to measure the temperature accurately with a radiometer that has a limited number of broad bands. Because of this, operational systems measuring ocean temperature have been successful for years. When applying these algorithms to land, however, the result is not the same. Unless *a priori* knowledge of the target exists, the emissivity of a particular pixel in a remotely sensed image is unknown.

In many applications, it is the emissivity itself that is of interest. The emissivity can be

used, for instance, in a spectral classi_cation algorithm. The underlying assumption is that a material can be uniquely described by a spectral emissivity curve. This inherent property of the material is then used to classify objects in a scene. To do this, classification algorithms must have a library of spectral curves that correspond to each material of interest. The measured emissivity is then compared to the curves in the library and the curve that matches it the most is selected.

The major challenge with this problem, however, is that the spectral surface radiance consists of N measurements while the unknown spectral emissivity and temperature add up to $N + 1$ unknowns. Therefore, this is an underdetermined mathematical problem (different materials with different kinetic temperatures, can emit the same radiance).

The general radiative transfer equation in the thermal region of the electromagnetic spectrum, is given below (the equation is wavelength dependent):

$$L_{sensor} = \underbrace{e_{earth}*E_{earth}*t/p}_{\text{Surface emission}} + \underbrace{e_{sky}*E_{sky}*r*t/p}_{\text{Surface reflected}} + \underbrace{L_{path}}_{\text{Path thermal}}$$

where,

L_{sensor} = Radiance received at the sensor

e_{earth} = Surface emissivity (usually between 0.7 to 1)

E_{earth} = Emittance of radiance of the surface according to Plank's law (as a black body)

t = Transmittance factor (of the atmosphere)

r = Reflectance factor (of the surface; according to Kirchof's law, equals to $1-e_{earth}$)

e_{sky} = Sky emissivity (usually between 0.5 to 0.75)

E_{sky} = Emittance of radiance from the sky according to Plank's law (as a black body)

L_{path} = Upward Path emitted radiance

~~The contribution of the three components of the radiative transfer equati~~on int the thermal region (between 8.5 to 13mm), to the radiance at the sensor, as modelled by Modtran for sand at 45 degrees, is as follows:

- *surface reflected* – less than 3% - due to the low reflectance spectrum of most materials in the thermal region.
- *path thermal* – between 15% to 50%, increasing with longer wavelengths, as an effect of the summing together of emitted radiance from different atmospheric layers; the higher the layer, the colder it is, the peak of emitted radiance moving towards longer wavelengths (Wein's law).
- *surface emission* – between 50% to 85%, decreasing towards longer wavelength (this depends off coarse upon the temperature of the surface being mointored, and its emissivity).

There are several methods for atmospheric correction in the thermal region, trying to extract both the kinetic temperature of the surface and its emissivity. As these methods are more sophisticated (some already implemented in available softwares), they will not be dealt here, and one is invited to read in the apropriate journals.

4.3 Geometric corrections

- *http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/indexe.html*
- *http://www.ncgia.ucsb.edu/~good/176b/03.html*
  http://www.nr.usu.edu/Geography-Department/rsgis/RSCC/v6.2/6-2.html
- *Domenico Visintini, Digital data stretching techniques, visintini@dgt.uniud.it*
- *Sabins Floyd F. (1976), Remote Sensing – Principles and Interpretation, Freeman*

All remote sensing imagery are inherently subject to geometric distortions. These distortions may be due to several factors, including: the perspective of the sensor optics; the motion of the scanning system; the motion of the platform; the platform altitude, attitude, and velocity; the terrain relief; and, the curvature and rotation of the Earth. Geometric corrections are intended to compensate for these distortions so that the geometric representation of the imagery will be as close as possible to the real world. Many of these variations are **systematic**, or **predictable** in nature and can be accounted for by accurate modeling of the sensor and platform motion and the geometric relationship of the platform with the Earth. Other **unsystematic**, or **random**, errors cannot be modeled and corrected in this way. Therefore, **geometric registration** of the imagery to a known ground coordinate system must be performed.

Satellite images can be acquired at different processing levels regarding their geometric quality; Examples will be given for the Landsat and SPOT satellites. The methods for the geometric correction of satellite images are the ones that will be given first. Images acquired from airborne platform suffer from more severe geometric distortions. The methods for correcting these will be described at the end of this chapter.

Systematic Errors:

Scan Skew: Caused by the forward motion if the platform during the time required for each mirror sweep. The ground swath is not normal to the ground track but is slightly skewed, producing cross-scan geometric distortion

Mirror-Scan Velocity Variance: The mirror scanning rate is usually not constant across a given scan, producing along-scan geometric distortion.

Panoramic Distortion: The ground area imaged is proportional to the tangent of the scan angle rather than to the angle itself. Because data are sampled at regular intervals, this produces along-scan distortion

Platform Velocity: If the speed of the platform changes, the ground track covered by successive mirror scans changes, producing along-track scale distortion

Earth Rotation: Earth rotates as the sensor scans the terrain. This results in a shift of the ground swath being scanned, causing along-scan distortion.

Perspective: For some applications it is desirable to have images represent the projection of points on Earth on a plane tangent to Earth with all projection lines normal to the plan. This introduces along-scan distortion.

Nonsystematic Errors

Altitude Variance: If the sensor platform departs from its normal altitude or the terrain increases in elevation, this produces changes in scale

Platform Attitude: One sensor system axis is usually maintained normal to Earth's surface and the other parallel to the spacecraft's direction of travel. If the sensor departs form this attitude, geometric distortion results.

DISTORTION EVALUATED FROM
TRACKING DATA

DISTORTION EVALUATED FROM
GROUND CONTROL

A. Non systematic
distortions

Earth rotation

Altitude variation

Pitch variation

Spacecraft velocity

Roll variation

Yaw variation

B. Systematic
distortions

Scanner distortion

Mirror angle

Actual velocity

Nominal velocity

time

Mirror velocity
variations

Scan
skew
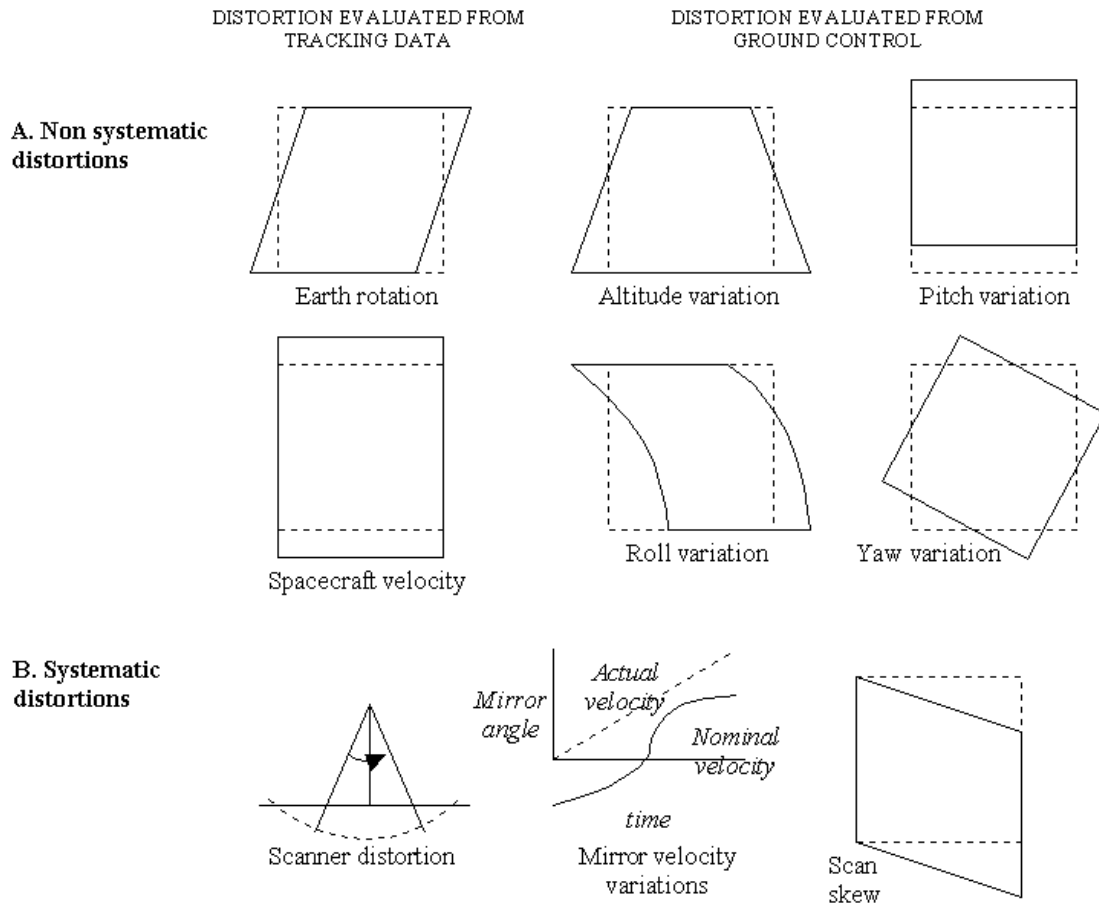
Geometric distortions of Landsat images

4.3.1 Geometric registration

The geometric registration process involves identifying the image coordinates (i.e. row, column) of several clearly discernible points, called **ground control points** (or **GCPs**), in the distorted image (A - A1 to A4), and matching them to their true positions in ground coordinates (e.g. latitude, longitude). The true ground coordinates are typically measured from a map (B - B1 to B4), either in paper or digital format. This is **image-to-map registration**. Once several well-distributed GCP pairs have been identified, the coordinate information is processed by the computer to determine the proper transformation equations to apply to the original (row and column) image coordinates to map them into their new ground coordinates. Geometric registration may also be performed by registering one (or more) images to another image, instead of to geographic coordinates. This is called **image-to-image registration** and is often done prior to performing various image transformation procedures, which involve comparing images from different sensors or dates.

A   B
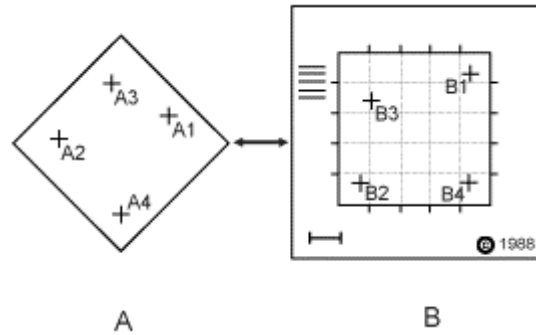
There are several types of transformations to be applied on image coordiantes, so that the image coordinates will transform into real world coordinates:

1. Plane transformations are those which keep lines straight, being on the first order.
2. Curvilinear (polynomial) transformations are higher order transformations that do not necessarily keep lines straight and parallel.
3. Triangulation.
4. Piecewise transformations break the map into regions, apply different transformations in each region.

The above transformations can be calssified as follows:

- They are all *point based* transformations (the Ground Control Points).
- The plane and polynomial transformations are *global interpolators*, determining a single function that applies to the whole region (a change in one pair of coordinates affects the entire image).
  Local interpolators (as the trinagulation, see below) apply an algorithm repeatedly to a small portion of the total set of points.
- *Exact interpolators* honor the data points upon which the interpolation is based, i.e., the GCP's are put exactly where supposed to (in triangulation).
  With *approximate interpolators* there is some uncertainty about the resulting location of the GCP's (in the plane and polynomial transformations). This utilizes the belief that the whole image suffers from a uniform distortion (a global trend).
- They are all *deterministic* methods, not using the theory of probability.
- The plane and polynomial transformations are *gradual interpolators*, while with the triangulation method, *abrupt* changes can be intorduced.
- They all asume the surface to be a plane, having relatively no relief variations. In mountainous areas, georectification of the image needs also the use of a Digital Elevation Model (DEM). – see below the discussion about *relief displacement*, and about *parametric geocoding*.

4.3.1.1 Plane transformations

Let (x,y) be th location of the object in the old coordinate system, and (u,v) the location of the object in the new coordinate system (after the transformation). The four basic types (primitives) of plane transformations keep parallel lines parallel:

1. Translation.
   origin is removed, axes do not rotate.
   $u = x – a; v = y - b$
   origin is moved a units parallel to x and b units parallel to y.

2. Scaling.
    both origin and axes are fixed, scale changes.
    $u = cx; v = dy$
    scaling of x and y may be different, if so, the shape of the object will change.
3. Rotation.
    origin fixed, axes move (rotate about origin).
    $u = x \cos(a) + y \sin(a); v = -x \sin(a) + y \cos(a)$
    a is measured counterclockwise.
4. Reflection.
    coordinate system is reversed, objects appear in mirror image.
    to reverse y, but not x: $u = x; v = c - y$
    this transformation is important for displaying images on video monitors as the default coordinate system has the origin in the upper left corner and coordinates which run across and down.

Usually a combination of these transformations (and some others) will be needed, as the distortion of the image are complex, and cannot be attributed to one of the primitives.

In the following table are given the different kinds of complex plane transformations, with their names, the unknown parameters, the distortion that will be experienced by the image, and the minimum number of ground control points. Up to 6 parameters, parallel lines stay parallel.

Kind of transformation	Deformation of the image	Unknown parameters	Minimum number of GCP's	
Congruency	"roto-translation", preservess shape and scale	3 (E_0, N_0, Q)	2	
Conformality	"roto-translation with (isotropy) scale variation", preserves shape	4 (E_0, N_0, Q, l)	2	
Particular Affinity	"roto-translation with two different scale variation"	5 (E_0, N_0, Q, l, m)	3	
Affinity	"roto translation with two different scale variation and a sliding angle"	6 (E_0, N_0, Q, l, m, g)	3	
Longitudinal / Transversal Homography	"roto-translation with two different scale variation, a sliding angle and a longitudinal / transversal convergence angle"	7 (E_0, N_0, Q, l, m, g, g/h)	4	
Homography	"roto-translation with two different scale variation, a sliding angle and two convergence angles" (projective / perspective transformation)	8 (E_0, N_0, Q, l, m, g, g, h)	4	

E_0 East coordinate of the image frame   g   Transversal convergence angle
N_0 North coordinate of the image frame   h   Longitudinal convergence angle
Q   Rotation angle from the image frame to the world coordinate frame
l   Scale variation from the image frame to the world coordinate frame in x-direction
m   Scale variation from the image frame to the world coordinate frame in y-direction
g   Sliding angle in the world coordinate frame

4.3.1.2 Polynomial Transformations

Simple linear transofrmation equations can be extended to higher powers:

$$u = a + bx + cy + gxy \text{ or } u = a + bx + cy + gx^2 \text{ or } u = a + bx + cy + gx^2 + hy^2 + ixy$$

equations of this form create curved surfaces, thus, straight lines may become curved. The higher the power of the equation, the more will be the minimum number of control points.

They usually give greater accuracy:
- greater accuracy in the sense that when used to transform the control points, the equations faithfully reproduce the known coordinates in the other system (the RMS value will be lower)
- however if error in measurement is present, and it always is to some degree, then greater accuracy may not be desirable
- a polynomial transformation may be more accurate for the control points, but less accurate on average (the image may suffer great distortions).

4.3.1.3 Triangulation

The triangulation method can be seen as a special case of a plane transformation, going on in the following way:

- A Triangulated Irregular Network (TIN) is constructed, one for the GCPs with image coordinates, and one for the GCPs with real world coordinates (connecting the same GCPs in both TIN's).
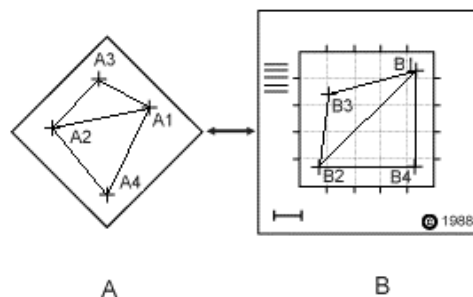
- Using the 3 pairs of coordinates in each triangle, a 6 parametrs *affinity* transformation is performed, a different one for each triangle.
- All GCP's coordinates are honored, the method assuming there is no error in chosing and locating them.
- Abrupt distortions may be observed in the transformed image, if there are errors in locating a GCP, or when adjacent triangles experience different distortions.
- Resampling of pixels will occur only inside the TIN, the areas outside them being "cut off" from the resampled image (unless another transformation is applied on them).

When an image was taken over a hilly area and we don't have the DEM of that area, or when an airborne image (with roll, yaw and pitch distortions) was taken and we don't have the navigation data of the airborne platform, the *Triangulation transformation* might be our best option. However, a very large set of GCP needs to be collected (~100), and the result will be not as accurate as the one that would be

achieved using the navigation data and the DEM (see the discussion about *parametric geocoding* below).

4.3.1.4 Ground Control Points

In order to compute the parameters fot the transformation, a set of Ground Control Points (or *tics*) must be identified. The following rules can be given for selecting the *tics*:

- The minimum number of *tics* should be equal to the number of the unknown parameters. However, the number of *tics* collected at practice, should be higher than the minimum, in order to use the *least squares adjustment method*. The basic idea, is to consider more information (observations) than those are strictly needed to determine the unknowns (as in our tics there are always errors); than, to solve a redundant linear equations system by performing an estimate (of the accuracy of our transformation).
- These control points must not be on a straight line (not collinear).
- Best located to cover the edge, but also include some in the middle if possible.
- Tics should be located much more carefully and accurately than the regular data.
- The following kinds of points may be used as tics:
  - grid intersections and corner points (relevant for scanned maps, not for images)
  - prominent features, e.g. major intersections, depending on scale
  - marks made on the ground for the project, e.g. white crosses for aur photos

To each GCP is calculated the residual value (from the least square estimation), and an overall RMS (Root Mean Square) error value is given. These indicate the accuracy of the analysis. The following points should be noticed:

- the spatial distribution of residuals may indicate weaknesses of the model.
- may show the image has been distorted unevenly
- magnitude of the residuals gives an estimate of the accuracy of the transformation
- the desired RMS value equals the higher of the following:
  - ➢ the spatial resolution of the image
  - ➢ the scale/accuracy of the map being used as the reference for the correction
  - ➢ the spatial accuracy of the real world coordinate reference map for the correction, should be highrt than that of the image
  - for example, if we are working on the PANchromatic band of the SPOT image (spatial resolution of 10 meters), the scale of the map being used for correcting the image should be at least of 1:20,000 , the greater the better (or using DGPS for locating the GCP's). An RMS value of half the pixel size is very good.

Quantifying geocoding quality is difficult. Methods typically applied are:

- calculate the location residuals of ground control points which were not used for the prior calculation,
- compare the image results with the DEM along terrain lines or in specific mountainous areas,
- overlay digital vector or raster maps on the geocoded results.

4.3.1.5 Resmapling

In order to actually geometrically correct the original distorted image, a procedure called **resampling** is used to determine the digital values to place in the new pixel locations of the corrected output image. The resampling process calculates the new pixel values from the original digital pixel values in the uncorrected image. There are three common methods for resampling: **nearest neighbour, bilinear interpolation**, and **cubic convolution**.

Nearest neighbour resampling uses the digital value from the pixel in the original image which is nearest to the new pixel location in the corrected image. This is the simplest method and does not alter the original values, but may result in some pixel values being duplicated while others are lost. This method also tends to result in a disjointed or blocky image appearance.

Nearest neighrbour   *Bilinear interpolation*
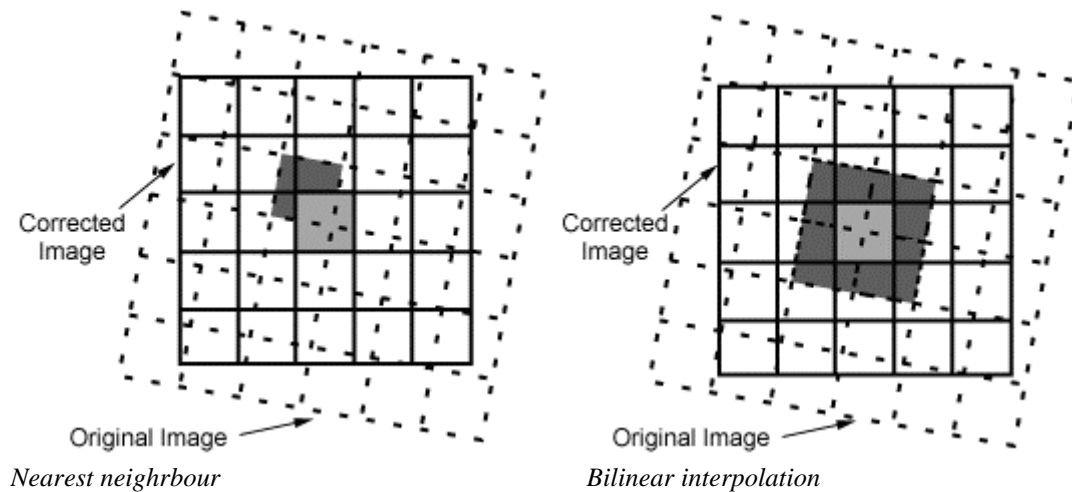
Bilinear interpolation resampling takes a weighted average of four pixels in the original image nearest to the new pixel location. The averaging process alters the original pixel values and creates entirely new digital values in the output image. This may be undesirable if further processing and analysis, such as classification based on spectral response, is to be done. If this is the case, resampling may best be done after the classification process.
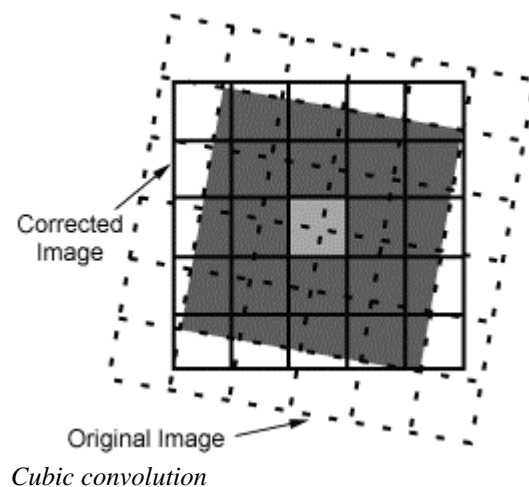
Cubic convolution

Cubic convolution resampling goes even further to calculate a distance weighted average of a block of sixteen pixels from the original image which surround the new output pixel location. As with bilinear interpolation, this method results in completely new pixel values. However, these two methods both produce images which have a much sharper appearance and avoid the blocky appearance of the nearest neighbour method.

4.3.1.6 Relief displacement

- *Antonio Arrighi, Elements of Photogrammetry and Air Photography with notes Photobathymetry, 1st Course on Port and Coastal Hydrography, IMO-IMA*
- *Domenico Visintini, Errori nel Raddrizzamento, visintini@dgt.uniud.it*
- *http://www.gis.psu.edu/geog121/photopo/relief_displacement.html*
- *http://www.ersc.wisc.edu/ersc/Instruct/IES301/slides03/sld013.htm*
- *http://dbwww.essc.psu.edu/lastop/run/las/doc/user/reliefx.html*
- *http://dbwww.essc.psu.edu/lastop/run/las/doc/user/relief2.html*
- *http://geog.hkbu.edu.hk/GEOG3610/Lect-04/sld054.htm*
- *http://dia.dri.edu/rscc/10.4.html*

The scale of an aerial photograph varies with the flying height of the aircraft. Thus, **variations in elevation of objects on the ground cause scale variations in aerial photographs**. The same can be said about images acquired by satellite or by airplanes.

In general, the higher the elevation of an object, the farther the object will be displaced from its actual position away from the *principal point* of the photograph (the point on the ground surface that is directly below the camera lens; in the case of a satellite image, there is no *prinicpal point* – rather, for each scanned line there the pixel located directly below the sensor, at the *nadir*). Conversely, the lower the elevation of an object, the more it will be displaced toward the principal point. This effect, is called **relief displacement.**

Compare the map and photograph below. Both show the same gas pipeline clearing, which passes through hilly terrain. Note the deformation of the pipeline route in the photo relative to the shape of the route on the topographic map. The deformation in the photo is caused by relief displacement.
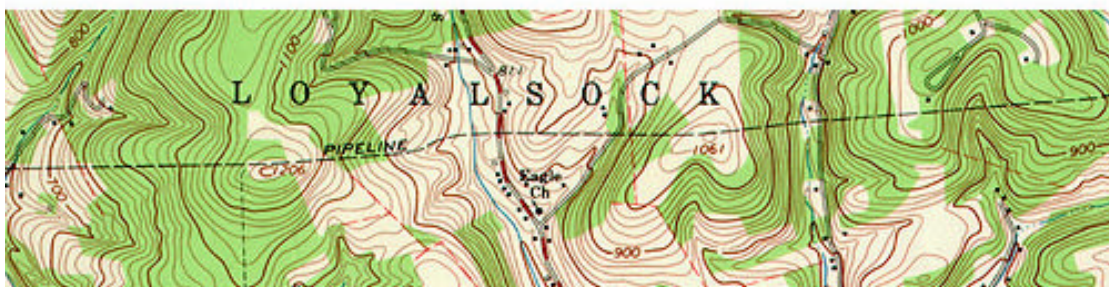
The equation for evaluating relief displacement is as follows:

$$d = r*h / H$$

where d = relief displacement
r = distance on the image from the nadir to the displaced object (d and r are expressed in the same units)
h = height above datum of the object point whose image is displaced
H = flying height above the datum selected for measurement of h

The characteristiscs of relief displacement are as follows:

* d increases with increase in distance from the nadir (r)
* d increases with increase in height of object (h) above datum (selected datum for the image; e.g. the lowest point in the image, or the average height of the image)
* d decreases with increase in flying height (H)
* at the *nadir*, $r = 0$, $d = 0$
* can be either outward (for points above datum) or inward (for points below datum)
* causes objects to lean
* causes straight lines on ground to appear crooked on photo

The following figure describes the basic geometry of relief displacement:
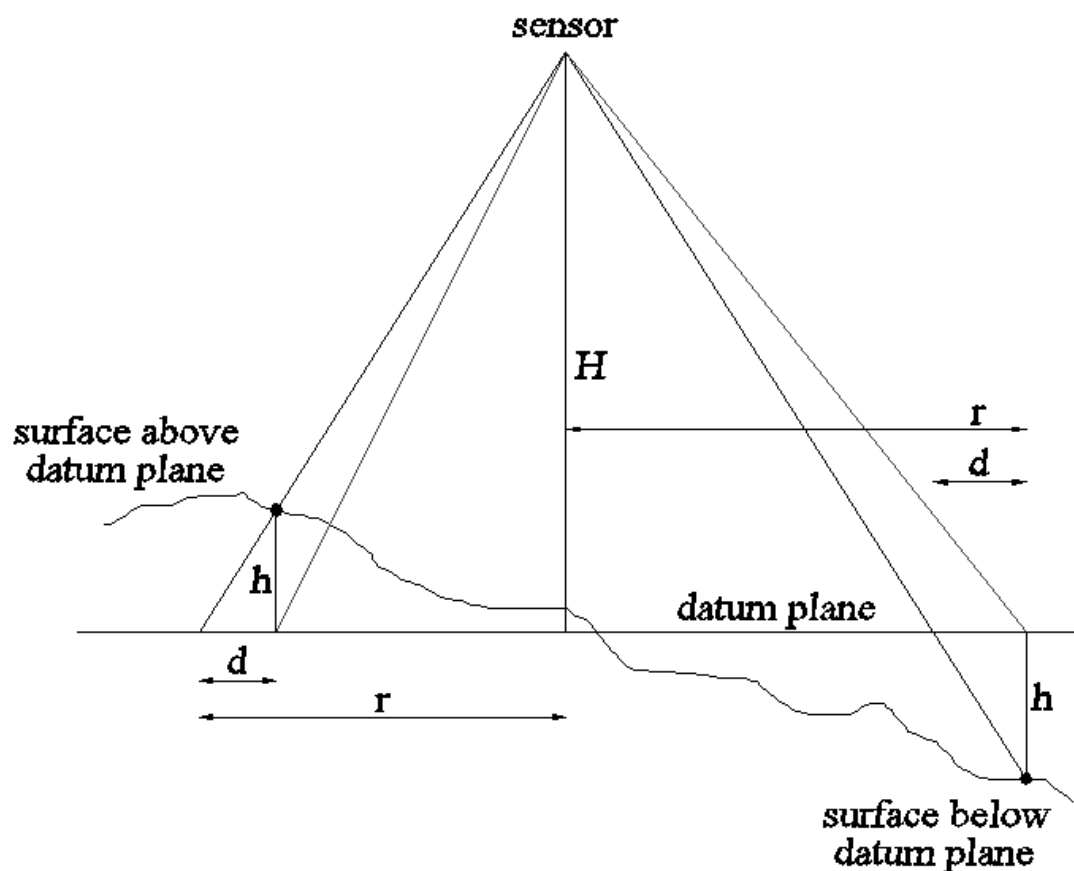
Applying the above equation for two sensors, one on board of a satellite (SPOT), the other airborne (DAIS), gives an example of the effect of relief displacement in Remote Sensing images (in this calculation the curvature of the Earth was not inculded).

Relief displacement as a function of object height above datum and viewing angle, SPOT (822 km above the Earth), ignoring Earth's curvature
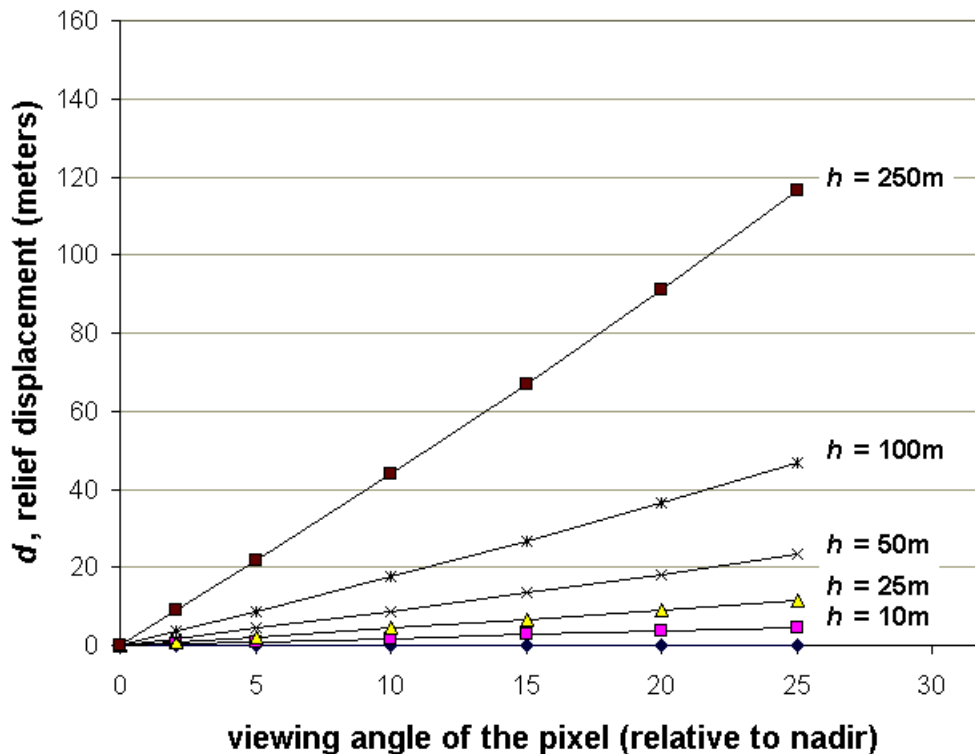
When SPOT is not directed to look obliquely, the far angle at the edge of the image is about 2.06º. It can be seen that for that angle, the effect of relief displacement is smaller than half the pixel size (10m for the mono-chromatic band of SPOT) for objects 100m above or below the image datum (thus, it is negligible). However, if the imaged area is mountainous, or, if the image was taken at some oblique angle, the effects of relief displacement cannot be neglected.

Relief displacement as a function of object height above datum and viewing angle,
DAIS (2 km above the Earth), Earth's curvature irrelevant
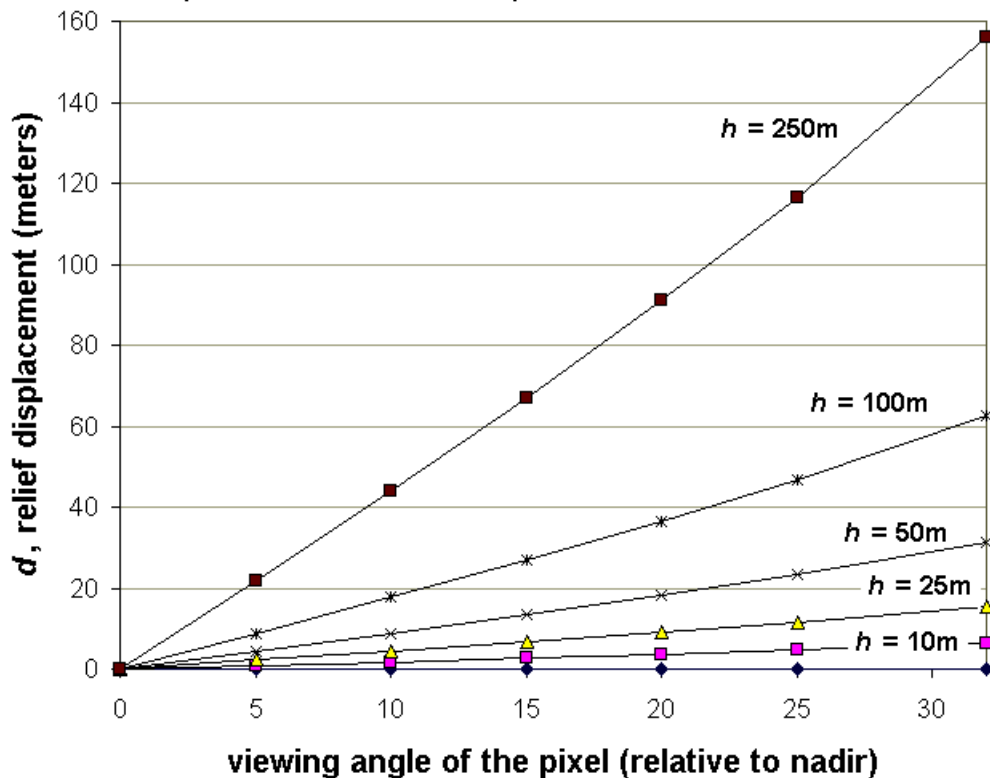
For airborne scanners, these effects of relief displacement are more severe, as the flying height is much lower, and the Field Of View is much larger (the FOV of the DAIS is about 64º, compared to the 4.1º FOV of the SPOT). Thus comes the need for *parametric geocoding*, discussed below.

When performing geometric correction of an image, GCP errors due to relief displacement in control points can be corrected, whenever elevations are available for each point pair. This will be done prior to the calculation of the image transformation from image coordinates to real world coordinates.

First, the image reference height datum needs to be determined. Then, for each GCP, its relief displacement on the image (relative to the *nadir* pixel for each line, for example) should be calculated, and then this should be applied to the image coordinates of the GCP (relief displacement in meters, divided by the pixel size in meters, gives the relief displacement in pixels – thus changing the image coordinates of the GCP in a positive or negative way –according to the relation between the object and the image datum [above or below]).

By correcting the control points, relief is not introduced into the geometric model; it should be remembered however, that relief in the imagery will not be corrected.

4.3.2 LANDSAT - Geometric Characteristics

When buying satellite images, it is often possible to acquire them already geometrically corrected, in different levels of correction. Here will be given an example for this, regarding the LANDSAT satellite.

4.3.2.1 TM Geometric Accuracy

The raw image contains geometric distortions induced by characteristics of the sensor. These characteristics include:

- non-linear mirror scanning velocity
- varying average mirror speed between scans
- sequential detector sampling (rather than "snapshot")
- detector offsets in the focal plane.

The resulting distortions are modelled and corrected during product processing.

Geometric correction removes geometric distortions in an image based on knowledge of the satellite and sensor, and re-maps the image to a regular grid in a standard map projection. This is accomplished by constructing a mapping between pixel coordinates in the image and geographic coordinates on the surface of the Earth.

The intrinsic geometric characteristics of the sensor, plus the knowledge of the precise orbital and attitude parameters at the moment of acquisition, make it possible to geolocate each pixel with an accuracy down to one half of the GSD (15m for Bands 1-5 and 7).

4.3.2.2 TM Data Processing Levels

Three basic levels of geometric processing are available: raw, system corrected or geocoded (levels A, B or C):

4.3.2.3 Raw Data

LANDSAT raw data is available, in a digital format, for full, quarter and mini scenes for European data, for full and sub-scenes for Worldwide data. A raw data product has no radiometric or geometric correction, but the scan lines are re-aligned in the same across-track direction. In addition, Band 6 pixels are replicated (16x) to unify the geometry of all seven bands.
RAW data needs to be processed with specialised equipment. This format is not suitable for inexperienced users

4.3.2.4 System Corrected Products

In addition to the processing of raw data, System Corrected products are corrected for the geometric distortion caused by the Earth's curvature and rotation, the satellite's attitude and the "panoramic distortion" inherent in the scanning geometry. Moreover, radiometric calibration is applied through a look-up table that models the detectors' known non-linearity. System correction uses one of two resampling algorithms (`Cubic Convolution' or `Nearest Neighbour') chosen by the customer when

ordering the product; the defaults are Nearest Neighbour for European data and Cubic Convolution for other data (more information about the algorithms is given below, see "Resampling Algorithm").

As the correction process is aiming to represent the curved Earth's surface in the flat plane of a two-dimensional data array, System Corrected products have to use a map projection and Earth ellipsoid to create a model of the earth's form in order to transform the pixels The built-in map projection of European data products is Space Oblique Mercator (SOM), which was developed specifically for use in LANDSAT images because there is no distortion along the relatively narrow band of the satellite track (path). The Earth ellipsoid is GRS 1980.

10-Year-Old U.S. data products are only available to this level of correction.

4.3.2.5 Geocoded Products

Further corrections are available with European and Worldwide Geocoded products. These are precision processed products which, in addition to being system corrected, are geometrically rectified according to the customer's specifications with respect to map projection, reference ellipsoid and pixel alignment; furthermore any residual striping is suppressed by filtering and other statistical processing operations. Full details of the projections and ellipsoids that can be used are given in the following pages.

Three levels of geocoding are available, providing different degrees of accuracy of the final product, as follows:

4.3.2.6 Level A - `Without Ground Control Points'

The information about the orbit and attitude of the satellite which is downlinked along with the image data, is used to geolocate each pixel to a map projection and a reference ellipsoid which are chosen by the customer. The absolute location accuracy is the same as for System Corrected (Path Corrected) products - about 500 m; however, the internal geometry of the image is improved. The orientation of the output image can be `true north', `along track' or any other specified angle.

4.3.2.7 Level B `With Ground Control Points'

Geometric rectification is carried out using ground control points from maps, or from geodetic or photogrammetric measurements. It is anticipated that maps will be provided by the customer, but Eurimage is able to supply such maps (at cost) if required. The projection and reference ellipsoid of the rectified image will correspond to that of the map unless the customer requests otherwise. The quality of the ground control points directly affects the accuracy of the rectified image. The rms error in the adjusted satellite model is typically half the nominal pixel size, provided that sufficiently accurate control points have been supplied. Terrain displacement errors, which are not corrected by Level B geocoding, are proportional to the instrument viewing angle and the true terrain variation within the image. It should be noted that this correction processing also corrects for non-systematic distortions such as those caused by instantaneous variation of the satellite's pitch and yaw.

4.3.2.8 Level C `With Ground Control Points plus DTM'

In addition to rectification using control points, images are corrected pixel-by-pixel for local terrain displacement errors by utilising a Digital Terrain Model (DTM); it is anticipated that this will be provided by the customer. The result is a satellite orthoimage. The location accuracy is 15m, even in areas of high terrain.

4.3.3 SPOT – processing levels

The raw data received from the satellite undergo preprocessing with the aim of:

- compensating for the instrument distortions,
- generating ancillary data which brings information needed for the use and/or the interpretation of the image data,
- getting the product into its final packaging.

The preprocessing levels range from the least to the most sophisticated level (1A to Ortho), covering a wide spectrum of utilization.

4.3.3.1 Level 1A – no geometric correction

Only detector equalisation is performed: it consists in compensating for the differences of sensitivities between the elementary detectors of the CCD (Charged Coupled Device) arrays, using a linear model. Absolute calibration coefficients posted in the ancillary data can be used to convert the pixel counts into irradiance values.

No geometric corrections are performed, and, when displayed, the SPOT Scene image is a square. Ancillary data (coordinates of the scene centre as well as the four corners) allow to locate the image with an accuracy better than 500m (rms). Other information (ephemeris, attitude, look directions) can be found and used to perform precise geometric processing. When displayed, the SPOT Scene is then a rectangle.

These images are recommended for all people wishing to apply their own geometric processing such as orthorectification or DEM (Digital Elevation Model) generation. Level 1A products are intended specifically for users requiring scene data that have undergone minimum processing. This is essential in cartographic applications, for further precision geometric corrections or for stereo plotting, and in detailed radiometric studies.

4.3.3.2 Level 1B – compensating internal geometric distortions

The same detector equalisation as for level 1A is performed (it consists in compensating for the differences of sensitivities between the elementary detectors of the CCD arrays). Absolute calibration coefficients posted in the ancillary data can be used to convert the pixel counts into irradiances. The geometric corrections are performed with the aim of compensating for the internal distortions of the image caused by the imaging conditions (variations of the attitude of the spacecraft, panoramic effect, earth curvature, rotation of the earth during the imaging of the scene, etc.). The transformation model used for the level 1B is such that two consecutive scenes along a same data strip match together (i.e., there is a perfect registration between the overlapping portions of the two images). The orientation of the processed image is the same as the one of the raw image (no rotation of the lines is performed). When displayed, the SPOT Scene image is a parallelogram. The ancillary data (coordinates of the scene centre and of the four corners, location model) allow to locate any point of the image with an accuracy better than 500m (rms). The distortion of lengths is less than 2×10^{-3} (for a flat terrain).

Suitable for doing geometric measurements (distances, angles and superficies), these images are recommended for people who need a first level of geometric correction.

4.3.3.3 Level 1AP - Photographic product for photogrametric use with analogue devices.

Same detector equalisation as level 1A is performed (compensation from the differences of sensitivities between the elementary detectors of the CCD arrays, using a linear model). In addition a filter enhancing high spatial frequencies is applied in order to highlight linear features on the photographic product.

The geometric distortions are not corrected, with the exception of a stretching of the lines with the purpose of coarsely compensating for the panoramic effect (due to non vertical viewing) for a better comfort of the stereoscopic viewing. When displayed, the SPOT Scene image is a rectangle. Same ancillary data as for level 1A can be delivered on a floppydisk; they allow to locate the image with an accuracy better than 500m (rms), and can be ingested into stereo plotters for geometric modelisation.

This product is designed for analogue photogrammetry (only on photographic films). The set of radiometric and geometric processing was designed specifically to ease the task of stereoscopic restitution of Spot imagery.

4.3.3.4 Level 2A – entry level for cartographic products – projected, no control points

The same detector equalisation as for level 1A is performed (it consists in compensating for the differences of sensitivities between the elementary detectors of the CCD arrays). Absolute calibration coefficients posted in the ancillary data can be used to convert the pixel counts into irradiances. The geometric corrections (geo-rectification) use a re-sampling model which takes into account the compensations for the distortions caused by the imaging conditions, as well as the transformations needed to put the image into the requested cartographic projection (conformal Lambert, UTM, oblique equatorial, polar stereographic, polyconic, etc.). This model is computed on the basis of "system data" known a priori (ephemeris of the spacecraft, attitude motion, etc.), and does not use external information (ground control point, for example). Therefore, the accuracy of the location of any point within the image is the same as for level 1B, i.e. better than 500m (rms). However, this error which is mostly a translation bias, can be reduced by the user with at least one ground control point. The resulting location accuracy depends on the parallax error which comes from the differences of altitude within the landscape, combined with the viewing angle: provided that the terrain is flat and that the image results from a vertical viewing, and after having removed the translation bias, the order of magnitude of the residual location error which has been currently noticed is 50m (rms).

Apart from a translation bias due to the location uncertainty (less than 500m, rms) which can be easily fixed by a user with at least one ground control point, the image is directly superimposable, with the accuracy mentioned herabove, to any other geographic information in the same cartographic projection such as vectors, scanned maps or other satellite images.

Typical users of this product are people wishing to combine geographic information of different types and sources and to apply their own image processing to level 2A images in order to extract specific information; this requires some remote sensing expertise.

4.3.3.5 Level 2B – geocoded product, projected, with ground control points

The same detector equalisation as for level 1A is performed (it consists in compensating for the differences of sensitivities between the elementary detectors of the CCD arrays). For SPOT Scene products, absolute calibration coefficients posted in the ancillary data can be used to convert the pixel counts into irradiances.

The geometric corrections use a re-sampling model which takes into account the compensations for the distortions caused by the imaging conditions, as well as the transformations needed to put the image into the requested cartographic projection (conformal Lambert, UTM, polar stereographic, polyconic, etc.). This is based on a modelisation of the flight dynamic of the spacecraft using "system data" (ephemeris of the satellite, attitude, etc.) plus additional geographic or cartographic data. These extra data are ground control points whose cartographic or geographic coordinates are measured on a map or on the ground (GPS points). This entails a major improvement of the location accuracy of any point within the image which can range from 10 to 30m, depending on the quality of the maps (provided that the landscape is flat).

Additional cartographic information is introduced as ancillary data (map projection, cartographic co-ordinates of the first pixel, etc.). When displayed, the SPOT Scene image is a parallelogram having undergone the rotation needed to match the cartographic reference frame (usually, the image is North oriented).

The availability of maps or GPS points is a pre-requisite for performing a level 2B preprocessing.

These products are designed to be used as digital maps. They provide a global, up-to-date, geographic information to the geographic information community. The products rectified to a precision 2B level can be used everywhere where the relief distortions are not too severe.

4.3.3.6 Level ORTHO - The ultimate level of preprocessing for the best cartographic accuracy: correction of the residual parallax errors brought by the relief.

The same detector equalisation as for level 1A is performed (it consists in compensating for the differences of sensitivities between the elementary detectors of the CCD arrays). For SPOT Scene products, absolute calibration coefficients posted in the ancillary data can be used to convert the pixel counts into irradiances.

The geometric corrections, called "Ortho-rectification", use a re-sampling model which takes into account the compensations for the distortions caused by the imaging conditions, as well as the transformations needed to put the image into the requested cartographic projection (conformal Lambert, UTM, oblique equatorial, polar stereographic, polyconic, etc.). This is based on a modelisation of the flight dynamic of the spacecraft using "system data" (ephemeris of the satellite, attitude, etc.) plus additional geographic or cartographic data. These extra data are ground control

points whose cartographic or geographic co-ordinates are measured on a map or on the ground (GPS points), plus a Digital Elevation Model which is used to fix the parallax error caused by the relief. This entails a major improvement of the location accuracy of any point within the image which can range from 10 to 30m, depending on the quality of the maps, irrespective of the relief and of the viewing angle. Additional cartographic information is introduced as ancillary data (map projection, cartographic co-ordinates of the first pixel, etc.). When displayed, the "SPOT Scene" image is a parallelogram having undergone the rotation needed to match the cartographic reference frame (usually, the image is "North oriented").

The availability of maps (or global positionning system points) and DEM is a pre-requisite for performing a level Ortho preprocessing.

They are thus ideal for mapping relief, offering the same accuracy as level 2B. A DEM of the region of interest must obviously be available to generate level Ortho products.

4.3.4 Parametric Geocoding (based on the PARGE alogrithm)

- Schläpfer D., Meyer P., and Itten K.I., 1998: Parametric Geocoding of AVIRIS Data Using a Ground Control Point Derived Flightpath. Summaries of the Seventh JPL Airborne Earth Science Workshop, published on the Web, JPL, Pasadena (CA), pp. 367-372
- Schläpfer D., Meyer P., and Itten K.I.: PARGE: Parametric Geocoding based on GCP-Calibrated Auxiliary Data,
- these articles and others are available at - *http://www.geo.unizh.ch/~dschlapf/*

- Wiemaker Rafael, Registration of Airborne Scanner Imagery using Akima Local Quintic Polynomial Interpolation

The position of scanning airborne systems (e.g. of the AVIRIS instrument) never is as stable as the behaviour of sensors on spacebrone platoforms. Thus, geometric distortions occur due to variations of the flight path as well as of the attitude (given by roll, pitch and heading angles) of the plane. These distortions can not be corrected by ground control point based traditional georeferencing procedures easily, since the movements can not be approximated satisfyingly by polynomial transformations of the image, due to the non-instantaneous image formation process, unlike the photogrammetric practice:

Photogrammetry which deals with digitized photographic immagery can easily model the geometric image formation process. Photographs are taken with a short shutter time and a center perspective optics, and only one sensor position and attitude is needed for the mapping equations. When terrain effects are neglected, photographic aerial imagery is often geocoded using affine, bilinear and second degree polynomial transformation functions. A small number of GCPs is sufficient in order to determine the necessary number of coefficients; a higher number of GCPs can increase the registration accuracy.

In contrast, multispectral remotely sensed imagery is often recoreded by line scanners which are mounted on airborne platforms. Thus the process of image formation is not instantaneous but depends on the flight path and the attitude of the instrument. Moreover, the mapping itself cannot be described by a lens camera imaging model. Therefore the ortho-rectification of scanner recorded images – especially aerial images from low flight altitudes with high spatial resolution – by conventional *global* coordinate transforms is not satisfactory.

A linewise calculation has to be performed instead, to consider the behaviour of the plane, and to allow for *local* rectification.

For an exact geometric rectification a variety of input data is required. The three categories of input data are:

1. *Navigation data*, consisting of location (longitude, latitude, height – from DGPS) and engineering data (roll, pitch and true heading). This data should be resampled exactly per line or per pixel of the scanner image.
2. *The Digital Elevation Model* has to be in the same coordinate system as the airplane data. The resolution has to be based on the image nominal pixel size. The Digital Elevation Model initiates the final geometry of the geocoded image.
3. *Image general information* consists of exact information on FOV (field of view) and IFOV (instantaneous field of view), scanning frequency, starting time, coordinates of first nadir point, missing lines, and dimensions of the image.
4. *Ground Control Points* are used to determine the uncertain absolute calibration of the airplane attitude angles, average height and position.

The full reconstruction of the geometry of the scanning process, can also provide the required data for the atmospheric correction – the viewing angle per pixel, the absolute distance from the sensor to each pixel location, or the relative air mass between sensor and pixel.

The following steps are performed during the main processing algorithm:

*   Calculate the current observation geometry (following figure); the theoretic look angle vector (L) is calculated between the airplane position and a supposed 'flat' DEM, using the instruments FOV and the pixel position information. This vector is then transformed to the effective look angle (L_t), applying the data from the roll, pitch and true heading.
*   Find the intersection point on the surface; the intersection procedure used in the PARGE algorithm calculates a height profile along the footprint of L_t and searches for a point of equal height L_t.
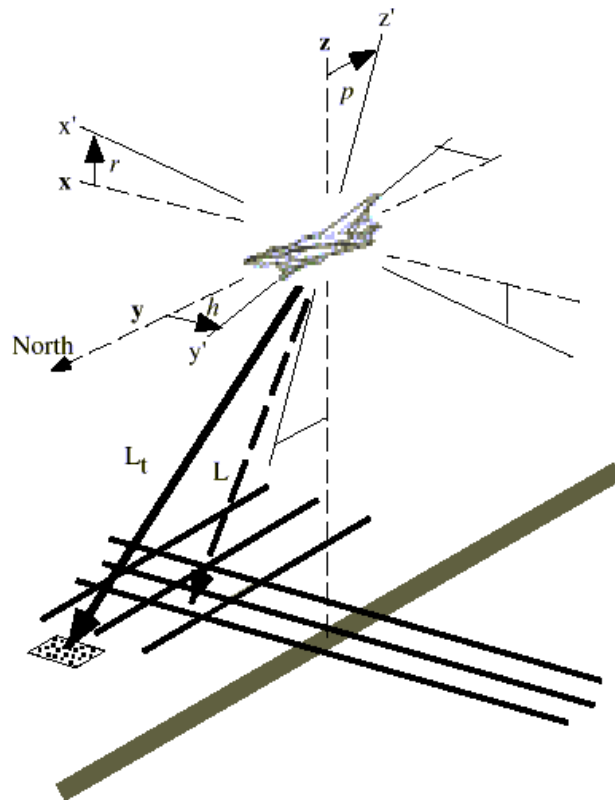
- Map the image coordinates; the pixel coordinates of the image (pixel and line number) are written to an array in DEM geometry at the intersection point position. The result of this procedure is a 'remapping array' which contains the indices of the raw image coordinates, mapped to the correct poistions on the DEM.
- Gap filling; triangulation and nearest neighbour techniques are used to create a spatially continous image (see below).

In order not to lose much information of the raw image, the resolution of the final Digital Elevation Model (and image) has to be taken slightly higher than of the original image data. This results with the creation of gaps, that is, pixels that contain no information from the scanner. In order to fill these gaps, the method prefered by the authors, is the triangulation method, as in the following figure:

Center Pixels   **Growing Method**   **Triangulation Method**

□ remaining gaps
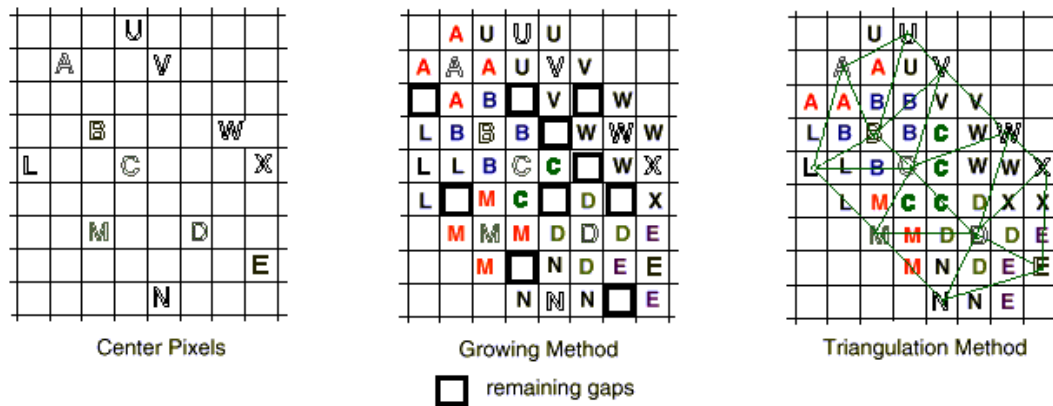
Figure 2: Gap filling methods for slightly oversampled output images. Residual gaps may occur with the fast growing method, whereas the triangulation method covers the area completely.

The flightpath normally is provided with a dataset (of the navigation data). If no path is available, a flightpath reconstruction procedure is applied based on a number of GCP's: the xy-position of the plane is determined for each GCP, and an average flight height is derived from the statistics of additional GCPs. The following assumptions allow such a flightline reconstruction based on a number of GCPs:

- the flight altitude is constant within the required accuracy
- the flight velocity is constant within GCP distances, and
- the flight is more or less straight without any hard turns.

The aircraft position is then calculated using a cubic spline interoplation between the position points.

5. Image Processing:

5.1 STORAGE FORMATS

Imaging Remote Sensing data is organized in a matrix (raster). The *columns* are usually termed as *samples*, and the *rows* as *lines*. As an image scene contains information from several bands, there can be different ways to organize the data. The data is stored as a binary stream of bytes in either Band Sequential (BSQ), Band Interleaved by Pixel (BIP), or Band Interleaved by Line (BIL) format.

- <u>BSQ</u> is the simplest format, with each line of data followed immediately by the next line of the same spectral band. BSQ format is optimal for spatial (X,Y) access to any part of a single spectral band.
- <u>BIP</u> format provides optimal spectral processing performance. Images stored in BIP format have the first pixel for all bands in sequential order, followed by the second pixel for all bands, followed by the third pixel for all bands, etc., interleaved up to the number of pixels. This format provides optimum performance for spectral (Z) access of the image data.
- <u>BIL</u> format provides a compromise in performance between spatial and spectral processing and is the recommended file format for most Remote Sensing processing tasks. Images stored in format have the first line of the first band followed by the first line of the second band, followed by the first line of the third band, interleaved up to the number of bands. Subsequent lines for each band are interleaved in similar fashion.

As an image file contains only pixel values, additional information is needed in order to display it, and to keep track of it. Depending on the software used, this metadata is located either at a separate file (usually an ASCII file with the same name, using a different extension), or in the same file containing the data, before the data. This information is known as *header*, or *documentation*.

Given below is the structure of an image documentation file, as is in the software Idrisi. They may be broken down into four major groups:

Information about the image as a whole:

- <u>title</u>: A descriptive name of the file. It is this text that appears at the bottom of the LIST dialog box when an Image Documentation file is highlighted.
- <u>data type</u>: The type of numbers stored in the file. Allowable entries are byte, integer and real.
- <u>file type</u>: The format in which the Image file is stored. Allowable entries are ASCII, Binary and Packed Binary.

- <u>columns</u>: The number of columns in the image. This is extremely important as it tells IDRISI for Windows modules how to construct the rectangular image from the stored values.
- <u>rows</u>: The number of rows in the image.

Information about the georeferencing system of the file:

- ref. system: The name of the geographic referencing system used with the file. This may be Plane, Lat/Long, or a specific referencing system defined by a Reference System Parameter file.
- ref. units: The unit of measure used in the specified reference system. Allowable entries are m, ft, mi, km, deg and radians.
- unit dist.: The scaling factor between the given coordinates and actual measurements on the ground. This will almost always be 1. The unit distance answers the question, "If I move one unit in the reference system described here, how far have I moved on the ground, measuring in reference units?"

- min X: The minimum X coordinate (left edge) of the image.
- max X: The maximum X coordinate (right edge) of the image.
- min Y: The minimum Y coordinate (bottom edge) of the image.
- max Y: The maximum Y coordinate (top edge) of the image.
- pos'n error: A measure of the accuracy of the positions in the image. This field can be used to record the RMS (Root Mean Square) error of locational positions in the reference system. At present, this field is not analytical, but rather is for informational purposes only.

- resolution: The inherent resolution of the image. In most cases, this should correspond to the result of dividing the range of reference coordinates in X by the number of columns in the image. However, there are some rare instances where it might differ from this result. An example is the case of LANDSAT Band 6 (Thermal) imagery. The resolution of those data is actually 120 meters, and would be recorded as such in this field. However, the data is distributed in an apparent 30 meter format to make them physically match the dimensions of the other bands of the scene. The resolution field is a way of correctly indicating the underlying resolution of these data.

Information about the values stored in the file:

- min value: The minimum value in the image.
- max value: The maximum value in the image.
- value units: The unit of measure of the values in the image. It is suggested that the term classes be used for all qualitative data sets, and that whenever standard linear units are appropriate, that the same abbreviations that are used for reference units should also be used (m, ft, mi, km, deg, rad).
- value error: This field is very important and should be filled out whenever possible. It records the error in the data values that appear in image cells. For qualitative data, this should be recorded as a proportional error. For quantitative data, the value here should be an RMS error figure. For example, for a DEM, an RMS error of 3 would indicate 68% of all values will be within ± 3 meters of the stated elevation, that approximately 95% will be within ± 6 meters, and so on. This field is analytical for some modules (e.g., PCLASS) and it is intended that it will become more so in the future.

- flag value: Any value in the image that is not a data value, but rather has a special meaning. If there is no flag value, this entry should remain blank.
- flag def'n: Definition of the above flag value. The most common data flags are those used to indicate background cells and missing data cells. This field is analytical for some modules (e.g., SURFACE) and will become more so in the future. The key words background and missing data are specifically recognized by some modules. Other terms are only informational in this version. If there is no flag value, this entry should remain blank.

- legend cats: The number of legend categories present. Legend entries are optional. If there is no legend, 0 should be entered in this field. If legend categories do exist, there will be as many lines following as there are legend categories. If no legend categories exist, these lines are not necessary. (See the example below.)

Other information about the file:

The following four entries are optional, and any number of each may be entered at the end of the file, so long as each has the correct term in the 14 character descriptive field to the left. These are all text fields and are included to facilitate complete documentation of the Image file. At present, these last fields are for information only and are not read by IDRISI for Windows modules (although some modules will write them).

- comment: Any additional information about the data may be recorded here.
- lineage: Description of the history by which the values were recorded/derived.
- completeness: The degree to which the values describe the subject matter indicated.
- consistency: The logical consistency of the file.

Note that the lineage, consistency and completeness fields are intended to meet the recommendations of the U.S. National Committee for Digital Cartographic Data Standards (NCDCDS). Along with the pos'n error and value error fields, they provide a means of adhering to the proposed standards for reporting digital cartographic data quality. For further information about the NCDCDS standard, refer to the January 1988 issue of The American Cartographer, 15(1).

Other software header files can have more (e.g. name of the sensor, storage format (BSQ,BIL,BIP), etc) or less metadata information.

Further discussion of this subject is given at the last chapter of this text, under the title of metadata.

5.2 Image enhancment:
- *Ilwis documentation: http://www.itc.nl/ilwis/dev/doc/index.htm*
- *Idrisi on line help*

Image enhancement deals with the procedures of making a raw image more interpretable for a particular application. In this section, commonly used enhancement techniques are described, which improve the visual impact of the raw remotely sensed data on the human eye.

Image enhancement techniques can be classified in many ways. Contrast enhancement, also called global enhancement, transforms the raw data using the statistics computed over the whole data set. Examples are: linear contrast stretch, histogram equalized stretch and piece-wise contrast stretch. Contrary to this, spatial or local enhancement takes into consideration local conditions only and these can vary considerably over an image. Examples are image smoothing and sharpening (These will be dealt here under the section on filters).

5.2.1 Histogram, stretching, colour palettes

5.2.1.1 Contrast enhancement

The objective of this section is to understand the concept of contrast enhancement and to be able to apply commonly used contrast enhancement techniques to improve the

The sensitivity of the on-board sensors of satellites, have been designed in such a way that they record a wide range of brightness characteristics, under a wide range of

utilizes the brightness range of the detectors. The goal of contrast enhancement is to improve the visual interpretability of an image, by increasing the apparent distinction

distinguishing and interpreting spatial features in an image, the eye is rather poor at discriminating the subtle differences in reflectance that characterize such features.

them readily observable.
Contrast stretch is also used to minimize the effect of haze. Scattered light that

objects at the earth surface, is called haze or path radiance. Haze results in overall higher DN values and this additive effect results in a reduction of the contrast in an

blue, and lowest in the infra red range of the electromagnetic spectrum.
Techniques used for a contrast enhancement are: the linear stretching technique and

types the piece-wise linear contrast stretch can be applied.
A computer monitor on which the satellite imagery is displayed is capable of

satellite images, as their digital numbers also vary within the range of 0 to 255. To produce an image of optimal contrast, it is important to utilize the full brightness

In order to decide which type of stretch to use, and to see the effect of a contrast stretch, in addition to just watching the image, it is very usefull to observe the

distribution where the widths of contiguous vertical bars are proportional to class widths of the variable, and the heights of the bars are proportional to the class

5.2.1.1.1 Linear Stretch

linear stretch is the simplest contrast enhancement. A DN value in the low end of the original histogram is assigned to extreme black, and a value at the high end is assigned to extreme white. The remaining pixel values are distributed linearly between these extremes. One drawback of the linear stretch, is that it assigns as many display levels to the

rarely occurring DN values as it does to the frequently occurring values. However, linear contrast stretch, putting (min, max) at (0,255) in most cases still produces a rather dull image. Even though all gray shades of the display are utilized, the bulk of the pixels are displayed in mid gray. This is caused by the more or less normal distribution, within the minimum and maximum values in the tail of the distribution. For this reason it is common to cut off the tails of the distribution at the lower and upper range (usually be defining the size of the tails by their percentage from the total).

5.2.1.1.2 Histogram equalization

The histogram equalization technique is a non-linear stretch. In this method, the DN values are redistributed on the basis of their frequency. More different gray tones are assigned to the frequently occurring DN values of the histogram.

A histogram of the resulting image will thus appear flat -- hence the name. In theory (i.e., Information Theory), this leads to an image that carries the maximum amount of information for any given number of classes. However, this does not imply that the resulting image is more meaningful. In fact, since the nature of the histogram has been altered, you will have lost one of the more informative characteristics of the image. However, in cases where it is difficult to develop a good visual display, this will usually provide an excellent visual result. Histogram equalization is not available for use with integer or real data.

5.2.1.1.3 Piece-wise linear stretch

The piece-wise linear contrast stretch is very similar to the linear contrast stretch, but the linear interpolation of the output values is applied between user defined DN values. This method is useful to enhance only a certain cover type, for example water. The data values for this feature are in the range of 5 to 18, and in order to be able to discriminate as much as possible, it is wise to use all available gray levels for this feature only. In this way detailed differences within the feature of interest appear, where as the remaining features are assigned to a single gray tone.

Following is an example of 3 types of stretching applied to band 1 of a SPOT image that was taken in 1992 (from the exercise files of the Idrisi software). Given are the images and their corresponding histograms.

Original

Linear from 16 to 60

Linear with saturation of 5%

Equalization
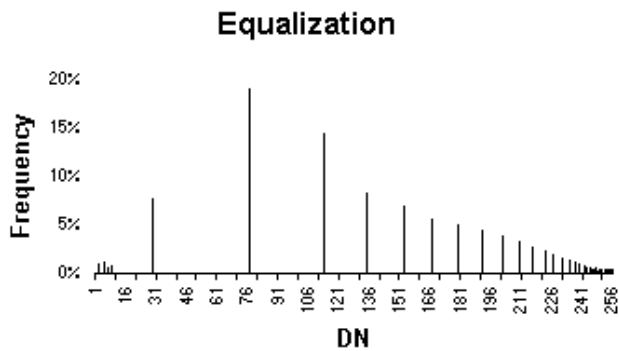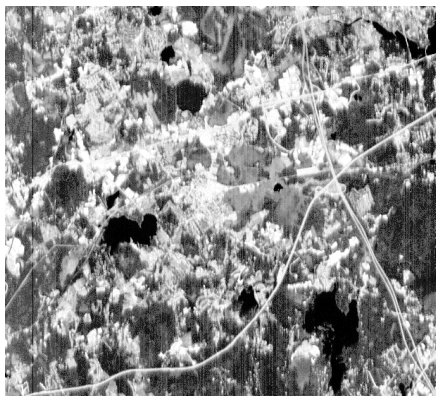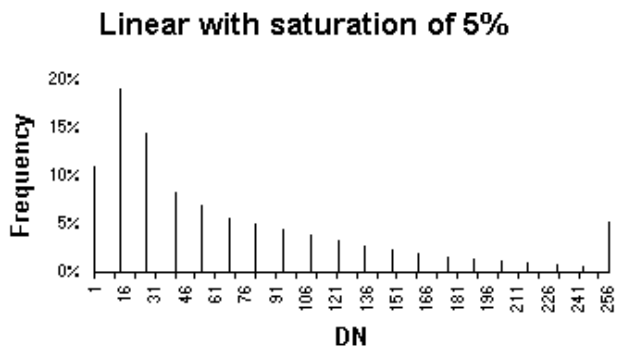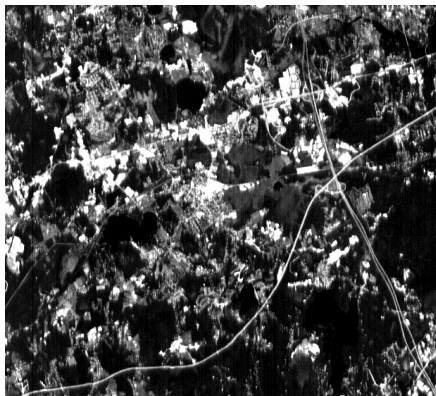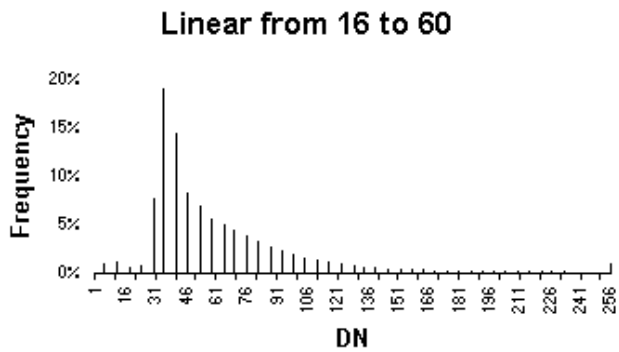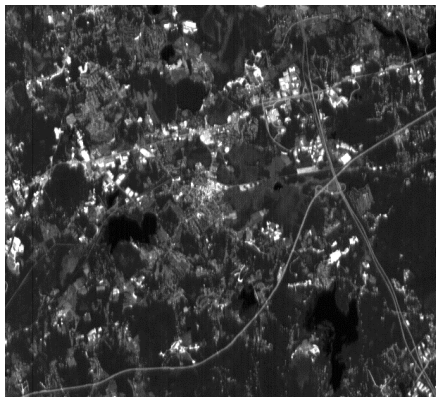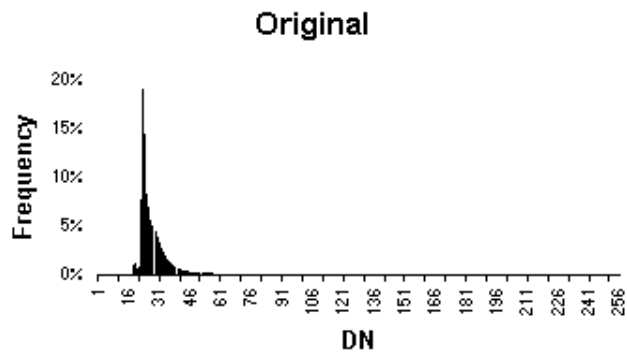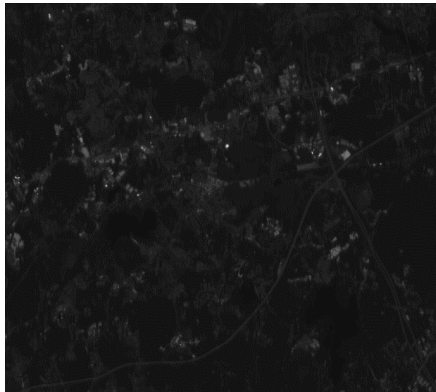
5.2.1.2 RGB false colour composit

In this section, the objective is to understand the concept of color composites and to be able to create different color composites.

The spectral information stored in the separate bands can be integrated by combining them into a color composite. Many combinations of bands are possible. The spectral information is combined by displaying each individual band in one of the three primary colors: Red, Green and Blue.

Every combination of bands used to create a color composite image is called a False Color Composite (FCC). In one of the more common FCC, the red color is assigned to the near-infrared band, the green color to the red visible band and the blue color to the green visible band. The green vegetation will appear reddish, the water bluish and the (bare) soil in shades of brown and gray. For SPOT multi-spectral imagery, the bands 1, 2 and 3 are displayed respectively in blue, green and red (for TM imagery this FCC will be composed of the bands 2, 3 and 4). Another combination used very often for TM imagery is the one that displays in red, green and blue the respective bands 5, 4 and 3. Other band combinations are also possible. Some combinations give a color output that resembles natural colors: water is displayed as blue, (bare) soil as red and vegetation as green. Hence this combination leads to a so-called Pseudo Natural Color composite.

Bands of different images (from different imaging systems or different dates), or layers created by band rationing or principal component analysis, can also be combined using the color composite technique. An example could be the multi-temporal combination of vegetation indices for different dates, or the combination of 2 SPOT XS bands with a SPOT PAN band, (giving in one color image the spectral information of the XS bands combined with the higher spatial resolution of the panchromatic band).

5.2.1.2.1 Colour Definition Models, fusion of TM and SPOT

5.2.1.2.1.1 Colour Definition Models

There are several methods for defining the projected colors that appear on a computer monitor. The following table outlines the Color dialog box red, green, and blue (RGB) and hue, saturation, and lightness (HSL) settings for the standard white light color spectrum.

Color			RGB Settings			HSL Settings		
Red	Green	Blue	Hue	Sat	Light			
Red		255	0	0	0	240	120	
Orange		255	128	0	20	240	120	
Yellow		255	255	0	40	240	120	
Green		0	255	0	80	240	120	
Azure		0	255	255	120	240	120	
Indigo		0	0	255	160	240	120	
Violet		255	0	255	200	240	120	

5.2.1.2.1.2 Red, Green, and Blue (RGB)

The most popular method for defining a projected color is as a combination of red, green, and blue, with values ranging from 0 to 255. For example, pure red is defined

by red = 255 (100%), green =0 (0%), and blue =0 (0%). Pure black has red, green, and blue values of 0%, and pure white has red, green, and blue values of 100%. The RGB model, one of the additive color models, is used on computer monitors.

5.2.1.2.1.3 Hue, Saturation, and Lightness (HSL)
A projected color can be defined by the three components of hue, saturation, and lightness:

Hue describes the color's shade or tint. It is measured on a circular spectrum running from red to green to blue and returning to red.
Saturation describes the hue's purity. A color with 100% saturation is bright and vivid, and a color with 0% saturation is a shade of grey.
Lightness (sometimes termed *value*) describes the color's brightness. A color with 100% lightness is always pure white, and a color with 0% lightness is always pure black.

5.2.1.2.1.4 CMYK Model
The CMYK model, a subtractive color model, is based on light being absorbed and reflected by paint and ink. This model is often used when printing. The primary colors, cyan, magenta, and yellow, are mixed to produce the other colors. When all three are combined, they produce black. Because impurities in the ink make it difficult to produce a true black, a fourth color, black (the K), is added when printing.

5.2.1.2.1.5 Fusion by HSV Transformation
▪ *http://kogs-www.informatik.uni-hamburg.de/PROJECTS/censis/satsim/node4.html*
Bands from different sensors with different spatial resolutions can be fused together, in order to take advantage of the higher spatial resolution of one of them, and the higher spectral resolution of the other.

The most widely applied fusion procedure is the merging of panchromatic SPOT imagery (10 m) with three-color SPOT imagery (20 m) or multispectral LANDSAT TM imagery (30 m). The simplest, most wide-spread and probably most intuitive technique works as follows:

1. Take three spectral bands from the multispectral imagery;
2. Register the low resolution color image to the high resolution panchromatic image (i.e. essentially to magnify the color image to the same pixel size as the panchromatic image);
3. Transform the magnified color image from an RGB color system into the HSV color system (Hue, Saturation, Value);
4. Replace the ``Value'' image by the high resolution panchromatic image;
5. Transform back into the RGB color system.

This technique is known to work well for moderate resolution ratios (such as 1:3 for SPOT + LANDSAT TM). The results are still helpful but less reliable for resolution ratios such as 1:20, e.g. for fusion of SPOT color images with panchromatic aerial photography.

It has to be noted, however, that fusion by HSV transformation can be applied only to multispectral imagery consisting of three bands, since the image has to be coded as an RGB image before fusion can take place.

Another method for fusing images, the *fusion by relative contribution*, will not be discussed here.

5.2.1.3 Spatial filters – noise reduction (low-pass), edge enhancment (high-pass)

- *Cracknell A.P. and Hayes L.W.B. (1993), Introduction to Remote Sensing, pp. 170-173*
- *Jain Anil K. (1989), Fundamentals of Digital Image Processing, pp. 53-54, 244-252, 347-357*
- *PCI (1996), Using PCI software vol. I, pp. 217-226*
- *Schott John R. (1997), Remote Sensing – The Image Chain Approach, pp. 241-248, 347-353*

The operation called filtering is an operation that changes the value of a pixel according to the "gray" values of its surrounding pixels (it is a so called "*focal operation*"). This might be called a spatial stretching operation, unlike a contrast stretch that is operating on the single pixel.

When filtering, we are actually analysing the variance of the pixel values in space, that is, we are dealing with the spatial frequency of an image. Any image can be seen as made of two basic components of information: a high frequency component (high pass) and a low frequency component (low pass), their sum makes up the original image. High frequencies describe sudden\big changes in pixel values along a short distance (linear features: geologic faults, rivers, roads, pipelines, etc or the borders of area features: buildings, etc). Low frequencies describe gradual\small changes in pixel values in space (such as: water bodies, agricultural fields, natural vegetation, etc).

The spatial operation of filtering receives its name due to the fact that it lets only those frequencies in the image that we want to remain. High-pass filters will emphasize spatial change (borders), while low-pass filters will generate an image that will appear smooth or blurred relative to the original (such a filter can be used to eliminate noises).

5.2.1.3.1 The basic principles of a filter

Although there are many kinds of filters that can be applied on an image, the basic principles of their operation are similar:

- The window size of the filter is to be defined, usually having an odd number of columns and rows (the number of the rows equals that of the columns).
- A weight is assigned to each pixel in the window. The pixel values in the window, with the window, are termed "*kernel*".
- The filtered image is obtained by moving the window we have defined, and multiplying each weight in that window, with the value of the corresponding pixel (concerning its location in the window) in the original image. The sumproduct (the sum of the multiplications) is the new value of the pixel in the new image.

Moving the window over the whole image, this operation is preformed on a pixel by pixel basis. Mathematically, this operation is known as "*convolution*". In order that pixels along the perimeter of an image can be a center for the window, the perimeter pixels are duplicated (temporarily) during the filter operation, as described in the following figure:

Original image | With duplicated edges to perform a 3*3 filter | The 3*3 filter window | Output filtered image

12	18	4	0
11	2	12	5
3	23	1	8

12	12	18	4	0	0
12	12	18	4	0	0
11	11	2	12	5	5
3	3	23	1	8	8
3	3	23	1	8	8

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

12	10.3	7	3.3
10.6	9.6	8.1	4.8
9.1	8.8	9.2	6.2
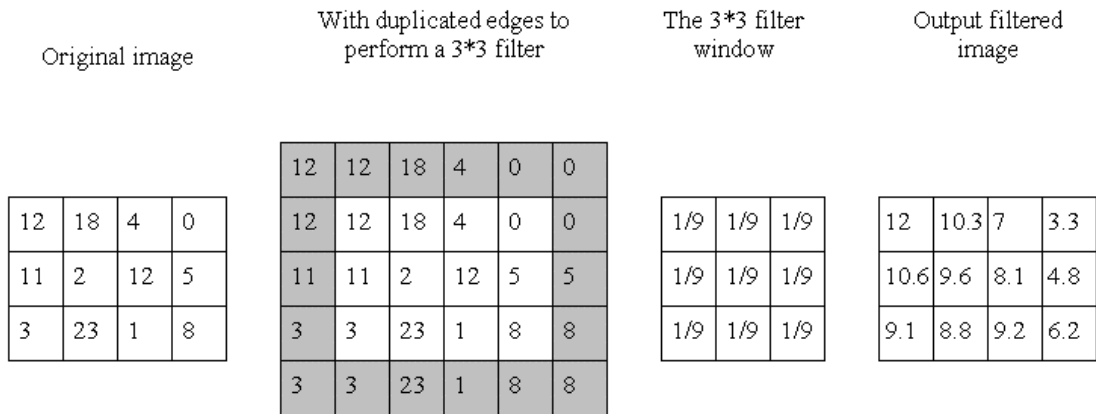
5.2.1.3.2 Low pass filters

Low pass filters remove high frequency features from an image, by diminishing the deviations from the central parameter. The bigger the moving window is, the more blurred will be the output image; sometimes it would be necessary to perform a contrast stretch after that filter, in order to use the full dynamic range of gray values.

These filters will be used mainly to remove noises from an image. Noises in an image can be classified in two according their structure: random noise, and systematic or repetitive noise. Systematic noise is usually due to problems in the sensor. Random noise might occur (for example) on a camera film, when some points on the film itself were damaged.

Such filters can be used also to lower the variability between pixels that belong to the same category, and thus might help in performing a better classification.

The low-pass filters of **MEAN** and **GAUSSIAN** are used to smooth an image. In the **MEAN** filter all pixels have an equal weight, while in the **GAUSSIAN**, the far a pixel is located from the central pixel, its weight is smaller (according to gaussian [bell] distribution); a filter where the weight of a pixel depends on its distance from the central pixel is termed "*distance weighted*" – these filters smooth more gently than "*equal weighted*" filters.

These filters are also applicable in GIS, for example:
- Smoothing a DEM created from contours.
- Computing density from point data.

MEAN 3*3   **GAUSSIAN** 5*5

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

1/121	2/121	3/121	2/121	1/121
2/121	7/121	11/121	7/121	2/121
3/121	11/121	17/121	11/121	3/121
2/121	7/121	11/121	7/121	2/121
1/121	2/121	3/121	2/121	1/121

In the **MEDIAN** and **MODE** filters there are no kernel values, and the value of the new pixel in the filtered output image is just the value of the median or mode of the pixels in the moving window. The **MEDIAN** filter is good for removing random noise. The **MODE** filter is good for filling in empty areas between polygons after a vector-to-raster conversion, or after performing classification in order to remove

isolated pixels that belong to a different category than their neighbors. A **MODE** filter does not create a new value, in the sense that, for example, the data type remains the same. If the values of the input image are integers, the same will be in the output image (notice the first figure presented in this section, where applying the **MEAN** filter changed the values from integer to real).

An example of applying filters will be given using a SONAR image in which we can see a pipeline. The SONAR image is quite noisy (similar to speckles seen on a Side Looking RADAR).

The image was taken from the following site: http://www.kleinsonar.com/imgal.html.
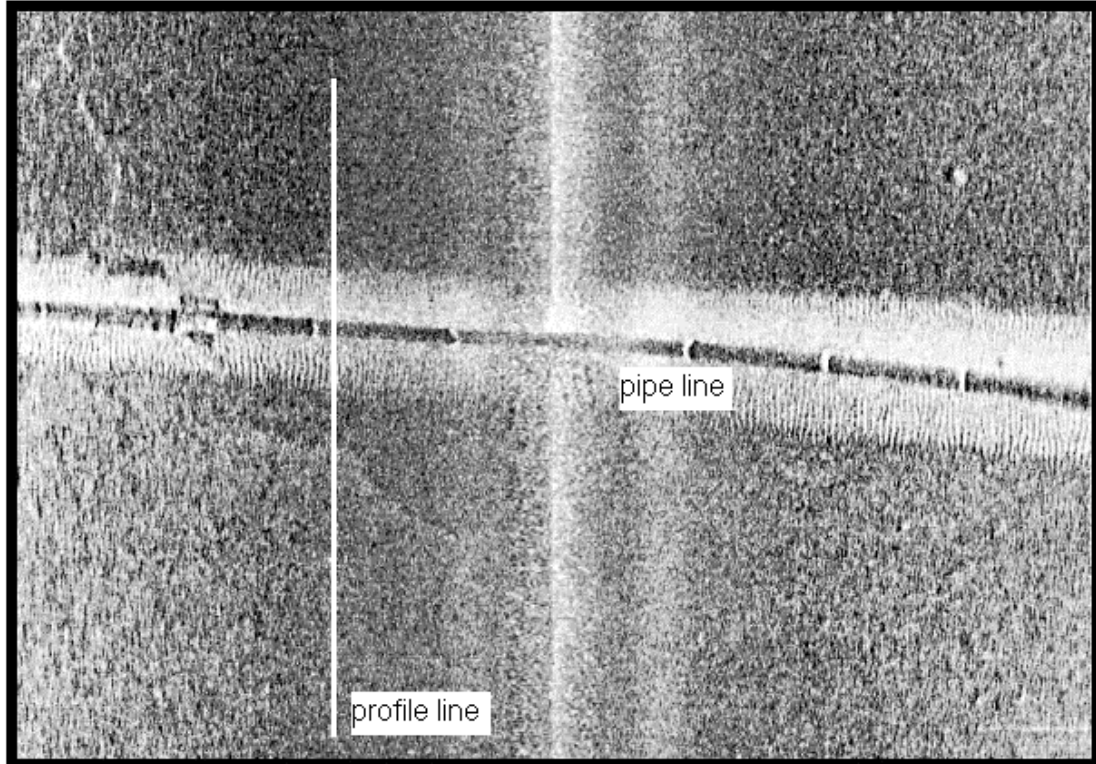
In the following figure the different effect of four different low-pass filters on the image can be seen, comparing them on a vertical profile line.

Low-Pass Filters Comparison
(offset for better view)
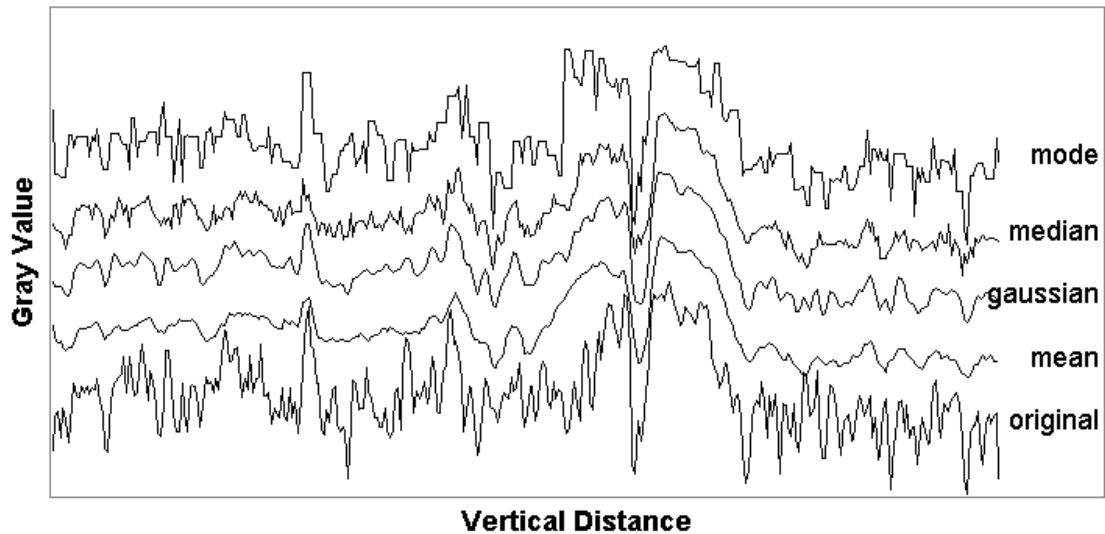
Vertical Distance

Notice the following:

- The pipeline's location on the profile is above the "a" in the X title "Vertical Distance".
- The high-pass features reduced to the maximum with the **MEAN** filter, leaving only low-pass features.
- The **GAUSSIAN** filter is following the original profile better than the **MEAN**, which might create a negative spike where it is actually positive.
- The **MEDIAN** filter leaving many small spikes.
- The **MODE** filter being the most unsuitable here, not removing much of the noise, and even creating grouped spikes.

For visual comparison are given the images produced by the **MEAN** and **MODE** filters:
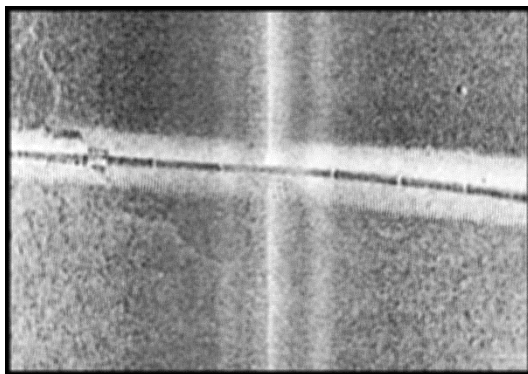
MEAN filtered image
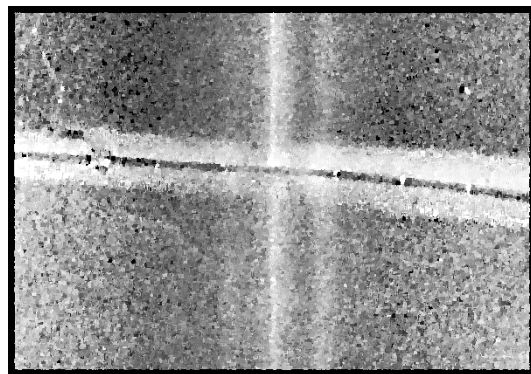
MODE filtered image

In low-pass filters the sum of the kernel values equals 1 (in some software, it is possible to define the kernel values in integer numbers – thus avoiding the need to calculate the exact numbers - and then define that the values should be normalized). Another characteristic of low-pass filters is that they maintain the units of the variable being filtered (unlike high-pass filters, as will be seen).

109

5.2.1.3.3 High pass filters

High pass filters emphasize sudden changes in the values of pixels over distance. Those abrupt differences stand for high spatial frequencies, termed "*edges*". In images, edges will be found in point features (random noise), linear features (road, river) or borders between objects (the border between a field and a forest).

Generally, a high-pass filter will emphasize objects smaller than half the size of the window being used, and blur objects larger than half the window size.

Before examining some high-pass filters, it would be fruitful to know the effect known as "*Mach bands*".

The spatial interaction of the reflected light from an object with its environment, creates a phenomena called the effect of Mach bands. This effect demonstrates that the brightness of an object, is not a simple function of its luminance.
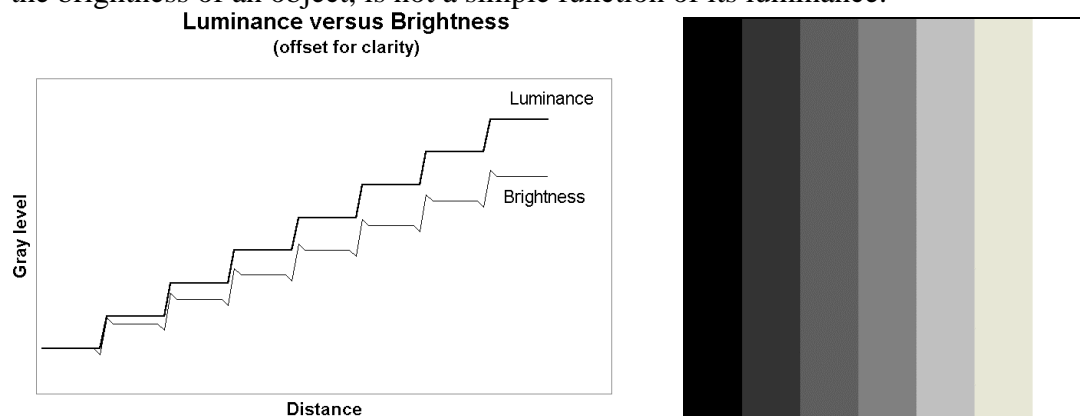
The gray bands in the image have a uniform reflectance, but the apparent brightness is not uniform. On the edge between two gray bands, we can see differences in the brightness value. The thin line in the chart demonstrates that effect, with the overshoots and undershoots. Some of the high-pass filters are actually creating such a phenomena, creating overshoots and undershoots of pixel values on the edges of different objects.

A graphical example of creating a high-pass filter is given in the following figure:
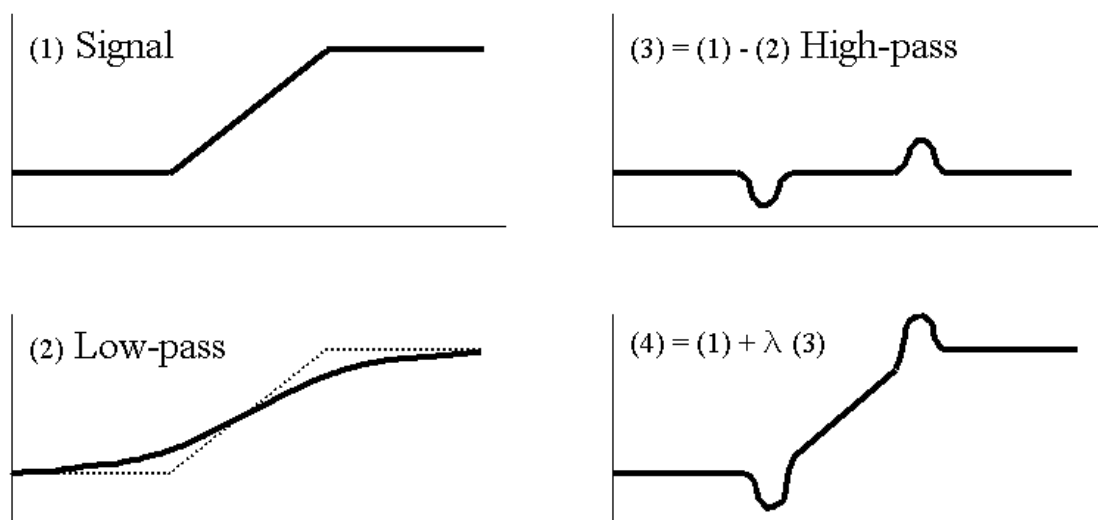
The procedure described in the figure is the most simple high-pass filter. It is based on the above mentioned idea that each image is built of two components, one of a low frequency, the other of a high frequency. By subtracting the low frequency component from the image, we receive the high frequency. This filter is called **HIGH PASS DIFFERENTIAL**.

High-pass filters examine the difference between adjacent pixels. As these differences are usually small, a contrast stretch should be performed on the output filtered image, in order to see better the details. These filters can be used in algorithms for pattern recognition, when the frequency of an "edge" in unit of area can indicate certain objects. The output filtered image can be also used as another band helping in the classification procedure.

One of the most known high-pass filters is the **LAPLACIAN EDGE ENHANCMENT**. Its meaning can be thus understood: We subtract the image from a blurred version of itself created from the averaging of the four nearest neighbors. This enhances edges and isolated pixels with extreme values. However, in the classic version of this filter, very bright values are obtained in dark edges, and dark values in bright edges. This is visually confusing, therefor we can use an alternative, the Modified LaPlacian. There are many other versions of this filter. One of them is given below too, on the right.

TRUE LAPLACIAN   MODIFIED LAPLACIAN

0	1	0
1	-4	1
0	1	0

0	-1	0
-1	4	-1
0	-1	0

-1/9	-1/9	-1/9
-1/9	17/9	-1/9
-1/9	-1/9	-1/9

There are two basic kinds for edge detection operators:
- Gradient operators, working in two orthogonal directions (horizontal and vertical).
- Compass operators, working in a user defined direction.

The basic idea, is to calculate the slope (that is, the difference in pixel values over a distance) of an image, and then to define a certain cutting value above which a pixel is considered as representing an edge point or an edge line.

There are many edge detection filters, one of them is the **SOBEL EDGE DETECTOR**, that is working in the following way:

New value = sqrt ($X^2 + Y^2$) where

X = the resulting image from applying the kernel K_x (below) to the input image
Y = the resulting image from applying the kernel K_y (below) to the input image

Kx =   Ky =

1	0	-1
2	0	-2
1	0	-1

1	2	1
0	0	0
-1	-2	-1

The Sobel edge detector is a gradient operator. Two high-pass filters are being calculated, one in the horizontal direction, the other in the vertical (their respective kernel values are rotated by 90 degrees the one to the other). The output images resulting are raised by the power of 2, and then we calculate the square root of their sum. Other gradient operators work basically in a similar way, only having different kernels (the weights given to the pixels, and the window size used).

Compass operators measure the slope in the requested direction. The resulting slope images are again used to define a "cutting" value, above which pixels are considered as edges. Below are given the 3*3 kernels of the 8 principal directions; in order to calculate directional edge enhancement in other angles, the moving window size should be enlarged.

<table>
<tr><td colspan="3">NW</td><td colspan="3">North</td><td colspan="3">NE</td></tr>
<tr><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>-1</td><td>-2</td><td>1</td><td>1</td><td>-2</td><td>1</td><td>1</td><td>-2</td><td>-1</td></tr>
<tr><td>-1</td><td>-1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>1</td><td>-1</td><td>-1</td></tr>
</table>

<table>
<tr><td colspan="3">West</td><td colspan="3">East</td></tr>
<tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>-1</td></tr>
<tr><td>-1</td><td>-2</td><td>1</td><td>1</td><td>-2</td><td>-1</td></tr>
<tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>-1</td></tr>
</table>

<table>
<tr><td colspan="3">SW</td><td colspan="3">South</td><td colspan="3">SE</td></tr>
<tr><td>-1</td><td>-1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>1</td><td>-1</td><td>-1</td></tr>
<tr><td>-1</td><td>-2</td><td>1</td><td>1</td><td>-2</td><td>1</td><td>1</td><td>-2</td><td>-1</td></tr>
<tr><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
</table>

An example of using a directional edge detector filter will be given, with the same SONAR image of the pipeline, with the chosen direction of North and East (it was preformed on the **MEAN** filtered image, to achieve better results). Notice the easy detection of the pipeline in the North filter.
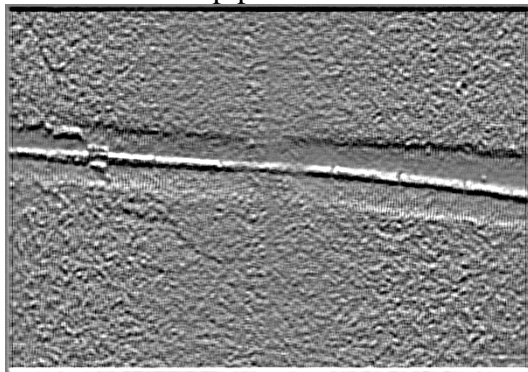
North edge enhancement
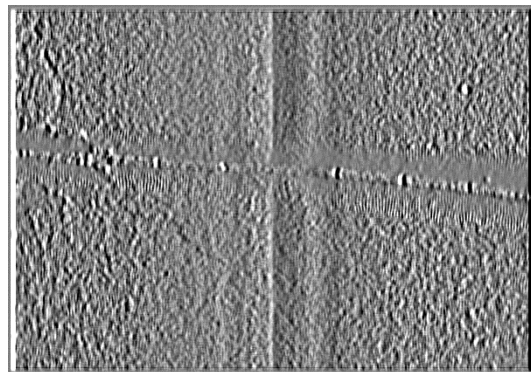
East edge enhancement

In high-pass filters, the sum of the kernel values is usually one, but can be any other number too. An important thing to remember is that the measurement units of the variable (image) being filtered are not maintained (as in contrast stretching), thus, the output image is being used to highlight the phenomena we are interested in, not more.

5.3 Multi-band operations

▪ *Ilwis documentation: http://www.itc.nl/ilwis/dev/doc/index.htm*

To enhance or extract features from satellite images which cannot be clearly detected in a single band, you can use the spectral information of the object recorded in multiple bands. These images may be separate spectral bands from a single multi spectral data set, or they may be individual bands from data sets that have been recorded at different dates or using different sensors. The operations of addition, subtraction, multiplication and division, are performed on two or more co-registered images (see previous exercise on image to image registration) of the same geographical area. This section deals with multi-band operations. The following operations will be treated:

- The use of ratio images to reduce topographic effects.
- Vegetation indexes, some of which are more complex than ratio's only.
- Multi-band statistics.
- Principal components analysis.
- Image algebra, and;
- Image fusion.

5.3.1 Image ratios: Brightness variations

When a satellite passes over an area with relief, it records both shaded and sunlit areas. These variations in scene illumination conditions are illustrated in the figure. A red silt stone bed shows outcrops on both the sunlit and the shadowed side of a ridge. The observed DNs are substantially lower on the shaded side compared to the sunlit areas. This makes it difficult to follow the silt stone bed around the ridge.
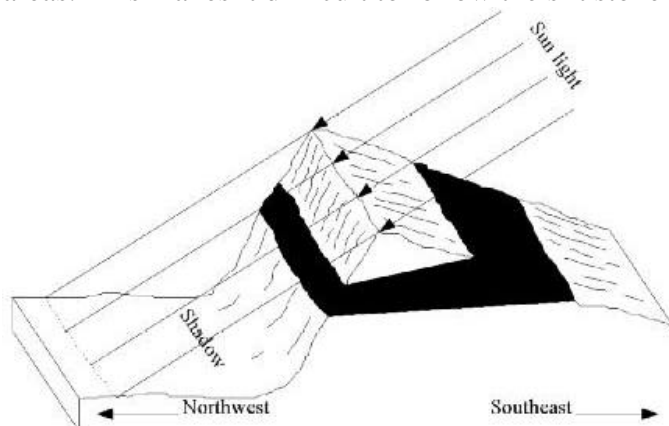
In the individual Landsat-TM bands 3 and 4, the DNs of the silt stone are lower in the shaded than in the sunlit areas. However, the ratio values are nearly identical, irrespective of illumination conditions. Hence, a ratioed image of the scene effectively compensates for the brightness variation, caused by the differences in the topography and emphasizes by the color content of the data.

Table: Differences in DN values of selected bands and the ratio values

	TM Band 3	TM Band 4	Ratio: Band3/Band4
Sunlit slope	94	42	2.24
Shaded slope	76	34	2.23

5.3.2 Normalized Difference Vegetation Index

Ratio images are often useful for discriminating subtle differences in spectral variations, in a scene that is masked by brightness variations. Different band ratios are possible given the number of spectral bands of the satellite image. The utility of any given spectral ratio, depends upon the particular reflectance characteristics of the features involved and the application at hand. For example a near-infrared / red ratio image might be useful for differentiating between areas of stressed and non stressed vegetation.

Various mathematical combinations of satellite bands, have been found to be sensitive indicators of the presence and condition of green vegetation. These band combinations are thus referred to as vegetation indices. Two such indices are the simple vegetation index (VI) and the normalized difference vegetation index (NDVI).

$$NDVI = (NIR - Red) / (NIR + Red)$$

Both are based on the reflectance properties of vegetated areas as compared to clouds, water and snow on the one hand, and rocks and bare soil on the other. Vegetated areas have a relatively high reflection in the near-infrared and a low reflection in the visible range of the spectrum. Clouds, water and snow have larger visual than near-infrared reflectance. Rock and bare soil have similar reflectance in both spectral regions. The effect of calculating VI or the NDVI is clearly demonstrated in next table.

Table: Reflectance versus ratio values

	TM Band 3	TM Band 4	VI	NDVI
Green vegetation	21	142	121	0.74
Water	21	12	-9	-0.27
Bare soil	123	125	2	0.01

It is clearly shown that the discrimination between the 3 land cover types is greatly enhanced by the creation of a vegetation index. Green vegetation yields high values for the index. In contrast, water yield negative values and bare soil gives indices near zero. The NDVI, as a normalized index, is preferred over the VI because the NDVI is also compensating changes in illumination conditions, surface slopes and aspect.

5.3.3 Principal components analysis

Another method (which is not spatial and applied in many fields), called principal components analysis (PCA), can be applied to compact the redundant data into fewer layers. Principal component analysis can be used to transform a set of image bands, as that the new layers (also called components) are not correlated with one another. These new components are a linear combination of the original bands. Because of this, each component carries new information. The components are ordered in terms of the amount of variance explained, the first two or three components will carry most of the real information of the original data set, while the later components describe only the minor variations (sometimes only noise). Therefore, only by keeping the first few components most of the information is kept. These components can be used to generate an RGB color composite, in which component 1 is displayed in red, component 2 and 3 in green and blue respectively. Such an image contains more information than any combination of the three original spectral bands.
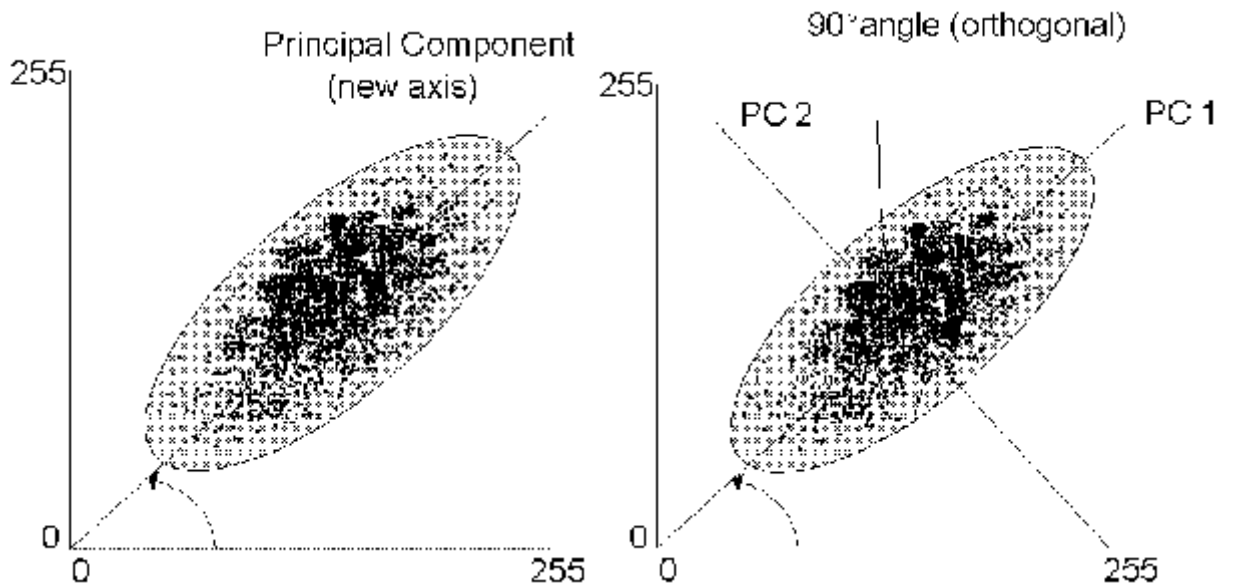
To perform PCA, the axis of the spectral space are rotated, the new axis are parallel to the axis of the ellipse (see figure). The length and the direction of the widest transect of the ellipse are calculated. The transect which corresponds to the major (longest) axis of the ellipse, is called the first principal component of the data. The direction of the first principal component is the first eigenvector, and the variance is given by the first eigenvalue. A new axis of the spectral space is defined by the first principal omponent. The points in the scatter plot are now given new coordinates, which correspond to this new axis. Since in spectral space, the coordinates of the points are the pixel values, new pixel values are derived and stored in the newly created first principal component.

The second principal component is the widest transect of the ellipse that is orthogonal (perpendicular) to the first principal component. As such, PC2 describes the largest amount of variance that has not yet been described by PC1. In a two dimensional space, PC2 corresponds to the minor axis of the ellipse. In n-dimensions there are n principal components and each new component is consisting of the widest transect which is orthogonal to the previous components.

The next figure shows graphically the result of a PC transform in which the data are presented without correlation. Through this type of image transform the relationship with raw image data is lost. The basis is the covariance matrix from which the eigenvectors and eigenvalues are mathematically derived. It should be noted that the covariance values computed are strongly depending on the actual data set or subset used.
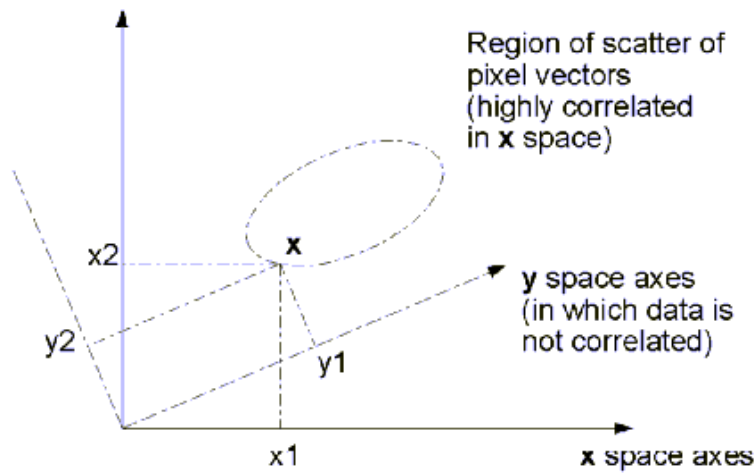
The following figure exemplifies the (possible) information content of the new components, and the ability to filter out some of the noise by performing an inverse PCA transformation, not including the new component containing only noise.
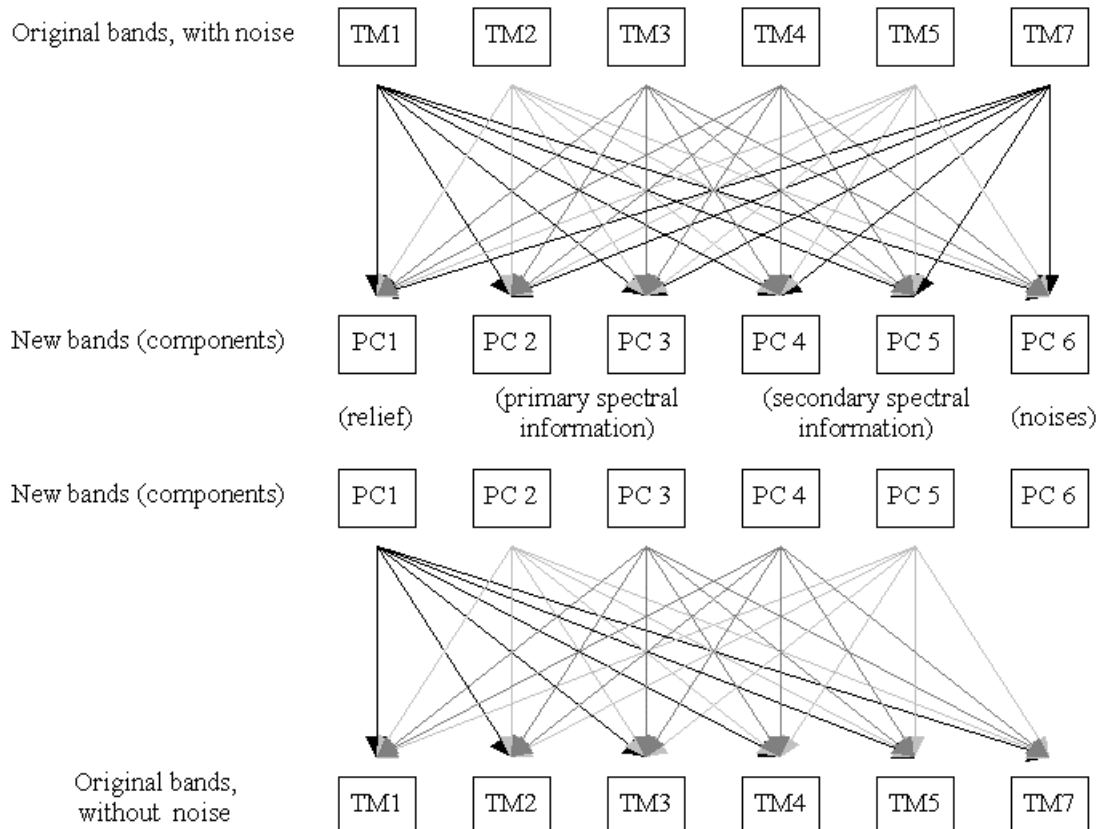
The output of a PCA contains the following tables:
- Variance/covariance matrix between the original bands (variables).
- Correlation matrix between the original bands.
- Component's eigenvalues – the amount of variance explained by each of the new bands (%variance = eigenvalue / S[eigenvalue]).
- Eigenvectors – the parameters for the linear combination of the new bands for an inverse transformation, back to the original bands.
- Component's factor loadings (factor pattern matrix) – factors with a high loading parameter with an original band, have a high correlation with it.

116

5.3.4 Image classification

Measured reflection values in an image depend on the local characteristics of the earth surface; in other words there is a relationship between land cover and measured eflection values. In order to extract information from the image data, this relationship must be found. The process to find the relationship is called classification. Classification can be done using a single band, in a process called density slicing, or using many bands (multi-spectral classification).

5.3.4.1 Density slicing

In theory, it is possible to base a classification on a single spectral band of a remote sensing image, by means of single band classification or density slicing. Density slicing is a technique, whereby the DNs distributed along the horizontal axis of an image histogram, are divided into a series of user-specified intervals or slices. The number of slices and the boundaries between the slices depend on the different land covers in the area. All the DNs falling within a given interval in the input image are then displayed using a single class name in the output map.

Firstly, the ranges of the DN values representing the different land cover types have to be determined. Secondly, a domain Group with the slice boundaries, names and codes has to be created.

By a visual interpretation of the spectral reflectance of the cover classes in the image, differences are observed with regard to their gray tones. Density slicing will only give reasonable results, if the DN values of the cover classes are not overlapping each other, as indicated in the following figure:

Digital range splitted in a number of slices
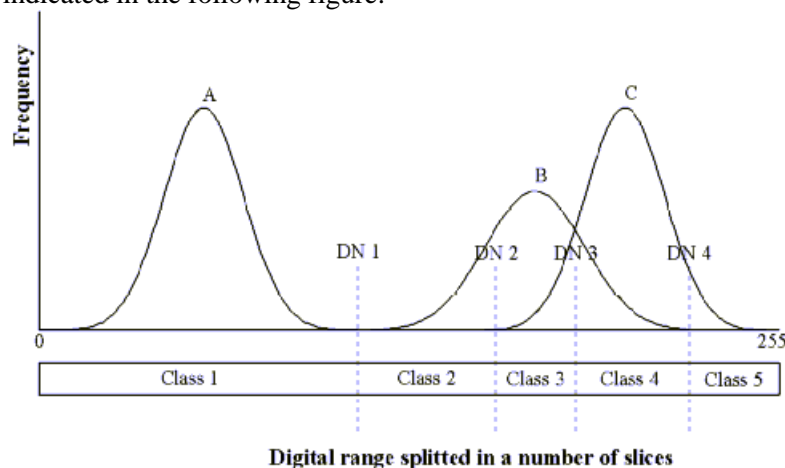
In the figure, a clear distinction is possible between cover type A and cover types B/C (slice with minimum number at position DN 1), as the distribution of the pixels over the digital range is different from cover type B and C. If the distributions of cover types B and C are studied, an overlap can be observed. To discriminate between B and C several options are possible, but all will result in an improper slice assignment. To classify cover class B, the slice DN 1 to DN 2 will include the lower DN values of C and the higher DN values of B are excluded. To include all the pixels belonging to B, the slice assignment is from DN 1 to DN 4. This slice assignment is including even more pixels that belong to cover class C. When trying to classify C, the same problem occurs.

5.3.4.2 Multi-spectral image classification

117

Multi spectral image classification is used to extract thematic information from satellite images in a semi-automatic way. Different methods for image classification exist; some of them are based on the theory about probabilities. Looking at a certain image pixel in M bands simultaneously, M values are observed at the same time. Using multi-spectral SPOT images, where M=3, three reflection values per pixel are given. For instance, (34, 25, 117) in one pixel, in another (34,24,119) and in a third (11, 77, 51). These values found for 1 pixel in several bands are called feature vectors. It can be recognized that the first two sets of values are quite similar and that the third is different from the other two. The first two probably belong to the same (land cover) class and the third belongs to another one.

In classification jargon it is common to call the three bands "*features*". The term features instead of bands is used because it is very usual to apply transformations to the image, prior to classification. They are called "*feature transformations*", their results "*derived features*". Examples are: Principal components, HIS transformations etc.

In one pixel, the values in the (three) features can be regarded as components of a 3-dimensional vector, the feature vector. Such a vector can be plotted in a 3-dimensional space, called feature space. Pixels belonging to the same (land cover) lass and having similar characteristics, end up near to each other in the feature space, regardless of how far they are from each other in the terrain and in the image. All pixels belonging to a certain class will (hopefully) form a cluster in the feature space. Moreover, it is hoped that other pixels, belonging to other classes, fall outside this cluster (but in other clusters, belonging to those other classes).

A large number of classification methods exist. To make some order, the first distinction is between unsupervised and supervised classification. For satellite image applications, the second is generally considered more important. In order to make the classifier work with thematic (instead of spectral) classes, some "*knowledge*" about the relationship between classes and feature vectors must be given.

Theoretically, this could be done from a database in which the relationships between (thematic) classes and feature vectors is stored. It is tempting to assume that in the past, enough images of each kind of sensor have been analyzed, as to know the spectral characteristics of all relevant classes. This would mean, for example, that a pixel with feature vector (44, 32, 81) in a multi spectral SPOT image always means: Grass, whereas (12, 56, 49) is always a forest pixel. Unfortunately, the observed feature vectors in a particular image are influenced by a large amount of other factors than land cover, such as: Atmospheric conditions, sun angle (as function of latitude/time of day/date and as a function of terrain relief), soil types, soil humidity, growing stage (vegetation), wind (affecting orientation of leafs), etc. The problems we meet when trying to take all these influences into account vary from quite easy to practically impossible; at least, vast amounts of additional data (DEM's, soil maps, etc.) would be required.

5.3.4.2.1 Supervised Classification

5.3.4.2.1.1 Sampling

Therefore, classifications methods are much more widely used, where the process is divided into two phases: a training phase, where the user "trains" the computer, by assigning for a limited number of pixels to what classes they belong in this particular image, followed by the decision making phase, where the computer assigns a class label to all (other) image pixels, by looking for each pixel to which of the trained classes this pixel is most similar.

During the training phase, the classes to be use are previously defined. About each class some "ground truth" is needed: A number of places in the image area that are known to belong to that class. This knowledge must have been acquired beforehand, for instance as a result of fieldwork, or from an existing map (assuming that in some areas the class membership has not changed since the map was produced). If the ground truth is available, training samples (small areas or individual pixels) are indicated in the image and the corresponding class names are entered. These training samples are also termed as "*Regions Of Interest*" (ROI).

Guidelines for selecting training areas:

- Training areas should be homogenous. This can be tested by graphic histograms, numeric summaries, 2-band scatter plot for investigating separability of feature classes by pairs of bands, 3-D plot of 3-band feature space (if the software allows!).
- One large 'uniform' training area per feature class is preferable to several smaller training areas, though this must depend upon the degree of variability within each class from site to site, and degree of variability within individual site.
- Easy to extract more than is needed, and then examine site statistics before making decision.
- Each training area should be easily located in the image: use a topographic map, nautical chart, or aerial photos to assist, though differential GPS observations may help.
- If a smaller training area is necessary, then the minimum size is critical. What should be the size of the training site?
  - Note CCRS statement for MSS: individual training area should be minimum of 3 - 4 pixels East-West by 6 pixels North-South.
  - Others [e.g. Swain and Davis, IDRISI] state (10 x # bands used), e.g. area of 40 pixels if all four MSS bands used (or approx 6 pixels x 7 pixels).

5.3.4.2.1.2 Classification

It is the task of the decision making algorithm to make a partitioning of the feature space, according to our training samples. For every possible feature vector in the feature space, the program must decide to which of the sets of training pixels this feature vector is most similar. After that, the program makes an output map where each image pixel is assigned a class label, according to the feature space partitioning. Some algorithms are able to decide that feature vectors in certain parts of the feature space are not similar to any of the trained classes. They assign to those image pixels the class label "unknown". In case the area indeed contains classes that were not included in the training phase, the result "unknown" is probably more realistic than to make a "wild" guess.

In the following scatter plots, the different symbols are used:

C – *corn,* **D** – *deciduous forest,* **S** – *sand,* **U** – *urban,* **W** - *water*

● *- ROI mean,* **1** – *pixel to be classified*
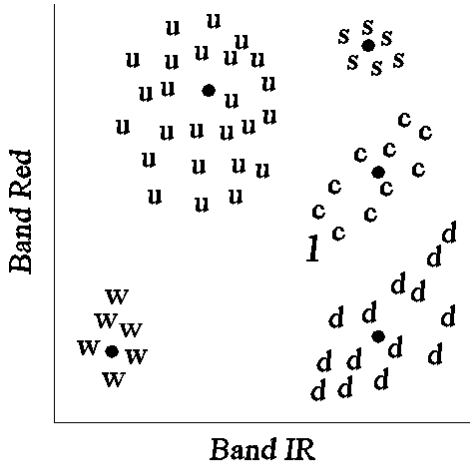
2d scatter plot of ROI's pixels
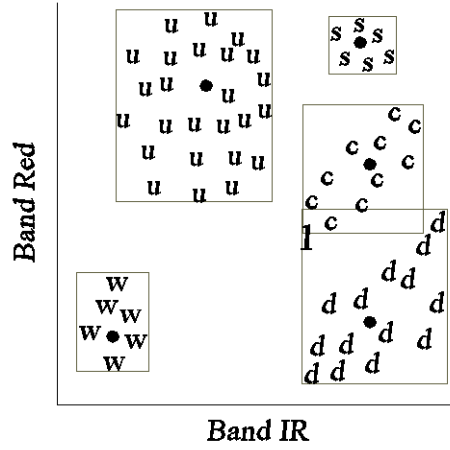
Box / Parallel piped

To find the relationship between classes and feature vectors is not as trivial as it may seem. Therefore, various decision making algorithms are being used; they are different in the way they partition the feature space. Three of them are:

The <u>Box classifier</u> is the simplest classification method: In 2-D space, rectangles are created around the training feature vector for each class; in 3-Dimension they are actually boxes (blocks). The position and sizes of the boxes can be exactly around the feature vectors (Min-Max method), or according to the mean vector (this will be at the center of a box) and the standard deviations of the feature vector, calculated separately per feature (this determines the size of the box in that dimension). In both cases, the user is allowed to change the sizes by entering a multiplication factor. In parts of the features space where boxes overlap, it is usual to give priority to the smallest box. Feature vectors in the image that fall outside all boxes will be "unknown".
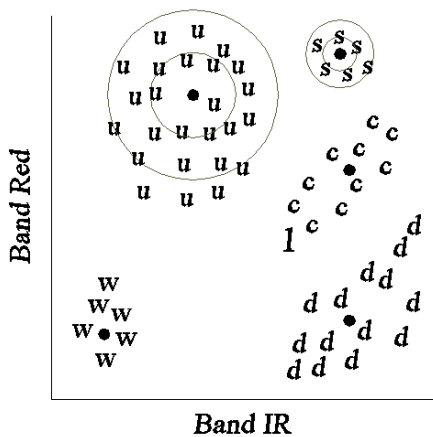
Minimum distance to mean
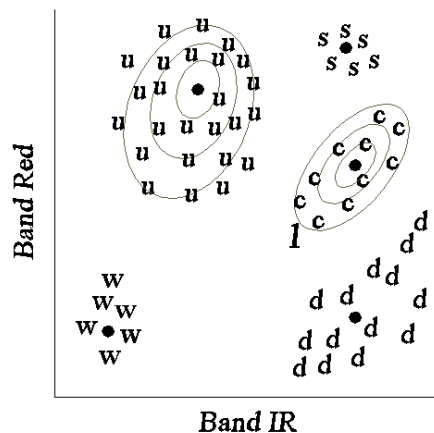
Maximum likelihood

The <u>Minimum Distance-to-mean</u> classifier, first calculates for each class the mean vector of the training feature vectors. Then, the feature space is partitioned by giving to each feature vector the class label of the nearest mean vector, according to Euclidean metric. Usually it is possible to specify a maximum distance threshold: If the nearest mean is still further away than that threshold, it is assumed that none of the classes is similar enough and the result will be "unknown".

<u>Gaussian Maximum Likelihood</u> classifiers assume that the feature vectors of each class are (statistically) distributed according to a multivariate normal probability density function. The training samples are used to estimate the parameters of the distributions. The boundaries between the different partitions in the feature space are placed where the decision changes from one class to another. They are called decision boundaries.

5.3.4.2.1.3 Accuracy Assessment

- *http://priede.bf.lu.lv/GIS/Descriptions/Remote_Sensing/An_Online_Handbook/Sect13/nicktutor_13 -3.shtml*

Accuracy may be *defined*, in a working sense, as *the degree (often as a percentage) of correspondence between observation and reality*. Accuracy is usually judged against existing maps, large scale aerial photos, or field checks. Two fundamental questions about accuracy can be posed: Is each category in a classification really present at the points specified on a map? Are the boundaries separating categories valid as located? Various types of errors diminish accuracy of feature identification and category distribution. Most are made either in measurement or in sampling. Three error types dominate:

1. Data Acquisition Errors: These include sensor performance, stability of the platform, and conditions of viewing. They can be reduced or compensated by making systematic corrections (e.g., by calibrating detector response with on-board light sources generating known radiances). Corrections, often modified by ancillary data such as known atmospheric conditions during overpass, can be made during initial processing of raw data.
2. Data Processing Errors: An example is possible misregistration of equivalent pixels in the different bands of the Landsat Thematic Mapper. The goal in geometric correction is to hold the mismatch to a displacement of no more than 1 pixel. Under ideal conditions, and with as many as 25 ground control points (GCP) spread around a scene, this goal can be realized. Misregistrations of several pixels significantly compromise accuracy.
3. Scene-dependent Errors: One such error relates to how the class is defined and established which, in turn, is sensitive to the resolution of both the observing system and the reference map or photo. Three variants of these are exemplified by a common geologic situation in which the sensor data are processed primarily to recognize rock types at the surface. In this there are pitfalls. First, geologists in the field seek to map bedrock but over large parts of a surface, soil and vegetation cover mask the bedrock at many places. The geologist makes logical deductions in the field as to the rock type most likely buried under the surface and shows this on a map (in which these masking materials are treated as invisible [ignored]). This, unfortunately, does not correspond to what the sensor "sees". Second, most geologic maps are stratigraphic rather than lithologic, i.e., consist of units identified by age rather than rock type. Thus, the same or similar rock types will be shown by different symbols or

colors on the map, so that checking for ground truth requires conversion to lithologies (often difficult because a unit may be diverse lithologically but was chosen for some other mode of uniformity). Third, a rock type may need to be considered in context with its surroundings in order to be properly named. Granite, and the sedimentary rock called arkose derived from it, both have similar spectral properties. The latter, however, typically is arranged in strata, being distributed as a depositional formation, whose spatial patterns (especially when exposed as folded or inclined layers) are usually quite distinct from those of massive granites and are often revealed by topographic expression.

These above points raise the question: Accuracy with respect to what? The maps we use as the standards are largely extrapolations, or more rightly, abstractions. They are often thematic, recording one or more surface types or themes - the signals - but ignoring others -the noise. But, the sensor - whether or not it can resolve them - sees all. When quantifying accuracy, the lack of equivalence and totality must be adjusted for, if possible. Another, often overlooked point about maps as reference standards concerns their intrinsic or absolute accuracy. The map itself requires an independent frame of reference to establish its own validity. For centuries, most maps were constructed without regard to assessment of their inherent accuracy; in recent years, some maps have been published with a statement of level of confidence. The U.S. Geological Survey has reported results of accuracy assessments of the 1:250,000 and 1:1,000,000 land use maps of Level 1 classifications based on aerial photos that meets the 85% accuracy criterion at the 95% confidence level.

As a general rule, the level of accuracy obtainable in a remote sensing classification depends on such diverse factors as the suitability of training sites, the size, shape, distribution, and frequency of occurrence of individual areas assigned to each class, the sensor performance and resolution, and the methods involved in classifying (visual photointerpretation versus computer-aided statistical classifier), and others. A quantitative measure of the mutual role of improved spatial resolution and size of target on decreasing errors appears in this plot:
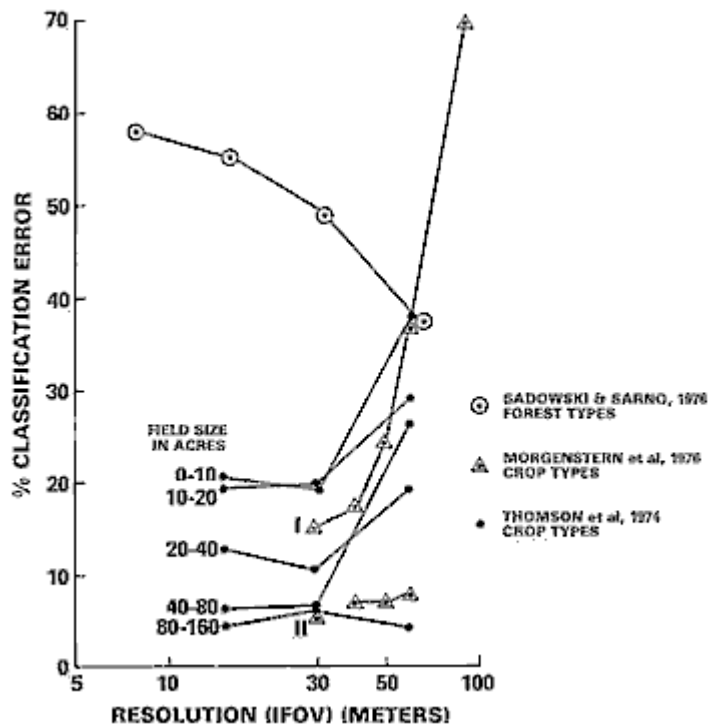
The dramatic improvement in error reduction around 30 m (100 ft) is related in part to the nature of the target classes - coarse resolution is ineffective in distinguishing crop types but high resolution (< 20 m) adds little in recognizing these other than perhaps identifying species. As the size of crop fields increases, the error decreases further. The anomalous trend for forests (maximum error at high resolution) may (?) be the consequence of the dictum: "Can't see the forest for the trees." - here meaning that high resolution begins to display individual species and breaks in the canopy that can confuse the integrity of the class "forest". Two opposing trends influence the behavior of these error curves: 1) statistical variance of the spectral response values decreases whereas 2) the proportion of mixed pixels increases with lower resolution.

A study of classification accuracy as a function of number of spectral bands shows this:

123

CLASSIFICATION ACCURACY
vs
NUMBER OF SPECTRAL BANDS
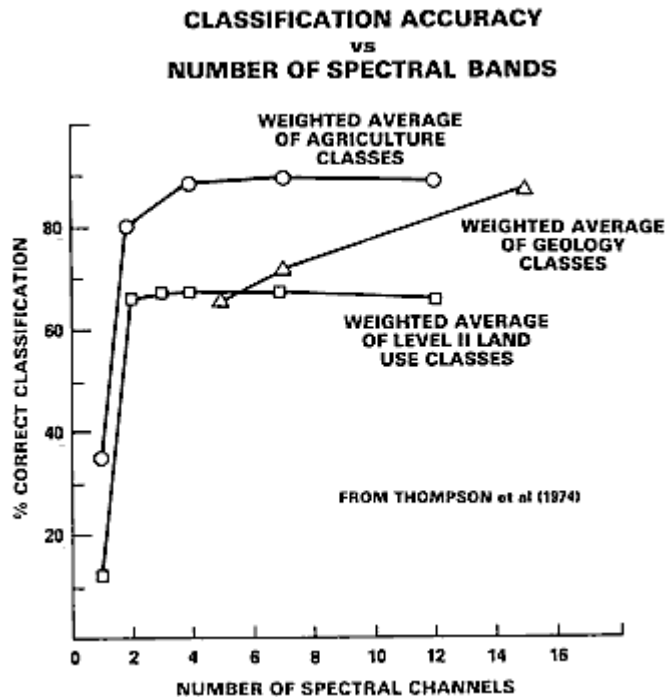
The increase from 1 to 2 bands produces the largest improvement in accuracy. After about 4 bands the accuracy increase flattens or increases very slowly. Thus, extra bands may be redundant, as band-to-band changes are cross-correlated (this correlation may be minimized and even put to advantage through Principal Components Analysis). However, additional bands, such as TM 5 and 7, can be helpful in rock type (geology) identification because compositional absorption features diagnostic of different types reside in these spectral intervals. Note that highest accuracy associates with crop types because fields, consisting of regularly-space rows of plants against a background of soil, tend to be more uniform .

In practice, accuracy of classification may be tested in four ways: 1) field checks at selected points (usually non-rigorous and subjective) chosen either at random or along a grid; 2) estimate (non-rigorous) of agreement on theme or class identify between class map and reference maps, determinedusually by overlaying one on the other(s); 3) statistical analysis (rigorous) of numerical data developed in sampling, measuring, and processing data, using such tests as root mean square, standard error, analysis of variance, correlation coefficients, linear or multiple regression analysis, and Chi-square testing (see any standard text on statistics for explanation of these tests); and 4) confusion matrix calculations (rigorous). This last approach is best explained by the writer's study made of a subscene from a July, 1977 Landsat image that includes Elkton, Maryland.

A 1:24,000 aerial photo that falls within this subscene was acquired from the EPA. Starting with a field visit in August, 1977 during the same growing season as the July overpass, the crops in many individual farms located in the photo were identified, of which about 12 were selected as training sites. Most were either corn or soybeans; others were mainly barley and wheat. A Maximum Likelihood supervised classification was then run.

124

With the class identities in the photo as the standard, the number of pixels correctly assigned to each class and those misassigned to other classes were arranged in the confusion matrix used to produce the summary information shown in the following table, listing errors of commission, omission, and overall accuracies:

Landsat Accuracy Assessment

Photo/Ground Classes	Landsat Classes				Total Possible	Omissions	Commissions	Mapping Accuracy*
	Corn	Soybeans	Forest	Other				
Corn	25	5	10	3	43	$\frac{18}{43} = 42\%$	$\frac{7}{43} = 16\%$	$\frac{25}{25 + 18 + 7} = 50\%$
Soybeans	2	50	6	5	63	$\frac{13}{63} = 21\%$	$\frac{11}{63} = 17\%$	$\frac{50}{50 + 13 + 11} = 68\%$
Forest	3	4	60	5	72	$\frac{12}{72} = 17\%$	$\frac{18}{72} = 25\%$	$\frac{60}{60 + 12 + 18} = 67\%$
Other	2	2	2	100	106	$\frac{6}{106} = 6\%$	$\frac{13}{106} = 12\%$	$\frac{100}{100 + 6 + 13} = 84\%$
Total	32	61	78	113	284			

$$\text{Overall Landsat Classification Accuracy} = \frac{25 + 50 + 60 + 100}{284} = 83\%$$

*Mapping Accuracy (MA) for any Class X:

$$MA = \frac{\text{Pixels of } X_{correct}}{\text{Pixels of } X_{correct} + \text{Pixels of } X_{omission} + \text{Pixels of } X_{commission}}$$

$$\text{Pixels of } X_{omission} = \text{All other classes in X Row}$$

$$\text{Pixels of } X_{commission} = \text{All other pixels in X Column}$$

(OMISSIONS → across; COMMISSIONS ↓ down)

Errors of commission result when pixels associated with a class are incorrectly identified as other classes, or from improperly separating a single class into two or more classes. Errors of omission occur whenever pixels that should have been identified as belonging to a particular class were simply not recognized as present. Mapping accuracy for each class is stated as the number of correctly identified pixels within the total in the displayed area divided by that number plus error pixels of commission and omission. To illustrate, in the table, of the 43 pixels classed as corn by photointerpretation and ground checks, 25 of these were assigned to corn in the Landsat classification, leaving 18/43 = 42% as the error of omission; likewise, of the 43, 7 were improperly identified as other than corn, producing a commission error of 16%. Once these errors are determined by reference to "ground truth", they can be reduced by selecting new training sites and reclassifying, by renaming classes or creating new ones, by combining them, or by using different classifiers. With each set of changes, the classification procedure is iterated until a final level of acceptable accuracy is reached.

It is not recommended to use the same sample map for both the classification and the accuracy assessment, because this will produce figures that are too optimistic.

5.3.4.2.2 Unsupervised classification (clustering)

One way to perform a classification is to plot all pixels (all feature vectors) of the image in a feature space, and then to analyze the feature space and to group the feature vectors into clusters. The name of this process is unsupervised classification. In this process there is no knowledge about "thematic" land cover class names, such as town, road, potatoes etc. All it can do, is to find out that there appears to be (for example) 16 different "things" in the image and give them numbers(1 to 16). Each of these "things" are called "spectral classes". The result can be a raster map, in which each pixel has a class (from 1 to 16), according to the cluster to which the image feature vector of the corresponding pixel belongs. After the process is finished, it is up to the user to find the relationship between spectral and thematic classes. It is very well possible, that it is discovered that one thematic class is split into several spectral ones, or, worse, that several thematic classes ended up in the same cluster.

Various unsupervised classification (clustering) algorithms exist. Usually, they are not completely automatic; the user must specify some parameters such as the number of clusters (approximately) you want to obtain, the maximum cluster size (in the feature space), the minimum distance (also in the feature space), that is allowed between different clusters etc. The process "builds" clusters as it is scanning through the image. Typically, when a cluster becomes larger than the maximum size, it is split into two clusters; on the other hand, when two clusters get nearer to each other than the minimum distance, they are merged into one.

5.3.5 Unmixing

▪ *http://priede.bf.lu.lv/GIS/Descriptions/Remote_Sensing/An_Online_Handbook/TofC/toc1.shtml*

Natural surfaces are rarely composed of a single uniform material. Spectral mixing occurs when materials with different spectral properties are represented by a single image pixel. A variety of factors interact to produce the signal received by the imaging spectrometer:

- A very thin volume of material interacts with incident sunlight. All the materials present in this volume contribute to the total reflected signal.
- Spatial mixing of materials in the area represented by a single pixel result in spectrally mixed reflected signals.
- Variable illumination due to topography (shade) and actual shadow in the area represented by the pixel further modify the reflected signal, basically mixing with a "black" endmember.
- The imaging spectrometer integrates the reflected light from each pixel.

5.3.5.1 Modeling Mixed Spectra

The simplest model of a mixed spectrum is a linear one, in which the spectrum is a linear combination of the "pure" spectra of the materials located in the pixel area, weighted by their fractional abundance.

This simple model can be formalized in three ways: a physical model a mathematical model, and a geometric model. The physical model as discussed above includes the Ground Instantaneous Field of View (GIFOV) of the pixels, the incoming irradiance, the photon-material interactions, and the resulting mixed spectra. A more abstract mathematical model is required to simplify the problem and to allow inversion, or *unmixing* (that is, seperating the "mixed" materials that make the pixel).

A spectral library forms the initial data matrix for the analysis. The ideal spectral library contains endmembers that when linearly combined can form all other spectra. The mathematical model is a simple one. The observed spectrum (a vector) is considered to be the product of multiplying the mixing library of pure endmember spectra (a matrix) by the endmember abundances (a vector). An inverse of the original spectral library matrix is formed by multiplying together the transposes of the orthogonal matrices and the reciprocal values of the diagonal matrix (Boardman, 1989). A simple vector-matrix multiplication between the inverse library matrix and an observed mixed spectrum gives an estimate of the abundance of the library endmembers for the unknown spectrum.

The geometric mixing model provides an alternate, intuitive means to understand spectral mixing. Mixed pixels are visualized as points in n-dimensional scatter-plot space (spectral space), where n is the number of bands. In two dimensions, if only two endmembers mix, then the mixed pixels will fall in a line (Figure a). The pure endmembers will fall at the two ends of the mixing line. If three endmembers mix, then the mixed pixels will fall inside a triangle (Figure b). Mixtures of endmembers "fill in" between the endmembers.

(A).   band i   (B).   band i
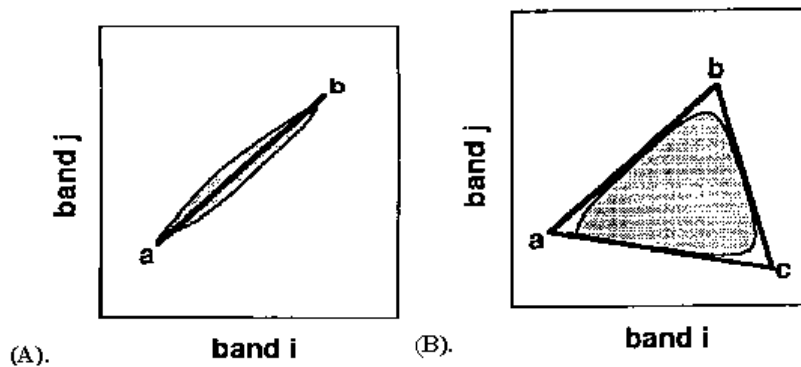
All mixed spectra are "interior" to the pure endmembers, inside the simplex formed by the endmember vertices, because all the abundances are positive and sum to unity. This "convex set" of mixed pixels can be used to determine how many endmembers are present and to estimate their spectra. The geometric model is extensible to higher dimensions where the number of mixing endmembers is one more than the inherent dimensionality of the mixed data.

5.3.5.2 Practical Unmixing Methods

Two very different types of unmixing are typically used: Using "known" endmembers and using "derived" endmembers.

Using known endmembers, one seeks to derive the apparent fractional abundance of each endmember material in each pixel, given a set of "known" or assumed spectral endmembers. These known endmembers can be drawn from the data (averages of regions picked using previous knowledge), drawn from a library of pure materials by interactively browsing through the imaging spectrometer data to determine what pure materials exist in the image, or determined using expert systems (image processing methods not covered here – MNF, PPI, and others) or other routines to identify materials.

The mixing endmember matrix is made up of spectra from the image or a reference library. The problem can be cast in terms of an overdetermined linear least squares problem. The mixing matrix is inverted and multiplied by the observed spectra to get least-squares estimates of the unknown endmember abundance fractions. Constraints can be placed on the solutions to give positive fractions that sum to unity. Shade and shadow are included either implicitly (fractions sum to 1 or less) or explicitly as an endmember (fractions sum to 1).

The second unmixing method uses the imaging spectrometer data themselves to "derive" the mixing endmembers (Boardman and Kruse, 1994). The inherent dimensionality of the data is determined using a special orthogonalization procedure related to principal components:

- A linear sub-space, or "flat" that spans the entire signal in the data is derived.
- The data are projected onto this subspace, lowering the dimensionality of the unmixing and removing most of the noise.
- The convex hull of these projected data is found.
- The data are "shrink-wrapped" by a simplex of n-dimensions, giving estimates of the pure endmembers.
- These derived endmembers must give feasible abundance estimates (positive fractions that sum to unity).

Spectral unmixing is one of the most promising hyperspectral analysis research areas. Analysis procedures using the convex geometry approach already developed for AVIRIS data have produced quantitative mapping results for a a variety of materials (geology, vegetation, oceanography) without a priori knowledge. Combination of the unmixing approach with model-based data calibration and expert system identification capabilities could potentially result in an end-to-end quantitative yet automated analysis methodology.

6. Active Remote Sensing

- *Sabins Floyd F. (1976), Remote Sensing – Principles and Interpretation, Freeman*
- *http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/indexe.html*

An *active* Remote Sensing system supplies its own source of energy, which is directed at the object in order to measure the returned energy. Active microwave sensors are generally divided into two distinct categories: **imaging and non-imaging**. The most common form of imaging active microwave sensors is RADAR. **RADAR** is an acronym for **RA**dio **D**etection **A**nd **R**anging, which essentially characterizes the function and operation of a radar sensor. The sensor transmits a microwave (radio) signal towards the target and detects the backscattered portion of the signal. The strength of the backscattered signal is measured to discriminate between different targets and the time delay between the transmitted and reflected signals determines the distance (or **range**) to the target.

Non-imaging microwave sensors include **altimeters** and **scatterometers.** In most cases these are profiling devices which take measurements in one linear dimension, as opposed to the two-dimensional representation of imaging sensors. Radar altimeters transmit short microwave pulses and measure the round trip time delay to targets to determine their distance from the sensor. Generally altimeters look straight down at nadir below the platform and thus measure height or elevation (if the altitude of the platform is accurately known). Radar altimetry is used on aircraft for altitude determination and on aircraft and satellites for topographic mapping and sea surface height estimation. Scatterometers are also generally non-imaging sensors and are used to make precise quantitative measurements of the amount of energy backscattered from targets. The amount of energy backscattered is dependent on the surface properties (roughness) and the angle at which the microwave energy strikes the target. Scatterometry measurements over ocean surfaces can be used to estimate wind speeds based on the sea surface roughness. Ground-based scatterometers are used extensively to accurately measure the backscatter from various targets in order to characterize different materials and surface types. This is analogous to the concept of spectral reflectance curves in the optical spectrum.

Here will be dealt four active systems which have interesting applications for Hydrography: the Side Looking Airborne Radar (SLAR), radar altimeters, Airborne Laser Scanning (ALS), and also the Global Positioning System (GPS) - being used not for positioning!.

6.1 Side Looking Airborne Radar (SLAR):

- *http://priede.bf.lu.lv/GIS/Descriptions/Remote_Sensing/An_Online_Handbook/Sect8/nicktutor_8-2.shtml*
- *Sabins Floyd F. (1976), Remote Sensing – Principles and Interpretation, Freeman*

Radar is the acronym for Radio Detection and Ranging, indicating that it operates in the radio and microwave bands of the Electromagnetic spectrum ranging from a meter to a few milimeters in wavelength.
RADAR systems can be operated independently of lighting conditions (i.e. also at night) and largely independent of weather. In addition, the terrain can be illuminated in the optimum direction to enhance features of interest.

A typical radar system consists of (1) a pulse generator that discharges timed pulses of microwave/radio energy, to (2) a transmitter, then through (3) a duplexer, to (4) a directional antenna that shapes and focuses each pulse into a stream transmitted to the target; (5) returned pulses are then picked up by the same antenna and sent to a receiver that converts (and amplifies) these into video signals carried to (6) a recording device which can store these digitally for later processing and/or produce a realtime analog display on a cathode ray tube (CRT) or drive a moving light spot to record on film. Each pulse lasts only microseconds (typically there are about 1500 pulses per second); pulse length - an important factor along with band width in setting the system resolution - is the distance traveled during the pulse generation. The duplexer serves to separate the outgoing and returned pulses (i.e., eliminate their mutual interferences) by functioning as an on-off switch that blocks out reception during transmission and vice versa. The antenna on a ground system is generally a parabolic "dish".

RADAR was developed during World War II for navigation and target location, using the familiar rotating antenna and circular cathode-ray display. The conaissance strip-mapping capability of Side-Looking Airborne Radar, or SLAR, was developed in the 1950s to aquire recnnaissance imagery without the necessity of overflying unfriendly regions.

A typical SLAR is composed of an antenna that both transmits the radar pulse and receives the return from the terrain. The bursts of Electromagnetic energy from the transmitter are of specific wavelength and duration, or pulse length. The timing of the returned energy pulse determines the position of terrain features on the image.

A real aperture SLAR system operates with a long (~5-6 m) antenna *usually* shaped as a section of a cylinder wall. This type produces a beam of noncoherent pulses and utilizes its length to obtain the desired resolution (related to angular beamwidth) in the azimuthal (flight line) direction. At any instant the transmitted beam is propagated outward within a fan-shaped plane perpendicular to the flight line.

A second type of system, Synthetic Aperture Radar (SAR), is exclusive to moving platforms. It uses an antenna of much smaller physical dimensions, which sends forth its signals from different positions as the platform advances, simulating a real aperture by integrating the pulse "echoes" into a composite signal. It is possible through appropriate processing to simulate effective antennae lengths up to 100 meters or more. This system depends on the *Doppler* effect to determine azimuth resolution. As coherent pulses transmitted from the radar source reflect from the ground to the advancing platform (air- or spacecraft), the target acts as though in apparent (relative) motion. This motion results in changing frequencies, which give rise to variations in phase and amplitude in the returned pulses. These data are recorded for later processing (employing optical [using coherent laser light] or digital correlation methods), in which the moderated pulses are analyzed and recombined to synthesize signals equivalent to those obtained by a narrow beam, real aperture system.

The pulses of energy transmitted from the antenna iluminate narrow strips of terrain oriented normal to the aircraft flight direction. In the following figure such a strip of terrain and the shape of the energy pulse that it returns to the antenna, are illustrated. The mountain front facing the antenna has a strong return because of its orientation with respect to the radar antenna. The mountain blocks the transmitted pulse from the terrain immediately downrange, and there is no return from that terrain. The resulting dark signature on the image is called a radar shadow. Because of the diverse shapes and orientation, vegetation produces a speckled signature of intermediate intensity. Metallic objects, such as the bridge, produce very strong returns and bright signatures because of their geometry and electrical properties. Radar energy striking calm water is almost totally reflected with the angle of reflection equal and opposite to the angle of incidence; very little energy is returned to the antenna and a dark signature results. Smooth surfaces such as calm water, dry lake beds, highways and airport runways are called specular targets because of their mirror reflection of radar energy. When these surfaces are located in the extreme near range position, a bright return can result; at other locations they produce dark returns.
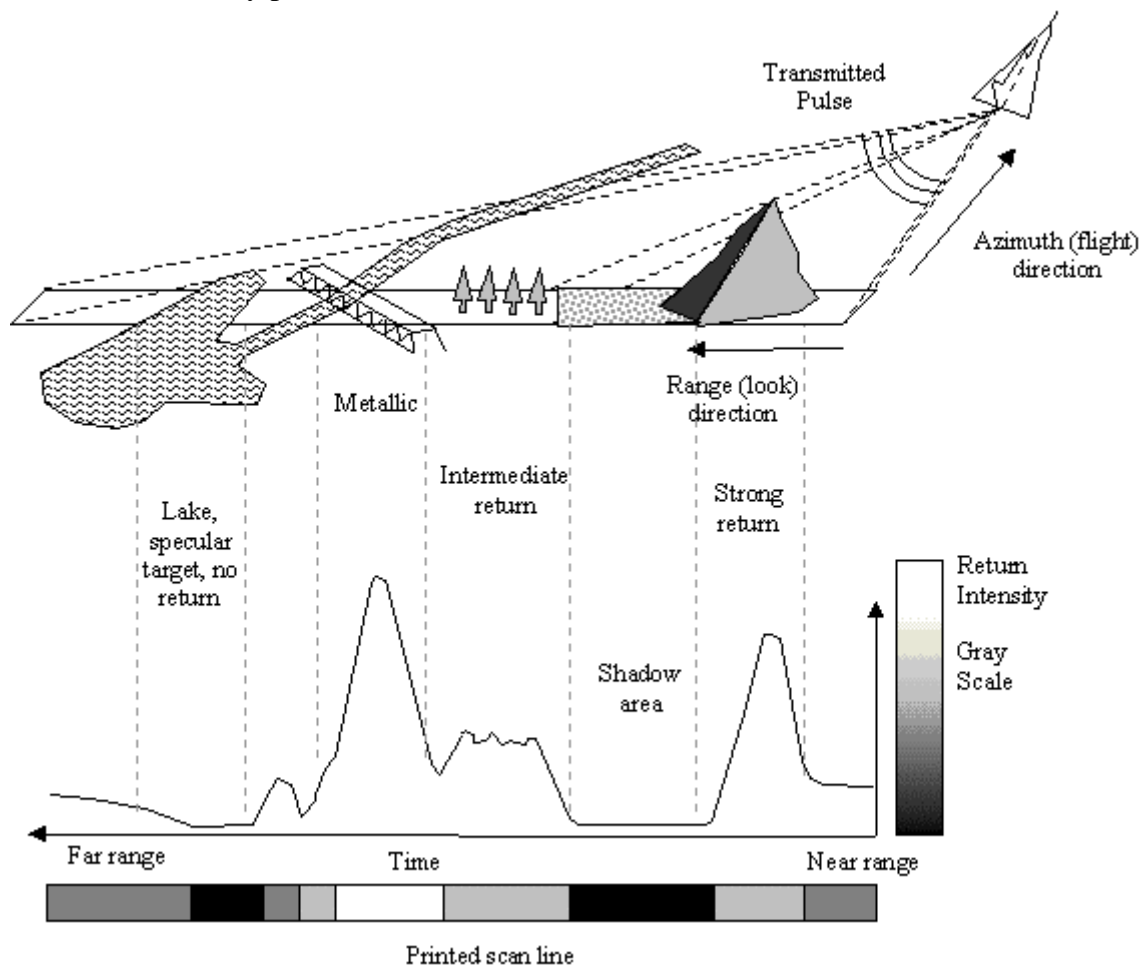
The return pulse is converted to scan line by assigning the the darkest tones of a gray scale to the lower intensity returns and the lightest tones of a gray scale to the highest intensity returns. In addition to being an active system, radar differs from other Remote Sensing because data are recorded on the basis of time rather than angular distance. Time can be much more precisely measured and recorded than angular distance; hence, radar imagery can be acquired at long ranges with high resolution.

Also, atmospheric absorption and scattering are minimal at most microwave wavelengths.

6.1.1 Frequencies

- *http://priede.bf.lu.lv/GIS/Descriptions/Remote_Sensing/An_Online_Handbook/Sect8/nicktutor_8-1.shtml*
- *http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/indexe.html*

Radar operates in part of the microwave region of the electromagnetic spectrum, specifically over the frequency interval from 40000 to 300 megahertz (MHz), the latter extending just into the higher frequency end of the radio (broadcast) region. Commonly used frequencies and their corresponding wavelengths are specified by a band nomenclature (these code letters were given during World War II, and remain to this day), as follows:

- K-alpha Band: 40000-26000 MHz (0.8 - 1.1 cm); K Band: 26500-18500 MHz (1.1 - 1.7 cm) - very short wavelengths used in early airborne radar systems but uncommon today.
- X Band: 12500-8000 MHz (2.4 - 3.8 cm) - used extensively on airborne systems for military reconnaissance and terrain mapping.
- C Band: 8000-4000 MHz (3.8 - 7.5 cm) - common on many airborne research systems (CCRS Convair-580 and NASA AirSAR) and spaceborne systems (including ERS-1 and 2 and RADARSAT).
- S Band: 4000-2000 MHz (7.5 – 15.0 cm) - used on board the Russian ALMAZ satellite.
- L Band: 2000-1000 MHz (15.0 - 30.0 cm) - used onboard American SEASAT and Japanese JERS-1 satellites and NASA airborne system.
- P Band: 1000- 300 Mhz (30.0 - 100.0 cm) - longest radar wavelengths, used on NASA experimental airborne research system.
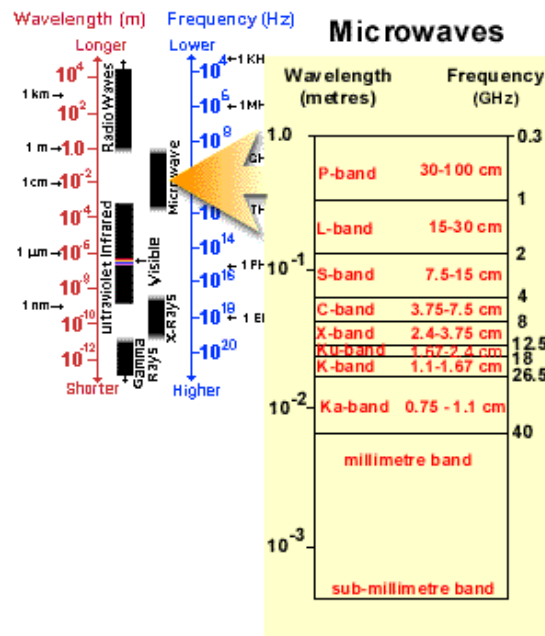
6.1.2 Polarization

When discussing microwave energy, the **polarization** of the radiation is also important. Polarization refers to the orientation of the electric field (recall the definition of electromagnetic radiation from). Most radars are designed to transmit microwave radiation either horizontally polarized (H) or vertically polarized (V). Similarly, the antenna receives either the horizontally or vertically polarized backscattered energy, and some radars can receive both. These two polarization states are designated by the letters H for horizontal, and V, for vertical. Thus, there can be four combinations of both transmit and receive polarizations as follows:

- HH - for horizontal transmit and horizontal receive,
- VV - for vertical transmit and vertical receive,
- HV - for horizontal transmit and vertical receive, and
- VH - for vertical transmit and horizontal receive.

The first two polarization combinations are referred to as like-polarized because the transmit and receive polarizations are the same. The last two combinations are referred to as cross-polarized because the transmit and receive polarizations are opposite of one another. Similar to variations in wavelength, depending on the transmit and receive polarizations, the radiation will interact with and be backscattered differently from the surface. Both wavelength and polarization affect how a radar "sees" the surface. Therefore, radar imagery collected using different polarization and wavelength combinations may provide different and complementary information about the targets on the surface.

6.1.3 Viewing Geometry and Spatial Resolution

The imaging geometry of a radar system is different from the framing and scanning systems commonly employed for optical remote sensing described in Chapter 2. Similar to optical systems, the platform travels forward in the **flight direction** (A) with the **nadir** (B) directly beneath the platform. The microwave beam is transmitted obliquely at right angles to the direction of flight illuminating a **swath** (C) which is offset from nadir. **Range** (D) refers to the across-track dimension perpendicular to the flight direction, while **azimuth** (E) refers to the along-track dimension parallel to the flight direction. This side-looking viewing geometry is typical of imaging radar systems (airborne or spaceborne).
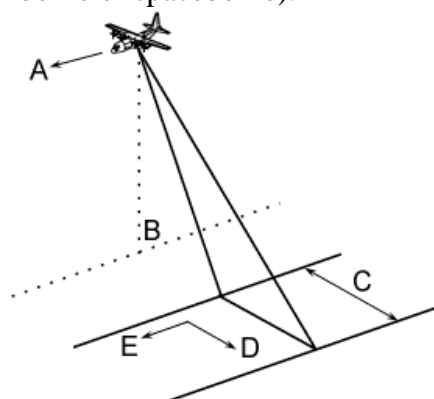
Imaging geometry of a radar system

The portion of the image swath closest to the nadir track of the radar platform is called the **near range** (a) while the portion of the swath farthest from the nadir is called the **far range** (b). The incidence angle is the angle between the **radar beam and ground surface** (A) which increases, moving across the swath from near to far

range. The **look angle** (B) is the angle at which the radar "looks" at the surface. In the near range, the viewing geometry may be referred to as being steep, relative to the far range, where the viewing geometry is shallow. At all ranges the radar antenna measures the radial line of sight distance between the radar and each target on the surface. This is the **slant range distance** (C). The **ground range distance** (D) is the true horizontal distance along the ground corresponding to each point measured in slant range.
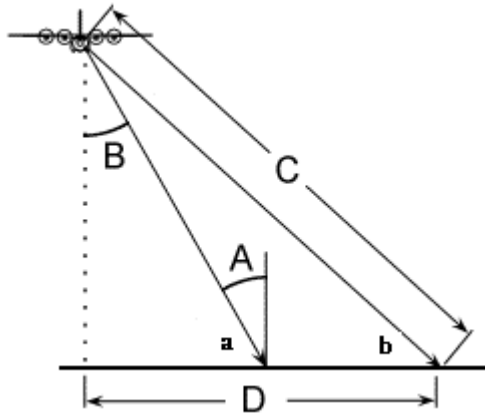
Near\far range and incidence angles

Unlike optical systems, a radar's spatial resolution is a function of the specific properties of the microwave radiation and geometrical effects. If a Real Aperture Radar (RAR) is used for image formation (as in Side-Looking Airborne Radar) a single transmit pulse and the backscattered signal are used to form the image. In this case, the resolution is dependent on the effective length of the pulse in the slant range direction and on the width of the illumination in the azimuth direction. The range or across-track resolution is dependent on the length of the pulse (P). Two distinct targets on the surface will be resolved in the range dimension if their separation is greater than half the pulse length. For example, targets 1 and 2 will not be separable while targets 3 and 4 will. Slant range resolution remains constant, independent of range. However, when projected into ground range coordinates, the resolution in ground range will be dependent of the incidence angle. Thus, for fixed slant range resolution, the ground range resolution will decrease with increasing range.
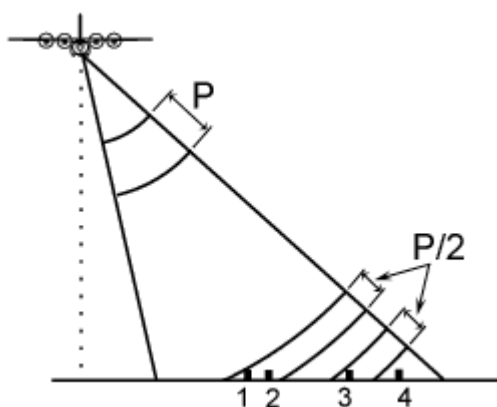
Range or across-track resolution

The azimuth or along-track resolution is determined by the angular width of the radiated microwave beam and the slant range distance. This **beamwidth** (A) is a measure of the width of the illumination pattern. As the radar illumination propagates to increasing distance from the sensor, the azimuth resolution increases (becomes coarser). In this illustration, targets 1 and 2 in the near range would be separable, but targets 3 and 4 at further range would not. The radar beamwidth is inversely

proportional to the antenna length (also referred to as the aperture) which means that a longer antenna (or aperture) will produce a narrower beam and finer resolution.
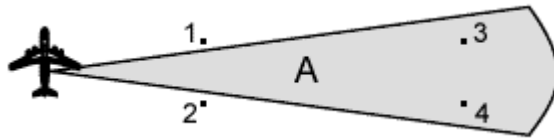
Azimuth or along-track resolution

The equation for range resolution is:

$R_r = tc \, / \, 2\cos g$

And for azimuth resolution:

$R_a = 0.7Sg \, / \, D$

t	*Pulse length*
c	*speed of light*
g	*the angle between the horizontal plane and the line connecting the target and the antenna (1 – incidence angle)*
D	*antenna length*

Finer range resolution can be achieved by using a shorter pulse length, which can be done within certain engineering design restrictions. Finer azimuth resolution can be achieved by increasing the antenna length. However, the actual length of the antenna is limited by what can be carried on an airborne or spaceborne platform. For airborne radars, antennas are usually limited to one to two metres; for satellites they can be 10 to 15 metres in length. To overcome this size limitation, the forward motion of the platform and special recording and processing of the backscattered echoes are used to simulate a very long antenna and thus increase azimuth resolution.
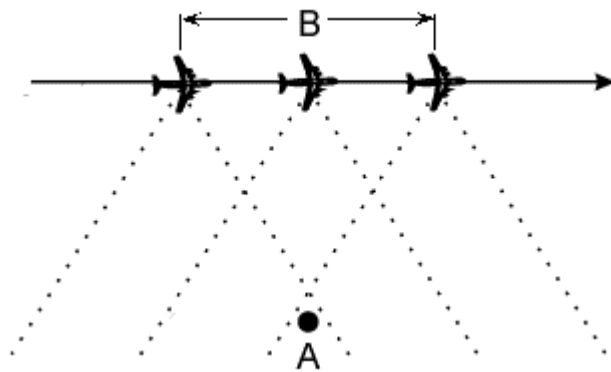
Increase azimuth resolution

This figure illustrates how this is achieved. As a target (A) first enters the radar beam (1), the backscattered echoes from each transmitted pulse begin to be recorded. As the platform continues to move forward, all echoes from the target for each pulse are recorded during the entire time that the target is within the beam. The point at which the target leaves the view of the radar beam (2) some time later, determines the length of the simulated or **synthesized** antenna (B). Targets at far range, where the beam is widest will be illuminated for a longer period of time than objects at near range. The expanding beamwidth, combined with the increased time a target is within the beam as ground range increases, balance each other, such that the resolution remains constant across the entire swath. This method of achieving uniform, fine azimuth

resolution across the entire imaging swath is called **synthetic aperture radar**, or **SAR**. Most airborne and spaceborne radars employ this type of radar.

6.1.4 Radar Image Distortions

As with all remote sensing systems, the viewing geometry of a radar results in certain geometric distortions on the resultant imagery. However, there are key differences for radar imagery which are due to the side-looking viewing geometry, and the fact that the radar is fundamentally a distance measuring device (i.e. measuring range). Slant-range scale distortion occurs because the radar is measuring the distance to features in slant-range rather than the true horizontal distance along the ground. This results in a varying image scale, moving from near to far range. Although targets A1 and B1 are the same size on the ground, their apparent dimensions in slant range (A2 and B2) are different. This causes targets in the near range to appear compressed relative to the far range. Using trigonometry, ground-range distance can be calculated from the slant-range distance and platform altitude to convert to the proper ground-range format. This conversion comparison shows a radar image in slant-range display (top) where the fields and the road in the near range on the left side of the image are compressed, and the same image converted to ground-range display (bottom) with the features in their proper geometric shape.
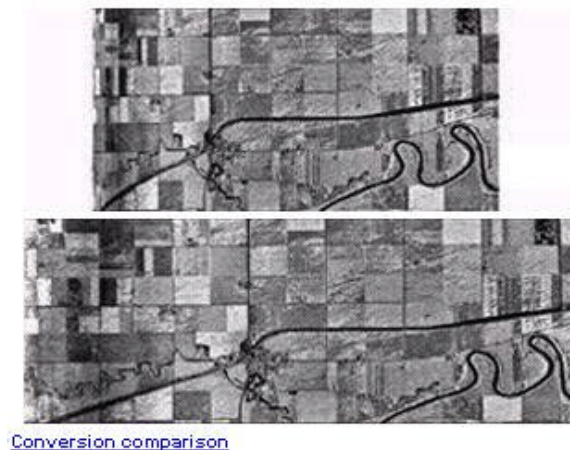
Conversion comparison

Similar to the distortions encountered when using cameras and scanners, radar images are also subject to geometric distortions due to **relief displacement**. As with scanner imagery, this displacement is one-dimensional and occurs perpendicular to the flight path. However, the displacement is reversed with targets being displaced towards, instead of away from the sensor. Radar **foreshortening** and **layover** are two consequences which result from relief displacement.

When the radar beam reaches the base of a tall feature tilted towards the radar (e.g. a mountain) before it reaches the top foreshortening will occur. Again, because the radar measures distance in slant-range, the slope (A to B) will appear compressed and the length of the slope will be represented incorrectly (A' to B'). Depending on the angle of the hillside or mountain slope in relation to the incidence angle of the radar beam, the severity of foreshortening will vary. Maximum foreshortening occurs when the radar beam is perpendicular to the slope such that the slope, the base, and the top are imaged simultaneously (C to D). The length of the slope will be reduced to an effective length of zero in slant range (C'D'). This figure shows a radar image of steep mountainous terrain with severe foreshortening effects. The foreshortened slopes appear as bright features on the image.
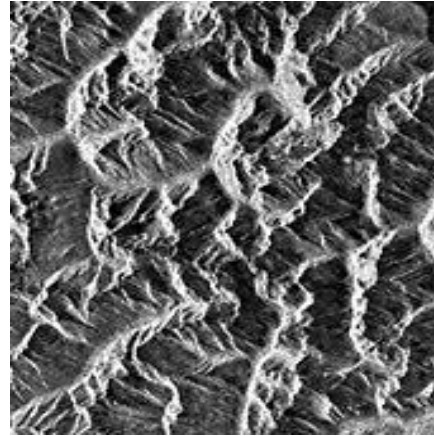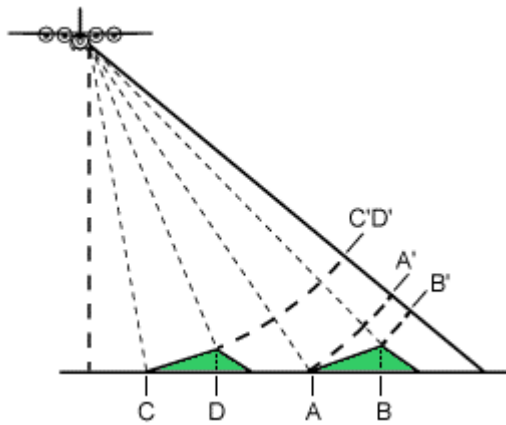
Layover occurs when the radar beam reaches the top of a tall feature (B) before it reaches the base (A). The return signal from the top of the feature will be received before the signal from the bottom. As a result, the top of the feature is displaced towards the radar from its true position on the ground, and "lays over" the base of the feature (B' to A'). Layover effects on a radar image look very similar to effects due to foreshortening. As with foreshortening, layover is most severe for small incidence angles, at the near range of a swath, and in mountainous terrain.
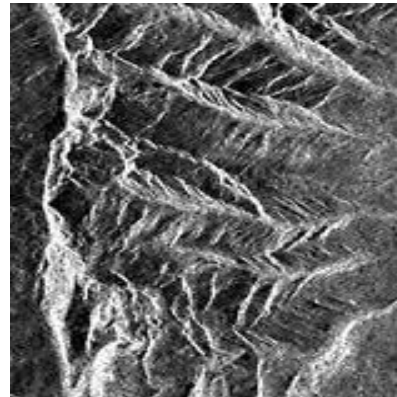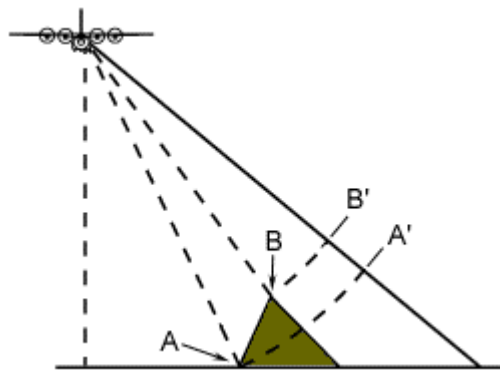
Both foreshortening and layover result in radar **shadow.** Radar shadow occurs when the radar beam is not able to illuminate the ground surface. Shadows occur in the down range dimension (i.e. towards the far range), behind vertical features or slopes with steep sides. Since the radar beam does not illuminate the surface, shadowed regions will appear dark on an image as no energy is available to be backscattered. As incidence angle increases from near to far range, so will shadow effects as the radar beam looks more and more obliquely at the surface. This image illustrates radar shadow effects on the right side of the hillsides which are being illuminated from the left.
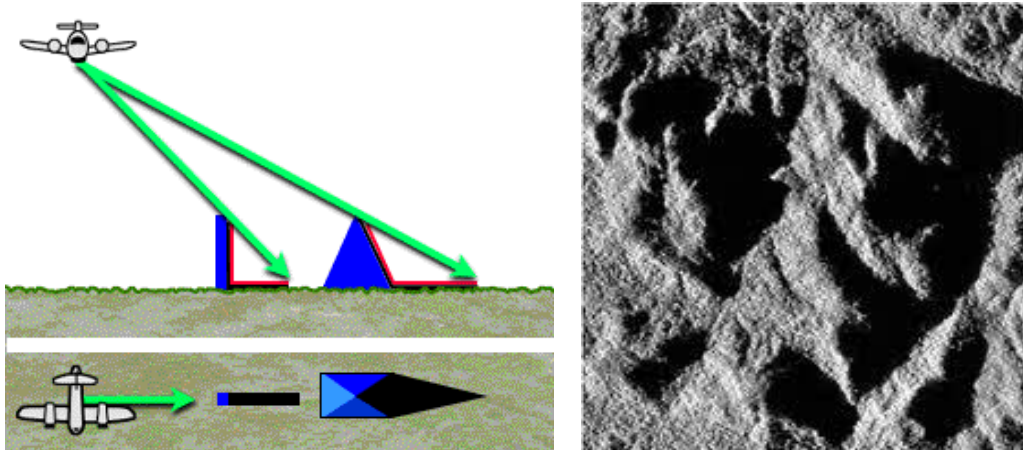
6.1.5 Target Interaction and Image Appearance

The brightness of features in a radar image is dependent on the portion of the transmitted energy that is returned back to the radar from targets on the surface. The magnitude or intensity of this backscattered energy is dependent on how the radar energy interacts with the surface, which is a function of several variables or parameters. These parameters include the particular characteristics of the radar system (frequency, polarization, viewing geometry, etc.) as well as the characteristics of the surface (landcover type, topography, relief, etc.). Because many of these characteristics are interrelated, it is impossible to separate out each of their individual contributions to the appearance of features in a radar image. Changes in the various parameters may have an impact on and affect the response of other parameters, which together will affect the amount of backscatter. Thus, the brightness of features in an image is usually a combination of several of these variables. However, for the purposes of our discussion, we can group these characteristics into three areas which fundamentally control radar energy/target interactions. They are:

- Surface roughness of the target
- Radar viewing and surface geometry relationship
- Moisture content and electrical properties of the target

The surface roughness of a feature controls how the microwave energy interacts with that surface or target and is generally the dominant factor in determining the tones seen on a radar image. Surface roughness refers to the average height variations in the surface cover from a plane surface, and is measured on the order of centimetres. Whether a surface appears rough or smooth to a radar depends on the wavelength and incidence angle.
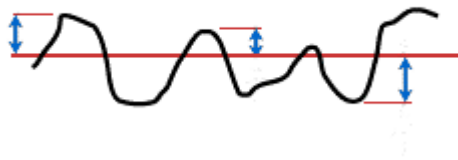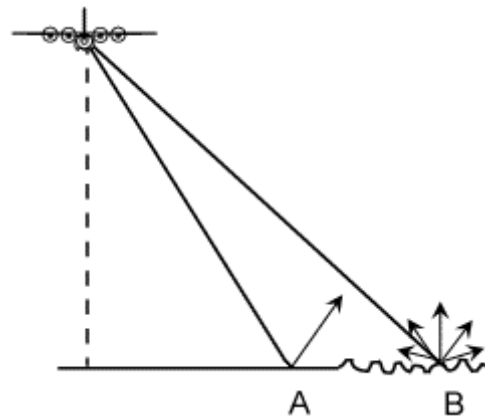
Surface roughness   *Spectral diffusity of the surface*

Simply put, a surface is considered "smooth" if the height variations are much smaller than the radar wavelength. When the surface height variations begin to approach the size of the wavelength, then the surface will appear "rough". Thus, a given surface will appear rougher as the wavelength becomes shorter and smoother as the wavelength becomes longer. A smooth surface (A) causes **specular** reflection of the incident energy (generally away from the sensor) and thus only a small amount of energy is returned to the radar. This results in smooth surfaces appearing as darker toned areas on an image. A rough surface (B) will scatter the energy approximately equally in all directions (i.e. **diffusely**) and a significant portion of the energy will be backscattered to the radar. Thus, rough surfaces will appear lighter in tone on an image. Incidence angle, in combination with wavelength, also plays a role in the apparent roughness of a surface. For a given surface and wavelength, the surface will appear smoother as the incidence angle increases. Thus, as we move farther across the swath, from near to far range, less energy would be returned to the sensor and the image would become increasingly darker in tone.

The relationship of radar wavelength and depression angle to surface roughness may be described by the *Rayleigh criterion* that considers a surface to be smooth if

$$h < \mathtt{l} \, / \, 8 \sin \mathtt{g}$$

where

h = the height of surface irregularities, or surface roughness
$\mathtt{l}$ = the radar wavelength
$\mathtt{g}$ = the grazing angle between the terrain and the incident radar wave

both h and $\mathtt{l}$ are given in the same units, usually centimeters. For horizontal terrain surfaces, the grazing angle equals the antenna depression angle. For an X-band radar ($\mathtt{l}$ = 3 cm) at a depression angle of 45º, the roughness value at which the surface will appear smooth is: $h < 0.53$ cm
This means that a vertical relief of 0.53 cm is the theoretical boundary between smooth and rough surfaces for the given wavelength and depression angle. The Rayleigh criterion does not consider the important category of surface relief that is intermediate between definitely smooth and definitely rough surfaces. The Rayleigh criterion was modified by Peake and Oliver to define the upper and lower values of h

for surfaces of intermediate roughness. Their *smooth criterion* considers a surface to be smooth if

$$h < 1 / 25 \sin g$$

For the above parameters, a vertical relief of 0.17 cm is the boundary between smooth surfaces and surfaces of intermediate roughness.
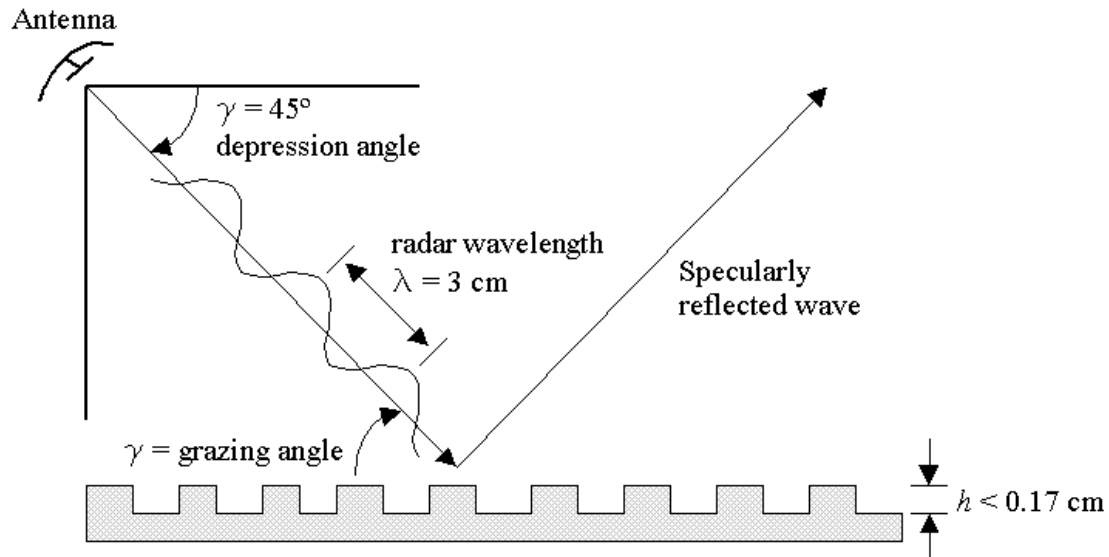
Smooth surface with specular reflection; no return

They have also derived a *rough criterion* that considers a surface to be rough if

$$h > 1 / 4.4 \sin g$$

Which for the above parameters gives the result of 0.96 cm.

Smooth surfaces produce strong returns at depression angles near vertical, but little or no return at lower angles. For a rough surface, the relatively uniform return decreases somewhat at low depression angles because of the longer two way travel distance.
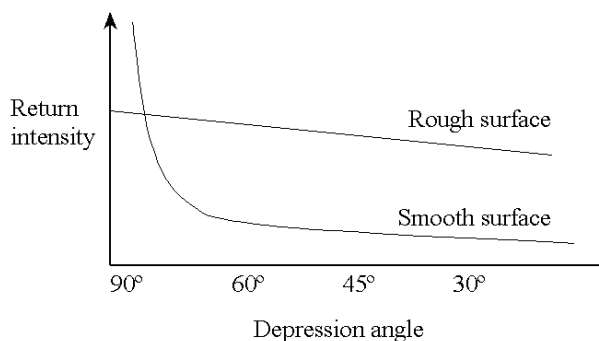
Return Intensity as a function of depression angle   *Local incidence angle*
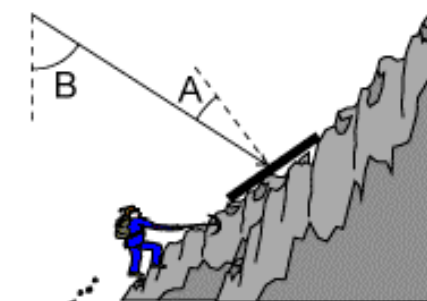
We have already discussed incidence or look angle in relation to viewing geometry and how changes in this angle affect the signal returned to the radar. However, in relation to surface geometry, and its effect on target interaction and image appearance, the local incidence angle is a more appropriate and relevant concept. The local

140

incidence angle is the angle between the radar beam and a line perpendicular to the slope at the point of incidence (A). Thus, local incidence angle takes into account the local slope of the terrain in relation to the radar beam. With flat terrain, the local incidence angle is the same as the look angle (B) of the radar. For terrain with any type of relief, this is not the case. Generally, slopes facing towards the radar will have small local incidence angles, causing relatively strong backscattering to the sensor, which results in a bright-toned appearance in an image.

As the concept of local incidence angle demonstrates, the relationship between viewing geometry and the geometry of the surface features plays an important role in how the radar energy interacts with targets and their corresponding brightness on an image. Variations in viewing geometry will accentuate and enhance topography and relief in different ways, such that varying degrees of foreshortening, layover, and shadow (previous section) may occur depending on surface slope, orientation, and shape.

The look direction or aspect angle of the radar describes the orientation of the transmitted radar beam relative to the direction or alignment of linear features on the surface. The look direction can significantly influence the appearance of features on a radar image, particularly when ground features are organized in a linear structure (such as agricultural crops or mountain ranges). If the look direction is close to perpendicular to the orientation of the feature (A), then a large portion of the incident energy will be reflected back to the sensor and the feature will appear as a brighter tone. If the look direction is more oblique in relation to the feature orientation (B), then less energy will be returned to the radar and the feature will appear darker in tone. Look direction is important for enhancing the contrast between features in an image. It is particularly important to have the proper look direction in mountainous regions in order to minimize effects such as layover and shadowing. By acquiring imagery from different look directions, it may be possible to enhance identification of features with different orientations relative to the radar.
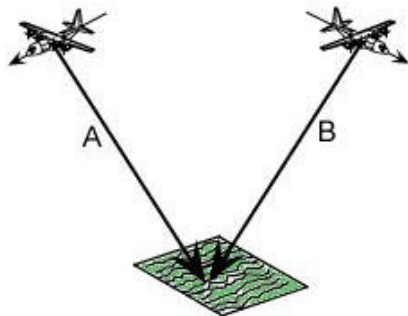
Features which have two (or more) surfaces (usually smooth) at right angles to one another, may cause corner reflection to occur if the 'corner' faces the general direction of the radar antenna. The orientation of the surfaces at right angles causes most of the radar energy to be reflected directly back to the antenna due to the double bounce (or more) reflection. Corner reflectors with complex angular shapes are common in urban environments (e.g. buildings and streets, bridges, other man-made structures). Naturally occurring corner reflectors may include severely folded rock and cliff faces or upright vegetation standing in water. In all cases, corner reflectors show up as very bright targets in an image, such as the buildings and other man-made structures in this radar image of a city.
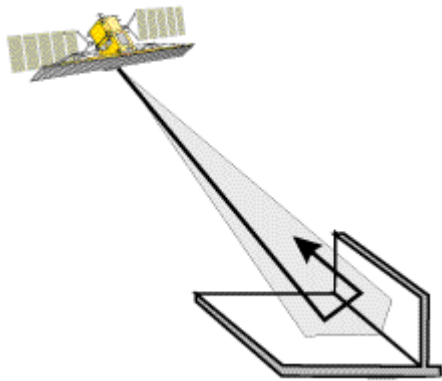
The presence (or absence) of moisture affects the electrical properties of an object or medium. Changes in the electrical properties influence the absorption, transmission, and reflection of microwave energy. Thus, the moisture content will influence how targets and surfaces reflect energy from a radar and how they will appear on an image. Generally, reflectivity (and image brightness) increases with increased moisture content. For example, surfaces such as soil and vegetation cover will appear brighter when they are wet than when they are dry.

When a target is moist or wet, scattering from the topmost portion (**surface scattering**) is the dominant process taking place. The type of reflection (ranging from specular to diffuse) and the magnitude will depend on how rough the material appears to the radar. If the target is very dry and the surface appears smooth to the radar, the radar energy may be able to penetrate below the surface, whether that surface is discontinuous (e.g. forest canopy with leaves and branches), or a homogeneous surface (e.g. soil, sand, or ice). For a given surface , longer wavelengths are able to penetrate further than shorter wavelengths.

If the radar energy does manage to penetrate through the topmost surface, then volume scattering may occur. Volume scattering is the scattering of radar energy within a volume or medium, and usually consists of multiple bounces and reflections from different components within the volume. For example, in a forest, scattering may come from the leaf canopy at the tops of the trees, the leaves and branches further below, and the tree trunks and soil at the ground level. Volume scattering may serve to decrease or increase image brightness, depending on how much of the energy is scattered out of the volume and back to the radar.
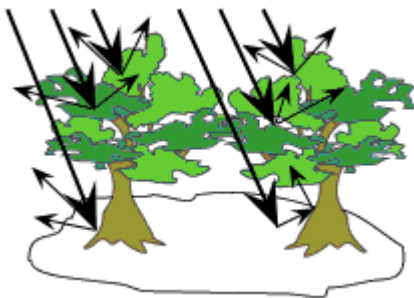
6.1.6 Radar Image Properties

All radar images appear with some degree of what we call radar **speckle**. Speckle appears as a grainy "salt and pepper" texture in an image. This is caused by random constructive and destructive interference from the multiple scattering returns that will occur within each resolution cell. As an example, an homogeneous target, such as a large grass-covered field, without the effects of speckle would generally result in light-toned pixel values on an image (A). However, reflections from the individual blades of grass within each resolution cell results in some image pixels being brighter and some being darker than the average tone (B), such that the field appears speckled.
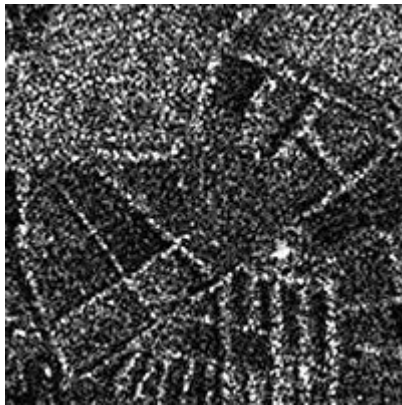
Speckle
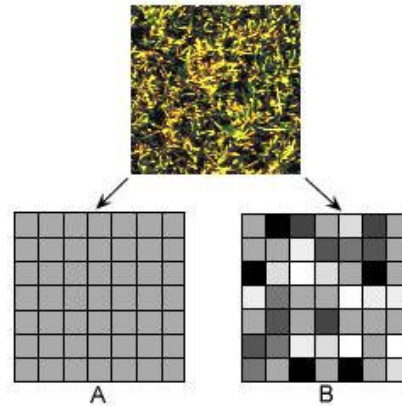
Grass covered field

Speckle is essentially a form of noise which degrades the quality of an image and may make interpretation (visual or digital) more difficult. Thus, it is generally desirable to reduce speckle prior to interpretation and analysis. **Speckle reduction** can be achieved in two ways:

- multi-look processing, or
- spatial filtering.

Multi-look processing refers to the division of the radar beam (A) into several (in this example, five) narrower sub-beams (1 to 5). Each sub-beam provides an independent "look" at the illuminated scene, as the name suggests. Each of these "looks" will also be subject to speckle, but by summing and averaging them together to form the final output image, the amount of speckle will be reduced.
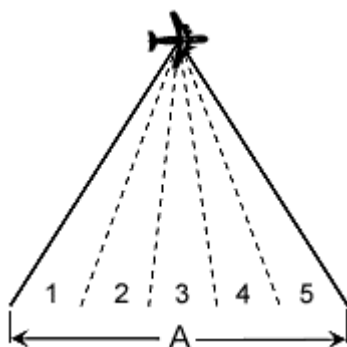
While multi-looking is usually done during data acquisition, speckle reduction by spatial filtering is performed on the output image in a digital (i.e. computer) image analysis environment. Speckle reduction filtering consists of moving a small window of a few pixels in dimension (e.g. 3x3 or 5x5) over each pixel in the image, applying a mathematical calculation using the pixel values under that window (e.g. calculating

the average), and replacing the central pixel with the new value. The window is moved along in both the row and column dimensions one pixel at a time, until the entire image has been covered. By calculating the average of a small window around each pixel, a smoothing effect is achieved and the visual appearance of the speckle is reduced. This graphic shows a radar image before (top) and after (bottom) speckle reduction using an averaging filter. The median (or middle) value of all the pixels underneath the moving window is also often used to reduce speckle. Other more complex filtering calculations can be performed to reduce speckle while minimizing the amount of smoothing taking place.
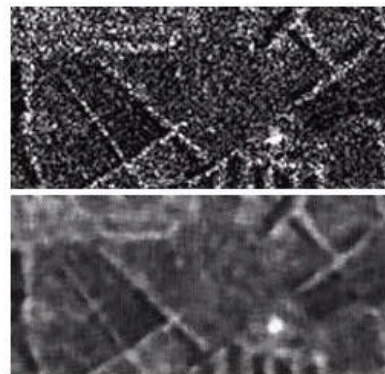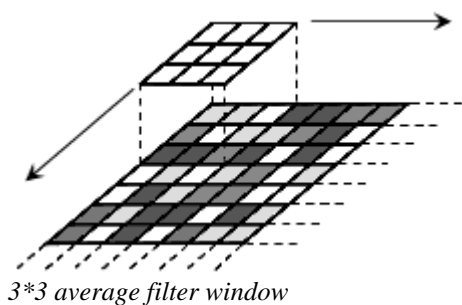
*3*3 average filter window*   *Speckle reduction using an average filtering*

Both multi-look processing and spatial filtering reduce speckle at the expense of resolution, since they both essentially smooth the image. Therefore, the amount of speckle reduction desired must be balanced with the particular application the image is being used for, and the amount of detail required. If fine detail and high resolution is required then little or no multi-looking/spatial filtering should be done. If broad-scale interpretation and mapping is the application, then speckle reduction techniques may be more appropriate and acceptable.

Another property peculiar to radar images is slant-range distortion, which was discussed in some detail in section 3.4. Features in the near-range are compressed relative to features in the far range due to the slant-range scale variability. For most applications, it is desirable to have the radar image presented in a format which corrects for this distortion, to enable true distance measurements between features. This requires the slant-range image to be converted to 'ground range' display. This can be done by the radar processor prior to creating an image or after data acquisition by applying a transformation to the slant range image. In most cases, this conversion will only be an estimate of the geometry of the ground features due to the complications introduced by variations in terrain relief and topography.

A radar antenna transmits more power in the mid-range portion of the illuminated swath than at the near and far ranges. This effect is known as **antenna pattern** and results in stronger returns from the center portion of the swath than at the edges. Combined with this antenna pattern effect is the fact that the energy returned to the radar decreases dramatically as the range distance increases. Thus, for a given surface, the strength of the returned signal becomes smaller and smaller moving farther across the swath. These effects combine to produce an image which varies in intensity (tone) in the range direction across the image. A process known as **antenna pattern correction** may be applied to produce a uniform average brightness across the imaged swath, to better facilitate visual interpretation.

The range of brightness levels a remote sensing system can differentiate is related to radiometric resolution (section 2.5) and is referred to as the **dynamic range.** While optical sensors, such as those carried by satellites such as Landsat and SPOT, typically produce 256 intensity levels, radar systems can differentiate intensity levels up to around 100,000 levels! Since the human eye can only discriminate about 40 intensity levels at one time, this is too much information for visual interpretation. Even a typical computer would have difficulty dealing with this range of information. Therefore, most radars record and process the original data as 16 bits (65,536 levels of intensity), which are then further scaled down to 8 bits (256 levels) for visual interpretation and/or digital computer analysis.

Calibration is a process which ensures that the radar system and the signals that it measures are as consistent and as accurate as possible. Prior to analysis, most radar images will require **relative calibration.** Relative calibration corrects for known variations in radar antenna and systems response and ensures that uniform, repeatable measurements can be made over time. This allows relative comparisons between the response of features within a single image, and between separate images to be made with confidence. However, if we wish to make accurate **quantitative** measurements representing the actual energy or power returned from various features or targets for comparative purposes, then **absolute calibration** is necessary.

Absolute calibration, a much more involved process than relative calibration, attempts to relate the magnitude of the recorded signal strength to the actual amount of energy backscattered from each resolution cell. To achieve this, detailed measurements of the radar system properties are required as well as quantitative measurements of the scattering properties of specific targets. The latter are often obtained using ground-based scatterometers, as described in section 3.1. Also, devices called **transponders** may be placed on the ground prior to data acquisition to calibrate an image. These devices receive the incoming radar signal, amplify it, and transmit a return signal of known strength back to the radar. By knowing the actual strength of this return signal in the image, the responses from other features can be referenced to it.

6.1.7 Advanced Radar Applications

In addition to standard acquisition and use of radar data, there are three specific applications worth mentioning.

The first is **stereo radar** which is similar in concept to stereo mapping using aerial photography (described in section 2.7). Stereo radar image pairs are acquired covering the same area, but with different look/incidence angles (A), or opposite look directions (B). Unlike aerial photos where the displacement is radially outward from the nadir point directly below the camera, radar images show displacement only in the range direction. Stereo pairs taken from opposite look directions (i.e. one looking

north and the other south) may show significant contrast and may be difficult to interpret visually or digitally. In mountainous terrain, this will be even more pronounced as shadowing on opposite sides of features will eliminate the stereo effect. Same side stereo imaging (A) has been used operationally for years to assist in interpretation for forestry and geology and also to generate topographic maps. The estimation of distance measurements and terrain height for topographic mapping from stereo radar data is called **radargrammetry**, and is analogous to photogrammetry carried out for similar purposes with aerial photographs.
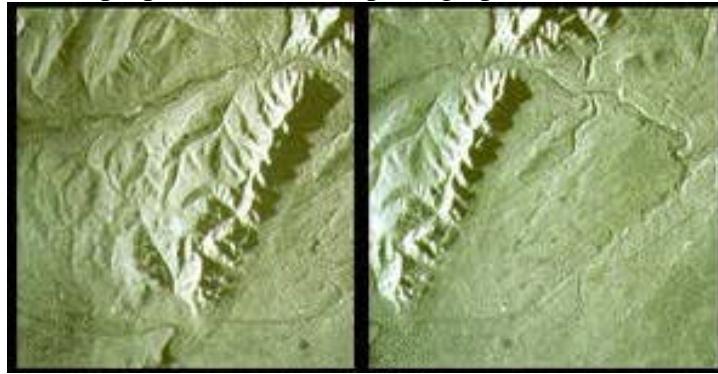
Radargrammetry is one method of estimating terrain height using radar. Another, more advanced method is called **interferometry.** Interferometry relies on being able to measure a property of electromagnetic waves called **phase.** Suppose we have two waves with the exact same wavelength and frequency traveling along in space, but the starting point of one is offset slightly from the other. The offset between matching points on these two waves (A) is called the **phase difference.** Interferometric systems use two antennas, separated in the range dimension by a small distance, both recording the returns from each resolution cell. The two antennas can be on the same platform (as with some airborne SARs), or the data can be acquired from two different passes with the same sensor, such has been done with both airborne and satellite radars. By measuring the exact phase difference between the two returns (A), the path length difference can be calculated to an accuracy that is on the order of the wavelength (i.e centimetres). Knowing the position of the antennas with respect to the Earth's surface, the position of the resolution cell, including its elevation, can be determined. The phase difference between adjacent resolution cells, is illustrated in this interferogram, where colours represents the variations in height. The information contained in an interferogram can be used to derive topographic information and produce three-dimensional imagery of terrain height.

Phase difference *Interferometric systems* *Interferogram*
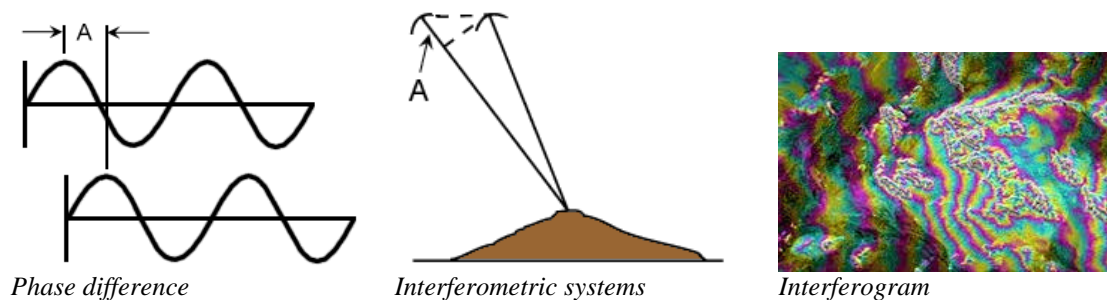
The concept of radar **polarimetry** was already alluded to in our discussion of radar fundamentals in section 3.2. As its name implies, polarimetry involves discriminating between the **polarizations** that a radar system is able to transmit and receive. Most

radars transmit microwave radiation in either horizontal (H) or vertical (V) polarization, and similarly, receive the backscattered signal at only one of these polarizations. **Multi-polarization** radars are able to transmit either H or V polarization and receive both the like- and cross-polarized returns (e.g. HH and HV or VV and VH, where the first letter stands for the polarization transmitted and the second letter the polarization received). **Polarimetric radars** are able to transmit and receive both horizontal and vertical polarizations. Thus, they are able to receive and process all four combinations of these polarizations: HH, HV, VH, and VV. Each of these "polarization channels" have varying sensitivities to different surface characteristics and properties. Thus, the availability of multi-polarization data helps to improve the identification of, and the discrimination between features. In addition to recording the magnitude (i.e. the strength) of the returned signal for each polarization, most polarimetric radars are also able to record the **phase** information of the returned signals. This can be used to further characterize the polarimetric "signature" of different surface features.

7. Remote Sensing Applications for the sea (passive and SLAR)

7.1 Sea Surface Temperature

- *http://nit.colorado.edu/remsens/payne/index.html*

Estimation of Sea Surface Temperature (SST) is important for weather models, estimation of heat content and heat flux, and also for climate monitoring and climate change detection.

SST can be measured either by Thermal Infrared Remote Sensing or by Passive Microwave Remote Sensing, the former (Thermal IR) being more applied. Thermal infrared remote sensing is the oldest technique used to measure SST. This is because the thermal infrared is near the peak of the Planck function at terrestrial temperatures, and the signal is thus strongest at the wavelengths coupled with the fact that thermal infrared detectors are relatively easy to build and reliable. A middle infrared channel is also often used at night (when not contaminated by reflected solar), because this is in a spectral region of strong variance in the Planck function at terrestrial temperatures.

- Strengths
  This strong signal of temperature in the thermal wavelengths provides the thermal infrared with the ability to measure with good accuracy and high resolution. This combination is very important for analysis of long term climate variability, as well as for shorter term weater forecasting. The time series of available infrared SST measurements (20 years) also provides a long term record which is very valuable.
- Weaknesses
  The biggest problem with thermal infrared remote sensing is the presence of an atmosphere. This atmospheric interference takes on two forms, total disruption of the signal in the presence of clouds and atmospheric attenuation of infrared radiation by gases, water vapor, and aerosols. If a cloud is present in a satellite field of view, the signal recieved by a satellite sensor is a measurement of the temperature of the cloud top, and thus cannot be used to calculate an SST. Even when no clouds are present, atmospheric corrections are required to compensate for radiation scattered into or out of the field of view or absorbed by atmospheric constituents (this subject was dealt in the chapter about corrections, previously).

The following steps are followed in taking a satellite bit stream and then converting it to an SST map:
- Radiance calibration (from DN to radiance), followed by a temperature calibration using the information provided by on board blackbody, we get the brightness temperature (radiant, not kinetic).
- Geometric correction.
- Cloud detection and filtering
- Atmospheric correction (for water vapor, gaseous constituents and aerosol attenuation) and SST calcualtion.
- Quality check using a climatology/algorithm intercomparison.

The most used sensor for SST is the NOAA AVHRR, which has the following advantages:

- Global Coverage Daily
- Long Heritage (20 years)
- Source for US Operational SST
- 3.7, 11, and 12 Micron Thermal Bands
- 1 and 4 km Resolution Data
- Accuracy ~ 0.6K rms

Other sensors used are the Along Track Scanning Radiometer (ATSR), the Geostationary Operational Environmental Satellite (GOES), and the TRMM Microwave Imager (TMI). Each one of them has its own advantages and weaknesses. Regarding the GOES, two points are interesting:

On the one hand, as it stays over a fixed point, it enables a study of the diurnal variation in the cycle of SST.

On the other hand, its position in a fixed orbit also causes problems as it goes into and out of the shadow of the earth each day and one side "freezes" as the other side "cooks", creating a (known) bias of up to 0.5K .

The following problems exist in the Remote Sensing of SST:

- Bulk/Skin Temperature Difference

A temperature gradient usually exists across the millimeter immediately below the ocean's surface or skin. This temperature gradient is related to the net air-sea heat flux and influences air-sea interactions and temperature measurements taken at or near the surface.
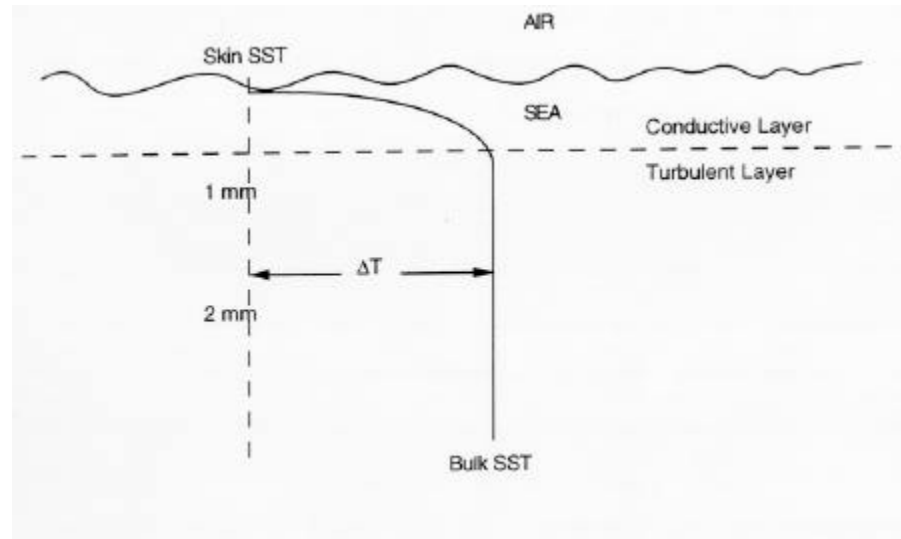
Typically the 1mm skin is about 0.3-0.5K cooler than the bulk of the water beneath. This is mainly a problem at thermal wavelengths, emission from top micron of ocean. As microwave emission penetrates deeper, there the effect is less important.

- Emissivity Variations of Seawater

The emissivity of seawater in the thermal region of the electromagnetic spectrum is very nearly one. However, when seeking an accuracy of .2K or better from a signal of approximately 300K, an error of .1% is enough to. The emissivity of water is affected by the salinity of the water, and as this may vary throughout the ocean, it becomes a factor that is not readily corrected for, even though its magnitude is very small. The

emissivity of water is also a function of viewing angle, and this must be corrected for. Normally it is implicitly absorbed in the viewing angle dependent terms of the SST retrieval equations.

The variation of emissivity of seawater at microwave frequencies is much more substantial than in thermal wavelengths. It is dependent not only on salinity here, but on surface roughness as well. The fact that surface roughness is dependent upon wind speed has led to the development of wind speed retrieval algorithms based on microwave emission and its variation due to wind induced roughness emissivity effects.

7.2 Oil Spill Detection

- *http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/indexe.html*
- *http://www.earthsat.com/geol/oil_seep.html*
- *Sabins Floyd F. (1976), Remote Sensing – Principles and Interpretation, Freeman*

Oil spills can destroy marine life as well as damage habitat for land animals and humans. To limit the areas affected by the spill and facilitate containment and cleanup efforts, a number of factors have to be identified:

- Spill location
- Size and extent of the spill
- Direction and magnitude of oil movement
- Wind, current and wave information for predicting future oil movement

Remote sensing offers the advantage of being able to observe events in remote and often inaccessible areas. For example, oil spills from ruptured pipelines, may go unchecked for a period of time because of uncertainty of the exact location of the spill, and limited knowledge of the extent of the spill. Remote sensing can be used to both detect and monitor spills. Detecting oil seeps can be also used in exploration programs.

For ocean spills, remote sensing data can provide information on the rate and direction of oil movement through multi-temporal imaging, and input to drift prediction modelling and may facilitate in targeting clean-up and control efforts. Remote sensing devices used include the use of infrared video and photography from airborne platforms, thermal infrared imaging, airborne laser fluourosensors, airborne and space-borne optical sensors, as well as airborne and spaceborne SAR. SAR sensors have an advantage over optical sensors in that they can provide data under poor weather conditions and during darkness. Users of remotely sensed data for oil spill applications include the Coast Guard, national environmental protection agencies and departments, oil companies, shipping industry, insurance industry, fishing industry, national departments of fisheries and oceans, and departments of defence.

The key operational data requirements are fast turnaround time and frequent imaging of the site to monitor the dynamics of the spill. For spill identification, high resolution sensors are generally required, although wide area coverage is very important for initial monitoring and detection. Airborne sensors have the advantage of frequent site specific coverage on demand, however, they can be costly. Spills often occur in inclement weather, which can hinder airborne surveillance.

Laser fluorosensors are the best sensors for oil spill detection (where it is needed to separate between the quite similar spectra of oil and water; with UV oil films as thin

as 15 mm can be detected), and have the capability of identifying oil on shores, ice and snow, and determining what type of oil has been spilled. However, they require relatively cloud free conditions to detect the oilspill.

On daytime and nighttime IR images both refined and crude oils have cooler signatures than the adjacent clean water. The oil and water have the same kinetic temperature beacause the two liquids are in direct contact. However, the emissivity of pure water is 0.993, and a thin film of petroleum reduces the emissivity to 0.972. For a kinetic emperature of 18ºC, the radiant temperature of pure water will be 17.5ºC, while that of the oil film will be just 15.9ºC. IR images have the advantage of being available both for day and night operations. However, rain and fog preven temperature image acquisition, and the interpreter must be careful to avoid confusing cold water currents with oil slicks.

SAR sensors can image oilspills through the localized suppression of Bragg scale waves. Oilspills are visible on a radar image as circular or curvilinear features with a darker tone than the surrounding ocean (the small waves that cause backscatter on radar images of the sea are dampened by a thin film of oil). The detection of an oilspill is strongly dependent upon the wind speed. At wind speeds greater than 10 m/s, the slick will be broken up and dispersed, making it difficult to detect. Another factor that can play a role in the successful detection of an oilspill is the difficulty in distinguishing between a natural surfactant and an oilspill. Multi-temporal data and ancillary information can help to discriminate between these two phenomena.

7.2.1 Case study 1 – oil slick
A supertanker, the Sea Empress, was grounded near the town of Milford Haven, Wales on February 15, 1996. After hitting rocks, the outer hull was breached and approximately 70,000 tonnes of light grade crude oil was dispersed southward under storm conditions.
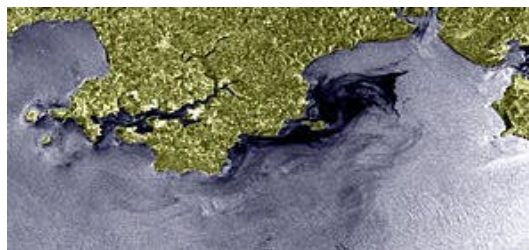
In this RADARSAT image taken a week after the spill, the extent of the oil is visible. The dark areas off the coast represent the areas where oil is present and areas of lighter tone directly south are areas where dispersant was sprayed on the oil to encourage emulsification. Oil, which floats on the top of water, suppresses the ocean's capillary waves, creating a surface smoother than the surrounding water. This smoother surface appears dark in the radar image. As the oil starts to emulsify and clean-up efforts begin to take effect, the capillary waves are not as effectively damped and the oil appears lighter. Size, location and dispersal of the oil spill can be determined using this type of imagery.

7.2.2 Case study 2 – oil seep
EarthSat developed the Seep Enhancement Algorithm (SEA), to select and process satellite images for offshore hydrocarbon seep detection. SEA operates with image

data from a variety of satellites (primarily Landsat, ERS-1 and SPOT), detecting seeps under a broad range of environmental and imaging conditions.

Offshore Oil Seeps
Seeps are direct evidence of hydrocarbon occurrences. Seeps are associated with many major onshore discoveries. Seeps provide significant targets for more intensive exploration efforts. Many prospective offshore areas have yet to be searched for seeps.

Finding Seeps by Satellite
Thin oil slicks form where bubbles of oil and gas reach the surface. Oil slicks affect water in two important ways that are readily detected by imaging satellites:

- Spectral: oil slicks increase reflectance in the visible through near-infrared portion of the electromagnetic spectrum.
- Textural: oil slicks smooth the sea surface, reducing the amount of reflected sun glint ("glitter") and radar backscatter.

The type of detection (spectral vs. textural) depends on seepage rate, oil composition, sea state and illumination. Only satellite imagery provides detailed data on the shape and size of natural oil slicks to pinpoint seep location and estimate seepage rates.

7.3 Ice motion and monitoring

- *http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/indexe.html*
- *Sabins Floyd F. (1976), Remote Sensing – Principles and Interpretation, Freeman*

Ice moves quickly and sometimes unpredictably in response to ocean currents and wind. Vessels can be trapped or damaged by the pressure resulting from these moving ice floes. Even offshore structures can be damaged by the strength and momentum of moving ice. For these reasons it is important to understand the ice dynamics in areas of construction or in the vicinity of a shipping/fishing route.

Remote sensing gives a tangible measure of direction and rate of ice movement through mapping and change detection techniques. Ice floes actually have individual morphological characteristics (shape, structures) that allow them to be distinguished from one another. The floes can be mapped and their movement monitored to facilitate in planning optimum shipping routes, to predict the effect of ice movement on standing structures (bridges, platforms).Users of this type of information include the shipping, fishing, and tourism industries, as well as engineers involved in offshore platform and bridge design and maintenance.

Monitoring of ice movement requires frequent and reliable imaging. The revisit interval must be frequent enough to follow identifiable features before tracking becomes difficult due to excessive movement or change in appearance. Active microwave sensing (radar) provides a reliable source of imaging under all weather and illumination conditions. RADARSAT provides this type of sensor and is a spaceborne platform, which is advantageous for routine imaging operations. The orbital path ensures that Arctic areas are covered daily which meets the requirement for frequent imaging.

Surface roughness explains most of the signature characteristics of sea ice, but the brine content of the ice also influences the radar return. Young ice has a high content

of brine that is progressively reduced with time to a minimum in multi-year ice. Brine has a higher dielectric constant than pure ice, which reduces the radar return.

The resolution and imaging frequency requirements for ice motion tracking vary with the size of floes and the ice dynamics in a region. In areas of large slow moving floes (e.g. Beaufort Sea), 1km resolution data over 10 day intervals is adequate. In dynamic marginal ice zones (e.g. Gulf of St. Lawrence), 100m resolution data over 12-24 hr intervals is required.

It is also possible to use passive Remote Sensing for identifying sea ice, thoush it is prone to atmospheric attenuation, and the orbital and temporal of most of these satellites (TM, SPOT) are not adequate. In order to distinguish ice from clouds, the following criteria can be used: the more uniform brightness of ice, the sharp contacts with water of the ice floes, and clous shadows.
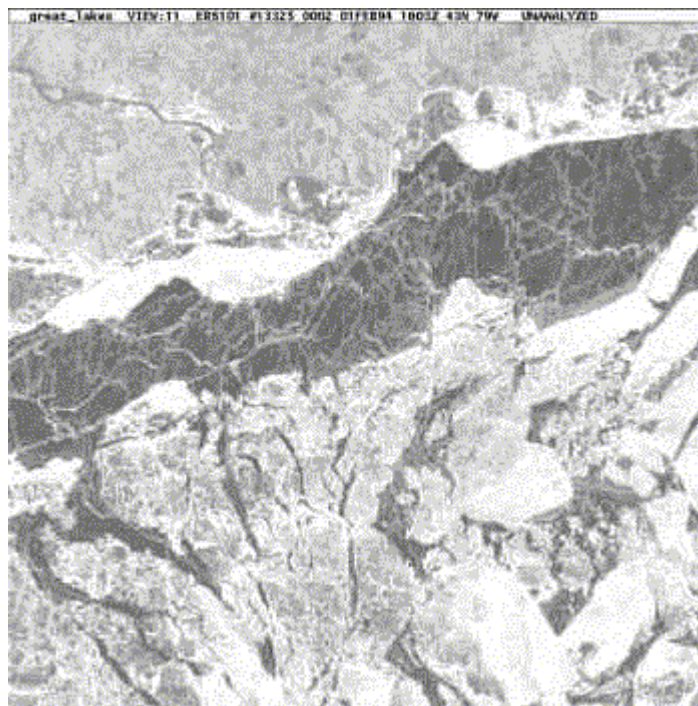
ERS-1 image, Northern Lake Erie, 1 Feb 1994, http://www.cis.ec.gc.ca/pictures/picture.html

7.3.1 Case study (example): *E.-A. Herland,* **Operational ice monitoring with ERS-1 SAR***, Space Technology, Technical Research Centre of Finland*, http://earth.esa.int/l2/0/eeo1.1

The ice season in the northern Baltic Sea lasts for more than six months in the northernmost part during normal winters, with the maximum extent occurring in the January-March period. This severely affects marine traffic in the area, and the Finnish Board of Navigation operates nine icebreakers in the area in order to support marine traffic and keep important harbours open during the ice season. The Ice Service compiles ice charts daily during the ice season, and information on the ice condition is distributed to shipping and harbour authorities and to ships in the area. Since 1989 icebreakers have used an image workstation, developed by VTT (Technical Research Centre of Finland), for viewing NOAA images. This workstation has later been expanded to incorporate ERS-l SAR images, too.

Each ERS-1 SAR image covers an area of 100 x 100 km with 30 m resolution and is produced in near real time at the receiving stations in Kiruna, Sweden, and Fucino, Italy. During the winter 1994 the images used were low-resolution (100 m) images generated by lowpass filtering FD-images. During the winters 1992 and 1994 ERS-1 was in the 3-day cycle orbit and in 1993 in the 35-day cycle orbit.

Images were transmitted in digital form to the icebreakers, where the images are displayed on an image workstation. In orderto be able to combine different types of images, map information, and other information available to the icebreakers, all images are transformed to the Mercator projection before transmission to the icebreakers. For this purpose a special image format has been developed, which includes all necessary information on image acquisition time, sensor and geographical coordinates.

The experiences from the three winters of 1992-4 have shown that it is technically feasible to use satellite SAR images for real-time ice monitoring. The main bottleneck has turned out to be time delay from satellite overpass until the user has the image. With the ERS-1 coverage, it is also not possible to cover all interesting areas in a timely fashion. With systems like Radarsat, the coverage problem is alleviated, but large areal coverage will most likely require a satellite distribution channel even for 100 m resolution images. In order to allow near real-time use of the images, this distribution must be done immediately after the processing at the ground station, and the processing capacity must be sufficient for the required areal coverage. The usefulness of the images has been clearly demonstrated, both for icebreaker operation and ice charting. If a recent satellite image is available, the icebreaker workstation lets the captain optimise the route for the icebreaker and advise ships by radio on how to find useful leads in the ice, thus minimising the need for icebreaker assistance. For ice charting, SAR images offer a unique capability, especially during cloudy weather conditions.
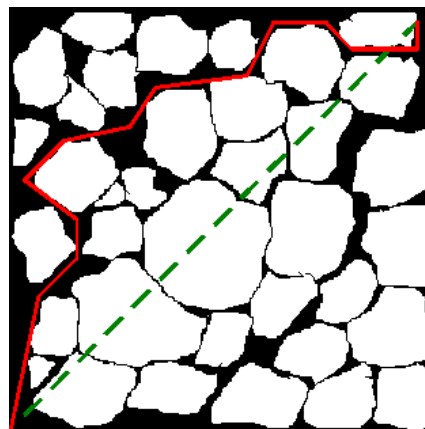
Concept of optimal ice routing between ice floes in a 10 by 10 km area using SAR images (bold red line) compared with theoretical straight line between two waypoints (dashed green line).

Taken from:

 O. M. Johannessen, ICEWATCH - REAL-TIME SEA ICE MONITORING OF THE NORTHERN SEA ROUTE USING SATELLITE RADAR TECHNOLOGY, *http://earth.esa.int/l2/0/eeo1.1*

7.4 Mapping the sea floor (bathymetry)

▪ *Sabins Floyd F. (1976), Remote Sensing – Principles and Interpretation, Freeman*

Many parts of the world's sea floor are charted only in a general fashion or the charts are inaccurate (being out of date, etc). Accurate charts are especially needed for shallow shelf areas where submarine deposition, erosion, and growth of coral reefs can change the bottom topography within a few years after a bathymetric survey is completed.

Remote Sensing of the sea floor from aircraft or satellites is restricted by the fact that water absorbs or reflects most wavelengths or electromagnetic energy. IR energy is absorbed by waterm, and only visible wavelengths penetrate water. The depth of penetration is influenced by the turbidity of the water and the wavelength of the light. A 10-m layer of clear ocean water transmits almost 50 percent of the incident radiation from 0.4 to 0.6 ㎜ in wavelength, but transmits less than 10 percent of radiation from 0.6 to 0.7 ㎜ wavelength. The increased turbidity results in a decrease of light transmittance and a shift in the wavelength of maximum transmittance to the 0.5 to 0.6 ㎜ region.
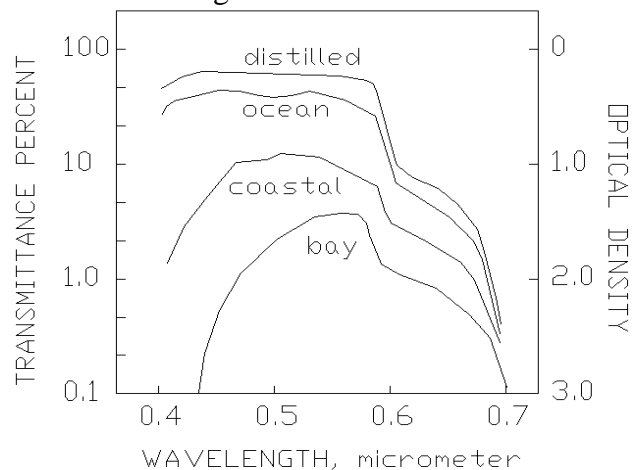
Spectral transmittance for 10m of various types of water

The signatures on satellite images are influenced not only by water depth, but also by water clarity, reflectance of the bottom sediment, and by atmospheric conditions. Therefore the depth associated with image signatures in one island will not necessarily be the same in other areas.

7.4.1 Case study – SHOM's nautical space chart (La spatiocarte marine)

▪ *http://www.shom.fr/sci/spatio/spatio_e.html*

The **SHOM** (Service Hydrographique et Océanographique de la Marine) is in charge of the hydrography and of the nautical cartography in areas under French responsability. Thus, Overseas French Regions (DOM) and Territories (TOM), Southern lands and Antarctica are as concerned as the metropolitan territory is.

But, if the hydrography and cartography of the metropolitan territory and of the Overseas Regions of France (DOM) have reached a high level, numerous regions of the Southern Pacific and of the Indian Ocean are not completely hydrographied. Topographic or bathymetric data are too often fragmented or isolated, even obsolete, in order to allow the publication of nautical charts of high quality.

With the present classic tools, several years would be necessary to carry out hydrographic works on these areas and enable to cartography them matching acceptable standards.

Since several years, satellite imagery is one of the source of data used by the SHOM for nautical chart elaboration. Mariners do not always notice it. Indeed, once processed, the informations provided by satellites are reproduced according to the graphics of classic cartography.

With nautical space charts, a further step was reached because spatial data are becoming the chart basis (terrestrial topography, sea beds geomorphology). These space charts are distributed as the other **SHOM** products.

The satellite images analysis will enable the SHOM to reactualize, within the ten years to come, the nautical charts covering the whole of the South Pacific islands and French atolls.

The studies carried out since the Preliminary Evaluation Program of SPOT (**PEPS** - 1984) on bathymetry have led to the implementation of a water height measurement model:

The results of the studies on bathymetry are directly linked to the characteristics of the SPOT satellite.

The sensors enable to measure the radiometry. Radiometry is the intensity of the radiation reflected by the floor and by the sea floor through the water layer. This intensity is measured within the visible spectrum for well determined frequency bands called "channels". On the images, it can be noticed that deeper the sea bed is, greater the radiance is absorbed and lower is the measured radiometry level.

Spectral resolution enables the observation of objects, in clear waters, up to 22 to 25 meters depths on the XS1 channel and up to 5 to 7 meters depths on the XS2 channel. As for the XS3 channel, it doesn't provide any bathymetric information because the infrared is absorbed by the water. This channel indicates, for the coastal strip, the waters limit when the satellite acquires the scene.
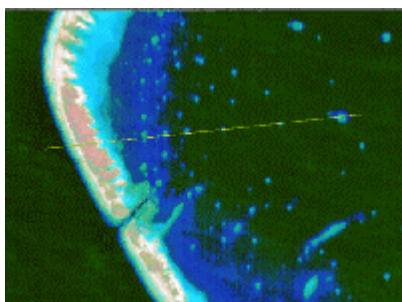
Extract of the SPOT 512/382 scene (12-09-95) Katiu atoll
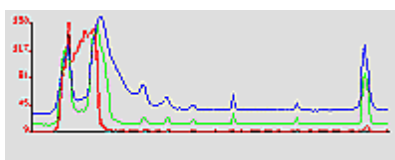
This diagram represents the variations of the levels of radiometry of the three channels (blue for XS1, green for XS2 and red for XS3) along the line (yellow color) of the small image.

Diagram analysis:

A strong variation for the three channels is to be noticed on the left part. This area corresponds to the terrestrial part. It is composed of a barrier reef on which there is some red, it corresponds to the vegetation (coconut trees). On this object, the response

on the channel 3 remains strong. On the contrary, it is to be noticed that the response on the two other channels is very low on this sector of vegetation while it is very strong on both sides of the coral sands.

On the right part of the diagram, an important peak is to be noticed, it corresponds, on the small image, to a well visible coral macro-structure, indeed the most influenced are the channels 1 and 2.

On the central part of the diagram, peaks of variable intensity are to be noticed, they correspond to isolated objects (coral pinnacles). At their level, variations more or less marked are noticeable on channels 1 and 2; these variations correspond to the water height above the shallows. The variation on channel 3 is very low, if not non-existent; it means that the detected pinnacles were awash or submerged when the satellite acquired the scene.

Different shades of blue are also noticeable on the small image (ranging from light to dark blue); these shades show the variations of the water height (from shallow to deep waters). On the diagram showing the different variations for each of the three channels, the XS1 channel is the most influenced.

Modelization of the bathymetry:

The principle of the bathymetric measurement is based on a simple reflection model. This model links the intensity of the radiometric signal measured by the satellite to the depth. It can call on two methods:

* the physical method. It requires the knowledge of all the parameters which preside to this model (water optical properties, seabed reflection coefficient, transparency of atmosphere, ...). But, most of time, failing anything better, these parameters are valued and considered as constant on areas which are supposed to be homogeneous.

* the empirical method. It consists in adjusting the coefficients of the model on the in-situ data taking, if possible, the nature of seabed into account.

The second method is used at the **SHOM**.

In order to avoid the bias due to the choice of the calibrate points of bathymetry, the average values of the soundings, which correspond to homogeneous areas of constant radiometry, are introduced into the model.

Some twenty values (from known soundings values of Ground Control Points) are necessary for an adequate gauging of this model. Its mathematic formula is as follows:

$$Z = A * LN (R1 - R1inf) + B * LN (R2 - R2inf) + C$$

Z is for the depth calculated for the **Ri** radiometry pixel, the suffixes **1** and **2** refer to the **XS1** and **XS2** channels, **Riinf** represents the radiometry in deep waters for the **i** channel, the coefficients **A**, **B** and **C** are obtained with the least squares method from the calibrate measures.

The calculated model provides a single-channel image, on 256 levels, each pixel of the maritime area is represented by a calculated depth and not by a measured radiometry as it used to be.

7.5 Vessel detection

- *http://www.amrs.org/forumvessel.html*
- *http://www.rsi.ca/adro/adro/abstracts/089.html*
- *http://www.rsi.ca/adro/adro/ger_97/abstracts/136.html*
- *http://www.rsi.ca/adro/adro/abstracts/RELATEDiiiiiiiiiiii.html*
- *http://www.ccrs.nrcan.gc.ca/ccrs/tekrd/rd/apps/iceocn/rsatship/shipe.html#paper*
- *Vachon, P.W., J.W.M. Campbell, C. Bjerkelund, F.W. Dobson, and M.T. Rey, 'Validation of Ship Detection by the RADARSAT SAR' submitted to Proc. Pacific Ocean Remote Sensing Conference (PORSEC'96), 13-16 Aug. 1996, Victoria, Canada.*

The use of remote sensing techniques for observing ships has gained substantial momentum over the past decade. Several data sources exist: satellite and airborne radar and optical instruments, as well as ground-based HF radar. Here we focus on spaceborne synthetic aperture radar (SAR). The ERS and Radarsat satellites have provided the first civilian long term sources of space based image data that present an efficient means of wide area monitoring of ocean traffic, both day and night, independent of visibility. Developments in computer power, storage space and communications technology have made it possible to utilize the observations in a matter of minutes or hours if necessary, by processing raw data into imagery, analyzing and disseminating images and/or analysis results to operational organisations both ashore and at sea.

While the vessel detection in SAR data problem initially could be regarded as primarily a military area of interest, a number of civilian issues have emerged where such observations can be very useful: fisheries management and enforcement (see figure), oil pollution monitoring, search and rescue operations, monitoring of illegal transportation of goods and people, and also collection of statistical information to assess requirements for navigational aids and insurance risks.
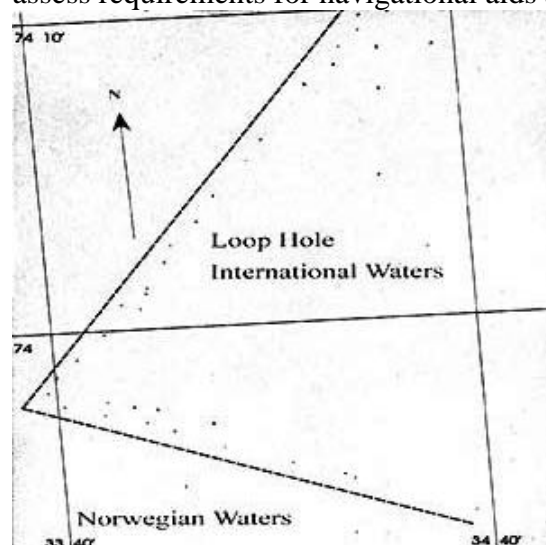
Radarsat image acquired in ScanSAR Narrow Far, 19 Aug. 1996 (Copyright CSA 1996.) Fishing vessels (black dots are lined up along the border between the "Loop Hole" and Norwegian waters (the colors are inverted).

These issues have spurred the establishment of R&D and pilot programs in several countries, aimed at developing efficient methods of exploiting satellite SAR data for vessel traffic monitoring: Australia, Canada, Japan, The Netherlands, Norway, the United Kingdom and the USA to mention a few. Some of these programs have transitioned to operational status. Certainly, it has come to be recognized that SAR observations, when used in conjunction with other sources of information, can provide

important support in making go/no go decisions for deployment of patrol aircraft or vessels.

Key questions from the operational agencies are:
1) What information is available, and how often and how rapidly is it available?
2) What is the minimum detectable vessel size?
3) How reliable is it? i.e. what is the false alarm rate, and what about the ones that are missed?

In the mid 1990's, several algorithms for vessel detection in SAR images were published. This established some baseline estimates, that together with an analysis of satellite orbits and ground station locations, provide some site specific answers to the first question. In addition to vessel position, methods for estimating vessel size, orientation, velocity and type have been developed.
The last four or five years have seen a large effort towards operationalizing algorithms and answering questions two and three. However, validation of marine observations is expensive and effort intensive, and remains a key issue. While significant progress has been made in verifying detection rates and size estimates for large vessels, we have not seen any published results extending to vessels smaller than 20 m under well documented environmental conditions. Other parameters remain mostly unvalidated, at least in the open literature (*the last four www sites listed at the top of this section, give examples*).

A statistical model for RADARSAT's expected ship detection performance was developed based upon experience with ERS-1 SAR data. The model includes ocean clutter, SAR image probability density function, and ship cross section elements. Ship detectibility predictions from the model are shown in the following figure.
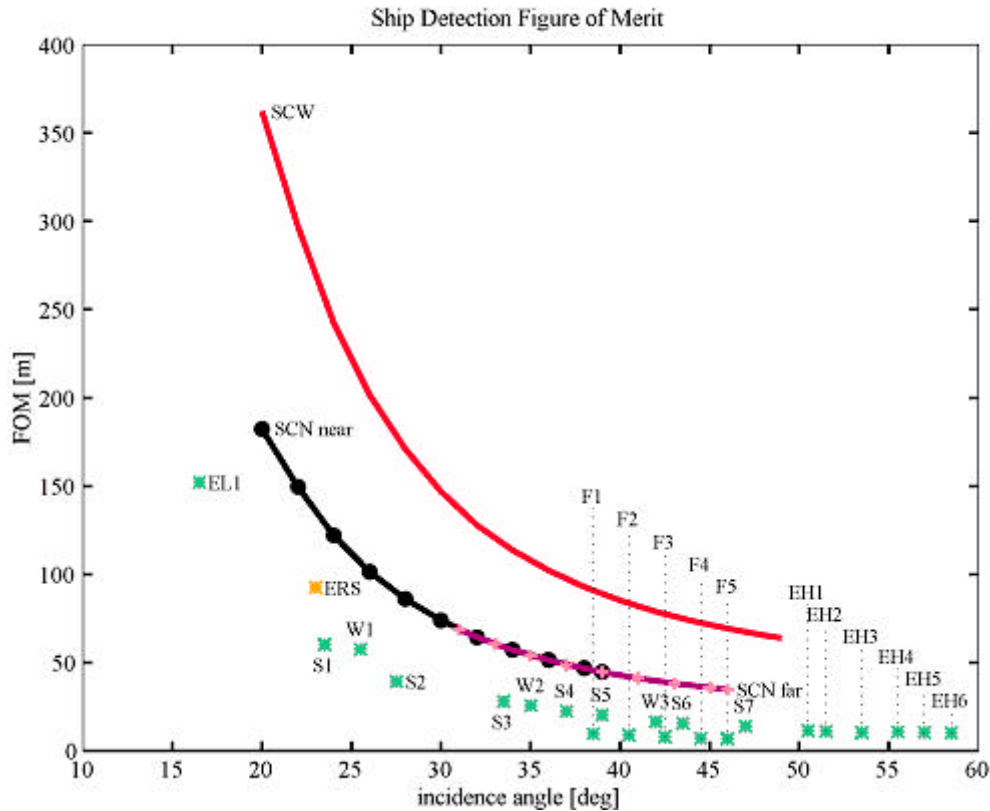
Plot of predictions of minimum detectable ship length as a function of beam mode and incidence angle subject to the assumptions of the model described by Vachon Paris (from the CCRS) et al., 1996. The minimum detectable ship length is expressed as a figure of merit for a 10m/s wind speed and allows for comparison between ERS-1 SAR and RADARSAT's various beam modes.

Key issues for the future include further validation of existing algorithms, and continued access to suitable SAR data. Envisat and Radarsat-2, with their abilities to collect co- and cross-polarized data, will represent significant improvements in vessel detection reliability.

Plans for other commercial SAR missions are also emerging, some with a focus on higher spatial resolution (one to three metres) at the expense of area coverage, possibly at frequencies other than C-band. While this may support extraction of more information about observed vessels, it will not have a similar operational impact as multi-polarized data.

The AMRS Forum on vessel detection can help address some of these issues through facilitating co-ordination of validation efforts and exchange of results and experience. Hopefully, we can initiate exchange of views between people working with different instrument types. Also, there is a role to play in providing market driven requirements to the commercial consortia planning new missions.

8. Digital Video Remote Sensing

▪ *http://www.minvenw.nl/projects/airborn/vid_appl.htm*

8.1 A little history

The use of video as a remote sensing tool started in the late 1960s and early 1970s. The sustained interest of video in the remote sensing is illustrated by the increasing number of papers dealing with video as remote sensing tool and the first 1988 international workshop dealing only with videography. Airborne video systems have been applied in a wide range of topics like forestry, biology, agriculture, coastal zone monitoring, landscape analysis and also in urbanised areas.

The increasing interest is partly due to new technologies such as the use of solid state Charged-Coupled Devices (CCDs) sensors instead of tubes and the introduction of the Super-VHS (S-VHS) recorder. The use of CCDs improved the geometric accuracy and the radiometric performance of the video camera. The development of S-VHS increased the image to a 400-lines resolution instead of the typical VHS (colour) 230-lines resolution. New technologies such as digital video and the advances in computer hardware and processing techniques make the use of video as a survey tool even more attractive.

8.2 General advantages of using video

Today, a number of airborne video systems exists. They all have their specific pros and cons but the general characteristics making video to a valuable sensor can be summarised as follows:

- low cost due to off-the shelf systems,
- real-time and near real-time availability of imagery for visual assessment or image analysis,
- immediate potential for digital processing on the signal,
- acquired data can be checked in real-time making real-time surveys possible,
- high sensitivity of CCD enabling the collection of spectral data in narrow bands (10 nm),
- linking of video data with ground control points and *Global Positioning System* (GPS) information,
- wide spectral sensitivity of modern silicon detectors,
- data redundancy in which images are acquired every 1/25 second producing multiple views of a target or scene.

Basic video technical aspects are itemised as follows :

8.3 The video image

This part discusses the principles of the video image, what is its structure and what kind of information does it contain?

Video signals are subjected to very strict norms. In Europe the so called PAL (Phase Alternate Line) standard defined in the 1950's, is used. Other countries like the USA use NTSC as standard. The different video standards are not compatible with each other. Throughout these pages we will discuss the European PAL video standard.

The PAL standard uses the principle of interlacing as illustrated in the figure. A video image (also called video frame) consists of two fields. The even field contains the even line numbers of a frame, the odd field the odd line numbers. With the PAL standard 25 frames per second (25 Hz) are built up. Consequently fields are produced with a rate of 50 fields per sec (50 Hz).
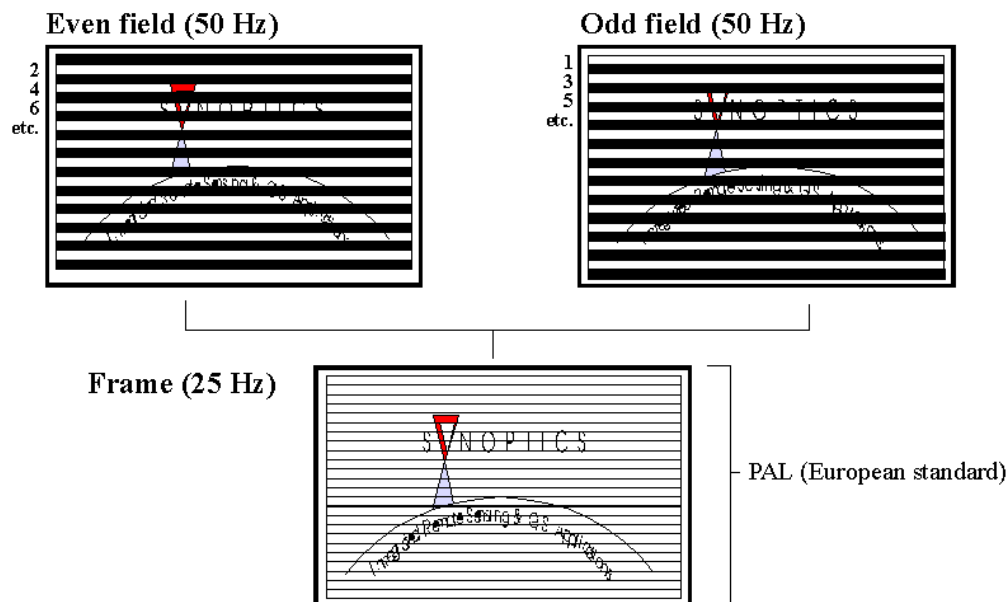
Two aspects are of importance to build up video images; the <u>radiometric</u> and the <u>geometric</u> information.

<u>The video image radiometry</u>
In the camera photosensitive <u>CCD elements</u> build up electronic charges caused by incoming light. The more light 'hits' the CCD element the higher the electronic charge with a maximum of 0.7 Volt - the peak white level - and a minimum of 0 Volt - the blanking level (= black). This information is read out by registers in the camera line by line for each field. Synchronisation pulses (see below) are placed after each line of radiometric information.

<u>The video image geometry</u>
Video image geometry is very important. Each subsequent field should be aligned properly as well as the lines with radiometric information within a field. Synchronisation pulses take care of that. In contrast to the radiometric information the pulses are characterised by negative voltages. This way the recording and display devices can separate geometric information form radiometric information. Roughly there are two types of synchronisation pulses of importance:

- VSync : The start of each field is indicated by a field blanking period or Vertical Synchronisation pulse (VSync); a pulse used to trigger the vertical retrace of a scanning electron gun from the bottom of a field back to the top left.

162

- Hsync : Each line within a field is separated by a negative voltage of -0.3 Volt called the Horizontal Synchronisation pulse ; a pulse used to trigger the horizontal retrace of a scanning electron gun from right to left and one line down.

The following figure shows the information a video line contains:
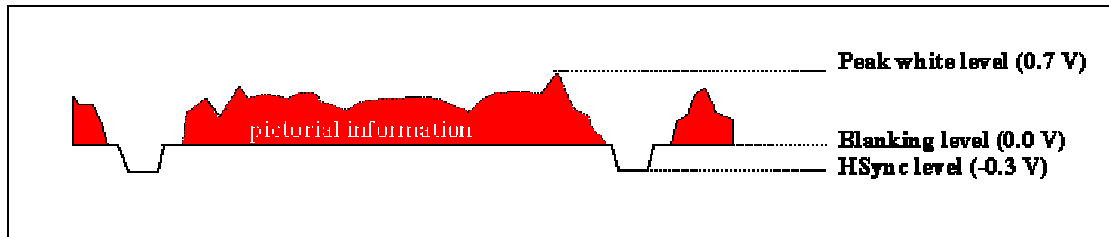
8.4 Charge-Coupled Devices

There are a lot of video cameras available but the modern ones have one thing in common: Charge-Coupled Devices function as photosensitive plates and have replaced the old-fashioned tube cameras. The physics of CCDs is based on the principle of a Metal-Oxide Semiconductor (MOS). Such MOS capacitor is formed by positioning a metal electrode, insulated by a film of silicon dioxide, onto a silicon substrate.
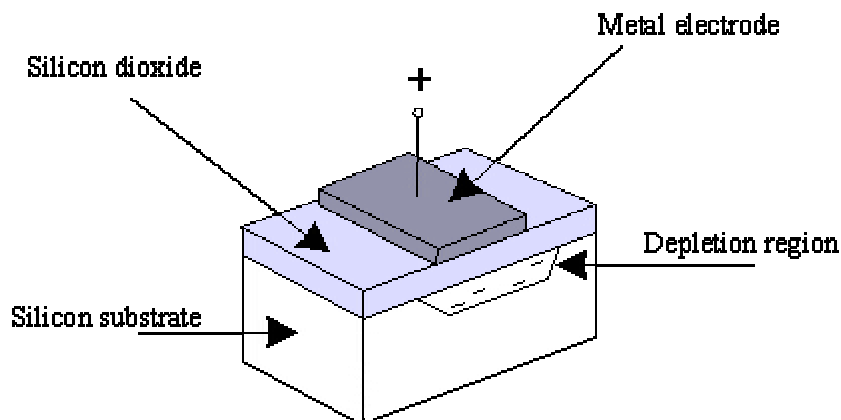
Example of an MOS capacitor generating a depletion region which is proportional to the amount of incoming light.

Semi-conductor light sensors and the use of charge-coupling techniques to move charge around within the device structure are applied in video and photography. Incident light is proportional converted by these sensors to an electronic charge and trapped in the depletion region. The ability to store a quantity of charge is fundamental to the operation of CCD devices. It corresponds to a memory device, but one that stores analogue quantities. A camera requires an Area Imaging Device that can sense the whole picture focused on it by the lens and the optical system: a CCD array. The CCD array analyses the picture brightness with the charge value on each photo sensing element representing the brightness at that point.

The dimension of an CCD array is standardised and is usually given in inches of its diagonal (see Table). These figures you can usually find in the camera specifications as well as the number of CCD elements present in an array.

CCD array size specifications

CCD array (inch)	Horizontal (mm)	Vertical (mm)	Diagonal (mm)
1" CCD	12.6	9.5	15.8
2/3" CCD	8.8	6.6	11.0
1/2" CCD	6.4	4.8	8.0

Each array is made up of thousands of CCD elements. The more elements used in a given picture size, the smaller each will become increasing the geometric resolution but reducing its light sensitivity. The more common cameras usually contain up to 1 CCD array, while the professional (and more costly) cameras may have 3 CCDs arrays. In the latest case a CCD array is available for each of the RGB (Red, Green and Blue) video channels.

The time to which the CCD array is exposed to the light source is determined by an electronic shutter. As CCD elements are more sensible to light then photographic film very high shutter speeds can be used usually ranging from 1/50 sec. to 1/1000 sec. Nowadays even cameras with shutter speed up to 1/10,000 sec exist.

Dependent on the material the CCDs are made of they have a wide radiometric spectrum ranging from the Blue Band (ca. 400 nm) up to Mid Infra-red (ca. 2400 nm). This makes that video can be applied in many different surveys from true colour observations to thermal surveys of the earth surface.

8.5 The geometric resolution of video

One should use common sense and apply video for what it is suited for. One can obtain centimetre geometric resolution but this will go at the expense of the area one video frame is covering, also called field of view (=fov).

If tens of kilometres have to be recorded you can consider to apply airborne video for giving a fast overview of the area under study. If you are interested in a small specific area some kilometres wide you can use video at a higher geometric scale.

To give you some ideas about the geometric resolution of the airborne CIR video system (where CIR stands for Colour Infra-Red) we use you can have a look at the following table based on a camera focal length of 7 mm.

Flying altitude (mtr)	Field of view (mtr)	Geometric resolution (cm)
100	90 x 70	12
200	140 x 105	25
300	275 x 210	38
400	370 x 275	51
500	460 x 350	64
1000	900 x700	125
1500	1350 x 1050	190
2000	1750 x 1300	255
2500	2250 x 1700	315
3000	2750 x 2100	385

8.6 Airborne video data acquisition

Airborne video data acquisition is quite easy. For vertical recording you can either mount the video camera above a hatch present in the bottom of the aeroplane or outside the aeroplane. The latter option requires that the camera should be relatively small and that changes in the camera settings can be set by means of remote control. In any case the camera as well as the videorecorder (=VCR) should be stabilised in the mounting.

Another possibility is of course recording out of hand. In most cases this is not recommended as it may cause severe camera movements. On the other hand it enables freedom of camera handling independent of unexpected movements of the airplane.

After the camera system is set up and various devices are connected the airborne survey can start. During the survey the analogue video signal is stored on. One can also perform a real time survey of the area by looking at the monitor that is connected to the camera. If the camera is linked to a GPS *(Global Positioning System)* receiver every video frame is not only time coded but also GPS coded.

For digital video image analysis a so called frame grabber is required. This device converts the analogue video signal on the tape into a digital format. This process is also called A/D conversion.
A/D conversion is simple. A VCR is linked to a framegrabber present in a PC or Workstation. Running the tape one can see what has been recorded including GPS and /or time code information. This GPS information is very useful for mapping large areas or areas with little points of recognition. If the area of interest has been reached one pushes the button and the analogue video information is converted into a digital video frame ready to be analysed

By means of this technique the areas of interest can be selected in a fast and easy way out of several Gigabytes of potential material. This 'newspaper approach' is very time and cost effective.

8.7 Types of airborne video systems

In the airborne video remote sensing basically two types of video systems exist when we consider multispectral information; multiplex camera systems and single camera systems. On this part we look at some of the characteristics of these two systems.

8.7.1 Multiband information

In the spectral analysis of images we are especially interested in multiband information as the use of more bands makes discrimination between objects more easy. Especially the use of a NIR band (700 nm - 1300 nm) in the system makes that different vegetation types can be discerned from each other.

In the USA initially single band systems were used using black & white (B&W) cameras. These cameras had visible (400 - 700 nm) or NIR sensitivity (700 - 900 nm). The systems were portable, cheap and easy in operation. Due to the single band information these systems were not very versatile and limited in their application.

8.7.2 Multiplex systems

In a following stage multiband systems has been developed by combining several B&W video cameras. Each camera represents one spectral band by placing a narrow band filter in front of the camera. Each camera is connected to a VCR and a monitor. An encoder is integrated to provide false colour information out of the single bands.

Multiplex video system
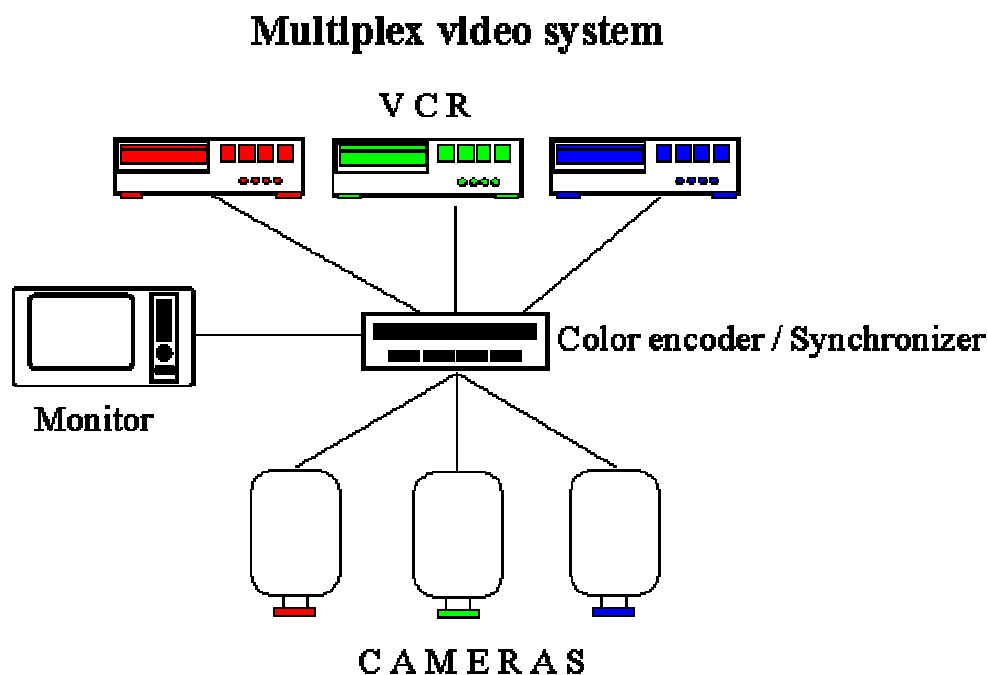
These so called multiplex video systems have the following advantages with respect to multiband single camera systems:
- Flexible spectral band choice
- Narrow spectral bands (10 nm)

The disadvantages are:

- Bulky/heavy
- Difficult camera alignment
- Moderate to high costs (c. USD 20,000)
- No oblique recording possible

8.7.3 Single camera systems

Another line was set out by modifying standard true colour cameras into Colour Infra-Red (CIR) video systems. One camera yields three band information in the Green, Red and NIR band. The camera is connected to a VCR and a monitor.

MULTIBAND SINGLE CAMERA SYSTEM

Monitor

VCR

Three band spectral information
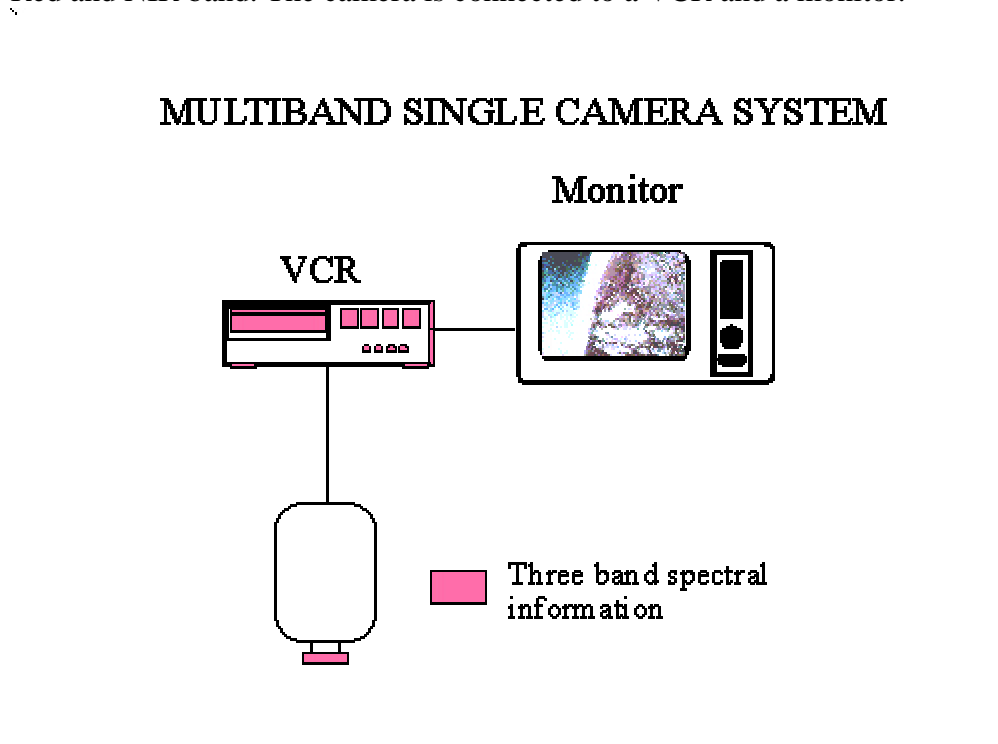
Synoptics uses - when required - an industrial S-VHS 3 CCD camera called Silvacam.

A single multiband camera system has the following advantages:
- Very flexible in use
- No camera alignment required
- Oblique recording possible
- Very high image quality
- Real time survey in the False Colour spectrum

The disadvantages are:
- Fixed spectral bands
- Relative broad spectral bands
- Moderate to high costs (c. USD 20,000)

8.8 The Silvacam camera

Synoptics uses a single camera Colour Infra-Red (CIR) video system called Silvacam for surveys requiring false colour data The system is very simple and flexible and looks like a true colour industrial camera.

The camera is converted from a true colour system into a false colour system by using the same principles as CIR photography; the green spectral information is diverted to the Blue video channel, the Red spectral information to the Green video channel and the NIR information to the red video channel. Each video channel (Red, Green, Blue) is represented by a CCD array. The incoming light is split into the three respective colours and lead to the respective channels by means of dichroic mirrors.
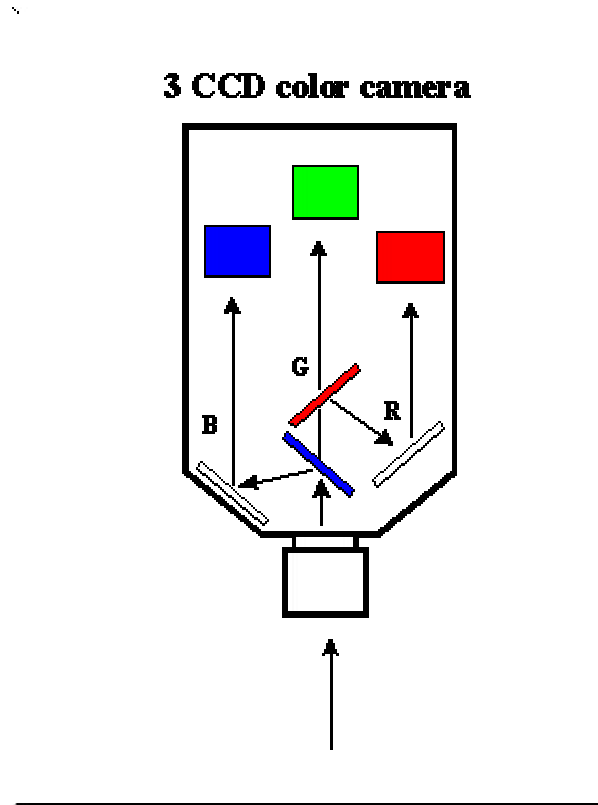
By placing internal filters in the camera it obtains its spectral False Colour characteristics; a passband in the Green (505 - 580 nm), one in the Red (580 - 680 nm) and one in the NIR (770 - 845 nm).

LAI = Leaf Area Index
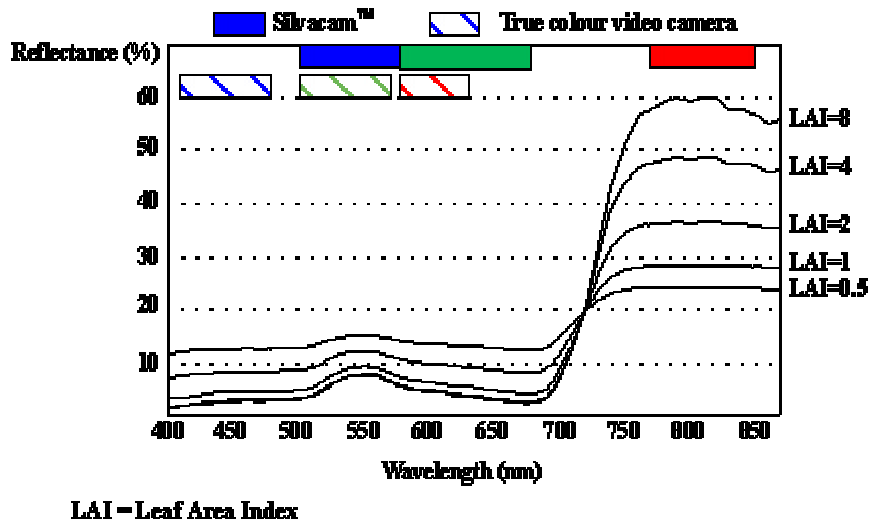
The video information is stored on tape and can be analysed by running the tape by means of a VCR which is connected to a framegrabber. Once the area of interest is visible, a push of the button is sufficient to convert the analogue video signal into a digital image.

8.8 Applications

8.8.1 Coastal monitoring in the Netherlands

In 1995 the Survey Department of the Netherlands launched a proposal for the introduction of new techniques for the acquisition of elevation data and thematic imagery of the coastal zone. The techniques consist of airborne laser scanning for elevation measurement and airborne video for thematic data.

The laser scanning (see in the chapter about Airborne Laser Scanner) is to replace, in due course, the traditional photogrammetric approach to obtain profiles of the beach and foredunes. With the introduction of laser scanning techniques and video the Survey Department aims to improve efficiency and performance. As of 1996, a DTM of the beach and foredunes of the Dutch coast will be provided yearly, based on laser scanning data.

Also, once every three year the complete coastal zone (incorporating dunes) will be surveyed. Video data will provide information on e.g. vegetative cover, development and migration of blow- outs. By means of draping the video images over the laser-DTM a good visual impression of the coastal area is obtained. With the provision of a laser-DTM in combination with video as basic data sets the Survey Department aims to contribute to integrated management of the coastal zone.

8.8.2 The Southwest Washington Coastal Erosion Study, and the ARGUS program

The Southwest Washington Coastal Erosion Study is a United-States Federal-State-Local cooperative research program to address the coastal geology, processes, and natural hazards of the Southwest Washington coast.

Remote Video Monitoring (RVM) sytems provide a means of automatically acquiring video data from remote locations and returning them to a central laboratory computer for processing. Since 1991, the U.S. Geological Survey's Center for Coastal Geology has been building RVM capabilities through a cooperative agreement with Oregon State Univerisity where video data acquisition and processing tecniques have been under development for 10 years.

Two video cameras have been positioned on Willapa Bay near North Cove, Washington to monitor coastal erosion and wave processes on the northern shore of the bay. Video images are taken hourly by two cameras. The data are down-loaded daily by the U.S. Geological Survey. The images are stored in the archive by year, day, camera, and hour.

This project began as a result of the ARGUS program, developed under the guidance of Professor Rob Holman of the Coastal Imaging Lab, Oregon State University. **Video-based monitoring stations** are being placed at important coastal sites around the world. The initial information that they have provided has proved as surprising as the early satellite shots

Critical to progress in understanding nearshore dynamics is the ability to make physical measurements under natural conditions. This is traditionally accomplished with surveys of beach response and with arrays of fixed-point measurements of fluid motions. To carry out even a single survey of the underwater bathymetry of the beach is expensive and often impossible under any but the mildest wave conditions. A current meter for measuring waves and currents at a single point in space costs about $7,500. To make sense of the measurements requires installing an array of at least ten current meters and attempting to maintain them in a hostile environment. Manpower costs are high and specialized equipment is usually needed. Thus, our knowledge of the nearshore has been limited by the difficulties and cost of measurement.

In contrast, a single Argus station costs approximately $10,000. From a station, **useful data can be collected over miles of beach** and, in time, **every hour for years**. Stations are unmanned, data are transferred over internet and maintenance costs are virtually zero. While a measurement taken with an Argus station is usually not as accurate as that from in-situ measurements, the collection of data over years has revealed characteristics about beaches that were never suspected.

The key part of an Argus station is one or more cameras pointed obliquely along a beach. The camera is connected to an image processor contained in a small personal computer. An Argus station needs a home, with power, protection from the elements,

and a phone line to communicate with the lab computers. Once installed, the station begins to collect images. Currently, stations are typically programmed tot take one snapshot and one time exposure image every hour. This occurs automatically. **The data are downloaded daily**, usually during the middle of the night when no images are being taken and phone rates are low. Once in the lab, the data are processed and added to the growing database from different Argus sites. **Support data from close-by tide gages and wave-measuring instruments are also incorporated into the database.**

Currently Argus stations are installed in the U.S., Hawaii, Australia, New Zealand, England and the Netherlands, and the images from these stations are available at: **http://cil-www.oce.orst.edu:8080**.

Beach camera, North of Noordwijk, the Netherlands

At this point the primary use of Argus data lies in time exposure images. By averaging the pattern of wave breaking over about ten minutes, the locations of submerged sand bars are clearly seen. From each image, measurements can be extracted of the sand bar scales and the behaviour of the beach system can be compared to the varying wave and tide forces.

A second application concerns the extraction of beach contours. From the time exposure images the 10 minute-mean shorebreak or waterline can be located. If the tide level is known, the contour of the beach at that level can be determined. By doing this at different tide levels, a contour map of the beach can be composed.

The use of time exposures is only one of many techniques that are now or have been developed and tested. Pilot studies show the ability to measure the strength of longshore currents as well as the period and angle at which incident waves approach the beach.

9. Altimetry

Both aircraft- and space-based altimeters yield direct measurements of elevations along narrow path footprints. These are readily plotted as profiles; a series of parallel profiles serve as a framework for contouring. Altimeters work by sending self-generated (active) signals to reflecting surfaces and measuring total roundtrip times from the targets (solid land, tree tops, ice, or water) over which they move. The signals can be either radio pulses (radar included) or light pulses (laser).

9.1 Laser Altimety

Laser is a shortened term for light amplification by stimulated emission of radiation. A typical solid laser device is a chromium-doped ruby (Al_2O_3). When an external source of radiation (as from an enclosing flash tube) acts on a shaped (usually cylindrical) ruby crystal, one end of which is silvered to act as a mirror, the chromium (Cr) ions dispersed in its lattice are excited to a new energy state (electrons raised to some new orbital level), The laser state is generated by stimulated emission when the electrons drop back to a lower energy level (remember, from the Introduction, the formula: E = hc/). Much of the light passes sidewards out of the crystal but light moving along the axial zone of the cylinder encounters other Cr ions as it undergoes repeated reflections, further exciting them and causing a buildup or amplification, called optical pumping, of light energy until discharged as a pulse (controlled by the flash lamp) of intense coherent radiation at some discrete wavelength (actually a narrow frequency range; for the ruby laser the light is pinkish) that is collimated to form a unidirectional beam which can be aimed. Other laser materials include gallium arsenide and excited gases such as neon or helium. Both visible and infra-red wavelength light can be generated in this way.

Laser altimetry has been conducted from aircraft platforms for several decades. This type of Remote Sensing is sometimes referred as lidar (for LIght Detection And Ranging). Timing devices allow extremely precise determination of transit times, so that accuracies of a few centimeters in determining elevations (and their differences or relief) along the traverse are attainable. Aircraft can be scheduled to fly in good weather, offsetting the main disadvantage of using lasers coming from cloud interference with the light beam. Lidar instruments can be operated both as profiliers and as scanners, and can, of course, be operated both day and night. Lidar can serve either as a ranging device to determine altitudes (topography mapping) or to conduct particle analysis in air.

Laser altimeters are beginning to be flown on space vehicles. The Orbital Profiling Laser Altimeter is a pulsed, time-of-flight optical (1.024 μm) sensor that sends 10 pulses per second (pps) in a narrow beam (footprint 30-150 m diameter; sampling in 150 to 700 m [492-2296 ft] intervals). Operated on the Shuttle Endeavor in January, 1996, it achieved a vertical precision of 0.75 m [29.5 inches]. Each laser shot fired has a dwell time of only 2-10 nanoseconds, within the 1-10 nsec resolution of the timing electronics; with this rapid return rate, ground positions are readily determinable provided space vehicle position is known.

Essentially the same sensor has been launched as the Mars Global Altimeter to the Red Planet at the end of 1996. A dedicated satellite (ALT) in the EOS, scheduled for launch in the year 2003, will mount the Geoscience Laser Altimeter System (GLAS) designed to measure ice-sheet topography as well as cloud and atmospheric properties, but will also survey selected land and water surfaces. The 40 pulse/sec beam is generated from a neodymium:yttrium-aluminum- garnet (Nd:YAG) crystal producing a two level emergy output :1.064 μm (infrared) for surface surveys and a 0.532 μm signal for atmospheric (clouds and aerosols) measurements. Each pulse will spread over a ground spot of 70 meters (230 ft), with separation between pulses amounting to 170 m (558 ft). From its 705 km (440 miles) orbital altitude, the instrument can measure height differences on the ice of 10 cm (4 inches) precision; characteristics of the altered return pulse indicate surface roughness.

Lasers are also reliable and very accurate tools for determining distances to satellites in orbit. Retroreflectors are small prisms (typically quartz) mounted on the spacecraft. Even at their high orbital speeds, these spacecraft can be targeted to intercept a laser beam pulse directed at them from Earth-based stations which is then bounced back by the reflectors. Along with GPS, this is one way in which the orbital position of a satellite can be precisely fixed. The astronauts in later Apollo missions left retroreflectors on the Moon's surface as a means of learning more about its orbital motions, including its recession from Earth.

9.2 Radar Altimetry

From space, real aperture microwave (radar) altimeters have served to measure various aspects of sea state, such as heights relative to the geoid, wave geometries, sea ice, and, indirectly, circulation characteristics such as currents and eddies. Like lasers, the signals are dispatched as short pulses, but at the longer radar wavelengths these signals can penetrate cloud cover. A typical radar altimeter operates in the 13.5 Gigahertz (radio/microwave) portion of the Electromagnetic Frequency Spectrum. Again, using roundtrip times the instrument acts as a ranger to determine distance to target. Radar altimeters designed to secure data from the ocean surface use small antennas that dispatch long wavelength pulses through a wide beam that produces a broad footprint whose size is determined by pulse length. This pulse-limited type operates best on smooth surfaces but analysis of the degree of "stretching" of backscattered pulses (echoes) yields information on surface wave heights (roughness). For land measurements, especially on surfaces of high relief, the beam-limited altimeter requires a larger antenna (a practical limitation) capable of generating a narrow beam (hence, smaller footprint that better discriminates changes in slope) and shorter wavelengths. The orbital position of the sensor platform must be determined with high precision and accuracy to establish the position of the geoid and local departures of the surface from it; modifications of the signal by the atmosphere need to be accounted for, usually through corrections made from data acquired by an accompanying radiometer.

The first spaceborne altimeter experiment was conducted during the 1973 Skylab mission. Next was GEOS-3 (Geodynamics Experimental Ocean Satellite) in 1975 which measured height differences greater than 60 cm. This was followed in 1978 by JPL's Seasat (which failed after 99 days but in the interim returned extremely valuable data over sea and land). Its altimeter, one of five instruments including a SAR, performed at 10 cm range precision. Among its remarkable products was a map of the morphology of the ocean floor, reproduced here:
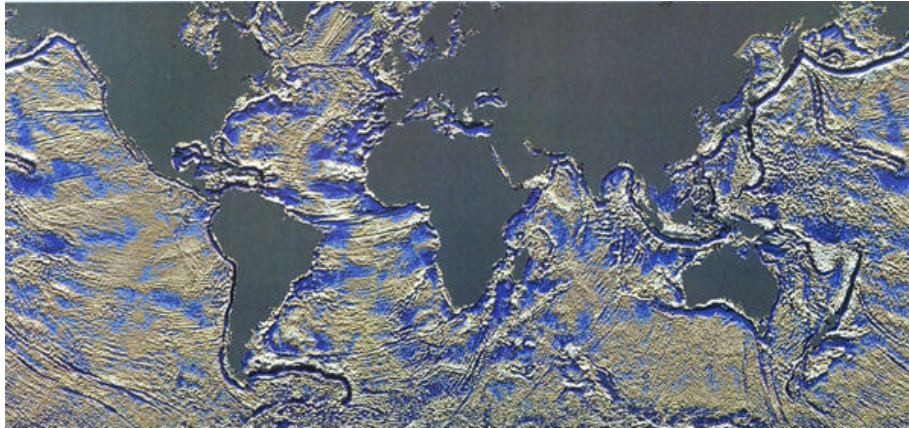
The actual data source for this was the global map of sea surface elevation, which was digitally adjusted to remove effects of the lunar and solar tides and further corrected for bathymetric variations. What appears to be a topographic expression of the floor is actually the distortion of the surface gravitational field, expressed as broad variations in sea surface heights, as controlled by subsurface mass variations and their distribution under variable floor relief. Discernible are the many transform fault zones in the ocean plates, along with spreading ridges, trenches, and seamounts.

9.3 Radar Altimetry over the oceans

- *Chelton, Dudley B., WOCE/NASA Altimeter Algorithm Workshop, U.S. WOCE Technical Report No. 2, 70 pp., U.S. Planning Office for WOCE, College Station, TX 1988.*
- *http://topex-www.jpl.nasa.gov/education/altimeter-basics.html*
- *http://earth.esa.int/l2/4/eeo4.64*
- *http://www7300.nrlssc.navy.mil/altimetry/*

9.3.1 Measuring Ocean Topography for Understanding and Predicting Climate Change

- *http://topex-www.jpl.nasa.gov/jason1/sci_and_data/sci_and_data.html*

Most of the heat stored in the earth's hydrosphere resides in the ocean. The upper 3 meters of the ocean contains the same amount of heat as that stored in the entire atmosphere. This enormous reservoir is moved around the world by ocean currents, affecting the earth's climate and its variability in a profoundly complicated manner. The ocean circulation is fundamentally turbulent, always changing, and its observation, especially on a global scale for addressing its climatic impact, has posed a tantalizing challenge to oceanographers for a long time.

Surface geostrophic currents, strongly linked to the circulation of the deep ocean, can be determined from the ocean topography, defined as the height of ocean surface relative to one of the earth's equi-geopotential surfaces, called the geoid. In addition, ocean topography directly reflects the heat content of the ocean and its changes. Knowledge of the ocean topography is thus very useful to the determination of ocean circulation and its heat transport.

The concept of using a spaceborne radar altimeter to measure ocean topography was formulated in the late 1960s. The concept was first demonstrated by Seasat (1978), followed by Geosat (1985-89) and reached its current state-of-the-art via the Joint US/France TOPEX/POSEIDON Mission (1992-present). The measurement principle is straightforward but the challenge is to reach the exceptionally demanding accuracy and precision at the level of one centimeter for adequately determining the oceanic

transport of mass, heat, freshwater, and chemicals, to which the earth's climate system is extremely sensitive.

9.3.2 Data sources

Information about the different altimeters from which the altimetry data is gathered. These include:

9.3.2.1 Geosat Follow-On

The Geosat Follow-On satellite data is sent directly to the Payload Operation Center (POC) at the Naval Oceanographic Office (NAVO). Orbit solutions are produced by both the Naval Satellite Operations Center (NAVSOC) and at the Jet Propulsion Laboratory (JPL) based on the Global Positioning System receivers onboard the satellite. The orbit solutions are merged with the satellite data and additional geophysical corrections (atmospheric pressure, tides, ...) at the Altimeter Data Fusion Center (ADFC) in NAVO within 48 hours of measurement. The GFO satellite has an exact repeat period of approximately 17 days.

9.3.2.1 TOPEX/Poseidon

- *http://podaac-www.jpl.nasa.gov/toppos/*

The TOPEX/POSEIDON mission was designed to provide information about the changing topography of the world's oceans which, in turn, helps scientists to understand the ocean's role in the global climate. The TOPEX/POSEIDON satellite was launched in August 1992 and was expected to operate through September, 1998. The satellite is currently in operation. TOPEX/POSEIDON measures the global ocean topography every 10 days.

TOPEX/POSEIDON is jointly conducted by the United States' National Aeronautics and Space Administration (NASA) and the French Space Agency, Centre National d'Etudes Spatiales (CNES), for studying the global circulation from space.

The specific goals of the TOPEX/POSEIDON mission are:

1) Measure sea level in a way that allows the study of ocean dynamics, including the calculation of the mean and variable surface geostrophic currents and the tides of the world's oceans.

2) Process, verify and distribute the data in a timely manner, with other geophysical data, to science investigators.

3) Lay the foundation for a continuing program to provide long-term observations of the oceanic circulation and its variability.

9.3.2.2.1 Mission requirements

To ensure that science and mission goals are accomplished by the TOPEX/POSEIDON Mission, the following requirements were established:

Accuracy of Sea-level Measurements

Each measurement of sea level shall have a precision of +2.4 cm and an accuracy of +14 cm (1 standard deviation) for typical oceanic conditions, with small geographically correlated errors. In this context, precision is the ability to determine changes in sea level over distances of 20 km, and accuracy is the uncertainty of each measurement of sea level when expressed in geocentric coordinates.

Sampling Strategy

Sea level shall be measured along a fixed grid of subsatellite tracks such that it will be possible to investigate and minimize the spatial and temporal aliases of surface

geostrophic currents and to minimize the influence of the geoid on measurements of the time-varying topography.

Tidal Aliases

Sea level shall be measured such that tidal signals will not be aliased into semiannual, annual, or zero frequencies (which influences the calculation of the permanent circulation) or frequencies close to these.

Duration and coverage

Sea level shall be measured for a minimum of three years, with the potential to extend this period for an additional two years.

9.3.2.2.2 Sensors on board the TOPEX/POSEIDON

The science and mission goals are carried out with a satellite carrying six science instruments, four from NASA and two from CNES. They are divided into operational and experimental sensors as follows:

(A) 4 operational sensors

(1) Dual-frequency Ku/C band Radar Altimeter (NRA) (NASA)

The NRA, operating at 13.6 GHz (Ku band) and 5.3 GHz (C band) simultaneously, is the primary sensor for the TOPEX/POSEIDON mission. The measurements made at the two frequencies are combined to obtain altimeter height of the satellite above the sea (satellite range), the wind speed, wave height and the ionospheric correction. This instrument is the first spaceborne altimeter that uses two-channel measurements to compute the effect of ionospheric free electrons in the satellite range measurements. It is redundant except for the microwave transmission unit and the antenna.

(2) Three-frequency TOPEX Microwave Radiometer (TMR) (NASA)

The TMR measures the sea surface microwave brightness temperatures at three frequencies (18 GHz, 21 GHz and 37 GHz) to provide the total water-vapor content in the troposphere along the altimeter beam. The 21 GHz channel is the primary channel for water-vapor measurement. It is redundant (21A and 21B). The 18 GHz and 37 GHz channels are used to remove the effects of wind speed and cloud cover, respectively in the water-vapor measurement. TMR data are sent to CNES for processing along with their altimeter data. The measurements are combined to obtain the error in the satellite range measurements caused by pulse delay due to the water vapor.

(3) Laser Retroreflector Array (LRA) (NASA)

The LRA reflects signals from network of 10 to 15 satellite laser tracking stations to calibrate NRA bias and to provide the baseline tracking data for NASA precise orbit determination.

(4) Dual-frequency Doppler tracking system receiver (DORIS) (CNES)

The DORIS system uses a two-channels receiver (1401.25 MHz and 2036.25 MHz) on the satellite to observe the Doppler signals from a network of 40 to 50 ground transmitting stations. It provides all-weather global tracking of the satellite for CNES precise orbit determination and an accurate correction for the influence of the ionosphere on both the Doppler signal and altimeter signals.

(B) 2 experimental sensors

The two experimental instruments are intended to demonstrate new technology.

(1) Single frequency Ku band Solid State ALTimeter (SSALT) (CNES)

The SSALT, operating at a single frequency of 13.65 GHz (Ku band), will validate the technology of a low-power, light-weight altimeter for future Earth-observing missions. It shares the antenna used by the NRA; thus only one altimeter operates at any given time. Measurements give the same geophysical information as NRA's.

However, since this sensor uses a single frequency, an external correction for the ionosphere must be supplied.

(2) Global Positioning System Demonstration Receiver (GPSDR) (NASA)

The GPSDR, operating at 1227.6 MHz and 1575.4 MHz, uses a new technique of GPS differential ranging for precise orbit determination.

9.3.2.2.3 Orbit of the TOPEX/POSEIDON

The orbit chosen for the TOPEX/POSEIDON mission is a compromise among a number of conflicting requirements. It provides broad coverage of the ice free oceans as frequently as possible without aliasing the tides to unacceptable frequencies, and it is high enough to ease the precision of the orbit determination process in minimizing the atmospheric drag. The reference (equatorial) altitude of the satellite is 1,336 km, the equatorial cross-track seperation being 315 km.

9.3.2.2.4 Data retrieval

The TOPEX data is retrieved from a NAVO computer set up in the TOPEX data stream at JPL. This computer converts the raw satellite telemetry data into physical units. The data is then passed on to the ADFC at NAVO. Orbit solutions produced at JPL are also passed on to the ADFC. At NAVO, additional data fields are added to the satellite data. These include estimates of the atmospheric pressure. The orbit solutions from JPL are merged with the altimeter data and the geophysical corrections, and the sea surface height measurements are made available within 48 hours.

The POSEIDON instrument is a solid state altimeter built by the French space agency CNES. The data from POSEIDON are processed at the French Processing and Archival Facility (PAF).

9.3.2.3 Jason

- *http://topex-www.jpl.nasa.gov/jason1/*

Jason is an oceanography mission to monitor global ocean circulation, discover the tie between the oceans and atmosphere, improve global climate predictions, and monitor events such as El Niño conditions and ocean eddies.

The Jason-1 satellite carries a radar altimeter and it is a follow-on mission to the highly successful TOPEX/Poseidon mission. It is joint mission between France and USA. The satellite will be launched in May 2000.

The Sea Level measurement accuracy of Jason is required to be better than 4.2cm, with 2.5cm set as the goal. Data coverage will be between 66 deg N and 66 deg S.

9.3.2.4 ERS-2

- *http://earth.esa.int/#4br*

The ERS-2 data is provided by the European Space Agency (ESA) through the National Oceanic and Atmospheric Administration (NOAA) Laboratory for Satellite Altimetry. Orbit solutions are provided by Delft University. The data are passed from NOAA to the ADFC, and are available within 48 hours. The ERS-2 satellite has an exact repeat period of approximately 35 days.

On-board systems of the ERS-2:

AMI - active microwave instrument consisting of a synthetic aperture radar (SAR) and a wind scatterometer (both in the C-band).

The purpose of the Wind Scatterometer is to obtain information on wind speed and direction at the sea surface for incorporation into models, global statistics and climatological datasets. It operates by recording the change in radar reflectivity of the

sea due to the perturbation of small ripples by the wind close to the surface. This is possible because the radar backscatter returned to the satellite is modified by wind-driven ripples on the ocean surface and, since the energy in these ripples increases with wind velocity, backscatter increases with wind velocity.

The three antennae generate adar beams looking 45deg. forward, sideways, and 45deg. backwards with respect to the satellite's flight direction. These beams continuously illuminate a 500 km wide swath as the satellite moves along its orbit. Thus three backscatter measurements of each grid point are obtained at different viewing angles and separated by a short time delay. These `triplets' are input into a mathematical model to calculate surface wind speed and direction.

Spatial resolution: >=45 km (along and across track)

Localisation accuracy: +-5 km (along and across track)

Wind direction range/accuracy: 0 - 360deg. / +-20deg.

Wind speed range/accuracy: 4 m/s - 24 m/s / 2 m/s or 10 %

RA- radar altimeter: takes precise measurements of the distance from the ocean surface and of wave heights. The Radar Altimeter is a Ku-band (13.8 GHz) nadir-pointing active microwave sensor designed to measure the time return echoes from ocean and ice surfaces. Functioning in one of two operational modes (ocean or ice) the Radar Altimeter provides information on significant wave height; surface wind speed; sea surface elevation, which relates to ocean currents, the surface geoid and tides; and various parameters over sea ice and ice sheets.

The Radar Altimeter operates by timing the two-way delay for a short duration radio frequency pulse, transmitted vertically downwards. The required level of accuracy in range measurement (better than 10 cm) calls for a pulse compression (chirp) technique. In ocean mode a chirped pulse of 20 micro-s duration is generated with a band width of 330 MHz. For tracking in ice mode an increased dynamic range is used, obtained by reducing the chirp bandwidth by a factor of four to 82.5 MHz, though resulting in a coarsr resolution.

ATSR - along-track scanning radiometer (operating in the infrared and visible ranges): measures sea surface temperatures and the vegetation cover of land surfaces. The ATSR consists of two instruments, an Infra-Red Radiometer (IRR) and a Microwave Sounder (MWS). Both are nationally funded experiments resulting from an ESA Announcement of Opportunity for a scientific add-on package.

The IRR is a four-channel infra-red radiometer (at wavelengths of 1.6, 3.7, 10.8 and 12 ㎜) used for measuring sea-surface temperatures (SST) and cloud-top temperatures . It was designed and constructed by a consortium, consisting of Rutherford Appleton Laboratory , Oxford University , Mullard Space Science Laboratory , UK Meteorological Office and CSIRO in Australia.

The MWS is a two channel passive radiometer (at 23.8 and 36.5 GHz) designed and built under the responsibility of Centre de Recherche en Physique de l'Environnement (CRPE). The MWS is physically attached to the IRR and its data is merged with that of the IRR prior to transmission to the ground.

GOME - global ozone monitoring experiment, an absorption spectrometer which measures the presence of ozone, trace gases and aerosols in the stratosphere and troposphere. GOME, a nadir-scanning ultraviolet and visible spectrometer for global monitoring of atmospheric Ozone, was launched on-board ERS-2 in April 1995. Since summer 1996, ESA has been delivering to users three-day GOME global observations of total ozone, nitrogen dioxide and related cloud information, via CD-ROM and internet. A key feature of GOME is its ability to detect other chemically active atmospheric trace-gases as well as aerosol distribution.

MS - microwave sounder: supplies data on atmospheric humidity.

PRARE - precise range and range rate equipment: ERS orbit and trajectory determination.

LRR - laser reflector: determines satellite position using ground-based laser stations.

IDHT - instrument data handling and transmission: temporary on-board data storage by means of two 6.5 GBit tape recorders, equivalent to the data volume acquired in one orbit. Recording, formatting and transmission of data at 105 Mbit/s (transmission of SAR imaging data in real time), or 15 Mbit/s (transfer from tape recorder).

9.3.2.5 NRL Layered Ocean Model

The NRL Layered Ocean Model (NLOM) is a primitive equation numerical ocean model with the vertical coordinate system provided through set of layers with differing densities. The horizontal resolution used here is 1/4 degree, though present manifestations are run at 1/16 degree resolution globally. The bottom topography is also included. The model is implemented at the Fleet Numerical Meteorology and Oceanography Center (FNMOC). The altimeter data produced daily at NRL from the ADFC are sent to FNMOC where they are assimilated into the NLOM. The model is forced by 6-hourly wind stress fields produced at FNMOC.

9.3.2.6 Modular Ocean Data Assimilation System (MODAS)

MODAS performs quality checking and optimum interpolation of ocean observations, including temperature, salinity and Sea Surface Height (SSH). The SSH and SST appearing in the plots were gridded by optimum interpolation using altimeter and AVHRR MCSST products from the Naval Oceanographic Office. The covariance functions used in the interpolations were derived from several years of satellite-based SST and SSH observations. MODAS also computes synthetic three-dimensional grids of ocean temperature and salinity based on the gridded SST and SSH using climatological relationships between subsurface temperature and SST and SSH derived from historical profile observations. Salinity is computed from the derived temperature using local climatological relationships between temperature and salinity.

9.3.3 Data Processing

After data aquisition, the data from each altimeter is processed in a similar manner:

9.3.3.1 Initial Data Processing

The Data processing begins with the Geophysical Data Records (GDRs) of each satellite. These are the data records distributed by the ADFC, and they contain all the information needed to correct the satellite SSH measurement for atmospheric effects as well as information to remove processes that are not of interest (inverse barometer pressure loading, solid earth tides, and ocean tides).

The following geophysical corrections are made to the altimeter data:

- dry troposphere path delay
- wet troposphere path dely
- ionosphere path delay
- electromagnetic bias
- static inverse barometer correction

Significant wave height and wind speed are also extracted from the GDRs.

The data are quality controlled at this step by verifying the quality flags, the RMS variability of the high rate measurements about the one sample per second measurements, detecting exceptionally high AGC quantities, detecting questionable

water vapor corrections based on onboard instruments, and detecting questionable ionosphere corrections based on onboard instruments.

The sea surface height resulting from the GDRs is a measurement of the position of the sea surface relative to the reference ellipsoid. Corrections are made to ERS-2 so that the data are relative to the same reference ellipsoid as TOPEX and GFO.
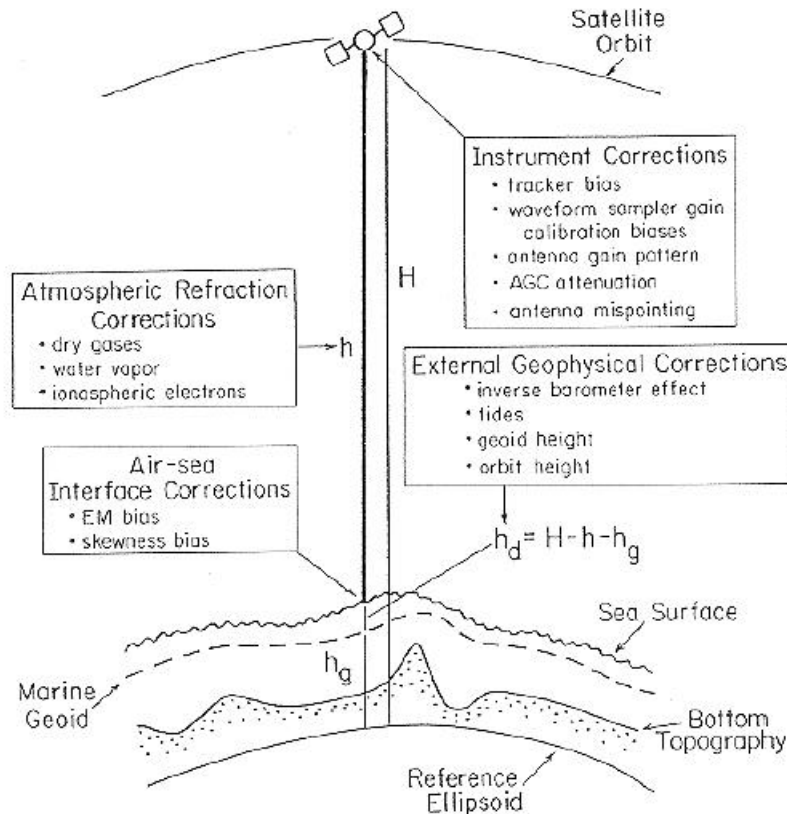
A schematic summary of altimeter measurements and the corrections that must be applied to obtain the dynamic sea surface elevation h_d. The altimeter range measurement is h, and H and h_g are the orbit height and geoid height, respectively, relative to a reference ellipsoid approximation to the earth's surface.

9.3.3.2 Interpolation

The measurement positions of the altimeter change from one repeat pass to another. Thus the SSH data must be interpolated along the ground tracks to fixed reference points. These reference points were produced by the University of Colorado Center for Astrodynamics Research (CCAR). The points are derived from past altimeter missions and are spaced by one second intervals along ground track solutions.

A quality control is made prior to interpolation to detect gross outliers. First, for each SSH measurement, other surrounding SSH values are used to interpolate to the point using a second order quadratic. The RMS difference between the interpolated and actual values is then computed, and values that are significantly different from the interpolated value are removed. The remaining SSH are then interpolated to the the reference ground tracks.

9.3.3.3 Tide removal

A global numerical model for the dynamic ocean tide, the tidal loading, and the long period tides is then removed from the altimeter data. The tide solution is the Grenoble 95.2.1. We use this to insure a consistent tide model is removed from all data. One

drawback to this is that the model solution is provided on a ¼ degree grid. Thus, points near land do not have a tide correction made. This is not a problem for deep water regions.

9.3.3.4 Orbit error removal

One of the largest error sources associated with real time altimeter data is errors in the orbit solution. These errors are so large that the altimeter data usefulness would be questionable if they were not removed. Fortunately, the structure of the orbit error is very long wavelength. Thus if we remove the long wavelength component of the SSH (40,000 km, or once cycle per satellite orbit revolution), we remove the orbit error. However, there is significant annual variability that resembles this long wavelength. The out of phase annual heating of the northern and southern hemispheres causes a large steric anomaly in one hemisphere of the opposite sign to the steric anomaly in the other hemisphere. Removing this steric signal can produce errors in the resulting interpretation of the satellite SSH.

To minimize the damage to the oceanography when removing the orbit error, we take into account the expected seasonal variations in SSH based on climatological data. For the seasonal dynamic height variations we use the Navy Generalized Digital Environmental Model (GDEM). This model based on climatological bathythermograph data allows us to maintain the seasonal steric signal in the SSH data while removing orbit errors. The orbit error removal is done on one satellite revolution's worth of data at a time using a weighted least squares technique. The weighting is based on the variability observed by prior altimeter missions. This minimizes the impact of large amplitude features in the areas dominated by mesoscale variability.

9.3.3.5 Referencing to a consistent mean

The most difficult problem in using multiple altimeter data sets is referencing all the data sets to the same mean sea level. Each satellite measures variability about the mean sea level along its own ground tracks. From previous altimeter missions we can produce a mean sea level along the ground tracks, and use present data relative to these prior means. However, the time periods of previous missions do not overlap, and significant changes in sea level occur due to ocean circulation. We use the time period from Jan 1, 1993 through Jan 1, 1998 (the first full 5 years of TOPEX/POSEIDON) as the time period to which we reference all the altimeter data.

Two different methods are used to reference ERS-2 and GFO to this time period. Because there is an overlap between the ERS and TOPEX/POSEIDON satellites, the TOPEX anomalies about the 5 year mean may be used in conjunction with the ERS data. First the T/P SSH anomalies about the 5 year mean are interpolated in space in time. These T/P deviations from the 5-year mean are interpolated to the ERS measurement positions at the exact measurement times and then subtracted from the ERS data. This transforms the ERS data into measurements of the 5 year mean, and the average of these data minimizes interpolation errors. The resulting ERS mean allows use of the real time ERS data to provide anomalies relative to the 5-year mean time period.

GFO presents a much more difficult challenge. There is no time overlap between the original Geosat-Exact Repeat Mission and other altimeters. The only place where the SSH may be compared to present altimeters is at the points where the Geosat ground track crosses other ground tracks since the geoid is the same at these points. By examining the mean SSH from the Geosat-ERM and from T/P at the crossover points,

we can determine the SSH change between the two time periods. This provides a correction to the original Geosat-ERM mean to allow the GFO data to be referenced to the common 5 year mean.

9.3.4 Altimetry products and derived products

Through signal processing and interpretation, researchers have identified several different types of geophysical parameters which can be derived from the altimeter data provided by TOPEX/Poseidon. These same parameters can then be directly associated with various oceanographic products as detailed in the table below:

Measurement	Geophysical parameter	Oceanographic product
Elapsed Time of Return	Altitude	Marine Geoid, Current Location & Velocity and Ice Topography
Waveform Leading Edge	Sea/Surface Height, Distribution & Standard Deviation	Significant Wave Height
Waveform Leading Edge	Skewness	Dominant Wavelength, Significant Slope and Thermocline Information
Amplitude	Backscatter Coefficient: Sigma (s)	Sea/Ice Boundaries and Sea Surface Windspeed
Trailing Edge	Satellite Pointing and Sea/Surface Tilt	Ice Slope
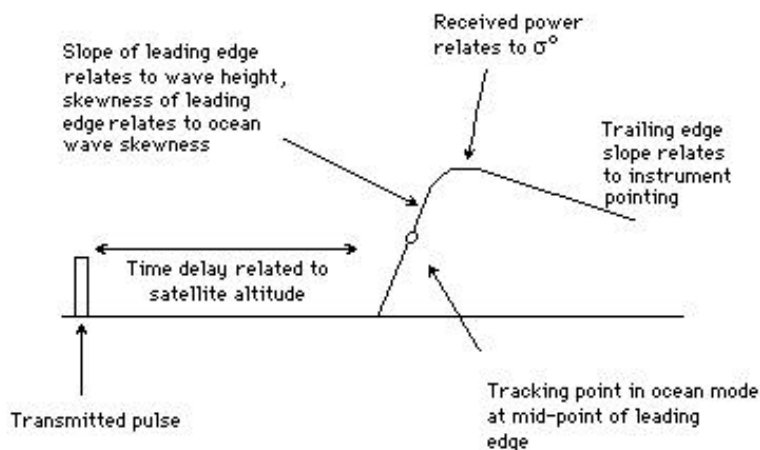
Figure of the average profile of an Ocean ReturnWaveform (taken from ERS-2 site). Over ocean surfaces, the distribution of the heights of reflecting facets is gaussian or near-gaussian, and the echo-waveform has a characteristic shape that can be described anayltically, as a function of the standard deviation of the distribution, which is closely related to the ocean wave height.

The key principle behind the altimeter is that the information required is in the shape and timing of the returned radar pulse.

The figure shows a pulse being reflected from a flat surface. As the pulse advances, the illuminated area grows rapidly from a point to a disk, as does the returned power. Eventually, an annulus is formed and the geometry is such that the annulus area remains constant as the diameter increases. The returned signal strength, which depends on the reflecting area, grows rapidly until the annulus is formed, remains constant until the growing annulus reaches the edge of the radar beam, where it starts to diminish.

Irreguarities on the surface, larger than the pulse width, cause the returned pulse to be distorted and stretched. The effect of this is to impose an additional slope on the leading edge of the returned signal strength curve (see the figure). This slope is related to the ocean wave height and the mid-point of this leading edge slope is equivalent to the reflection from the average position of the surface (i.e. the mean sea surface). By measuring the total area under the curve, the average reflectivity of the surface may be obtained.
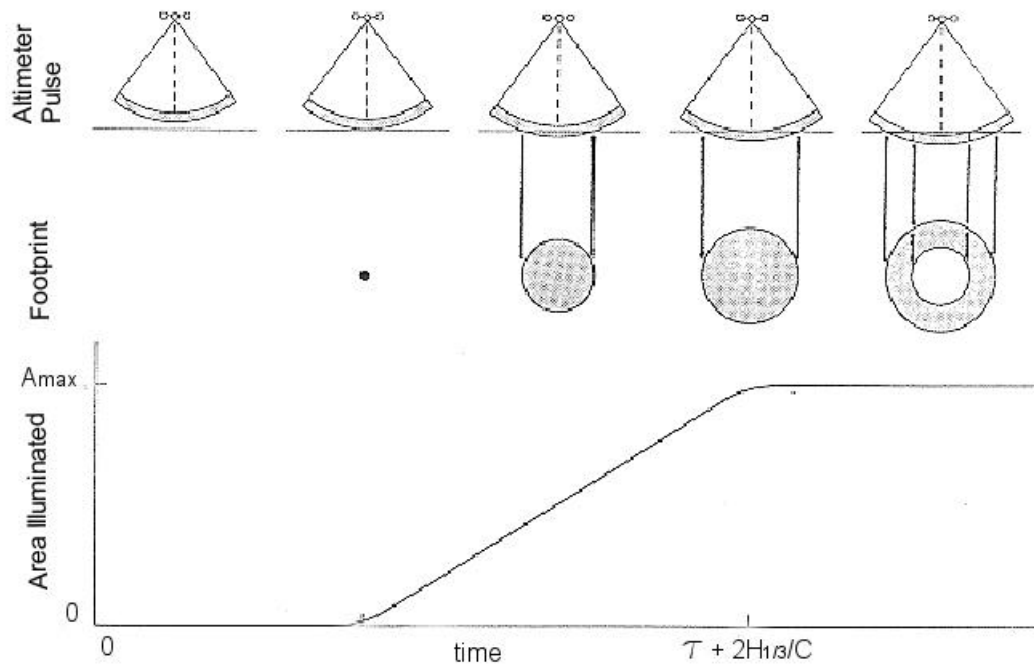
Schematic representation of a wide beamwidth, short pulse propagating from the satellite to the sea surface (upper row). The antenna footprint on the sea surface is shown as a function of time in the middle row. The area of the footprint is shown as a function of time in the bottom panel. For a calm sea surface, the area rise time is equal to the pulse duration t. For a rough sea surface with significant wave height H 1/3 , this rise time increases by amount 2c-1H 1/3 .

Clearly the altimeter requires a very short pulse width and full analysis of the pulse shape must be carried out quickly since the pulse shown in the figure is so short. Both of these time requirements have been avoided by translating from time to frequency, using a technique called full-deramp . The altimeter therefore operates with frequency-modulated rather than time-modulated signals and the analysis of pulse shape is a spectrum analysis , performed by the processor. The characteristics of this processor and the control of the radar are varied according to the mode of the Radar Altimeter. The most important is the acquisition mode, during which the radar finds the approximate distance to the surface and then switches to one of the tracking modes - ocean or ice (as is done in ERS-2).

9.3.4.1 Dynamic Sea Surface Topography - (from Altitude)
A measure of sea level relative to earth's geoid from which oceanographers can calculate the speed and direction of winds. Two very precise distance measurements must be established in order to acquire reliable topographical ocean surface maps. First, the height of the satellite above the reference ellipsoid is measured by tracking the satellite in orbit from a globally-distributed network of lasers and/or Doppler stations. The trajectory and height of the satellite are further refined by using orbit dynamic equations. Second, the height of the satellite above the closest ocean surface

is measured with a microwave radar altimeter. The difference between the height above the reference ellipsoid and the altitude above the ocean surface is approximately equal to the geoid height. Depicted in the image below* is the difference between the geoid height and mean sea level, commonly referred to as sea surface topography. This sea surface height has two main components, the geoid height, which reflects the gravitational field of the earth and the dynamic sea surface height, which reflects the ocean currents and tides.

Image is from Jet Propulsion Lab's Physical Oceanography Distributed Active Archive Center
Date of Image: January 20, 1996
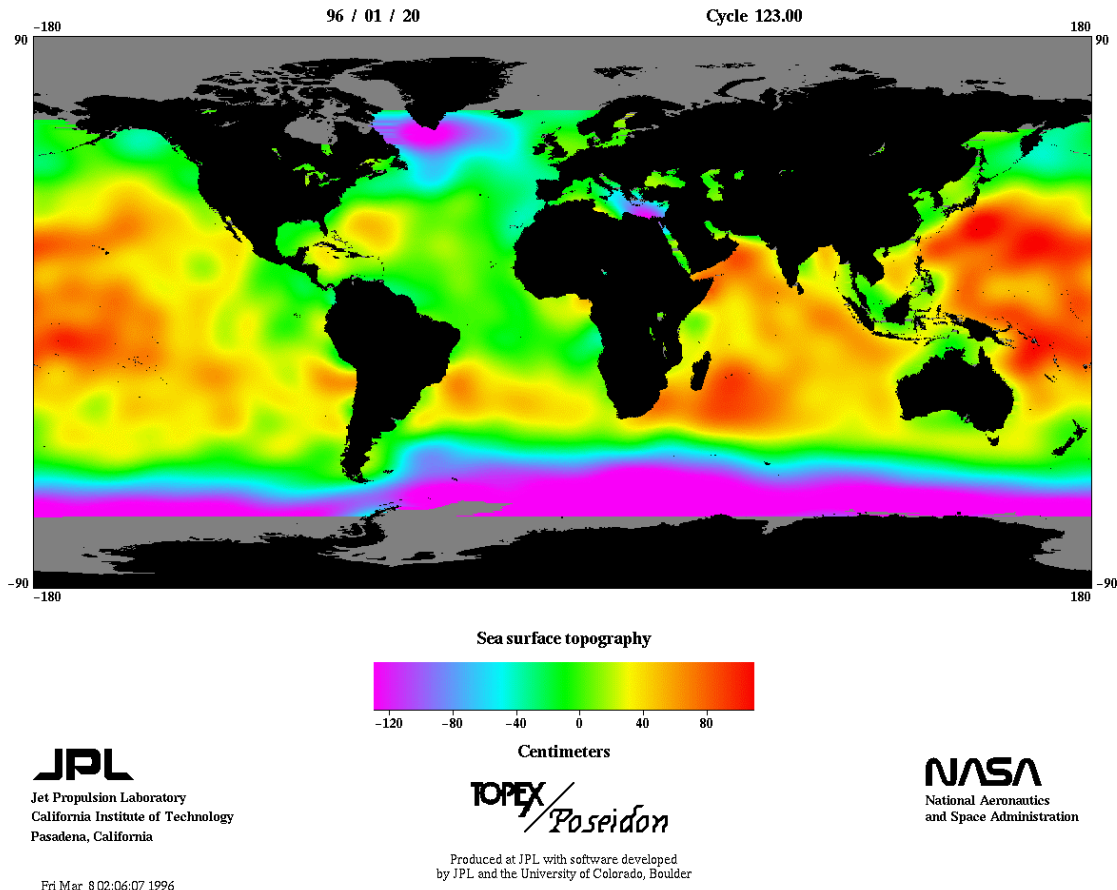
9.3.4.2 Sea Surface Variability –
Illustration of Dynamic Sea Surface Topography via time-dependent anomalies relative to a mean sea surface.

9.3.4.3 Wind Speed - (from backscatter coefficient)
- *http://www.oce.orst.edu/po/satellite.html*
- *http://winds.jpl.nasa.gov/scatterometry/apps.html*

By measuring global sea-surface wind speed and direction, ocean scatterometer data can help meteorologists more accurately predict the marine phenomena that affect human life on a daily basis. This is important for weather forecasting, storm detection, ship routing, safer end efficient off shore oil and gas drilling operations, and more.

The wind speed is measured by examining the radar signal generated by the satellite and reflected from the ocean surface. The wind on the ocean surface generates roughness that affects the scattering of the radar signal. The radar backscatter is used with laboratory measurements to infer the wind speed. There is no direction associated with this speed (however, wind direction can be derived from scatterometer

data – see in the details about ERS-2 above). A calm sea is a good reflector and returns a strong pulse, while rough seas scatter the signal thereby returning a weak pulse. In general there is a high degree of correlation between wind speed and wave height. The amount of scattered power can be used to observe the sea-ice extent, the boundaries of currents, oceanic wind speed, and the rate of accumulation of snow on glaciers.

Surface winds in the Pacific Ocean observed by the NSCAT scatterometer during September 1997.
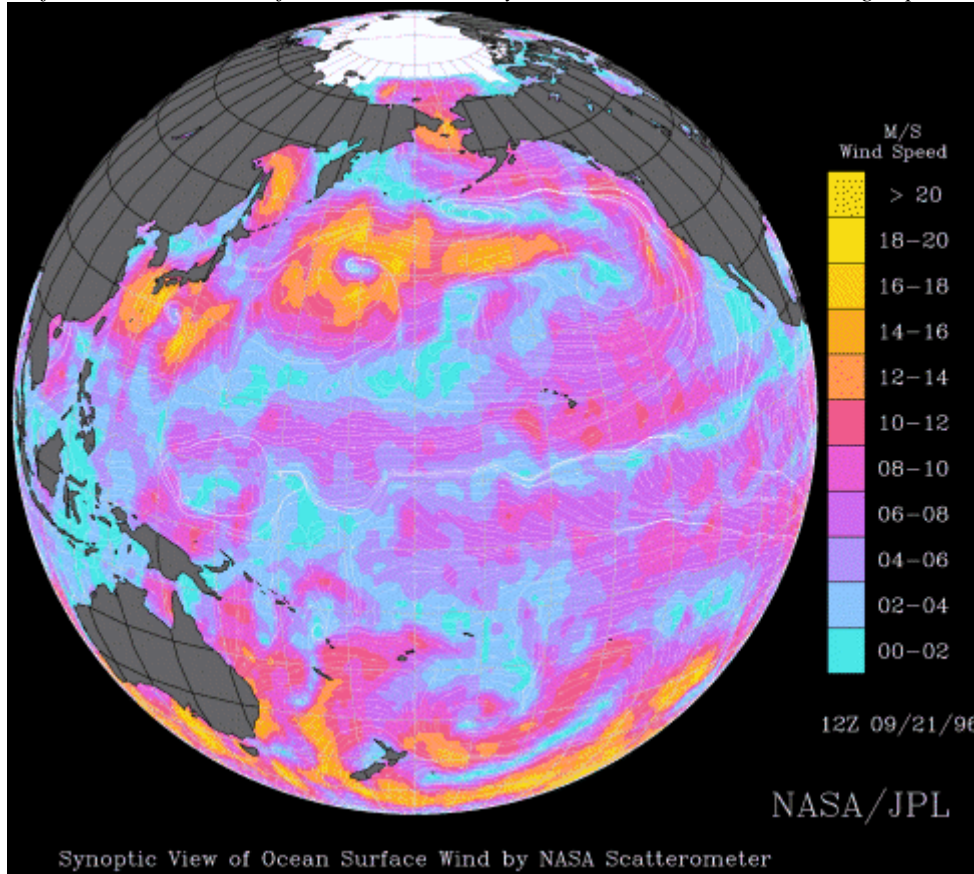
Zoom in on the central area of the figure above:
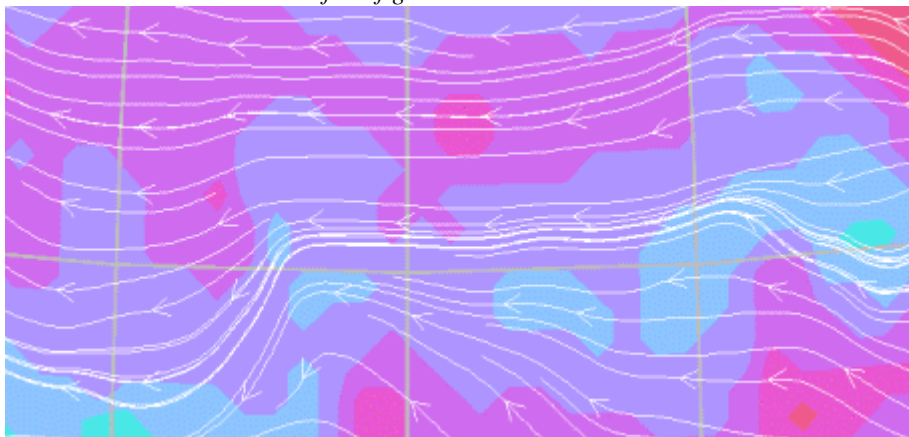
A scatterometer is a high frequency microwave radar designed specifically to measure ocean near-surface wind speed and direction. As the wind blows over the ocean, the surface is roughened by the generation of cat's paws (centimeter scale capillary waves). These, in turn, modify the surface backscatter (reflected signal or echo) properties.

Scatterometery has its origin in early radar used in World War II. Early radar measurements over oceans were corrupted by sea clutter (noise) and it was not known at that time that the clutter was the radar response to the winds over the oceans. Radar response was first related to wind in the late 1960's. The first scatterometer flew as part of the Skylab missions in 1973 and 1974, demonstrating that spaceborne scatterometers were indeed feasible. The Seasat-A Satellite Scatterometer (SASS) operated from June to October 1978 and proved that accurate wind velocity measurements could be made from space. A single-swath scatterometer flew on the European Space Agency's Remote Sensing Satellite-1 (ERS-1) mission.

The NASA Scatterometer (NSCAT) which launched aboard Japan's ADEOS-Midori Satellite in August, 1996, was the first dual-swath, Ku-band scatterometer to fly since Seasat. From September 1996 when the instrument was first turned on, until premature termination of the mission due to satellite power loss in June 1997, NSCAT performed flawlessly and returned a continuous stream of global sea surface wind vector measurements. Unprecedented for coverage, resolution, and accuracy in the determination of ocean wind speed and direction, NSCAT data has already been applied to a wide variety of scientific and operational problems. These applications include such diverse areas as weather forecasting and the estimation of tropical rain forest reduction. Because of the success of the short-lived NSCAT mission, future Ku-band scatterometer instruments are now greatly anticipated by the ocean winds user community. The NSCAT mission proved so successful, that plans for a follow-on mission have been accelerated to minimize the gap in the scatterometer wind database.

Winds over the ocean modulate air-sea changes in heat, moisture, gases and particulates (matter in the form of small liquid or solid particles), regulating the crucial bond between atmosphere and ocean that establishes and maintains global and regional weather and climate. Data derived from ocean scatterometers is vital to researchers in the their studies of air-sea interaction and ocean circulation, and their effects on weather patterns and global climate. In the past, weather data could be acquired over land, but our only knowledge of surface winds over oceans came from infrequent, and sometimes inaccurate, reports from ships and buoys. This data are also useful in the study of unusual weather phenomena such as El Niño, the long-term effects of deforestation on our rain forests, and changes in the sea-ice masses around the polar regions. These all play a central role in regulating global climate.

Wind affects the full range of oceanic motion – from individual surface waves to complete current systems. The tropical Pacific Ocean and overlying atmosphere react to, and influence each other. Easterly surface winds along the equator control the amount and temperature of the water that upwells (moves or flows upward) to the surface. This upwelling of cold water determines sea-surface temperature distribution,

which affects rainfall distribution. This in turn determines the strength of the easterly winds – a continuous cycle.

9.3.4.4 Ocean circulation

- *http://www.oce.orst.edu/po/research/strub/index.html*
- *Marth, Paul C., et al., **Prelaunch Performance of the NASA Altimeter for the TOPEX/Poseidon Project,** IEEE Transactions on Geoscience and Remote Sensing, Volume 31, Number 2, Pages 315-332, March 1993.*

The way in which global mean sea level is related to ocean circulation is best described by Marth, et al (1993): "Movement of water in the sea on spatial scales exceeding ~30 km and temporal scales exceeding ~1 day are manifested by deflections of the sea surface; that is, by changes in mean sea level associated with the strength and direction of the flow. This change in topography varies from the 1 meter increase in mean sea level in 100 km over the Gulf Stream to a 10 cm change over 1000 km over an El Nino event in the tropical Pacific. Given the strength of the sea surface topography signature, one can infer the magnitude and direction of the oceanic water movement."

Ocean currents are mapped by studying the "hills" and "valleys" in maps of the height of the sea surface relative to the geoid. This height is called "Dynamic Ocean Topography." Currents move around ocean dynamic topography "hills" and "valleys" in a predictable way. Note that a clockwise sense of rotation is found around "hills" in the Northern Hemisphere and "valleys" in the Southern Hemisphere. This is because of the "Coriolis effect." Conversely, a counterclockwise sense of rotation is found around "valleys" in the Northern Hemisphere and "hills" in the Southern Hemisphere. In general, our major ocean currents are stable and so maps of dynamic ocean topography change very little over time. Before satellites, it took many years of ship data to create a map of global oceans. Now, an updated map of ocean circulation is available every 10 days.

The following figure shows altimeter field heights of the California current:
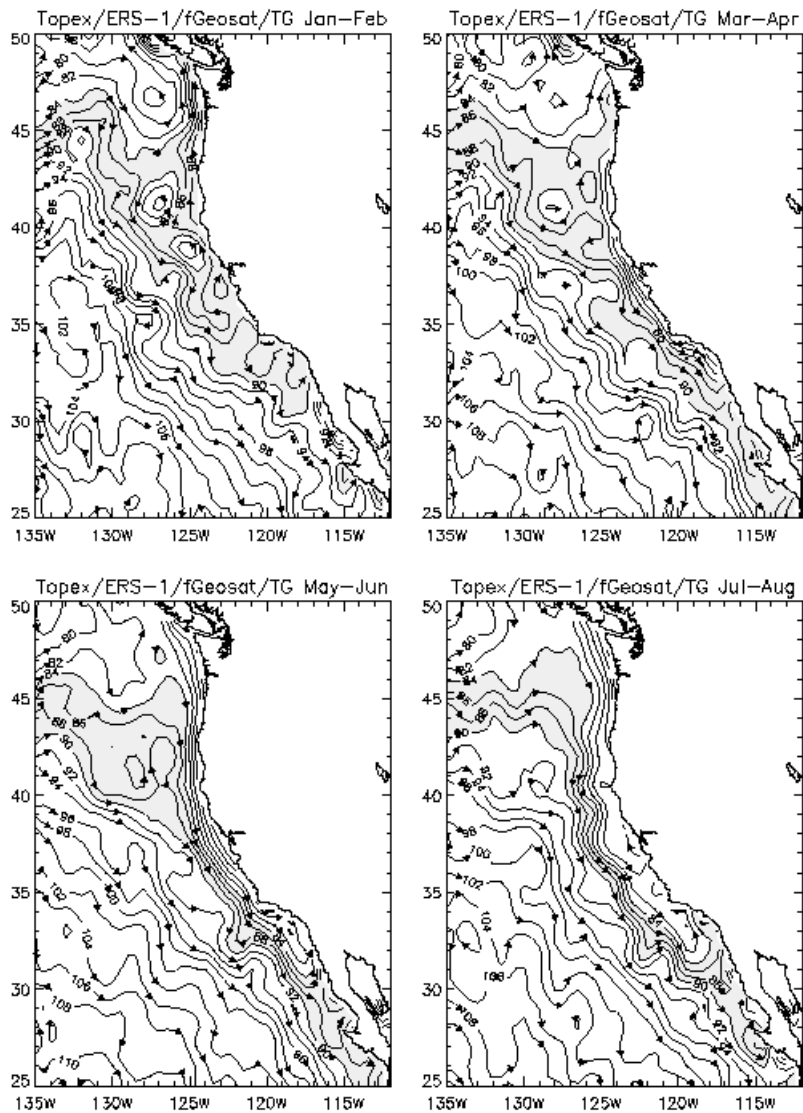
The altimeter height fields show the poleward Davidson Current in winter and the equatorward California Current (CC) jet that begins near the coast in spring and moves steadily offshore until fall. Shading shows that water from the west wind drift only affects the region offshore of the CC jet core during most of the year, except in the Southern California Bight.

9.3.4.5 Significant Wave Height - (from Wave Form Leading Edge)
Determined from the shape of the return pulse of the radar altimeter; a calm sea with low waves returns a condensed pulse, while rough seas with high waves return a stretched pulse.

9.3.4.6 Watervapor
Watervapor in the atmosphere delays the return of the radar altimeter pulse and thereby produces a false reading of sea level. Onboard microwave radiometers measure the content of water vapor and allow for adjustments.

9.3.4.7 Marine gravity and sea-floor topography

- *Chelton, Dudley B., WOCE/NASA Altimeter Algorithm Workshop, U.S. WOCE Technical Report No. 2, 70 pp., U.S. Planning Office for WOCE, College Station, TX 1988.*
- *Global Bathymetric Prediction for Ocean Modelling and Marine Geophysics, David T. Sandwell and Walter H. F. Smith*
- *David T. Sandwell, Walter H. F. Smith, Chapter 11 - BATHYMETRIC ESTIMATION, (To appear in Satellite Altimetry and Earth Sciences, Academic Press, 1999)*
- *http://topex.ucsd.edu/marine_topo/text/topo.html*

Variations in gravitational acceleration over the earth's surface result in an uneven distribution of water mass in the oceans. There is a latitudinal variation associated with the oblateness of the earth. In addition, there are gravity anomalies associated with topographic features on the earth's surface. The gravitational acceleration at the sea surface is slightly stronger over bumps on the ocean bottom and slightly weaker over depressions in the bathymetry. In the absence of other forcing (e.g., pressure gradients, wind forcing, or tides), the sea surface would be a surface of constant gravitational potential (the marine geoid).
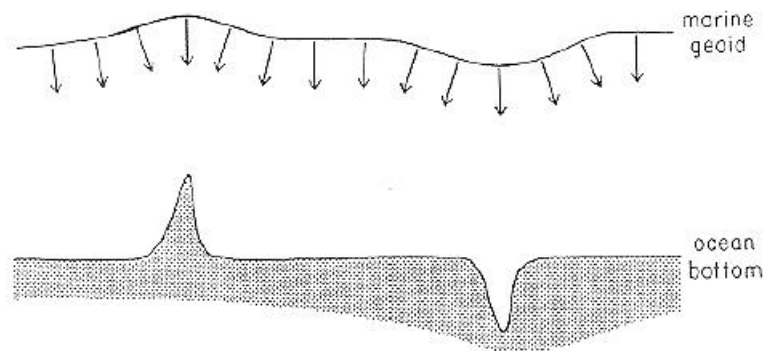
Schematic diagram of a bump and a depression on the ocean bottom and the corresponding marine geoid. The vectors indicate the gravitational acceleration along the geoid. The gravitational acceleration is locally deflected toward the bump and away from the depression and is tangentially perpendicular to the geoid.

To first order then, the marine geoid is a low-pass filtered image of the bathymetry. Global estimates of the marine geoid are obtained from combined satellite tracking data and ship-based gravity measurements. The long wavelength components are obtained from measurements of the perturbed motion of near-earth satellites and short wavelength components are determined from ship-based gravity measurements. These gravity measurements are unevenly distributed globally, so there is large uncertainty of approximately 1 m in the global geoid. Except in limited geographical regions of densely-sampled shipboard gravity surveys (e.g., the northwest Atlantic), uncertainty in the geoid height h_g *is* the largest source of error in altimeter estimates of absolute dynamic sea surface topography h_d . Globally, the height h_g of the marine geoid has a range of about ±100 m about a reference ellipsoid approximation to the surface of the earth. By comparison, the sea surface topography $h\ d$ from dynamic ocean currents has a range of at most a few meters globally. Thus, to a very close approximation, the mean sea surface measured by an altimeter coincides with the marine geoid. Because the marine geoid is a low-pass filtered version of the bathymetry (as noted above), altimetric mean sea surfaces closely resemble bathymetry maps.

David T. Sandwell and Walter H. F. Smith report about the following work:

We are constructing a complete bathymetric map of the oceans at a 3-10 km resolution by combining all of the available depth soundings collected over the past 30 years with high resolution marine gravity information provided by the Geosat, ERS-1/2, and Topex/Poseidon altimeters. Detailed bathymetry is essential for understanding physical oceanography and marine geophysics. Currents and tides are controlled by the overall shapes of the ocean basins as well as the smaller sharp ocean ridges and seamounts. Because erosion rates are low in the deep oceans, detailed bathymetry reveals the mantle convection patterns, the plate boundaries, the cooling/subsidence of the oceanic lithosphere, the oceanic plateaus, and the distribution of off-ridge volcanoes. Current global digital bathymetry maps (e.g. ETOPO-5) lack many important details such as a 400 km-long ridge that rises to within 135 m of sea level in the sub-antarctic front. Moreover, they are contaminated by long-wavelength errors (~2000 km) which prevent accurate identification of seafloor swells associated with mantle plumes.

The tasks include:

- Accumulate all available depth soundings collected over the past 30 years. (funded by NSF)
- Use the short wavelength (< 160 km) satellite gravity information to interpolate between sparse ship soundings.
- Improve the resolution of the marine gravity field using enhanced estimates along repeat altimeter profiles together with the dense altimeter measurements.
- Refine/Improve bathymetric predictions using the improved resolution gravity field and also by investigating computer-intensive methods for bathymetric prediction such as inverse theory.
- Produce a *Globe of the Earth* similar to the globe of Venus prepared by the NASA Magellan investigation. This will also include the best available digital land data.

In the wavelength band 15 to 200 km, variations in gravity anomaly are highly correlated with seafloor topography. Since many southern ocean areas and some northern ocean areas are sparsely surveyed, these new satellite altimeter data reveal many previously unsurveyed features such as ridge axes, seamounts and fracture zones.
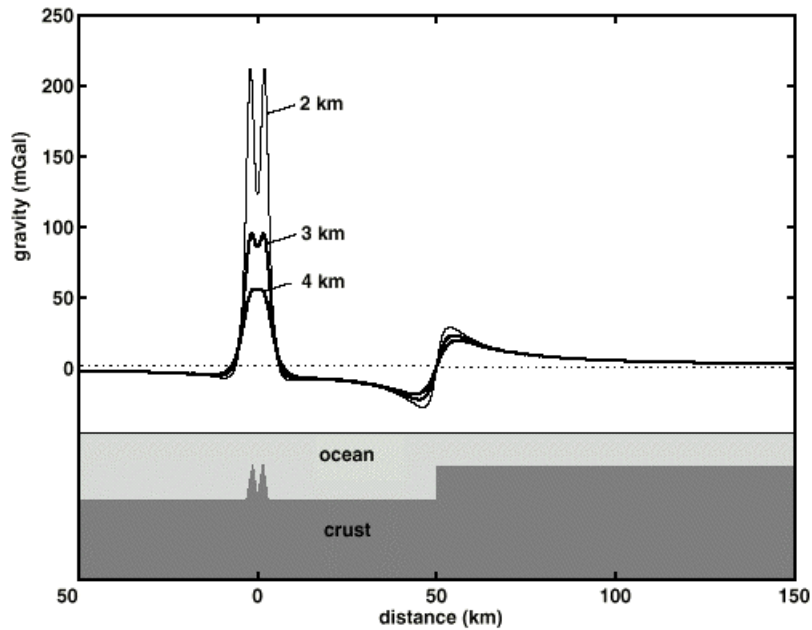
Fundamental limitations of topographic recovery from gravity anomaly measurements are illustrated by seamounts (left) and plateau (right). The gravity signatures of the closely-spaced seamounts (4 km apart and 1 km tall) are strong and distinct when the average ocean depth is 2 km or less but their signatures combine and become weak when the ocean depth is 4 km. The isostatically compensated step in depth produces a local gravitational edge effect that is strongly attenuated at a distance of 150 km from the step; thus the gravity far from the step does not provide information on the overall depth offset across the step.

The conceptual approach is to use the sparse depth soundings to constrain the long-wavelength depth while the shorter-wavelength topography is predicted from the downward-continued satellite gravity measurement. Over the short wavelength band, the topography/gravity ratio is regionally calibrated using available soundings.
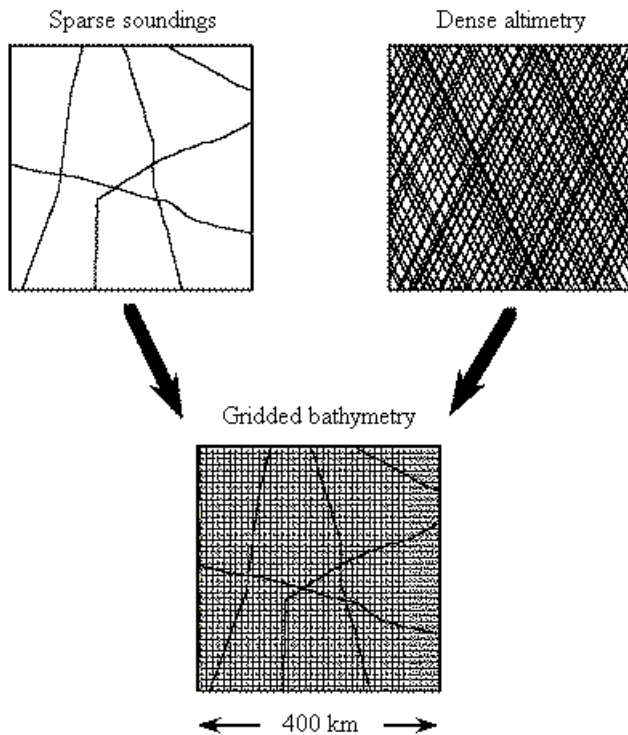
Data quality is the most important aspect of bathymetric prediction. High resolution satellite gravity data is needed not only to interpolate among the sparse soundings but also to identify which soundings are bad. Many soundings from remote areas of the oceans were collected before shipboard computers and satellite navigation were available. Moreover, all of the ship sounding data were collected over a 30 year period on a variety of ships from many countries and with many different chief scientists aboard. Thus the quality of the data is highly variable and many entire cruises or sections of cruises are bad; only the most recent (since ~1987) GPS-navigated multibeam data are reliable. Typical errors include: navigation errors, digitizing errors, typographical errors due to hand entry of older sounding, reporting the data in fathoms instead of meters, incorrect sound velocity measurements and even computer errors in reading punch cards One bad section of a cruise in an isolated region will introduce a seafloor topographic feature that does not exist.

The high resolution gravity fields provides the information needed to assess the accuracy of the ship sounding data. Our approach is to identify the bad cruises through a comparison with an initial prediction based on the gravity and either eliminate them or attempt to fix problem areas (data rescue); rescue is especially important for soundings that fill a large data gap.

Data preparation and assembly is an ongoing process; the current data are sufficiently good to construct a global bathymetric grid. Here is one recipe (Nettleton's Method) that we are developing.

NETTLETON'S METHOD (to understand the concepts mentioned below, review the chapter about filters):

1) Grid available bathymetric soundings on a 2 minute Mercator grid that matches our gravity anomaly grid. To avoid seams, all work is done on a global grid between latitudes of ±72deg. Coastline points from GMT provide the zero-depth estimates. A finite-difference, minimum-curvature routine is used to interpolate the global grid. This gridding program requires at least 256 Mbytes of computer memory.

2) Separate the grid into *low-pass* and *high-pass* components using a Gaussian filter (0.5 gain at 160 km). Filtering and downward continuation are performed with a multiple strip, 2-D FFT that spans 0-360deg longitude to avoid Greenwich edge effects.

3) Form *high-pass filtered gravity* using the same Gaussian filter.

4) Downward continue the *high-pass filtered gravity* to the *low-pass filtered bathymetry* assuming Laplace's equation is appropriate. A depth-dependent Wiener filter is used to stabilize the downward continuation.

5) Accumulate *high-pass filtered soundings* and corresponding *high-pass filtered/downward-continued gravity* into small (160 km) overlapping areas and perform a robust regression analysis. In sediment-free areas, the topography/gravity transfer function should be flat and equal to *1/2[pi]G[Delta][rho]* so in the space domain, a linear regression is appropriate. This works well on young seafloor but not on old seafloor where sediment cover destroys the correlation between topography

and gravity. In these cases we assume the seafloor is flat and set the *topography/gravity ratio* to zero. Finally there are intermediate cases where topographic depressions will be sediment filled while the highs protrude above the sediments so the topography/gravity relationship is non-linear. It is these partially sedimented areas that make the bathymetric problem difficult and inherently non-linear. Continental margins and shelves pose similar problems.

6) Regional *topography/gravity ratio* estimates are gridded and multiplied by the *high-pass filtered/downward-continued gravity* to form *high-pass filtered predicted bathymetry*.

7) The *total predicted bathymetry* is equal to the sum of the *high-pass filtered predicted bathymetry* and the *low-pass filtered bathymetry*.

8) Finally, the pixels constrained by ship soundings or coastline data are reset to the measured values and the finite-difference, minimum curvature routine is used to perturb the predicted values toward the measured values. Measured depths are flagged so they can be extracted separately. This final step dramatically increases the accuracy and resolution of the bathymetric grid in well surveyed areas so it agrees with the best hand-contoured bathymetric charts.
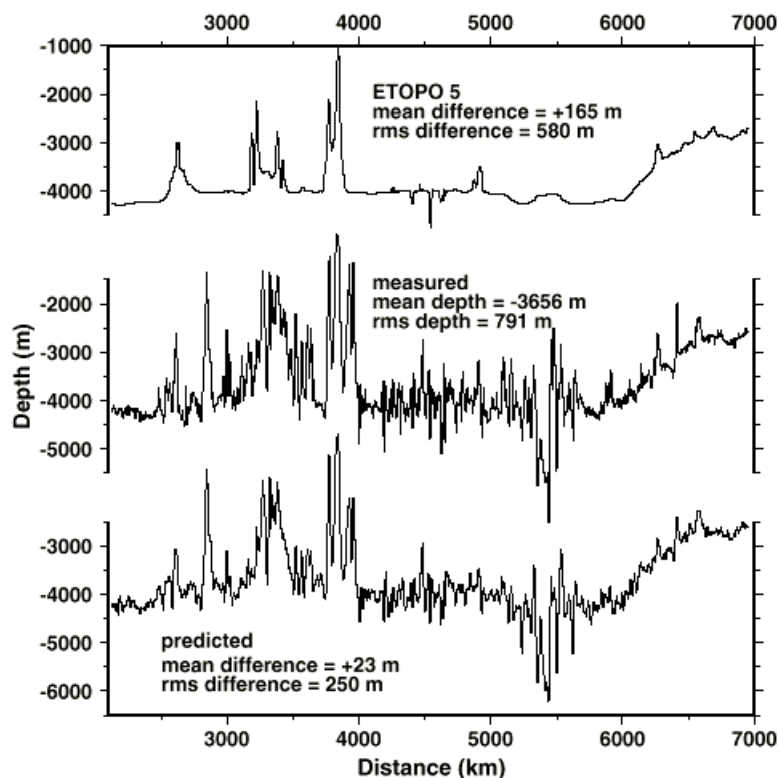
Comparison of measured depth along ship track (middle) with ETOPO-5 (upper) and our predicted depth (lower). These ship data were acquired after the V6.2 of the bathymetric grid was completed.

We have assessed the accuracy of the prediction through a comparison with soundings from a recent cruise to the Foundation Seamount Chain in the South Pacific. This poorly charted area contains a 1600-km long volcanic chain as well as topography associated with microplate tectonics. Based on depth predictions from our earlier study [1994], an initial mapping and sampling expedition was carried out in 1995 aboard the *R/V Sonne* where they charted 44 volcanoes with height ranging from 1500

m to 4000 m; eleven of the uncharted volcanoes come to within 500 m of the ocean surface. These *Sonne-100* sounding data were included in our global seafloor topography map and provide good definition of the summits of the seamounts. In January and February of 1997, one of us (Sandwell) participated in a second expedition to the Foundation Seamounts area aboard *R/V L'Atalante*. The cruise track covers very high relief topography areas that were not surveyed during the *Sonne-100* cruise and thus offers an excellent test of the accuracy of the predicted seafloor depth. The results are shown in above figure where the center beam of the Simrad 12D multibeam echo sounder is plotted versus distance from Tahiti (center profile). A number of large seamounts and ridges were surveyed between distances of 2500 and 4000 km while a 6500-m deep trench was surveyed at a distance of 5800 km. These measurements are compared with depths sampled from the ETOPO-5 grid (upper profile) and our version 6.2 predicted depths (lower profile). The depths from ETOPO-5 show a poor correlation with the measured depths and have an rms misfit of 590 m in an area where the rms signal is large 747 m. The poor fit is due to a lack of ship soundings in the area. The prediction offers a much better match to the observed depths (259 m rms) because the high-resolution gravity field information from the Geosat and ERS-1 satellite altimeters provides most of the depth information. In other well charted areas of the northern oceans, the grid is more tightly constrained by ship soundings. On average, 12% of the grid cells are constrained by ship measurements while the remaining 88% are estimated.

The global map of sea floor topography can be accessed in the following site:
http://raven.grdl.noaa.gov/cgi-bin/meg/bathD.pl

9.4 Airborne Laser Scanning (ALS):

- *Crandall Clifford J. (1976), Coastal Aerial Photo-Laser Survey (CAPS) – A Near Shore Charting System, International Hydroraphic Review, LIII (1), pp. 53-64*
- *Nairn R. (1994), Royal Australian Navy Laser Airborne Depth Sounder – the first year of operations, International Hydrographic Review, LXXI (1), pp. 109-119*
- *Koppari Kurt, Karlsson Ulf and Steinvall Ove (1994), Airborne Laser Depth Sounding in Sweden, International Hydrographic Review, LXXI (2), pp. 69-90*
- *Aloysius Wehr, Uwe Lohr, Airborne laser scanning—an introduction and overview, ISPRS Journal of Photogrammetry & Remote Sensing 54 1999 68–82*
- *Emmanuel P. Baltsavias, A comparison between photogrammetry and laser scanning, ISPRS Journal of Photogrammetry & Remote Sensing 54 1999 83–94*
- *Jennifer L. Irish, W. Jeff Lillycrop, Scanning laser mapping of the coastal zone: the SHOALS system, ISPRS Journal of Photogrammetry & Remote Sensing 54 1999 123–129*
- *Peter Axelsson, Processing of laser scanner data – algorithms and applications, ISPRS Journal of Photogrammetry & Remote Sensing 54 1999 138-147*
- *http://www.hydro.navy.gov.au/surveying/index4.htm*
- *IHO standards for hydrographic surveys, 4th edition, April 1998, Special Publication no. 44*
- *http://www.ehis.navy.mil/dstp/11jul94.htm*
  European Naval Oceanography, by David B. St.Pierre
- *Axelsson Rune and Alfredsson Mats, Capacity and Capability for Hydrographic Missions*
- *AN ASSESSMENT OF NASA'S AIRBORNE TOPOGRAPHIC MAPPER INSTRUMENT FOR BEACH TOPOGRAPHIC MAPPING AT DUCK, NORTH CAROLINA, United States Department of Commerce National Oceanic and Atmospheric Administration Coastal Services Center, Coastal Services Center Technical Report CSC/9-98/001 Version 1.0 September 1998*
- *http://www.csc.noaa.gov/crs/ALACE/techcd/htm/laser.htm*
- *http://www.fema.gov/mit/tsd/lidar_4b.htm*

9.4.1 Introduction – laser principles

Laser scanners utilise opto-mechanical scanning assemblies just as many multispectral scanners like the spaceborne scanner on Skylab, the LANDSAT MSS and the Thematic Mapper TM. However, they are active sensing systems using a laser beam as the sensing carrier. Therefore, two optical beams - All laser systems measure by some means the distance between the sensor and the illuminated spot on ground.

The narrow divergence of the laser beam defines the instantaneous field of view IFOV. Typically, the IFOV ranges from 0.3 mrad to 2 mrad. The theoretical physical limit of the IFOV is determined by diffraction of light, which causes image blurring.

Therefore, the IFOV is a function of the transmitting aperture D and the wavelength of the laser light l. For spatially coherent light, the diffraction-limited IFOV is given by:

$$IFOV_{diff} = 2.44 * l / D$$

The IFOV of the receiving optics must not be smaller than that of the transmitted laser beam. Due to the very narrow IFOV of the laser, the optical beam has to be moved across the flight direction in order to obtain area coverage required for surveying. The second dimension is realised by the forward motion of the airplane. Thus, laser scanning means deflecting a ranging beam in a certain pattern so that an object surface is sampled with a high point density.

9.4.1.1 Pulse lasers

Laser light is very special. The acronym laser stands for *L*ight *A*mplification by *S*timulated *E*mission of *R*adiation. In current ranging laser systems, mostly pulsed lasers are used.

The most direct ranging measurement is determining the time-of-flight of a light pulse, i.e., by measuring the travelling time between the emitted and received pulse. According to the following Figure, the travelling time (t_L)of a light pulse is:

$t_L = 2R/c$

with *R* the distance between the ranging unit and the object surface and *c* the speed of light.

Range:	$R = \dfrac{1}{2} c \cdot t_L$
Range Resolution:	$\Delta R = \dfrac{1}{2} c \cdot \Delta t_L$
Max. Range:	$R_{max} = \dfrac{1}{2} c \cdot t_{L_{max}}$
Range Accuracy:	$\sigma_R \sim \dfrac{c}{2} t_{rise} \cdot \dfrac{1}{\sqrt{S/N}}$
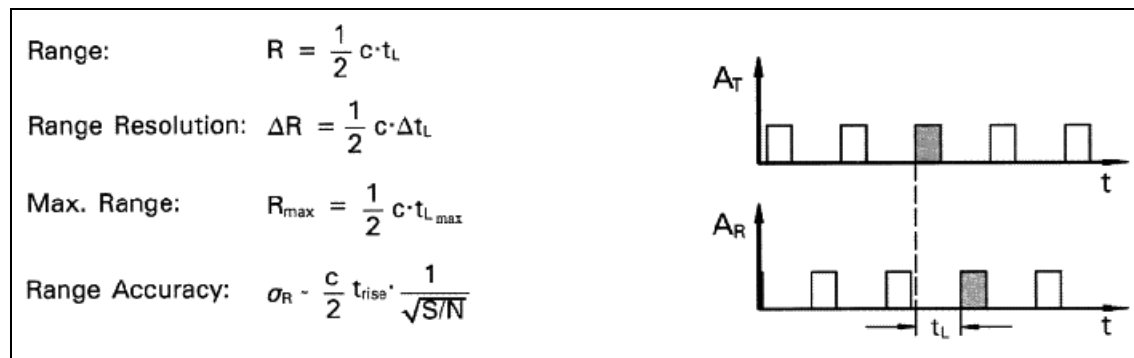
Figure: measuring principles of pulse lasers. The upper figure shows the transmitted signal, the lower the received one. A_T and A_R are the amplitudes of the transmitted and received signal, respectively.

9.4.1.2 Wavelength

In current ranging ALS, semiconductor diode lasers and Nd:YAG lasers pumped by semiconductor lasers are used, covering the optical band between 800 nm and 1600 nm. As current laser scanners work by the direct energy detection principle, a coherent light source for ranging is not required. Therefore, the following physical laser properties are used in laser scanning: high power, short pulses, high collimation, and narrow optical spectrum in which lasers emit. The narrow optical spectrum or in other words, the narrow spectral laser line is advantageous, because narrow optical interference filters (usually with 10 nm bandwidth) can be mounted in the receiving path to suppress disturbing background radiation, caused e.g., by backscattered sunlight.

The selection of the optical wavelength of the laser is dependent on the overall laser scanning system design. The most sensitive detectors are available between 800 nm and 1000 nm. Therefore, the first laser scanners worked with a wavelength of 900 nm. However, at this wavelength of the optical spectrum, eye safety is still a concern. If higher laser pulse energy is required, wavelengths have to be considered in which the eye is less sensitive.

When discussing laser wavelength, one should also consider the backscattering properties of the target, e.g., the object surface. The reflectivity of a target, for a given wavelength, also influences the maximum range. The following table shows the typical reflectivity of various materials for 900 nm wavelength:

Material	Reflectivity %
Snow	80–90
White masonry	85
Limestone, clay	Up to 75
Deciduous trees	Typ. 60

Coniferous trees		Typ. 30
Carbonate sand dry		57
Carbonate sand wet		41
Beach sands, bare areas in dessert		Typ. 50
Concrete, smooth		24
Asphalt with pebbles	17	
Lava		8
Black neoprene synthetic rubber	5	

9.4.1.3 Scanning

For a given flying height above ground h, the laser footprint mainly depends on the divergence of the laser beam g, and the swath width on the scan angle Q, which is also called the FOV.

Scans can be uni- or bidirectional. Typical scanning mechanisms that are used for airborne surveying are shown in the following figure. Oscillating mirrors usually produce a zigzag-line bidirectional scan or with two-axis galvanometers a bidirectional meander-type scan of parallel lines or arcs, rotating polygon and multifaceted mirror scanners produce parallel lines (unidirectional scan), nutating mirrors (Palmer scan) produce an elliptical pattern and the fiber scanner produces a parallel line scan. The scan pattern on the ground depends not only on the laser scan pattern but also the flying direction and speed and the terrain topography. The points along a line are usually scanned in equal angle steps, i.e., their spacing on the ground is not constant. Due to acceleration or slow down of the scan mechanism, the points at the swath borders exhibit other characteristics and are sometimes removed from the raw data set.
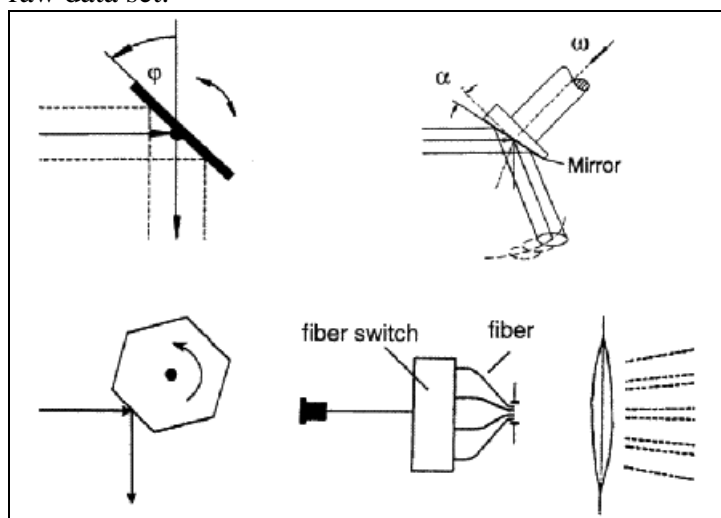
Figure: Scanning mechanisms from top left, clockwise : oscillating mirror, Palmer scan, fiber scanner, rotating polygon.

9.4.1.4 Position and orientation system

The laser scanner measures only the line-of-site vector from the laser scanner aperture to a point on the earth surface. The 3-D position of this point can only be computed, if at any time, the position and orientation of the laser system is known with respect to a coordinate system, e.g., WGS84. Thus, to obtain accurate range measurements in a given coordinate system, a laser scanner system must be supported by a POS. As laser scanners have a potential range accuracy of better than 1 dm, POS should allow at least the same accuracy. Such an accuracy can be achieved only by an integrated POS

197

consisting of a DGPS and an IMU. Geocoding of laser scanner measurements requires an exact synchronisation of all systems: IMU, DGPS and laser scanner data.

9.4.1.5 Typical processing steps

After a surveying flight, basically two data sets are available: the POS data and the laser ranges with the instantaneous scanning angles. Assuming that the accuracy of POS data is better than 1 dm in position and 0.02° in orientation, already very precise laser measurement points in an earth-fixed coordinate system can be calculated. However, some systematic parameters must be considered. These are e.g., the three mounting angles of the laser scanner frame, described by the Euler angles roll, pitch and yaw, with respect to the platform-fixed coordinate system (usually with origin at the IMU), the position of the laser scanner with respect to the IMU, and the position of the IMU with respect to the GPS. This so-called calibration data can be derived from laser scanner surveys, whereby certain reference areas are flown-over in different directions. Reference areas are e.g., flat terrain like large sport fields or stadiums, buildings and corner of buildings. From the relative orientation and position of the different surveys and their absolute orientation and position with respect to an earth-fixed coordinate system, calibration data can be derived.

The main objective of the further processing is the calculation of digital elevation models. First, the laser range data must be transformed from WGS84 to the desired local map coordinate system. The result is a cloud of randomly distributed laser points in elevation and position. The distribution of the measurement points depends on the scanning pattern of the laser scanner system.

Now, the elevation measurements are sorted with respect to their position. After the sorting process, ground points must be separated from non-ground points like buildings and vegetation. By classifying the different points, one can create a Digital Terrain Model - DTM, from the Digital Surface Model – DSM (all visible object top surfaces), obtained by the laser. For this task, different filter algorithms and procedures are applied.

Since the generated data amount is very high for certain applications and commercial programmes for DTM interpolation, visualisation, etc., can handle only a limited number of points, a thin-out is often applied. This is (or at least should be performed after filtering and DTM interpolation.

9.4.1.6 Some extended laser capabilities

Apart from measuring ranges, some lasers can also record the intensity of the backscattered laser light (with intensity measured as the maximum of the returned pulse, or signal integration over the returned pulse width). The intensity can be used for visualisation of the scene but also to improve filtering r removal and classification / separation of objects in combination with the range and other information, if available.

Some lasers can also record multiple echoes for each pulse sent, i.e., the range (and possibly intensity) of various objects along the pulse path and within the laser footprint. Such systems usually record both the first and last echo of a pulse, although there is one commercial system that can record up to four echoes per pulse. There are also experimental systems that record the whole form of the returned signal (called waveform) with waveform digitisers sampling the incoming signal with very high frequency, e.g., 250 MHz. Lasers that record only one echo, do so for the first or last one (some systems allow switching to first or last, depending on the application).

Recording of multiple echoes can be useful, when the vertical profile of multiple objects within the laser footprint (or at least the highest and lowest object), as for example, in the case of trees, is needed. Finally, there are bathymetric lasers, that are based on the same principles as the topographic lasers, but emit in two wavelengths, usually 1064 nm and 532 nm. The infrared wave-length is reflected on the water surface, while the green one penetrates the water and is reflected by the bottom surface or other objects in the water.

9.4.1.7 A short overview of applications
Some of the major applications of ALS include:
- Mapping of corridors, e.g., roads, railway tracks, pipelines, waterway landscapes
- Mapping of electrical transmission lines and towers including ground / tree clearance
- DTM generation, especially in forested areas in forests also for road and path planning, study of drainage patterns, etc.
- Measurement of coastal areas, including dunes and tidal flats, determination of coastal change and erosion
- High accuracy and very dense measurement applications, e.g., flood mapping, DTM generation and volume calculation in open pit mines, road design and modelling
- DTM and DSM generation in urban areas, automated building extraction, generation of 3-D city models for planning of relay antenna locations in wireless telecommunication, urban planning, microclimate models, propagation of noise and pollutants
- Rapid mapping and damage assessment after natural disasters, e.g., after hurricanes, earthquakes, landslides, etc.
- Measurement of snow- and ice-covered areas, including glacier monitoring
- Measurement of wetlands
- Derivation of vegetation parameters, e.g., tree height, crown diameter, tree density, biomass estimation, determination of forest borders
- Hydrographic surveys in depths up to 70 m.

9.4.1.8 Airborne Laser Scanning vs Photogrammetry – a comparison
Common aspects between photogrammetry and ALS include:

1. use of GPS, and with digital photogrammetric sensors, especially linear ones, GPS / INS;
2. methods for processing of raw data, like filtering of large errors, removal of non-DTM objects like buildings, data reduction (thin-out) and compression, and detection of breaklines, are shared between ALS and image matching for DSM / DTM generation;
3. furthermore, when laser data are regularly interpolated, they can be treated as images and various image analysis / processing techniques can be applied to them.

Thus, sensor integration and image (or digital signal) processing and analysis are two important topics that unify the two technologies.

The main common application, and competition field, between photogrammetry and ALS is the 3D measurement of surfaces and single objects. The reason is that classification and

identification of objects with ALS, without use of additional optical sensors, is very difficult to impossible.

However, ALS has some strengths which can be favourably exploited in certain applications, the most important of which are listed below:

- Mapping of surfaces with very little / no texture or poor definition. There, image matching delivers very poor results, and manual measurements are also poor or slow / cumbersome. Examples include ice / snow surfaces, sand (coasts, dunes, desserts), swamps and wetlands.
- Mapping of forests and vegetated areas. ALS systems can provide measurements on the ground. The penetration rate mainly depends on type of trees (deciduous or coniferous) and season. Useful results, depending also on the terrain roughness, can be achieved even with penetration rates of 20–30%. In addition, through appropriate data process-ing, both ground and tree height can be determined.
- Mapping of long, narrow features. This includes road mapping, planning and design, powerline corridor planning and tower design, coastal erosion monitoring, coastal zone management, traffic and transport, riverways and water resources and traffic management, mapping of railway lines, fiber-optic corridors, pipelines, dikes, etc. Since ALS systems have a narrower swath in comparison to optical sensors, they are more cost-effective in capturing the information needed for such applications.
- DSM generation of urban regions for urban planning, roof-top heights for communication antennas, etc.
- Mapping of very small objects, e.g., power lines (probably THE killer application of ALS), which are hardly visible in optical images, or whose measurement cannot be automated.
- Fast response applications. Since ALS provides digital range measurements, this information can be quickly converted to 3D coordinates. This can be important in some cases, e.g., involving natural disasters.

9.4.2 Laser remote sensing of forest structure

- *http://www.fsl.orst.edu/~lefsky/slicerpage.html*
- *M.A. Lefsky, W.B. Cohen, S. A. Acker, T.A. Spies, G.G. Parker, D. Harding, LIDAR REMOTE SENSING OF FOREST CANOPY STRUCTURE AND RELATED BIOPHYSICAL PARAMETERS AT THE H.J. ANDREWS EXPERIMENTAL FOREST, OREGON, USA, Pages 79-91 in: Greer, J.D. ed. 1997 Natural Resources Management using Remote Sensing and GIS. ASPRS, Washington D.C.*
- *http://ltpwww.gsfc.nasa.gov/division/VCLhome/VCLInst.html*

Scanning lidar remote sensing systems have recently become generally available for use in ecological applications. Unlike microwave and conventional optical sensors, lidar sensors directly measure the distribution of vegetation material along a vertical axis and can be used to provide three-dimensional characterizations of vegetation structure. Ecological applications of scanning lidar have previously used uni-dimensional indices of canopy height. A new three-dimensional approach to interpreting lidar waveforms was developed to characterize the total volume of vegetation and empty space within the forest canopy, and their spatial organization. These aspects of the physical structure of canopies have been infrequently measured, either from field or remote methods.

Characterization of forest structure in moderate to high biomass systems is one of the key challenges in remote sensing. The ability to remotely sense both the total biomass and physical structure of these forests would provide one way to meet the need for forest inventory in support of research and management of both carbon balance and habitat conditions.

The SLICER (Scanning Lidar Imager Of Canopies By Echo Recovery) instrument is one of a new generation of lidar remote sensing systems that augment traditional first-return laser altimetry with a surface lidar capability. Laser altimeters measure the distance between the sensor and a target through the precise measurement of the time between the emission of a pulse of laser light from the sensor, and the time of detection of light reflected from the target. In surface lidar, the power of the entire returning laser signal is digitized, resulting in a waveform that records the vertical distribution of the backscatter of laser illumination from all canopy elements (foliar and woody) and the ground reflection, at the wavelength of the transmitted pulse (1064 nm, in the near-infrared). The use of relatively large footprints (5-25 m) is optimized to recover returns from the top of the canopy and the ground in the same waveform, yet be small enough to be sensitive to the contribution of individual crowns.
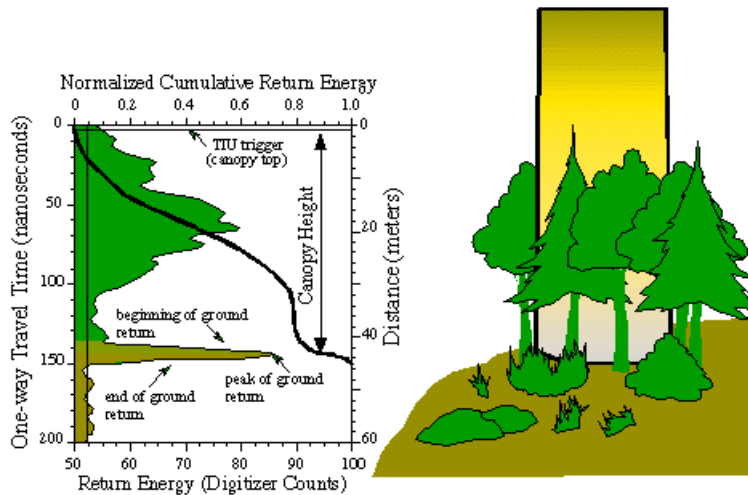
SLICER RETURN PULSE WAVEFORM

Figure adapted from Blair, Harding, NASA GSFC

Most work in remote sensing of forested systems has been analysis in the spectral domain, specifically the analysis of multi-spectral images. The spectral qualities of forest stands are due both to the electromagnetic properties of the elements that are present in each pixel (foliage, soil, woody debris, etc.) and to their three-dimensional organization, which determines the total cover of each element in the pixel and the distribution of illumination and shadow on those elements. While these type of images have been extremely valuable for general vegetation mapping, it has not been possible to derive from them detailed biophysical information for moderately-high to high-biomass. A second approach to the use of
multi-spectral images is analysis in the spatial domain. This approach uses the two-dimensional spatial pattern of spectral variability to infer aspects of the physical organization of forests, and has had considerable success in predicting forest structure. However, these approaches have been constrained by the requirement that they infer three dimensional structure from its two-dimensional projection, in the form of a remotely sensed image.

The failure of optical sensors to adequately measure the structural properties of forests is not incidental; it is a consequence of the imaging process itself. These sensors integrate energy and matter fluxes along the axis between the sensor and scene. Integration along this dimension, roughly corresponding to the "z-axis" or height of vegetation, combines the overstory, understory, and soil measurements into a single measurement per pixel. This results in a two-dimensional remotely-sensed image, with the spectral data for each pixel most influenced by cover of canopy structure. More detailed structural information from closed canopies is only possible when there are associated changes in the horizontal dimension (both within and among pixels), such as varying proportions of tree and shadow with changing tree density, or by the increased presence of new scene components (such as lichen) in the forest canopy. Lidar does not suffer from this limitation, because it directly measures the vertical distribution of aboveground plant biomass.

The first generation lidar sensors for vegetation studies were designed to record distance to the first reflective surface intercepted by a laser pulse over a relatively small sampling area, or footprint, approximately 1 m in diameter. Returns from the top surface of a forest canopy were combined with subsequent measurements of distance to the forest floor, obtained through gaps in the forest canopy, to infer the height of the dominant trees. A later, more sophisticated technique involved recording the distance to the first and last reflective surface for each footprint, giving a direct height measurement for each observation. Such techniques have proven useful for predicting canopy height, timber volume and forest biomass, and percent canopy cover. However, the relatively small geographic area covered in these data sets, challenges in analyzing the data, and the lack of standardized methods for their geolocation have limited the use of conventional lidar sensors within the ecological community.

The new generation of lidar instruments developed at NASA's Goddard Space Flight Center and elsewhere have minimized these barriers to a wider application of the technology. Whereas earlier devices used a small footprint and most often measured the distance to the first reflective surface, the newer devices send out a laser pulse over an approximately 5-25 m diameter footprint, and record the timing and power of backscattered light over the full height profile. Although the power of the return signal falls as the signal is intercepted by canopy structure, return energy from the ground is recorded in nearly all waveforms, which allows an estimate of the total height of the stand, and indicates that some energy is available for the detection of understory foliage, where present. Using an algorithm developed by Drs. D. Harding and M. Lefsky, the lidar waveform can be transformed to estimate the bulk canopy transmittance and the vertical distribution of reflective canopy surfaces.
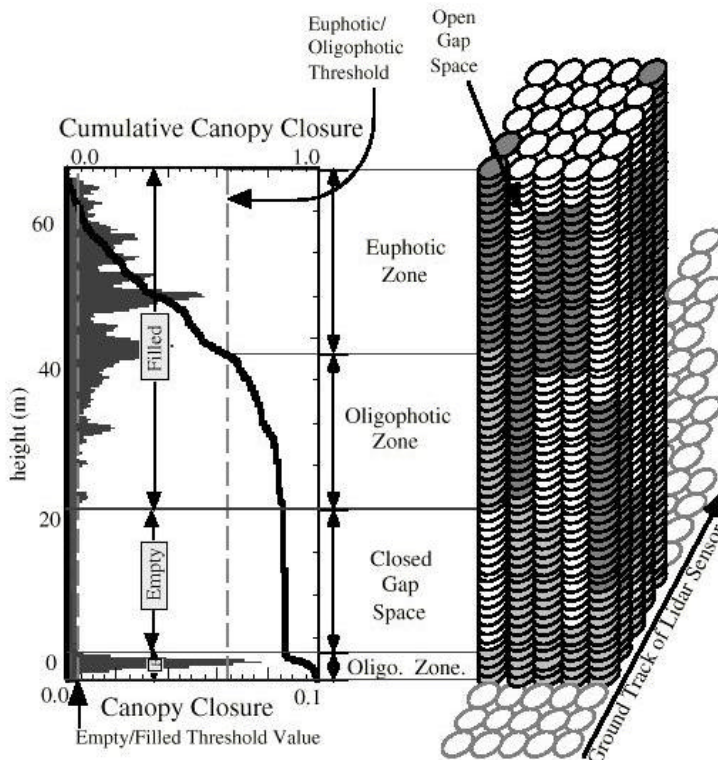
The canopy colume method of M. Lefsky and D. Harding – transforming the waveform into an estimate of the canopy height profile (CHP), the relative distribution of the canopy as a function of height. A threshold value was then used to classify each element of the CHP into either "filled" or "empty" volume, depending on the presence or absence (in the profile) of canopy material. A second step classified the filled elements of the matrix into an "euphotic" zone which contains all filled elements of the profile that are within the uppermost 65 % of canopy closure, and an "oligophotic" zone, consisting of the balance of the filled elements of the profile. These two classifications were then combined to form three classes; empty volume beneath the canopy- (ie. closed gap space), filled volume within the euphotic zone, and filled volume within the oligophotic zone.

Using these methods, very good estimates of aboveground biomass, Leaf Area Index, number of stems greater than 100cm, Diameter at Breast Height, and distinctions can be made between young, mature and old-growth plots.

9.4.2.1 Multi-Beam Laser Altimeter (MBLA) - The Lidar Instrument for the Vegetation Canopy Lidar (VCL)

Motivation for work relating forest attributes to lidar sensed canopy structure has been enhanced by the announcement that VCL, the Vegetation Canopy Lidar mission, has been

funded by NASA's Earth System Science Pathfinder (ESSP) program, scheduled to be launched in mid 2000.

The Vegetation Canopy Lidar (VCL) Mission utilizes pulsed laser radar (i.e. lidar) from a single-instrument, small spacecraft in a 65 degree inclination, 400 km altitude, circular, Earth orbit for continuous, global remote sensing of tree canopy height, vertical distribution of intercepted surfaces in the canopy, and ground topography. The lidar instrument for VCL is called the Multi-Beam Laser Altimeter (MBLA). It has 5 laser transmitters, a large receiver telescope, a set of 5 detectors and laser-pulse analysis electronics, computer data/command electronics, and a pointing angle measuring system. The primary role for MBLA in measurement of the Earth's vegetation canopy is enabled by the related techniques of laser pulse time-of-flight measurement and laser pulse shape analysis that constitute a lidar capability for remote sensing of the Earth's surface.
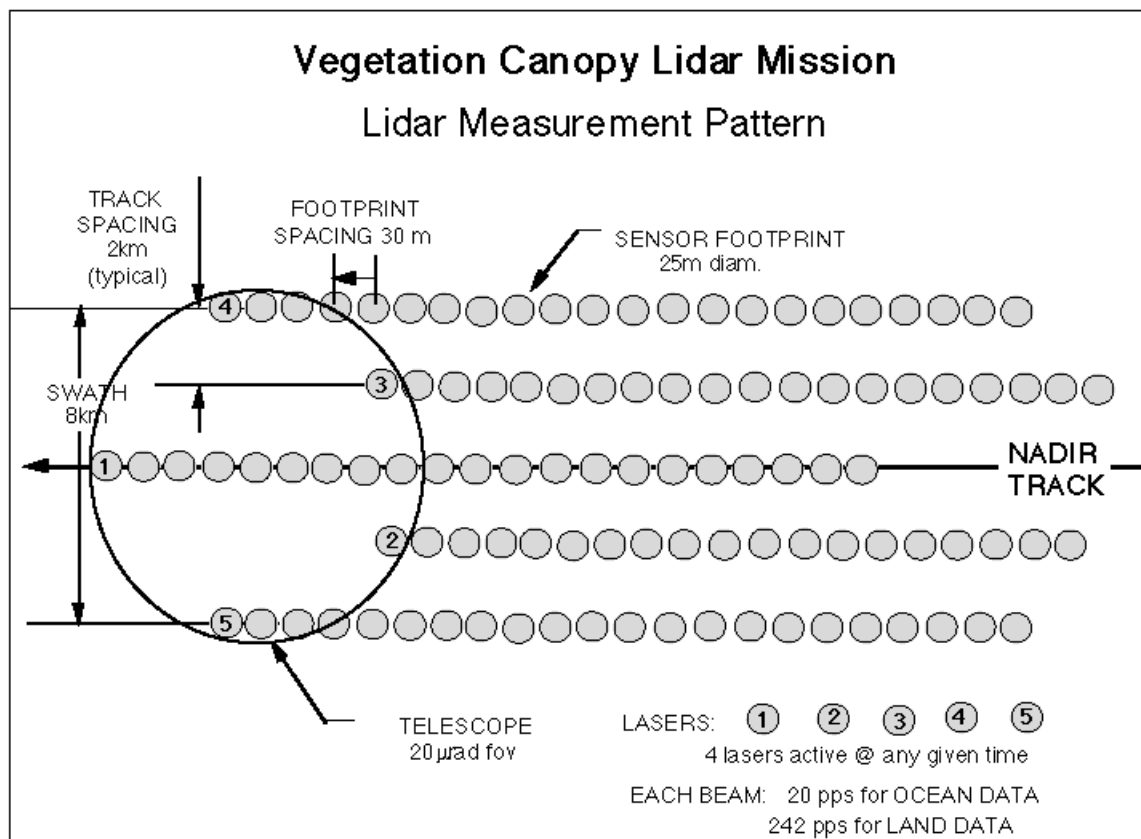
The MBLA lidar data have a greater spatial extent than those of conventional space-based nadir-profiling lidar systems due to increases in both coverage and sampling of the Earth's vegetation and topography. Spatial coverage is increased across-track by providing information with multiple beams, each of which samples independently varying landforms. Sampling is increased by using higher laser pulse rates in each beam. In our MBLA measurement concept, 5 beams, each operating at the 1064 nm fundamental wavelength of the Nd:YAG solid-state laser, are arranged in a pentagon pattern within a 20 urad circle that is centered on nadir. When one point of this pentagon is oriented along the flight direction, VCL simultaneously produces 5 tracks of Earth surface information; 1 track at nadir and 2 tracks each to the left and right that are spaced apart by 5 urad. For the nominal VCL orbital altitude of 400 km the

across-track separation between adjacent tracks is 2 km. The result is an 8 km wide strip image of Earth surface vertical structure.

9.5 Scanning laser mapping of the coastal zone

9.5.1 Historic development

The utilization of an airborne laser as a depth "sounder" has been studied and tested at the U.S. Naval Oceanographic Office since 1966. The experimental laser system at the early 1970's was able to measure depths down to 10m. Since then much development has taken place.

Today there are three fully operational airborne lidar bathymetry systems in operation: SHOALS (Scanning Hydrographic Operational Airborne Lidar Survey) , LADS (Laser Airborne Depth Sounder), and Hawk Eye.

The SHOALS system, developed for the US Army Corps of Engineers (USACE), employs lidar technology to remotely collect accurate, high-density measurements of both bathymetry and topography in coastal regions. SHOALS has been in full operation since March 1994 and to date surveyed more than 230 projects totalling 5000 km 2 . These surveys cover a variety of project types including coverage of maintained channels and harbours, coastal structures, and dredged material placement areas as well as adjacent beaches. SHOALS data collected for the US Navy and for the National Oceanic and Atmospheric Administration (NOAA) were used for creation of nautical charts. Other SHOALS surveys were for beach nourishment and erosion monitoring and for emergency response to hurricanes and ship groundings.

Originally developed by Australia's Defence, Science and Technology Organisation (DSTO) for the Royal Australian Navy, the Laser Airborne Depth Sounder (LADS), fulfils a crucial need. Two hundred years after Cook, Flinders and other explorers of the Southern Hemisphere started charting Australia's vast coastline, nearly half of the continental shelf, or around one million square kilometres, remains unsurveyed or incompletely surveyed. LADS was developed to help speed up surveying and charting progress.

The LADS system has been providing valuable data for The Australian Hydrographic Service over the last seven years (since 1993). 10,000 soundings per square kilometre, spaced 10 metres apart in a swath 240 metres wide; LADS provides accurate, high density digital depth and positional data of coastal waters up to 50 metres in depth. Flying at 145 knots, 500 metres above the sea, unhindered by reefs or shallows, LADS surveys the sea floor at a rate in excess of 50 square kilometres an hour.

The swedish helicopter borne lidar called FLASH (FOA Laser Airborne Sounder for Hydrography – started operating in 1986) has been further developed into two operational systems called Hawk Eye. Saab Instruments is the main contractor and Optech Inc. as the main subcontractor. FOA (Swedish Defence Research Establishment) is member of the Hawk Eye project team together with the Swedish Hydrographic Department, the Swedish Navy and the Swedish Material Administration. The first Hawk Eye unit was delivered in 1993.

A Hydro Optical Transmission Sensor (HOTS) was developed in conjunction with Hawk Eye to obtain depth profiles of beam attenuation, for predicting the effect of the optical properties of water on Hawk Eye performance. HOTS is used in the preplanning phase of Hawk Eye operations. The design of the instrument features

easy operation by one man from any surface or helicopter platform. The system is lowered into the sea by hand and sinks at a rate of 0.5 m/sec. Reading are taken and recorded every 0.5 m. When the bottom is touched the instrument is recovered by the line. The stored data is transferred to a PC for evaluation.

9.5.2 Working principles

An airborne lidar bathymeter uses lidar technology to measure water depths. A laser transmitter / receiver mounted on an aircraft transmits a laser pulse which travels to the air–water interface where a portion of this energy reflects back to the receiver.

The remaining energy propagates through the water column and reflects off the sea bottom. The water depth comes directly from the time lapse between the surface return and bottom return, and each sounding is appropriately corrected for surface waves and water level fluctuations (see figure). In practical application of this technology, laser energy is lost due to refraction, scattering, and absorption at the water surface, sea bottom, and as the pulse travels through the water column. The combination of these effects limits the strength of the bottom return and therefore limits the maximum detectable depth. Optical water clarity and bottom type are the two most limiting factors for depth detection. Typically, lidar bathymeters collect through depths equal to three times the site's Secchi (visible) depth.
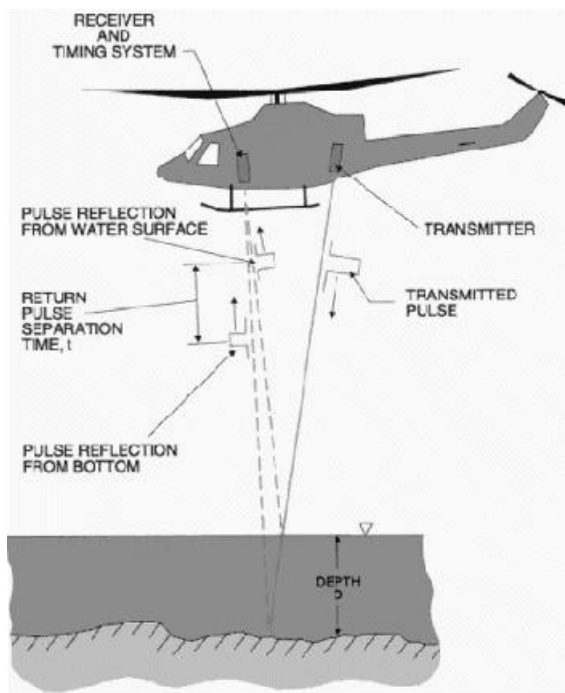
Fig. 1. Lidar operating principle.

206
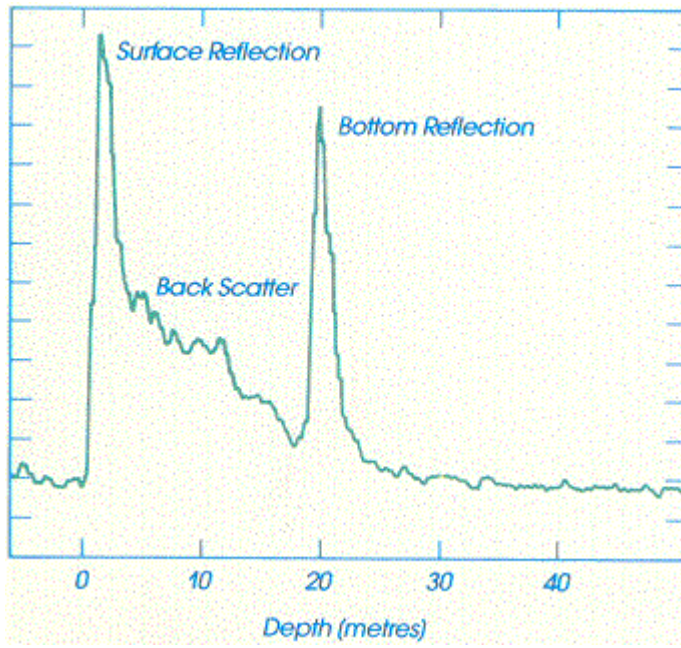
figure2: This is an example of a waveform that the Ground-Based Unit receives

The following table summarises the performance of ALS used for bathymetry (the performance depends off course on flight height and speed):

ALS system	LADS	SHOALS	Hawk Eye
Survey height	500m	200m	50 to 800m
Survey speed	145 knots (75m/s)	60 m/s	
Wavelength (IR)	1,064nm	1,064nm	1,064nm
Wavelength (green)	532nm	532nm	532nm
Sounding rate	168 per second		200 per second
Swath width	240m (nominal)	110m	programmable, max 0.85*altitude
Sounding spacing	10m (nominal) – thus, providing full coverage at depths shallower than 12m	4m	1 to 15m
Depth range	2 to 50m	Max 60m	1 to 40m
Area coverage	54 km^2/hour	16 km^2/h	18 km^2/h for 300m altitude and 5m density
Depth sounding accuracy	better than 0.3m (one standard deviation) over depth range 2-30m	±15cm	0.3m
Positional accuracy of sounding	15m	±3m (with DGPS), ±1m (with KGPS)	3m
Diameter of laser footprint at the bottom	2m		

Weather conditions can affect the data collected (as reported for the LADS):

1. *winds*
    - strong winds increase turbidity, particularly in inshore areas and the rougher sea increases beam refraction and scattering, degrading the overall quality of bottom returns and adversely affecting sounding accuracy.
    - glass calm conditions on the other hand, cause loss of surface returns from the extremity laser spots with consequent degredation of the surface model and possible increased depth errors due to undetected platform tilt.
    - optimum conditions for LADS are light to moderate winds with sea states between 1 and 3.
2. *clouds*
    - high clouds, particularly stratus, can improve performance by reducing ambient light levels and sun glint.
    - low clouds (base below 1640 feet) prevents data collection, yielding bright reflections and obscuring firstly the infra-red and then the green laser beam.

The accuracy of ALS soundings is dependent off course both on the accuracy of the ALS depth measurement, and on the accuracy of tidal corrections applied.

9.5.3 Benefits
Airborne data collection eliminates the difficulties associated with shipborne survey of shallow, dangerous or complex waters such as reef areas. It also permits the fast identification of key areas of navigational significance, such as undiscovered channels and passages.

Powerful analysis and conversion capabilities provide same day turn-around of data. This enables efficient, flexible survey sortie operations.

4 Seconds of Ship Survey
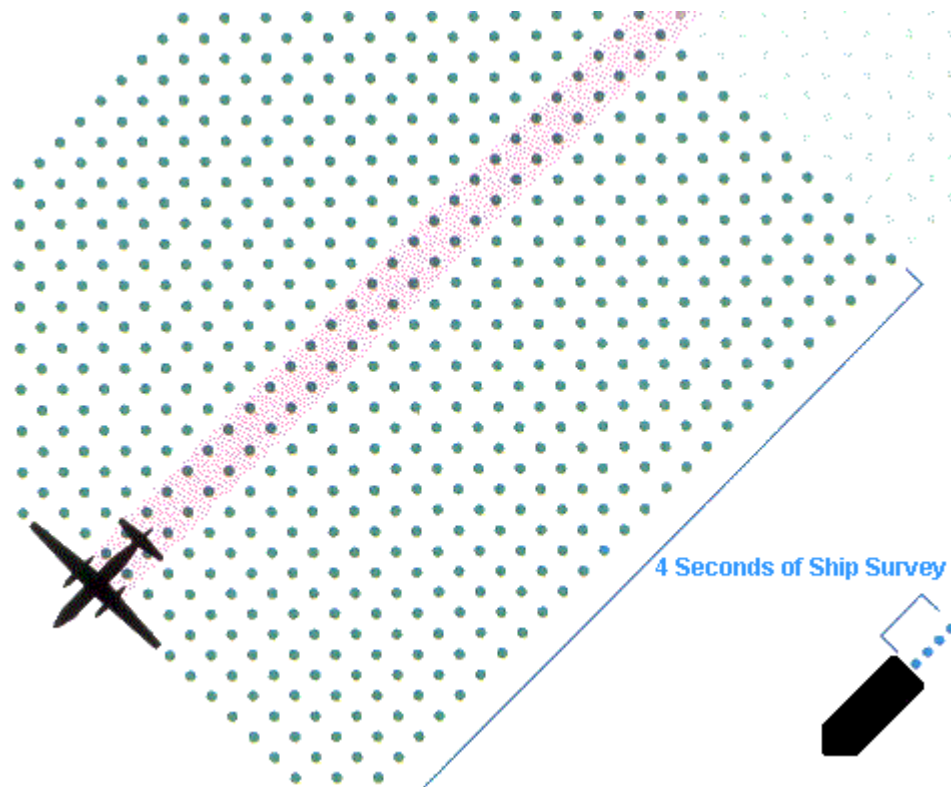
The above figure (taken from the LADS pages) compares the coverage of soundings of an LADS and a conventional echo-sounder technology, the productivity gains are readily understood. However, ALS will never replace ships, due to its limitations.

9.5.4 Case study: The ALACE project - Airborne LIDAR Assessment of Coastal Erosion Project

Beaches are some of the earth's most dynamic geologic features. Beach morphology fluctuates over a wide range of time scales, varying from periods of hours associated with diurnal tides and storm events, to years and decades associated with long-term erosional trends. Human actions, especially during the last 100 years, have created a situation in which beach erosion can have severe economic consequences.

Accurate and timely assessment of erosion conditions and storm impacts is needed to assist decision making on land use, beach renourishment, erosion calculations, insurance compensation, and property value estimation. Proper storm damage assessment is an enormous task for emergency and disaster response agencies and personnel.

The ALACE project is a partnership between the National Oceanic and Atmospheric Administration (NOAA) Coastal Services Center (CSC) in Charleston, South Carolina; NASA Goddard Space Flight Center, Wallops Flight Facility (WFF) in Wallops, Virginia; and U.S. Geological Survey (USGS) Center for Coastal Geology in St. Petersburg, Florida. The project's goal is to establish the capability of aircraft laser swath mapping to provide highly accurate, cost-effective information on coastal topography, erosion, and shoreline position. In working toward this goal, NOAA, NASA, and USGS have conducted several mapping missions along significant portions of U.S. coast using the NASA Airborne Topographic Mapper (ATM) flown aboard a NOAA Twin Otter aircraft.

The ATM is currently operated with a Spectra Physics TFR laser transmitter that provides a 7-nanosecond wide, 250-micro-joule pulse at a frequency-doubled wavelength of 523 nanometers (nm) in the blue-green spectral region. A scan mirror with an off-nadir angle of 15 degrees was used in the ALACE beach mapping surveys, producing an elliptical scan pattern with a swath width equal to approximately 50 percent of the approximately 700-meter aircraft altitude. The usual density of points on the ground is 30 per 100 square meters.
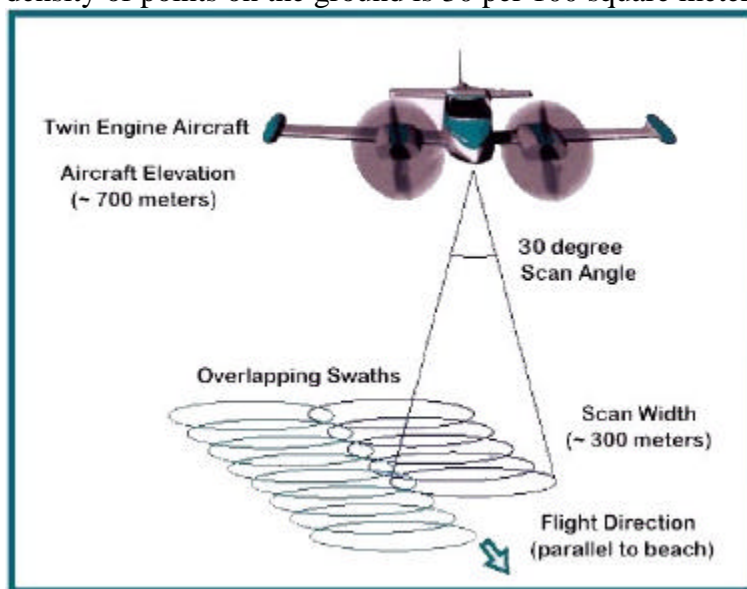
Figure: ATM Laser Mapping Scan Swath

For the fall 1997 mapping missions, a passive channel sensor was added to the ATM. This sensor collects geo-referenced panchromatic (excluding 523 nm) data along the same elliptical

scan path as the active laser. Images created from the passive channel data help identify ground features, and are used to assist in the delineation of the beach region.

Comparisons were made between same day ATM surveys, multiple days ATM surveys, and with data collected by buggy trucks, Reiss total station survey, and others.

The statistical results from the ATM comparisons are summarized in the following tables (m = mean, s = standard deviation).

ATM Comparisons	μ (cm)	σ (cm)	RMS (cm)	# Points	Comments
09/26/97 ATM minus 09/26/97 ATM	3.4	7.6	8.3	7,113	Over Wright Memorial
09/27/97 ATM minus 09/27/97 ATM	8.0	11.4	14.0	21,085	Over Wright Memorial
09/26/97 ATM minus 09/27/97 ATM	9.2	11.0	14.3	48,154	Over Wright Memorial
135929 ATM minus all 09/26/97 ATM	-6.1	12.1	13.5	8,939	Constrained to List buggy track
182303 ATM minus all 09/27/97 ATM	0.9	13.8	13.8	4,691	Constrained to List buggy track
Individual 09/27/97 ATM passes minus combined 09/26/97 ATM	-12.1	14.7	19.1	32,700	Constrained to List buggy track
135929 ATM minus all 09/26/97 ATM	-6.6	14.8	16.2	1,097,154	Beach surface
182303 ATM minus all 09/27/97 ATM	-0.8	19.0	19.1	535,803	Beach surface
Individual 09/27/97 ATM passes minus combined 09/26/97 ATM	-9.4	20.0	22.1	3,788,375	Beach surface

Table 11. Summary of statistics from ATM to ATM survey comparisons

ATM Comparisons	μ (cm)	σ (cm)	RMS (cm)	# Points	Comments
09/26/97 ATM minus List buggy	8.7	10.4	13.6	29,588	
09/27/97 ATM minus List buggy	-4.2	13.0	13.6	19,783	
09/26/97 ATM minus Holman buggy	14.1	14.4	20.1	7,957	
09/26/97 ATM minus Holman buggy	14.9	13.5	20.1	163	Constrained to List buggy track
09/27/97 ATM minus 09/27/97 Reiss	8.5	16.0	17.5	11	Beach surface
09/27/97 ATM minus 09/27/97 Hansen	8.6	12.8	15.4	2,592	Beach surface

Table 12. Summary of statistics from ATM to ground survey comparisons

It can be seen that ATM measurements were found to match the desired accuracy about 10 cm, in most of the cases, amounting to +/- 15 cm over the beach. The horizontal accuracy is 1.5m in the worst case.

The next figure shows a preliminary raw data image (note the elliptic pattern of the scanner):

Airborne LIDAR Assessment of Coastal Erosion (ALACE)

Raw beach survey elevation data acquired in early October 1996 over north Assateague Is. and Ocean City, Maryland from a laser altimeter, the NASA Airborne Topographic Mapper (ATM). Elevations from -38 to -31 meters (WGS-84 ellipsoid height) are continuously shaded from purple/blue (low elevation) to orange/red (high elevation).
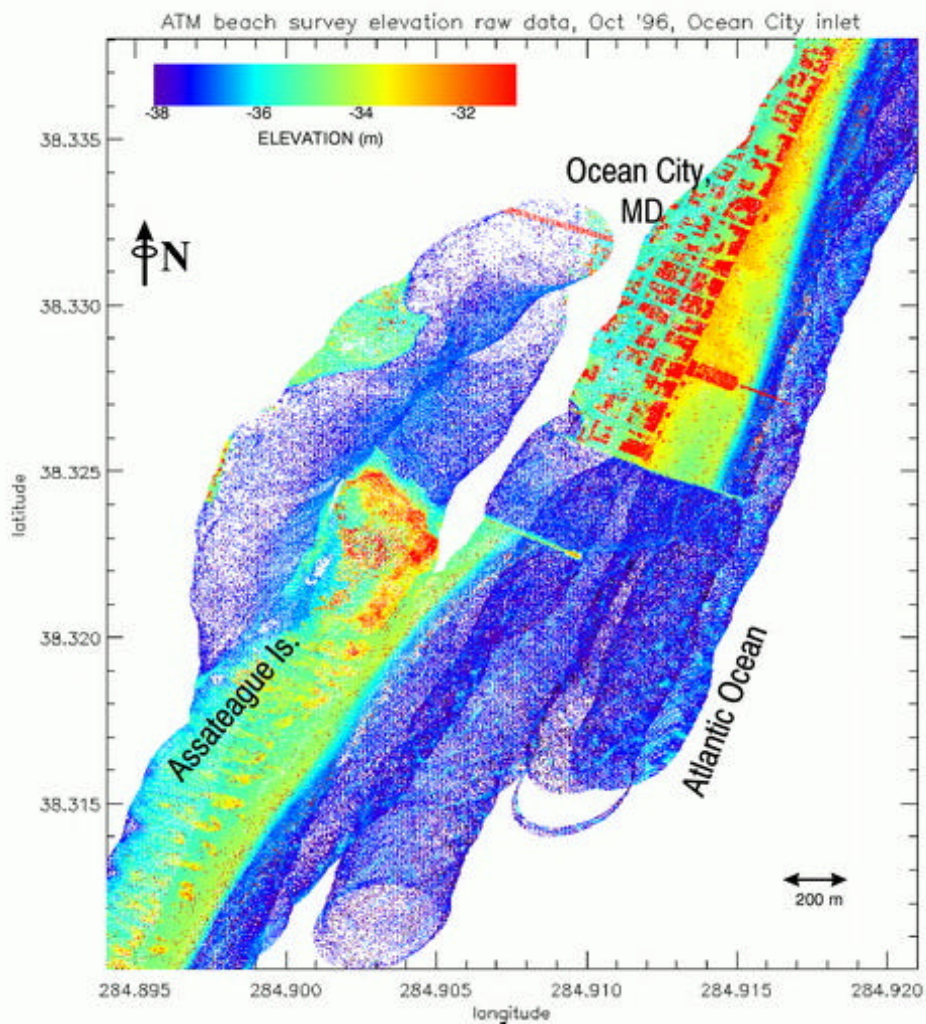
The following figure presents a beach profile of Assateague Island, as surveyed by the ATM-2 in october 1996:

213
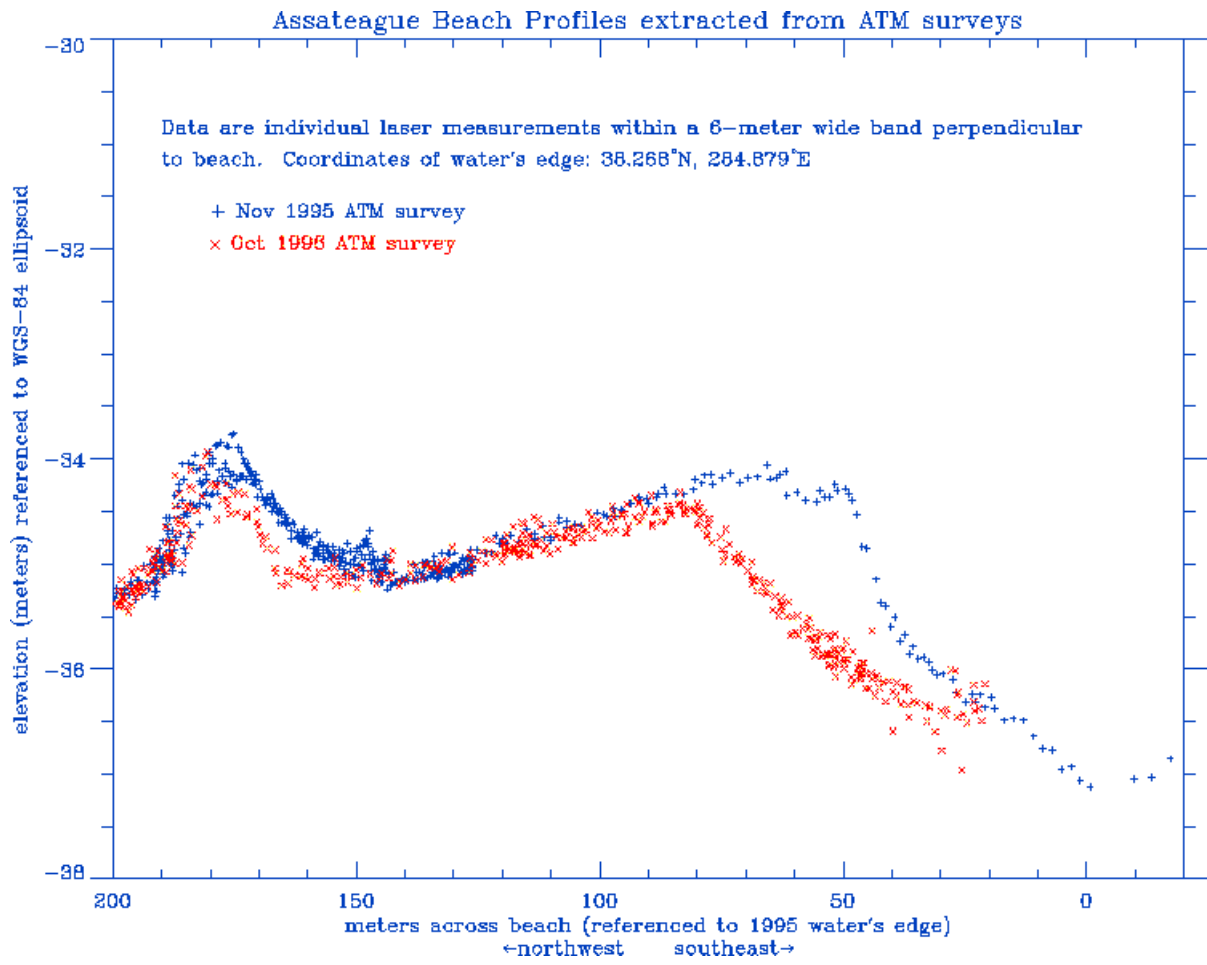
9.5.5 IHO and FEMA standards, and their relation to ALS technology

9.5.5.1 IHO S-44

The following table is a Summary of the Minimum Standards for Hydrographic Surveys, as required by S-44 (special publication no. 44) of the International Hydrographic Organization (IHO):

Order	Special	1	2	3
Examples of typical areas	Harbours, berthing areas, and associated critical chanells with minimum underkeel clearences	Harbours, harbour approach channels, recommended tracks and some coastal areas with depth up to 100m	Areas not described in Special Order and Order 1, or areas up to 200m water depth	Offshore areas not described in Special Order, and Orders 1 and 2
Horizontal accuracy (95% confidence level)	2m	5m + 5% of depth	20m + 5% of depth	150m + 5% of depth
Depth accuracy for reduced depths (95% confidence level)	0.25m	0.5m	1.0m	Same as order 2
100% bottom search	Compulsory	Required in selected areas	May be required in selected areas	Not applicable
System detection capability	Cubic features > 1m	Cubic features > 2m in depths up to 40m; 10% of depth beyond 40m	Same as order 1	Not applicable
Maximum line spacing	Not applicable, as 100% search compulsory	3*average depth or 25m, whichever is greater	3-4*average depth or 200m, whichever is greater	4*average depth

ALS is in fact the only Remote Sensing technology included in the S-44, and I quote:

"Airborne laser sounding is a new technology which can offer substantial productivity gains for surveys in shallow, clear water. Airborne laser systems are capable of measuring depths to 50 m or more."
"Airborne laser systems are capable of measuring depths to 50 m or more provided the water is clear. Hazards to navigation detected by airborne laser should be examined using SBES (Single Beam Echo Sounder), MBES (Multi Beam Echo Sounder) or high density airborne laser. All swaths should be crossed, at least once, by a checkline to confirm, by this method, the accuracy of positioning, depth measurement and depth reductions."

A comparison analysis between MBES and ALS in performing hydrographic surveys in shallow waters, was done by Axelsson and Alfredsson (both from Saab). Their main conclusions are as follows:

- The cost for MBES survey mainly depends on the depths of the waters surveyed, with very high costs for shallow waters, decreasing a lot for deeper waters. The cost for surveying with ALS is mainly dependent on the IHO order and not on the depths.
- Configurations using combinations of MBES platforms and ALS systems have much lower annual costs, with the ALS covering most of the shallow waters.
- ALS surveys offer the ideal complement to sonar surveys, for several reasons:

214

- ➢ Survey efficiency, ALS provides in shallow waters outstanding productivity with minimum crew requirements. High velocity of the carrier and a large swath width give fast coverage of the sea bed. Narrow passages, archipelagos, reefs and the coast line can be surveyed in just one mission.
- ➢ Survey accuracy, ALS gives full bottom coverage in all areas even in very shallow areas where other methods is very expensive to use. Local variations in salinity and/or temperature present no problem to ALS.
- ➢ Survey safety, the survey platforms of an ALS is airborne and readily clears narrow passages, shoals and reefs. Unknown waters and areas subject to mine hazards are no safety problems for the survey crew or the survey equipment.
- ➢ Versatility, an ALS may be adapted for environmental control, e.g. erosion, contamination or reef growth/breakdown periodical surveys. In its environmental control role, the ALS also benefits from the carrier's short re-deployment time.

9.5.5.2 FEMA

The United States Federal Emergency Management Agency (FEMA) has published guidelines and specifications that must be used for the application of Airborne LIght Detection And Ranging (LIDAR) systems for gathering the data necessary to create digital elevation models (DEMs), digital terrain maps, and other National Flood Insurance Program (NFIP) products.

Given below are the main guidelines and specifications, to give the reader an idea about the complications involved

9.5.2.2.1 AIRBORNE LIGHT DETECTION AND RANGING SYSTEMS - General Guidelines for Use

Two important factors in the LIDAR system mission planning are the point density of the randomly spaced LIDAR points and the point spacing of the uniformly spaced DEM points derived from the randomly spaced LIDAR returns. The correct point density necessary to accurately represent terrain and terrain features will depend on flight conditions, mission purpose, and other variables. As discussed later in this Appendix, DEM point spacing of 5 meters and vertical accuracy of 30 centimeters are required.

Flight-path planning is another important factor in the LIDAR system mission. The flight path shall include both parallel and cross flight lines to cover the study area satisfactorily.

Unlike aerial photogrammetry, LIDAR missions can be flown without regard to sun angle. Flights may take place at night, if conditions otherwise allow.

Elevation and measurement information related to subsurface channel and hydraulic structure geometry must be obtained through the use of other mapping technologies over deep or turbid water. In some instances, shallow water and near-shore coastal surveys can be accomplished using LIDAR systems equipped with lasers operating in portions of the light spectrum that allow transmission through water.

LIDAR system tolerance for inclement weather conditions (e.g., high winds, wet snow, rain, fog, high humidity, low cloud cover) generally is higher than that of other photogrammetric methods. However, such conditions have been known to degrade the accuracy of laser return data. Therefore, the contractor shall generally avoid missions during inclement weather.

High point densities may allow satisfactory data collection in areas of dense foliage. Still, care must be taken in planning missions with regard to both natural (vegetative) and manmade (structure) ground cover. Pulse width, beam divergence, first and last pulse return discrimination, and choice of the post-processing algorithms may all affect the accuracy of LIDAR-derived data in areas of dense foliage.

9.5.5.2.2 Performance Standards

The appropriate mapping standards in Appendix 4 of these Guidelines shall apply to NFIP maps and map products derived from LIDAR systems. LIDAR-derived data must have the accuracy required to produce topographic maps and products that meet the National Standard for Spatial Data Accuracy (NSSDA).

FEMA is not aware of any existing LIDAR system performance standards. Current information about

LIDAR systems is available from the National Oceanic and Atmospheric Administration (NOAA), National Aeronautic and Space Administration, U.S. Army Corps of Engineers, LIDAR system manufacturers and vendors, and private firms that provide LIDAR system services. As professional or trade associations issue specifications and standards, FEMA may adopt them and amend this Appendix.

A. Performance Standards

The contractor must furnish all necessary materials and equipment. The contractor also must supply the supervisory, professional, and technical services personnel required to manage, survey, document, and process all data associated with LIDAR system mapping, scanning, and digital image processing. All deliverables must be provided in accordance with the contract and the requirements in this Appendix.

B. System Calibration

LIDAR system components are most effectively tested and calibrated by the equipment manufacturer. Therefore, the contractor must provide FEMA with evidence of manufacturer calibration. In addition to evidence of manufacturer calibration of system components, the contractor must submit evidence that the total LIDAR system was calibrated prior to project initiation for the purposes of identifying and correcting systematic errors. Proper system calibration requires repetitive overflight of terrain features of known and documented size and elevation using flight paths similar to those that will be used in the study area.

C. Flight Planning

Planning a flight path that considers all aspects of data collection is critical to the success of the mission. An analysis of the project area, project requirements, topography, proximity to restricted air space, and other factors will determine the flight path configuration. The mission should include parallel flight lines and at least one cross flight line. The spacing between the flight lines will depend on the desired amount of sidelap between swaths.

The density and accuracy of data generated by different equipment vary widely. The contractor shall have the flexibility of providing a flight path to create the necessary point density to minimize the occurrence of data voids.

The contractor must check the Position Dilution of Precision (PDOP) in the study area. The PDOP is an indicator of the positional accuracy that can be derived from the current GPS satellite geometry, which varies continuously; the smaller the PDOP number, the higher the data quality.

The contractor must document mission date, time, flight altitude, airspeed, scan angle, scan rate, laser pulse rates, and other information deemed pertinent. For a sample mission data recordation checklist, refer to Table A4B-1.

D. GPS Base Stations

The contractor must select the GPS base station(s) carefully to ensure reliable differential processing of airborne GPS data. The National Geodetic Survey (NGS) recommends the simultaneous use of two GPS base stations during the mission. (Note: Either public- or private-domain GPS base stations are suitable for use for this purpose.) Where possible, GPS base stations shall have ellipsoid height to an accuracy of 2 centimeters relative to the Continuously Operating Reference Stations (CORS) or the High Accuracy Reference Network (HARN), both operated by the NGS. The contractor must use a high-quality, dual-frequency GPS receiver and associated antenna at the GPS base stations.

9.5.5.2.3 GPS Control

Part 1, "Reporting Methodology (FGDC-STD-007.1)," and Part 2, "Standards for Geodetic Networks (FGDC-STD-007.2)," of the Geospatial Positioning Accuracy Standards, published by the FGDC in 1998, provide a common methodology for determining and reporting the accuracy of horizontal and vertical coordinates for geodetic control points (survey monuments). Additional guidance is included in NOAA Technical Memorandum NOS NGS-58, "Guidelines for Establishing GPS-Derived Ellipsoid Heights (Standards: 2 cm and 5 cm)," dated November 1997. The GPS control guidance in FGDC-STD-007.1 and FGDC-STD-007.2 and in Appendix 4 of these Guidelines shall apply to LIDAR-derived data submitted to FEMA.

9.5.5.2.4 Post-Processing of Data

For hydraulic modeling, the contractor must provide high-resolution, high-accuracy, "bare-earth" ground elevation data. To restrict data to ground elevations only, the contractor must remove elevation points on bridges, buildings, and other structures and on vegetation from the LIDAR-derived data. In addition to randomly spaced LIDAR points, before and after removal of data associated with structures and vegetation, the contractor must produce a bare-earth DEM, with regular 5-meter point spacing in eastings and northings. In accordance with NSSDA, the contractor must use Triangular Irregular Network (TIN) linear interpolation procedures when validating the vertical accuracy of the DEM.

In addition to DEMs, the contractor shall produce breaklines for stream centerlines, drainage ditches, tops and bottoms of streambanks, ridge lines, road crowns, levees, bulkheads, road/highway embankments, and selected manmade features that constrict or control the flow of water (e.g., curb lines).

9.5.2.2.5 Quality Control/Quality Assurance
Quality Control/Quality Assurance (QC/QA) of the LIDAR-derived data is primarily the responsibility of the contractor. This QC/QA process shall include reviews of flight alignments and completeness of supporting data (e.g., cross sections, profiles). FEMA or its designee may perform additional QC/QA testing.

NSSDA uses the root mean square error (RMSE) to estimate both horizontal and vertical accuracy. RMSE is the square root of the average of the set of squared differences between dataset coordinate values and coordinate values from an independent source of higher accuracy for identical points. If those differences are normally distributed and average zero, 95 percent of any sufficiently large sample should be less than 1.96 times the RMSE. Therefore 15-centimeter RMSE is often referred to as "30-centimeter accuracy at the 95-percent confidence level." Following that convention, the vertical accuracy of any DEM is defined as 1.96 times the RMSE of linearly interpolated elevations in the DEM, as compared with known elevations from high-accuracy test points.

DEMs should have a maximum RMSE of 15 centimeters, which is roughly equivalent to 1-foot accuracy. The contractor must field verify the vertical accuracy of this DEM to ensure that the 15-centimeter RMSE requirement is satisfied for all major vegetation categories that predominate within the floodplain being studied. The main categories of ground cover that the contractor must separately evaluate and report on the DEM accuracy for shall be:

a) Bare-earth and low grass (plowed fields, lawns, golf courses);
b) High grass and crops (hay fields, corn fields, wheat fields);
c) Brush lands and low trees (chaparrals, mesquite, swamps);
d) Fully covered by trees (hardwoods, evergreens, mixed forests); and
e) Urban areas (high, dense manmade structures).

The contractor shall evenly distribute sample points throughout each category area being evaluated and not group the sample points in a small subarea.

The RMSE calculated from a sample of test points will not be the RMSE of the DEM. The calculated value may be higher or it may be lower than that of the DEM. Confidence in the calculated value increases with the number of test points. If the errors (lack of accuracy) associated with the DEM are normally distributed and unbiased, the confidence in the calculated RMSE can be determined as a function of sample size. Similarly, the sample RMSE necessary to obtain 95-percent confidence that the DEM RMSE is less than 15 centimeters can also be determined as a function of sample size.

For each major vegetation category, the contractor must test a sample of points and show the test points have an RMSE less than

$$RMSE_{sample} \leq 15\sqrt{\frac{(n-1)-2.326\sqrt{n-1}}{n}}$$

where n is the number of test points in the sample.

The contractor must select a minimum of 20 test points for each major vegetation category identified. Therefore, a minimum of 60 test points must be selected for three (minimum) major vegetation categories, 80 test points for four major categories, and so on. The contractor should consider establishing test points when planning field surveys to gather cross section data for hydraulic modeling. If more than two test points are outside the range of two times the RMSE, the contractor must make the appropriate adjustment using guidance in Appendix 4 of these Guidelines.

The contractor shall select the test points carefully in areas of the highest PDOP to evaluate DEM accuracy under trees and in vegetation representative of the study area. Test points on sloping or irregular terrain would be unreasonably affected by the linear interpolation of test points from surrounding DEM points and, therefore, shall not be selected.

Because the definition and criterion for measuring accuracy are derived from the assumption that the test point samples come from a uniformly distributed population with zero mean, the contractor must calculate other statistics. In particular, the mean and the coefficient of skew must be calculated for each sample. Values of the mean of the test points outside of the interval +/- 2 centimeters and/or values of the coefficient of skew outside of the interval +/- 0.5 centimeter may indicate systematic error; the contractor should discuss such values with the FEMA Project Officer (PO).

All data and products associated with contract deliverables must meet or exceed relevant NSSDA and fully comply with the FGDC metadata format standard with the provisions in the contract. The contractor shall use Appendix 7, "Digital Product Delivery Specifications," of these Guidelines as a guide for preparing and submitting deliverables in digital format.

A. Pre-Project Deliverables

Prior to data collection, the contractor must submit:

1. A map (typically, U.S. Geological Survey maps are desirable for this purpose) showing the study area boundaries and flight path, at a medium scale (1:50,000) or small scale (1:100,000);
2. Documentation specifying altitude, airspeed, scan angle, scan rate, LIDAR pulse rates, and other flight and equipment information deemed appropriate; and
3. A chart of areas of high PDOP, or a list showing the time of the beginning and end of high PDOP.

B. Post-Project Deliverables

Following project completion, the contractor must submit:

1. A LIDAR system data report;
2. A flight report;
3. A ground control report;
4. Data processing procedures for selection of postings, and all orthometric values of x, y, and z coordinates for LIDAR returns; and
5. A system calibration report.

The LIDAR system data report must include discussions of: data processing methods used; final LIDAR pulse and scan rates; scan angle; capability for multiple returns from single pulses; accuracy and precision of the LIDAR data acquired; accuracy of the topographic surface products; any other data deemed appropriate; and companion imagery, if any.

The flight report must document mission date, time, flight altitude, airspeed, and other information deemed pertinent. The report must include information about GPS-derived flight tracks, provide a detailed description of final flight line parameters and GPS controls (i.e., benchmarks), and include ground truth and complementary reference data.

The ground control report must include, at a minimum, all pertinent base station information and mission notes, including information on GPS station monument names and stability.

C. Delivery of Digital Data

In addition to the pre- and post-project deliverables described above, the contractor must submit the following:

1. All raw datasets, bare-earth DEM data, and breaklines in separate data files; and
2. Uniformly spaced DEM(s), on ISO 9660 standard CD-ROM media in a format specified in Appendix 7 of these Guidelines.

The contractor must deliver raw datasets and LIDAR system data, including orthometric values for each point, in standard, comma-delimited ASCII files in x, y, and z format. The contractor also must flag raw datasets from sidelap and overlap areas of separate flight lines. Breaklines must be produced, and breakline files must contain a flag record that identifies them as breakline features. The contractor must submit raw datasets in tiles or data models matching file delivery format.

All deliverables must conform to the projection, datum, and coordinate system specified in the contract. File sizes cannot exceed 1 gigabyte, unless otherwise specified by the FEMA PO. Each file must be organized to facilitate data manipulation and processing.

218

10. Metadata

- *http://terra.geo.orst.edu/users/heymano/MetaData.htm*
- *http://www.fgdc.gov/standards/documents/proposals/csdgm_rs_ex_pro.html*
- *http://rat.lic.wisc.edu/metadata/metaprim.htm*
- *http://www.fgdc.gov/metadata/csdgm/*
- *http://www.giconnections.vic.gov.au/content/anzlic/index.htm*

The growing mass of digital data in the GIS marketplace is fuelling a demand for information about geographic data, i.e. geospatial metadata. Recent efforts toward standardization of digital geographic data around the world (particularly the US Spatial Data Transfer Standard (SDTS)) constitute a key element in utilizing existing spatial data. Through metadata clearinghouses, such as the National Geospatial Data Clearinghouse (http://www.fgdc.gov), spatial data users can find what data exists, its quality and condition, and the terms for getting and using it.

10.1 GIS metadata

The major sections of a GIS metadata include the following (the relatively unfamiliar terms are explained, the full standards can be found in the *FGDC* site):

1. **Identification Information -** *basic information about the data set.*
    data set title, area covered, keywords, purpose, abstract, access and use restrictions

2. **Data Quality Information** - *a general assessment of the quality of the data set.*
    horizontal and vertical accuracy assessment, data set completeness and lineage
    > Completeness Report -- information about omissions, selection criteria, generalization, definitions used, and other rules used to derive the data set.
    > Lineage -- information about the events, parameters, and source data which constructed the data set, and information about the responsible parties.

3. **Spatial Data Organization Information** - *the mechanism used to represent spatial information in the data set.*
    raster, vector, or an indirect (e.g. address) link to location

4. **Spatial Reference Information** - *the description of the reference frame for, and the means to encode, coordinates in the data set.*
    lat/long, horizontal and vertical coordinate system, map projection, datum

5. **Entity and Attribute Information** - *details about the information content of the data set, including the entity types, their attributes, and the domains from which attribute values may be assigned.*
    definitions of the attributes of the data set
    > Attribute Domain Values -- the valid values that can be assigned for an attribute.
    > Attribute Value Accuracy Information -- an assessment of the accuracy of the assignment of attribute values.

6. **Distribution Information** - *information about the distributor of and options for obtaining the data set.*
    distributor, file format of data, off-line media types, on-line link to data, fees

Offline Media -- name of the media on which the data set can be received (CD-ROM, 8 mm cartridge tape, etc)

7. **Metadata Reference Information** - *information on the currentness of the metadata information, and the responsible party.*
   who created the metadata and when
   Currentness Reference -- the basis on which the time period of content information is determined.

10.1.1 GIS metadata worked example – Bathymetry of the Gulf of Carpentaria and the Arafura Sea

The following worked example of a GIS metadata was taken from ANZLIC (Australia New Zealand Land Information Council), and is constructed according to their metada guidelines.

Dataset TITLE - Bathymetry of the Gulf of Carpentaria and the Arafura Sea, Edition 1

Dataset CUSTODIAN - Royal Australian Navy Hydrographic Service

Dataset JURISDICTION - Australia

Description ABSTRACT -
The Bathymetry of the Gulf of Carpentaria and the Arafura Sea, Edition 1, is a dataset that contains digital bathymetric information of the Gulf of Carpentaria and the Arafura Sea, Australia, Papua New Guinea, and Indonesia. The digital data in this data base are the latitude, longitude coordinates of the end points of vectors that represent bathymetric contours. Contours were hand drawn using digital systematic-survey soundings from the Royal Australian Navy Hydrographic Office and from a variety of other bathymetric contour maps.
Data accuracy varies. Older data were collected using techniques and navigation that were less accurate than those used at the time this map was prepared. Although nautical charts are periodically updated, some contain soundings that are over 100 years old. Accuracy estimates of the location and depth of soundings cannot be made because information on collection and processing procedures was not available for much of the source data. These data have been published in analogue form as USGS Miscellaneous Investigations Series Map 1-2550.

Description SEARCH WORD(S) -
MARINE
MARINE Geology & Geophysics

Description GEOGRAPHIC EXTENT NAME(S) -

Description GEOGRAPHIC EXTENT POLYGON(S) -
-3.0 130.0, -3.0 149.0, -18.0 149.0, -18.0 130.0, -3.0, 130.0

Data Currency BEGINNING DATE - Not Known

Data Currency ENDING DATE - 31DEC1994

Dataset Status PROGRESS - Complete

Dataset Status MAINTENANCE AND UPDATE FREQUENCY - Irregular

Access STORED DATA FORMAT(S) -
DIGITAL data are stored in separate files for each contour level. Each file has a header that describes the format and the number of records in the file. The files are in ASCII, approximately 5.5 Mbytes, for map scales 1:750,000 to 1:2,500,000

Access AVAILABLE FORMAT TYPE(S)

DIGITAL - ASCII
NONDIGITAL - Plotted Maps

Access ACCESS CONSTRAINT - None

Data Quality LINEAGE
Contours were hand drawn using digital systematic-survey soundings obtained from the Royal Australian Navy (RAN), Hydrographic Office (unpublished data 1992), and using a variety of nautical charts from the following sources: Australian Department of the Navy, Hydrography Branch (1970); [Great Britain] Admiralty (1983,1990); [Great Britain] Hydrographer of the Navy) (1988); Indonesia Angkatan Laut, Djawatan Hidrografi (1961); and the Royal Australian Navy Hydrographic Service (1968. 1974, 1979, 1991). Also contours were scanned from an unpublished map (1987) by T Chase, B. Seakins, J. Young (all of the U.S. Geological Survey), and H. Prasentyo (Marine Geological Institute of Indonesia); they were modified and scanned from published geological studies (Jongsma, 1974; Torgensen and others, 1983); and they were modified and scanned from the Australian National Bathymetric Map series (Australian Division of National Mapping, 1981, 1983, 1984, 1986, 1989, 1990), which was compiled by the division of national Mapping (NATMAP) from the bathymetric data that now comprise the RAN digital data set.
The older data were gathered using less accurate techniques and navigation. Some of the data are over 100 years old. Sea level corrections, when applied, were estimated to about 0.5 metres from values for distant tidal stations. The hand contoured maps were scanned using an auto-vectorising scanner and vector end-points were converted to latitude-longitude values using the USGS software package MAPGEN which is also available as the vector part of the GIS, GRASS.

Data Quality POSITIONAL ACCURACY -
Accuracy estimates of the location and depth of soundings cannot be made because information on collection and processing procedures were not available for much of the source data.

Data Quality ATTRIBUTE ACCURACY - As for positional accuracy

Data Quality LOGICAL CONSISTENCY -
All lines were visually checked at 1:1 000 000 and 1:250 000 scale to verify that no lines crossed, that there were no extraneous line segments and that all lines had the correct contour value. Multiple and dangling lines were edited using in-house software.

Data Quality COMPLETENESS -
All lines and polygons are complete where there was sufficient depth sounding data except for three features where lines are not shown for clarity because of the steepness of the gradient.

Contact Information CONTACT ORGANISATION - Australian Oceanographic Data Center

Contact Information CONTACT POSITION - Data Manager

Contact Information MAIL ADDRESS 1 - Level 2, MHQ Annex

Contact Information MAIL ADDRESS 2 - Wylde St.

Contact Information SUBURB/PLACE/LOCALITY - Potts Point

Contact Information STATE/LOCALITY 2 -NSW

Contact Information COUNTRY - Australia

Contact Information POSTCODE - 2011

Contact Information TELEPHONE - 61 2 563 4812

Contact Information FACSIMILE - 61 2 563 4820

Contact Information ELECTRONIC MAIL ADDRESS - ben@AODC.gov.au

Contact Information CONTACT ORGANISATION - National Geophysical Data Center

Contact Information CONTACT POSITION - Data Manager, NOAA - NGDC E/GC3

Contact Information MAIL ADDRESS 1 - 325 Broadway

Contact Information SUBURB/PLACE/LOCALITY - Boulder

Contact Information STATE/LOCALITY 2 - Colorado

Contact Information COUNTRY - United States

Contact Information POSTCODE - 80303-3328

Contact Information TELEPHONE - 1 303 497 6945

Contact Information FACSIMILE - 1 303 497 6513

Contact Information ELECTRONIC MAIL ADDRESS - gfs@mail.ngdo.noaa.gov

Metadata Date METADATA DATE - 01MAR1996

Additional Metadata ADDITIONAL METADATA -
Maps of the Australian National Bathymetric Map Series for the area referenced.

10.2 Remote Sensing metadata

A special case of spatial data is Remote Sensing (RS) data that require some unique considerations, though the fundamental concepts for arranging their metadata are essentially similar to those of any other geospatial data.

There seems to be a lack in papers dealing with the theoretical and conceptual aspects of remote sensing metadatabases. Most of the materials deal with general geospatial data.

The following text describes the special attributes that Remote Sensing metadata should include, as there are yet no standards.

Proper use of remote sensing data requires an understanding of how those data were obtained. While ground-based data are often compiled from existing data sources without change of form or are obtained by direct *in situ* measurement, deriving geospatial data from the measurements made by remote sensing instruments is often much less direct. To do so may require knowledge of the observing geometry, the instrument behavior, and the processing methods and history. In addition, remote sensing measurements produce large volumes of data, and users typically do not access the entire data set, only selected files or frames.

Information about the viewing geometry and the properties and behavior of the instrument in the FGDC *Metadata Content Standard* is limited to the description of the number of points along the raster axes. The draft ISO metadata standard also includes solar elevation and azimuth angles and the angle of an image to the vertical. However, many user needs a more detailed viewing geometry: satellite orbit or aircraft flight path, platform orientation, and orientation of instruments relative to the platform. While the proposed ISO standard includes a number of items in the section

of spatial data representation describing instrumentation not present in the FGDC *Metadata Content Standard*, the only calibration information is whether camera calibration information is available. More information on the calibration of the instrument, including its dependence on wavelength and time, is usually required. A standard description of such metadata should be defined.

Processing of remote sensing data passes through several stages. The instrument calibration must be applied to the readings communicated by the raw telemetry and the resulting physical measurements located geographically. In some cases, what the instruments measure is not the final product; for example, radiation measurements may be used to infer temperatures. Maps and grids may be generated from data at individual points. Information on the algorithms used to for these steps should accompany the data. In addition, information about the processing itself, such as what stage a given processing represents, or which version of processing is represented, is needed. The FGDC *Metadata Content Standard* allows for this information an entry for lineage, which the draft ISO standard has expanded this item to an entire section on lineage information, but in both cases the content is unspecified free text. These extensions will define the specific items that are needed in remote sensing metadata.

The dataset containing results from a remote sensing mission is large and heterogeneous. Necessary descriptive metadata may not apply to the entire dataset, but only to individual pictures or files. While the FGDC *Metadata Content Standard* has no specific provision for such granularity, the informative Appendix F to the ISO draft provides but does not define granule-specific metadata. These extensions will define the granule-level metadata appropriate to remote sensing.

11. References and links

Following is a *selected* list of useful references for further reading:

11.1 Basic books about Remote Sensing:

- **Remote Sensing : Principles and Interpretation**, by Floyd F. Sabins, 3rd edition (August 1996), W H Freeman & Co.; ISBN: 0716724421
- **Remote Sensing and Image Interpretation**, by Thomas M. Lillesand, Ralph W. Kiefer, 4th edition (October 1999) , John Wiley & Sons; ISBN: 0471255157
- **Remote Sensing : Models and Methods for Image Processing**, by Robert A. Schowengerdt, 2nd edition (July 1997), Academic Pr; ISBN: 0126289816

11.2 Remote Sensing journals:

- **ISPRS Journal of Photogrammetry and Remote Sensing**
  http://www.asprs.org/
  http://www.elsevier.nl/inca/publications/store/5/0/3/3/4/0/
- **Remote sensing of environment**
  http://www.elsevier.com/inca/publications/store/5/0/5/7/3/3/
- **IEEE transactions on geoscience and remote sensing**
  http://ewh.ieee.org/soc/grss/trans.html
- **International journal of remote sensing**
  http://www.tandf.co.uk/journals/default.html
- **Canadian journal of remote sensing**
- **Backscatter – observing the marine environment**
  http://www.amrs.org/magazine.html

11.3 On-line tutorials:

Following are listed some of the tutorials (and university courses) I have used preparing this text. Given is the main subject of each tutorial.

- **The Remote Sensing Tutorial – NASA**
  http://priede.bf.lu.lv/GIS/Descriptions/Remote_Sensing/An_Online_Handbook/TofC/toc1.shtml
- **Canada Center for Remote Sensing tutorial**
  http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/indexe.html
- **Remote Sensing Core Curriculum, volume 3, Introductory Digital Image Processing**
  http://www.cla.sc.edu/geog/rslab/rsccnew/rscc-no-frames.html
- **Centre National d'Etudes Spatiales** - **The science of Remote Sensing**
  http://www-projet.cnes.fr/ceos/cdrom-97/ceos1/science/science.htm
- **University of Tennessee - Astronomy course 162 - Radiation laws**
  http://csep10.phys.utk.edu/astr162/lect/light/radiation.html
- **Washington university in St Louis – EPSc course 407 – Remote Sensing**
  http://wully.wustl.edu/epsc407/outline.html
- **USGS – Spectroscopy and imaging spectroscopy**
  http://speclab.cr.usgs.gov/
- **USGS - Surface Reflectance Calibration of Terrestrial Imaging Spectroscopy Data**
  http://speclab.cr.usgs.gov/PAPERS.calibration.tutorial/calibntA.html.bak3
- **Iowa State University - Remote Sensing in Precision Agriculture**
  http://www.amesremote.com/Contents.htm
- **Synoptics and the Dutch survey department - Airborne Digital Video**
  http://www.minvenw.nl/projects/airborn/vid_appl.htm
- **NOAA - Airborne Laser Beach Mapping**
  http://www.csc.noaa.gov/crs/ALACE/techcd/htm/laser.htm
- **University of Texas - Radar altimetry**
  http://www.ae.utexas.edu/courses/ase389/sensors/alt/alt.html
- **The NCGIA Core Curriculum in GIScience**
  http://www.ncgia.ucsb.edu/giscc/cc_outline.html
- **GIS lessons (each html page is a lesson) – by Michael Frank Goodchild**
  http://www.ncgia.ucsb.edu/~good/176b/

11.4 Remote Sensing softwares:

Among the leading Remote Sensing softwares, the following names can be given:
ENVI   http://www.envi-sw.com/
PCI   http://www.pcigeomatics.com/
ERDAS   http://www.erdas.com/

There are also some low-cost Remote Sensing-GIS raster softwares:
IDRISI   http://www.clarklabs.org/
ILWIS   http://www.itc.nl/ilwis/

And, there are some Remote Sensing-GIS raster freewares (downloadable):
GRASS   http://www.baylor.edu/~grass/
SPRING   http://sputnik.dpi.inpe.br/spring/english/home.html
PIT
http://priede.bf.lu.lv/GIS/Descriptions/Remote_Sensing/An_Online_Handbook/Appendix/nicktutor_A-5.shtml

11.5 Other Remote Sensing links:

- **SPOT IMAGE**
  http://www.spotimage.fr/
- **Airborne Laser Mapping**
  http://www.airbornelasermapping.com/ALMNews.html
- **Canada Center for Remote Sensing**
  http://www.ccrs.nrcan.gc.ca
- **Landsat 7**
  http://landsat.gsfc.nasa.gov/
- **Center for the Study of Earth from Space**
  http://cires.colorado.edu/cses/
- **Table of Fundamental Physical Constants**
  http://www.fpl.uni-stuttgart.de/fpl/physchem/const.html
- **European Space Agency**
  http://earth.esa.int/
- **TOPEX/Poseidon**
  http://topex-www.jpl.nasa.gov/
- **IKONOS**
  http://www.spaceimaging.com/
- **RADARSAT**
  http://www.rsi.ca/