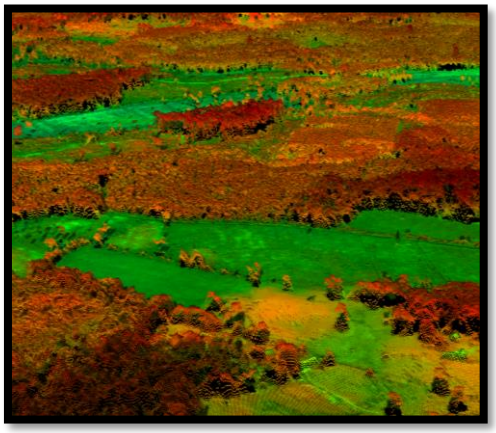
Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Institutionen för skoglig resurshushållning

LJUNGBERGS FONDEN

FOREST REMOTE SENSING



Version: 1.1 (30 augusti 2017)

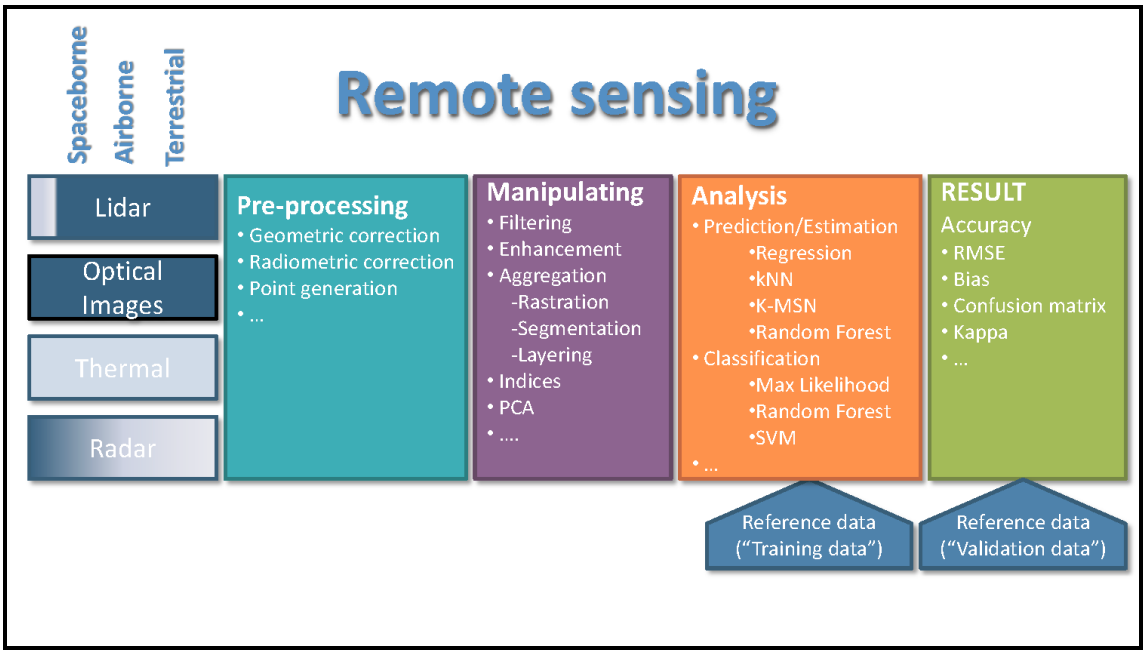
PREFACE

This compendium has been generously financed by the Erik Johan Ljungbergs Educational fund. It has been part of a larger effort to introduce more technical competence into the Swedish University of Agricultural Sciences (SLU) Forestry Master's program. The contributors to the compendium come from the Department of Forest Resource Management at SLU in Umeå. The compendium consists of three main parts: Forest Inventory, Remote Sensing of Forests, and Forest Planning. It is intended to be used as literature in the Department's courses, primarily at the basic undergraduate level, but is also freely available on the internet and can be used by general readers interested in the topics. The compendium should give the reader an understanding of basic concepts in the three areas individually, and as a whole, and give insight into the process of creating forest information that could be used in assessment and planning of forest resources, including input to Heureka or GIS.

In the remote sensing part of the compendium, we aim to answer the following questions: How can forest data be effectively yet accurately collected? What data do forest industry or government authorities need? What are advantages or pitfalls of different data and methods? Sometimes a sample of the landscape via a field-based inventory is all that is needed. In other cases, full area coverage data are needed. Remote sensing (RS) can provide full area coverage on, for example, tree height, location of harvested stands, stand boundaries, forest health, tree species, and forest structure. The current trend in RS is 3D analysis with laser scanning, digital photogrammetry, or radar data, while vegetation types can be seen using optical satellite data. To process RS data, inventory data are needed to link ground data with what can be seen from the air. On the other hand, samples can be taken from RS data to produce statistics. Advances in sensor-based field inventory are being made with terrestrial laser scanning. What do these data deliver compared to field-based inventory?

The remote sensing compendium will first address background concepts and history of remote sensing, as well as remote sensing's relationship to forest inventory and forest planning in Chapter 1. The physics of remote sensing is covered in Chapter 2. Basic information regarding sensors, platforms and digital data are discussed in Chapter 3. Chapter 4 goes into more detail about electro-optical sensors, and Chapter 5 discusses traditional aerial photography interpretation, as well as digital manipulation into three-dimensional point data. Chapter 6 takes up airborne and terrestrial laser scanning, while Chapter 7 addresses radar data. Chapter 8 focusses on issues of combining reference data and remote sensing data, and desirable properties of the reference data. Chapter 9 gives details on the methods used to combine reference data and remote sensing data in order to process the data and produce new information. Chapter 10 discusses the methods and importance of accuracy assessment of the information produced. Chapter 11 gives a short introduction to programming used currently in remote sensing.

The remote sensing part of the compendium also will use a framework we see as helpful in referring to remote sensing image processing. That framework is given below and will be referred to in the text.



We hope that this compendium informs the reader about the past and current knowledge, as well as the future potential, of Forest Remote Sensing.

Umeå, Sweden
December 12, 2016

Heather Reese
Senior Lecturer in Forest Remote Sensing, SLU

Håkan Olsson
Professor in Forest Remote Sensing, SLU

Table of Contents

1. INTRODUCTION TO REMOTE SENSING

Remote sensing. The term used to describe the act of observing a phenomenon from a distance. Remote sensing has often dealt with the imaging of areas on the Earth's surface taken by sensors placed on airplanes or satellites; however, it is becoming all the more common with sensors placed on moving vehicles or on ground-based platforms.

Electromagnetic spectrum. The entire range of wavelengths or frequencies of electromagnetic radiation extending from gamma rays to the longest radio waves and including visible light.

Sensor. The instrument which detects and stores observations.

Platform. The carrier of the sensor. This may be, for example, a stative, a small airplane, or a satellite.

1.1. What is remote sensing?

Remote sensing is the name of the subject which deals with observing objects remotely. Remote sensing is used in many applications including forestry, weather, nature conservation, cartography, just to name a few. Remote sensing is often performed via a sensor which observes and records properties of the electromagnetic spectrum. The sensor is placed on a platform, which can be, for example, a satellite or an airplane. The information recorded from the sensor is then analyzed in order to produce information usable as maps or within GIS or planning systems.

There are several definitions of remote sensing found in different textbooks, each with a little different wording. The following two definitions are similar and rather good:

> *"**Remote sensing** is the science and art of obtaining information about an object, area, or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area or phenomenon under investigation[1]."*

This in fact leaves a rather wide-open definition. It can include many different aspects of remote sensing such as walking through a forest looking with your eyes, taking photographs with your pocket camera, or acquiring data with laser sensors flown in a plane or from a satellite flying hundreds of kilometers above Earth in space.

[1] Lillesand,T., R. Kiefer and J. Chipman, 2008. Remote sensing and image interpretation. Wiley Press, 738 p.

In Swedish, the word for remote sensing is “fjärranalys”, which can be broken down into two English words, “fjärr = remote” and “analys = analysis”. In Lillesand et al.’s definition we find the important word “analysis”. While some definitions define remote sensing as an act which collects data, the definition can be widened to include the act of analyzing data in order to obtain useful information.

As a rule, this analysis is done by computer-based methods. For example, in a first step special software can be used to calculate the characteristics collected from remote sensing data, such as the ratio between the values of two wavelengths recorded within each pixel (e.g., red value/near-infrared value), or the difference between the highest and lowest laser return within an area. Within forest remote sensing, it is usual that the information from the remote sensing data is translated into forest measurements with the support of field reference data, so that certain combinations of colors are translated to forest species classes, or the distance between the laser returns translated, for example, into basal mean height. Often, conventional statistical software, such as Minitab or R can be used for calibration of the remote sensing data against field data.

1.2. Why use remote sensing?

Use of remote sensing data can sometimes require technical knowledge and can incur some costs, so there should be reasonable arguments for the use of remote sensing data. Some of the benefits of remote sensing are that they can provide:

- A cost-effective way to get information
- Large area coverage, and coverage of remote areas
- Better and new information can be obtained as compared to traditional data collection methods
- Better spatial resolution
- Objective analysis of digital data is possible
- Repeated images

Cost efficiency is often an argument for use of remotely sensed data, as it can mean less field work than otherwise required and therefore less costs. However, note that field work IS still often needed for work with remote sensing (it’s not obsolete). Cost efficiency is likely the most important reason to use remote sensing in forestry primarily because the need for manual measurements in the field can be reduced, even if the field references are also needed for making estimates based on remote sensing data. A prerequisite for the acceptance of remote sensing in practical forestry has proved to be that it has been cost effective and provide at least as good data as provided by traditional inventory methods -- this is currently the case for the estimation of several forest variables using airborne laser scanner data.

Large area coverage, or "Synoptic coverage" refers to getting a "bird's eye view" over an area. A dictionary definition of "synoptic" is "giving a general view of a whole or displaying conditions as they exist simultaneously over a broad area". Depending on the platform used to carry the sensor, the users' needs and also the sensor's capabilities, remote sensing data may be used to cover very large areas, or be concentrated on smaller areas. *Remote areas* can also be covered more easily than when using traditional methods, for example, when the sensor is mounted on a plane or satellite, images can be obtained regardless of whether there are roads or footpaths to travel to that area.

Better and new information from remote sensing data can, for example, be gained from infrared bands that provide a great deal of information about vegetation that our eyes do no see. In addition, laser data give us highly accurate height measurements – in many cases more accurate than can be manually measured.

Objective analysis of digital data is possible. Data are collected without the initial interpretation by subjective human judgment, and can provide a raw, objective source of digital data. Analyses such as vegetation classification or estimation of wood volume can then be applied to the digital data in an objective manner. It is also an advantage that remote sensing estimates can be impartial and independent of the person who is carrying out the measurements.

Most remote sensing methods provide data with *better spatial resolution* than traditional forest maps, which can give, for example, indications that only parts of a stand need to be thinned, or are damaged. Remote sensing data often provide what is called "wall-to-wall" coverage, or "spatially explicit" coverage.

Assuming that the proposed collection method is cost effective, it is of course also beneficial if *you can get even better data than with traditional methods*. In the case of estimation of tree height or wood volume with laser scanner data, one can also expect better estimation accuracy than with subjective manual methods.

Most remote sensing satellites systematically cover all the Earth with a repetitive time interval. Currently, new small satellite systems can cover each spot on Earth every day (e.g., Planet Labs), while Sentinel-2 satellite covers the same spot on Earth every 5 days at the Equator (ever 3 days in Sweden). Satellite-based remote sensing is particularly important for nationwide or international applications. As long as the same sensor and same settings are used, images over longer time periods can be used to look at a landscape over time, which can be considered as another dimension (the temporal dimension).

A regular supply of digital data that can be digitally processed also makes it quite easy to automatically detect areas that have changed in an unusual way, e.g., areas of damaged forest. However it is good to keep in mind that

additional information (e.g., field or reference data) is required to determine the cause and exact nature of detected changes.

1.3. History

In the late 1800s people attempted taking photographs from hot air balloons. Aerial photography technology was then developed rapidly during the First World War (WWI). After WWI aerial photographs began to be used for forest mapping in e.g., Canada in the 1920s. In Sweden aerial photography was tested for the first time in 1930 and even then, the application for forest use was central. However, it would take until after World War II before aerial photography use for forestry would really take off. In the 1950s, a government committee was formed that would promote the use of aerial photographs for forestry. Even today, the manual interpretation of aerial photographs is the dominant use of remote sensing in forestry. Mainly orthophotos are used, but the interpretation can also be made with different types of stereo-viewing instruments.

In the 1980s, Sweden launched the so-called LMV-method of forest variable estimation. It means that the forest is divided into compartments, the tree height is measured, and the estimate of tree density and species mix is done manually on a photogrammetric workstation. From these data, the timber volume is then estimated. The method was used among other things as SCA and Holmen redid the division of their holdings in the 1990s. Since the early 2000s, aerial photography has been acquired via digital cameras, which among other things has resulted in increased access to aerial photographs in color infrared.

Since the mid-00s the use of drones (also called UAVs or UAS) for aerial photography has increased. Contributing to the development of UAV remote sensing is a series of leaps in technology: the development of improved battery technology that permits the use of electric motors, as well as small navigation systems and digital cameras.

The first military remote sensing satellites came in 1960, and in 1972 Landsat 1 came which was the first civil earth resource satellite. Landsat 1 had pixels with an 80 x 80 m size. This spatial resolution was improved by the Landsat TM sensor which came in 1982 and had 30 x 30 m pixels, and SPOT 1 which was launched in 1986 and had 10 x 10 m pixels. These older satellite images are archived, and this archive can be very useful for the study of landscape development over time. During the early 2000's, Landsat data became available free of cost, which has increased their use.

In the 1990s, satellite imagery with 1 x 1m pixels became available, however, these satellites were frequently operated by commercial actors (the data cost money), and have covered generally smaller areas than the Earth resource satellites such as Landsat.

Satellites are taking frequent pictures of the Earth, but a limitation of photographic-like images (i.e., color or color infrared imaging) from space is that the images are often cloudy. There is therefore great interest in radar satellites, which can record data through clouds. The testing of different types of radar data for forest mapping has been going on since the 1990s; while the potential is great, the practical use of this technology is still limited.

In the mid-1990s airborne laser scanning also developed. This technique involves laser beams sent out from an instrument (placed within an airborne platform such as a plane or helicopter), where the distance is measured from the source of the laser beam down to the ground or canopy (or whatever object that laser beam is reflected from). The result is a three-dimensional (3D) point cloud. The measurement is very accurate and an important factor for the airborne laser technology has been developed is that although the aircraft's position and movement in the air is calculated using GPS and inertial navigation. Measuring forest with airborne laser was first tested in Leningrad in the late 1970s and in the United States and Canada in the mid-1980s. They used the instrument to just measure a profile from the aircraft (i.e, a profiling laser). In Sweden we were the first in 1991 to test an experimental system with a scanning laser, but it was not until commercial systems with integrated positioning came in the mid-1990s as the development of laser data use for forestry (and many other applications) took off.

Laser scanning first began to be used as an operational method for forest inventory in Norway in 2002. The method then took off with the newly introduced so-called “area-based approach”, which means that field measured reference plots are used to assign forest data to a combination of metrics calculated from the laser point cloud. In Sweden the first attempt was made on a larger scale by the Forest Agency and the SLU (10,000 ha for re-parcelling in Dalarna) in 2003. This was followed by a period when several forest companies tested the laser scanning in areas of approximately 10,000 ha. When the National Land Survey laser scanning started in 2009 forestry companies then had new access to cheap laser data, which led to several forestry companies ordering laser data (or the finished product of estimates) for their entire holding. First was Bergvik, who in 2011 decided to make airborne laser data estimates for their entire forest holding.

1.4. Current role of remote sensing to provide forest information

Aerial photographs, primarily digital orthophotos, are the main source of information for forest data. Orthophotos can be used for manual classification of compartments and as a background for a variety of digital map products. To some extent, manual measurement and interpretation of stereo aerial photographs in digital photogrammetric workstation is still performed according to the so-called LMV-method. Forest data estimation using this method, however, requires great skill of the photo-interpreter if it is to replace field work. An alternative is to use aerial photos or orthophotos

for a preliminary evaluation of the area before field work.

Digital stereo aerial photographs can also be processed automatically by the computer to create a 3D point cloud similar to that obtained by airborne laser data. The Swedish National Land Survey starts in 2016 to produce such 3D point clouds from their aerial photography product which they will call a "surface model from aerial photographs." They are also referred to as 3D data from image matching. These data can be used to calculate a raster with tree canopy heights. They can also be used for forestry estimates in the same way as 3D point cloud from laser data. However, the data obtained from aerial photography does not give as good information about forest density as laser data does, and therefore, estimates of timber volume will be worse from 3D image-matching data.

Drones (Unmanned Aerial Vehicles or UAVs) are being put into use in the preparation of forest management plans for individual properties. It is so far mainly orthophotos created from mosaiking drone flyovers that are being used.

The availability of frequent and free satellite images has increased. In 2014 Landsat 8 was launched and in 2015 the European Sentinel 2A satellite with 10 m pixels was launched. Since about 2000 Skogsstyrelsen (The Swedish Forest Agency) has used satellite images to annually map all new clearcuts. The Swedish Forest Agency also uses satellite images to estimate the forest cleaning/thinning needed.

For the years 2000, 2005 and 2010 SLU has made national forest maps with 25 x 25 m grid cells by joint processing of satellite images and reference plots from the National Forest Inventory (NFI). These forest maps have been widely used among researchers and authorities, but estimates of forest data based on the use of "2D" optical satellite images from a time are generally not accurate enough to be used in forestry planning. Forest data for the year of 2015 will be created by SLU from airborne laser scanning and Sentinel-2 satellite data.[2]

Lantmäteriet's (The National Mapping Agency) laser scanning throughout Sweden that started in 2009 will in practice be finished for the forested part of the country in 2016. What remains is a few areas in the mountains. This laser data can be purchased for a marginal cost by commercial actors and is free to governmental authorities. The national laser scanning has thus been a catalyst for the use of laser data in Swedish forestry. Bergvik and Holmen have ordered estimates for their entire holdings and SCA has ordered laser estimates based on National Land Survey data for their forests in Norrland. Together, the Forest Agency and SLU worked in the context of a government commission to make forest estimates for the country, based on Lantmäteriets laser scanner data and reference plots from the National Forest Inventory. Together with other consulting companies, the Forest Agency has also

[2] https://www.slu.se/centrumbildningar-och-projekt/riksskogstaxeringen/statistik-om-skog/slu-skogskarta/

developed raster layers from laser data showing soil wetness maps, as well as a raster layer with tree heights at a 2 x 2 m grid cell size. These map layers can be viewed and downloaded from the Forestry Board's website. They have been widely used, both in the forestry sector and by other actors, including banks and insurance companies.

Currently there is much ongoing research and testing on a range of technologies with great potential to streamline the future measurement of forest. An example of this is that laser scanners can also be placed on a tripod on the ground (the name of this technology is often shortened to TLS for Terrestrial Laser Scanning), or mounted in a backpack or on a vehicle (often abbreviated to MLS for Mobile Laser Scanning), or even carried in the hand (also MLS). With TLS and MLS detailed information can be obtained about tree stem position and shape, which, inter alia, can be used as reference data to create estimates from the airborne data.

1.5. Products from remote sensing

Remote sensing provides raw data which can be used simply as a visual background image, such as in Google Earth. However, remote sensing data are digital data, and can be analysed and manipulated and therefore converted into map data that give users the information they need.

We can categorize map data as being *thematic* or *continuous*. Thematic maps have *discrete* classes, such as the Swedish Land Cover map with the thematic classes of "forest", "mire", "water", for example. Continuous variable maps consist of a range of values for a single phenomenon. An example of this is a map of timber volume, with values ranging from 0 up to the maximum value. These two map output types are shown in Fig. 1.1.

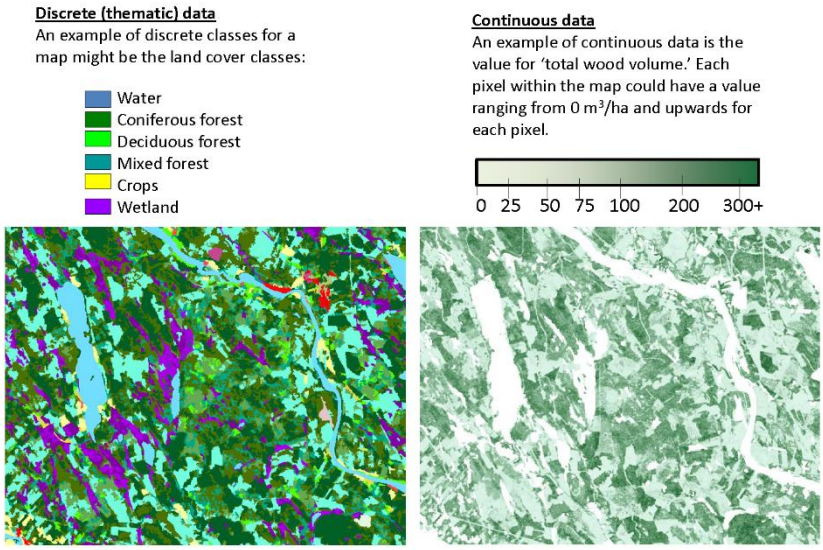


Fig 1.1. Example of a discrete or thematic value map on left (e.g., land cover classes), and a continuous value map on right (e.g., total wood volume).

Another common product from remote sensing data is a change map, which can show the differences between two or more dates of remote sensing data.

Another product from remote sensing data is as input to visualization; the remote sensing data input may be as raw data, such as from Terrestrial Laser Scanning, or it may be map products that form the baseline for visualization.

1.6. The relationship between remote sensing and GIS, forest inventory and forest planning

The field of remote sensing has a relationship with other fields, such as Geographical Information Systems/Technology (GIS/GIT), forest inventory, and forest planning. It is worth clarifying the definition and roll of remote sensing in relation to these other fields.

A GIS consists of four components, namely data acquisition, data storage, data analysis, and map production[3]. Remote sensing fills the function of "data acquisition" in a GIS; the remote sensing data input may be raw data (e.g., images) or processed data (e.g., map data derived from remote sensing).

The field of forest inventory is about techniques and methods for measuring and estimation of forest resources. "Traditional" forest inventory has been done with manual methods, but remote sensing plays an increasing role for providing both wall-to-wall data, as a tool for in-field measurements, as well as remote sensing data acting as ancillary data for statistical estimations of forest resources. Remote sensing and forest inventory have an intertwined relationship, which can be described as having two main interactions:

- Remote sensing data acquisition for forest inventory purposes, and
- Forest inventory data use as reference data to help interpret remote sensing data (i.e., training data) or to use in validation of the map or product from remote sensing (i.e., validation data).

Note that the subject area of remote sensing differs from the subject areas of inventory, planning, and GIS. Remote sensing data are used within the process of forest inventory, and they are analysed or displayed within a GIS. However, the subject of remote sensing includes not just measuring (i.e., inventory), but also knowledge of how to acquire and process the raw remotely sensed in a correct way. This may require background knowledge in physics, statistics, photogrammetry, programming, and certainly geography. The subject of remote sensing also involves knowing which remote sensing data source is best suited for the purpose (i.e., strengths and

[3] Harrie, L. 2013. Geografisk Informations Behandling: Teori, metoder, och tillämpningar. Studentlitteratur AB. 326 s.

limitations), and how to perform and present an accuracy assessment of the map products from remote sensing data.

1.7. Remote sensing for global to individual tree applications

The choice of remote sensing data source is dependent upon the aim and goal of the project at hand, the availability of remote sensing data, and the cost, among other factors. For example, if we consider remote sensing data acquisition for forest inventory purposes, one parameter that we need to consider is "at what scale we wish to produce information?" This can be at the landscape scale and may range downwards to the individual tree scale.

The large area, landscape scale coverage that can be provided by remote sensing (synoptic coverage) makes it a useful tool. Aerial photographs have been used between the 1930's to present for delineating forest stands and measurement of tree height. Satellite data have also played a role in providing forest information over estates, or whole countries. With the current innovation of airborne LiDAR, which measures tree heights and forest density with high accuracy, the use of remote sensing for forest inventory over large areas is increasing rapidly. The availability of data to both private forest owners and larger companies means that they are used by both.

Another consideration to take when deciding on the remote sensing data source is, of course, cost. Some sensors may provide high spatial resolution (e.g., small pixel size), but can cost a great deal of money per image. The user needs to know the advantages and disadvantages of the different sensors available so that costs are not higher than necessary.

At the individual tree scale, remote sensing technologies are providing data from both the air and from the ground. The ground-based remote sensing includes terrestrial laser scanning, and ground-based photogrammetry. These sensors may be placed on platform which is a stative, or may be mobile (e.g., placed on a car or hand-held). The level of spatial detail in the remote sensing data, the accuracy of the map products, and cost-effectiveness have had an influence on whether the data will be useful for forestry applications.

1.8. The growth of remote sensing for forestry purposes

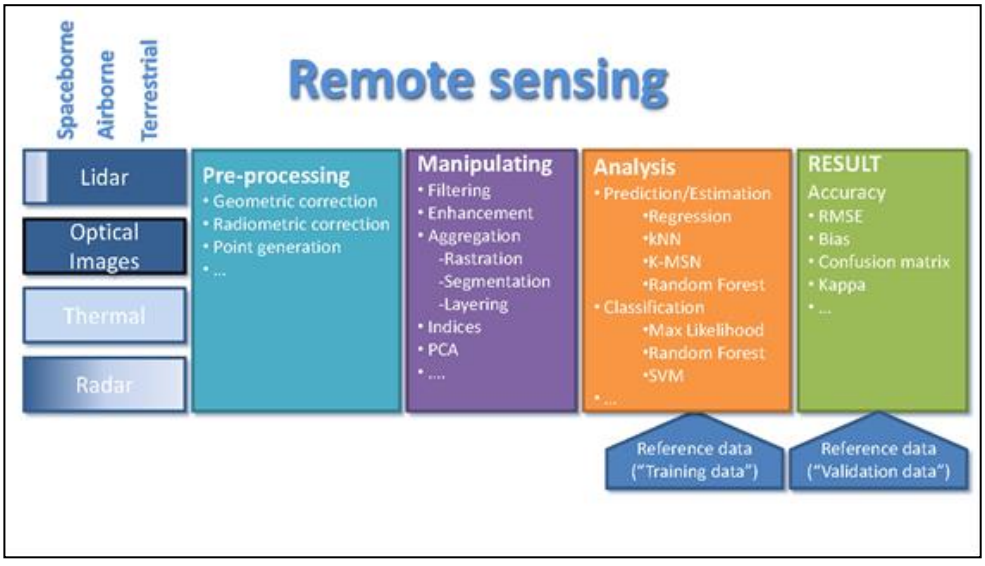
Developments in the subject and use of remote sensing is growing exponentially due to:

- the ability to acquire highly accurate and useful 3D data (from laser, radar, and digital photogrammetry) ;
- access to free open-source remote sensing and geographic data (with the ability to use free, open-source software)
- economic cost-effectiveness of remote sensing data in businesses (e.g., forestry);
- increased computing power; and
- increased knowledge about remote sensing and GIS by the public with access to sites such as Google Earth, and personal use of GPS.

The geospatial data era is well underway – for researchers, for governmental agencies, for businesses, and the private citizen -- at the time of writing this compendium.

Framework

There are basic components involved in acquisition, processing and product assessment of remotely sensed data. If we simplify this in a diagram, it could be represented as shown below. This compendium will refer back to this diagram, when discussing the different elements and processes.



Self study questions

1. What is a good definition of remote sensing? Can you add to the definitions given in this chapter? Do you think the definition of remote sensing will change over time?

2. In which situations might it be good to use remote sensing?

Further reading

Harrie, L. (red.), 2013. Geografisk Informations Behandling: Teori, metoder, och tillämpningar. Studentlitteratur AB. 326 s.

Lillesand, T., R. Kiefer and J. Chipman, 2008. Remote sensing and image interpretation. Wiley Press, 738 p.

2. PHYSICAL BASIS OF REMOTE SENSING

Electromagnetic spectrum. Electromagnetic radiation is magnetic and electric fields that travels in the form of waves, with the speed of light. The spectrum stretches from gamma rays with 10^{-11} m wavelengths over visible light with 400 – 700 nm waves, to radio waves that are more than 1 m long.

Irradiance (E). Radiant flux from all directions, per unit area (W/m^2).

Radiance (L). Radiant flux per unit area of surface per unit solid angle (W/m^2 Sr).

Reflectance (ρ). The ratio between incoming and emitted flux, observe however that the flux might be emitted differently in different directions, of which only one, or a few, directions are measured as radiance registered by the sensor. Reflectance is always between 0 – 1.

Reflectance factor (R). The ratio between the actually measured radiance from a target at a given direction and illumination situation and the radiance that is measured at a reference target that reflects fully and equally in all direction. The reflectance factor is related to reflectance, but easier to measure, and might vary between 0 – ∞.

The definition of remote sensing can be quite broad, as it is simply based on the act of observing objects from far away. In theory, one could "remotely sense" objects based on several different techniques, for example, sound waves (e.g., sonar). Here we concentrate primarily on remote sensing which makes use of recording the properties of *electromagnetic radiation* being *reflected* from objects on the Earth. This chapter primarily addresses the physical basis for registration of optical images from airplanes or satellites, including an introduction to the basic properties of light and other electromagnetic radiation, and in particular how sunlight interacts with the atmosphere and the target. The basic principles are also applicable to remote sensing using laser and radar that also use electromagnetic energy. However, laser and radar are active techniques, where the sensor sends out its own energy and records the response, rather than recording reflectance from the sun. Chapters 6 and 7 address the specifics of laser and radar, respectively.

2.1. The electromagnetic spectrum

Electromagnetic radiation can be understood both with wave theory and with quantum theory. In quantum theory, the energy of a *photon* or a *quantum* is given in Equation 2.1:

$$Q = h * v \qquad (2.1)$$

where
Q = the energy of a quantum in Joules (J)
h = Planck's constant $6.626 * 10^{-34}$ J sec
v = frequency of the energy

One obvious source of electromagnetic radiation is the sun, but there are many other sources as well. All objects with a temperature above absolute zero emit energy, and the energy emitted is in proportion to its temperature, which will be described later.

Electromagnetic radiation has both electric and magnetic field components (seen in Figure 1.2 in Lillesand et al.). The radiation travels through a vacuum at the velocity of light, *c* (where *c* is a constant equal to 299,792,458 m/s or for simplicity 3×10^8 m/sec).

The time point at which there was a breakthrough in understanding light's electromagnetic properties came in 1864, when James Clerk Maxwell published his paper "A dynamical theory of the electromagnetic field". Maxwell derived a wave form of the electric and magnetic equations, thus uncovering the wave-like nature of electric and magnetic fields, and their symmetry. Because the speed of the electromagnetic waves predicted by the wave equation coincided with the previously determined and measured speed of light, Maxwell concluded that light itself was an electromagnetic wave. It followed thereafter that

$$c = v * \lambda \qquad (2.2)$$

where c = speed of light, ν is the energy wave frequency, and λ is the energy wavelength. *Wavelength* is the distance between successive wave peaks and *frequency* is the number of cycles passing a fixed point in a given period of time (see Fig 1.2 in Lillesand et al.).

By combining equations (2.1) and (2.2) we get:

$$Q = hc / \lambda \qquad (2.3)$$

Therefore, what is useful to remember is that *the energy of electromagnetic radiation is inversely proportional to the frequency of the radiation*, or in other words, short wave radiation, for example blue light, will generally carry more energy than longer waves.

Note that since the wavelength and the frequency are inversely proportional to each other, only one of them needs to be stated to describe the wave. In some cases, frequency is used to describe the wave (as in radar remote sensing) and in other applications (as in optical remote sensing), the wavelength is most often used. For convenience, the wavelengths can be ordered and arranged along the *electromagnetic spectrum* (Figures 2.1, 2.2).

Significant ranges of wavelengths have been categorized by giving a name to that range, for example, the ultraviolet spectrum, visible spectrum, infrared spectrum, microwaves, and radio, to name a few. Some spectral ranges of importance are named Table 2.1.

*Table 2.1 Examples of electromagnetic wavelength bands used in remote sensing**

Name	Approximate	Actual	Use in remote sensing
Visible spectrum	0.4 – 0.7 µm	0.380 - 0.750 µm	Is the name for the full visible light spectrum. Can be used as a so-called panchromatic band, often of higher spatial resolutions.
- Blue	0.4 – 0.5 µm	0.450 – 0.495 µm	Used together with green and red wavelengths to get a natural color images. Blue wavelength registration is however most sensitive to haze and atmospheric influences. It can be used for atmospheric correction purposes. But in analyses it is often omitted, since it is highly correlated with information in the green wavelength band.
- Green	0.5 – 0.6 µm	0.495 – 0.570 µm	Used in most optical sensors + LiDAR for water depth
- Red	0.6 – 0.7 µm	0.620 – 0.750 µm	Used in most optical sensors. Photosynthesis results in use and absorption of red light.
Near Infrared light	0.7 - 1.3¤ µm	0.78 – 1.3¤ µm	Important band for most optical sensors + LiDAR.
Shortwave Infrared light†	1.3¤ – 3.0 µm	1.3¤ - 3.0 µm	Use in some optical sensors + Multispectral LiDARs. Short-wave is sensitive to shadows from trees in a forest.
Thermal Infrared	3.0 – 14.0 µm	3.0 – 14.0 µm	Sensed as heat, sometimes used in a separate sensor. Often on an airplane or satellite platform, such as Landsat-8.
Microwave	1 mm – 1 m	1 mm – 1 m	Used in radar and also for so-called passive microwave radiometry.

* *It is useful to know that there are differing opinions on exactly where to draw boundaries and names for different ranges of wavelengths. See for example the "International Commission on Illumination" or different ISO standards or different textbooks.*

¤ *Note that some texts use 1.4 µm as an upper limit for NIR/lower limit for SWIR.*

† *Note that "Shortwave Infrared" is sometimes also referred to as "Mid-Infrared", depending on the field of study.*

There are other categories (listed below) that are also of significance in remote sensing. Note that these are not necessarily mutually exclusive, and that the categories are often defined for convenience and relevance to certain applications. Also note that there is sometimes confusion about use of the name of these categories, and you may see mistakes on the internet and other places.

- *optical spectrum* (0.30 – 15.0 µm) are the wavelengths that can be reflected and refracted with lenses and mirrors.
- *reflective spectrum* (0.38 – 3.0 µm) are the wavelengths whose reflectance can be measured, and are commonly used in "optical" remote sensing.
- *emissive spectrum* or *thermal spectrum* (3.0 – 100.0 µm), which are wavelengths which emit heat, and are measured by thermal remote sensing instruments.

The Electromagnetic spectrum is shown in Figure 2.1. Table 2.1 gives wavelength measurements in micrometers (µm), while Figure 2.1give the wavelengths in nanometers (nm). One µm is 1000 nm.

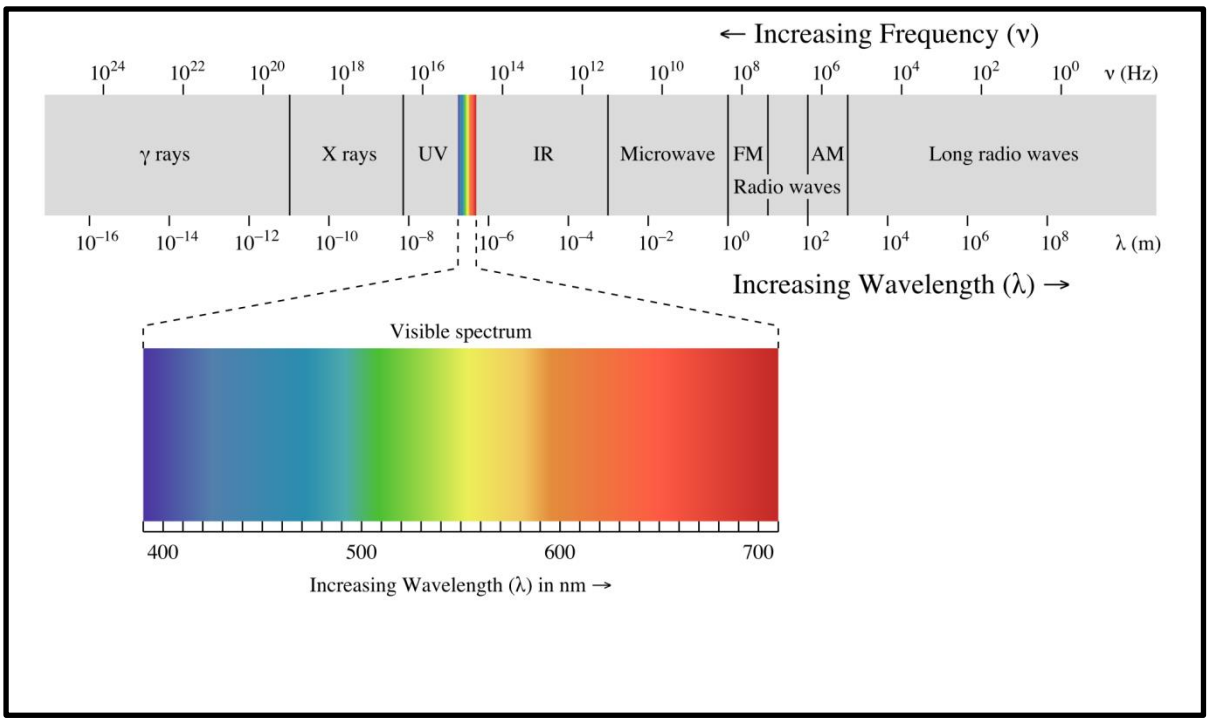


Figure 2.1 – The electromagnetic spectrum (image from Wikipedia).

We said that all objects *emit* some amount of radiation. How much? Why? The amount of energy an object emits is partly a function of the surface temperature, which is expressed by the *Stefan-Boltzmann law* which states that *blackbody* radiators (which is a hypothetical "ideal" radiator - an object which absorbs all and re-emits all of the energy incident upon it) have an amount of emitted radiation or *radiant exitance* (M_e) which can be calculated as:

$$M_e = \sigma T^4 \qquad (2.4)$$

where
M_e = total radiant exitance (i.e., "emitted radiation") from the surface in Watts (W)/m^{-2}

σ = the Stefan-Boltzmann constant (5.6697×10^{-8} W $m^{-2}K^{-4}$)
T = the temperature of the object's surface in degrees Kelvin.

Since any material that has a temperature above absolute 0 will emit energy, the Stefan-Boltzman Law shows that it will emit energy in proportion to its temperature raised to a power of 4. We will first perceive this electromagnetic energy as heat, and then if the temperature rises, also as light. For example, a piece of iron that is being heated will be red at a certain temperature and then become even white and maybe bluish. Why blue? There is a law for this phenomenon called *Wien's displacement law*. It says that the dominating wavelength for the energy emitted from an object will be shorter and shorter as the temperature of the object rises, so that:

$$\lambda_m = A / T \qquad (2.5)$$

where
λ_m = the wavelength of maximum spectral radiant exitance in μm
A = 2897.8 (or rounded up to 2898) μm K (a constant)
T = temperature in degrees K

6000˚K, which is the temperature of the surface of the sun, will result in a radiation maximum in the wavelength region of 0.4 – 0.7 μm, which is exactly the wavelengths the human eye can register! We also see that the normal temperatures of the Earth (around 300° K) will result in a radiation maximum in the wavelength region around 10 μm and this is in fact wavelengths that we perceive as thermal heat.

2.2. Interaction of energy with the atmosphere and Earth

To understand just what the remote sensor is recording, it is necessary to understand the interactions between light, the atmosphere, and the objects being observed. Figure 2.2 below illustrates the topics to be discussed in the following sections.

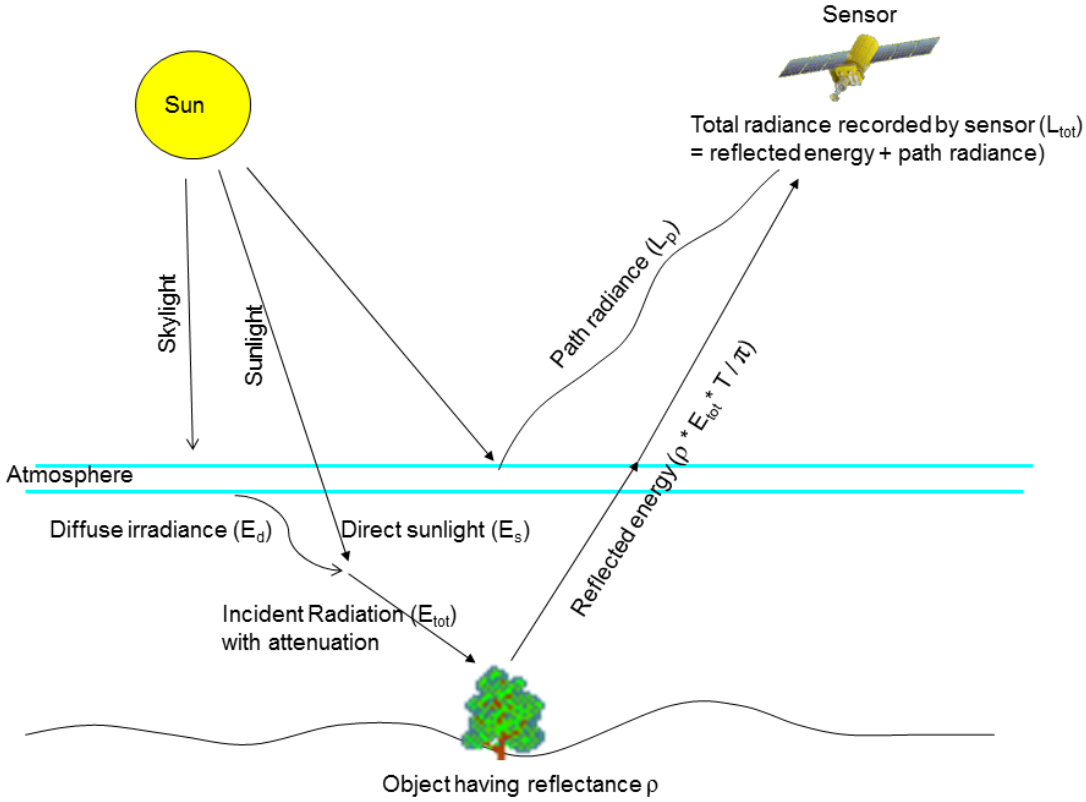


Figure 2.2 – Explanation of total radiance that is recorded by the sensor. The total incident radiation upon an object (E_{tot}) comes from a combination of the direct but attenuated sunlight (E_S) ("attenuated" means "reduced intensity") and skylight (or diffuse irradiance, Ed) which hits an object and reflects with reflectance ρ. The attenuated radiance reflected from the object (which is equal to $\rho*E_{tot}*T / \pi$, where T = the transmission of the atmosphere) is combined with path radiance (Lp), to equal the total radiance (L_{tot}) recorded by the sensor (in other words, $L_{tot} = (\rho*E_{tot}*T / \pi) + Lp$).

2.2.1. Atmospheric influence

Let's first understand the influence of the atmosphere on the radiation coming from the sun, as well as its influence on radiation reflected back from the objects on Earth. During optical remote sensing of the Earth's surface, the atmosphere will interfere with the light in several different ways.

One important mechanism is that incoming and outgoing electromagnetic radiation (light) is *absorbed* by molecules in the atmosphere (such as carbon dioxide, oxygen, ozone). In this process, shortwave light energy is converted to longer-wave heat energy. Different gas molecules interact with different wavelengths of the electromagnetic spectrum. For this reason, there are only certain "windows" (or wavelength bands) that are open for remote sensing through the atmosphere (Figure 2.3). Consequently, remote sensing sensors are most often designed to operate within these "atmospheric windows" where absorption of light energy is minimal.

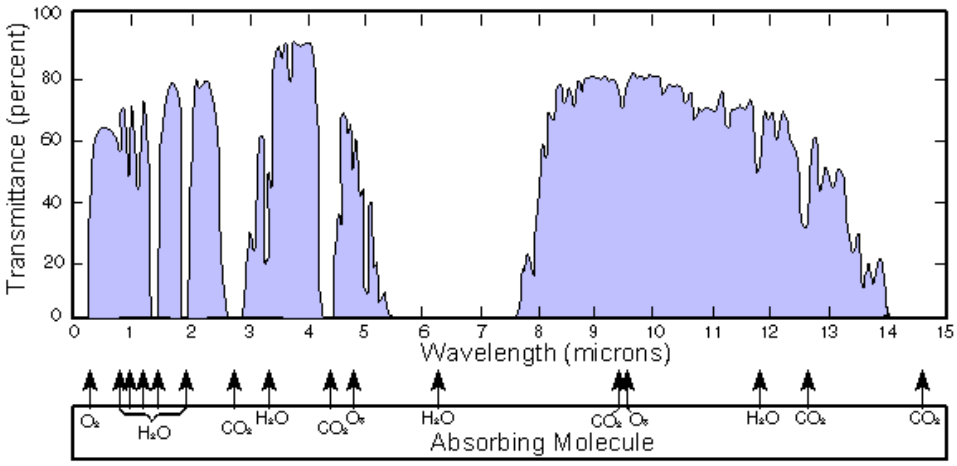


Figure 2.3 Atmospheric transmission in different wavelengths shown in purple. The transmission is zero in several wavelength areas due to gas absorbtion. Figure from Wikipedia.

The other major mechanism present in the atmosphere is *scattering*, which means that light interacts with molecules in the air and changes direction. The way in which light scatters is related to the wavelength, but in a more continuous way than with absorption. Scattering is also most apparent for shorter wavelengths, such as ultraviolet and blue light, and to some degree in the green and red wavelengths.

There are three main scattering mechanisms:

- Rayleigh scattering
- Mie scattering
- Aerosol scattering

Rayleigh scattering is caused by small molecules in the atmosphere. Shorter wavelengths (blue and violet) are scattered much more, because the strength of the Rayleigh scattering mechanism is roughly proportional to the wavelength. This is the main mechanism that causes the sky to look blueish. Since Rayleigh scattering is due to interaction with air molecules, barometric pressure can be used to estimate the amount of Rayleigh scattering.

Mie scattering is caused by atmospheric particles of about the same size as the wavelength of the light.

Aerosol scattering (or nonselective scattering) is caused by larger atmospheric particles (dust, haze, smoke, etc). Aerosol scattering has a larger day-to-day variation than the other scattering mechanisms. Since it is caused by large particles near the ground, aerosol scattering will be dependent on the weather situation, as well as the amount of air pollution in the area.

Scattering and absorption of energy can occur on the path from the sun to the Earth. When it is scattered and then hits the target it is referred to as *sky irradiance* or *diffuse irradiance*. Refer back to Figure 2.2 and observe that the main radiation components reaching the ground are a combination of *direct sunlight* and *diffuse irradiance*. The sum of these two components upon an object is called *incident radiation*.

2.2.2. Energy interaction with objects on Earth

What happens to the incident energy that hits an object? Total incident energy = *reflected* energy + *absorbed* energy + *transmitted* energy (Fig 2.4). In other words, no incoming energy will "disappear" as it interacts with an object; all of it is either reflected, absorbed or transmitted. The reflected light is what is seen by our eyes and recorded by sensors (e.g., cameras, satellite sensors, etc.). Different amounts of light are reflected in the different wavelengths, which results in our seeing different colors. The sunlight is aborbed and converted to heat and chemical energy in photosynthesis. Photosysnthesis uses primarily red and blue light and the green light, which isn't absorbed as much, is therefore reflected to a higher degree than the red and blue wavelengths. Transmitted energy is the component of light that passes through the object. Leaves are surprisingly transparent, and about the same quantity of sun light (in each wavelength) will be transmitted through the leaf as reflected by it.

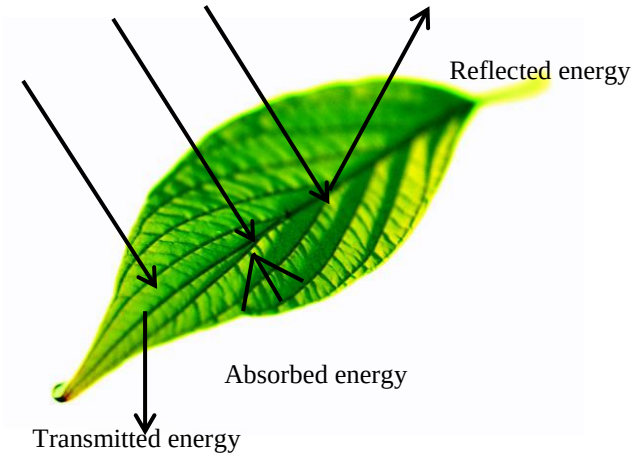


Figure 2.4 The incident energy which meets the leaf is either absorbed and used in photosynthesis, or is transmitted or reflected. (adapted from Figure 10:3 in F&F).

2.2.3. Reflectance and Reflectance factor

"Reflection is when incident energy (incoming light) hits an object and bounces off. The color of an object is actually the wavelengths of the light reflected while all other wavelengths are absorbed or transmitted. The physical and chemical composition of matter determines which wavelength (or color) is reflected." - *NASA*

Reflectance (denoted as "ρ") is easy to define in theory:
Reflectance = reflected energy / incident energy

Reflectance (ρ) is always given as a number between 0 and 1. However, to measure reflectance is more difficult, since both the incident radiation (coming in from all directions), and the outgoing radiation (going out in all directions), need to be figured into the calculation. In practice, a concept known as the *Reflectance factor* (R) is therefore used more often than Reflectance (ρ).

The Reflectance factor (R) = the radiation from a target that is actually measured from a given position for given illumination conditions, divided by the radiation that would have been measured if the target had been an ideally reflecting Lambertian surface (described in next section). In practice, the Reflectance factor can be measured by making a radiometer measurement over specific objects of interest (for example, a small plot of crops, or grass). Then a plate made of a material with near Lambertian properties is placed in front of the radiometer instead and a second radiometer measurement is made. The Reflectance factor is the ratio between the first and the second radiometer reading. Different objects have different reflectance factors, depending on the physical and chemical composition of the matter (as the NASA statement above says).

The proportions of reflected light in different wavelengths causes *color*. It is however only the diffuse part of the reflection, also called *Lambertian reflection* (= that reflects equally in all directions), that carries color. The opposite of Lambertian reflection is *specular reflection*. This means that the light is reflected like a mirror, (the angle of incidence has an equal size as the angle of reflection). Most objects in nature, such as vegetation, have properties that are a mix between ideal specular reflectors and ideal Lambertian reflectors. Figure 2.5 shows an example of Lambertian reflection and specular reflection.

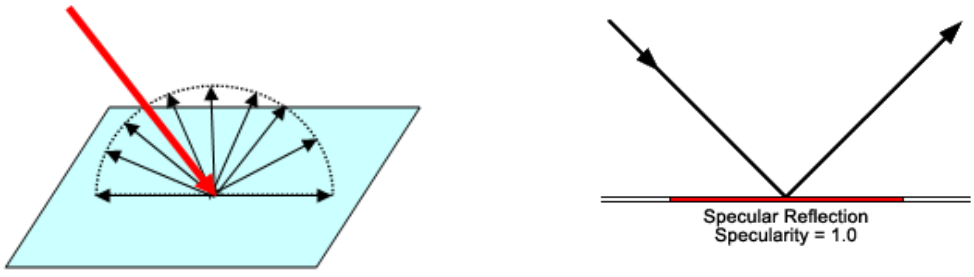


Figure 2.5. Demonstration of Lambertian reflectance (left) and Specular reflectance (right).

2.2.4. Calculation of radiance received by a sensor

It is only a few percent of the light energy hitting the Earth that is reflected back up to the atmosphere. The reflected light that reaches the sensor is also combined with the *path radiance*, which is radiation that comes from scattering of light in the atmosphere, and that have reached the sensor without having been reflected from the target.

Formula (Eq 2.6) show how to calculate what is being measured by the sensor. The solar radiation that reaches the Earth is measured as the total incident radiation per area unit, and called *Irradiance*, which in physics is abbreviated (E) and has the unit measure Wm^{-2}. The two main components of Irradiance (E) are the diffuse irradiance from the sky (E_d) and the direct sunlight (E_s). These two components are often summarized to a measure of total irradiance (E_{tot}) hitting the target: ($E_{tot} = E_d + E_s$).

A sensor measures the radiation that is received from a given direction, in a cone angle. The sensor measures Radiance, which in physics is abbreviated (L) with units $Wm^{-2}\ Sr^{-1}$ where Sr stands for steradian, which is the unit of a cone angle.

The total radiance reaching a sensor can be written:

$$L_{tot} = (\rho * E_{tot} * T / \pi) + Lp \qquad (2.6)$$

L_{tot} = Total radiance (Wm^{-2}/ Sr) reaching the sensor
ρ = Reflectance for the studied object
E_{tot} = Total irradiance hitting the studied object at ground
T = Transmission in the atmosphere (0 – 1)
L_p = Path radiance (Wm^{-2}/ Sr) which is a radiance component from the atmosphere that reaches the sensor without having reached the target.
π = 3.14159… (the need for a π in the formula can be derived from the fact that E_{tot} considers Irradiance from all directions and Ltot measures radiance from one direction).

The formula in Eq 2.6 could be written for a defined wavelength region (e.g., for example visible light from 0.4 – 0.7 µm). However, it is most usual to use a general "radiance per unit wavelength" term instead. This term is spectral radiance (L_λ) and the unit is Wm^{-2}/ Sr µm.

There are computer programs that can calculate Es, Ed, T and Lp for given atmospheric conditions. Theoretically, we could then solve the relationship between radiance registered by the sensor (L_{tot}) and the reflectance properties of the studied surface, ρ. The most well-known of these programs is "6S" and also "ATCOR". In practice however, this approach does not give accurate enough results to create good forest variables. Atmospheric correction is most important to do if one is measuring physical variables, such as Leaf Area Index. Often, however, it may be sufficient to use "relative

methods" with remote sensing data that are not severely affected by atmospheric haze, and take the approach of interpreting the remote sensing data by building relationships between the image data and field-measured forest data.

2.2.5. Practical implications of atmospheric influences

Sky irradiance leads to the shorter wavelengths of light, particularly the blue light, coming from all directions in the sky and not just from one direct path straight from the sun. This has important implications for aerial photography as well as for satellite remote sensing. One such implication is positive, since sky irradiance makes it easier to obtain information from the shadowed parts of an image, e.g. "between the trees", if the blue light is present. Thus a "*true colour*" (an image with blue, green and red wavelengths – the wavelengths our eyes see) image is better for this purpose than a "*false colour*" image with a near infrared band but no blue band. A further implication is that when taking photos from high altitudes, there will be a lot of blue light reaching the sensor which is reflected from air molecules and not necessarily from the object or ground. This will cause high altitude true colour photos to have a bluish haze, which is a negative effect.

The effects of the atmosphere can be seen in satellite imagery when one looks at viewing radiometrically uncorrected remote sensing images (i.e., without an atmospheric correction) or by inspecting the image histograms (image histograms will be discussed later in the compendium. We haven't covered the subject of histograms yet, but for now remember that histograms are simply a plot of the frequency of the values of a chosen variable). The effect of a thick atmosphere compacts the histograms primarily in the short wavelengths (e.g., blue wavelength the most), resulting in a very poor dynamic range. Path radiance adds a background brightness which shifts the histograms to a higher value range, again with a stronger effect in the short wavelengths. True-colour composites will reveal a bluish haze over the image with very poor contrast. Under clear atmospheric conditions, shadows will be much deeper, water will appear darker and there will be much better overall contrast. The short visible wavelengths, such as blue and green will be most severely affected by poor atmospheric conditions while the shortwave-infrared bands are least affected and typically have dark shadows and good contrast even in relatively poor atmospheric conditions.

The main problem with optical remote sensing, both using satellites and aircrafts as platforms, is that clouds make most days useless for image acquisition. This is one of the main advantages for radar sensors which operate with cloud-penetrating wavelengths. However, even if images are acquired during an apparently cloud free day, it should be remembered that the optical thickness of the atmosphere varies from day to day due to different weather conditions.

2.3. Reflectance from boreal forest

Everything in nature has its own unique distribution of reflected, emitted, and absorbed radiation. These spectral characteristics can—if ingeniously exploited—be used to distinguish one thing from another or to obtain information about shape, size, and other physical and chemical properties. (Parker and Wolff, 1965, p. 21)

Reflectance from vegetation will have variations characteristic of specific vegetation types, such as tree species or difference between shrubs, soils and grass, which can be used for interpretation of remotely sensed data. Living vegetation will absorb blue and red light, but absorbs less green light (i.e., reflects more green light in proportion to red and blue), which is the reason why vegetation appears green for our eyes. If our eyes would have been sensitive also for near infrared light (= the light just beyond 0.7 μm) we would see that the reflectance from living vegetation is particularly high in this wavelength region. This phenomenon can be seen when displaying color infrared images, where the sensor response from the near infrared light is most often displayed using the red display color.

The typical reflectance curves shown in Figure 2.6, usually taken from isolated object (e.g., perhaps taken from single leaves) will be less distinct when the measurements are taken over a larger area that consists of a mix of that vegetation types, such as full trees, or groups of the trees including their shadows, and branches.

For example, the color of a 10 x 10 m satellite image pixel in a forest will be composed of four components: sunlit canopy, sunlit ground, shadowed canopy, and shadowed ground. The fraction of shadows is an important factor for the brightness of a forest pixel, in particular at our northern latitudes, where the sun angles are low.

There is a correlation between how dark a forest pixel is and tree stem volume (or correlated variables like tree biomass or basal area). This might be explained by the fact that large trees, or many trees, causes more shadows than small or few trees. The correlation exists for blue to SWIR wavelength regions and stem volume, and it becomes stronger with longer wavelengths -- with exception of the Near-Infrared band. The reason for increase in correlation between the wavelength bands is probably that the shadows are darker and “sharper” in the longer wavelength bands (e.g., SWIR) where there is less scattered light from the atmosphere (as compared to the blue band). See Figure 2.7 and 2.8.

In the NIR wavelength region, the reflectance is affected not only by shadows, but is also highly affected by vegetation activity and structure. The correlation with photosynthetic activity within the pixel will be stronger than the correlation with the tree biomass. NIR can also help distinguish tree species (particularly to distinguish between broad-leaved and needle-leaved) and tree health (e.g, dead or diseased trees). If one wants to study tree-species differences, a ratio of NIR and another band (often SWIR or Red) can be useful, as the effect of shadows can be reduced by use of a ratio. With the

availability of high temporal frequency of images, phenological changes can be used from a time series to identify tree species or other vegetation. If on the other hand one wants to predict stem volume, linear combinations of several spectral bands is best to use, in particular the longer wavelengths, so that one can use the advantage of having information about the tree shadows (which can be related to tree size). An obvious example of how field layer vegetation and ground conditions influence reflectance is shown in Fig 2.9.

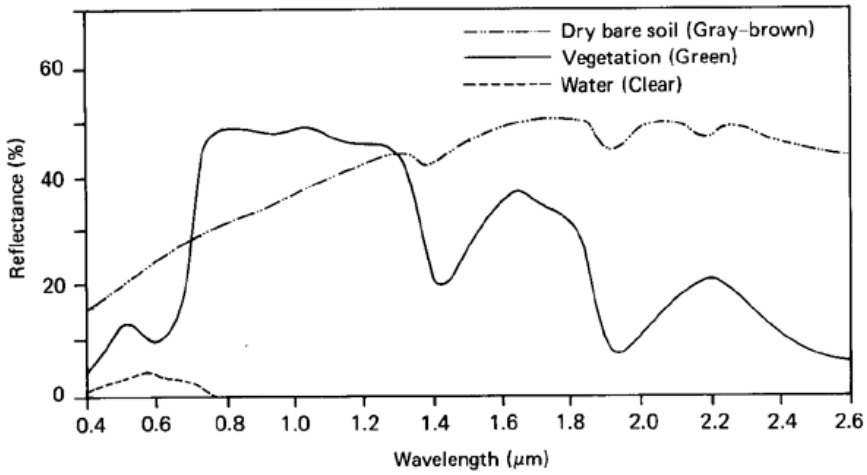


Figure 2.6. Typical reflectances from vegetation, soil and water, as a function of wavelength.

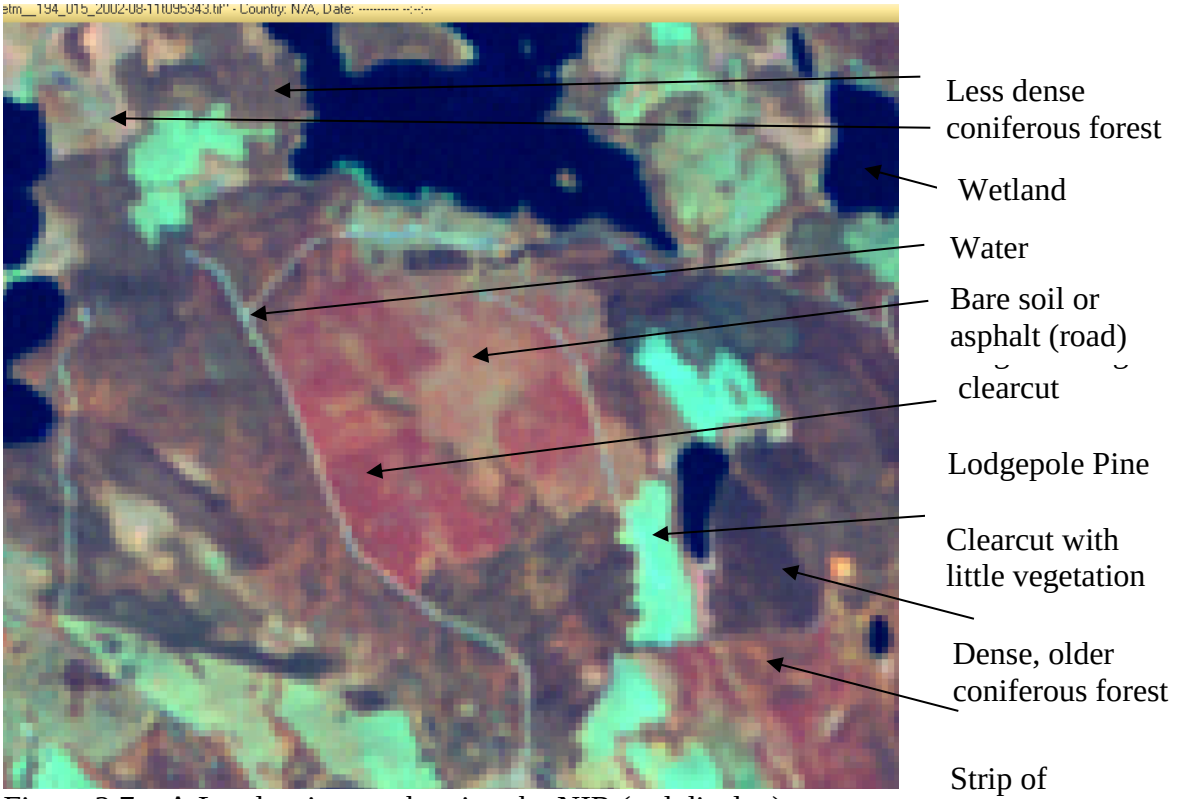


Figure 2.7 – A Landsat image showing the NIR (red display), SWIR (green display), and Red (blue display) wavelengths bands. Examples of different forest and vegetation types easily distinguished visually are pointed out.

Figure 2.8 – Color-IR Orthtophoto (left), SPOT satellite image (right, with NIR, SWIR and Red as RGB display) over a mountain birch forest, where darkening effect of tree shadows can be seen in both images. For orientation, the red circle is the same plot in both images.

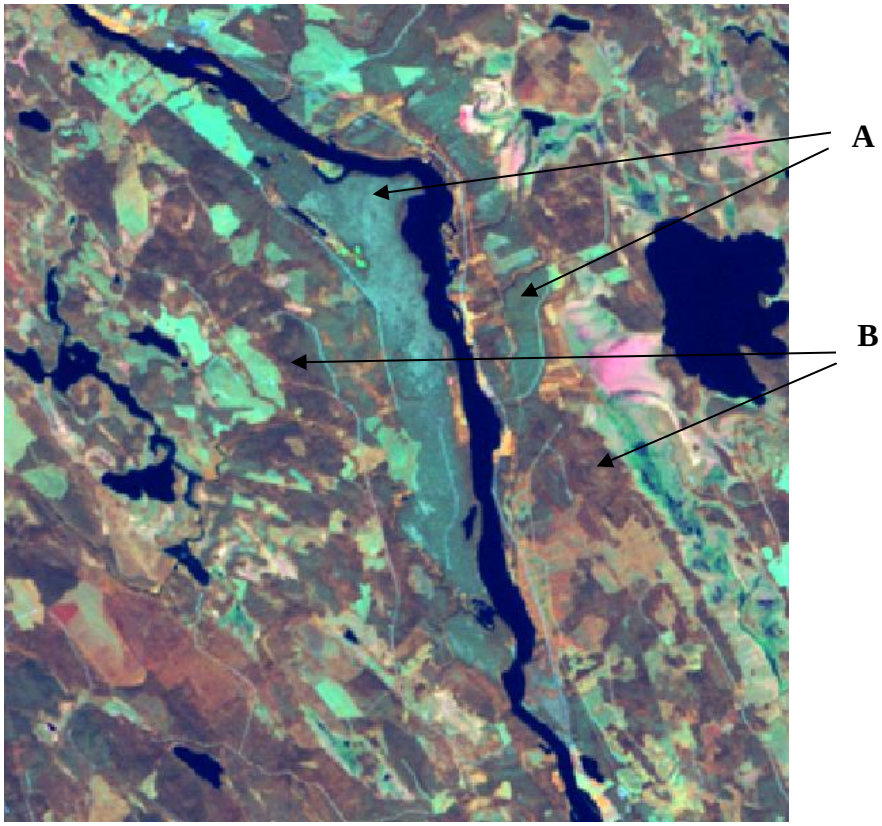


Figure 2.9. SPOT image (with NIR, SWIR and Red as RGB display) along the Vindel River. Example areas A show Coniferous forest with reindeer lichen and sand in the field and ground layer (appearing blue-ish), while example areas B indicates Coniferous forest where there is a shrub field layer.

Self study Questions

1. Blue light is scattered more in the atmosphere than longer wavelengths such as for example red light. Which advantage and which drawbacks does the scattering of blue light cause when using aerial photography?

2. In boreal forests, areas with high stem volume will appear as dark in most wavelength bands when analyzing optical satellite images. Mention a possible explanation for this.

3. FOUNDATIONS IN REMOTE SENSING

Optical sensor: The term is used here to collectively define sensors which register the reflected sunlight, for example, aerial photography, airborne optical scanners, and satellite-borne electro-optical sensors.

LiDAR: A system which measures distances by registering the time between sending away a laser pulse, and when it is returned. A scanning system is often used, and therefore it is often referred to as laser scanning. Airborne laser scanning (ALS) is an important data source for forestry.

Radar: A system which creates an image through registration of how a returned microwave wavelength is returned.

Passive sensor: A sensor which registers reflectance whose source radiates from nature (e.g., reflected sunlight, or warmth radiated from an object on earthe), as opposed to a sensor which sends out it's own signal. Examples are aerial photography cameras, multi-spectral optical sensors placed on satellites, and thermal cameras.

Active sensor: A sensor which sends out its own signal which is then also recorded by the sensor. Examples are LiDAR and radar.

This chapter is an overview and introduction to sensors commonly used to perform remote sensing of forests, which platforms are commonly used, and what type of data are produced by these sensors. A short introduction to forest information collected by these sensors is also given, but more details will be given in the individual chapters about the different remote sensing data types.

3.1. Sensors

Remote sensing can be performed using a wide range of sensors. We can categorize these sensors by their properties, often defined by the mechanism they use or the wavelength in which they operate. The following three categories – Optical, LiDAR, and Radar – are those currently most used in forest remote sensing today. Optical sensors refers to "Electro-optical" sensors which are cameras or scanners that register the sunlight's reflectance from objects.

3.1.1. Active and passive remote sensing techniques

One important distinction between sensors is identification of them as *passive* or *active* remote sensing techniques. Passive techniques mean that the sensor registers energy that is either reflected by or emitted from an object. *Active sensors* send out energy themselves and record the signal

that comes back as well as the time it takes for the signal to travel back and forth from the sensor. One example of an active sensor is the *laser sensor* which sends out very controlled light pulses and record the time delay and intensity for the reflected light pulse. Laser scanners also register the angle of each laser pulse sent out. Another important example is *radar*, which sends out very controlled radio signals and creates an image from the returned radio signal. Table 3.1 exemplifies some passive and active remote sensing techniques and comments on their usability for forestry.

Table 3.1 Examples of passive and active remote sensing techniques and comments on their usability for forestry.

	Optical wavelength region	Microwave wavelength region
Passive techniques	Optical sensors on airplanes (e.g., Aerial photography) Traditionally the most important remote sensing techniques for forestry. Today photographic film in aerial cameras has been replaced with electro-optical sensors such as CCD arrays and matrixes. Optical sensors on satellites Are of great importance for obtaining overviews of forest resources over large areas, and for detection of changes. Values within each pixel are a measure of reflected light. Three dimensional data can also be computed from overlapping images taken from different observation points.	Passive microwave – There are sensors on board satellites and aircrafts that register radio waves emitted from the earth, however due to the low energy emitted, the footprint (pixel size) is so large that these sensors have no practical importance for forestry.
Active techniques	Laser scanning – Laser systems work with light, often in the near infrared wavelength range, but also the visible and SWIR wavelengths. Currently, multi-spectral laser systems are available. Distance measuring laser scanners are used for creating digital elevation models, as well as for measuring tree heights and canopy structure. They can produce a 3D point cloud, which has proven to be very useful for estimation of forest variables.	Radar – There are many radar systems, both onboard aircrafts and satellites. Their importance to forestry is often dependent on the wavelength. Systems with longer wavelengths have a higher potential for forestry (see the radar chapter for more information). The pixel value shows the electrical properties from the site of the pixel, but registrations from different view points can also be combined to form 3D data.

3.1.2. Optical sensors

Cameras with film are a form of passive optical sensors. In the 1960's electro-optical sensors were developed where incoming light is converted to an electrical signal which is then stored in memory.

There are different mechanisms for scanning from an aircraft or satellite. The oldest principle is to have a rotating mirror in front of the sensor that direct swaths of light across the scene into the sensor. However, this sensor had many moving mechanical parts.

The technique used most today is a *"push-broom"* technique, where there is an array of sensor elements, often of a type that is called a CCD array (the inventors of the CCD array received the Nobel prize in Physics 2009). The sensor array senses a line of pixels at a time, and is dependent on the forward speed of the platform in order to feed data in from the next line (i.e., array) which might be only 0.03 seconds later (Fig. 3.1). Particularly when these sensor types are mounted on an airplane, geometry problems may arise. For example, a power-line corridor that is straight (in reality) may appear curvy due to distortions caused by the plane being blown around by turbulent wind. It was for this reason that first in the 1990s that a combination of Inertial navigation systems (INS) and Global Positioning Systems (GPS) could be used to record and then correct for the plane's movement when using a scanning camera. It was after this technological development that this type data from electro-optical sensors on airplanes could really be used. Today, the best aerial photography cameras are the push-broom type.

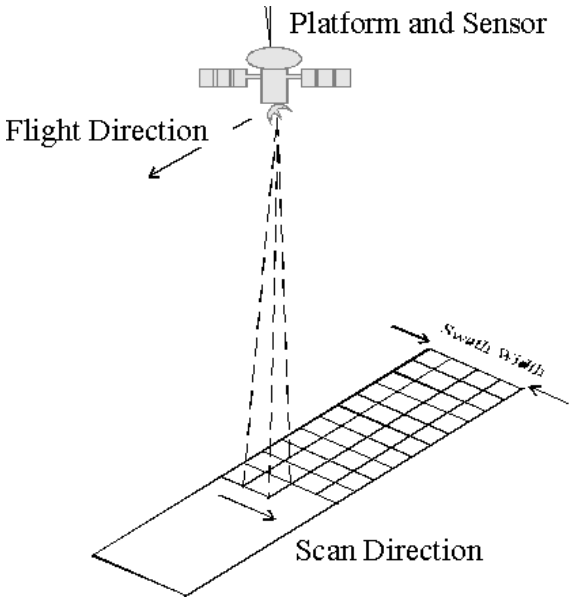


Figure 3.1. Demonstration of a push-broom sensor (image from Wikimedia Commons).

Another electro-optical sensor type is to have a *matrix* of sensor elements which instantaneously registers an image. An advantage of this sensor is the ability to have more accurate positioning of all the cells in the raster image in relation to each other. It is also a smaller system than a push-broom sensor.

Most sensors record multiple wavelengths, and record each specific wavelength range separately. The design and technical specifications (which wavelength bands, what pixel size) is often based on the eventual use of the data collected (e.g., for ocean, or land, or ice), but also is based on what is technically possible, avoidance of certain wavelengths affected by the atmosphere, as well as cost. The sensors may record energy in a few wavelengths (e.g., the four bands blue, green, red, near infrared is a common choice) and be referred to as a *multi-spectral sensor*, or many narrow ranges (e.g., 260 bands), in which case it is called a *hyperspectral sensor.*

More details on optical sensors are given in Chapter 4 on Optical Satellite data and in Chapter 5 on Aerial Photography.

3.1.3. LiDAR

Laser scanning has been established as a new and effective approach to data capture for forestry purposes. Airborne laser scanning measures the location of points on the ground and in the canopy with decimeter accuracy. Airborne laser scanning uses the laser's ability to measure distance by using the time from when a laser pulse is sent out, until the laser light is reflected from the ground or vegetation comes back to the sensor. The technology to measure distances with laser is often called *LiDAR* from the English “Light Detection And Ranging”. More details on LiDAR are given in the Chapter 6 on LiDAR.

3.1.4. Radar

Radar (acronym for RAdio Detection And Ranging) is an active sensor for detecting, locating, tracking, and identifying objects even at a considerable distance. The Radar system sends out and receives back microwaves and is a technique that is not influenced by cloud coverage (as sensors that use shorter wavelengths are). Typical wavelengths are 3 cm (called “X-band”), 5cm (called “C-band”) and ca 25 cm (called “L-band”). In the case of Radar, the size of the wavelength is related to what size objects can be detected. That means that for a simple, single radar image, the L-band radar system is that one that will be most correlated with wood volume. If radar images from multiple time points or multiple registrations are used, then the shorter X and C wavelengths can be of interest for forestry. Currently, radar data is still not widely used for forestry purposes. More details about Radar are given in the Chapter 7 on Radar.

3.2. Platforms

The sensors used in remote sensing often need to be placed on a larger carrier in order to give a good position over the objects of interest. The carrier of the sensor is referred to as a “platform” and is a separate component of a remote sensing system. It may be anything from a satellite, to an airplane, to a stative. There are categories referring to the position of the platforms, namely spaceborne, airborne, and terrestrial. The choice of platform is based on the

purpose of the sensor. For example, if larger areas need to be imaged, then the sensor needs to be high above the landscape, requiring a spaceborne or high-flying airborne platform.

3.2.1. Satellites

Satellite platforms are a common carrier of sensors, providing stable and repetitive measurements of the Earth. There are satellites used for communication, GPS, and many other purposes, but here we concentrate on what are called “Earth Observation (EO) satellites”. The orbit characteristic refers to the movement or placement of the satellite in relation to the Earth. There are two main types of orbits used for remote sensing satellites:

Most EO satellites often go in a *polar orbit* (or near-polar orbit), which means that the satellite travels north to south over the poles and images the area of the Earth under it, which varies due to the rotation of the Earth. The altitude is typically between 600 km and 2000 km above the earth. When the satellite crosses the equator at the same time with every orbit, this is called a *sun-synchronous polar orbit*, which is very common for most EO satellites.

There are some satellites whose orbit doesn’t reach all the way to the northern lattitudes, such as the International Space Station (ISS) that turns at 52 degrees north, and therefore doesn’t pass over Sweden.

Some satellites have a *geo-stationary orbit* meaning that the satellite travels over the equator at a distance of 36 000 km at the same speed and direction as the rotation of the Earth, allowing it to always image the same area. Weather satellites are often geo-stationary. The images produced from that great distance do not have good enough spatial resolution for forestry.

The electricity needed for operating the sensor and sending the data to earth is obtained from solar panels. Most satellites also have small rockets that are used for correcting the orbit when needed. The fuel for these rockets is often a factor that limits the lifetime of a satellite. A typical lifetime is around five years, but some satellites have been operated much longer.

3.2.2. Airplanes/Helicopters

Aerial photography is often made with purpose-specific planes that have one or two central openings in the floor of the plane. These openings have standardized dimensions where the aerial cameras or other instruments, e.g., a laser scanner, can be placed. A "normal” flying altitude for aerial photography is usually between 2000 m and 4500 m. Even the acquisition of “high altitude images” may occur at around 10 000 m flying altitude.

Aerial photography planes can also be used for low altitude photography around 1000 m flying altitude, but for this kind of shooting, often done for smaller rather than larger area coverage, there are also several other options, such as drones, sport planes, or helicopter. At low heights, these platforms would allow a safer and more controlled data acquisition than if using an airplane.

3.2.3. Drones or UAVs

UAV is an abbreviation for *Unmanned Aerial Vehicle* and a common synonym is *drones*. Other common names are model aircraft, UAS (Unmanned Aerial Systems), or RPAS (Remotely Piloted Aircraft Systems).

Since the early 2000s drones have been used for forest applications, but it is only in recent years (2010s) that it has had an impact. A contributing factor to the breakthrough is that the systems have become cheaper, easier to fly and there is small cameras and software that makes it easy for the user to create orthophotos and 3D point clouds from images.

Today drones are used in forestry primarily for inspection and documentation of smaller areas, as well as for creating orthophotos from series of over-lapping images, over smaller areas. It is even possible to create 3D point clouds from the digital images, based on exactly the same principles as for over-lapping digital images acquired from airplanes and satellites. Drones can be used as platforms for other sensors as well, such as a laser scanner. However, placement of a laser system on a drone is currently rather expensive and sensitive to movements of the drone, and is not currently used in practical forestry. Developments are occurring quickly in this field, but perhaps the biggest "limitation" for typical "personal-sized" drones is the size of the area that is effective to cover. One must weigh the options of the application when considering which platform is right to use.

Regulations

In Sweden, the Swedish Transport Agency issues flight rules regarding drones. The main limitation of the current regulations for forestry purposes is that the flight must take place within the operator's sight. The drone platforms are divided into categories based on the starting weight and impact energy. A heavier platform requires higher security and is limited to a maximum flying altitude of 120 meters above ground level.

Permission needs to be sought from the Transport Agency for flying on missions, research or testing, but not if you are flying as a private individual for non-commercial purposes. The regulations must be followed even if you do not have to seek permission. Read on the Transport Agency website (www.transportstyrelsen.se) for the current rules. Besides permits for flight, you may have to seek permission for taking and dissemination of any photos acquired from a UAV. However, the rules around this are currently changing in Sweden and it is recommended to check up to date information.

Safety

There are often a number of safety features built into the autopilot system that can be triggered by various events that can, in turn, compromise security. Examples of such triggering events are

- Flying off the defined flight range (geo-fence)
- Loss of radio contact (returning and landing)

- Loss of data link

The action to be taken when a triggering event occurs, the pilot can usually take some counter-action. Here are some examples of actions:

- Return to the starting point and land
- Continue the mission despite the incident
- Reconnect the manual pilot function and allow the pilot to take over the flight.

3.2.3.1. UAV Airplane

An airplane or flying wing has a technically simple design. Figure 3.1 shows a flying wing UAV from the Swedish manufacturer SmartPlanes. The sensors are usually mounted inside the fuselage, which also provides protection on landing.

Figure 3.1 This is a typical setup when flying. The pilot performs a pre-flight check. Data Modem (1) communicates with the aircraft (2) in real time and via the ground station software in the computer (3) the pilot can obtain information on the plane's location, speed, incline, etc. The remote control (4) is used to control the plane during takeoff and landing, but at any time during the flight the autopilot can be used to manually control the aircraft.

Control

Control can be done completely manually through a standard remote control, but usually tend to make use of the autopilot stabilization at takeoff and landing. With the help of the autopilot function, the plane can also take off and land automatically. How the start and take off are performed will vary between manufacturers. Some deliver a startup ramp, while others require the user to throw the plane into the air.

Advantages

There are a number of aircraft weighing less than 1.5 kg and thus fall under the Transport Agency's "Class 1A", which has no limit in flying altitude (however, must always be in the pilot's sight). The fuselage is often manufactured in polystyrene and is thus a very cheap and durable construction. It can even land without the engine turned on because it can glide to the landing. The sensor is usually mounted inside the fuselage and is well protected by a hard landing or crash. The platform can be flown without the support of the autopilot.

An advantage of an airplane UAV platform is that electronics (stabilization) aren't needed to fly -- it is enough that the pilot can control two servos. Even if these were to malfunction that doesn't necessarily mean the aircraft would crash, and the sensor would be destroyed.

Because the wings provide lift capability, very little battery is required. This gives a long flight time. Currently the regulations are more of a limiting factor rather than the battery time to fly larger areas.

Disadvantages

The speed depends largely on the wind and one should plan missions with the wind coming from the side of the flight direction in order to get the most constant possible air speed. However, one problem with the plane platform is the constant lean and therefore you will not be taking photos in nadir orientation if you do not have a gimbal mounted. During takeoff and landing, a relatively large area is required, for example, a clearing or a field. For the experienced pilot take-off and landing can occur on a road (untrafficked, of course!) even though the trees are close.

3.2.3.2. Multirotor UAVs

A multirotor platform is able to lift straight up by pushing air downwards. This technology has been further developed in recent years (2010s) thanks to cheaper inertial navigation systems. To keep a multirotor platform stable the motors must be controlled with high precision. If the vehicle starts to tilt on take-off, one or more motors must change speed to correct for this. There are a plethora of manufacturers and Figure 3.2 shows an example of a quadrocopter from the American company 3D Robotics.

Figure 3.2 Quadrocopter (1) of 3D Robotics model Solo. On this platform, there is a Sony Alpha SLR 5100 camera (2) with a 20mm lens.

Control

A multirotor UAV platform assumes that you have some kind of stabilization or autopilot that adjusts fan speed as the platform stays level. Many autopilots help the user to have control by simplifying the procedure, e.g., by keeping the platform stationary using inertial navigation (INS) and GNSS (GPS). Control can be done via both the autopilot and manually via the remote control.

Advantages

Take-off and landing goes very fast because the craft lifts straight up and can quickly get up to altitude where data collection occurs. The flight assignment (mapping) can be done quickly thanks to the fact that multirotor platforms do not need to fly a circle at the turn otherwise made at the end of each flight path. The speed can be kept constant, however, there will be a different pitch, but this can be solved with a gimbal.

Disadvantages

Battery time is in the current situation one of the major constraints and depending on altitude, you cannot cover such areas. Another disadvantage is that you basically cannot use a platform which will be in the Transport Agency's Class 1A; this is due to the weight being higher than 1.5 kg with the camera.

With a motor malfunction, the quadrocopter will generally crash, but a hex- or octocopter may be able to fly despite an engine malfunction. The body is also very sensitive to hard landings or crash landings. The sensor is usually mounted hanging under the craft and is thus easily damaged during hard landing or crash. A multirotor is completely dependent on autopilot or stabilization electronics and cannot fly without it.

3.2.4. Ground-based and mobile systems

3.2.4.1. Ground-based systems

Both laser scanning as well as images acquired with digital cameras can be made from sensors located on a stationary tripod standing on the ground. In this manner, the data can provide tree stem position and shape, etc. To get the 3D data from digital photogrammetry the photos must be acquired from more than one position, after which the images are co-processed with a special software. Even in laser scanning, there may be reason to scan from more than one position, so that the number of hidden trees are minimized (e.g., trees that cannot be seen because another tree stem blocks them from view). This type of ground-based registration has great potential, but is not yet operational in any significant way.

3.2.4.2. Mobile systems

It is also possible to record data about tree stems and positions via a mobile (moving) sensor located on the ground. A laser scanner can be mounted to a backpack, or on an ATV or a forest machine. The type of laser scanner suitable for this, however, does not give as accurate measurements as larger systems that are fixed to a tripod. It is also possible to acquire 3D information about the tree stems from a moving video recording. This technique is called "Shape from motion" (SfM).

When mobile data are acquired, we need to know the position of the sensor at any moment, which can be difficult in the forest where GPS signals provide limited positional accuracy. As a complement to GPS, inertial navigation systems are used, as well as techniques that use the trees depicted as reference points for the sensor movement, a technique known as Simultaneous Location and Mapping (SLAM).

3.3. Remote sensing data models

Digital geographic data can be represented in several formats, the primary types being raster and vector. *Raster* is the format used most for representation of image data. A raster has a defined number of rows and columns, with a set number cells (see figure 3.3). The grid cells have a defined equal size and may be square or rectangular. Raster data can consist of a single raster, or there may be several rasters layered in a single file. A raster may also be referred to as a grid or a matrix. *Vector* data comprises point, line, and polygon format.

Another characteristic of remote sensing data is dimensionality. We currently talk about data as having dimensions, and being two-dimensional (2D), three-dimensional (3D), and even four-dimensional (4D). Two-dimensional means that the data have East-West and North-South geographic coordinates, while 3D means that the data have X, Y, and Z (Elevation), and 4D refers to the previous three dimension, plus the dimension of time.

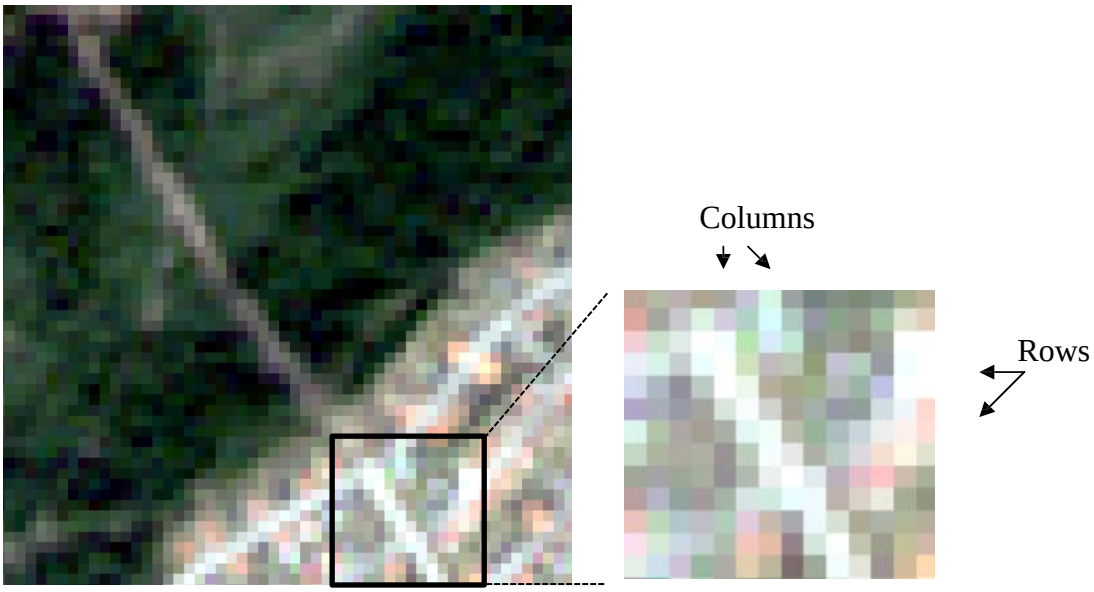


Figure 3.3 Illustration of a typical raster satellite image divided into grid cells (in this case, pixels) by a given number of columns and rows.

Laser scanning data are delivered as point data (Fig 3.4), and each point has three-dimensions of X, Y and Z. To each point, a number of attributes can be given, such as color. A common file format for ALS data is .las format, which you can read about in Chapter 6. Essentially, laser data is a type of vector data, although it is stored in its own .las dataformat. In Chapter 9 you will read about the "area-based method" of analyzing point cloud data in order to create forest estimates. The area-based method uses point cloud data and creates metrics within raster grid cells, thereby using converting vector data into new information that uses a raster format.

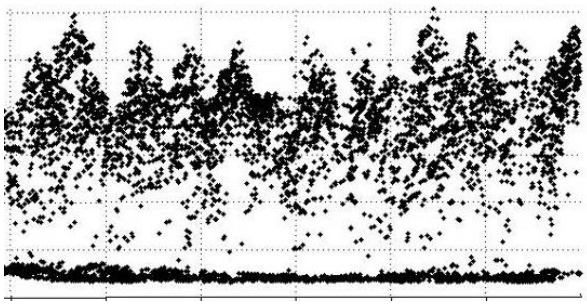


Fig. 3.4 Example of point data (in profile) from a high point density airborne laser scanning.

3.3.1. The principal data structure of a digital raster image

A digital raster image could be viewed as a matrix of grid cell values, where the raster has a certain number of rows and columns. The raster grid cells in the optical satellite data are called *pixels* (short for "Picture Elements"). The raster, and therefore the pixels, have geographic coordinates associated with

each corner of the pixel, and as such have a given dimension, e.g. 0.5 m × 0.5 m for a digital aerial photo, or 10 m × 10 m for a satellite image (Figure 3.3). Pixels may also be rectangular, but this is not as common.

Each pixel contains a value, often referred to as a *Digital Number, or DN*, which corresponds to the reflectance the sensors measure from the corresponding area on the earth. The DN value will have a limited range called the radiometric resolution, which may typically be 8-bit, 12-bit, 16-bit (for example, 8-bit data have values between 0 and 255).

The conceptual data structure described here (a matrix or raster) is identical to those raster data structures used in geographical information systems (GIS). When several wavelengths are acquired, these are represented by having one "layer" for each color.

However, there is an important distinction between just the raster format and a remotely sensed image, since the values from the image is essentially a sampled signal, and depending on (among other things) the sensor type, the platform's speed, the topography, and the atmospheric effect, the DN can be a sample not just from the area delineated by the pixel, but also a bit of the surrounding area. This is an advanced topic in remote sensing, but may be important to be aware of. See, for example, the definition of "ground resolution cell" in the following section.

Optical satellite data are often delivered in 2D raster format, as a file that consists of several raster layers, where each raster layer represents a different wavelength. For example, optical satellite data from a sensor which records reflectance in the green, red, near-infrared and short-wave infrared bands will have four raster layers. The rasters are (should be) geographically co-registered to each other, and often have the same grid cell size.

Image files can have a large range of different data formats such as TIFF, GEOTIFF, JPG, BIL, BSQ, IMG, JP2000 format, among others. These formats are common, and it is possible to export from one format to another using standard computer programs. One should be take care not to lose data quality or other information when exporting to new formats. GeoTIFF is currently a standard format. Raster files can be very large, and the data format can influence the speed of processing the data files.

3.4. Data resolution

When describing different characteristics of remotely sensed data, there are different types of "resolution" that are important. Each sensor has its own characteristics of importance, and you often need to know these in order to determine whether it will be useful for an application. We will first concentrate on passive sensors, since terminology for active sensors and passive sensors is different.

The primary terms to describe resolution characteristics are

- Spatial resolution
- Spectral resolution
- Temporal resolution
- Radiometric resolution.

3.4.1. Spatial resolution

Spatial resolution refers primarily to pixel size or *ground sampling distance* (GSD; Fig 3.5). Although there are many concepts that are similar between photographic film-based cameras and digital electro-optical sensors, it is good to note that the concept of "resolution" is different between these two systems.

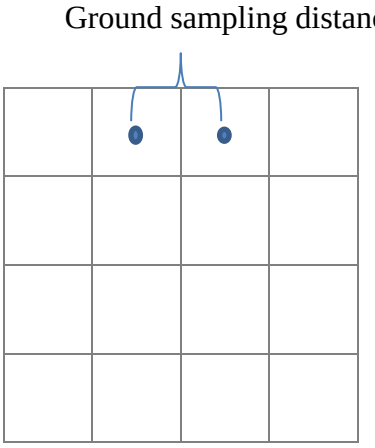


Fig 3.5 – Schematic of raster showing the concept of ground sampling distance

The pixel size, based on the pixel spacing (the GSD) in an image is determined by the platform (and scanning) speed and the sampling rate. The GSD tells the distance between pixel readings, which might be different along track and across track. It is important to remember that digital images have often been "*resampled"* to a new pixel size through some pre-processing steps, such as orienting the images into a geographic coordinate system. Different resampling methods can be used and have different effects on the original data values, and are discussed in Chapter 4 on Optical data.

The area that the sensor actually "sees" at one moment is called the *ground resolution cell* (GRC). This area could be both slightly larger than the area given by the GSD (Figure 3.6).

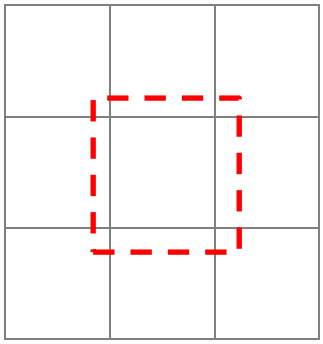


Fig 3.6 – Illustration of the concept of the ground resolution cell (GRC), which may be larger than the Ground Sampling Distance (GSD).

The size of the GRC depends on the characteristics of the sensor as well as flying characteristics (height, look angle, etc). Depending on the properties of the sensor, the actual shape of the GRC area can be a circle or a square/rectangle.The GRC is calculated as the segment of the ground that is measured, which is circular and has a radius of D (Eq. 3.1), where

$$D = H' * \beta \qquad \text{Eq 3.1}$$

where H´ = flying height above the terrain and β = the *Instantaneous Field of View (IFOV)* of a system, which is expressed as an angle expressed in radians.[1] This is an advanced topic within remote sensing.

Image extent refers to the area on the ground that a single image covers. The pixel size often has a relationship with Image Extent, in that large pixel sizes often allow for larger image extents whereas small pixel sizes tend to be collected over smaller image areas. In some cases the image extent can be described as having a limited area in both width and height (e.g., 60 km x 60 km), or in the case of some scanners, only a width is referred to, often called the *swath width*, as the scanner is continually moving in one direction and is only limited by the swath that it covers. Here, the data are often extracted according to the user's needs and according to how the delivering company wishes to sell the data.

A combination of the sensor properties and the flying height of the sensor (in the case of the satellite sensor, the *height of orbit*) will play a role in determining GRC and Image extent.

[1] Please refer to Lillesand et al. (Figure 5.2) to see a graphic demonstration of this concept.

3.4.2. Radiometric, spectral, and temporal resolution

Radiometric resolution refers to the quantization used by the sensor to record the incoming radiance. It indicates the sensitivity of the instrument. The radiometric resolution is expressed in the number of bits used in the quantization. While many previous sensors have had 8-bit radiometric resolution (with values from 0 – 255), the new trend is towards higher radiometric resolution, such as 12- and 16-bit.

One advantage of increased radiometric resolution is the potential to have a higher dynamic range of values. However, even if sensor can store data over a wide range of values, the actual meaningful radiometric resolution might be less, for example, due to technical limitations in the sensor system.

The *spectral resolution* for a wavelength band refers to the width (in nm) of the wavelength sensed by the sensor. This is not just a matter of indicating whether they are blue, green, red, or Near-IR, but refers to the exact region of the electromagnetic spectrum that the sensor is constructed to be sensitive to. It is a designed technical specification of the instrument. Hyperspectral sensors, for example, sense hundreds of narrow ranges of wavelengths, and are therefore said to have a high spectral resolution.

Temporal resolution refers to the *revisit time* of a sensor. For example, Landsat satellites pass over the same spot on Earth every 16 days, while the Sentinel-2 satellite passes over the same spot every 3 to 5 days. The satellite that passes over more often has a higher temporal resolution. Since optical images are affected by clouds and atmospheric conditions, a high temporal frequency will hopefully give more opportunities to collect cloud-free images.

A characteristic connected to temporal resolution is the ability of some satellite sensors to change the view angle of the sensor. Some satellite sensors (e.g., Landsat) are constantly looking straight down, and acquiring what are called “nadir-pointing images”, meaning that at the mid-point of the image, the sensor is looking straight down, without any angle. In contrast, some sensors (e.g., SPOT or Pléiades) can acquire data by pointing their sensor at an angle, and these are called “off-nadir images”. The advantage to a pointable sensor is the flexibility of taking images in several areas. If it is cloudy at nadir, and a non-cloudy area is viewable at +20° viewing, then that image can be acquired instead. The ability to point the sensor allows for a higher temporal resolution.

3.5. Display of raster images

To understand the display of optical data from remote sensing, it is necessary to discuss some basic concepts. The human eye sees color only in the “visible wavelengths”, that is, the red, green and blue wavelengths. Remember that most raster files will contain multiple spectral bands or layers of data within a single image file. Often, there is one or more wavelength represented that we do not usually see with the human eye.

The displays that we use on computer screens are made for our human eyes, which view in the visible spectrum, consisting of colors from violet to blue to green to yellow to red. Most image processing and GIS software that is used to look at image data uses a display with three "color guns" – Red, Green, and Blue.

The following is a descriptive example of how this works: We can add one wavelength into a specific color gun – an easy example is to display the Blue reflectance into the blue color gun display, the Green reflectance using the green color gun display and the Red reflectance using the red color gun display. This is a "true color" display, because this is what our eyes would actually see. We could have used any combination we wanted to, such as the putting the Blue wavelength in the red color gun display, and so on, but this would look unnatural and unintuitive to us. The software will use the Digital Numbers (DNs) in the raster to display this amount of red for that layer, in a gradient. If you were to look at the red wavelength, in just the red band, it would be a red image of varying red hues.

3.5.1. Additive and subtractive color theory

The human eye has three types of rod cells: short-, middle- and long-wavelength sensitive (blue, green and red). The *trichromatic theory of color* states that when all three types are stimulated equally, we see white light. Other colors are perceived when the three rod cell types are stimulated with different amounts of light (Eq 3.2):

$$C = (b_R \times R) + (b_G \times G) + (b_B \times B) \qquad \text{Eq. 3.2}$$

where C is the perceived color, R, G, B are the three primaries (red, green, blue) and the b's are the proportions of each. To use the R-G-B primaries on a computer screen or a projector is an *additive* principle.

Other processes are fundamentally *subtractive*, such as most inks on white paper, which use the subtractive primaries Cyan-Magenta-Yellow. These primaries are complementary colors of red, green, and blue. The complementary color for each primary color is obtained by mixing the other two primaries (e.g. cyan is the complementary color of red, and is produced by adding red and green).

Color additive theory says that when we take the colors Red, Green, and Blue, and add one to another, we get new colors (cyan, yellow, magenta), and if we add them all together, in equal amounts, we perceive white (Fig 3.4).

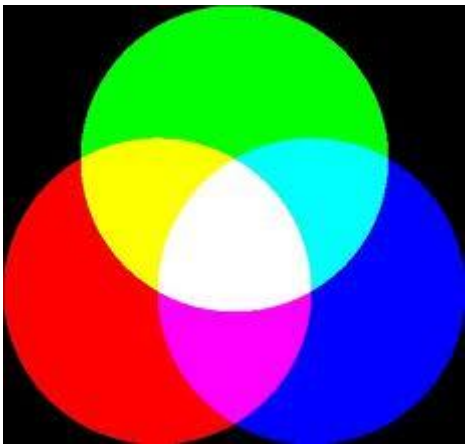


Figure 3.4 – A demonstration of the concept of additive color theory

A *Color Space* is a coordinate system for measuring and specifying colors. Some color spaces are designed to Euclidean distances between coordinates represent approximately perceptual differences between colors. The RGB color cube (Fig 3.5) used in image display on computer monitors is defined by the tristimulus coordinates of the phosphor response.

All colors that can be realized by the display are represented by (R, G, B) coordinates in the cube:

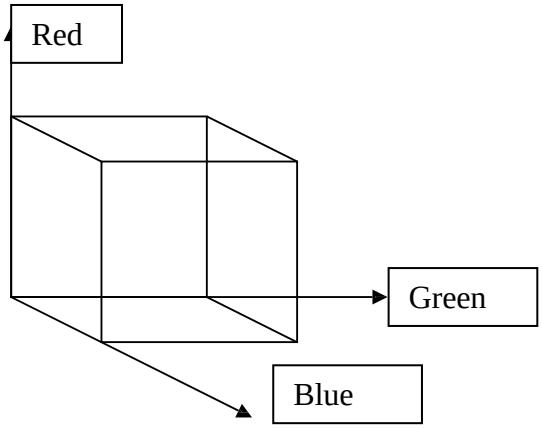


Fig 3.5. The RGB color cube[2].

In general, the R-G-B color space is not perceptually uniform; it cannot display all colors; and it is very device dependent.

A Red-Green-Blue coordinate triple for each cell determines the intensity of each primary color on the display. A large number of colors can be created by mixing the intensities of R-G-B primaries (256)3 =~16.5 Million unique colors!

[2] See also figure 7.36 in Lillesand et al.

3.5.2. "True color" and "False color" images

Most aerial cameras records colour in four spectral bands: *blue, green, red and near infrared*. Often there is also a *panchromatic* band with a broader spectral bandwidth and smaller pixel size recorded. During later postprocessing of the images, the panchromatic band might be combined with the multispectral bands in order to obtain a "pan-sharpened" image with both colour and high resolution. The four colours above are sufficient for producing both true colour and false colour imagery (Table 3.2 below).

Table 3.2. Standard schemes for displaying a true colour and a false colour infrared image on a colour screen.

True colour (vegetation appears green)			False colour (vegetation appears red)	
Sensed wavelength	Displayed as		Sensed wavelength	Displayed as
Blue	Blue		Blue	-
Green	Green		Green	Blue
Red	Red		Red	Green
NIR	-		NIR	Red

A "true colour" image might be displayed from Landsat 8 data, by displaying B2 (blue wavelength) with the blue colour gun, B3 (green wavelength) with the green colour gun and B4 (red wavelength) with the red colour gun. A "true colour" image is not possible with SPOT data, since there is no blue band.

A "false colour" image where vegetation will have reddish colours for vegetation, such as in a colour infrared aerial photo, can be displayed using the following bands

	Landsat 8 band	SPOT band
Blue colour gun on display	B3 (green)	B1 (green)
Green colour gun on display	B4 (red)	B2 (red)
Red colour gun on display	B5 (Near-IR)	B3 (Near-IR)

Often, in remote sensing we talk about displaying bands in R,G,B (Red, Green, Blue), so we may say, for example, that we have "displayed TM band 5,4,3 in R,G,B" (as the false colour display above).

3.7. Forest information from remote sensing data

Different remote sensing data sources are more or less appropriate given the application. To determine certain forest characteristics, you will need to choose one data source over another, and therefore should know the possibilities and limitations of different data sources. What can and cannot be determined, given the available data?

3.7.1. Optical 2D data

The main spectral components in a forest stand are sunlit canopy, shadowed canopy, sunlit ground and, shadowed ground. Pixels that have dimensions larger than a tree will have values made up of several trees, their shadows, and ground cover.

Concerning what type of reflectance might be expected, the Short-wave infrared band (SWIR) is most influenced by tree shadows, which makes it useful for determining characteristics (particularly) for coniferous forest. If using satellite data with visible and near infrared as well as shortwave infrared wavelengths, data from the blue, green, red, and the SWIR bands will be highly correlated with each other. The darkness seen in the pixels in these bands over forest areas will be mostly a function of the size of the trees and the stem number, since shadows are a dominant factor behind the reflectance.

The SWIR bands are the ones that are best correlated with forest biomass and thus also with forest stem volume. However, regardless of the choice of spectral band, the images will not be much darker when the stem volume becomes more than 300 m^3/ha. This is because the canopy closes, and growth often occurs for the stem, which cannot be detected with optical sensors, which look primarily at the surface/canopy. We say that the signal is "saturated" at this point, and doesn't provide the detail of information. In a similar way, the general dependence between stand age and spectral signature becomes weak after the first thinning cutting.

The Near Infrared (NIR) band is more correlated with the structure of the leaves. NIR reflectance increases sharply when vegetation is photosynthetically active. Deciduous forest and grass have higher reflectance in this band than coniferous forest, or the reflectance coming from the field layer with lingonberry or blueberry.

The contrasting information in NIR and SWIR/Visible bands often makes a combination of these most useful. A difference, or ratio, measure between the Near Infrared band and the red band, (as well as other combinations of near infrared and the sum of the other bands), is well correlated with the photosynthetic activity.

Altogether, spectral data from optical satellites over boreal forest do not contain much more than two dimensions of uncorrelated data (as stated above: the NIR and all other bands), but since many forest variables also are correlated and related to different tree size variables, estimates of many more than two forest variables are often presented in remote sensing studies.

The thermal band, particularly acquired from satellite platforms, is influenced by factors that are difficult to control, like moisture, and is seldom used for forestry studies.

The ability to separate and classify coniferous and deciduous forests is well established in the remote sensing literature, as deciduous leaves reflect more

of the NIR wavelength than coniferous trees (Lillesand et al., 2008). Mixed deciduous/coniferous forests are more difficult to classify correctly due to the spatial arrangement and uneven mixtures of the deciduous and coniferous species within pixels. For boreal and managed coniferous forests, optical satellite data have a primarily negative correlation with wood volume in the visible and mid-infrared bands. Correlation with the near- infrared band is more varied and sometimes there is no correlation. Forest canopy self-shadow has a large effect on the spectral response, yet it can be used to help derive stand parameters. For Swedish forest conditions, shadow is particularly important due to several factors including that coniferous tree species' crowns tend to cast more shadows than deciduous species and that the low sun angles occurring at Sweden's relatively high latitude produce more shadow and illuminate ground vegetation less. The SWIR bands have been shown to be of significance in forest parameter estimation, most likely due to their sensitivity to shadow patterns.

As forests mature, the canopy tends to close which results in a weakening ability to determine forest parameters from optical satellite data. Trotter et al.[3] found poor correlation between the SWIR band (TM band 5) and plantation forests, however, this was most likely due to the lack of complexity and shadows within the even-height plantation forests. The correlation between spectral data and wood volume tends to be stronger for younger stands than older stands. Horler and Ahern[4] found Landsat's two SWIR bands to be the most sensitive to forest vegetation density, particularly in the case of regenerating forest stands.

Clear-cuts are easily identifiable, with increased reflectance from the visible and SWIR bands and changes in the NIR reflectance (NIR may increase or decrease depending on the management activities and ground/field layer reflectance [5]). Determination of wood volume using Landsat and SPOT data has been hampered in several studies by the limited dynamic range of the spectral data, as reflectance of forest tends to be relatively low.

In an article by Nilson et al.[6], results are presented from work with computer models of radiation in forest canopies, and verified with radiometer measurements in field and satellite data. The work has been carried out by a leading group at Tartu Observatory (partly in co-operation with SLU). According to the research presented in the article, the driving factors behind the reflectance from forest canopies are canopy closure, leaf area index, tree species composition, and the understory vegetation. Unfortunately, these factors are only partly correlated with forestry relevant information like stem volume, diameter and stem number. It is also stated in the article that the spectral signatures not changes much after 30-40 years.

[3] Trotter, et. Al., 1997.
[4] Horler and Ahern, 1986
[5] Olsson, H., 2009
[6] Nilsson et al., 2003

Satellites which take frequent spectral images (e.g., an image every day, or every 5 days) create the opportunity to use phenology information to help identify different species. An example is that tree species which generally leaf-out earlier in the spring than others, can potentially be identified using this information. In other words, the difference in the development of vegetation types can give us yet another dimension of information to use in their identification.

3.7.2. Three-dimensional data sources

The most important data source to be introduced in the past decade for characterizing the forest has been three-dimensional data from airborne LiDAR. LiDAR allows a very accurate measurement of object locations (x,y) and height (z). For this reason, LiDAR is extremely useful for measuring of tree height, which in turn can be used in allometric equations for determination of other height related forest variables such as wood volume or stem diameter. Due to LiDAR's ability to determine x,y position as well as penetrate the canopy, vegetation density can also be determined. The measure of vegetation density can be dependent on the number of laser hits per m^2 and also the characteristics of the vegetation.

LiDAR is currently often acquired using a laser in the near-infrared wavelength, and reflectance intensity is a parameter recorded in the returned data. However, due to a lack of information about the processing of these data (often due to company secrets), they often cannot be used to determine information about the vegetation types. Multi-spectral LiDAR are currently being introduced, and with further research, the multi-spectral information may be used to determine both forest structure characteristics accurately (height, density), as well as vegetation information (vegetation type).

It is not uncommon to combine the spectral information from satellite images with three-dimensional information from LiDAR, image-matching, or radar, in order to get the optimal information from each system.

3.8. The issue of scale

The importance of scale issues deserves special. Satellite data are often categorized by the pixel size (i.e., spatial resolution):

- Low or coarse resolution: 200-1000 m pixel size
- Medium or moderate resolution: 10-200 m,
- High or fine resolution: 1-10 m,
- Very high: < 1 m.

Landsat, Sentinel-2 and SPOT are examples of moderate resolution sensors, while Pléiades, PlanetLabs, World-View, and most images from aerial photos are considered very high resolution. Specific satellite sensors are described more in their respective following chapters.

Strahler *et al.*[7] described two different types of models representing the interaction between satellite spatial resolution and the scale of the objects being observed in the landscape. These were *H*-resolution and *L*-resolution, in which *H*-resolution image pixels are smaller than the objects observed, and *L*- resolution where the objects are smaller than the image pixels. As an example, for forests, *H*-resolution may translate to "several pixels per tree", while *L*- resolution translates to "many trees per pixel."

In general, coarse resolution (or *L*-resolution) satellite data will have more pixels containing mixtures of cover types, which can make estimation of vegetation parameters more difficult. *H*-resolution imagery does not however assure higher classification accuracies, as aggregated spectral information of the landscape may be necessary for accurate classification. Several studies have shown decreasing thematic classification accuracy with increasing spatial resolution. This may occur, as an example, because not only the spectral reflectance from tree crowns but also the shadow cast by the tree crowns is an important source of information for the estimation of forest parameters. A moderate resolution pixel will capture both tree crown and its associated shadow. Higher spatial resolution data may provide more thematic detail, but may perhaps not have all the information needed (e.g., shadow and tree). In this case different analysis approaches may be used (e.g., Object-based classification). In addition, the trade-off in pixel size is generally paid for with a smaller scene area coverage, resulting in a potentially more costly and complex mapping project.

The processing algorithms required to produce map information are likely to differ based on the spatial resolution of the data. For example, identifying tree cover with very high resolution data may require an aggregation of pixels into tree crowns (e.g., segmentation or so-called "Object-based classification"), while coarse resolution data require a breaking-down of the information in the pixel (e.g., spectral unmixing). The production of global land cover data often requires the use of coarse resolution data, which has stimulated the use of sub-pixel estimation methods to capture information about the heterogeneity within the larger pixel.

Different landscapes present different levels of heterogeneity and transitions between land cover types can be distinct or fuzzy, and hard to define even in the field. The spatial composition of the landscape structure, such as the forest stand sizes present, and the diversity within it such as presence of elements like bedrock outcrops, wetlands, water bodies, and roads, will have an effect on the result. The properties of the landscape and the goals of the mapping project exert important influences on the appropriate choice of remotely sensed data that have an "optimal spatial resolution"[8]. Some research has been done on determining "optimal pixel size", while others have proposed using multiple-scale remotely sensed data for land cover classification.

[7] Strahler, 1986
[8] Woodcock *et al.*, 1988

Study questions

1. What are the differences between active and passive remote sensing systems? Give an example of eath sensor type. What are the advantages or disadvantages of active and passive systems?

2. Give an example of platforms used for remote sensing of forest.

3. Under what conditions can a drone/UAV be an appropriate platform for collection of data?

4. Give an example of when a multirotor drone/UAV should be used instead of a fixed-wing UAV.

5. If you have a satellite image with 10 x 10 m pixels (e.g., Sentinel 2), how appropriate aret hese data for a) delineating forest stands? b) finding new clearcuts? c) finding new thinnings? d) estimating high stem volumes?

Further Reading

Horler, D. N. H. & Ahern, F. J. (1986) Forestry information content of Thematic Mapper data, International Journal of Remote Sensing, 7:3, 405-428, DOI: 10.1080/01431168608954695

Nilson, T. et al. (2003). Forest reflectance Modeling: Theoretical Aspects and Applications, Ambio 32:535-541

Olsson, H (2009). A method for using Landsat time series for monitoring young plantations in boreal forests, International Journal of Remote Sensing, 30:19, 5117-5131

Strahler, A., et al., (1986). On the nature of models in remote sensing. Remote Sensing of Environment; p. 121-139

Woodcock, C. E., Strahler, A. H., & Jupp, D. L. (1988). The use of variograms in remote sensing: I. Scene models and simulated images. *Remote Sensing of Environment*, *25*(3), 323-348.

4. OPTICAL SATELLITE DATA

Classification scheme. A classification scheme consists of the class names and definitions of the land cover types or characteristics to be included in a thematic map. The classes and definitions should be mutually exclusive (i.e., it should be possible for each pixel to be logically assigned to one and only one class according to the definitions).

Copernicus program, including the Sentinel program. The European Union launched Copernicus, which is an ambitious program for earth observation. It consists of several series of earth observation satellites, including the Sentinel series of satellites. Sentinel-2 is a medium resolution optical satellite sensor.

Field of View (FOV). A specification of a sensor that results in the imaging of a given distance on the Earth's surface (from east-to-west). Satellites often register images in a continuous strip whose swath width is determined by the sensor's Field of View. The FOV often determines the maximum size of the image when it is delivered (i.e., image extent).

Ground Sampling Distance (GSD). Distance on the ground between the center of two adjacent pixels. The GSI is the pixel size of the un-processed image.

Instantaneous Field of View (IFOV). Not to be confused with FOV, IFOV is the area sensed by each detector element, and is closely related to the concept of pixel size.

Land cover and Land use. Land cover refers to the actual cover on the ground, while Land Use refers to the use. An example is "Wheat" would be the Land Cover class, while "Agriculture" is the Land Use class.

Minimum Mapping Unit (MMU). The MMU is the smallest area of a class to be included in the map product.

Path/Row. A common identifier for satellite images is through use of an indexing system (e.g., the World Reference System or WRS), often containing a path number and row number – in particular for the Landsat data series. The identifier system is unique for each satellite program.

Pre-processing of images. Images from optical satellite sensors may require work before analyzing them. Such corrections may be geographic (registration to geographic coordinates), atmospheric (correcting for haze or removing clouds and cloud shadows), and topographic (correcting for sun angles and differences due to topography).

SPOT eller Satellit Pour Observation de la Terre. The French sensor that spanned from SPOT-1 to -7.

Thematic Mapper (TM). The sensor aboard Landsat 4 and 5, which with its repeat coverage and 28.5 m ground resolution cell, made the optical satellite data from Landsat TM widely used. ETM+ refers to the Enhanced Thematic Mapper sensor on Landsat 7, while OLI refers to the Operational Land Imager on Landsat 8.

When the term "optical" remote sensing data is used, it is meant that the sensor operates within the optical spectrum (visible, near-infrared, and/or short-wave infrared), and use a type of lens or electro-optical device to acquire the data. Longer wavelengths, requiring additional equipment to obtain the data, such as antennas, are not included in this definition.

The spectral data recorded by these sensors are often of the vegetation or other objects on the Earth's surface. By analyzing characteristics of the spectral reflectance, land cover types and their characteristics can be identified and mapped.

4.1. Introduction to optical satellite data

Operational acquisition of optical satellite data has been taking place since the 1970s. Sensors often record information in the visible and near-infrared spectrum, and may include information in the short-wave infrared and thermal infrared spectrum. There are a number of different satellite data programs which provide data to the user community, and these are described in this chapter.

Countries often build their own sensors, where the characteristics of the sensors are often decided based on what is needed, what is possible, what has historically been used, costs, and market niches. Satellites may be state-owned and the data may be distributed free or at a cost, or corporately owned, where data are distributed often at a cost. The building of sensors and the cost of state-owned data has historically been politically influenced, yet there is currently a trend toward a free-data policy, as the use of data has been seen as beneficial to science and the economy. Since each satellite has the potential to view the whole globe, and since many countries want to have the technical capability to operate remote sensing satellites, but few countries are willing invest in redundant monitoring systems, there is a large need for international cooperation where satellite operators coordinate their plans. Organisations for such coordination are CEOS, and on a more general level also GEO.

The basic characteristics of remote sensing data have been described in the previous chapters (e.g., Chapter 3) but are shortly repeated here. Recall that important properties optical satellite data include spectral resolution, the number of wavelengths, spatial resolution, radiometric resolution, temporal resolution, and image extent. Recall also that satellites have orbits that can vary in characteristics. They can have a polar or geo-stationary orbit, and they can have different heights above the earth's surface.

4.3.1. Common optical satellite data programs

4.3.1.1. The Landsat program

Landsat 1-3

In the 1960s and 70s, for the persons that knew about military satellites, and for the larger number that had insights in the state of the art technology used to map the surface of the moon before the Apollo landings, it became more and more frustrating that this technology could not be used to map the resources of the earth as well.

To meet this criticism, the Landsat program was created. The Landsat 1 satellite was launched in 1972 (Table 4.1). It had a multispectral scanner (MSS) that produced digital images with 79 m pixels (often rounded up to 80m) in four spectral bands (nominally Green, Red, NIR1 and NIR 2).

The reason why the counting started on MSS 4 was that the first three bands were devoted to another sensor, a TV camera-like system called Return Beam Vidicon. These data are seldom used, whereas the MSS data can be used as an early reference for land cover change studies.

Landsat 1 was followed by the quite similar Landsat 2, launched 1975 and Landsat 3, launched 1978. MSS is still the best source for old satellite data before 1984.

Landsat 4-8

The US Landsat program became a success story, and Landsat 4 that was launched 1982, had an improved sensor named *Thematic Mapper, with six reflective bands* from blue to mid infrared, with 30 m pixels, and a Thermal band with 120 m pixels. The Thematic Mapper (TM) had better spatial resolution (30 m pixels) in six reflective bands, nominally in blue, green, red, NIR, SWIR1 and SWIR2. In addition, the TM sensor had a thermal band with 120 m pixel size: TM 6. The reason why the order between TM 7 and TM 6 doesn't follow the order of the wavelength was that TM 7 was decided at a very late stage. This numbering has then been kept on the following TM sensors.

Landsat 4 soon got technical problems, but the similar Landsat 5 was launched 1984 and worked for more than 25 years, which is exceptional, since the design lifetime for a remote sensing satellite is normally about 5 years. Thus, Landsat 5 Thematic Mapper (TM) data is a very important source of old and current satellite data. Landsat 5 was collecting imagery into the year 2011 when it developed more problems, requiring that its operation be turned to an extremely minimum (and non-operational) level.

Landsat 6 was launched by a private company, but never reached orbit. It's said to have crashed into the ocean after launch.

Landsat 7, launched 1999, was again a US government owned satellite. The sensor was now called ETM+. The main differences were a panchromatic

band (TM 8) with 15 m pixel size, and thermal band (TM 6) with 60 m pixels. In May 2003 the Landsat ETM+ sensor developed a problem with its Scan Line Corrector, which resulted in images where the data had "line-drops". Although there have been many attempts to correct these data and fill them in with data from adjacent images, Landsat 7 could not be seen as an operational instrument after 2003.

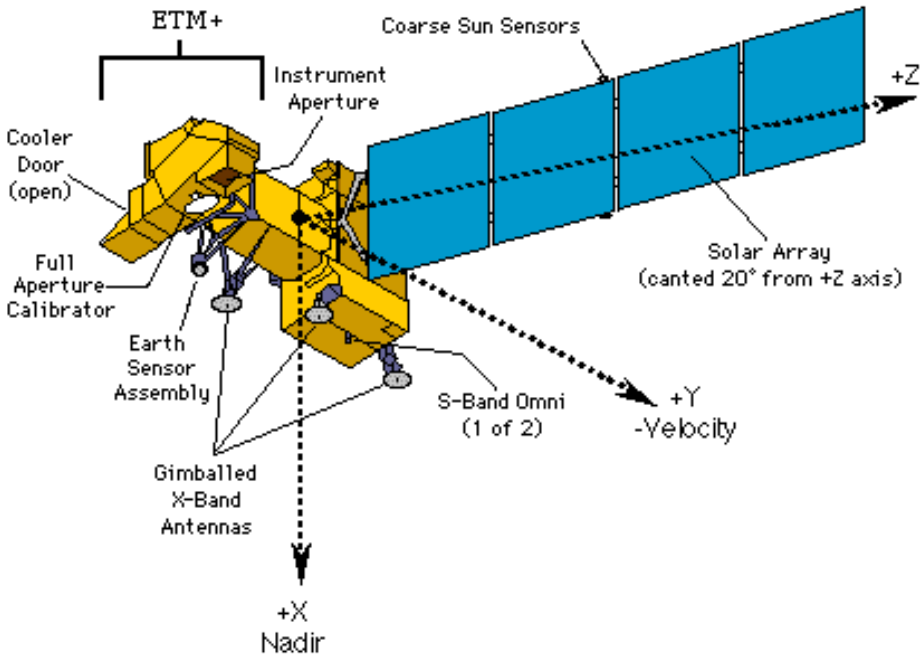


Fig 4.1 Landsat-7 satellite platform and its components.

As of 2012, neither Landsat 7 nor Landsat 5 were able to produce usable data. This resulted in what researchers call "the Landsat data gap", where the continuity of the Landsat program, operating since the 1970s, had been broken. The launch of Landsat-8 (also called the Landsat Data Continuity Mission or LDCM) occurred in April 2013, solving this problem. Landsat-8 has a push-broom sensor called OLI (Operational Land Imager) but with specifications much like the previous Landsat TM sensors, with the main difference being higher radiometric resolution, and the addition of a coastal-blue and thermal wavelength bands. A subset of a Landsat-8 image is shown in Figure 4.2.

Table 4.1. Satellites in the Landsat program

Satellite	Launched	Decommissioned	Main sensor	Bands and Pixel sizes
Landsat 1	1972	1978	MSS	4 bands with 80 m pixels
Landsat 2	1975	1982	MSS	
Landsat 3	1978	1983	MSS	
Landsat 4	1982	2001 (last data 1993)	TM	6 bands with 30 m pixels + one 15 m panchromatic + one thermal band
Landsat 5	1984	last data 2011	TM	
Landsat 6	1993	1993 (failed at launch)	ETM	
Landsat 7	1999	sensor problem 2003	ETM+	
Landsat 8	2013		OLI	8 bands with 30 m pixels + one 15 m panchromatic + two thermal bands with 100 m pixels

Table 4.2 Band specifications for Landsat 8

Bands	Wavelength (micrometers)
Band 1 - Coastal aerosol	0.43 - 0.45
Band 2 - Blue	0.45 - 0.51
Band 3 - Green	0.53 - 0.59
Band 4 - Red	0.64 - 0.67
Band 5 - NIR	0.85 - 0.88
Band 6 - SWIR 1	1.57 - 1.65
Band 7 - SWIR 2	2.11 - 2.29
Band 8 - Panchromatic	0.50 - 0.68
Band 9 - Cirrus	1.36 - 1.38
Band 10 - Thermal Infrared (TIRS) 1	10.60 - 11.19

Band 11 - Thermal Infrared (TIRS) 2	11.50 - 12.51

Figure 4.2 – Subset of a Landsat-8 image 2014/08/13 over Umeå shown as NIR (band 5), Red (band 4), Green (band 3) in RGB display.

4.3.1.2. The SPOT program

The first satellite in the French SPOT satellite series, partly owned also by Sweden, was launched 1986. SPOT had a more modern push-broom scanner, based on a CCD-array, and better spatial resolution (10 m panchromatic pixels and 20 m colour). The SPOT program was an attempt to create a commercial market, and there have been seven SPOT systems launched. Data for the government owned SPOT 1-5 is summarized in Table 3.2. The SPOT satellites have now been taken over by the company Airbus and the recent SPOT 6 and 7 belongs to the very high resolution VHR category of commercially operated satellites and is therefore listed in Table 4.4 instead. The SPOT satellites have two push-broom sensors that are pointable sidewise. The original HRV sensors could deliver either a 20 m pixel IR colour image based on three spectral bands (Green, Red, NIR), or a panchromatic image with 10 m pixels. The HRVIR sensor onboard SPOT 4 had a mid-infrared band as well (which is good for forestry). The HRG sensors onboard had better geometric resolution (10 m colour, 5 m panchromatic). SPOT 5 also carries the HRS, a 3-line scanner for along track stereo. A subset of a SPOT-5 image is shown in Figure 4.3.

Table 4.2. Satellites in the SPOT program

Satellite	Launched	Main sensors	Bands and pixel sizes
SPOT 1	1986	2 HRV	3 multispectral bands green - NIR with 20 m pixels or 1 panchromatic band with 10 m pixels
SPOT 2	1990	2 HRV	
SPOT 3	1993	2 HRV	
SPOT 4	1998	2 HRVIR	3 multispectral bands green – SWIR with 20 m pixels and the red band with 10 m pixels
SPOT 5	2002	2 HRG, +1 HRS for stereo images	3 multispectral bands green – NIR with 10 m pixels and one SWIR band with 20 m pixels Pixels. One panchromatic band with 2.5 – 5 m pixels.

Table 4.3 Band specifications of SPOT-5

Bands	Wavelength (micrometers)
Band 1 - Green	0.50 – 0.59
Band 2 – Red	0.61 – 0.68
Band 3 – NIR	0.78 - 0.89
Band 4 - SWIR	1.58 - 1.75
Panchromatic	0.48 - 0.71

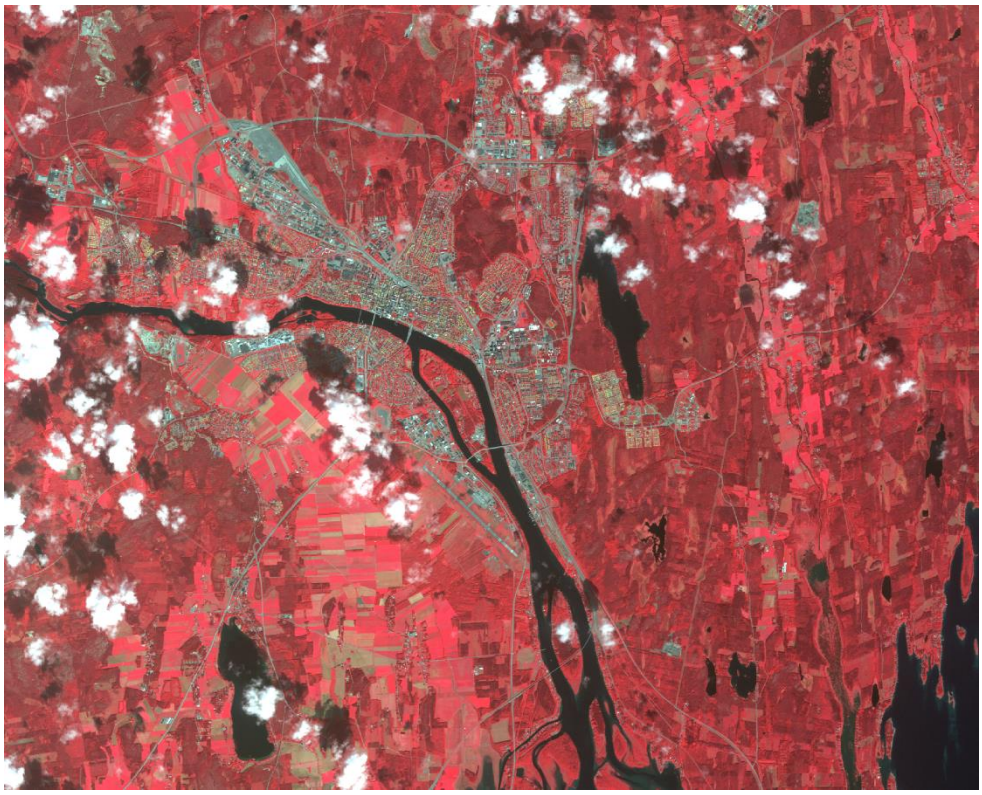


Figure 4.3 – Subset of a SPOT-5 image 2013/08/16 over Umeå shown as NIR (band 3), Red (band 2), Green (band 1) in RGB display.

4.3.1.3. The Sentinel Program

The European Union has launched Copernicus, which is an ambitious program for earth observation. It consists of several series of earth observation satellites, as well as connection to *in situ* observations and operational services. The satellite series are called Sentinels, and are developed by the European Space Agency (ESA). Of most interest for forestry is data from Sentinel 2 which is a series of Landsat-like satellites. Sentinel 2a was launched 2015 and Sentinel 2b is planned for launch 2017.

The sensor has 13 bands with 10, 20, or 60 m pixel size, depending on wavelength (Table 4.4). The field of view is 290 km, which enables a Sentinel 2 satellite to pass a given point every 5 days, or even more frequently at high latitudes, like Sweden. The data are free and data over Sweden will be obtainable from a national archive. Thus, the Sentinel 2 data should be of high interest for forestry. A subset of a Sentinel-2 image is shown in Figure 4.4.

Table 4.4. Spectral bands provided by the Sentinel 2 satellites

Sentinel-2 Bands	Central Wavelength (µm)	Resolution (m)
Band 1 - Coastal aerosol	0.443	60
Band 2 – Blue	0.490	10

Band 3 – Green	0.560	10
Band 4 – Red	0.665	10
Band 5 - Vegetation Red Edge	0.705	20
Band 6 - Vegetation Red Edge	0.740	20
Band 7 - Vegetation Red Edge	0.783	20
Band 8 – NIR	0.842	10
Band 8A - Red Edge	0.865	20
Band 9 - Water vapour	0.945	60
Band 10 - SWIR – Cirrus	1.375	60
Band 11 – SWIR	1.610	20
Band 12 – SWIR	2.190	20

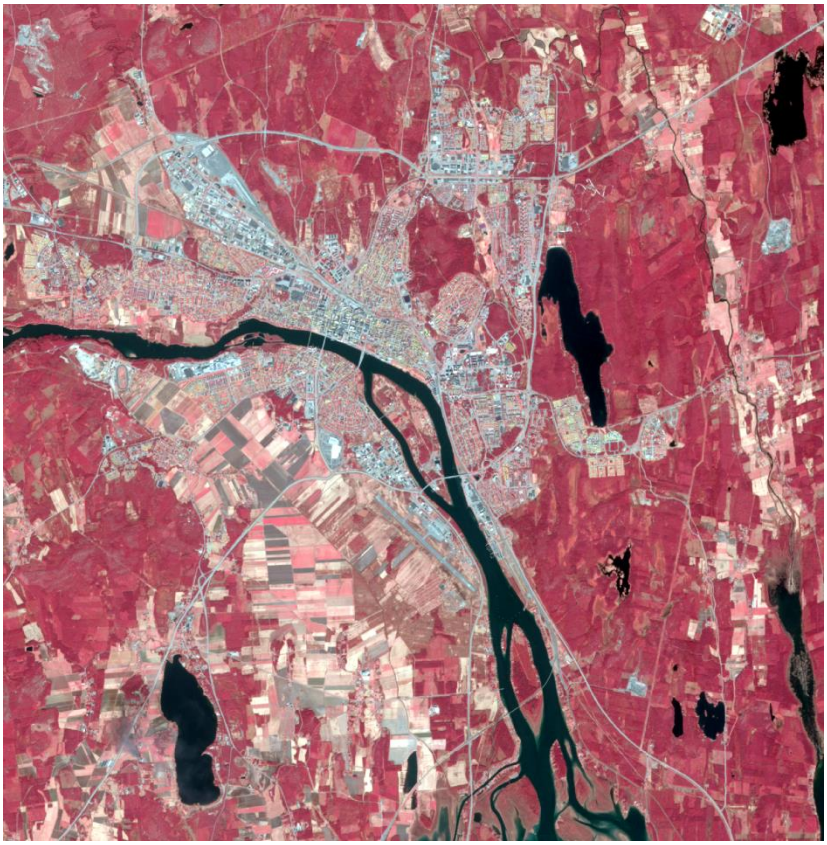


Figure 4.4 – Subset of a Sentinel-2A image 2016/05/09 over Umeå shown as NIR (band 8), Red (band 4), Green (band 3) in RGB display.

4.3.1.4. Very high resolution satellites

Remote sensing satellites with sensors that produce images with much higher spatial resolution than Landsat, and later SPOT, were from about 1960 to the beginning of the 1990's only available for military intelligence. This changed however with the end of *the Cold War* and the collapse

of the Soviet Union. Russian space authorities then started to sell spy satellite images on the open market. The pixel size had been degraded to 2 m, but this was still much better than any other satellite images available on the open market at that time.

The USA responded to this in two ways. By releasing their old CORONA images for sale on internet, and by *licensing out the right to build "civilian spy satellites"* to private companies. The first of these very high resolution satellites was IKONOS, that was launched by Lockheed 1999 and produced panchromatic images with 1 m pixel size and IR colour images with 4 m pixels. Later in 2001 followed Quick Bird that produced images with 0.5 pixel size in the panchromatic band. There has since been many follow on satellites of similar type.

Today, there are many nations that have launched remote sensing satellites which produces images with pixel sizes in the order of 1m. Governments are willing to invest in this technology, much for military reasons, but since it is not secret any more, they are often also willing to sell images on the open market as well, this strategy is called *dual use*. Thus, the "very high resolution segment" can be regarded as stable. Table 4.4 lists some of the more important high resolution satellite sensors.

These very high resolution satellites produce imagery that is almost like orthophotos from aerial cameras. Each image also covers however a very limited area (often in the order of 100 to 400 km^2). Thus, they compete mostly with aerial photos and it is difficult to obtain cloud free data for large areas like whole countries. The market for very high resolution satellites is probably limited in countries with a regular supply with governments subsidized air photos, such as Sweden.

The WorldView-4 sensor from the company DigitalGlobe has 31 cm pixels in panchromatic and 124 cm pixels for multi-spectral data. Currently Google Earth often uses WorldView images when zooming in to high resolution (and often Landsat or other medium resolution data are used at less detailed levels).

Table 4.5. Some very high resolution satellite programs that produces imagery for the civilian market.

Satellite	**Year Launched**	**Mulispectral bands**	**Mulispec bands pixel size**	**Pan band pixel size**	**Image Swath**
IKONOS	1999	B, G, R, NIR	3.2 m	0.8 m	11,3 km
QuickBird	2001	B, G, R, NIR	2.4 m	0.6 m	16,5 km
WorldView-1	2007	-		0.46 m	17.6 km
WorldView-2	2009	B, G, R, NIR + 4 more	1.84 m	0.46 m	16.4 km
WorldView-3	2014	28 bands 400-2365nm	1,24-30 m	0.31 m	13.1 km

WorldView-4	2016	B, G, R, NIR	1,24 m	0.31 m	13.1 km
GeoEye-1	2008	B, G, R, NIR	1.65 m	0.4 m	15.2 km
RapidEye (5 satellites)	2008	B, G, R, NIR + 1 "red edge"	6.5 m	-	77 km
Pleiades 1A	2011	B, G, R, NIR	2 m	0.5 m	20 km
Pleiades 1B	2012	B, G, R, NIR	2 m	0.5 m	20 km
SPOT 6	2012	B, G, R, NIR 6 m		1.5 m	60 km
SPOT 7	2014	B, G, R, NIR 6 m		1.5 m	60 km

4.3.1.5. Microsatellites

One of the latest developments is coming from PlanetLabs, which is sending up "shoebox satellites" or "*microsatellites*", which are very small, inexpensive satellites. PlanetLabs aims to send up 100-200 shoebox satellites so that they can achieve high temporal resolution, namely to cover every spot on the Earth, every day. The data, provided by a commercial company, are not free. Another such initiative is called "BlackSky" which plans to send up 60 satellites for multispectral data with 1 m pixels. We can expect to see more of this type of development in the future. In fact, it is estimated that more than 900 remote sensing satellites are planned to be sent up between 2017 and 2021, and many of these will be microsatellites.

How many satellites are there in orbit?

There are all types of satellites, some of them used for communication, and some for Earth Observation (Fig 4.5). As of 2014 there were approximately 192 Earth Observation satellites in orbit, not all of which were functioning[1]. As of 2017, there was a very large increase of EO satellites, with a total of 620![2] One reason for this increase is large numbers of microsatellites being launched.

Figure 4.5. Depiction of objects orbiting in space around the Earth, courtesy of Wikimedia Commons (D. Shikomba & T. Haiduwa).

[1] https://www.pixalytics.com/how-many-eo-space/
[2] https://www.pixalytics.com/eo-sats-in-space-2017/

4.3.2. Satellite scenes and image extents

Satellites often register images in a continuous strip (or "*swath*") whose swath width is determined by the sensor's Field of View (FOV). Satellite images are then created by taking (traditionally) a square image from the swatch. Most satellite systems have a *World Reference System* (WRS) that assigns a unique number to each potential satellite scene, often based on a Path (north to south) and Row (east to west) number (Fig 4.6). Often, the WRS identifier along with the date of acquisition makes up the file name of the satellite image.

Examples are

- Landsat: the Landsat image center closest to Umeå is from Path 193 and Row 15, and is referred to as 193/15. Landsat images have a fixed WRS identifier (the same area will always have the same WRS number). The image extent is 185 × 185 km.
- SPOT: these images in contrast do not have a constantly fixed area, and the WRS number is approximate. The use of oblique viewing with the pointable sensor may result in off-nadir images, and the image size will be variable depending on the viewing angle. The nadir-viewing image extent is 60 × 60 km
- Sentinel-2: the Sentinel-2 swath width is 290 km, however, due to the size of the images, the data are delivered in 100 km × 100 km "granules".

Fig 4.6 – The Landsat WRS-2 path and row system for identification of images. Image outlines are shown in white, and paths overlap.

Getting the huge amounts of image data from the satellite down to the earth (so-called "downlinking") was a bottleneck for satellite remote sensing. The early spy satellites used film capsules sent down with parachutes. Later earth resource satellites with digital sensors (such as the early Landsat satellites) downlinked data to a few receiving stations with large parabolic antennas. Some remote sensing satellites, since Landsat 4, had the possibility to send the data to a central receiving station that might be on the other side of the earth, by using a telecommunication satellite in geostationary orbit as a link.

The network of receiving stations has successively become denser and denser, but onboard recording of data is still important. For downlinking of data and also for sending control commands to the polar-orbiting satellite, it is an advantage to have receiving stations with a northern location. Two such stations are Esrange outside of Kiruna and the Norwegian station at Svalbard that in 2010 had 18 parabolic antennas for receiving satellite data.

4.2. Geometry of a raster satellite image

Geometric correction is the process of transforming image pixel positions to account for the optical geometry and compensate for various distortions caused by the motion of the image platform and shape of the terrain.

All of the image distortions that can be modeled or measured are collected together in a mathematical equation to compute the true position of each image pixel. This equation may include residual distortions that can be calculated with the aid of a number of *ground control points* (GCPs). The image is then re-projected to the proper location by changing the location and arrangement (and therefore the DNs) of the pixels in a process called *resampling*.

Georeferencing is the process of assigning map coordinates to an image. You may see several different names in the literature for this process, such as "geocoding". A map coordinate system has both an Earth model (also called a geodetic datum, an example is WGS84), and a projection (e.g. a Transverse Mercator). There is also another term called *rectification,* which involves the assigning of map coordinates to an image, but also changes the shape of the raster, possibly the pixel size, as well as possibly values within the pixels. This is done through the process called *resampling.*

There are many sources of geometric distortion in satellite image data. Some examples are

- Earth rotation - The earth rotates by a fixed amount from east to west during the recording of an image frame.
- Panoramic distortion - Pixels at the edges of the image are larger and more spaced apart than those at nadir.

- Platform motion - Changes in altitude, velocity, pitch, roll, and yaw, affect the position and/or scale of each pixel.
- Terrain variation - Higher elevations give effectively smaller pixel sizes.

Raw images are acquired and stored in a rectangular grid without any information about the ground location. The true position of each image pixel can be determined only if information is available for:

- The optical geometry (IFOV, FOV)
- The platform position (X, Y, Z)
- The platform attitude, or orientation (roll, pitch, yaw)
- The terrain elevation (DEM)

4.2.1 Geometric correction levels

Geometrically "raw" images are sensor data, essentially without any geometric correction, they are used only for further processing in other systems;

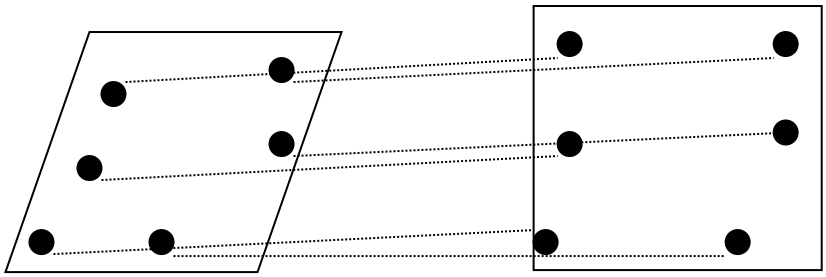
Geometrically system corrected images means that the pixels have been adjusted to have coordinates in a map projection, using what is known about the image geometry from the satellite system itself (e.g. GPS coordinates for the satellite orbit when the image was registered). This is the most common form of delivery from distributors. Note that geometrically system corrected data acquired on the international market is seldom resampled to the map projection and geodetic datum used in a particular country (e.g. Sweden).

Geometrically precision corrected images are those that have been geometrically corrected to a map projection with high accuracy, using both the mathematical models applied in the system correction, and information from GCPs. The geometric precision correction is often done with advanced computer programs that model the orbit and movements of the satellite to reconstruct correct map geometry. A standard for acceptable geometric error is less than half the pixel size.

Geometrically terrain corrected or *precision terrain corrected* or *orthorectified* images use the above process but also involves using a DEM in the correction. This may be especially necessary in mountainous areas with a lot of distortions due to large elevation differences in the image.

While working with satellite data, it is most convenient to order geometrically precision corrected images delivered in the same co-ordinate system as other data used in the project. However, sometimes only system corrected data is available and it might be necessary for a user to project the

satellite data into the desired map projection. A common way to do this is to develop a pair of polynomial equations that express geographic coordinates (x, y) in terms of the uncorrected image pixel positions (u, v). The coefficients of the polynomial equations can be estimated by collecting a number *Ground Control Points* based on coordinate pairs for identical locations in the uncorrected image and the map.



Uncorrected satellite data
with known coordinates (x, y)

Geometrically corrected data
(map or satellite data) with pixel
coordinates (u, v)

A second-order mapping polynomial is generated automatically by the computer and has the following format, which solves for the coefficients $\mathbf{a_i}$ and $\mathbf{b_i}$ using the GCP pairs and a least-squares fit.

$u = a_0 + a_1x + a_2y + a_3xy + a_4x^2 + a_5y^2$
$v = b_0 + b_1x + b_2y + b_3xy + b_4x^2 + b_5y^2$

where x and y are the known coordinates in the already geometrically corrected data, and u and v are the pixel coordinates of the uncorrected satellite data. All of these (x, y, and u, v) are given, and you are solving for the coefficients a and b in these equations. This then becomes the equation used to make the geometric correction and to determine new "correct" coordinates (u', v') for the uncorrected satellite data.
This process is somewhat easier if another, already precision-corrected image over the same area is available. Then the GCP pairs can be collected from the two images and the map coordinates read directly from the corrected image. This is called *image-to-image geometric correction* and can in some systems be used automatically.

Once the mapping polynomials have been determined, a new image is created by a process called *resampling*. For each position in the new grid (x,y) the pixel value must be recalculated for the original image. This is determined by first calculating the u,v position from the mapping polynomials, then looking up the original pixel values in the uncorrected image at the u,v coordinates. The problem is the u,v coordinates will not, in general, fall on even pixel boundaries and you must decide how to calculate the pixel value at the non-integer location based on the closest surrounding pixels.

Three common resampling methods, used to decide how you will recalculate the pixel value, are available:

- *Nearest neighbour*: take the pixel value of only the closest pixel.
- *Bilinear interpolation*: take a distance weighted average of the surrounding 4 pixels.
- *Cubic convolution*: use the surrounding 16 pixels to calculate a weighted average for the new cell. The weights are determined from considerations in signal processing.

The resampling method used depends somewhat on the type of image and your objectives. Nearest neighbour resampling has the advantage of not altering original data values but may lead to jagged edges in linear features. Bilinear interpolation generally smoothes the output image, and cubic convolution often produces the highest visual quality. Generally, cubic convolution is preferable, since it handles the data most correctly from a signal processing perspective.

The number of GCPs needed for the geometric correction will depend on the order of the polynomial transformation, and on the terrain properties of the area in the image. Depending on the number of variables in the transformation, the minimum number needed is equal to the number of variables +1. If a 2nd order polynomial transformation is used, then a minimum of least 7 GCPs are needed (= the number of variables in the equation + 1).

There are some pixels in the image which make better GCPs than others. These features are seen easily in the imagery, and should be objects that will not change over time or between seasons. For example, water levels can often change and for this reason are not always good GCP sources. Road crossings are an excellent GCP source, as well as clearly identifiable parts of roads. GCPs should be taken over the entire scene, and not all gathered in one area of the scene. They should be spread out and have at least one GCP in each corner of the satellite data. If the GCPs are not spread out, the areas without GCPs will not be geometrically corrected well.

When ordering data, there are often different levels of geometric processing to chose from, referred to as Level 0 to Level 3.

- *Level 0* or Raw image data – As collected at the sensor (not usually distributed)
- *Level 1A* – Corrected for detection variations within the sensor – so this is actually a *radiometric correction*
- *Level 1B* – Corrects for errors such as non-uniform pixels sizes or gaps in scan-lines due to the factors above. Also called *Path-oriented correction.* However, position accuracy is still not correct.

- *Level 2A* – Image is referenced to a cartographic map projection based on a prediction. Location errors can still be quite large. Often called a Geo-referenced or Geometrically system corrected mage.
- *Level 2B* - Through a process called geometric correction or image rectification, the image is "registered" to known coordinates (e.g., via an existing base map) known as ground control points. This is called a Precision corrected image.
- *Level 3* - geometric terrain corrected or orthorectified images use the above process but use not just the x,y coordinates, but x,y and z from a DTM in the correction.

4.3 Basics in image interpretation

In the following section, we give a brief overview of the process of analyzing optical satellite data to create information on forests and other land cover. More detail about analysis methods is given in Chapter 9.

When beginning a project, there can be a number of steps involved, which we have summarized as the following:

- Project goals;
- Determine data sources (remotely sensed data as well as reference data);
- Pre-processing of remotely sensed data;
- Decide analysis method;
- Analysis;
- Initial control of the map product;
- Post-processing of product; and,
- Final accuracy assessment

4.3.1. Project goals

The final goal of a project is often known at the outset of the project. It may be a thematic map of land cover types, or a wall-to-wall raster of stem volume, or large area detection of changes (e.g, clear-cuts). External data users may order a product, and turn to remote sensing experts to deliver this. It is up to the remote sensing expert to know whether the data at hand can meet the goals and costs of the project. The first step is to identify the goals, the potential data sources, the time and cost, and estimate the potential accuracy (based on previous studies) of the outcome.

If you want to create a thematic map, with a number of land cover classes, then you will need to develop a *classification scheme*. A classification scheme consists of the class names and class definitions. The classification scheme should cover all possible cases and be mutually exclusive so that every case should belong to one and only one class.

The project may develop its own classification scheme, or use existing ones. Examples of existing ones are the FAO classification system, the Anderson system, or the ISO system, among others. It is often useful if classification schemes are hierarchical, and depending on outcomes, may be collapsed into broader classes if necessary.

An example of a classification scheme is given below, where the Forest class is an example of a hierarchical scheme. The forest classes may be defined by, for example, canopy closure percentage, mean forest height, and the percent of conifer and deciduous tree species. The defintions arespecified at the beginning of the project.

1. Artificial surfaces
2. Agricultural areas
3. Forests and semi-natural areas
    - 3.1 Forests
        - 3.1.1 Clearcut
        - 3.1.2 Young coniferous
        - 3.1.3 Young deciduous
        - 3.1.4 Mature coniferous
        - 3.1.5 Mature deciduous
        - 3.1.6 Mature mixed forest
    - 3.2 Shrub or herbaceous
    - 3.3 Open spaces with little vegetation
4. Wetlands
5. Water

It is important to distinguish here between *land cover* and *land use*. A classification scheme often consists of either land cover classes, or a mix of land cover and land use classes (often abbreviated LC/LU). Land cover refers to the actual cover on the ground, while Land Use refers to the use. An example is “Wheat” would be the Land Cover class, while “Agriculture” is the Land Use class. It is often more difficult to determine what the use for an area is as opposed to the cover.

It is also of importance to know what your *Minimum Mapping Unit (MMU)* in your final map product will be. The MMU is the smallest area of a class to be mapped. The MMU is dependent on the original pixel size of your data, and the ability to detect this phenomenon. For example, if you state that your final map MMU is 0.5 ha, then you will need to generalize your final map product to areas of this size. Any landscape object smaller in size than the decided MMU will be assimilated or ignored. The MMU and the properties of the landscape can affect reference data collection, post-processing actions, and accuracy assessment results.

4.3.2. Determine the data source(s)

Different remote sensing data are suitable to different tasks. A global coverage of forest? It will be important to use data with large scene extent,

and global coverage. A smaller area mapping of urban expansion? A long term availability of higher resolution data may be most suitable. A map over a rainforest area? Here it is not enough to know that a sensor exists, but whether usable data really exist over the area to be mapped. Currently Sentinel-2 is a free and high quality data source that is a natural first choice.

It's important to think about the most appropriate season or seasons (spring, summer, fall?) for the imagery. It is also important to chose images as free from clouds, cloud shadows, and haze as possible. When viewing imagery, note that while the blue band may have haze, the NIR bands may be much less influenced by haze. Cloud Cover percent limits can be set in search engines, but remember that 25% clouds may be covering only one corner of an image and not be distributed over the whole image (and hopefully not over your study area).

Often, a bigger challenge is to obtain the reference data needed for both interpretation and accuracy assessment of a mapping project. This is particularly true if the remote sensing data were acquired some time ago (i.e., earlier in time than the current date). Refer to the section on Reference Data to explore this issue more.

4.3.3. Consider a potential analysis method

There are many different methods used to process remotely sensed data into map data. The method chosen often depends on the input data available, and the desired result. Some analysis methods require more reference data than others – which is the reason to consider this step before reference data collection. some analysis methods are used only for creating thematic maps and others used only for creating continuous value maps. Some analysis methods must be chosen based on the characteristics of the data (e.g., coarse resolution vs high resolution data). Chapter 9 goes into more detail about the characteristics and uses of different analysis methods.

4.3.4. Reference data

Reference data are most often needed in some form to work with remote sensing data. The collection of these data is addressed in Chapter 10 on Reference data. Even reference data themselves require a pre-processing step, which might consist of controlling the quality of the reference data, forecasting or back-casting the reference data, determining the suitability, removing outliers, and perhaps adding more reference data if deemed necessary.

4.3.5. Pre-processing of optical satellite data

The radiance recorded by a satellite sensor is affected by several factors, including the atmosphere, topography, vegetation composition, solar illumination angles, and sensor characteristics. Optical satellite data have characteristics that require some pre-processing steps which other remotely sensed data do not. For example, due to the visible wavelengths that may be influenced by atmospheric haze, there may be a need to reduce atmospheric effects. Some of the common pre-processing steps, namely atmospheric correction and topographic correction, are mentioned below. The need of

doing these steps depends on the goals of the project, as well as time and cost limitations.

4.3.5.1. Factors affecting the DN values

The way we use optical satellite data in forestry is most often to create a model which relates the digital numbers (DN) to the target variables of interest from field-based reference data. The DNs are the values delivered in the satellite image product, and are related to the true reflectance from the objects plus including the influence of atmospheric attenuation (as explained in Chapter 2). There are computer programs for converting the DN values for each spectral bands to units of reflectance. In practice, this will be a linear transformation of the type:

$$\text{Reflectance} = a + b * DN$$

One advantage with this procedure is that reflectance is a physical unit that is more comparable between sensors, and with data in the literature than the DN values. However, since the atmospheric correction is just a linear transformation, it will not improve estimates of forest variables or classifications into discrete classes, when they are made by the aid of ground reference plots.

It is difficult to replace the need for field-based reference plots with a so-called “library of typical reflectance values”. One reason for this is that the estimation of the coefficients *a* and *b* in the above function seldom is accurate enough. The coefficients depend both on the atmospheric optical thickness on that actual day, and on the calibration of the sensor and both of these factors vary over time and are difficult to estimate accurate enough for the purpose of creating general relationships for forest estimation.

Furthermore, the reflectance from a forested pixel depends not only on the size and amount of trees, but also on the length of the shadows, and the status of the field layer vegetation. Both these factors vary considerably with the season of the year, which also complicates the use of reflectance as a general measure for forest status.

Factors such as atmosphere, sensor calibration, shadow length, solar angles and vegetation season, will also influence both scenes in a bi-temporal change detection. Thus, the best way to find changed areas is to normalize the DN values in both scenes to each other with statistical techniques rather than using physically based atmospheric correction.

The atmospheric optical thickness might vary over different parts of a satellite scene. Most atmospheric correction methods do not address variations over an image, but rather treat atmospheric haze as uniform over the area. You can best see haze by contrast stretching of the shortest wavelength (e.g. the blue band). It is best to totally avoid using scenes with visible haze for computer based analysis. It is wise to exclude the blue band for classification or estimations (if possible) for this reason.

4.3.5.2. Topographic normalization, C-correction

The topographic characteristics of an area, such as slope and aspect, in combination with the angles of the sun's position (typically the solar zenith and azimuth angles) result in illumination differences within a satellite image. Topographic correction is used to directly manipulated the DNs in the satellite image so that a vegetation class will have similar spectral response whether on a north facing slope (facing away from the sun) or on a south facing slope (towards the sun).

Topographic normalization methods are often grouped as either photometric/photometric-empirical which includes the semi-empirical corrections such as Minnaert correction[3], *b*-correction[4], and C-correction[5], or physically-based models[6]. Gu and Gillespie[7] suggested using different topographic normalization methods for forested vegetation, namely the more forest-appropriate Sun-Canopy-Sensor (SCS) correction. While for non-forested areas the C-correction or statistical-empirical correction is most appropriate. The SCS model was later modified by Soenen et al.[8] to include the *c*-parameter, called the SCS+*C* method.

A relatively simple method for topographic normalization is the C-correction[9][10]. This example may help understand the process of topographic normalization. The C-correction was developed to perform topographic correction for non-Lambertian reflectance behavior, as the cosine correction was suitable for Lambertian reflectors only[11]. It is based on the linear relationship between the spectral reflectance recorded for a pixel and the corresponding cosine of the solar illumination incidence angle, *i*. The cosine of *i* can be calculated as a function of the local terrain slope and aspect, as well as the solar illumination angles upon the surface at the time of satellite data acquisition (Eq. 1).

$$\cos i = \cos z * \cos s + \sin z * \sin s * \cos(\Phi_a - \Phi_n) \quad (1)$$

where *i* is the solar illumination incidence angle with respect to surface normal, *z* is the solar zenith angle, *s* is the terrain slope angle, Φ_a is the solar azimuth angle, and Φ_n is the terrain aspect angle (Fig. 4.7).

Linear regression with cosine of *i* as the independent variable and reflectance ($\hat{\rho}_{\lambda t}$ is the topographically-influenced (*t*) reflectance of band

[3] Smith, Lin, and Ranson 1980
[4] Vincini and Frazzi 2003
[5] Teillet, Guindon, and Goodenough 1982
[6] Shepherd and Dymond 2003, Soenen et al. 2008
[7] Gu and Gillespie, 1998.
[8] Soenen et al., 2005
[9] Tellet et al., 1986
[10] Reese and Olsson, 2011
[11] Teillet et al. 1982

λ) as the dependent variable (Eq. 2) is used to estimate intercept (b) and slope (m).

$$\hat{\rho}_{\lambda t} = b + m * \cos i \qquad (2)$$

The c-parameter is calculated as b divided by m (Eq. 3). The relationship between reflectance and cosine of i is wavelength dependent[12], therefore, a c-parameter is calculated for each wavelength band.

$$c_\lambda = \frac{b}{m}, \qquad (3)$$

The c-parameter is then added to the numerator and denominator of the cosine correction to form the C-correction equation (Eq. 4).

$$\hat{\rho}_{\lambda h} = \hat{\rho}_{\lambda t} \frac{\cos\theta_z + c_\lambda}{\cos i + c_\lambda}, \qquad (4)$$

where $\hat{\rho}_h$ is the topographically normalized reflectance (h indicating "horizontal") for band λ, and c_λ being the c-parameter for band λ. According to Teillet et al.[11] and Meyer et al.[13], the c-parameter is used as an approximation of diffuse sky irradiance.

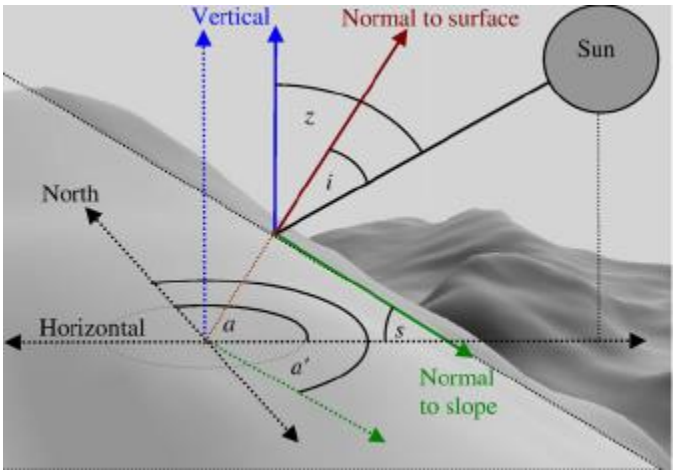


Figure 4.7 Angles used in C-correction (from Reese and Olsson, 2011).

4.3.6. Analysis

See Chapter 9 for more details on the analysis phase of an image processing project.

4.3.7. Initial control of the map product

Once you have run the first iteration of your classification, you may be excited to check the result. It's likely though that the first run of your

[12] Teillet, Guindon, and Goodenough 1982
[13] (Meyer et al., 1993)

classification will not be the last. You will want to check your work by comparing the map product to some reference data anchored in reality. This may come from existing reference data, and/or that the remote sensing analyst goes out in the field. You may find that a certain class is poorly classified, and you need to adjust your input data or method to obtain the end result which meets the requirements of your project. The value of the remote sensing analyst visiting the study area and observing the actual vegetation and input data cannot be underestimated, as a deeper understanding of the relationship between the vegetation and the remote sensing data will only improve the outcome of the mapping process.

4.3.8. Post-processing map manipulation

Post-classification *smoothing* where isolated pixels are re-assigned with the class label of nearby pixels, with the aim of creating a "smoother" looking map. The result from supervised classification may require smoothing because it often looks more "pixelly" or has a "salt-and-pepper" look than does the outcome from unsupervised classification.

Different algorithms can be used for smoothing, such as a majority filter (also called a mode filter). There are also "eliminate and fill" algorithms. The window filters, such as a majority filter work like this:

- move a 3 by 3 pixel "filter box" over the classified image,
- for each pixel position, replace the center image in the 3 * 3 box with the most usual pixel value in the box.

The smoothing effect will be stronger if a larger filter, e.g. 5 by 5 pixels are used.

4.4 Access to optical satellite data

Image data can be downloaded from archives searched via internet sites, or some acquisitions may be ordered specifically for a certain time and area.
Citizens today may use satellite data that are ready-made seamless mosaics, and not handle the raw-data in the individual scenes themselves (e.g., Google Earth). For professional users, there is often a need to obtain single images, either in raw or processed format, as well as metadata about the image. If it is important to obtain images over a certain area a certain year with such a sensor, it is therefore often possible, and advisable, to order a *programming* of the satellite in advance. Such a programming should preferably be ordered a few months in advance and might cost extra.

In Sweden, a new national archive for free satellite data, called SWEA, is funded by the Space Agency to replace the current SACCESS archive at Lantmäteriet. Landsat images can also be downloaded from USGS in the US and Sentinel photos can be downloaded from ESA. For SPOT images there is a directory called Sirius. There are several commercial companies selling VHR satellite images through their web pages and through retailers,

including DigitalGlobe, which sell images from the WorldView satellites. Links to the web pages are given in the Appendix.

The remote sensing community is trying to also make as many images as possible available from the archives, including historical images. Freely available data can often be found from several websites. It is often beneficial to look at more than one website. For commercially available satellite data, websites might also be used if satellite images have been archived by the company, or they may need to be contacted directly in order to program the desired image, which will be sent most likely via internet.

4.5 The future of optical satellite data remote sensing

The European Space Agency plans to continue with the Sentinel space program, sending up new versions of Sentinel-2 satellites in the next decade. There are plans to send up a Landsat 9 OLI sensor in the year 2020. However, the fate of satellite can be dependent on the economy and goals of ruling political parties. The Landsat series aims to keep the basic characteristics of the sensor through time, in order to have long time series of comparable data. Google Earth works closely with Digital Globe Inc. which has the WorldView satellite series, and has plans for future WorldView satellites. There is also a trend to commercialization of the space industry, and from this, PlanetLabs has started sending up “showbox” satellites. Other nations that have strong space programs with future plans for optical satellite data are India, with its ResourceSAT satellite series, and China. A good source of information on future satellite launches is provided in the internet links, and can often be found on the internet.

5 Automated Digital Photogrammetry (J. Bohlin & K. Nordkvist)

Even aerial photographs can be used to produce three-dimensional information about the forest. Pictures taken in stereo can be used to create point clouds similar to those from laser scanning, with two important differences: i) points from aerial images also contain information on color, usually blue, green, red and near infrared, ii) because the camera does not "see through" the trees, there are virtually no points from the undergrowth and forested areas (Figure 23). In Sweden there is a good supply of aerial photographs thanks to Lantmäteriet's photography program, and with the help of the new, national elevation model (NNH) which supplies the height of the ground elevation, which would otherwise be missing. This technology has not yet received the same circulation as laser scanning in forestry, but the research results are promising and of great interest.

Inspiration to the figures below have been downloaded in Anders Boberg Introduction to photogrammetry [25], which is also an excellent source for those who want to know more about photogrammetrys grounds.

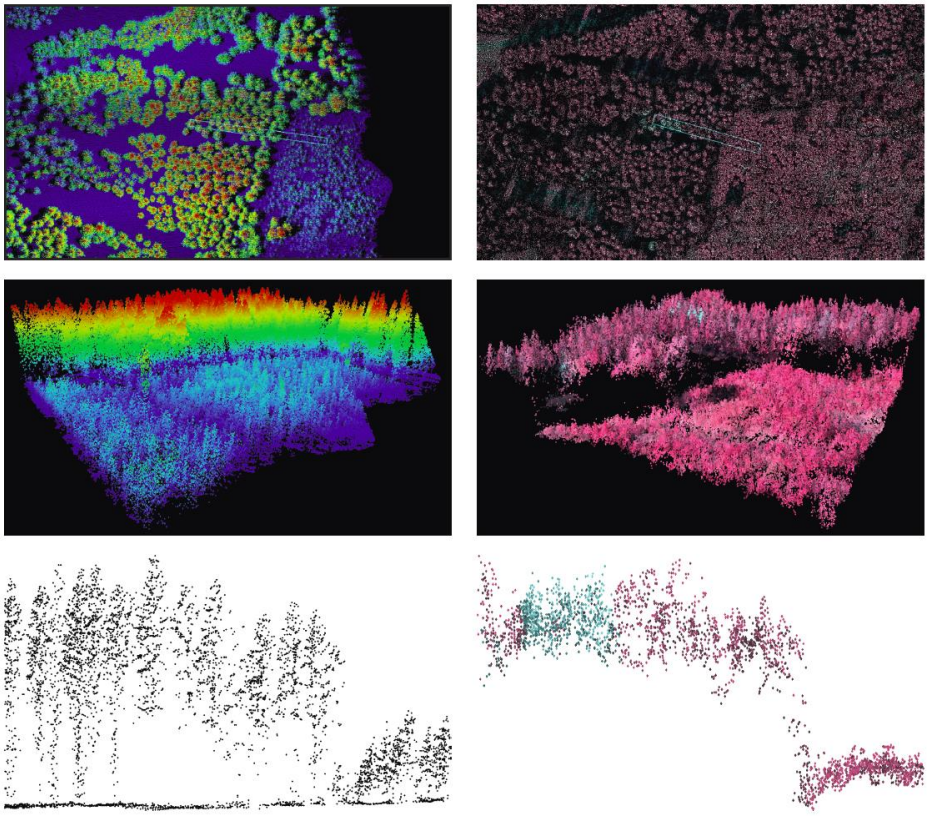


Figure 23. 3D point clouds from laser data (left) and from the matching of digital aerial photos (right). The pictures show data from the same area. The profiles (bottom row), it is clear that the laser data provides more information on the ground and tree cover's vertical structure. The digital photogrammetry, however, gives useful information on the tree cover height and also the color information in this case for example showing dead trees as blue dots. Photo: Jonas Bohlin, SLU. © National Land Survey, i2012/107. Reproduced with the author's permission.

5.1 Color information from aerial imagery

Digital aerial cameras record usually the four colors blue, green, red and near infrared. Furthermore registered often panchromatic channel where light from a large part of the visible spectrum are used. The panchromatic band usually has a smaller pixel size and is used to improve the spatial resolution of the image. The colors blue, green and red can be "normal" color images produced where vegetation appears in

green. One disadvantage of these images is that the blue band can be influenced by haze in the atmosphere, especially at altitudes above 3000 m. Often green, red and near infrared bands are used instead, but are displayed using the blue, green and red colors in the produced photographs. In these so-called “false color images” vegetation appears in red. An advantage of these images is that different types of vegetation can be identified more easily in the near infrared wavelength band.

5.2 Central Projection and orthophoto

In the type of digital camera that Lantmäteriet has used so far, as well as analog cameras, the image is a central projection. The image that this camera creates gives the landscape a so called “central projection”. Light rays from each point in the landscape pass through the camera lens (projection center) and continue to the detector, thus creating a mirror image (Figure 24).

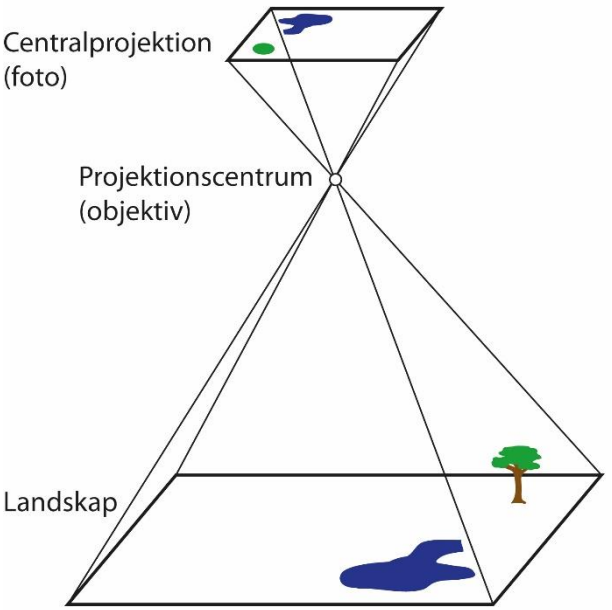


Figure 24. Central Projection.

If the landscape is flat and horizontal, and the camera is looking perpendicular to the ground, the central projection (photo) will be a map-accurate picture of the landscape. In practice this ideal case never occurs, - camera tilt, down gradient and hills, hollows, trees and buildings influence this. This has the consequence that the image geometry does not match the map’s geometry. Figure 25 a) and b) show the scale of the image varies due to elevation differences in the terrain and tilting camera axis. Heights such as trees, houses and mountains seem to lean out from the center, known as radial displacement (Figure 25 c). The offset is proportional to the object height. Similarly displaced dips in elevation appear to sink towards the center.

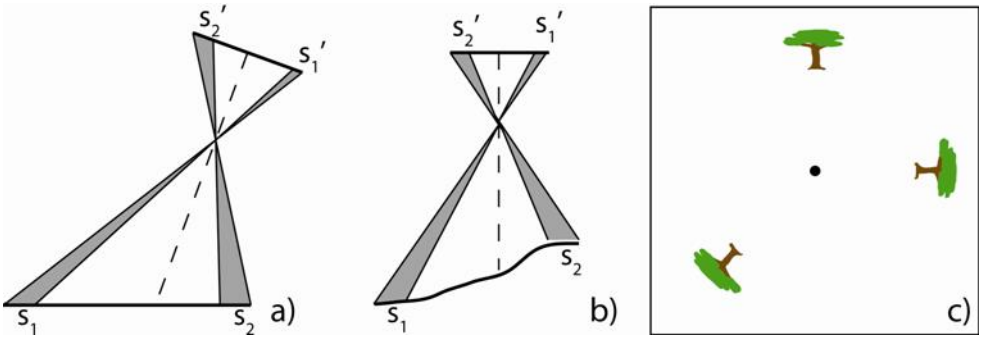


Figure 25. a) Camera axis tilt and b) the height differences in the terrain cause scale variations in the image, so that the actually equal distances s1 and s2 are depicted differently (s1 '<s2'). c) Taller object heights and trees are displaced radially from the center.

There are several commercial software programs for digital production of orthophotos (map accurate images). The programs convert the images pixel by pixel. One can imagine the orthophoto as a grid of pixels which start with no content. The pixels in the original aerial photo should be given new positions in the orthophoto. This means that each pixel of the orthophoto will be filled with color from one or more pixels in the aerial photo. Orthophot pixels have known planar coordinates. The height of the corresponding point on the ground is obtained from a height model. If the camera position and angle are known, one can calculate the flight image for this part of the landscape that has been depicted. The color of that part of the flight picture is now transferred to the current pixel in the orthophoto. The conditions for making a digital orthophoto is that you know the orientation of the camera and have access to a digital elevation model, or create one from aerial photos.

5.3 Line Scanning and cylindrical projection

An alternative to the traditional camera is a line scannern. Instead of taking a sequence of two-dimensional images the landscape is scanned row by row. To obtain stereo images scanners are used that look in two, or more often, three directions: diagonally backward, straight down and slightly forward (Figure 26 a). Thus every point on the ground is photographed from three angles. Scanning lines are perpendicular to the direction of flight, but even in this case, differences in height and tilting camera axes will cause scale differences. As with the central projection heights and hollows are displaced, but not radially from the image, but out to the sides perpendicular to the flight line (Figure 26 b). Such a projection is called cylindrical projection. The offset is proportional to the object height.

5.4 Stereo photogrammetry

The principle of stereo photogrammetry is the same as for human depth perception: two images from different angles, combined into a three-dimensional model with information about the distances to different objects. Commonly at least a 60% overlap between images is used, which by some margin provides stereo coverage across the image strip.

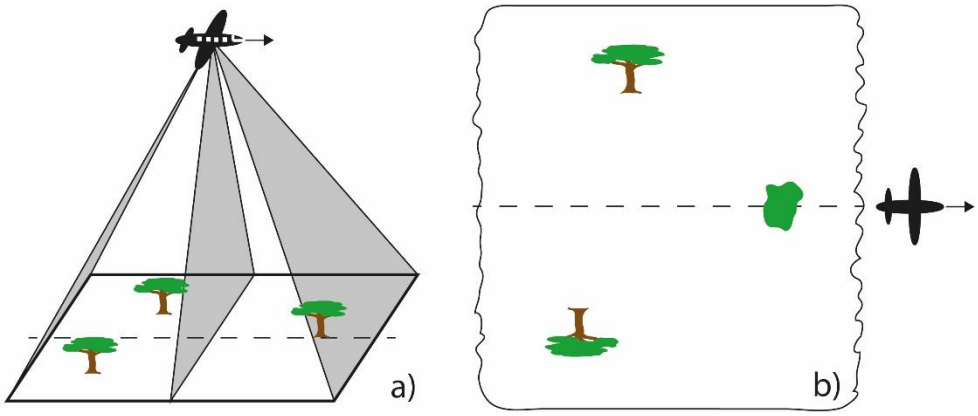


Figure 26. a) Line scan results b) cylindrical projection. In a cylindrical projection shifted tall items out to the sides, perpendicular to the flight line.

In the two-dimensional images the objects are displaced out from the center, see figure 25. The displacement is greater the closer the eye or camera object is located. An object is displaced in different ways in the two pictures, because they have different central points. The difference in shift is called parallax and allows stereo vision and range estimation. Aerial photos used for stereo interpretation overlap in the direction of flight. Orthophotos cannot be used because parallax caused by the topography has been removed. Nowadays, there are special computer screens that are used in combination with special glasses to enable stereo interpretation of digital aerial photographs. Digital photogrammetric workstations with such displays have replaced the earlier analog instruments to identify correct interpretation of aerial photographs.

Height and elevation can be calculated analytically for objects that are visible in both images of the pair. This is done by using the so-called parallax formulas. Figure 27 illustrates how the parallax formulas are derived when the camera axes are parallel to each other and perpendicular to the photographing base (distance between the two points when the camera made its exposure). It's called “photogrammetrys normal case” but, despite its name, an ideal case in practice only occurs approximately. In the figure, the image plane is drawn in so-called positive position, i.e., showing the same side of the projection center of the imaged object (compare with Figure 24 where the image plane and the depicted landscape are drawn on either side of the projection center

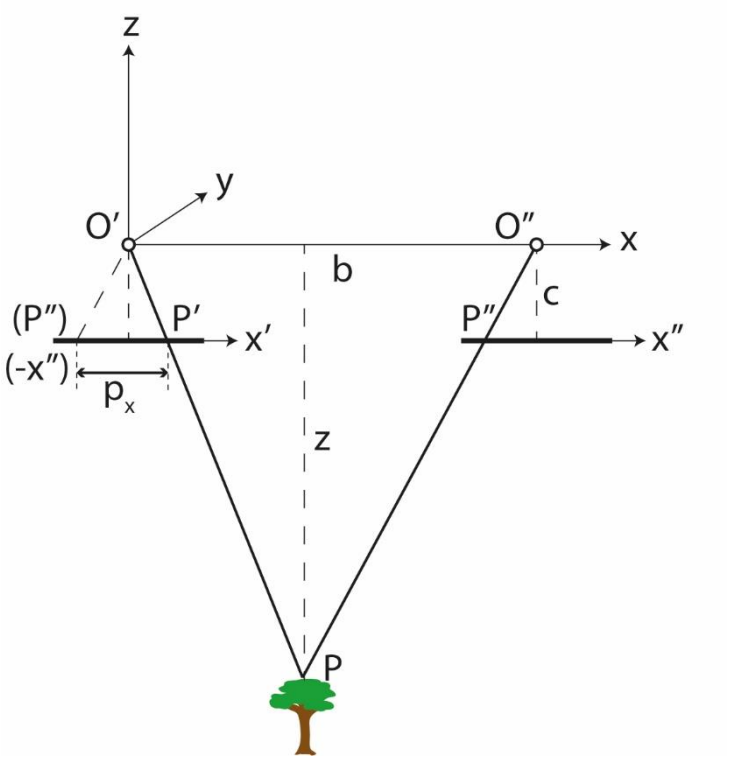


Figure 27. An illustration of how parallax formulas are derived. The coordinate system xyz is the origin of the left image projection center O '. Object point P has the coordinates x, y, z coordinate systems marked 'and'' have the origin in each image center. The mapping of point P has the coordinates x ', y', z 'in the left image, and x'', y'', z'' in the right. Camera constant c is the image z-coordinate, and is negative. Photographing base is denoted b

Similar shaped triangles gives

x / x' = z / -c’ (Eq 7)

which in turn gives

x = x' * z / -c’ (Eq 8)

p_x or x-parallax, is defined as x '- x''. Again, similar triangles give

z /-c’ = b / p_x. (Eq 9)

Combing Equation 9 and 8 finally gives

x = x' * b / p_x (Eq 10)

In a similar manner one can show that
y = y'* b / p_x (Eq 11)

and (because the y-parallax normally is zero and y '= y'')

y = y'' * b / p_x (Eq 12)

and that

$z = -c * b / p_x$ (Eq 13)

The height z is the object's altitude. To determine the height above the ground the ground height must either be measured in the same way as above, or you need to have access to an ground elevation model.

5.5 Point clouds from stereo images

The conditions needed to generate point clouds from aerial photographs are i) the images overlap, ii) the camera interior orientation is known and iii) the camera position and angle are known with great accuracy. The camera's interior orientation is measured in the laboratory. The external orientation (position and angle) are recorded with GNSS and INS (section 3.4), but these values are often not accurate enough for a good image match to be made. Fine adjustment is done with the block triangulation. A common method for this is called “bundle adjustment”. Clearly visible and accurately measured points in the terrain, for example, painted white squares or manhole covers, are used to connect the images together into an image block. Bundles from these objects are used to fine-tune (adjust) the cameras external orientation.

When the conditions above are met, the images within the block are matched to the point cloud. There are some different methods for matching and the results are slightly different. Object-based matching requires that as many objects as possible are identified and matched. Pixel-wise matching (semi-global matching) is a newer method that ensures that every pixel is matched, resulting in very dense point clouds. After matching the objects, the x-, y-and z-coordinates are calculated using parallax formulas (section 9.4). The result is a point cloud very similar to that obtained with laser scanning, but one important difference is that the points can be complemented with color information from aerial photographs. WHich principle that is used to do this varies between different software for image matching.
As the number of ground points is very low for the matching of aerial images, especially in forests, in practice it requires that you have access to an existing elevation model. Starting in 2009, Lantmäteriet has been conducting a national laser scanning in order to create a new national elevation model (NNH), and the possibilities of using stereo matched aerial photography is good.

5.6 Applications in forestry

Orthophotos are widely used as background maps in forestry, as well as delineation and updating of forest management plans. Even mapping in photogrammetric instruments has been used, particularly for forestry planning of major holdings. In the so-called "LMV" method, photogrammetric instruments were used to make the delineation of stands and to measure tree height, and to estimate the volume density (in Swedish “massaslutenhet”) and tree species distribution. The height and volume density is then used to estimate the timber volume. With the introduction of digital images and digital photogrammetric instruments, this approach has become easier. It is not so widespread in the forestry companies but is mostly used by consulting firms.

There is a possibility to use orthophotos, for example to perform segmentation of stands, but different viewing angles in different parts of the image require special

software. In addition, a two-dimensional image gives only limited information on forest height.

The technique of matching aerial images that is used to create 3D point clouds with the ability to depict the canopy similar to laser data provides new opportunities to automatically obtain forest information from aerial photographs. This is especially interesting for smaller properties because Lantmäteriet regularly photographs the country and distributes the digital images at a marginal cost.

If the points are converted into raster data, then they can be used for segmentation of populations with different heights and texture. It should also be possible to use 3D models of this kind in order to detect changes.

Point clouds from photogrammetry can also be used for estimates of height, timber volume and basal area in the same way as with the area-based method. A recently reported trial gave good estimation results when images from Lantmäteriets DMC camera with altitude 4800 m were used (Table 5).

Table 5. Accuracies at the stand level for the estimation of tree height and stem volume from point clouds created from the National Land Survey digital aerial imagery, from Bohlin et al 2012 [26].

Variable	Metric used from point data	Accuracy at Stand level (RMSE)
Basal area weighted mean height	h_{80}	8,8 %
Stem volume	h_{80} and a density metric	14,6 %

The experiments reported in Table 5 are very promising because it indicates that forest estimates can be made automatically with relatively inexpensive technology, with at least as good results as the manual methods. The condition is, however, just as for laser-based estimates, that there are reference plots of forest data that can be used as independent variables in the estimation. The forest used in the study above had tall forest arranged in well managed stands. Attempts of this kind must be repeated on more properties before firm conclusions can be drawn. It is clear that the point cloud from photogrammetry is a very valuable source of data for future forestry purposes.

6. LASER DATA FOR REMOTE SENSING

LiDAR. The technology to measure distances with laser is often called LiDAR from the English "Light Detection And Ranging".

Global Navigation Satellite Systems (GNSS). GNSS is a generic term for all satellite-based navigation systems.

Inertial Navigation System (INS).

Laser data in remote sensing

Laser scanning provides accurate three-dimensional measurements of vegetation and other surfaces. Airborne laser scanning has revolutionized forestry opportunities to efficiently produce forest maps with forest variables such as timber volume, basal area, tree height and trunk diameter. Through terrestrial (ground-based) laser scanning, it is possible to rapidly obtain field data with extensive information in comparison to traditional manual measurements. These instruments can be placed on a stative or can be a mobile system (eg., on an ATV or a backpack with an operator). Information from terrestrial systems can be used to calibrate remote sensing data collected from airborne systems.

6.1. Introduction to airborne laser scanning

Laser scanning has been established as a new and effective approach to forest data capture. Airborne laser scanning measures with decimeter accuracy the position of points both on the ground and in the canopy. With the help of special computer programs the ground height can be calculated. After this, statistical measures describing the distribution of the remaining laser points in the canopy are calculated. Using for example regression analysis the forest data from plots can be transferred to the raster cells which cover all forest land covered by the laser scanning. This so-called area-based approach was introduced in 2002 as a commercial approach to forest management planning in Norway and is now used by almost all forestry planning there. Foremost the area-based method provides measurements of tree size and number within each raster cell (stem volume, height, basal area, etc.). Delineation of stands and estimation of species distribution is done in Norway usually by manual interpretation of digital aerial photographs in stereo.

The Finnish state forestry centers have switched from a traditional, field-intensive forest mapping to making automated estimation of laser scanning. The area estimated by laser-based methods now amounts to 1-2 million hectares per year. Unlike in Norway, Finland tries to estimate tree species automatically, for example by combining the laser data with digital aerial photographs. The methods for automatically classifying detailed tree species from digital aerial photographs over large areas for operational projects, however, is still relatively uncertain.

In Sweden, most of the major forestry companies and some forest owners' associations tried to make laser estimates in the early 2000s, often on study areas of about 10 000 ha. The results were generally good in terms of variables related to tree size. According to a statement made by OL Forest Inventory AB the estimation accuracy (relative RMSE) of stand level with the area-based approach was 3-6% of basal medium height, 6-14% of the timber volume and 7-13% for basal stem diameter. These accuracies are much better than that achieved by traditional inventories made in connection with forestry planning. The accuracy of the estimate of the number of stems is about 12-24% (Brethvad and Iversen 2012).

The establishment of laser scanning as an operational method in Swedish forestry took momentum with the national laser scanning that the National Land Survey (Lantmäteriets) started in 2009. In 2011, the forestry company of Bergvik was the first major forest company in Sweden to order laser estimates for all its holdings based on Lantmäteriets laser scanner data. Since then, among others, Holmen has done the same and SCA has ordered the estimates for all of Norrland north of Sundsvall, even outside their own holdings. Even the National Property Agency (Fastighetsverket) and Sveaskog have ordered laser products from the consultancy that worked on Lantmäteriets laser scanner data.

In 2013 the National Board of Forestry was commissioned by the Government, together with SLU and other stakeholders, to make forest products over all of Sweden based on Lantmäteriets national laser scanning data. As a result, SLU made forest estimates based on raster data with 12.5 x 12.5 m size grid cells throughout Sweden except for the mountains. The variables created were estimated growing stock, tree biomass, basal area, basal area mean diameter and mean height. In addition, the Board of Forestry together with consulting companies developed additional products such as a 2 × 2 m tree height grid, and raster layers based on the DEM model from laser data showing hill shading, and ground wetness areas. These data layers can be accessed from the National Board of Forestry's website under the name of “skogliga grunddata” (in English translating loosely to “basic forest data”). The production of the first version of these forest data was completed in 2016, as all ALS acquisitions over the forest were not ready until then.

Initially the spread of the operational use of laser scanning occurred more quickly in Norway and Finland, especially for private properties. A contributing factor was probably that in these countries there was an established system of coordinating forestry planning for private properties over larger areas, while the state partially funded the plans.

6.1.1. Laser scanning's development

Laser light is single color, directional light wave that is in phase. The first functioning laser was constructed in 1960. Today, lasers are used in several areas: in CD and DVD readers, printers, laser spectroscopy, the study of gases in the atmosphere, distance and speed measurement, topographic mapping, mapping of the sea depth in shallow areas and much more. Airborne laser scanning (ALS) uses the laser's ability to measure distance based on the time from when a laser pulse is sent out, until the laser light reflected from the ground or vegetation comes back to the sensor. The technology to measure distances with laser is often called LiDAR from the English LIght Detection And Ranging.

The first attempt to measure tree heights with LiDAR was made in the former Soviet Union in the late 1970s. Measuring tree height profiles with airborne laser was also tested in Canada and the United States in the early 1980s (Nelson, Krabill, and MacLean 1984, MacLean and Krabill 1986). The first known studies of using an airborne scanning laser system for forest inventory in Sweden was carried out in collaboration between the then FOA (now the FOI, or the Swedish Defence Research Institute) and the Swedish University of Agricultural Sciences, 1991 (Nilsson 1996, 1994). These early experiments showed that stem volume and tree height could be measured by laser scanning. One problem with the experimental system used in this early test was the low accuracy in positioning. In the mid-1990s commercial laser scanning systems for airborne systems over land were developed. In these systems the GPS was integrated with inertial navigation (Inertial Navigation Systems, INS) which made it possible to determine the position of the laser measurement with 0.5 meter accuracy or better.

One of the first commercial systems was Swedish TopEye, developed by Saab in 1993. This helicopter-borne system produced data of high quality. The TopEye system was designed for detailed measurement of small areas and was widely used for the measurement of various types of infrastructure and to experiment with estimates of forest. The early attempts of helicopter-borne laser scanning did not lead to its use within operational forest inventory because the cost of data collection was perceived as too high. In Norway, Erik Næsset in 1995 attempted to measure forest with a scanning laser mounted in an airplane. The tests were successful, leading to the development of the so-called area-based approach for inventory at stand level with the support of field measured reference surfaces. In the late 1990s Finnish and Swedish research showed the ability to also detect and measure individual trees in laser data with high point density. An account of the early development of the Nordic countries is given in the following reference[1].

[1] Næsset et al. 2004

Today, laser scanning is an established method for forest inventory. Examples of countries or regions where extensive laser scanning for forestry purposes are made are Norway, Finland, Austria, Spain, northern Italy, USA, Canada, Chile and Tasmania off Australia. Individual projects are in another wide range of countries, such as Nepal, Tanzania and Brazil to name a few. In 2011, Professor Erik Næsset from Norway received the Wallenberg Prize (also called the forest Nobel Prize) because he launched the area-based approach as an operational method for forest inventory. The year 2011 can also be counted as a breakthrough year for large-scale operational use of the method in Sweden, as the Bergviks project to underestimate its entire forest holdings with laser scanning started at that time. In Sweden, it is therefore 20 years from the first trials in 1991, to the full-scale deployment.

6.1.2. Basic characteristics of laser

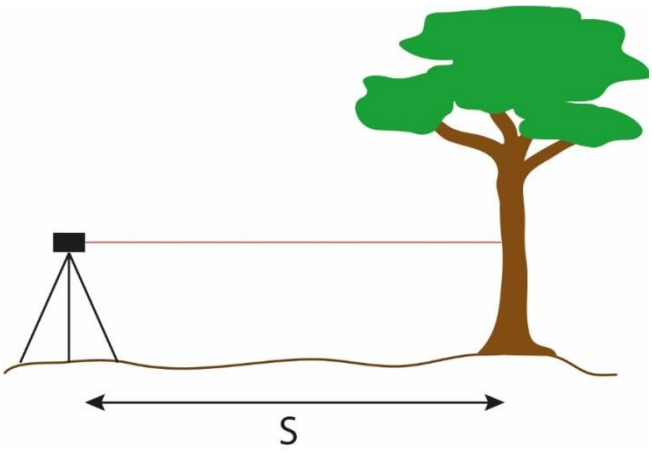
This chapter describes the main principles of laser scanning and the main components used in today's systems. Foremost is the chapter on airborne laser scanning. A distance-measuring laser is placed in an airplane or a helicopter. The laser emits pulses to measure the distance to points on the ground or in the vegetation, and a scanning mechanism is used to distribute pulses in a wide swath beneath the airplane. For accurate positioning, the laser measurement points are required to be of high precision to register the distance to the point and the laser position and direction in which the pulses are transmitted.

6.1.2.1. Distance measurement

The method used for measurement of the distance (on the order of 100 meters or more) means that the laser emits a short (about 4-10 ns, corresponding 1.2-3 m), but intense light pulse. The pulse travels through the air, hits an object and is reflected by the object back to the instrument, which records how much time has passed since the pulse was sent out (Figure 6.1). The detector converts the light into an electric voltage where the signal strength is a function of time. Since the speed of light is known, the distance between the instrument and the reflecting object is calculated as

$$S = v \times t / 2 \quad \text{(Eq 1)}$$

where S is the distance to the object, v is the speed of light and t is the time measured. The division by 2 is because light travels the same route twice: first from the instrument to the object and back. The method with the timing of the return pulse is common in both ground-based and airborne instruments.

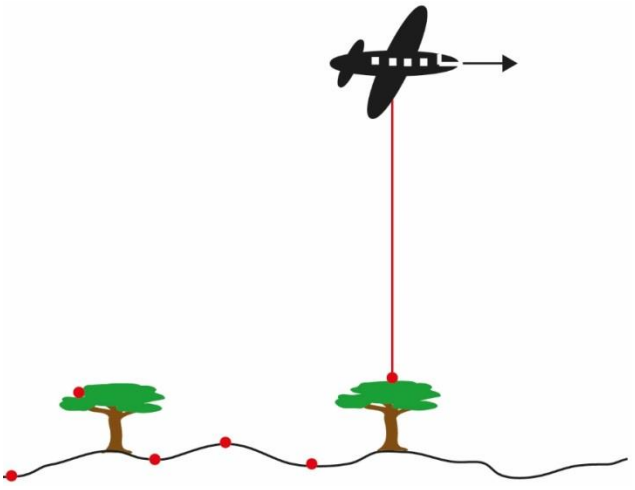


Figur 6.1. Avståndsmätning med laser.

There is also a type of ground-based instruments that instead measures the phase shift of the returned signal from an amplitude modulated laser pulse. This method is very accurate but is only suitable for short distances.

6.1.2.2. Profiling laser

By placing a distance-measuring laser in an aircraft altitude profiles of ground and vegetation can be created. While the aircraft moves forward laser pulses are emitted with high frequency and hit the ground along a line in the plane's flight direction (Figure 6.2). The early experiments with laser measurement of forest in the 1980s were made with this technology and potentially it could be a cheap way to get forestry statistics for large areas.

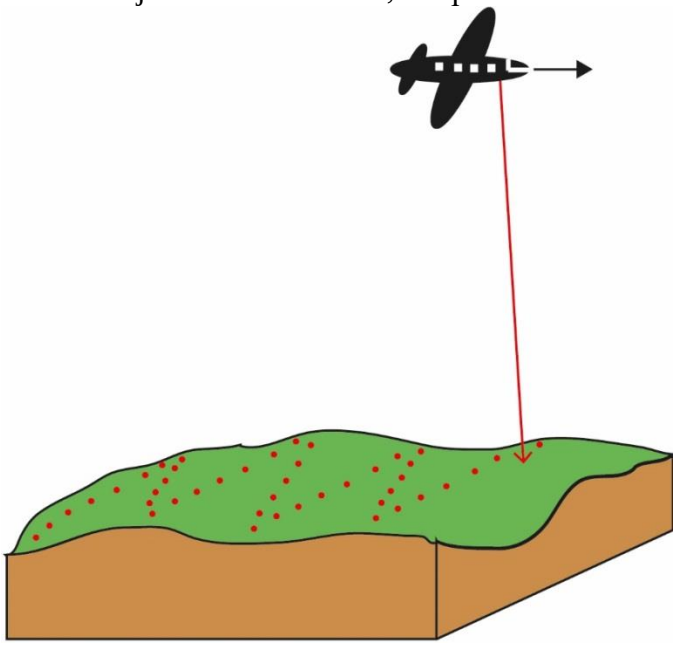


Figur 6.2. Profilerande laser.

6.1.2.3. Scanning

The methods for laser inventory of forest that are now used in practical forestry are based, unlike the profile measurement, on the scanning of the entire area to be inventoried. The height measuring laser is complemented by a scanning mechanism that distributes the measurements in a path under the aircraft (Figure 6.3). The scanning mechanism often consists of a

rotating or oscillating (wobbling) mirror. By flying several paths or strips which are adjacent to each other, it's possible to cover larger areas.



Figur 6.3. Flygburen laserskanning.

Strip width is determined by the altitude and angle of the scan, that is, the maximum angle between the laser beam and the vertical. String width and air velocity then determines how large area can be covered per unit of time. The faster an area can be covered, the lower the cost, and for that would use a high altitude, air speed and scan angle. For a given combination of these parameters point density on the ground is determined by the pulse rate (number of transmitted pulses per second). Because users often want to have a sufficiently high point density and large area coverage per unit of time, development has moved towards scanners with a high pulse rate. A technical challenge is to combine the increased frequency of the greater power needed for scanning from a high altitude, because the laser needs more time between each pulse when the power per pulse is high.

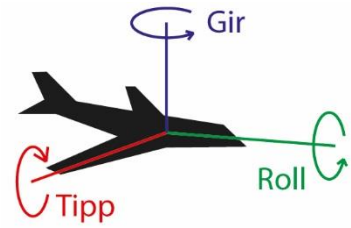
The development, however, has been very rapid, and the pulse frequency of the current commercial systems are often several hundred thousand pulses emitted per second, compared with about 2000 pulses per second by the earliest commercial systems in the mid-1990s. A high frequency relative to the altitude results in a pulse does not have time to be received back at the sensor before the next is sent out, and therefore systems have also been developed that can keep track of multiple pulses while in the air.

For large scanning angles the proportion of the laser pulses that hits the vegetation instead of the ground increases. This means both lower coverage of the ground, and that part of the laser measurements used in forestry estimates become less representative. Often therefore there is a limit set on a useful scanning angle – which is approximately +/- 20 ° for laser scanning primarily for the production of ground models, and approximately +/- 15 ° for laser scanning primarily for forestry estimates.

6.1.2.4. Positioning of laser point

In order to calculate the position of the measurement point in horizontal and vertical direction the laser scanner's position must be known as well as the orientation at the time of each transmitted pulse. The position is defined by three coordinates in space (x, y, z) and the orientation of the three angles called tilt, roll and yaw (Figure 6.4). Moreover, one must know the emitted laser pulse and direction relative to the instrument.

.



Figur 6.4. Tipp, roll och gir (på engelska pitch, roll, yaw).

When the first experiments with airborne laser scanning were done, the technology was lacking to measure these parameters with sufficient precision, but in the mid-1990s, GNSS (Global Navigation Satellite Systems) started to be integrated with inertial navigation systems (INS). GNSS is a generic name for all satellite navigation systems, in the current situation the US GPS and Russian GLONASS systems. In practice, GPS is used more than any other system. GNSS provides good measurements of position and velocity, but the frequency is so low that multiple laser pulses can sent out between each measurement.

To measure the plane's orientation and fill in the gaps between the GNSS measurements used INS. An INS, or inertial system, often consists of three gyroscopes and three accelerometers. Gyroscopes measure the angular velocity around the three axes, and by integrating with respect to time, the change in orientation compared to the original orientation can be calculated. The accelerometers measure force and acceleration. Double integration with respect to time gives the position change from the starting point. GNSS and INS complement each other by the fact that INS fills the gaps between the relatively sparse GPS position and velocity measurements, while the GPS is used to correct for drift in the inertial system.

6.1.2.5. Mätnoggrannhet

For hard surfaces, the accuracy and resolution of the distance measurements primarily on tidtagningens accuracy and laser pulse length. Since the distance to an object is

$$S = v \cdot \frac{t}{2}, \qquad (1)$$

where t is the accuracy of the counting. The speed of light is approximately 300,000 km / s, so that a measurement accuracy of 1 dm requires a timing accuracy of 0.67 nanosecond (1 ns = 10-9 s). The speed of light is well appointed but slightly affected by air temperature, so a measurement accuracy better than anyone cm is difficult to achieve using airborne laser scanning, no matter how well time can be measured. The laser beam has low divergence, that is cohesive - compare the narrow beam of a laser pointer with the light beam from a flashlight. At large distances is the divergence nevertheless sufficient to pulse should see a small area on the ground rather than a point. A divergence of 0.5 mrad (0,029 °) and an altitude of 1000 meters gives a face (footprint) by approximately 50 cm in diameter. Larger divergence and altitude leads to the laser pulse, a larger hitting area, which affects the accuracy of the measurement negatively. Even large terrain slope and scanning angle can reduce measurement accuracy.

For hard surfaces, the accuracy and resolution of the distance measurements depend primarily on time-taking accuracy and the laser pulse length. Since the distance to an object is

$S = v \times t / 2$ (Eq 2)

so, the accuracy of distance measurement can be approximated to:

$\Delta S = V \times Dt / 2$ (Eq. 3)

where Dt is the accuracy of the timing. The speed of light is about 300,000 km / s, which means that a measurement accuracy of 1 dm requires a timing accuracy of 0.67 nanoseconds (1 ns = 10^{-9} s). The speed of light is well defined but slightly affected by air temperature, so a measurement accuracy better than a few cm is difficult to achieve using airborne laser scanning, no matter how well the time can be measured.

The laser beam has low divergence, in otherwords it is coherent – as an example to demonstrate this, compare the narrow beam from a laser pointer with a beam from a flashlight. At large distances, the divergence is still sufficient so that the pulse should hit a small area on the ground rather than a point. A divergence of 0.5 mrad (0.029 °) and an altitude of 1000 meters results in measuring an area of about 50 cm diameter.

A part of the laser pulse can be reflected by an object that does not block the entire pulse beam path, while the rest of the pulse continues and is reflected towards items that will occur later in the path, for example, the ground. Many sensors can register more than one echo of the transmitted laser beam. For most sensors there must be at least one meter between echoes recorded from the same laser pulse.

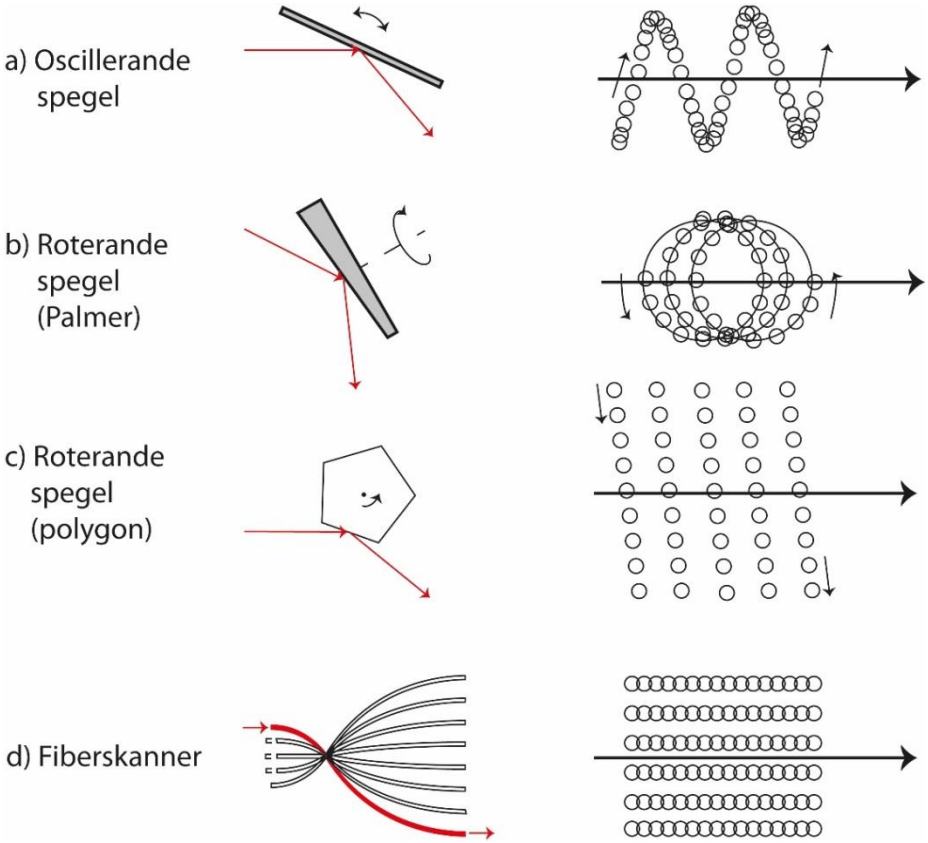
6.1.2.6. Different types of scanners

The scanning mechanism usually consists of an oscillating or rotating mirror. A third type is a fiber scannern. An oscillating mirror oscillates between two positions and the laser pulses hits the ground in a zigzag pattern (Figure 6.5a). At the turn, the mirror slows down and the point density perpendicular to the direction of flight is higher in the strip's outer edges.

One type of scanner rotating mirror is Palmerskannern (Figure 6.5b). The mirror surface is not perpendicular to the rotation axis, and the transmitted pulse direction changes with the position of the mirror. The pulses hit the ground in a circular pattern that moves forward in the direction of flight. The density is highest in the corridor's edges. A Polygon scanner (Figure 6.5 c) is also based on a rotating mirror. The mirror does not need to accelerate and brake, resulting in a smoother point dentsity. Because of the aircraft's forward motion, the pattern is slightly angled towards the direction of flight.

The Fiber scanner (Figure 6.5 d) is made up of a bundle of optical fibers. A laser pulse is emitted into a fiber bundle in one end, where the fibers are assembled in a circle. At the other end of the bundle, where the pulse is emitted, the fibers are in a line. In this way the circular scanning motion is translated into a linear pattern. A rotating mirror divides the pulses between the fibers. The mirrors used are smaller than in other scanners and can therefore rotate faster. Point density is significantly higher in the flight direction than the strip width. One way to compensate for this is to allow the fiber-bundle emitting-end to swing back and forth so that the pulses from each fiber forms a zigzag pattern[2].

[2] Petrie and Toth 2009a



Figur 6.5. *Various types of scanning mechanisms and the resulting scan pattern*

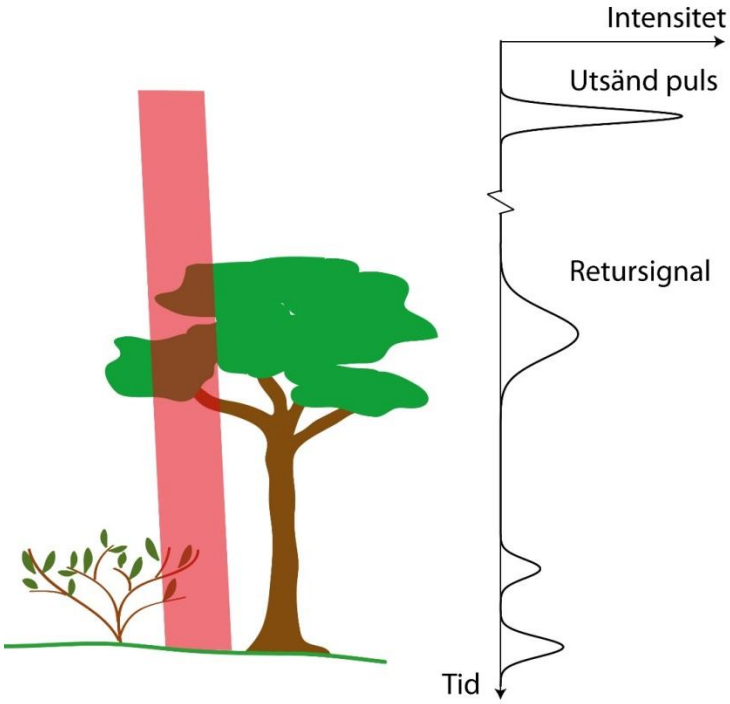
6.1.3. Laser pulses interaction with ground and vegetation

On its way to the ground, the transmitted laser pulse can be reflected by the ground or objects above the ground such as tree crowns, trunks, other vegetation, rocks, buildings and more. A hard and compact surface results in a single distinct return pulse, but when the laser beam hits for example a tree crown or the edge of a roof, just a part of the pulse is reflected while the remainder continues travelling. Water has low reflectance and therefore has virtually no return.

Also a measure of the reflected radiation referred to as "intensity" is often recorded together with the coordinates for each detected laser return. Many laser systems use near-infrared light and a surface that reflects much light in this wavelength range (such as living vegetation) therefore appears "bright" in the intensity data. The intensity of the return pulses is difficult to interpret, because the amount of light reflected from vegetation depends on leaf angles, vegetation density, and other reasons.

Figure 6.6 illustrates how the reflection from different layers of vegetation and finally from the ground can give multiple returns from a transmitted laser pulse. The number of registered returns can vary between scanners. The forest often gives the first return from the canopy and the last returns from

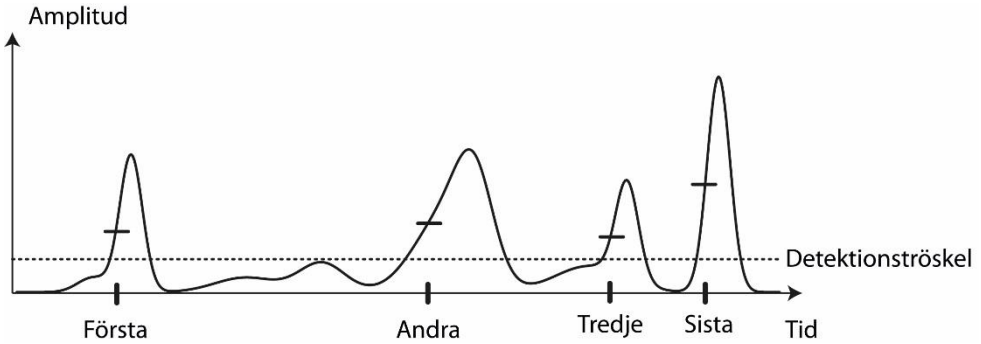
the ground. (There are also scanners that sample the whole laser return with high frequency, so called full-waveform laser).



Figur 6.6. *Laser pulse path to ground. This situation generates multiple returns - first from the tree, then from the bush and finally off the ground. The curve next to the image represents the return pulse's intensity as a function of time, with peaks due to the different objects.*

The height distribution of the laser points (the returns) occurs below the actual tree height distribution. This is partly because the laser beam penetrates the part of the tree crown before enough energy is reflected to trigger a measurement in the detector, and partly because the pulses not only hit the treetops, but also the tree crowns' sides. Dense vegetation can also cause returns from the ground and vegetation to blend together so that the ground return will be partially masked. It may lead to the return being registered early and for the ground height to be overestimated.

The transmitted laser pulse duration is in the order of meters, but the uncertainty in the measurements against a hard surface should not exceed a few inches. Of great importance for the measurements is therefore the algorithm (calculation rule) that determines at what signal strength a detector will register a return. A common method is to register a return is when the peak of the return signal has reached a certain percentage, for example 50%, of its maximum value (Fig 6.7). In this way timing is not affected by the pulse peak power, but only by its width Peaks that do not reach above a certain threshold, the detection threshold, are not recorded at all.



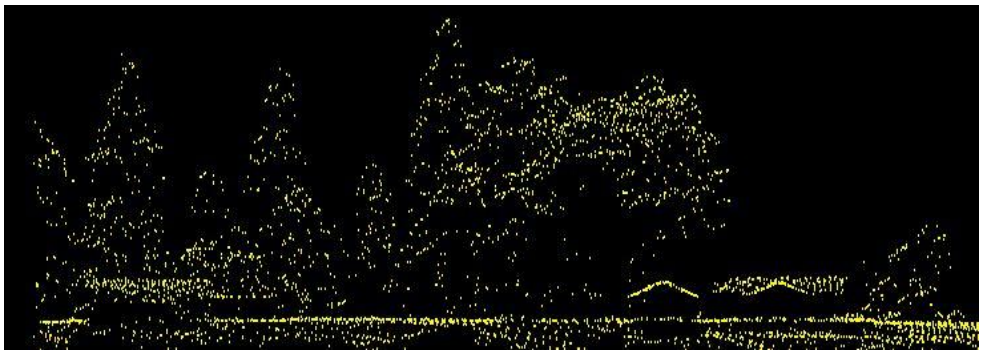
Figur 6.7. A common method for recording the return signal is a peak in the signal which triggers an alert when it reaches a certain percentage of its maximum value, for example 50% (horizontal line in graph). Only peaks that reach above the detection threshold (dotted line) is recorded.

Laser systems are calibrated to give correct readings when hitting hard surfaces such as asphalt. The vegetation is however half permeable to the laser, and the fact that different systems use different principles for determining when a return pulse should be detected contributes to the values of vegetation of different laser systems being not directly comparable. The documentation available on the method used by the system to acquire and process the laser scanning data is often inadequate.

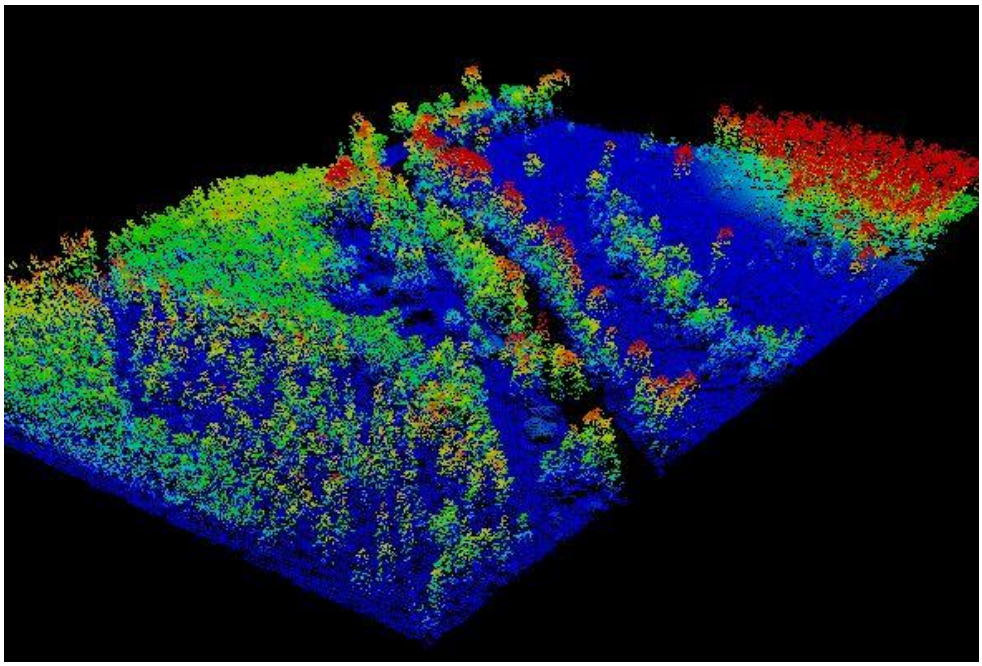
.

6.1.4. Processing of laser data

The laser returns recorded from the scan together form a so-called point cloud, which is a group of points which are coordinate-set in three dimensions (Figure 6.8 and 6.9). Most laser scanner systems also save the return signal intensity with the coordinates of the detected points. Before the data are used for further analysis they require some preprocessing. This is usually done by the data provider and is not specific to forestry applications.



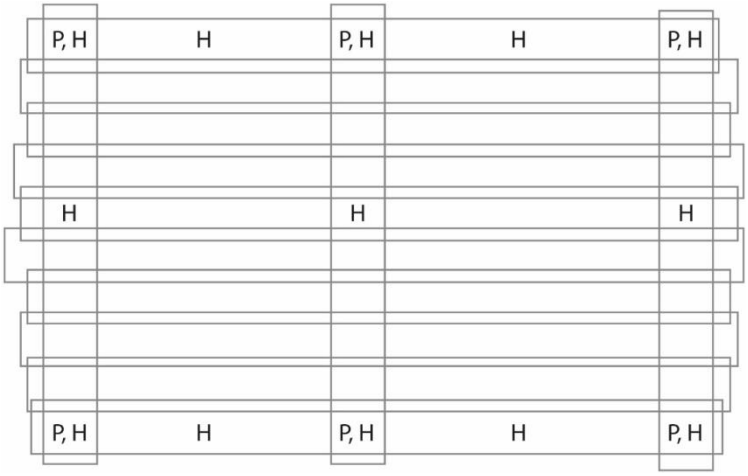
Figur 6.8. Profile of laser point cloud where trees and buildings are clearly visible. The image is based on data from the National Land Survey. *© Lantmäteriet, i2012/107.*



Figur 6.9. Point clouds in three dimensions, seen obliquely from above. The points are colored from blue to red after rising height measurement. Through the area runs a stream. The image is based on data from the National Land Survey. © *Lantmäteriet, i2012/107.*

6.1.4.1. Stråkutjämning och inpassning mot kontrollpunkter

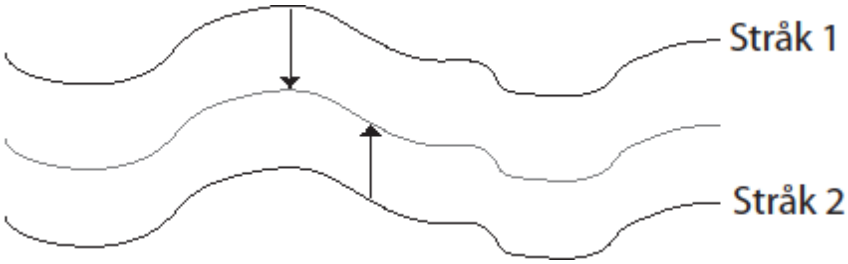
A number of factors that affect the positioning of the laser points, and an important source of error is the navigation system which consists of the GPS and INS. When an area is scanned it is usually in the strips which go back and forth, with some overlap between the lateral strips. They also tend to cross paths at the beginning, end and possibly the middle of the area (Figure 10). The uncertainty in navigation varies with time.



Figur 6.10. *Outlays of the strips, cross strips, planar support (P) and the height support (H) in a scan area.*

Data from different strips fit into a single seamless block via a process called strip equalization (Figure 6.11). Then the block is fit with the data against

the control points on the ground. Planar support is used to fit the data in the horizontal direction, and is composed of items that can be easily identified in the point cloud. Examples of suitable planar support are rooftops and ditches. Another possibility is to use white painted markings visible in the intensity data. Height support is used to fit the strips vertically, and should be done after planar alignment. Here open, even slightly inclined surfaces are desirable. Hard surfaces like asphalt and gravel are suitable because they give high accuracy in height measurements. [3].



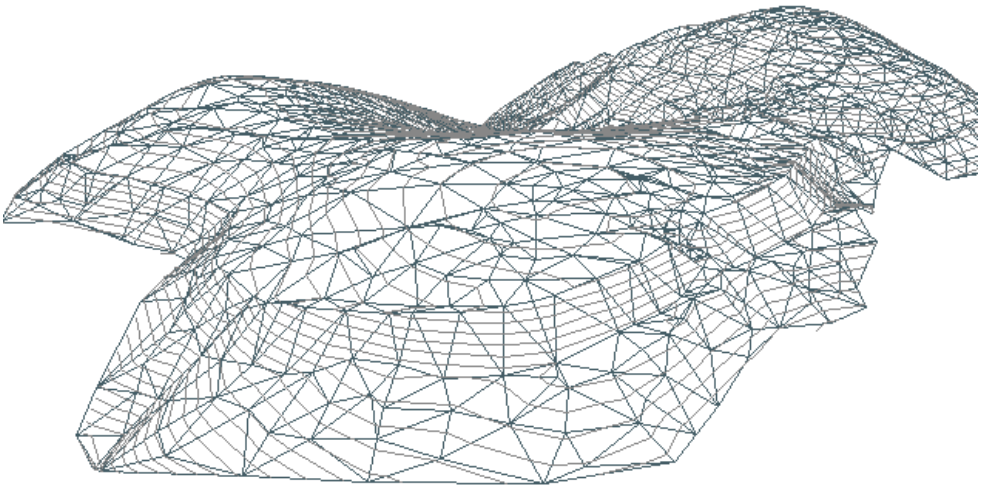
Figur 6.11. *At strips equalization fit the different flight routes toward each other, whether in an average area or against a known surface.*

Classification of ground points

The erroneous data points removed are those that are high in the air as a result of reflection from such clouds, fog and birds, and those that clearly are below ground level. Low points can occur when a pulse is reflected in several steps so that the return signal is delayed, and the point seems to be farther away than it actually is. Classification of ground points can be made in different ways, two of which are described below.

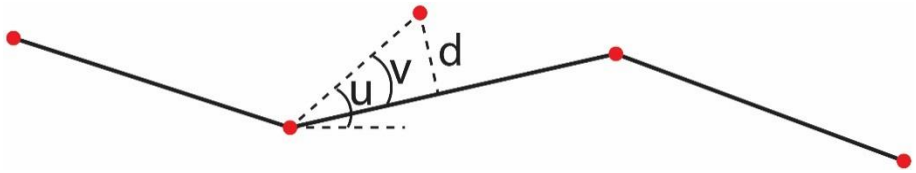
One method was developed at KTH by Peter Axelsson and is used in TerraSolids software Terra Scan, and has received very widespread recognition [10]. To begin with, it uses a rectangular grid over the point cloud, with a cell size as determined by the user. In each cell, the lowest measuring point out is selected and classified as a ground point. The selected points are tied together in a network of triangles which provides an approximate representation of the surface. This data structure called a TIN (Triangulated Irregular Network) (Figure 12).

[3] Olsson, Rost, and Reshetyuk 2013



Figur 6.12. *Ground elevation model in the form of TIN (dark gray lines). The lighter lines are contours. (After the picture of Robert Kropf at wikipedia. Licensed by the author under the GFDL).*

Now begins a new process of gradually adding new points in the network so as to densify and follow the surface details. Points in the cloud are examined one by one, and rejected or accepted as a new ground points according to certain criteria. One criterion is based on the distance between the point and the existing surface, another uses the difference in the slope of the ground surface of the new point added, and a third criterion is based on the maximum ground slope (Figure 6.13). What values are appropriate for these parameters is partly due to the topography.

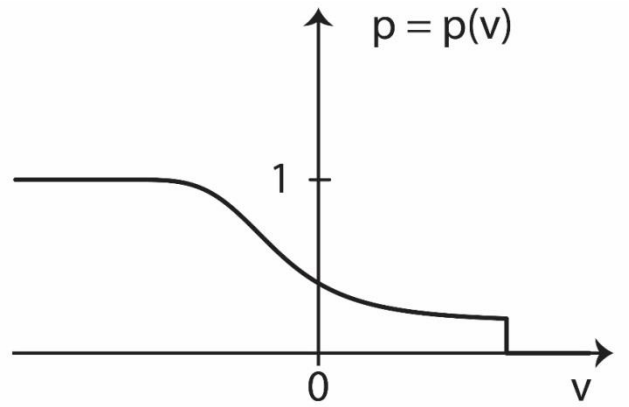


Figur 6.13. *In Axelsson's method of land classification accepted a new point land point if the distance d to the existing surface, the angle v to the surface and ground slope u below the selected thresholds.*

Another method is to first create a surface with the help of all points, where each point is given equal weight[4]. The surface falls somewhere between the ground and the canopy, and ground points are more likely under the surface than above. Each point is now given a weight which depends on its distance and direction from the surface (Figure 6.14). The lower the point is in relation to the surface, the higher the weight it gets, going down to a limit below which all points will get a maximum weight equal to 1. It also puts a distance above the surface where all points get a weight of 0 and therefore do not affect the surface. The weighting functions slope between the two extreme positions can be adjusted. The new weights

[4] Kraus and Pfeifer 1998

are used to create a new surface, by points with higher weight "pulling" the surface to itself. The weights are updated again, and another new surface created. This continues a number of times - about 3-5 iterations may be required. After the last iteration the last computed surface determines which items should be classified as ground hits. This method is used in FUSION, a software for processing of laser data for forestry applications.



Figur 6.14. *Weight p as a function of distance from the surface from Kraus and Pfeifers method*[5] *. Over the surface (v> 0) the weight decreases of above a certain value 0. Points below the surface (v <0) get higher weight, the lower they are, but never more than 1.*

6.1.4.2. Development of ground elevation model

Following the classification of ground points, the digital ground elevation model (or the Digital Elevation Model, DEM), is produced in the form of a TIN or raster. If the Axelsson method described in the previous section is used for ground classification, then there is already a complete TIN. If instead a method that only categorizes points into ground and other is used, for example Kraus and Pfeifer's method describe above, it has a number of ground classified points that can be connected to a TIN. The benefits of a TIN is that all input ground points retain their coordinates in the plane, so that the accuracy of the point data is retained. However, it is a complex data structure that takes longer to create and process than a grid..

A grid can be generated from a TIN by is assigned a height to a raster cell from the TIN data structure has for the center of the grid cell. You can also create a grid directly from a collection of ground classified points either by using ground points within the grid cell or by an interpolation which also uses points just outside the grid cell. In a raster the points and their exact positions are not retained and hence this format does not taking advantage of the measurements' full accuracy. Rasters are still the most common form of ground elevation model for use in forestry applications because they are very convenient to work with.

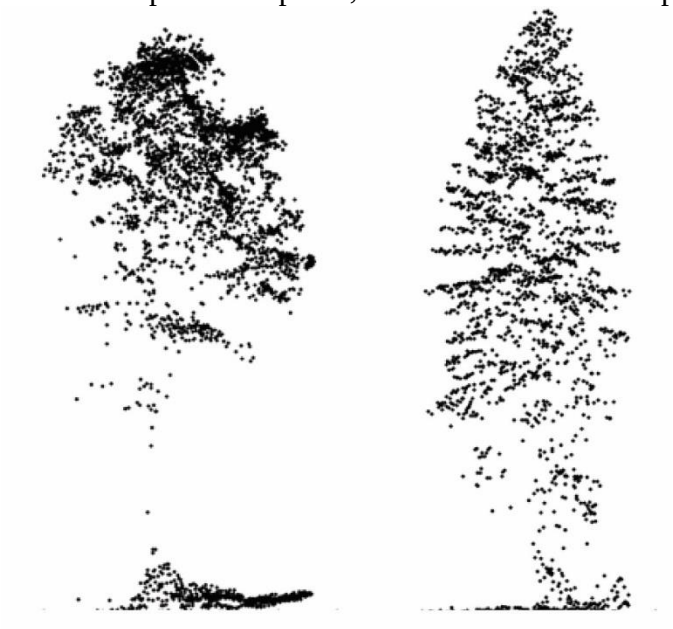
Laser data's point density on the ground is very important for both the level of detail and accuracy of the final ground elevation model. Point density is

[5] Kraus and Pfeifer 1998

affected by the pulse rate, flying altitude and vegetation. Dense vegetation means that fewer pulses hit the ground. The ground elevation model's accuracy is also affected by the accuracy of the distance measurements.

6.1.5. Estimation of variables for single trees

If the pulse density of the laser data is high enough then there will be many returns per tree crown. One can then detect the individual trees and estimate the characteristics for these. There is of course no limit to how many points are needed, but often at least 10 pulses/m2 are needed, although lower densities can also work. Figure 6.15 shows the point cloud profiles from pine and spruce, scanned with about 50 pulses / m2.



Figur 6.15. Point clouds profiles from pine and spruce, the pulse density of about 50 m^{-2}. Image: Johan Holmgren, SLU.

The preprocessing of the data is done in the manner described in Section 6.1.4. Then the work can be roughly divided into the following steps:

• Assign height to all the laser returns over ground level

• Detecting individual trees in the laser data and delineate tree crowns using segmentation

• Calculate the measure from the laser data that describes the individual trees

• Connect the trees in the laser data with trees inventoried and coordinate set in the field

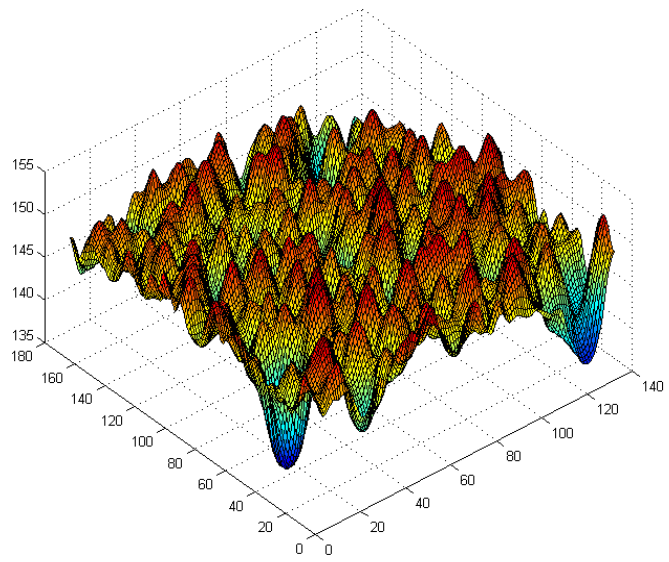
• Develop regression functions for variables to be estimated

• Applying functions on all detected trees.

Many of the processing steps above require special software. Analysis of individual trees in the laser data is therefore primarily done by researchers, as well as a few specialized companies.

6.1.5.1. Detection of trees

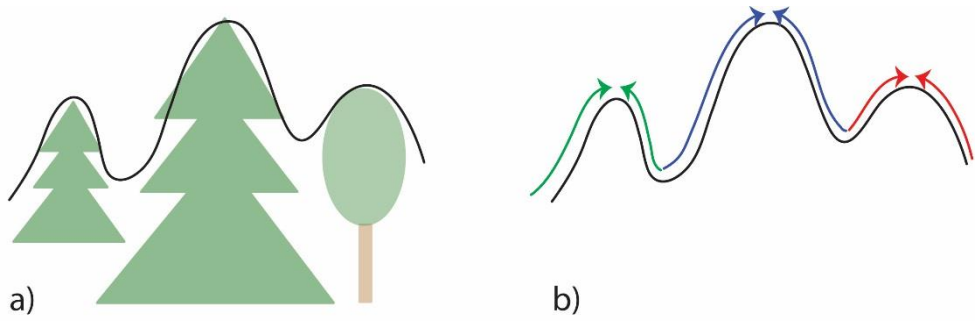
A common method for detecting trees is to first create a canopy height model (Digital Canopy Model, DCM), see Figure 6.16. Local maxima (peaks) in the crown elevation model is used as the basis of a segmentation. The goal is that each tree crown is represented by a segment (not to miss a few trees), and no crown should have more than one segment (to avoid "fake" trees). This means that the crown height model ideally should have exactly one local maximum per tree. A simply prepared crown height model that only follows the highest points in each grid cell does not usually meet this condition because it contains local height variations within the canopy. Therefore it uses some form of smoothing filter. The filter is adjusted to remove height variations within the canopy without removing the entire tree. The trees still missed are often small, hidden by the larger. Two trees standing close together can also be taken for one, while trees with broad crowns and several peaks can be interpreted as several trees.



Figur 6.16. *Crown Height Model with exposed tree tops. © Åsa Persson. FOA, Reproduced with the author's permission.*

A common method for the segmentation of individual trees is called “watershed segmentation” and is illustrated in Figure 6.17. In each grid cell over a certain height a "seed” is placed. The seeds then “climb” upwards in the direction of the steepest slope, until it reaches a point where all the surrounding cells have a lower value. This point is a local maximum and interpreted as a treetop. All seeds that climbs to the same point from the same segment, is assumed to correspond to the propagation of a crown. Alternatively, the segmentation is done in three dimensions, by some form

of clustering in the point cloud or by using so-called voxels, which is the term for three-dimensional pixels. There are a few studies which compare different methods of detecting individual trees.



Figur 6.17. *Delineation of tree crowns using watershed segmentation. a) crown height Model filtered to give a maximum per tree. b) in each pixel of the crown height model is a "seed" that can climb the steepest path to a peak. A segment, or trees, defined as those pixels whose seeds climb to the top.*

Alternatively, the segmentation can be made in three dimensions, through a kind of clustering of the point cloud or by using so-called voxels, which are three-dimensional pixels. There are a few studies comparing different methods of detecting individual trees[6].. The advantage of three-dimensional methods is that small trees in the canopy can be identified and that the segmentation of the canopy is more accurate even in height. The disadvantage is that the methods are often more complicated and that there is a greater risk that individual trees are divided into several segments.

6.1.5.2. Metrics of individual trees

Each segment may have an ID number, which is also given to the laser returns occurring within the segment. You can then connect all laser returns belonging to a tree and calculate different measurements of the tree. This is done for the trees. If segmentation is done in two dimensions, such as the above watershed segmentation, one must also find the crown lower limit in order to not get the points from the undergrowth. Among the benefits of laser estimation of individual trees it can be seen that species can be estimated from the crown shape, and that better information about the tree size distribution can be obtained. Some measures that have proven useful for estimating individual trees are:

- Tree height and crown area from segments
- Height Distribution dimensions in Table 6.1, calculated for individual trees, and the proportion of returns from tree crown

[6] Kaartinen et al. 2012, Vauhkonen et al. 2012, Eysn et al. 2015, Wang et al. 2016

- Proportion of different return types, such as single-, first-and second returns from tree crown
- Texture Measurements derived from DCM and aerial photos
- Intensity (has no clear physical meaning if no calibration can be done, because it depends on many factors including system characteristics)
- Geometric measurement of the tree crown shape, such as the parameters of a parabolic surface that is adapted to a tree crown.

The measurements are used in later steps for estimating stem diameter, stem volume, tree height and tree species classification. Other measures that may be useful for tree species classification is color information from aerial photographs. (Holmgren, Persson, and Söderman 2008). One way is to project segments of the canopy on the aerial images using both color information and measurements from the laser data for each segment.

6.1.5.3. Linking field and remote sensing data

In order to estimate the tree variables from the segments, they must be connected to the field measured tree with known characteristics (height, diameter, etc.). The estimation with individual tree methods is not enough to know the sample plots position - one must also know the individual tree's placement relative to the sample plot. It can be measured for example by ultrasound equipment connected to the autoclave (Haglöfs Postex) or compass. The position of the sample plot center should be measured with DGPS.

So that the pattern of laser measured tree and field measured tree should have the correct geometry, special programs are used to match field surfaces to the laser data on tree level. One method is to match two synthetic images with each other, and are summarized in the following points: [7]:

• Create a synthetic image of the tree pattern according to the laser data from the area around the sample plot

• Create a synthetic image of the tree pattern on the sample plot from field measurements

• Move and rotate "field image" to fit as good as possible with the "laser image
'

• Connect the laser segment with field measured trees so that treetop distances are minimized.

The synthetic images are composed of a dark background where the trees are represented by brighter spots. The brightness is determined in the laser

[7] Olofsson, Lindberg, and Holmgren 2008

image based on segment height, and in the field image by the trees measured diameter because you usually do not measure the height of the trees in the field.

6.1.5.4. Estimation of tree variables

Estimation of height, diameter, volume, etc. of individual trees can be done using, for example, regression or k-MSN, similar to the area-based estimation (Section 6.2). Regression models describing the relationship between the dependent (forest) and independent variables (from lasers or aerial photographs). The relationship is then applied to all detected trees within the scanned area.

Among the benefits of laser estimation of individual trees is that species can be estimated from the tree crown shape, and that better information about the tree size distribution can be obtained. To get unbiased results, meaning that the wood volume, number of stems, etc., on average, will be estimated correctly at the stand level, one must compensate for the trees that are not detected in the laser data. One method called semi-ITC achieves this by connecting all field measured trees to a segment, although in some cases it means that a segment has several field measured tree associated with it (Breidenbach et al. 2010).

6.2. Area-baserade metoder

6.2.1. Calculation unit

The area-based methods use a specific area unit upon which calculations are made, which is often grid cells in a grid covering the laser-scanned area. Grid cell size is determined typically based on a number of practical reasons, such as which field method was/will be used to collect reference data. The aim is often to estimate the totals and averages, for example, timber volume, basal area, average stem diameter and average tree height.

As reference data, circular plots having a size of a few hundred square meters are often used. The reason that a circular sample plot is used is that it is practical to measure in the field in all trees within a certain distance from the sample plot center. The circular sample surface area should be as large as the grid cell area that makes up the grid that covers the laser scanned area. The reason is that the variables should be calculated from laser data from the circular and square units which are as similar as possible.

What size is the best for raster cells depends on several things. They must be large enough to fit a group of trees in them and there should not be too many problems with edge effects, such as a tree outside the plot still has a canopy that covers part of the surface, which may result in the laser measurements indicating that there is a tree height, although there is a lack of trees in the field data plot.

If small plots are used, there may be errors in the positioning of the sample plot which would have a greater impact on the estimates. If large sample plots are used, the risk increases that the trees found in the sample plot / grid

cell have completely different properties, such as the edge of a forest between the old and young forest. The size of the units (grid cells / plots) that is suitable also depends on the point density of the laser data.

The variables for each grid cell from laser data used in the models for estimating forest variables are calculated from the height distribution of the laser data reflections from the forest canopy and requires a certain number of observations so that calculations will be stable. It may be few laser reflections in the canopy if the laser data do not have a hight point density, and if, at the same time, there is also a small number of trees within the calculation unit, and if these trees are small.

6.2.2. Variables calculated from laser data

To calculate the variables from the laser data for each unit, the height of the ground needs to be calculated first. The most common is that the height model is calculated as a grid with cells having a size determined by the distance between the laser measurement on the ground. Then the vertical distance to the ground is measured for each laser reflection. In the calculation of height measurements, one uses as a rule only vegetation hits above a set threshold, eg, 2 m above the ground, to avoid returns from the ground and low vegetation which affect measurements. There may also be reason to calculate the measurements separately, for example, to calculate them separately for the first and last returns.

The term “height percentiles” refers to the height within an area, such as a grid cell, in which a certain percentage of points in the tree layer is found. More explicitly, 10% of the laser points are below the tenth percentile which is expressed as h10. Then 20% are below h20, and so on up to h100 corresponding to the highest point. The strongest correlation with forest height is usually around h90 (the 90^{th} height percentile). The metric of h100 is often avoided because it is susceptible to occasional high trees and faulty laser points, and thus h100 is more dependent on point density than the lower height percentiles.

The proportion of all returned laser pulses reflected from trees is often highly correlated with crown closure. This percentage is called the vegetation ratio, here denoted *V*, and is usually calculated as the number of vegetation hits divided by the total number of laser hits.
So called crown density metrics are calculated by dividing the height difference between the lowest and highest vegetation hits in a number equal fractions, often in intervals of 10. Then the proportion of returns over the respective fraction.Even the vegetation maximum height, average height, standard deviation and coefficient of variation are metrics that are sometimes used.

6.2.3. Factors that influence the laser returns distribution in the canopy

The technology used for the airborne laser scanning has been especially developed for the measurement of infrastructure and land surface. A flat surface is easy to define, and we get a clear return signal from it if we are

sending a short laser pulse to the surface. The technology was developed for measuring ground surfaces even in tall vegetation by detecting multiple echoes from each sending of a laser pulse. It is important to remember that the technology was not developed to measure vegetation and instead treated the data from the vegetation as noise that would be filtered out. However, it was soon clear that airborne laser scanning could also be used for estimating forest variables because the laser beams penetrate the vegetation and describe the vegetation vertical structure.

The distribution of the laser reflection will depend on a variety of factors that are not always correlating with the forest variables that we want to estimate. We can describe the canopy of a variety of surfaces of varying size and location where each surface is not sufficient to reflect a laser pulse. If we have very frequent laser data (more than about five points per m2), we get a large number of measurements for each tree crown and it then becomes possible to geometrically model the individual trees. For area-based methods, we usually have sparser point density laser data and we model therefore not individual tree crowns geometry but use only the vertical distribution of the laser returns from the forest over the area of a grid cell.

We can think about how this vertical distribution arises. A laser pulse is sent down to the vegetation and a photo detector records the return of reflected laser light intensity. The system records the return pulse of an intensity peak which is detected above a certain threshold. Multiple returns can be detected from each sending of a pulse but the residual energy that is provided to detect the intensity peaks is then lower as the pulse has slowed in the upper layers of vegetation.

If an intensity peak is registered, it is due to a variety of system-specific factors: the algorithm used for detecting the intensity peaks, the sensitivity of the photodetector, the amount of energy in each sending of a laser pulse, and the wavelength used for the laser. Often, the system cannot register two return pulses which are close together from the transmitted laser pulse. This phenomenon also makes laser data from many systems less reliable for the measurement of shrub and field layers, as returns from these can interfere rebounds from the ground. For a given system, it is also dependent on the project: for example, a high flying altitude and a high pulse rate means a low energy of each laser pulse before it hits the vegetation.

If a high maximum scanning angle is used, the distribution of the laser reflections also depend on how far from nadir the pulse hits -- as a laser pulse that hits the vegetation at a very sideways angle must travel farther through the vegetation layers to reach the ground. There are also forest-dependent properties which have an influence: for example, distribution of laser returns may be dependent on the tree canopy shape, tree crown density and crown height limit.

We can now imagine the situation that we underestimate average height and timber volume for raster cells with variables calculated from the vertical distribution of the laser returns in the vegetation. We note that the

distribution of the laser returns depends on the distribution of reflective material in vegetation layers and a host of other factors.

When we estimate the average height of the trees, we usually estimate the basal area mean height. This weighted average has traditionally been used for forest inventory long before laser scanning was developed. It is calculated the same way as all weighted averages: Summarize the product of the measured values and the weight and dividing this sum by the sum of the weights. In this case, the weight of the cross sectional area of the tree (which we calculate from the value of the diameter with the assumption that the stem is circular). This average has been useful in forestry estimates because larger trees are important for eg timber volume.

There are studies showing that tree trunk cross-sectional area is proportional to the mass of the tree above the cross-sectional area. The probability that a laser pulse is returned from a certain tree is proportional to the tree crown mass which can be simplified as the size and density. So we have more laser returns from large trees than small trees. We would, according to this reasoning, expect a correlation between the basal area mean tree height and the average height of the laser returns in the canopy.

This relationship can be observed, but it complicates the situation that the laser light penetrates among the branches before a return is detected and the laser beams hits not only the tree tops but also the sides of the canopy. There may be a reason that a stronger correlation has been observed between the upper percentiles of laser returns for the height distribution and the basal area mean tree height.

It is important to note that a model is not used often to describes how the laser return distribution in the canopy is related to tree height distribution. If e.g. the basal area mean tree height is estimated in a two-layered forest, there is a high risk of underestimating because a large number of laser returns can come from the lower tree layer.

To estimate the timber volume, height measurements from laser data are often used, --- normally a percentile of height distribution of the laser returns in the canopy, along with a measure of density. As density measurements can the percentage of laser returns from the canopy be used. For estimates related to the density of the forest, we can use the connection between the total mass crown and the likelihood of a laser return from the canopy.

This relationship is also affected by several factors, of which perhaps the most obvious is illustrated by the difference between the proportion of the laser returns from the canopy of deciduous trees in winter and summer.

6.2.4. Planning of laser scanning

The variables we calculate from the laser data in the vegetation is a function of the vertical distribution of canopy and a host of other factors. Often we are interested in estimating forest variables describing tree stem properties eg, timber volume [m3] per hectare. We examine empirically the variables

calculated from the height distribution of the laser returns in the vegetation that may explain, for example, as much of the variation of the timber volume as possible and use regression or another method to calculate estimates.

If we do not get good estimates it can be useful to think about the physical characteristics of the laser measurement of vegetation and how the technology works. It is good to check as many influencing factors as possible. If we lack models to explain how different factors affect the estimates, it may be desirable to plan the theh laser data acquisition so that as many factors as possible are constant, such as using the same laser scanner systems, altitude, etc, for the entire project.

In Scandinavian coniferous forests dominated areas, laser data estimates have functioned at least as well for estimating forest variables from laser data acquired before leaf-out and after leaves have fallen off as compared to making estimates from laser data acquired during the summer. Laser Data collected from the period of ongoing leaf-out season or when leaves are falling off the trees should be avoided, because the amount of leaves may vary within the project area and between different species. In addition, laser scanning should be avoided when there is snow on the ground, as this affects the height model of the ground.

It is most difficult to control the factors that are determined by the forest that we measure. For this, there may be additional information from other data sources that can be used for stratification. If we have a stand register containing species information, it is possible to use different estimation functions for different kinds of forests, eg., forests dominated by different species.

6.2.5. Planning of field inventory

For both raster cells and plots, variables are calculated from the height distribution of the laser returns in the canopy. Laser and field data from test areas are then connected together so that it is possible to use the laser data to estimate forest variables in grid cells across the scanned area. The results can then be aggregated to the population level by calculating the average of all grid cells contained in a forest.

Sample plots for field data collection are laid out so that the forest variation of the area to be estimated is represented in the material field. A purely systematic sampling, for example in the form of a regular grid of sample plots throughout the region, often resutls in too few plots in the stands of unusual character. One way to get a sufficient number of sample plots from all types of forest, without costs being too high, can instead be to first stratify the area, such as by tree species and/or age. Then there will be a certain number of plots in each stratum according to some objective method. Field and remote sensing (laser and aerial) data should be gathered close together in time. However, there are several advantages to delay the collection of field data until after the scan is made. Only then do you know what areas were really covered with useful data. Laser Data can also be an aid in the selection of field references.

If the sample plots field measured position is not consistent with the laser data the estimate quality can deteriorate significantly. Coordinates of the plots should be made with differential measurements with satellite positioning systems. At differential survey two receivers are used: one located on the sample plot and another that is placed on a well-measured point nearby. Information on the deviation between the measured and known coordinates at the reference station is sent to the device that is used on the sample plot so that the measured position of the sample plot can be corrected. For the attempts within laser scanning of forest that have been reported in the literature, the coordinate of plots has been set to 1 meter or better.[8]. Positioning accuracy is important when the forest is more heterogeneous and when small plots are used.

6.2.6. Estimation methods

Various estimation methods can be used, e.g. Regression[9] and k-NN[10]. methods. For regression estimates, usually only laser data are used and the estimates are usually not broken down by species. Instead, stands with different dominant tree species are estimated for different strata. Regression theoretically gives unbiased estimates, meaning that the value of the estimated variables average is correct. Since regression methods are model-based and allow interpolation (and within reasonable limits even extrapolation) they work with relatively little field data. The sample plots must be representative - for example, there should be a preponderance of plots where the forest is unusually dense in relation to their height. A limitation common when multiple regression is used is that each variable is then estimated separately, which means that unusual combinations of estimated variables may occur.

It has become common to use KNN methods where the laser data and aerial images are combined to produce species-specific estimates. The method is based on imputation, which means that for each grid cell is calculated based on forest data from sample plots that have similar characteristics as the laser data. This requires considerably more field data than regression, evenly spread over the entire range of variation. One advantage of the k-NN methods are that they measure multiple variables simultaneously, which means that there will be a more natural relationship between the estimated variable values, than with regression.

6.3. Future

The lasers that have been used commercially for airborne laser emits a laser pulse in a predetermined direction and a photo detector is used to detect the light reflected back from the messenger pulse. In this way the distance to the reflective object is measured. The emitted pulses are distributed over the area with the help of a mirror and aircraft movement. This means that the number

[8] Næsset 2004; Packalen & Maltamo 2007
[9] Næsset 2002
[10] Packalen & Maltamo 2007

of laser measurements per square meter depends on the laser pulse rate, altitude and scanners sweeping width. There are now commercial systems using an array of detectors instead of a single detector.

In order to understand how they are constructed, they can be compared with the detectors in a digital camera where there are several individual pixels. With these it is possible to split the returned signal in several parts from a pulse sent out. This makes it possible, from a given altitude, to gain much higher resolution laser data compared to using conventional laser systems, which means higher resolution for the same cost as before.

For forestry applications, it is important to quite often be able to collect 3D data for large areas, which has been a limitation with airborne laser scanning using linear detectors. If we use multiple detectors in a matrix, laser data with higher resolution is collected within the same amount of time as before, for example, if an array of 32 × 128 photodetectors are used, we can get 4,096 measurements from each laser pulse sent out.

We can, as previously done, use a scanning system, but each laser beam is divided into sub-beams, with each pulse return detected by a single detector in an array of detectors. [11]. When the laser light is split, the energy that comes back is weaker for each detector. It is possible to use this technique because the detectors are very sensitive - so sensitive that they can detect single photons. That so sensitive detectors are available also means higher flying altitudes can be used because so little energy needs to come back to the sensor. The amount of light energy that comes back is inversely proportional to the square of the distance. There are so-called photon-counting systems that can be used from approximately 10 000 m flyging altitude.

There are two different types of photon-counting systems that can be used to efficiently scan large areas of land. They have various types of detectors. A sort of system has been photo-multiplier tubes (PMT) which are very sensitive photodetectors. However, it is difficult to manufacture these for detecting near-infrared light which is the standard for the measurement of vegetation. Instead green lasers are often used which unfortunately do not reflect equally well from vegetation.

It is possible to measure time very accurately when the photons arrive at the detector and the sensor can detect multiple photons that are close together in time, which means that multiple returns can be detected for each detector and sent laser pulse.

Another kind of system has Geiger-mode Avalanche Photo Diodes (APD) for detection. Geiger-Mode is a way to use the APD so that a fast electrical pulse of several volts can occur when a single photon comes to the sensor. It is not as difficult to manufacture such detectors for near-infrared light. They can also measure time very accurately when a photon comes to the sensor but the sensor cannot usually measure the photons that are too close together

[11] Stoker et al., 2016

in time, for example, if a strong signal coming from the upper part of the crown of the tree, the lower part of the tree cannot be detected from the same pulse. [12]. Both types of detectors are extremely sensitive and, therefore, noise such as caused by sunlight arise, but because so many photons are detected, there are algorithms that can effectively filter the data.

There are now demonstrations where the new systems have shown an efficiency (per unit area) that are up to 30 times compared with traditional technology for airborne laser scanning. [13].

6.4. Terrestrial Laser Scanning

6.4.1. Terrestrial Laser Scanning and other ground based methods

Ground-based laser scanning and terrestrial laser scanning (TLS) is a technique widely used in the industry today. In the past decade there has been an increased interest in using these methods in forestry. There are mainly two aims: to get a quick field inventory and to get more information from the field survey. Considering the weight and measurement speed of the instruments today TLS solutions are nearly as fast as manual measurements. With technological development systems can probably be built that are both easy to carry and faster than manual measurements. In addition, these systems can be mounted on vehicles and thus cover much larger areas than usual. Because TLS measurements sample from the entire tree including trunk, branches and needles / leaves models can be built with high accuracy, on everything from biomass and basal area to the stem volume and stem shapes.

Although the research is new, there has been technological development regarding extraction of forest information from TLS data. For example, Tanaka et al (1998) used a system in which a laser plane was projected on to the forest area and the 3D structure was obtained by triangulation. The parameters extracted were stem *position* and *diameter*, among others.

A widely used technique of TLS systems is to produce data in point form as a list of coordinates and intensity. Hopkinson et al (2004), demonstrated the possibility of using such systems in forest calculations to obtain variables, such as height, volume, and diameter at breast height (DBH). The study used manual methods to extract information but indicated how these forest variables could be obtained automatically in the future.

Other examples are: Tanaka et al (2004) and Hosoi and Omasa (2006), showed that methods of extracting leaf-area-index (LAI) from the TLS data; Gorte and Winterhalder, (2004), extracted branch structure of individual trees and Pfeiffer and Winterhalder (2004), made a more detailed measurements of stem form and measured characteristics such as ovality, and Thies et al (2004) implemented methods to extract taper, curvature and leaning from TLS data.

[12] Stoker et al., 2016
[13] Swatantran et al. 2016.

6.4.1.1. Equipment

A TLS instrument weighs today less than 20 kg and is equipped with a tripod, see Figure 6.18. Some instruments are mounted on vehicles and the technology is called Mobile Laser Scanning (MLS). The wavelengths range from green to infrared. The sampling rate is around 1 000 000 points / second today and technology is developing continuously. TLS instruments can be divided into a number of categories: short, medium and long distances; panoramic scanning and camera scanning; phase measuring system and pulse measuring systems.

Systems built for short distances work within the radius of 50-100 m while the systems for long distances can reach more than 500 m of highly reflective surfaces. Panorama types of instruments usually use rotating mirrors to cover a hemisphere with different sample patterns while the imaging system takes pictures like a conventional camera with the addition that the distance to the object is saved along with the intensity of each camera pixel.

Phase measuring systems transmit a continuous wave (CW) which is amplitude-modulated. The phase difference between the reflected wave and the transmitted wave measured. To determine the ambiguity problem that the distance to an object can be longer than the wavelength used, a number of frequencies and the longest wavelength are factors used to determine the maximum possible measurement distance.

Pulsmätande system eller ”time of flight” (TOF) system mäter tiden det tar för en laserpuls att färdas från instrumentet till föremålet och tillbaka. Eftersom ljusets hastighet är känd går det att extrahera avståndet. Se Shan och Toth (2009) för mer information om lidartekniker.
Pulse measuring systems or "time of flight" (TOF) systems measure the time it takes for a laser pulse to travel from the instrument to the object and back. Since the speed of light is known it is possible to calculate the distance. See Shan and Toth (2009) for more information about LiDAR techniques.

Figur 6.18 The Leica terrestrial laser scanner can be used to scan buildings, rock formations, forests, etc., to produce a 3D model consisting of

a list of coordinates. The instrument can direct the laser beam in a variety of angles: the head rotates horizontally and the mirror rotates vertically. The laser beam is used to measure the distance to the first object in the beam path. Source: Wikimedia Commons.

For imaging at close range, flash laser can also be of interest in the future. Such a system sends out a very short pulse that illuminates the entire surrounding area, unlike a conventional laser scanner that sends out a pulse in a certain direction with a small divergence (spread). The time for the laser pulse which is sent out is measured with a matrix of photodetectors and the time is also measured for the light which is reflected back. In this manner, a 3D image is created without a scanner needing to split single laser pulses over the area to be measured.

Therefore small and inexpensive systems can be constructed and these are sometimes called "time-of-flight cameras" because they can be compared with cameras which measures the time it takes for light to travel to the object and back to the sensor. One advantage is that the whole scene is measured simultaneously as opposed to a scanner where it takes some time to scan the surrounding objects. It can be an advantage if you use mobile systems and / or follow moving objects. Additionally, the entire scene is scanned with a high frequency.

One problem is that much energy or highly sensitive detectors are needed to measure objects at long distances because the energy sent out is distributed over a large area and not that of the scanned system only a small area ("footprint"). We can expect that in the future there are small systems that can be hand held or that we place on eg drones or forestry machinery.

6.4.2. Data characteristics

The output from the TLS systems is usually in the form of a 3D point cloud where each position also has an intensity value (sometimes color values red, green and blue). Some systems also save other metadata such as vegetation class and returns number of the pulse, etc. Data can be saved in text format or in a binary formats such as LAS. (read more on the LAS format on http://www.asprs.org/). The text-based formats are usually lined up in columns with a coordinate or color per column. For small amounts of data it can be read into text based files into a common spreadsheet software but often the data sets are so large that other analysis tools are required.

The TLS system measures the distances to surrounding objects with millimeter resolution based on the emission of laser light and detection of the reflected signals. Depending on the measurement resolution there can be several points per square centimeter covering an object which is ten meters from the sensor, which gives a very detailed map of the trunk and branches of a tree, Figure 6.19.

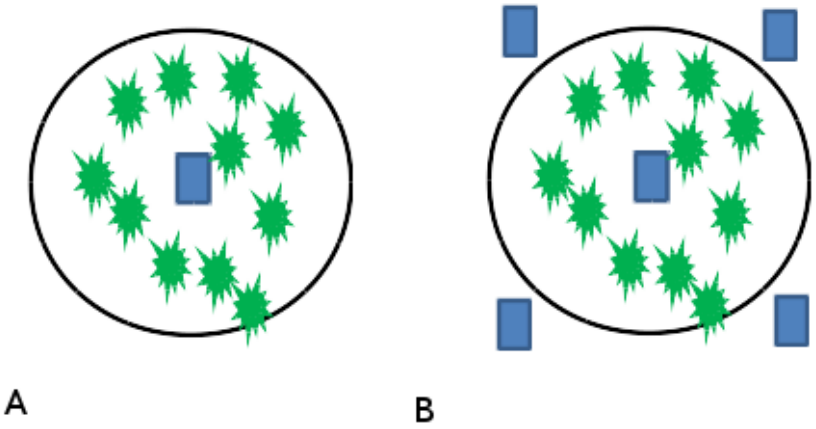
Figur 6.19 Examples of a TLS test area. The round hole is where the instrument has been, so that the self-shading has occurred. Behind the trees there will be shadows where the laser beams do not reach.

When forest information are to be extracted from TLS data, statistical and mathematical software are necessary that can read LAS file format, or the text-based formats. Alternatively, a self-made program developed that reads and analyzes the TLS data can be created and used.
A variety of visualization software may display the TLS data. An example is available online is fugroviewer.com.

6.4.3. Using TLS for field inventory

TLS can be used for field inventory in the same way as the traditional manual methods. The instrument can be placed in the middle of a sample surface and any size sample plot can be selected using the distance data from the scan. Both systematic and random sampling designs can be used. It is also possible to use line-intersect sampling or belt-transect sampling – in that case the scanner could be located on a vehicle or a portable frame.

Om TLS-instrumentet används vid provytesampling finns det två vanliga metoder: ”single scan” och ”multi scan”, figur 6.20. I ”single scan” metoden placeras instrumentet i mitten av provytan och skannar enbart den sida av träden som vetter mot sensorn. I ”multi scan” metoden placeras instrumentet på ett flertal positioner i provytan och ger därför många vyer för varje enskilt träd. Punktdata från de olika vyerna måste rektifieras till ett gemensamt koordinatsystem innan användning.
If a TLS instrument is used in plot inventory, there are two common methods: "single scan" and "multi-scan", Figure 6.20. In "single scan" method the instrument is placed in the middle of the plot and scans only the side of the tree facing the sensor. The "multi-scan" method places the instrument in a variety of positions in the plot and therefore provides many views for each individual tree. Laser points from the different views must be rectified to a common coordinate system before use.



Figur 6.20 (A) Examples of "single scan" method. (B) Example of "multi-scan" method. Position for TLS instrument is marked in blue and the sample plot is marked by a circle. Trees positions are marked with green.

Fördelen med ”multi scan” metoden är att punkttätheten blir större vilket ger en högre detaljeringsgrad. Det är möjligt att täcka alla träd om nog många instrumentpositioner används. Nackdelen är att reflektorer behövs för rektifieringen mot de övriga instrumentpositionerna och att mätningen tar längre tid.

The advantage of the "single scan" method is the measurement speed and ease of use. The instrument needs only one position per sample plot and there is no need for rectifying the second sensor positions. The downside is that some trees will be shaded by other trees, so then some trees will be missing in the analysis. There is a statistical method which can take into account the missing trees.

The advantage of the "multi-scan" method is that point density becomes higher, allowing a higher degree of detail. It is possible to cover all the trees if the instrument is placed in many different positions. The downside is that reflectors are needed for rectification among all the points from all the instruments positions, and that the measurement takes longer.

6.4.4. Automatic extraction of forest variables from TLS data

Output from the TLS instrument contains only the positions and intensities / colors in 3D space and not variables such as basal area and stem volume. To calculate (or extract) forest variables from TLS data, signal and image analysis techniques are used. The most common technique is to extract the tree geometry from the 3D point cloud. If for example a cylinder is adapted to the data points that are far down on the trunk of a tree, basal area and DBH can be calculated with a given accuracy dependent on the quality of the algorithm. If multiple cylinders are adapted along the trunk of the tree, a stem profile can be calculated and if the branches are modeled, the number of branches on the trunk can be calculated. The amount of laser point hits in the branches and foliage can be used to estimate biomass.

Different algorithms are differently refined from simple models of DBH to advanced models where virtually every twig is represented. However, there are four sub-steps that usually are represented in most of the algorithms:

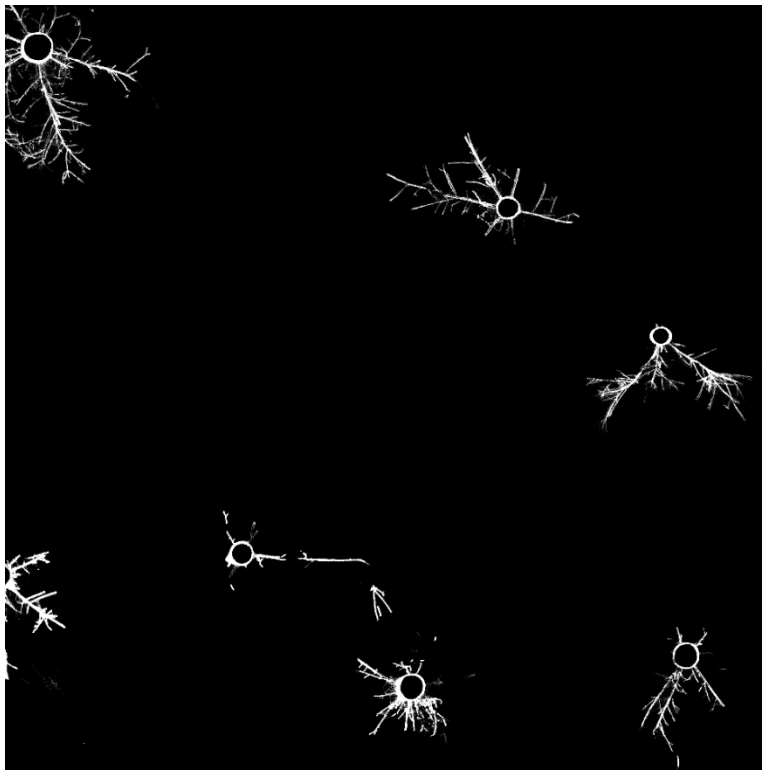
- Calculate a digital elevation model (DEM)
- Separate data points for each tree in the examined the sample plot
- Classify data points for each tree in categories such as trunk, branches and leaves / needles
- Model the detected trees using points for the different classes.

6.4.4.1. Calculating the DEM

Some of the algorithms for tree detection and modeling of TLS data needs a digital elevation model (DEM) as input data to estimate the DBH of the trees in the test area. Many of them create a raster where the height values of the minimum data point within each cell is recorded as land elevation at that position. Simonse et al. (2003), for example, created a grid of 50 × 50 cm 2 cell sizes. Many algorithms have been created to extract them from the airborne laser scanning; some of them can be adapted for TLS data. One of the earliest was developed by Axelsson (2000) and is based on a triangular irregular network adapted to the paragraphs below. The process is iterative and refined step by step where points are rejected if they give an unreasonable ground model.

Isolate data points for every individual tree

There are a number of techniques to separate each individual tree in the TDS data from the test surfaces. One method is to detect the part of the stem which is closest to the ground to find an initial position. It is possible, for example, cut out a section of the point cloud located at chest height Figure 6.21.

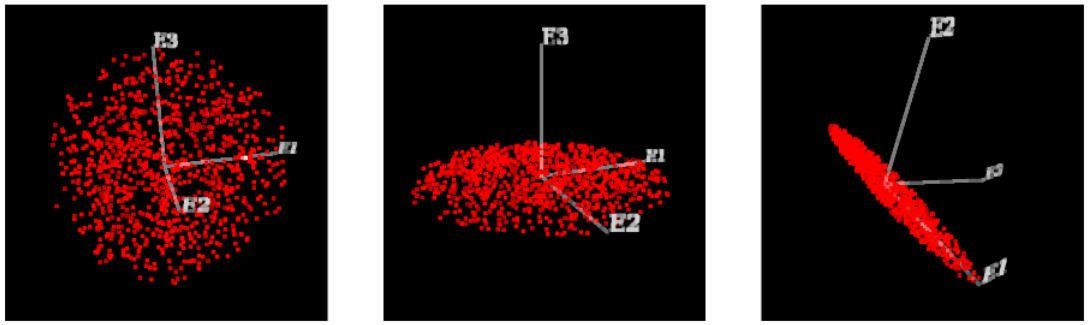


Figur 6.21 A cross-section at the height of 1-1.5 m from the ground level for a TLS data sets. The data points are projected into an image. A large number of points in each pixel provides a high intensity. Stems in the images can be seen as circles surrounded by branches.

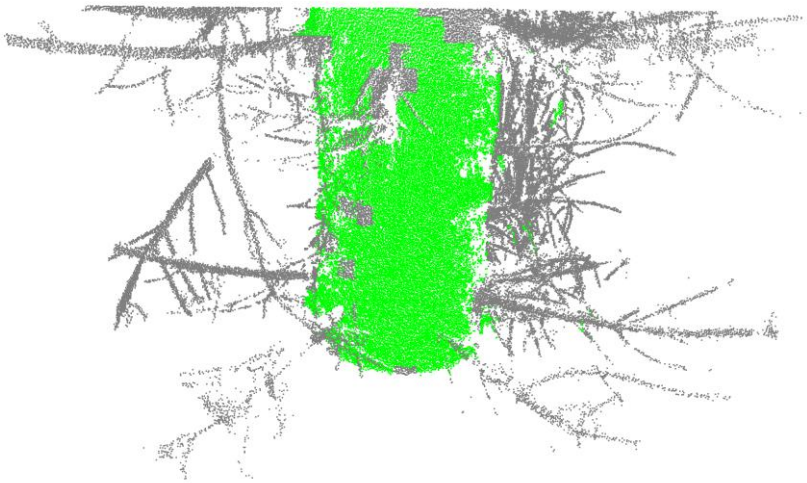
From these cross-sectional images it is possible to find positions for tree trunks. Some methods are very simple measuring only point density in different parts of the image. In those places where there are many measuring points algorithms assume that there is a tree in the middle. Other algorithms are more advanced and are looking for circular shapes in the image. Trees positions assumed to be in the middle of these circles. Hough is such an image-processing algorithm that is looking for geometric shapes in images such as circles.

Once the trees' positions are known, the data points closest to a strain believed to belong to the tree. All data points that are below a certain level can be cut away so that the data points from ground level are not included in the analysis. For the algorithms that have been detected circles tribal forms can be cut dataset belonging to tribes where most of the branches are filtered out. For these cuttings it is possible to model the stems more closely.

One way to classify and demarcate tribes, twigs and pine needles / leaves is to find the characteristics of the spatial distribution of the small parts of the point cloud. They can be classified as spherical, flat and linear using the eigenvalues, Figure 6.22. Flat areas of TLS data can be assumed to be parts of a tree trunk or branch, while coarse linear parts is probably small twigs. By linking many flat parts to the surfaces it is possible to find points belonging to tree trunks, Figure 6.23.

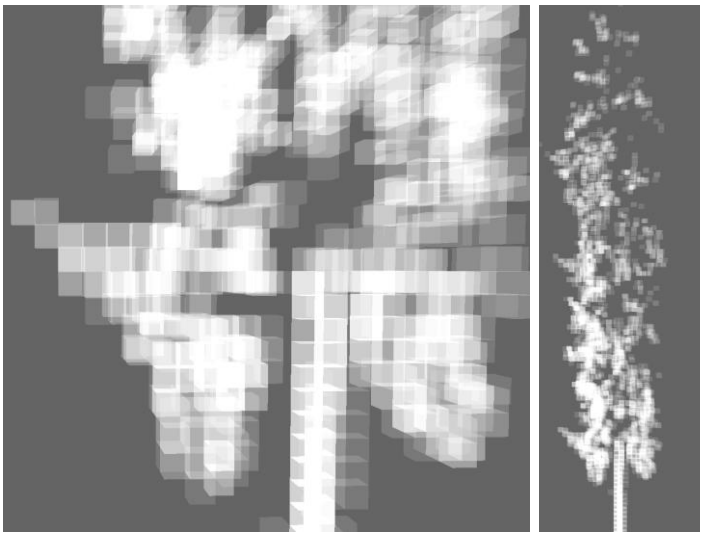


Figur 6.21. Examples of point clouds with different propagation. E1 stands on its own value in the direction of greatest proliferation, E2 between large and E3, the intrinsic value that has the smallest spread. A: A spherical distribution. E1 = E2 = E3. B: Distribution of a plane. E1 = E2> E3 C: Line-shaped distribution. E1> E2, E1> E3.



Figur 6.22. Picture of a part of a tree where the TLS data points belonging to the stem are highlighted in green. The data points located in a flat area is assumed to be part of a stem.

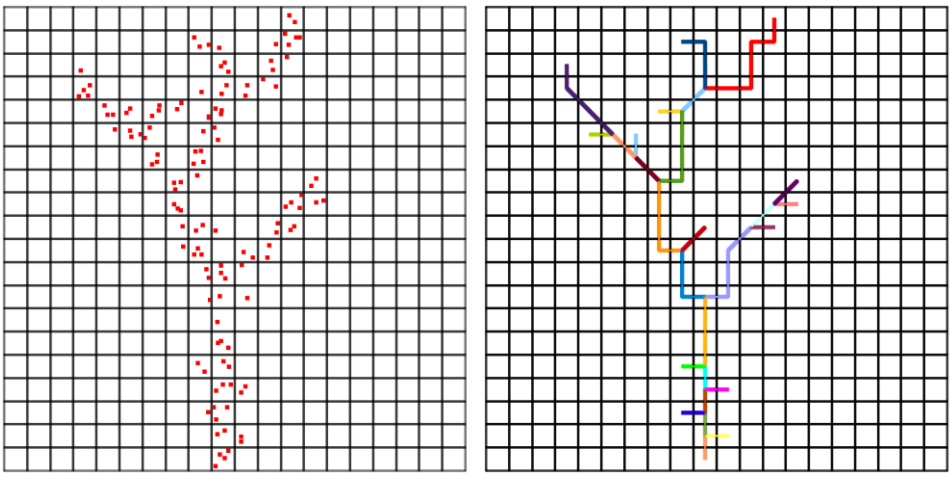
It is also possible to define trees by completing a three-dimensional grid (a voxel space) with point densities, Figure 6.23. By looking for volume pixels (voxels) that are connected to each other it is possible to find the trunks and branches that belong together.



Figur 6.23. Picture of a tree where the points from a TLS data sets stored in cells (volume pixels or voxels) with a high point density saved as a higher intensity. From the data it is possible to find trunks and thick branches.

6.4.4.2. Classification of data points for every tree in different categories

Once the data points for each tree are defined, they can be classified into trunks, branches and foliage. For both the voxel based and surface-based methods it is possible to find the first trunk lying at the bottom, and then follow the stem upward and connect the ramifications that arise. The line starts at the ground classified as trunk while the others become branches. The methods used are within the field of mathematical morphology where Dijkstra's "minimum spanning tree" is a common technique, Figure 6.24.



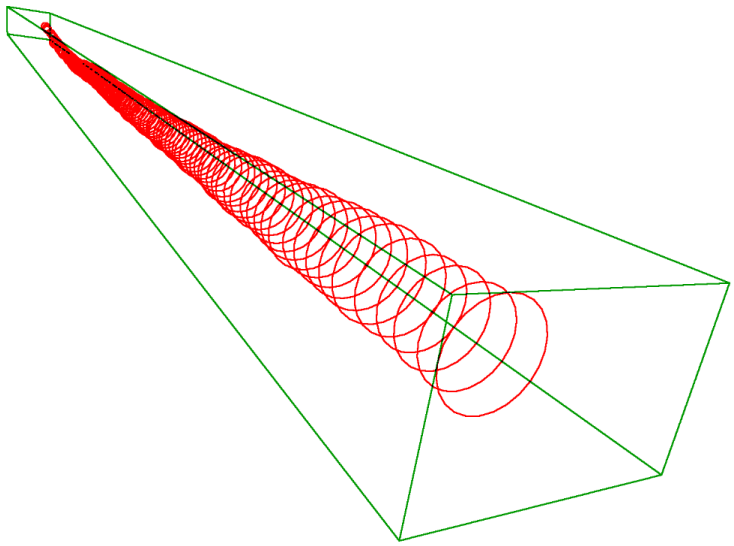
Figur 6.24 If the data points from a TLS measurements are saved in a 3D grid (a voxel volume) as in the picture to the left, it is possible to look for the trunk and branches by connecting the different volumes to a so-called "minimum spanning tree" as in image right.

6.4.4.3. Model the detected trees

When forest variables to be extracted from the point cloud, it is usually an advantage if the geometry of the scanned trees modeled figure 6.x9. Different studies have different refinement of these models: from the simple measurement of DBH to models with twigs and leaves. The most common technique is to adapt circles or cylinders to the data points that belong to the trunks and branches by nonlinear or linearized least squares. Some studies use robust adaptation of the cylinders, for example, Liang et al. (2009.2012) who used "Tukey's estimator".

There are also a number of other techniques to model the tribes. For example, it is possible to find the edges of the trunk of an image. By the distance to the strain is known it is possible to get an estimated diameter.

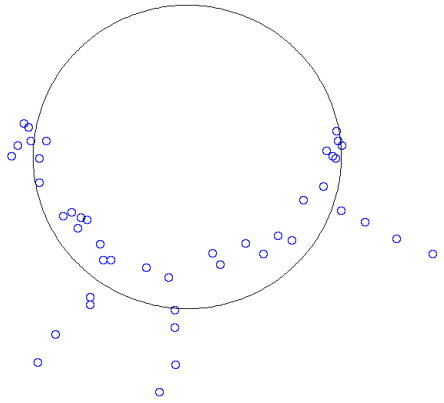
Some techniques model tree trunks with triangular irregular network or various kinds of splines. These more advanced models that show every bump on the surface can be used to obtain a volume of trunk alternatively to obtain an equivalent trunk diameter giving the same volume as the crooked model. For these advanced stem models it is possible to get the measure of quality, such as straightness, taper and ovality.



Figur 6.25. Examples of the stem profile of a tree that is modeled as circles extracted from TLS data.

6.4.4.4. Cirkel/cylinder adaptation of stems by regression

In the regression of circles to a point dataset minimizes the residuals of the points to the circumference of the circle. The position and size of the circle that gives the least error is selected as an estimate of the tribe, Figure 6:26. One problem with the regression is the sensitivity of the data points that unilaterally located on the outside of the circle. If stock items are not properly filtered in advance without a lot of branches and twigs are represented in the selected dataset errors in position and size to be included in the estimates of the stems. There is a need for a robust circle adjustment method for stems to be estimated with a good accuracy.



Figur 6.26. Examples of a circle adapted to a synthetic dataset TLS a strain with branches. The branches affects the position and size of the circle, thereby giving the errors affecting the estimation of the tree trunk. There is a need for a robust circle adjustment method or a good method to filter out branches before the stem's estimation.

THE FOLLOWING NEXT FEW PAGES ARE NOT TRANSLATED TO ENGLISH AND ARE NOT REQUIRED READING. SKIP AHEAD TO THE MOBILE LASER SCANNING SECTION.

6.4.4.5. Cirkelanpassning av stammar genom RANSAC algoritmen

RANSAC är en förkortning för ”Random Sample Consensus” och är en iterativ algoritm. Den är baserad på det faktum att det existerar datapunkter som tillhör modellen (inliers) och de som ej tillhör modellen (outliers). Om ett antal slumpmässiga punkter i ett dataset väljs så finns det en viss sannolikhet att alla tillhör modellen. Ekvation 6.x1 ger det antal iterationer som krävs för att få åtminstone ett sådant fall.

$$N = \log(1-p)/\log(1-w^n) \qquad \text{ekv 6.x1}$$

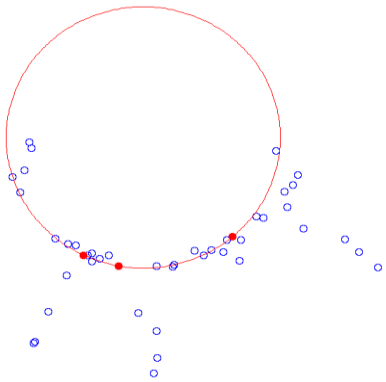
där *N* är antalet iterationer där åtminstone en bra modell är hittad med sannolikheten *p*. Antalet valda datapunkter är *n och w* är sannolikheten att en punkt tillhör modellen. Parameter *w* ges av ekvation 6.x2.

$$w = M / D \qquad \text{ekv 6.x2}$$

där *M* är antalet ”inliers” i modellen och *D* är antalet punkter i datasetet. Vanligtvis är denna kvot ej känd utan måste estimeras från ett datatset. Genom att studera ett typiskt dataset är det möjligt att erhålla en grov uppskattning av denna kvot.

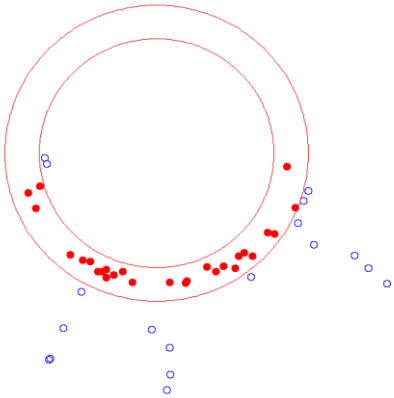
Trädstammar kan modelleras med cirklar eller cylindrar vilket gör dem till ett intressant fall för RANSAC algoritmen. Vid modellering av cirklar krävs minst tre datapunkter. Det gör parameter *n*=3 i RANSAC algoritmen, figur 6.x11. När en cirkel anpassas till tre datapunkter kan exempelvis en linje vinkelrätt mot kordan mellan två punkter på cirkelperiferin hittas.

Skärningen mellan två sådana linjer ger centrum för cirkeln om två korda används.
För att estimera parameter *w* i ett TLS dataset måste kvoten mellan antalet punkter i stammen och totala antalet punkter inkluderat grenarna bestämmas. Antal iterationer bestäms genom att sätta parameter *p* i ekvation x1 till en hög sannolikhet och genom att multiplicera parameter *N* med en säkerhetsfaktor beroende på hur många bra modeller som behöver hittas för att erhålla ett bra estimat.



Figur 6.x11. Välj tre slumpmässiga punkter ur ett dataset vid varje iteration i RANSAC algoritmen. Skapa en cirkel från de utvalda punkterna.

För varje iteration hittas en cirkel genom tre slumpmässiga datapunkter. Inom en given tolerans väljs datapunkter nära cirkelns periferi till möjliga ”inliers”, figur 6.x12. Om antalet möjliga ”inliers” är större än ett förvalt tröskelvärde behålls modellen som en bra modell och sparas för efterföljande beräkningar. För alla bra modeller cirkelanpassas punkterna som är ”inliers” exempelvis genom regression. Behåll den cirkel som ger det minsta felet vilket då blir den estimerade cirkeln av RANSAC algoritmen, figur 6.x13.



Figur 6.x12. I varje iteration i RANSAC algoritmen hittas ”inliers” med en given tolerans för den valda cirkeln. Om antalet ”inliers” är större än ett given tröskelvärde behålls modellen för vidare beräkningar.

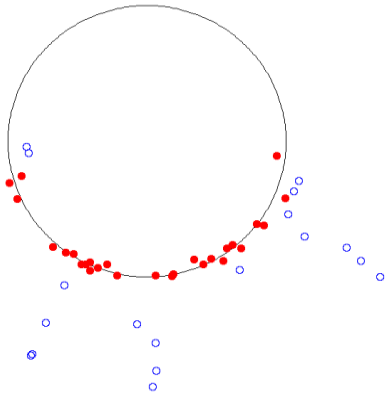
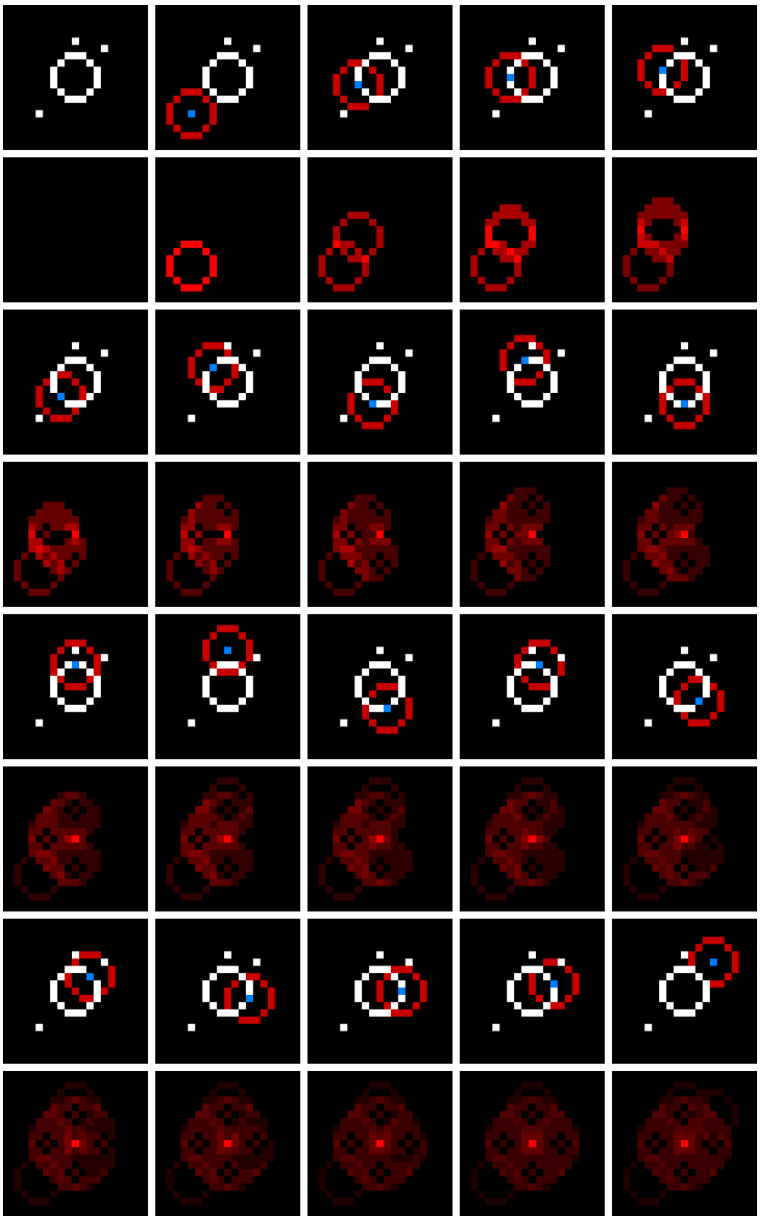


Figure 6.x13 Gör en cirkelanpassning för varje modell med ett högt antal ”inliers” i RANSAC algoritmen, genom att använda de utvalda punkterna exempelvis genom regression. Se om felen är mindre än i föregående iteration. Behåll modellen med de minsta felen.

6.4.4.6. Cirkelanpassning av stammar genom Hough-transformen

Att använda Hough-transformen till cirkelanpassning av trädstammar bygger på att ett antal datapunkter klipps ut vid ett visst höjdintervall och projiceras till en bild som exempelvis i figur x4. Houghtransformen bygger på att varje pixel i bilden röstar på var cirkelcentrum ska ligga. Om många pixlar röstar på samma position är det förmodligen en trolig kandidat till ett cirkelcentrum. En radieklass studeras åt gången, figur 6.x12. Om exempelvis cirklar med radien tre pixlar eftersöks så ligger ett troligt centrum tre pixlar ifrån den undersökta datapunkten. De troliga platserna för ett centrum kommer att ligga i en cirkel runt datapunkten, markerat i rött i figur 6.x14. Genom att summera alla troliga centrumpositioner kommer den mest rimliga positionen att ha flest röster.



Figur 6.x14. De vita pixlarna är det data där algoritmen letar efter cirklar. De röda pixlarna visar på möjliga cirkelcentrum. De blå pixlarna markerar den pixel som undersöks i respektive bild. Övre vänstra hörnet: Leta efter cirklar av en utvald storleksklass exempelvis med en radie på 3 pixlar. Nästa bild till höger: Undersök den blåfärgade pixeln. De röda pixlarna runt den blå markerar områden där ett möjligt centrum för en cirkel skulle kunna ligga. Spara de möjliga centrumkoordinaterna och fortsätt undersöka nästa vita pixel. Nedre högra hörnet: Efter att ha undersökt alla pixlar i bilden har många av pixlarna röstat för området i mitten som ett troligt ställe för centrum av en cirkel med radie 3.

6.5. Mobile laser scanning

A mobile laser scanner system consists of the following components:
• (1) laser scanner,
• (2) positioning, and
• (3) other equipment such as computer and batteries.

All parts including a scanner, for example, mounted on a frame that is attached to a sling and scanner mounted on top of the frame with a clear view in all directions (see Figure 6.30). The mobile system resembles an airborne system so that a positioning needed to continuously measure the position of the sensor. For an airborne system, a satellite navigation system combined with inertial (accelerometers and gyroscopes). Using inertial changes in movement and orientation calculated with high precision and high frequency system, but affected by the operation, which means that the wrong absolute position calculated after only a few seconds. Therefore, data from a satellite navigation system (which does not have as high frequency but a position calculation without operation) to continuously calibrate the absolute position.

Figur 6.30. An experimental mobile laser scanning system with laser scanner, stereo camera, inertial navigation and computer for data logging and wireless communication network to other devices.

For mobile systems that are portable or mounted on forestry for the problem of signals from satellites often are not as good under the canopy. There are various ways that this problem can be solved. One way is to use stereo cameras that collect a sequence of images with high frequency. Algorithms

can be used to automatically identify the pixels depicting the object in subsequent images and 3D coordinates can be calculated for points on these items. By minimizing the distances between points that are associated to the same object is calculated camera position at a time relative position of the camera at a different time. In this way, the sensor movement is expected in 6 dimensions (x, y, z and three rotational angles). This calculation can be used as the start value for further calibration.

From the system, we also get data from a scanner that sweeps 360 degrees several times in every second and sends out several hundred thousand laser pulses every second. If we have two point clouds from the subsequent sweep, the algorithm "Iterative Closest Point" (ICP) used. This algorithm consists of the following steps:

- (1) find correspondence - that associate points in pairs to each other,
- (2) find a function that minimizes the distances between the points of the two data sets,
- (3) calculate new coordinates of the feature from the previous step,
- (4) Repeat steps 1-3 until the error is small, or the maximum number of iterations.

The easiest way to associate points to each other is to connect the points that are closest to each other. There are other ways to associate points e.g. "Point-to-plane" in which a surface is determined by the surrounding dots. The distance between the pairwise registrations is then calculated from a point in one point cloud, and along the normal to the surface of the second point cloud. In this way the plane surfaces slide over each other so that the structures are better matched.

The algorithms used to detect and estimate the tree trunk diameter similar to the algorithms used for stationary ground-based laser scanning (TLS). However, it is not certain that it is best to use the exact same algorithms. One advantage of the mobile laser scanning is that the trees are measured from many different directions, which means that the risk is reduced for zones that are obscured by other trees and there is a greater chance of trees upper parts, such as tree top, can be measured well from any direction.

The errors that come from calculation of the position of the sensor means that the point cloud's accuracy will be lower, especially if satellite navigation systems are used to support the INS in the dense forest. Estimation Errors have proved to be greater than what is common in statistical measurement. [14]. If many measurements on the tree trunk and the error is normally

[14] Liang et al. 2014

distributed, it is not certain that the estimation error becomes large. A comparative study showed that data from a handheld laser scanner was better compared with data from static laser scanning to estimate trunk diameter, a probable cause is that the measurements from the handheld system were around the entire trunk. [15].

As examples of how data from mobile systems can be analyzed is the "region growing" algorithm. There is an algorithm to segment the point cloud so that each 3D point given an ID number is defined as belonging to a particular object. First a threshold is determined for the mean square deviation (s) from a surface, and then the maximum angle difference (α) between two normals. The two variables r and α are calculated for each 3D point using a local area. This neighborhood may be determined in two ways: K-Nearest Neighbors (kNN), and Fixed Distance Neighbors (FDN). For kNN the k nearest neighbors to the current point are selected.

If we assume that the point density indicates the measurement accuracy, which is often the case because the point density is inversely proportional to the distance between the scanner, this is a good way to define the calculation area normals because a larger area selected for the longer distance of the sensor. It is also possible to avoid cases where the algorithm collapses because of the low point density in a certain area. For FDN selected all neighbors within a certain distance. This method is suitable for point density is constant within the measuring range.[16].

The following steps are used:

- (1) A surface is calculated using points and the normal and the mean square deviation of the surface is calculated,

- (2) The point that has a surrounding surface with a minimum mean square error is selected as the first starting point,

- (3) The items found in the area included in the region of the angular difference between their normals and the current punkterns normally not greater than α. The points have a mean square error which is less than r is placed in a list for possible start values,

- (4) If the list of possible starting values are not even going to the next available start value and repeat steps 3,

- (5) Include the region segmentation result and go to step 2.

[15] Bauwens et al. 2016
[16] Rabbani et al. 2006

This is repeated until the list of items not included in the segmentation is empty then the list of 3D points that have been segmented are sorted in order of magnitude. One advantage of this algorithm is that it has so many parameters and these parameters can be explained physically. Depending on how the parameters are selected there may be an over-segmentation (many segments) or under-segmentation (few segments). In steps after segmentation geometric objects are modeled, and it may be easier to detect under-segmentation compared to the over-segmentation[17].

Litteraturen

Albota, M.A., Aull, B.F., Fouche, D.G., Heinrichs, R.M., Kocher, D.G., Marino, R.M., Mooney, J.G., Newbury, N.R., O'Brien, M.E. and Player, B.E., 2002. Three-dimensional imaging laser radars with Geiger-mode avalanche photodiode arrays. Lincoln Laboratory Journal, 13(2): 351-370.

Aschoff, T. and Spiecker, H., 2004 B. Algorithms for the automatic detection of trees in laser scanner data . In proceedings of the ISPRS workshop Laser-Scanners for Forest and Landscape Assessment, October 3-6, 2004, Freiburg, Germany.

Aschoff, T., Thies, M., and Spiecker, H., 2004 A, Describing forest Stands using Terrestrial Laser-scanning. In proceedings of the XXth ISPRS Congress, July 12-23, 2004, Istanbul, Turkey, 237–241.

Aull, B.F., Loomis, A.H., Young, D.J., Heinrichs, R.M., Felton, B.J., Daniels, P.J. and Landers, D.J., 2002. Geiger-mode avalanche photodiodes for three-dimensional imaging. Lincoln Laboratory Journal, 13(2): 335-349.

Axelsson, P. 2000. "DEM generation from laser scanner data using adaptive TIN models." International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Amsterdam, Netherlands, 16-22 July, 2000.

Axelsson, P. 2000. DEM generation from laser scanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing, Vol. 35, Part B4/1:236-241.

Bauwens, S., Bartholomeus, H., Calders, K. and Lejeune, P., 2016. Forest Inventory with Terrestrial LiDAR: A Comparison of Static and Hand-Held Mobile Laser Scanning. Forests, 7(6): 127.

Bienert, A., Scheller, S., Keane, E., Mohan, F., and Nugent, C., 2007, Tree Detection and Diameter Estimations by Analysis of Forest Terrestrial Laserscanner Point Clouds. In proceedings of the ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007, Espoo, September 12-14, 2007, Finland, 50-55.

Bienert, A.; Scheller, S.; Keane, E.; Mullooly, G.; Mohan, F. Application of terrestrial laser scanners for the determination of forest inventory parameters. International Archives of

Binney, J.; Sukhatme, G.S. 3D Tree Reconstruction from Laser Range Data. In Proceedings of the IEEE International Conference on Robotics and Automation, Robotics and Automation (ICRA'09), Kobe, Japan, 12–17 May 2009; pp. 1321–1326.

[17] Rabbani et al. 2006

Breidenbach, Johannes, Erik Næsset, Vegard Lien, Terje Gobakken, and Svein Solberg. 2010. "Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data." Remote Sensing of Environment 114 (4):911-924. doi: 10.1016/j.rse.2009.12.004.

Brethvad, Thomas, and Erik Heimdal Iversen. 2012. "Nyindelning av Bergvik Skog." Stockholm, Sverige, 24-25 april, 2012.

Brolly, G. and Király , G., 2009. Algorithms for Stem Mapping by Means of Terrestrial Laser Scanning . Acta Silv. Lign. Hung., Vol. 5 (2009) 119-130 .

Bucksch, A.; Lindenbergh, R. CAMPINO—A skeletonization method for point cloud processing. ISPRS J. Photogramm. Remote Sens. 2008, 63, 115–127.

Bucksch, A.; Lindenbergh, R.; Menenti, M. SkelTre: Robust skeleton extraction from imperfect point clouds. Vis. Comput. 2010, 26, 1283–1300.

Chum, O., 2005. Two-View Geometry Estimation by Random Sampling and Consensus. PhD-thesis. Center for Machine Perception Department of Cybernetics Faculty of Electrical Engineering Czech Technical University in Prague .

Clawges, R.; Vierling, L.; Calhoon, M.; Toomey, M. Use of a groundbased scanning lidar for estimation of biophysical properties of western larch (Larix occidentalis). Int. J. Remote Sens.

Côté, J.F.; Widlowski, J.L.; Fournier, R.A.; Verstraete, M.M. The structural and radiative consistency of three-dimensional tree reconstructions from terrestrial lidar. Remote Sens. Environ. 2009, 113, 1067–1081.

Eitel, J.U.; Vierling, L.A.; Long, D.S. Simultaneous measurements of plant structure and chlorophyll content in broadleaf saplings with a terrestrial laser scanner. Remote Sens. Environ. 2010, 114, 2229–2237.

Erikson, M. and Vestlund, K. 2003. Finding tree stems in laser range images of young mixed stands to perform selective cleaning. Scandlaser 2003, Umeå Sweden.

Eysn, Lothar, Markus Hollaus, Eva Lindberg, Frédéric Berger, Jean-Matthieu Monnet, Michele Dalponte, Milan Kobal, Marco Pellegrini, Emanuele Lingua, Domen Mongus, and Norbert Pfeifer. 2015. "A Benchmark of LiDAR-Based Single Tree Detection Methods Using Heterogeneous Forest Data from the Alpine Space." Forests 6 (5):1721.

Forsman, P. and Halme , A., 2005. 3-D Mapping of Natural Environments With Trees by Means of Mobile Perception . IEEE transactions on robotics, vol. 21, no. 3, june 2005.

Freiburg, Germany, 3–6 October 2004; Volume 36.

Gorte, B. and Winterhalder, D., 2004. Reconstruction of laser-scanned trees using filter operations in the 3d raster domain. In proceedings of the ISPRS workshop Laser-Scanners for Forest and Landscape Assessment, October 3-6, 2004, Freiburg, Germany.

Gorte, B.; Pfeifer, N. Structuring laser-scanned trees using 3D mathematical morphology. Int. Arch. Photogramm. Remote Sens. 2004, 35, 929–933.

Haala, N., Reulke, R., Thies, M., and Aschoff, T., 2004, Combination of terrestrial laser scanning with high resolution panoramic images for

investigations in forest applications and tree species recognition. In proceedings of the ISPRS Panoramic Photogrammetry Workshop, February 19-22, 2004, Dresden, Germany.

Hackenberg, J., Spiecker, H., Calders, K., Disney, M. And Raumonen, P., 2015, SimpleTree-An efficient Open Source Tool to Build Tree Models from TLS Clouds. 6(11) 4245-4294.

Henning, J.G. and Radtke, P.J., 2006, Detailed stem measurements of standing trees from ground-based scanning lidar. Forest Science, 52, 67–80.

Holmgren, J., Å. Persson, and U. Söderman. 2008. "Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images." International Journal of Remote Sensing 29 (5):1537-1552. doi: 10.1080/01431160701736471.

Hopkinson, C., Chasmer, L., Young-Pow, C., and Treitz, P., 2004, Assessing forest metrics with a ground-based scanning lidar. Canadian Journal of Forest Research-Revue Canadienne de Recherche Forestiere, 34, 573–583.

Hosoi, F, and Omasa, K, 2006. Voxel-Based 3-D Modeling of Individual Trees for Estimating Leaf Area Density Using High-Resolution Portable Scanning Lidar . IEEE Transactions on geoscience and remote sensing, vol. 44, no. 12, 3610-3618.

Hyyppä, Juha, Mathius Schardt, Henrik Haggrén, Barbara Koch, Uwe Lohr, Raito Paananen, H. U. Scherrer, Heikki Luukkonen, Michaela Ziegler, Hannu Hyyppää, Ulla Pyysalo, Hans Friedländer, Janne Uuttera, Stephen Wagner, Mikko Inkinen, Andreas Wimmer, Antero Kukko, Eero Ahokas, and Mika Karjalainen. 2001. "HIGH-SCAN: The first European-wide attempt to derive single-tree information from laserscanner data." The Photogrammetric Journal of Finland 17 (2):58-68. doi: citeulike-article-id:8833454.

individual tree using ground-based lidar data. Can. J. Remote Sens. 2008, 34, 320–332.

Jupp, D.L.; Culvenor, D.; Lovell, J.; Newnham, G.; Strahler, A.; Woodcock, C. Estimating forest LAI profiles and structural parameters using a ground-based laser called ‘Echidna(R). Tree Physiol.

Kaartinen, Harri, Juha Hyyppä, Xiaowei Yu, Mikko Vastaranta, Hannu Hyyppä, Antero Kukko, Markus Holopainen, Christian Heipke, Manuela Hirschmugl, Felix Morsdorf, Erik Næsset, Juho Pitkänen, Sorin Popescu, Svein Solberg, Bernd Michael Wolf, and Jee-Cheng Wu. 2012. "An International Comparison of Individual Tree Detection and Extraction Using Airborne Laser Scanning." Remote Sensing 4 (4):950-974.

Király, G.; Brolly, G. Tree height estimation methods for terrestrial laser scanning in a forest reserve. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2007, 36, 3, 211–215.

Kraus, K., and N. Pfeifer. 1998. "Determination of terrain models in wooded areas with airborne laser scanner data." ISPRS Journal of Photogrammetry and Remote Sensing 53 (4):193-203.

Lalonde, J.F.; Vandapel, N.; Huber, D.; Hebert, M. Natural terrain classification using three-dimensional ladar data for ground robot mobility. J. Field Robot. 2006, 23, 839–861.

Liang, X Litkey, P, Hyyppä, J, Kaartinen, H., Vastaranta, M., and Holopainen, M., 2009. Automatic stem location mapping using TLS for plot-wise forest inventory . SilviLaser 2009, Oct. 14-16, 2009 – College Station, Texas, USA .

Liang, X., Kukko, A., Kaartinen, H., Hyyppä, J., Yu, X., Jaakkola, A. and Wang, Y., 2014. Possibilities of a personal laser scanning system for forest mapping and ecosystem services. Sensors, 14(1): 1228-1248.

Liang, X.; Litkey, P.; Hyyppa, J.; Kaartinen, H.; Vastaranta, M.; Holopainen, M. 2012. Automatic Stem Mapping Using Single-Scan Terrestrial Laser Scanning. IEEE Transactions on Geoscience and Remote Sensing 2012, 50, 661–670.

Lindberg, E., Holmgren, J., Olofsson, K., and Olsson , H., 2010. Estimation of stem attributes using a combination of terrestrial and airborne laser scanning . Silvilaser 2010, Freiburg, Tyskland. Moorthy, I.; Miller, J.R.; Hu, B.; Chen, J.; Li, Q. Retrieving crown leaf area index from an

MacLean, G. A., and W. B. Krabill. 1986. "Gross-merchantable timber volume estimation using an airborne LiDAR system." Canadian Journal of Remote Sensing 12 (1):7-18.

Naesset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sensing of Environment, 80(1): 88-99.

Naesset, E., 2004. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. Scandinavian Journal of Forest Research, 19(2): 164-179.

Næsset, E., T. Gobakken, J. Holmgren, H. Hyyppä, J. Hyyppä, M. Maltamo, M. Nilsson, H. Olsson, A. Persson, and U. Söderman. 2004. "Laser scanning of forest resources: The Nordic experience." Scandinavian Journal of Forest Research 19 (6):482-499. doi: 10.1080/02827580410019553.

Nelson, R., W. Krabill, and G. Maclean. 1984. "Determining forest canopy characteristics using airborne laser data." Remote Sensing of Environment 15 (3):201-212. doi: 10.1016/0034-4257(84)90031-2.

Nilsson, M. 1994. "Estimation of tree heights and stand volume using airborne LiDAR systems." Rapport - Institutionen for Skogstaxering, Sveriges Lantbruksuniversitet (57).

Nilsson, M. 1996. "Estimation of tree heights and stand volume using an airborne LiDAR system." Remote Sensing of Environment 56 (1):1-7. doi: 10.1016/0034-4257(95)00224-3.

Olofsson, K., E. Lindberg and Holmgren, J., 2008. A method for linking field-surveyed and aerial-detected single trees using cross correlation of position images and the optimization of weighted tree list graphs. SilviLaser 2008, 8th international conference on LiDAR applications in forest assessment and inventory., Heriot-Watt University, Edinburgh, UK, SilviLaser 2008 Organizing Committee, Edinburgh: Forest Research.

Olofsson, K., E. Lindberg, and J. Holmgren. 2008. "A method for linking field-surveyed and aerial-detected single trees using cross correlation of position images and the optimization of weighted tree list graphs." Proceedings of SilviLaser 2008, 8th international conference on LiDAR

applications in forest assessment and inventory, Heriot-Watt University, Edinburgh, UK, 17–19 September, 2008.

Olofsson, K., Holmgren, J. and Olsson, H., 2014. Tree Stem and Height Measurements using Terrestrial Laser Scanning and the RANSAC algorithm, Remote Sensing, 6, 4323-4344.

Olsson, P., H. Rost, and Y Reshetyuk. 2013.

Packalen, P. and Maltamo, M., 2007. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. Remote Sensing of Environment, 109(3): 328-341.

Petrie, Gordon, and Charles K. Toth. 2009a. "Airborne and Spaceborne Laser Profilers and Scanners." In Topographic Laser Ranging and Scanning: Principles and Processing, edited by J. Shan and C. Toth. Boca Raton, FL: CRC Press/Taylor & Francis Group.

Petrie, Gordon, and Charles K. Toth. 2009b. "Introduction to Laser Ranging, Profiling, and Scanning." In Topographic Laser Ranging and Scanning: Principles and Processing, edited by J. Shan and C. Toth. Boca Raton, FL: CRC Press/Taylor & Francis Group.

Pfeifer, N. och Winterhalder, D., 2004, Modelling of Tree Cross Sections from Terrestrial Laser-Scanning Data with Free-Form Curves. In proceedings of the ISPRS workshop Laser-Scanners for Forest and Landscape Assessment, October 3-6, 2004, Freiburg, Germany.

Pfeifer, N.; Gorte, B.; Winterhalder, D. Automatic Reconstruction of Single Trees from Terrestrial Laser Scanner Data. In Proceedings of 20th ISPRS Congress, Istanbul, Turkey, 12–23 July 2004; pp. 114–119.

Photogrammetry, Remote Sensing and Spatial Information Sciences 2006, 36

Pueschel, P.; Newnham, G.; Rock, G.; Udelhoven, T.; Werner, W.; Hill, J. The influence of scan mode and circle fitting on tree stem detection, stem diameter and volume extraction from terrestrial laser scans. ISPRS J. Photogramm. Remote Sens. 2013, 77, 44–56.

Rabbani, T., van Den Heuvel, F. and Vosselmann, G., 2006. Segmentation of point clouds using smoothness constraint. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36(5): 248-253.

Raumonen, P.; Kaasalainen, M.; Åkerblom, M.; Kaasalainen, S.; Kaartinen, H.; Vastaranta, M.; Holopainen, M.; Disney, M.; Lewis, P. Fast automatic precision tree models from terrestrial laser scanner data. Remote Sens. 2013, 5, 491–520.

Rutzinger, M.; Pratihast, A.K.; Oude Elberink, S.; Vosselman, G. Detection and modelling of 3D trees from mobile laser scanning data. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2010, 38, 520–525.

Schütt, C.; Aschoff, T.; Winterhalder, D.; Thies, M.; Kretschmer, U.; Spiecker, H. Approaches for Recognition of Wood Quality of Standing Trees b Based on Terrestrial Laser Scanner Data. In Proceedings of the 2004 WG VIII/2 Laser-Scanners for Forest and Landscape Assessment,

Shan, J. and Toth, C, 2009. Topographic laser ranging and scanning. Principles and processing. Taylor & Francis Group, London New York.

Simonse, M., Aschoff, T., Spiecker, H. and Thies, M. 2003. Automatic determination of forest inventory parameters using terrestrial laser scanning. Proceedings of the ScandLaser Scientific Workshop on

Airborne Laser Scanning of Forests, Umea, Sweden, September 2003, pp. 251-257.

Stoker, J., Abdullah, Q., Nayegandhi, A. and Winehouse, J., 2016. Evaluation of Single Photon and Geiger Mode Lidar for the 3D Elevation Program. Remote Sensing, 8(9): 767.

Strahler, A.H.; Jupp, D.L.; Woodcock, C.E.; Schaaf, C.B.; Yao, T.; Zhao, F.; Yang, X.; Lovell, J.; Culvenor, D.; Newnham, G.; et al. Retrieval of forest structural parameters using a ground-based lidar instrument (Echidna). Can. J. Remote Sens. 2008, 34, S426–S440.

Swatantran, A., Tang, H., Barrett, T., DeCola, P. and Dubayah, R., 2016. Rapid, High-Resolution Forest Structure and Terrain Mapping over Large Areas using Single Photon Lidar. Scientific Reports, 6.

Tanaka, T., Park, H., and Hattori, S., 2004, Measurement of forest canopy structure by a laser plane range-finding method. Improvement of radiative resolution and examples of its application. Agricultural and Forest Meteorology, 125, 129–142.

Tanaka, T., Yamaguchi, J., and Takeda, Y., 1998, Measurement of forest canopy structure with a laser plane range-finding method - development of a measurement system and applications to real forests. Agricultural and Forest Meteorology, 91, 149–160

Thies, M., Pfeifer, N., Winterhalder, D. and Gorte, B.G.H., 2004, Three-Dimensional Reconstruction of Stems for Assessment of Taper, Sweep and Lean Based on Laser Scanning of Standing Trees. Scandinavian Journal of Forest Research, 19, 571–581

Vauhkonen, Jari, Liviu Ene, Sandeep Gupta, Johannes Heinzel, Johan Holmgren, Juho Pitkanen, Svein Solberg, Yunsheng Wang, Holger Weinacker, K. Marius Hauglin, Vegard Lien, Petteri Packalen, Terje Gobakken, Barbara Koch, Erik Naesset, Timo Tokola, and Matti Maltamo. 2012. "Comparative testing of single-tree detection algorithms under different types of forest." Forestry 85 (1):27-40. doi: 10.1093/forestry/cpr051.

Vonderach, C.; Voegtle, T.; Adler, P. Voxel-base approach for estimating urban tree volume from terrestrial laser scanning data. Int. Arch. Photogr. Remote Sens. Spat. Inf. Sci 2012, XXXIX-B8, 451–456.

Wang, Yunsheng, Juha Hyyppä, Xinlian Liang, Harri Kaartinen, Xiaowei Yu, Eva Lindberg, Johan Holmgren, Yuchu Qin, Clément Mallet, Antonio Ferraz, Hossein Torabzadeh, Felix Morsdorf, Lingli Zhu, Jingbin Liu, and Petteri Alho. 2016. "International Benchmarking of the Individual Tree Detection Methods for Modeling 3D Canopy Structure for Silviculture and Forest Ecology Using Airborne Laser Scanning." IEEE Transactions on Geoscience and Remote Sensing 54 (9):5011-5027.

Watt, P.J.; Donoghue, D.N.M. Measuring forest structure with terrestrial laser scanning. Int. J. Remote Sens. 2005, 26, 1437–1446.

Wezyk, P.; Koziol, K.; Glista, M.; Pierzchalski, M. Terrestrial laser scanning versus traditional forest inventory first results from the polish forests. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2007, 3, 424–429.

9. METHODS FOR ANALYSIS OF REMOTE SENSING DATA

Change detection. Method for detection changes by comparing remote sensing data from more than one timepoint.

Classification. Analysis methods for assigning the values of remote sensing data into specific categories or classes.

Estimation. Analysis methods used for obtaining continuous variables.

Supervised Classification. An analysis approach used when training data are available, and used to develop the model used for classification or prediction.

Creating new information from remotely sensed data is performed by using the values of the remote sensing data in a statistical model which then interprets the data into information relevant for the user. The choice of analysis method often depends on the desired outcome (thematic or continuous values) and the characteristics of the input data. The different models or algorithms involved in the analysis method have their own characteristics, with associated advantages and drawbacks. The method of analysis chosen is often a subject for research in remote sensing.

9.1. Classification and estimation

We can categorize the approaches used to produce new information from remote sensing data in a number of ways, but perhaps the most general way is to categorize them by whether the end product will be thematic (classes) or continuous values. Classification algorithms are used to produce thematic classes, while estimation methods are used to produce continuous values.

The remote sensing data values are interpreted and translated into other meaningful information (e.g., thematic classes or values of a given forest parameter) most often by having reference data and building a model, or by categorizing the remote sensing data based on its characteristics. This will be discussed later in the section on "*supervised*" vs. "*unsupervised*" classification.

Classification refers to the assigning of any group of things into their own categories, based on the data characteristics. In remote sensing, it refers to assigning the values of remote sensing data into specific categories or classes. It is an exercise in pattern recognition, and a number of analysis methods used to categorize remote sensing data come from the field of pattern recognition.

There are several types of classification methods, which use a statistical basis for class assignment of the pixels. Some of these will be discussed in the following sections of this chapter. There is no single correct classification method to use. However, some methods may be better than others depending on the goal of your project and the available data. Methods are continually being developed and tested.

Estimation methods are used for obtaining continuous variables, and this approach is commonly used for forest mapping (for example stem volume, age, or species proportions). Estimation of continuous forest variables is a particularly common method with laser data, radar data, and 3D aerial photo models. Examples of estimation methods are kNN, regression analysis, regression trees, or random forests. A more strict definition is that *estimation* is used for statistical methods that establish a relationship between variables, e.g., an estimation of a regression functions that gives the relation between a dependent variable (e.g. stem volume) and on or several independent variables (e.g. spectral bands in a satellite image and/or metrics from 3D point clouds). The process of applying that estimated function on all forest pixels is then called *prediction* in some texts, but *estimation* in other texts. Some of the methods used for establishing relationships between remote sensing data and forest variables, for example Neural Networks comes from other science fields than statistics, for example computer science; still, we often use the word estimation for all these techniques.

9.2. Reference data

Reference data can be taken from field visits, aerial photo-interpretation, or ground-based inventories, for example. The term “ground truth” is sometimes used to refer to reference data. Reference data have assigned or measured variables describing the properties of the sample plot or an area (e.g., land cover class, percent shrub cover, or basal area) and often have associated geographic coordinates. The common geographic locations allow associating the remote sensing data with the variables from the reference data (see Figure 9.1).

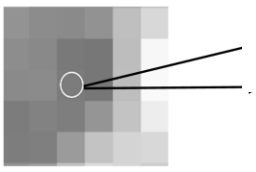


Figure 9.1. Example of a 10 m radius NFI plot corresponding to the same geographic location in Landsat-7 ETM+ data (25 m pixels).

9.3. Spectral classes, information classes, scatterplots and spectral curves

When classifying spectral data, there are some terms to be familiar with. These are *spectral* classes and *information* classes. The classes in your

classification scheme that you want to identify are called *information classes.* A single information class may be represented one or many *spectral classes.* This is demonstrated in a *scatter plot* below in Figure 9.2.

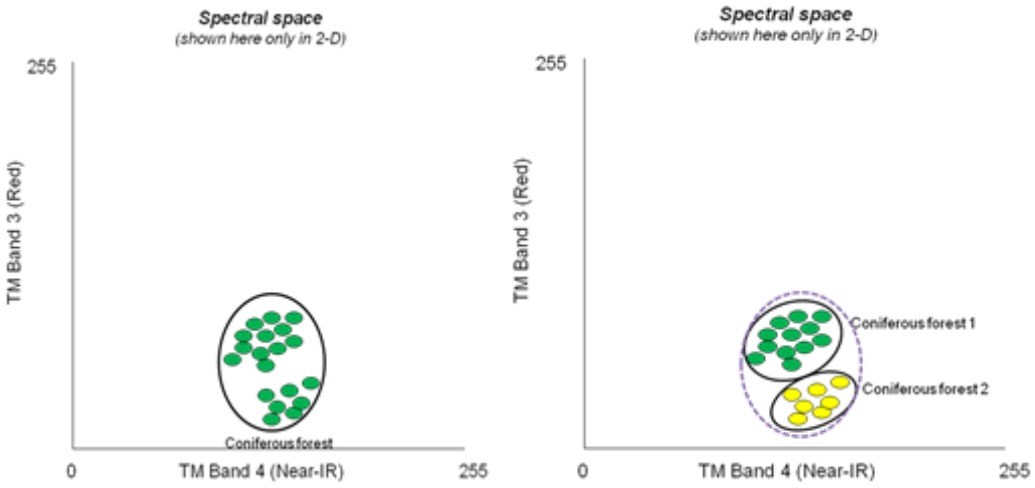


Figure 9.2. On the left the information class "Coniferous forest" is shown although you can see two distinct spectral groups, which are shown on the right as the two spectral classes that belong to the "Coniferous forest" information class. This might be due to training date collected from for example different slope directions.

Another useful way to look at spectral data is with *spectral curves*, meaning that you plot values for each wavelength band for different categories. It may be a mean value (and good to include standard deviation) for a vegetation type, or it may be many plots belonging to one vegetation type. An example is shown in Figure 9.3.

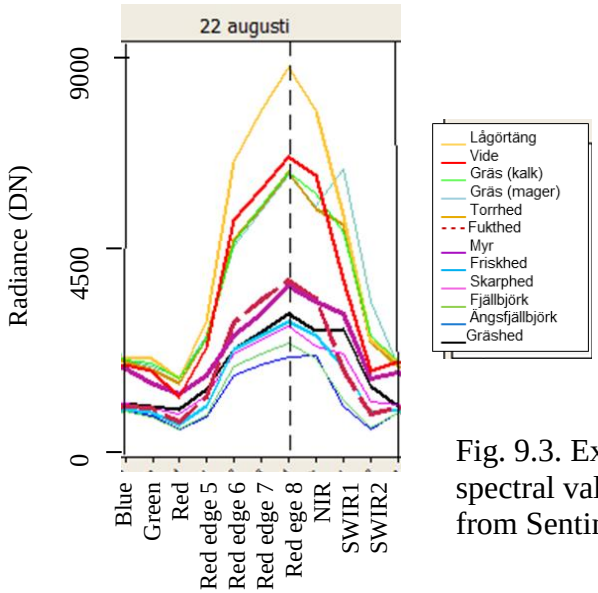


Fig. 9.3. Example of a spectral curve with the mean spectral value for different alpine vegetation types, from Sentinel-2 data (2015-Aug-22).

9.4. Classification methods

Classification has previously been applied mainly with spectral data from satellite sensors. However, it's important to realize that any data can be divided (with more or less success) into thematic classes.

When using spectral data, we can say that there are two basic categories of classification methods:

- spectral pattern recognition
- spatial pattern recognition

The main type of classification method used is based on spectral pattern recognition principles, meaning that the pixelwise colours of the multi-spectral images are used to separate different vegetation types. In this chapter, we will primarily discuss spectral pattern recognition methods.

There is also *spatial pattern recognition* where the spatial relationship between image pixels are used. Examples of spatial measures are:

- image *texture* (important with high resolution imagery, with 2 m pixel size, or better)
- object *shape* (e.g., using segmentation)
- *context* in the landscape (e.g., a set of rules how to interpret distance to water is used).

We can further categorize classification methods into different groups, which in addition help us choose which method we want to apply. The first main categorization is based on whether we have sufficient reference data to use in the analysis. This division is

- Supervised classification
- Unsupervised classification

Yet another categorization of classification methods is based on the statistical properties of the algorithm. This division is

- Parametric methods
- Non-parametric methods

Supervised classification is where reference data are available and used directly to identify the spectral values in the satellite image belonging to that vegetation type. The reference data are used to identify a subset of the pixels in the image to build a model for the classification algorithm. The model is then applied to the entire satellite image. These reference data are referred to as "training data", and a good training data set should assign class labels to the range of spectral values present within the satellite image. The result is a classified thematic map.

Unsupervised classification methods first cluster the remote sensing data automatically into a decided number of classes. The classification is made based on the statistics of the image, without use of reference data. Reference data are then used afterwards to assign a class label to each cluster. Unsupervised methods are often used when reference data are sparse or inadequate.

9.6.1. Unsupervised classification

Unsupervised classification can be divided into the following steps:

- Clustering (analysis)
- Assigning the clusters to information classes

Many different algorithms for this exist, for example, "K-means" clustering or the ISODATA algorithm. The following describes a typical process for unsupervised classification:

1) the user selects a number of classes or "clusters" based on the goals of the project and also the spectral properties of the image. For example, maybe only four clusters are needed if the information classes "water" and "other" are wanted (because water is so spectrally unique and easy to identify from most other land cover). However, if you want to have several information classes, you will most likely need to give a higher number of clusters than your number of information classes.
2) the computer first divides the multi-spectral satellite data into four clusters based on the mean values
3) each pixel in the image will belong to one of the clusters to which it has the spectrally closest class mean
4) new means are calculated for the classes according to the pixel values assigned to them
5) Procedures 3 and 4 are repeated until stable means are found.
6) Once the clustering algorithm has finished, the spectral classes in the output must then be manually assigned to the relevant information classes.

The following situations may occur when assigning an information class to a cluster:

- One spectral class will clearly belong to one information class (this is the easy case)
- Several spectral classes will belong to one information class (the spectral classes should then be labeled with the single information class)
- One spectral class will belong to several information classes (this means there will be problems obtaining an accurate map product based on spectral data only).

9.6.1. Supervised classification

Supervised classification is divided into the following steps, in this order:

- Training data set creation
- Classification (or analysis).

The training data creation stage is where a data set which associates real world vegetation characteristics with the remote sensing data, and the data set is then used to build a model. Individual training areas are identified and delineated (these can be single pixels or they can be polygons which enclose several to many pixels) in the image. Each one of these training areas is individually referred to as a *training data sample*, and a group of training samples is called a *training data set*. In a training data set, it is important that all spectral classes are included. In general, for each information class you will have several training data samples. The training data set can consist of all the training data samples for all the information classes in the image.

Training data samples can be obtained either automatically or manually. Automatically derived training data may be made when existing inventory data are used, where the remote sensing data are extracted from that coordinate. A manual way to obtain training data is by drawing a single point or polygon around a single pixel or group of spectrally similar pixels (Fig 9.3).

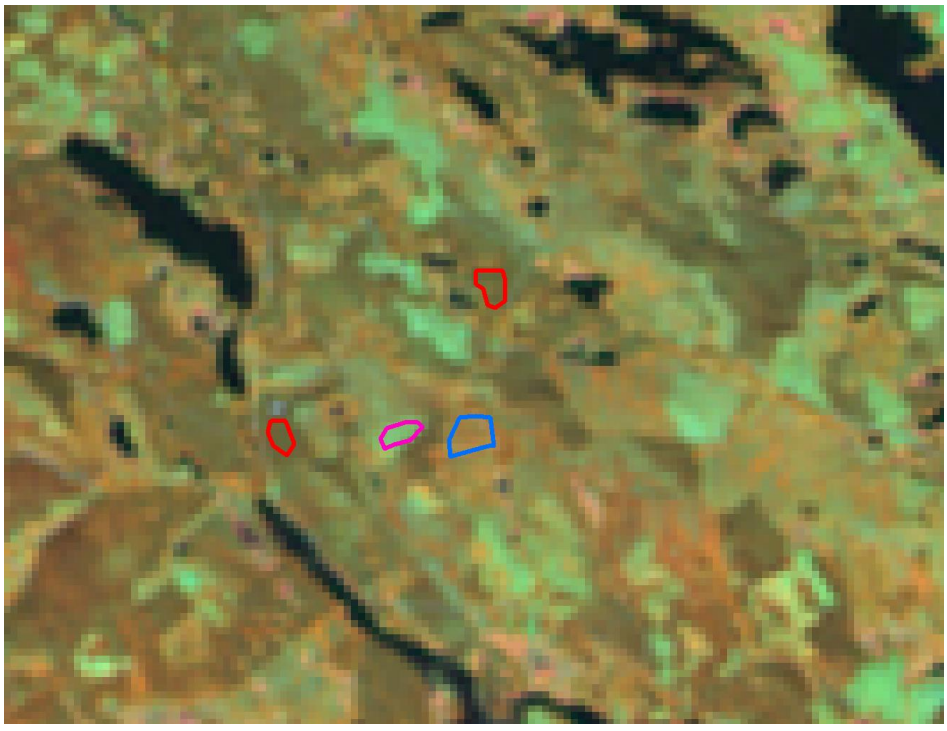


Figure 9.3. Four training data samples, where polygons are used to identify a group of spectrally similar pixels.

The aim of a training data set is to accurately represent the range of values in the remote sensing data that are associated with each vegetation type. For

example, several training data samples for a particular vegetation class will have basic statistical properties such as a mean value and standard deviation. This describes the range of values that are typical for this class. The aim is to have distinct values for the different classes, and for that reason, training data samples should be spectrally similar and not have a very large standard deviation.

Training data can be selected *subjectively* or *objectively*. Subjectively selected training data is done by purposely identifying optimal sites for developing a training data set. In other words, it is not based on a probability-sample and has been chosen subjectively by the analyst. This approach can lead to a well-behaved training data set, and (often) a better map product, and is allowed for training data. A disadvantage of this approach is that the analyst may be biased about which areas they choose.

In forestry, there is more of a tradition to carry out *objective* ground sampling. Therefore we mention that another source of training data can be, for example, NFI plots with GPS coordinates or other pre-existing inventory data. An advantage of this approach is that if the reference data plots are located in a systematic or random manner, proportions of classes can be estimated from proportions of plots. A disadvantage is that the sample may consist of many mixed (spectrally non-homogenous) pixels (Fig 9.3).

One cause of many misclassifications or difficulties in classification is the *mixed pixel (*sometimes called a *"mixel")*. Mixed pixels occur when there is more than one land cover type within a single pixel. This can occur especially when the pixel size is large, when the landscape is made up of small vegetation units, but in general, it can occur with any image just because it is impossible for a pixel to always fall evenly on clear boundaries of vegetation, nature doesn't work that way. The problems occur because mixed pixels can have spectral signatures that imitate the spectral signature of another class, or else have their own unique signature. For example, look along the borders between stands in an image. This is a typical area for mixed pixels, and a typical problem area. Consider what happens when a pixel covers an area of half water and half forest.

Although a subjective approach is acceptable for training data creation, it is worth repeating that accuracy assessment should be always be with an objectively collected dataset to evaluate the result.

The classification or analysis stage where a class membership is assigned to each pixel. There are many different supervised classification algorithms which can be used, for example, *Minimum distance to means Maximum Likelihood Classification, Decision Trees, Random Forest or Neural networks.* These algorithms will be discussed more in detail later in this chapter. Often the classification process is an iterative one, where the training data and the remote sensing input data are adjusted, as well as the parameters of the classification method, until the result is satisfactory.

Post-classification smoothing is a post-processing method which aims to create a "smoother" looking map. The result from supervised classification may require smoothing because it often looks more "pixelly" or has a "salt-and-pepper" look than does the outcome from unsupervised classification. This technique is described further in the section on manipulation of images.

9.3.1.2. Parametric and non-parametric classification methods

There are many different supervised classification methods, which can be separated into two groups: parametric methods and non-parametric methods. Parametric methods can be used if the training data meets the assumption that they have a multi-variate (i.e., all bands) normal distribution (Fig 9.4). With non-parametric methods the above assumption is not necessary to meet.

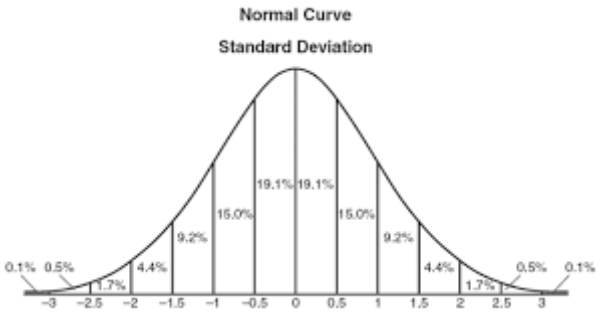


Fig 9.4. A normal distribution, also called Gaussian distribution.

Parametric methods are useful which the characteristics of the data are less complex. Typical methods include linear discriminant analysis (LDA), and maximum likelihood classification. Non-parametric methods include decision and regression trees, random forest, kNN, support vector machines, and neural networks. The previously named non-parametric classifiers all fall under a category called "machine-learning" algorithms, in which the data are iteratively processed until the accuracy of the outcome is optimized.

While non-parametric methods do not require having training data that follow a normal distribution, the results can be affected by having a training data set which is too small, or has too many of a single class (an "imbalanced data set"). In general, non-parametric methods require a larger number of training data sets.

9.4. More on training data for supervised classification

There are a number of characteristics of the training data used for supervised classification that need to be considered. These include

- number of training data samples needed,
- size of the training sample in relationship to the phenomena in the landscape as well as the remote sensing data spatial properties,

- statistical properties of the sampling scheme,
- interaction between the classification method and characteristics of the training data set, and
- quality of the training data.

9.4.1. Number of training samples

A sufficient number of training samples and their representativeness are critical for image classifications. However, precise determination of the number of training data samples needed to achieve an accurate classification is elusive. There is a complex interaction of different factors that influence how many training samples might be needed. The (spectral) variation within a defined class is one such factor, as well as the algorithm chosen for the analysis.

Determining the number of training data samples needed requires investigating the data set. One way to approach this is with statistical measures. The number of samples necessary to obtain adequate representation of the spectral variability present in a class can be determined statistically by sampling. Traditional distance measures of separation between classes are also often used (e.g., Jeffries-Matusita distance or Transformed Divergence), but in reality, these may be not be accurate predictors of expected classification accuracy with a certain number of training samples, as they don't provide information about the adequacy of spectral representation of an individual class.

It should be kept in mind that variations in field layer can also influence the spectral variability of a certain class. For example, there will be large spectral variations in the class "Pine Forest" if Pine forest on exposed bedrock, Pine forest with ground lichens, and Pine forest with shrubs are all contained within the Pine Forest class.

Not only the representation of the spectral variability within a class needs to be taken into consideration, but also the spectral similarity or dissimilarity to other classes. As an example, water has a spectral signature so distinct from many vegetation types, that a full description of the spectral variability of water may not be necessary for accurate classification. Two very spectrally similar classes, such as pine and spruce may require careful and full assignment of their spectral characteristics if one hopes to accurately classify these spectrally similar classes.

If one wants to use an approximate rule of thumb, a loose recommendation from the authors is to use no less than 10 training samples for a class, and preferably up to 30 training samples.[1],[2],[3]

9.4.2. Interaction between the analysis algorithm and the training data

The characteristics of the supervised classification method exert an influence on the requirements from the training data set. As an example, statistical classifiers using the mean vector for class assignment will be influenced less by outliers in the training data, while classifiers such as Neural Networks can be highly influenced by individual poor quality training data samples. Some classifiers, such as maximum likelihood, require a minimum of p+1 (p = number of input variables) training samples per class to build statistics (e.g., covariance matrices). Based on this, Mather[2] recommended that 10 to 30 times training samples per class should be used. Non-parametric methods don't face this restriction, but are still affected by the total number and frequency of classes in the training data. In particular, the output from Random Forests algorithms can be influenced by "imbalanced" training data[4].

In the estimation of continuous values, a larger training data set is generally required. Mathys *et al.*[5] found that for estimation of continuous parameters, a training data set consisting of the whole range (0-100%) of the parameter was necessary. In fact, training samples containing mixes or 0% of the parameter were more important than training samples with 100% representation of the parameter.

9.4.3. Dimensions of the training data sample

The dimensions (i.e., size) of the training data sample plot is also important. The plot may be circular or square, and have particular dimensions (i.e., radius or square side length). Some size plots are better suited to the remote sensing data to be used (e.g., circular radius plots that are as large or just slightly larger than the pixel size). If plots are larger than the pixel size, the data within the pixels can be used (e.g., mean value). It is also important to use plots that are of the same size in the training data, or at least take consideration to this.

---

[1] Reese, H. 2011. Classification of Sweden's Forest and Alpine Vegetation Using Optical Satellite and Inventory Data. Doctoral thesis no. 2011:86, Swedish University of Agricultural Sciences.

[2] Mather, P., 2004. Computer Processing of Remotely Sensing Images. 3rd ed. John Wiley & Sons, Inc.

[3] Foody, G., et al., 2006. Training set size requirements for the classification of a specific class. Remote Sensing of Environment. 104:1-14.

[4] Reese, H., et al., 2014. Combining airborne laser scanning data and optical satellite data for classification of alpine vegetation. Intl J of Applied Earth Observation and Geoinformation. 27:81-90.

[5] Mathys L. et al., 2009. Evaluating effects of spectral training data distribution on continuous field mapping performance. ISPRS J of Photogrammetry and Remote Sensing. 64(6):665-673.

Gong and Howarth[6] suggested that training data were best selected using single pixels and a systematic sample, however Chen and Stow[7] said this was more suitable for homogeneous land cover types. Chen and Stow tested training samples taken on single pixel level and in blocks of pixels, finding that training data set size mattered more when single pixel training samples were used, and that blocks of pixels produced higher accuracies for training in heterogeneous landscapes. However, their result may have been dependent on the urban land use classes particular to their study. When the landscape of a study area is complex and heterogeneous, selecting sufficient training samples becomes difficult[8].

9.4.4. Statistical properties of the sampling scheme

The frequency of the classes as represented in the training data has an influence on the results. In large area projects, it is desirable to have *a priori* information on the frequency of classes in the area[9], as input to the supervised classification. For this reason, prior probabilities are assigned in Bayesian classifiers. Additionally, weights can be added in Decision and Regression Tree models to counteract imbalanced data[10]. Rare classes are often a desired class in map products, and sufficient training data may need to be collected specifically for this purpose.

9.4.5. Quality of the training data

Quality control and refining training data in order to obtain accurate classifications is important to do. Some researchers suggest reducing the effects of outliers in the training data by weighting training samples according to their quality or by subjecting training data to majority-vote classifiers to detect mislabeled data. In the ideal case, training data should be evaluated before used in the classification/estimation process. To evaluate training data, the following can be done:

- Looking at scatter plots of training data
- Using spectral distance measurements between training data sets (e.g., Transformed divergence, Jeffries-Matusita)
- Investigating the classification result and the effect of inclusion/exclusion of potentially problematic training data samples
- Usually a combination of all of these is done

---

[6] Gong P. and Howarth, P.J. 1990. An assessment of some factors influencing multi-spectral land cover classification. PE&RS56(5):597-603.

[7] Chen, D.M. and Stow, D. 2002. The effect of training strategies on supervised classification at different spatial resolutions. PE&RS 68(11):1155-1161.

[8] Lu ,D. and Weng, Q. 2007. A survey of image classification methods and techniques for improving classification erformance. Intl J of Rem Sens 28(5):823-870.

[9] Cihlar, J., et al, 2000. Selecting representative high resolution sample images for land cover studies. Rem Sens of Env. 72(2):127-138.

[10] Xie et al., 2009. Customer churn prediction using improved balanced random forests. Expert Systems with Applications 36(3):5445-5449.

- To improve the result, you may need to add and delete training data samples…

The training data set almost always needs be evaluated and changed to achieve optimal classification accuracy. This process can take a considerable amount of time!

There are also temporal aspects to recognize when using training data. Reference data may be collected from dates differing from the satellite imagery and this may cause erroneous class assignment. The timing of the image acquisition and the vegetation phenology must be considered in relation to the training data. In the case of the alpine landscape, the natural seasonal dynamics and change in moisture conditions that can occur within the growing season as well as from one year to another in the alpine region pose challenges to using training data and satellite data from different time points. Within managed forest landscapes, silvicultural activities such as thinning and clear-cutting need to be identified.

9.5. Algorithms for Supervised classification

9.5.1. Maximum Likelihood Classification (or Discriminant Analysis)

In remote sensing, practitioners often refer to the "Maximum Likelihood" (ML) supervised classification method[11] (Lillesand *et al.*, 2008). Maximum likelihood classification was often used for land cover mapping projects in the 1980's and 90's, and is still used today. In essence, when remote sensing practitioners refer to "maximum likelihood classification", they are most often using what statisticians refer to as Bayesian quadratic discriminant analysis (QDA). Bayesian QDA uses statistics about the prior probability of a class occurring, based on the frequency of that class in the training data. We go to the length of explaining this because, it would actually be more correct to call "Maximum Likelihood Classification" by the same term used by statisticians (Bayesian QDA), however, this has not been commonly done in the remote sensing literature. The following should be read with that in mind.

Discriminant analysis (and our "Maximum Likelihood Classification" method) is a method by which one assigns class membership to an unknown observation based on a sample of data with known class memberships. The example here is that our known data comes from our training data, and the unknown data are all the other pixels in the image.

From the training data, a probability density function based upon the training data, as well as prior probability weights for each class are used to calculate the probability of the unknown pixels values "belonging" to the values in the sample data set. Then a posterior probability that an observation belongs to

[11] Lillesand, T., et al., 2008. Remote Sensing and Image Interpretation. John Wiley and Sons, Inc.

a certain class is calculated using the prior probabilities, the probability density function, and the probability of occurrence for that observation.

The remote sensing term “Maximum Likelihood Classification” is derived from the fact that a maximum likelihood estimator rule is used in Bayesian discriminant analysis, in which the observations are assigned to the most “likely” (i.e., highest probability) class, in order to maximize correct classification assignment. An illustrative example of how the Maximum Likelihood Classification algorithm works is shown in Figure 9.5.

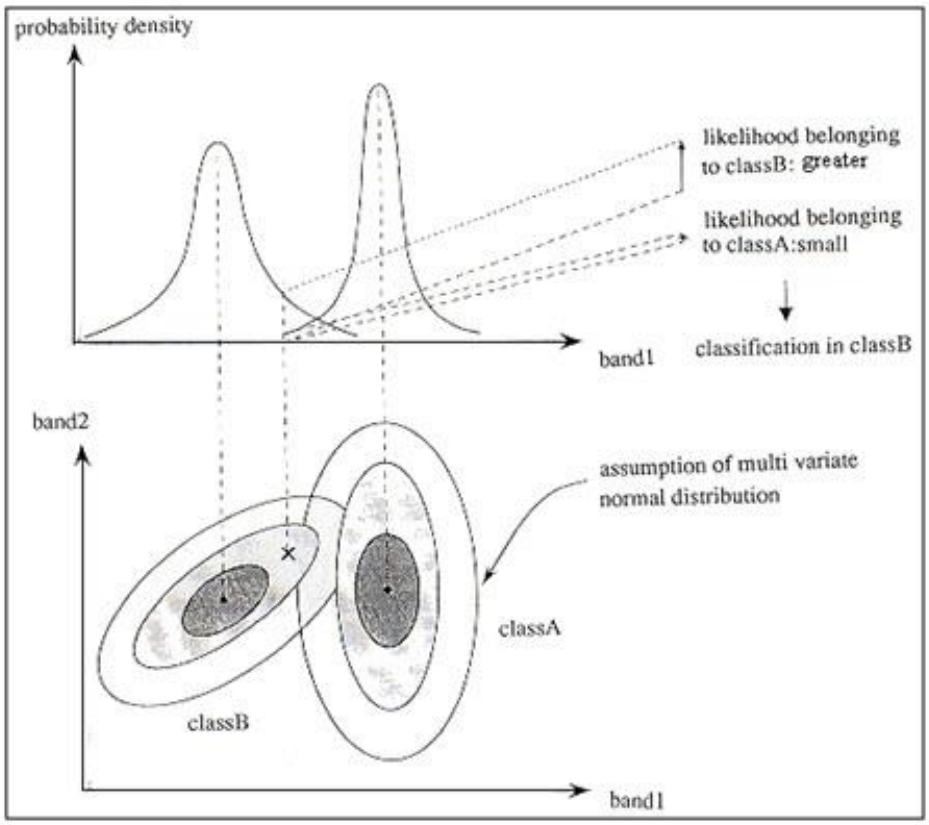


Figure 9.5 Maximum Likelihood algorithm.

Discriminant analysis works best following the assumption that the data for each class and variable are normally distributed (meaning that there is a Gaussian distribution of the spectral data values within the training data for a single land cover class (or whatever class you wish to map). This means that Discriminant analysis (i.e., our Maximum Likelihood classification) is a *parametric* method. In some cases, the covariance matrices of the different classes in the data may have the same distribution, and in this case, a linear discriminant analysis (LDA) can be used. In cases where covariance matrices are not the same between classes, quadratic discriminant analysis should be used.

In practice, the application of Maximum Likelihood classification in land cover mapping has often left out the use of prior probabilities of class occurrence due to lack of sufficient information about the frequency of class

occurrence. When "maximum likelihood" classification without prior probabilities is used, then this is actually "Bayesian discriminant analysis with equal prior probabilities for all classes". If equal probability is assumed and no weights are used in the training data, the result may be that more frequently occurring classes in the training data will be under-classified (errors of omission) in the resulting map and less frequently occurring classes will be over-classified in the map (errors of commission). Several remote sensing studies have pointed out the utility of including prior probabilities within the "maximum likelihood" classifier, finding that it improved land cover classification accuracy, particularly for spectrally similar classes.

The expression used for Maximum Likelihood classification is often based on Richards (1999), shown in Eq. 9.1 below.

$$g_i(x) = \ln p(\omega_i) - \tfrac{1}{2} \ln |\Sigma_i| - \tfrac{1}{2}(x - m_i)^T \Sigma_i^{-1} (x - m_i) \qquad \text{Eq. 9.1}$$

where:

i = class

x = n-dimensional data (where n is the number of bands)

$p(\omega_i)$ = probability that class ω_i occurs in the image and is assumed the same for all classes

$|\Sigma_i|$ = determinant of the covariance matrix of the data in class ω_i

Σ_i^{-1} = its inverse matrix

m_i = mean vector

In practice, with Maximum Likelihood, two approaches with the training data can be used (as demonstrated in Fig 9.6).

- separate training samples can be merged into a single training data statistic for a single information class, or
- individual training samples can be used (i.e., spectral classes) and thereafter be recorded in the end product belonging to an information class (e.g., Coniferous 1, Coniferous 2, etc…).

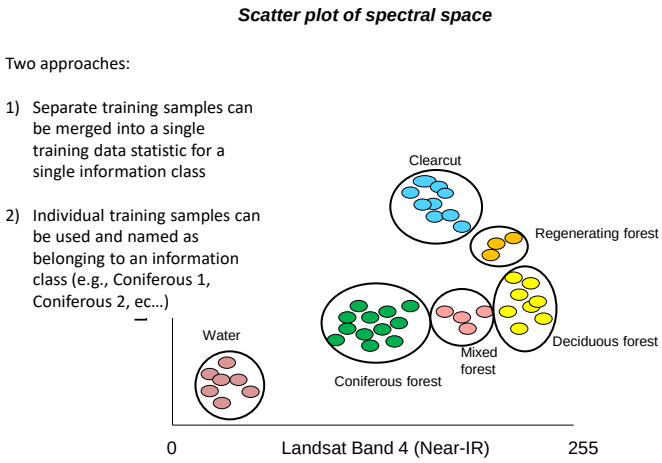


Figure 9.6. Example of a scatter plot of training data and the thematic classes. Red band is on the Y axis and NIR band is on the X axis.

An advantage (or potential disadvantage, depending on the data) of the latter approach is that the standard deviation of the training data will be smaller, and reduce the risk of including too many pixels in the satellite data (potentially not belonging to that class), resulting in misclassification. A disadvantage would be if the training data samples are too "narrow" having too low a standard deviation, to include the appropriate pixels. Your choice depends on the properties of your data.

Figure 9.6 is an ideal picture of a training set, and rarely are the classes so well separated. Figure 9.7 shows a training data set closer to reality.

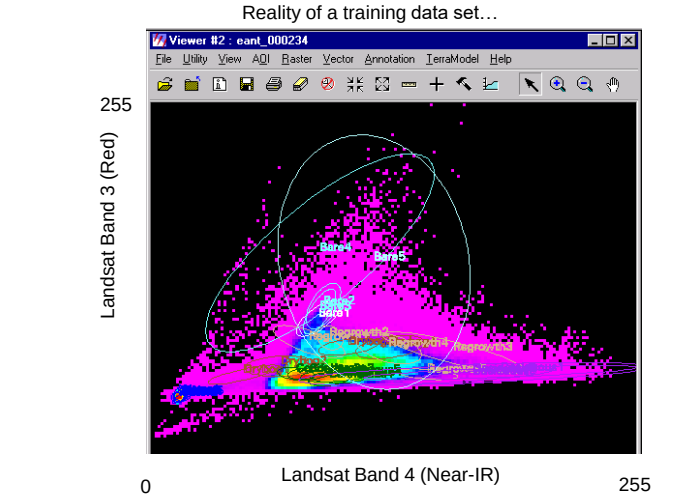


Fig 9.7 A feature space plot showing all values within a Landsat satellite image (in color), with Red band on Y axis, the NIR band on the X axis, and the training data shown as ellipsis with a mean value and the extent of the ellipse being 2 σ.

9.5.2. CART and Random Forests Classification

Classification and Regression Tree (CART) methods have been increasing in use for land cover classification over the past decades. Three methods being widely used are decision trees, regression trees and random forests. Decision trees produce a categorical output, regression trees produce continuous variables, and the random forests algorithm is capable of producing both. These non-parametric methods have an important advantage over maximum likelihood classification in that data from different sources (e.g., spectral data, elevation derivatives, map data) can be combined, without the need for normal distribution of the data. These methods are also particularly useful when a large number of input variables are to be used.

9.5.2.1. Decision (or Classification) and regression trees

A decision or regression tree is a simple format, on which random forests are based; although they are no longer widely used, it may be useful to understand the concept. A hierarchical tree is constructed (e.g., Fig 9.8) from the training data, and consists of what are called root nodes, interior nodes and terminal or leaf nodes[12]. Data splitting rules are constructed at each (non-terminal) node based on the training data's spectral and class values. Splitting rules depend on the particular decision tree, but are most often determined by achieving maximum information gain and the lowest Gini impurity index[13] based on the input variables at each node. In decision trees one variable is normally used for splitting at each node, although multivariate decision trees have been developed. Regression trees also consist of the same node system, however the difference is that regression functions are built to estimate continuous values.

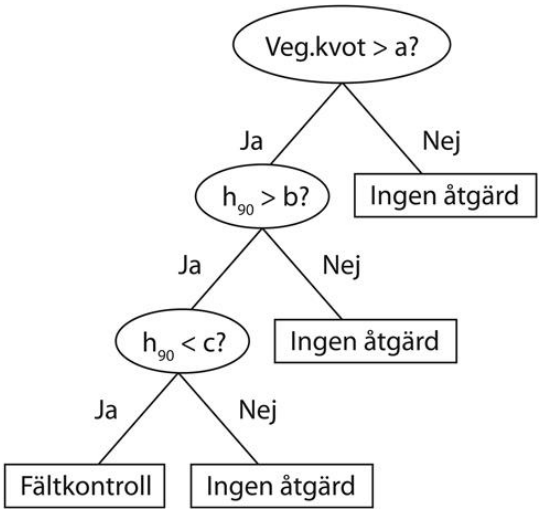


Figure 9.8. Example of a simple classification tree that can be used to identify areas of dense forest (Veg.kvot > a) within a certain height range (b <H90 <c).

[12] Tso & Mather, 2009.
[13] Breiman *et al.*, 1984

Refer back to Figure 9.8. Say that we are interested in finding forest that requires thinning. We believe that the stem/ha count is high and is within a certain height range, and we want to define such areas in a GIS to plan a control in the field. We create the classification tree and use it in conjunction with vegetation ratio, V, and the 90th height percentile, H90, in the form of breaks. Each question, or node, in the tree in a variable and can be answered with "yes" or "no". The tree flows into a number of end nodes, each of which corresponds to a class. In the example in the Figure, there are two classes: "no action" and "field check". A class can be found in multiple end nodes.

In the regression case, the data set is not divided into categories as illustrated in Fig 9.8, but instead constructed into trees so that each leaf contains a small, relatively homogeneous group of observations. The end node is a number (i.e., a continuous value) rather than a class. Training data, such as field inventory plots, are used to build the tree. The final result is calculated as the average number of the results from all the trees. The model obtained is then applied to all the “unknown” pixels in the image.

A tree is constructed by taking a so-called bootstrap sample of plots (it extracts a certain percentage of training samples which are used to internally quality check the model), by drawing with replacement. To build the first node it uses a user defined number of independent variables to test. The best variable that divides the sample clearly into two separate groups is selected. After that, each branch going out from the first node, leads to a new node which is constructed in the same manner as the first. The tree is considered complete when the number of sample plots in each leaf reaches a user-specified value.

9.5.2.2. Random Forests

Random Forests[14] is a non-parametric method that can be used for both estimation and classification. Note that the term random “forests” does not have to do specifically with the forest, but refers instead to a method that uses a large number of classification (or regression) trees. A model is constructed by inputting the training data, and giving a number of parameters, such as the number of trees (i.e., iterations) to be used. A final model is achieved automatically by using the best result from all of the trees. Each individual classifiction/regression tree produces either an estimation or classification result, and the final result is calculated as the mean of the continuous values (in the case of estimation) or by "majority voting" (the classification).

The advantages of using the Random Forests algorithm are that one can use a large number of independent variables, including those that are correlated with each other, as well as different types of data (spectral data and map data) with different properties. Random Forests have effectively shown to give almost as good results as regression, without having to manually build a model of the relationship between independent and dependent data. Use of the random forests classifier has also produced classification results that are

[14] Breiman, L., 2001. Random Forests.

equally accurate or more accurate than other classification methods, and it is relatively robust to outliers and noise[14]. It may still be advantageous, and produce better results, however, to perform variable selection. Variable selection is often done by using a combination of statistical methods and common sense (e.g., variable(s) known to be relevant should be kept in the model, such as the SWIR band for forest volume estimation, or the NIR band for vegetation type identification).

There is a final assessment of the model quality in both Random Forests and CART models which is given by the so-called OOB Error Estimate (OOB stands for "out-of-bag"). It is estimated by a given percentage of the sample plots that are not included in the training data that were used to create the Random Forest model. Random forests uses bagging[15] as well as a random selection of the variables to consider at each node, therefore pruning of trees in random forests is not required.

Probability of class membership is based on the frequency of classes in the training data. Therefore, all the tree classification methods discussed here are subject to misclassifications due to imbalanced data (i.e., having an uneven distribution of training samples among the classes), and for this reason, weights and other modifications to the training data sets are sometimes made.

9.5.2.3. Support Vector Machines and Relevance Vector Machines

Both Support Vector Machines (SVM)[16] and Relevance Vector Machines (RVM) are relatively new methods, which are based on so-called "kernel" methods. This means that they transform the original values into new values (go from the original feature space to a "kernel feature space" which has higher dimensionality and allows for a linear model to be applied). Support Vector Machines rely on a well-defined boundary (called the "hyperplane") between two or more spectral classes, which is determined iteratively. Relevance Vector Machines are a Bayesian extension of SVMs (include prior probability measures). An advantage of SVM and RVM is that it does not require a large number of training data samples, but can work with many bands of input remote sensing data. A figure demonstrating SVM is shown in Figure 9.9.

[15] Breiman, 1996

[16] Mountrakis, G., et al., 2011. Support vector machines in remote sensing: a review. ISPRS J. Photogramm. Remote Sens. 66, 247–259.

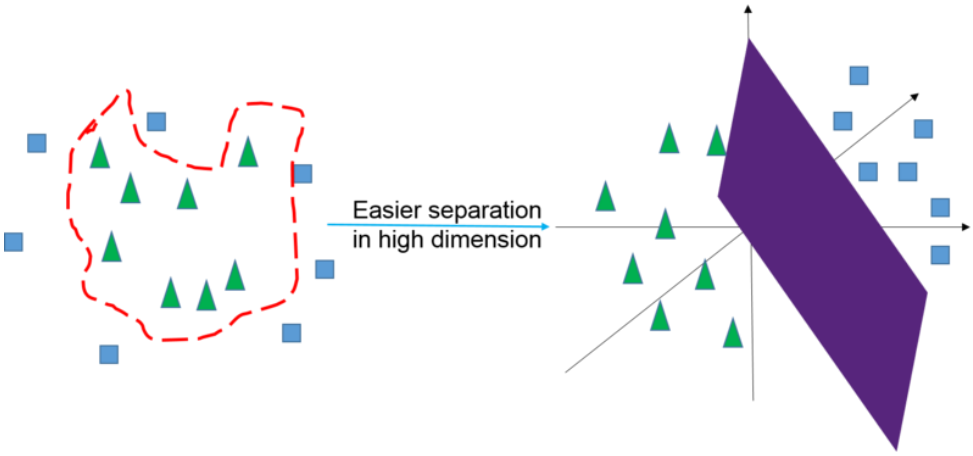


Fig 9.9 Illustration of Support Vector Machine algorithm with input data (left) and higher dimensional data using support vectors (right). Image from Wikimedia Commons.

9.5.2.4. Object-oriented image analysis (OBIA)

Classification can be carried out at the pixel level or at an aggregated level, such as within segments, which is referred to as an object-oriented image analysis, or OBIA or GEOBIA (i.e., Geo-OBIA) for short. This is often done when there are many pixels per object due to having remote sensing data with a small pixel size. An example would be when you have many pixels per tree crown, and using an automatic segmentation, a segment of the tree crown can be made. The spectral values within the segment can be analyzed (mean, standard deviation), and used to identify tree species or other vegetation classes.

9.6. Estimation algorithms

9.6.1. Regression

Regression is a statistical method for finding a relationship between two or more random variables. The method is model-based, i.e., it assumes that there is a special mathematical model that relates the two variables together. Although the method is based on a lot of theory, the basic idea is simple. Working with regression analysis is something of an art, but you can get far with fairly simple models. The following is a somewhat simplified description of regression.

Figure 9.10 shows a scatter plot of field measured basal area average height, H, on the y-axis and the 90th height percentile, H90, on the x-axis.

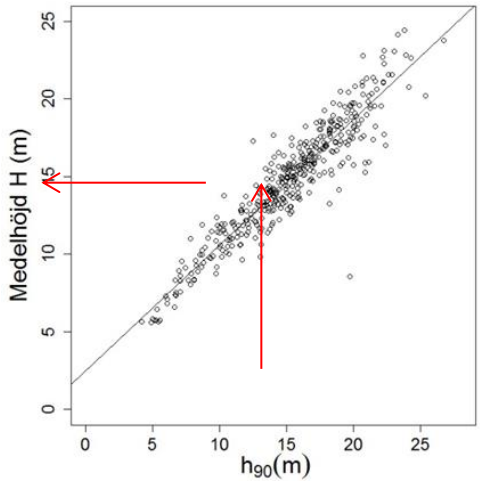


Figure 9.10. Scatter plots with 90th height percentile from laser data (X-axis) plotted against basal area weighted mean height (Y-axis). The arrows show how the regression function (solid black line) will result in use for elevation estimation.

Each point represents a sample area where you have both field measurements and laser data. There tends to be a strong relationship between forest height and H90, which is evident in the scatter plot above, where you can see the points are well grouped along a line. By investigating the relationship between average height and H90, you can then estimate H in areas where field data is missing. Note that H90 is not the same as the average height of the forest, but only correlated with this! Correlation is a measure of the strength of a linear relationship between two variables, and the correlation coefficient can have values between -1 and 1. If y increases with x and the points lie exactly on a line, the correlation is 1, which means a perfect, positive linear relationship. If instead y decreases with x, and points lie exactly on a line, the connection is perfect negative correlation is -1. The closer the correlation is to 0, the weaker the connection. In the data sets used in the Figure above, the correlation between H and H90 = 0.93, which means a very strong, positive relationship.

To begin with, you have to find the model that best fits the observed data. In the case of H and H90, it appears to be a straight line (Eq 9.2).

A Straight-line equation is written as

$$y = k \cdot x + m, \qquad \text{Eq 9.2}$$

where k is the slope of the line and m is the intercept where the line crosses the y-axis. The points on the graph are not exactly on a line, there is some random deviation or spread around the line. We write, therefore, the relationship between H and H90 as

$$H_i = \alpha + \beta \cdot h_{90,i} + e_i, \qquad \text{Eq. 9.3}$$

where the index i denotes a certain field plot location and e is the deviation from the line at that pair of observations from field and remote sensing data.

Using regression analysis, we can determine the values of α and β that minimize the sum of squares of the distances e, the so-called residuals, between the points and the line, under the constraint that the mean of all deviations is zero. Equation 9.3 describes H as a function of H90, and the variable H is called the dependent variable and the variable H90 is called the independent variable. H can also be called the response variable and H90 can be called the explanatory variable. The regression analysis also provides estimates of how strong the dependence is between the two variables. A common measure is the coefficient of determination. This quantity is often referred to as R^2 and indicates how much of the variance the regression model describes for the y-variable. The coefficient of determination is one of the tools usually used to select which variables should be included in the model. Another important measure of the model's accuracy is its' spread around the regression line.

When a reliable function has been made, it can then be used to estimate the average height of trees in areas where there are laser data but no field measurements have been taken.

The above example addresses the simple linear regression. In many cases, several independent variables are used to describe the subject, and this is then called multiple linear regression. You may also need to transform the independent variables by squaring it, multiplying the two independent by each other, and so on. Some relationships are non-linear and can be linearized by logarithmic transformation prior to doing the regression. This is not addressed here, and those who want to go deeper into the theory are referred to the statistical literature.

9.6.1.1. Regression in practice

Since the estimation results generally are aggregated into stand averages, it is important with a good stand delineation. In Norway, stand deleniation and species assessment is done with manual interpretation in digital photogrammetric workstations, then other forest data are estimated on the basis of laser data. Stand delineation can also be made on the basis of segmentation of laser data. Manual editing may be needed and aerial photos are of course also useful in this work. Stand boundaries are stored in vector format in a GIS. Stands and sample plots are sometimes divided into strata for which separate regression functions are estimated. The stratification might for example be based on tree species, site index and age. These characteristics affect crown shape, which in turn affects the relationships between laser measurements and forest variables. Estimates can get higher accuracy if done for more homogeneous groups, but there are also examples of experiments where stratification has not improved outcome.

When the field data are collected, as well as possible stratification and processing of laser data are done, it's possible to now do the regression

analysis. There are a variety of statistical software programs that can be used to perform calculations, but for the results to be good requires both statistical and expertise. A regression function should be developed for each forest variable to be estimated, possibly divided per stratum. A first step is to decide what type of model is to be constructed. Residuals of regression should be evenly distributed. If the residuals are seen to increase or decrease for the independent variable, the data may be linearized, for example, by using logarithms.

When selecting the independent variables, you can either start with one and gradually add new ones, or start with a lot of them and take away those which are not of importance. The former method is more manageable, then you should aim for a function with relatively few independent variables. One must also consider which laser measurements are reasonably associated with the variables to be estimated. For example, it is likely that wood volume is associated with both height and density measurements. Automatic methods such as stepwise regression and best subset regression, can provide guidance in the selection of variables. However, one should not rely blindly on the results because the methods are purely based on statistical relationships and do not necessarily say anything about the actual causation. The input laser measurements should not be too highly correlated with each other, such as two adjacent height percentiles. For each variable that is added or removed, a new, temporary, regression function is made. By analyzing the correlation between the residuals from the temporary regression function and each candidate for the new regression, it is determined which variable is to be added next. Table 9.2 shows examples of the regression functions which have been used in different studies to estimate different forest variables.

Table 9.2. Examples of regression functions for estimating the population level, from studies of Næsset et al. and Holmgren et al.

Study	Variable	Funktion[a]
Næsset	Mean height, H	$\ln(H) = 0,35 + 0,529 * \ln(h_{90f}) + 0,355 * \ln(h_{maxf})$
Holmgren	Mean height, H	$H = 1,46 + 0,95 * h_{95}$
Næsset	Stem volume, V	$\ln(V) = 3,151 + 3,027 * \ln(h_{80l}) – 1,66 * \ln(h_{maxf}) + 1,223 * \ln(d_{50f})$
Holmgren	Stem volume, V	$\ln(V) = -2,50 + 0,87 * \ln(D_v) + 1,49 * \ln(h_{90}) - 2,44 * relstd + 0,44 * D_p$
Næsset	Mean diameter, D_g	$\ln(d_g) = 0,406 + 0,892 * \ln(h_{90f}) – 0,374 * \ln(d_{1f})$
Næsset	Stem number, N	$\ln(N) = 10,33 - 0,487 * \ln(h_{0l}) – 0,667 * \ln(h_{cvf}) + 1,187 * \ln(d_{50f})$

a) Subskript f and l indicate that the metric is calculated for only the first (f) and last (l) returns. hmax is the maximum height of vegetation hits, d50f is the number of first returns from the treetops of H50 divided by the total number of returns. Dv is the number of first returns above 3 m divided by the total number

of returns. relstd is the standard deviation in height divided by H95. dp = (n1 + n3) / (n1 + n2), where n1 is the number one returns, n2 is the number of first returns, and n3 is the number of first returns where others return is over 3 m.

The complete regression functions (one for each estimated variable and strata) is applied to the grid of laser data for estimation of forest variables. The resulting screens in combination with stands limits in vector format used to calculate population averages. Results from some Swedish experiments with regression method are shown in Table 9.3.

So-called cross-validation is usually often used to evaluate the results. Cross validation will be more thoroughly addressed in Chapter 10 on Accuracy Assessment. But as an example "Leave-one-out-Cross-validation" works by leaving out one plot at a time (or sometimes many), estimates the functions parameters and uses these to estimate a value for the single plot or group of plots taken away. This is done for all plots and the estimation result is then compared with the plot field measurements.

Regression provides unbiased estimates, i.e., the value of the estimated variables on average are correct. Since the regression method is model-based and allows interpolation, regression works with relatively little field data. Sample plots must, however, be fairly representative - you should not, for example have too many plots that are unusually dense for their height.

Table 9.3. Results from some Swedish experiments with regression method, evaluated at the stand level and compensated for sampling error.

Estimated variable	Mean Error (%)
Volume/ha	6-14
Basal-area weighted mean diameter	7-13
Stem number/ha	12-24
Basal-area weighted mean height	3-6

9.6.2. k-MSN and k-NN

The k-MSN (k Most Similar Neighbors) method is based on imputation. The k-MSN, and other similar methods such as k-NN (k Nearest Neighbors), selects a number of plots which have similar spectral values as the raster cell to be estimated and calculates a weighted average of these. The *k* stands for the number of plots to use, and the value of k can vary depending on the purpose, but typical values are between 1 and 20. A high k gives better average estimation results, but may give less realistic relationships between the estimated variables in a given area.

The difference between kMSN and kNN is that kMSN uses canonical correlation analysis to select the neighbors. This part is a parametric method, which is based on the linear combination of the dependent variables (here

forest data) that are most correlated with a combination of independent data (eg laser measurement) is calculated. This gives the weights for the different laser metrics importance in imputation.

All dependent variables are estimated while allowing the natural relationships between the estimated variables retained better than the regression. The estimation of k-MSN is widely used for estimations using remote sensing data from spectral data as well as laser scanning.

It is important to have a large number of test surfaces evenly spread over the entire range of variation. The greater the variation in the forest, the more plots are needed. Extrapolation works poorly because a theoretical model is not calculated in this case, and only the actual data in the data set are used to assign values. To illustrate, consider how kMSN or kNN might work for estimating wood volume over a large area if the training data only has values up to 200 m^3/ha.

The k-MSN method requires significantly more field data, evenly spread over the entire range of variation in the data. Since the method is based on imputation to each grid cell of forest data from plots that have similar characteristics as the laser data, it is very important that the extremes of the data set are represented.

9.7. Combining estimation and classification

Note that output from continuous value estimation can also be used to assign a thematic class to a pixel. As a simple example, a raster output of percent forest cover can be converted to a two class thematic map of forest and non-forest by defining forest using a threshold in the percent of forest cover (e.g., > 10%). When continuous values are estimated with the goal of creating a thematic map, this is sometimes referred to as “soft” classification, whereas a thematic classification is referred to as a “hard” classification (Fernandes et al., 2004).

9.8. Selecting the remote sensing data variables

As you know by now, satellite images have several wavelength bands of data. Which bands do you want to use in the classification process? This is also a subject of research. The answer is to use a combination of statistical analysis and common sense to decide which bands to use. One should always investigate their data set, and plot the data. Are there certain wavelengths or variables from the remote sensing data that are highly correlated with the vegetation property you wish to map? Then include this band. Are there several bands of remote sensing data which are highly correlated? Then you may want to use only one of these. Correlation analysis, best subset regression, variable importance or step-wise addition of information are just some of the methods used to choose the remote sensing data which should be used in the analysis.

9.9. Change analysis

Change detection is one of the large operational applications of optical satellite remote sensing of forests. The Swedish Forest Agency for example acquires medium resolution satellite data (SPOT) annually for all of Sweden. The main reason for this is that the images are used for checking the location and year for new clear felled areas. The areas where changes are detected are then matched with the cutting permits.

In other countries, such as Estonia, change detection has also been used for checking how large proportions of the fellings that have been made without cutting permits. Many countries are also discussing the use of change detection techniques for the mandatory deforestation mapping to be made for Carbon reporting to the UNFCCC.

Change detection is also an important technique for damage mapping. In Sweden, it has for example been used for mapping the extent of Gremeniella damages and for mapping areas affected by storm damage (for example, after Gudrun in southern Sweden).

To date, most change detection studies have been performed with aerial photos, optical satellite data, and radar data. As the use of laser data is relatively new and few multitemporal data sets exist, change detection studies with laser data has not yet been extensively used operationally.

9.9.1. Prerequisites for change detection

The comparison of imagery from different time points is a powerful tool for detection of local changes in the forest, compared to the normal development. Before such change detection can be carried out, a number of prerequisites, listed below, should be considered.

- Time of year, and number of years between images
- Pixel size
- Geometry
- Sensor type and spectral bands

Regarding the time of year, and time between images, for Swedish conditions, it is preferable that both images are from the mid- or late part of the summer, when the vegetation greenness is stable. Images from spring, early summer, or autumn, might give variations in color that depend on the tree species and these variations might be mistaken for changes. In practice, this means from about June 15 to August 31 for Sweden.

An additional factor is considering the phenomena you are observing, and selecting a number of years between images that makes sense. For example, if you wish to detect clearcuts in a tropical area, where forest regrows much faster than in a boreal forest, you will need to have images with less time between the acquisition dates.

Regarding pixel size, in order to have as few "mixed pixels" as possible and to have a "many pixels per stand" situation, a small pixel size is preferable; on the other hand, "many trees per pixel" will give easily handled mean values. Thus, for the Swedish forest landscape, pixel sizes in the order of 20 m, and down to as small as approximately 5 m, is preferable.

Regarding geometry, the images should be registered to the same co-ordinate system with the same pixel size, e.g. 20 * 20 m pixels, oriented squarely using the Swedish coordinate system.

Regarding sensor type and spectral bands, it is preferable that the same sensor type and spectral band, is used at each time point, but that is not absolutely necessary; while most changes in the Swedish forest will be best visible using a shortwave infrared band, it might also be of some help to make change imagery from more than one spectral band.

9.9.2. Simple change detection techniques

The basic idea in most change detection applications within forestry is that within specific areas (often stands) changes can be seen when comparing earlier and later images. Some different techniques can be used, such as

- Visual display
- Post-classification change
- Image differencing.

Visual display involves quickly switching between satellite images from two subsequent summers (e.g. images from July 2015 and July 2016), new clear-felled areas will appear to be brighter in the later image, whereas the rest of the forest landscape will appear quite similar in both images.

Another simple technique is to display data from different time points using different colour display guns in an image processing system. For example, one band from the July 2015 image is displayed on both the blue and green display colour gun and one of the corresponding bands from the July 2016 image is displayed using the red display colour gun. In this way, areas that have been brighter in the later image (due to clear-felling, for example) will appear as red, while all other areas will have a rather common grey tone (note that colour balancing of the images is often needed to achieve the best result here).

The *Post classification change* method consists of using classifications from two different images and judging the difference between the classes. This method may be seen in the literature, but is not the most accurate way to do change detection. The reason why it isn't the most accurate is because each image will have classification errors. When using both images together, the errors will be compounded, resulting in even more error in the result.

Image differencing is the most common change detection analysis method. Major changes in the forest, like clear-fellings, will show up using very simple techniques listed previously. For detection of more marginal changes

in the forest, a change image might be useful, this is what is commonly meant with “change detection” in remote sensing.

If we let
T1j = the pixel value for pixel j in the early image (image T1)
T2j = the pixel value for pixel j in the later image (image T2)

A usual way for computing a change image is then:

$\Delta j = a\ (T2j - f\ (T1j)) + b$

Where Δj = the pixel value for pixel j in the change image, f is a function that adjusts the pixel values in image T1 to the same grey scale as the pixel values in image T2, and a and b are just coefficients for contrast enhancement of the change image. Often b is chosen to be equal to 127.

The next question is then, which method should be used for estimating the calibration function (f). The reflectance calibration approach would be to transfer each image to the reflectance scale, and then work with the differences between reflectance calibrated images. However, this far, this method has not been accurate enough for forestry. Furthermore, there are also natural reasons (like seasonal differences) for reflectance differences between images. The method currently recommended is therefore to use a statistically based “relative” calibration, based on the grey values in each of the two images.

9.9.3. Some statistical approaches for relative calibration of images to each other

The next question is how to compute the function f, that adjust the radiometry (= the pixel values) of the early image T1, to that of the later image T2. There are two main approaches: pixelwise methods and distribution based methods. Both methods are statistical and we can extract a number of pixel values from each image and do the computations on tabulated data, as illustrated below:

Pixel	T1	T2
1	23	25
2	22	26
3	12	14
etc		

In pixelwise methods, we use the pair-wise pixel values from T1 and T2 in an estimation method, most typically regression, where we give the pixel values for T2 as the dependent variable, and the pixel values of T1 as the independent variable, thereby getting a regression function. The residuals between expected and actually measured values will have the same scaling as T2 and could be used as a change image. Advantages with regression analysis are that it is a flexible and easy understood standard method, and several bands in the early image could also be used to predict one single band in the later image. However, all pixelwise methods are sensitive to geometric

errors between the two images T1 and T2. (One possible way to reduce this problem would be to work with segment mean values instead).

In distribution based methods, we only work with the summary statistics of the pixel values in each image, thus, these methods are not sensitive to geometric errors between the images. One such method, that works well and is implemented in most image processing systems, is histogram matching.

The steps are to:

- compute separate cumulative histograms of both T1 and T2 with pixel values on the X-axis and the percent (%) of pixels within the image on the Y-axis;
- for a DN level in T1, read which % of the cumulative distribution it corresponds to;
- read which DN in T2 has the same cumulative %;
- make a look-up table with T1 DN values on the X-axis, and the values in T2 that has the same cumulative distribution value, on the Y-axis.
- repeat the three previous steps for all DN (0-255)
- update T1 through the look-up table to obtain T1´, which is an image with same grey value distribution as T2
- we can then compute a change image as

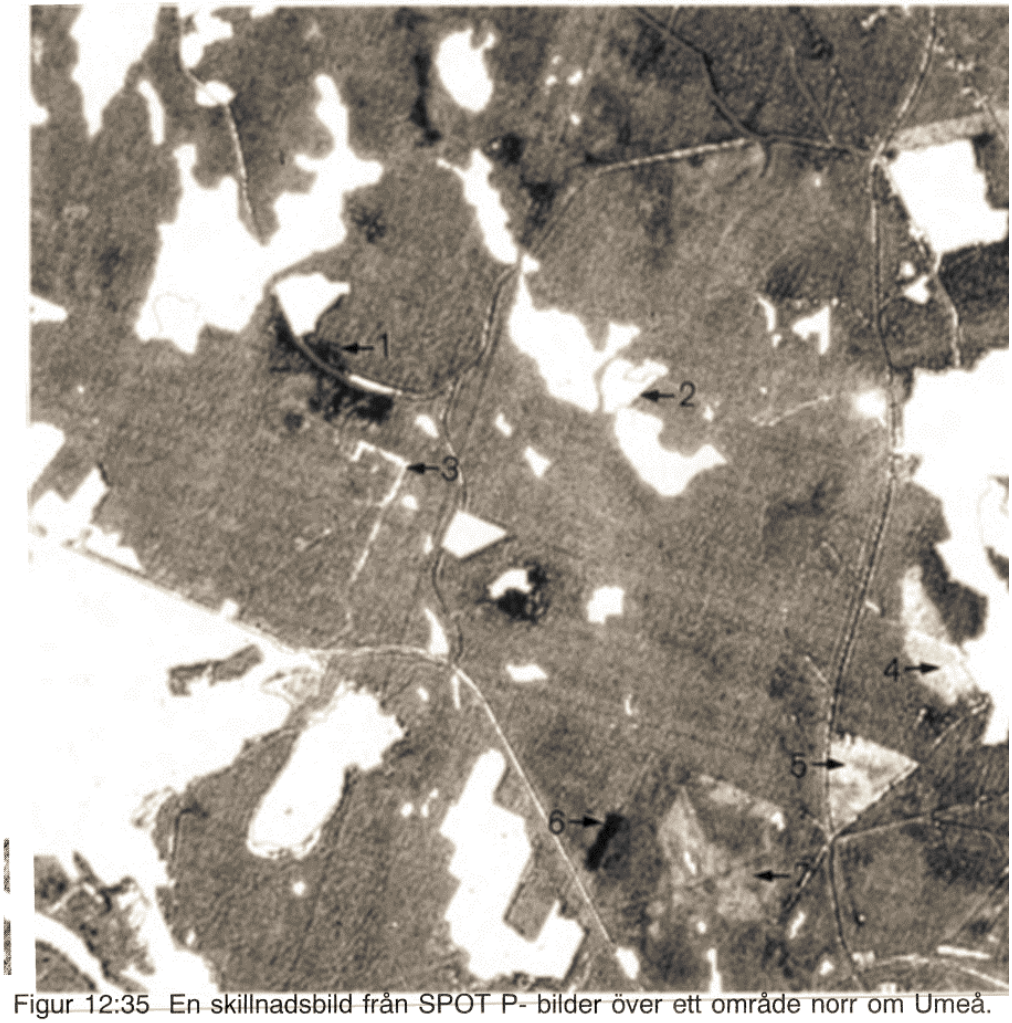
  $\Delta j = a\ (T2j - T1´j) + b$

The next question is which pixel values to base the above procedures on? In certain cases, it may be desirable to use a *forest mask.* They will work for a sample of all pixel values in the image. However, for forestry change detection, they will (at least for Swedish conditions) work even better if they are based on a sample of forest pixels only, which can be defined by a digital map mask, for example.

9.9.4. Which forest changes will be visible in a change image?

The following image (Fig 9:12) comes from the textbook “Flygbildsteknik and Fjärranalys”. The image shows a change image made using two SPOT Panchromatic bands over Brännland, just NW Umeå. A 10 m pixel image from 1986 has been histogram matched against a similar image from 1989, and a difference image has been computed where the adjusted 1986 image has been subtracted from the 1989 image. The middle grey areas shows normally developed forest; the brighter areas are changes that usually are associated with a biomass decrease, or exposure of mineral soil; the darker areas are associated with a biomass increase of a type that have a relatively large effect on the spectral signature.

The numbers in the image, are according to field checks, associated with the following changes:



Figur 12:35 En skillnadsbild från SPOT P- bilder över ett område norr om Umeå. Gråskalorna i bilden har först normaliserats till varandra, varefter den första bilden har subtraherats från den senare. Oförändrade områden har grå nyanser och områden som blivit ljusare mellan registreringarna är ljusa i bilden. För områden med kraftigt växande ungskog eller gräs förekommer även mörka partier i förändringsbilden. Icke skogsmark täcks i bilden av en gul mask. (Avd för fjärranalys, SLU).

1. Snabbt växande ungskog.
2. Kalaverkad skog.
3. Nytt dike.
4. Fröträdsställning.
5. Avverkade fröträd.
6. Hygge med kraftig gräsväxt.
7. Röjd ungskog.

1) the dark area is a young forest plantation with much deciduous trees, that during the 3 year period probably has passed the stage where it forms a closed canopy;

2) the bright area is a clearfelling;

3) the narrow white line is a newly dig ditch with exposed mineral soil, located on a clearfelled area;

4) the light grey area is a final felling in pine forest, with left seedling trees;

5) this light grey area is a final felling, where the seedling trees have been cut

6) this dark area was subject to final felling just before 1986, and since then, there have been a heavy increase in the grass cover;

7) this light grey area is young plantation where a deciduous shrub cleaning has been carried out between the image acquisitions.

In terms of reflectance factors (R), the following is typical for boreal forest in the Umeå area (Table 9.4):

Table 9.4. Reflectance Factors from Landsat TM data.

	Blue	Green	Red	NIR	SWIR1	SWIR2
Medium aged, pine dominated forest	0,03	0,04	0,03	0,16	0,08	0,04
Increase caused by strong thinning	0,001	0,002	0,004	-0,013	0,012	0,010
Increase because of seed tree stands	0,014	0,019	0,032	0,018	0,087	0,061
Increase because of clear felled areas	0,020	0,028	0,048	0,036	0,138	0,102

In other words, the increase in reflectance due to clearfelling is about 10 times greater than that of thinning cuttings. Still, thinning cuttings are often visible in change imagery, especially in pine forest. However, from a change image only, it might be difficult to tell what the type of change it is.

Damages, for example from the fungi Gremeniella that damage pines, snow break damages, or wind thrown trees, can also be detected in change imagery. Under favorable circumstances (good summer images, and a very carefully made analysis) it might be possible to detect damages where only 20 % of the basal area has been damaged. The spectral change of the above-mentioned changes are similar to those caused by thinning cuttings.

The detection of the existence of a change is quite easily done by interpretation of a change image. The delineation of the changed areas are also quite feasible. It is possible to use automated image processing techniques, like segmentation, or interactive line following. A quite crude procedure that often is used in practice, is interactive thresholding of the pixel values in the change image. It should however be observed that there are mixed pixels at the borders of changed objects. Furthermore, the forest surrounding a clear-felled area is sunlit on one side and causes shadows on the other side of the area, and therefore thresholding techniques may not necessarily result in correct delineation of clear-cut areas.

The labeling of the type of change is the most difficult part. Many changes will cause a similar spectral response, for example thinning cuttings, snow break damages, and gremmeniella fungi damages. The combined use of the change image and the early image, will contribute with some information about the type of forest before the damage. However, we have to accept that it is much easier to detect the existence of change than to be able to automatically assign the reason for the change. In reality, field visits or more detailed information (detailed field inventories) are required for certain knowledge of the reasons for change. Automatic procedures will require field references, or calibrated images and known thresholds

9.10. Time series analysis

Another way to use remote sensing data from different time points is to use data from several different time points to answer questions. For instance, by using a number of Landsat images over several decades (e.g., imagine a series of images from 1975, 1980, 1985, 1991, 1996, 2002, 2008, 2012, 2017), we may be able to distinguish the relative age of a forest after a clearcut, or can observe the development of the forest by looking at trajectories (creating a vector of values by observing a single pixel over many time points). We can expect in this type of application to observe forest that has grown in height and density at different rates over the landscape, and can then label them as productive or non-productive forest classes. In another application, 3D remote sensing data can be used to look at forest height development over time, and thereby say something about the forest productivity.

Another way to use a time series is to use satellite images from many dates over different seasons, and in this way, capture phenological changes of the vegetation that help to identify that class more accurately. An example would be the use of this tactic to identify different types of deciduous tree species that have different bud-break and senescence times (e.g., oak and alder). As data with high temporal frequency are acquired (e.g., Sentinel-2) and used increasingly, methods for using spatio-temporal variables will increase.

9.11. Data fusion

There are two ways to look at the term “data fusion”. One is quite literally, and an example of this is the process of *pan-sharpening*. Pan-sharpening is a technique used to combine spatial higher resolution bands with coarser spatial resolution bands. For example, Landsat 7 and 8 have a panchromatic band with a 15 m pixel resolution, while the multi-spectral bands have a 30 m pixel resolution. By using a weighting technique, the information in the four 15 m pixels contained in the multi-spectral 30 x 30 m pixel can be fused to create a multi-spectral dataset that has the appearance of a higher spatial resolution.

Another way to consider “data fusion” is more the approach of using multiple sources of data in a classification or estimation. This might be by using both Landsat-8 and Sentinel-2 data, as well as a DEM. This isn’t a “true” data fusion, as you could refer to this simply as using multiple data sources, but it may sometimes be referred to in the literature as data fusion.

9.12. Data assimilation

9.12.1. Motivation for use of data assimilation

Stand registers are often forecasted and updated after management actions, put the stand register tend anyhow to decay with time, e.g. as a result of major storms. There is a need to develop a new paradigm for keeping forest stand registers up to date, a paradigm that utilizes all available information about the forest, in proportion to the information content each data source can contribute with.

The forest sector can expect to have access to at least the following data types in the future:

- Satellite images from several occasions per year
- Ortophotos and 3D point clouds from digital photogrammetry with a few years interval
- Large area airborne laser scanning with 5-10 years interval
- Ground based data from field plots, terrestrial and mobile laser scanners and harvesters.

9.12.2. The Kalman filter

The *Kalman filter* is a method that combines a forecasted value, with a new measurement, and computes a weighted new value where the forecast and the measurement are given weights in inverse proportion to their variance. The new weighted value will then be likely to be more accurate than the forecast or the remote sensing estimate alone. Figure 9.13 illustrates how estimates from remote sensing and forecasts are assimilated each time a new remote sensing estimate from new remote sensing data becomes available.

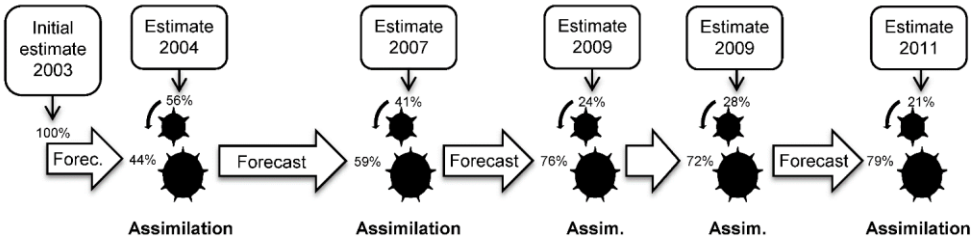


Figure 9.13 Illustration of continuous combinations of new estimates and growth forecasts.

The result from the above indicated *data assimilation* process is that we obtain a model of the forest that is continuously updated and where all remote sensing data has been utilized in proportion to its information content. An additional advantage is that the geographical model of the forest will be up

to date, even if remote sensing data is missing in some areas, e.g due to cloud cover. In those areas, the data will be forecasted from the previous assimilation time point when remote sensing data was available the last time.

Forecasting of forest growth is not a linear process. In order to handle such non-linear cases, a version of Kalman filtering, called the *Extented Kalman filter*, might be used instead.

The Kalman filter assumes that the errors obtained from the forecasting as well as in the remote sensing based estimation, are normally distributed. One way to handle more irregular error distributions is to use *Bayesian statistics* instead. This means that the full error distributions and not only normally distributed errors around an estimate, will be used, regardless of the shape of the error distribution.

The use of data assimilation is a new concept for handling forest inventory data. At the time of writing (January 2017), we are only aware of two studies, both from SLU in Umeå. In a simulation study by Ehlers et al.[17], it was shown that data assimilation was most useful when the remote sensing estimates were less accurate (e.g. the case of optical satellite data), and the forecasting functions were accurate. In the opposite case, with very good estimates from remote sensing (e.g. the case of airborne laser data) and less good forecasting functions, the benefits from the use of data assimilation was only marginal. In Nyström et al.[18], it was shown that data assimilation slightly improved the estimates based on a series of surface models from digital photogrammetry, although not as much that could be expected from a theoretical point of view. The six remote sensing data sets from 2003 to 2011 used in this study are illustrated in Figure 9.13 above.

The subsequent assimilation of each new remote sensing acquisition into a model of the forest which is constantly forecasted is a straight forward concept. It can however be assumed that additional logic beyond statistical weighting only, needs to be added in the future as well. One example might be that certain phenological patterns might be associated with certain tree species, and that series of images and not just the latest image acquisition is needed in order to reveal these patterns. Another example is that series of heights will be needed in order to determine site indices.

9.12.3. The role of field data in data assimilation

The most straight forward way for obtaining reference data for the remote sensing estimates to be used in the data assimilation would be to have a

[17] Ehlers, S., Grafström, A., Nystöm, K., Olsson, H., and Ståhl, G. 2013. Data assimilation in stand-level forest invnetories. *Canadian J. of Forest Research*, 43:1104-1113.

[18] Nyström, M., Lindgren N., Wallerman J, Grafström A., Muszta A., Nyström K., Bohlin J., Fransson J.E.S., Ehlers S., Olsson H., Ståhl G. 2015. Data assimilation in forest inventory: first empirical results. *Forests* 6:4540-4557.

permanent network of field plots that are regularly re-surveyed, e.g. every 5th year. However, field data collection is expensive and additional techniques that might reduce the need for field data should be evaluated as well. One such technique might be to use data from harvesters. Another possibility will be to use ground based or mobile (including person carried) laser scanning of tree stems. Data assimilation has been used in meteorology during more than 60 years now, and the experience from that field as that the need for field reference data will decrease as the models develops over time and becomes more and more complex.

9.12.4. The role of change detection in data assimilation

Data assimilation of for example stem volume assumes in the simple case that the growth over time can be modeled with a growth function, and that those estimates are continuously adjusted with remote sensing estimates. However, in order to accurately model more abrupt changes, like for example storm damages etc, there is a need to include a change detection routine in the data assimilation application. Since data assimilation is based on modelling of forest growth combined with the use of series of images, the change detection routines to be used can utilize more data than only the two latest images.

9.12.5. Application areas for data assimilation

The most obvious application for data assimilation in forestry, would be for keeping stand registers in large forest companies up-to-date. The assimilation might then work both on raster cell and stand level and make use of site index information etc from the stand register.

It will also be possible to use the data assimilation concept for keeping general forest maps made by authorities, e.g. Skogliga grunddata in the case of Sweden, up to date. This assumes however that some data that are missing in today's openly available wall-to-wall products. For example site index, would need to be modeled from series 3D remote sensing data, possibly in combination also with data about the terrain.

9.13. Software used for image processing

Currently there is a movement from use of commercial packages towards open source code for processing of remotely sensed data. For optical satellite data, some common commercial packages are Erdas Imagine; ENVI; IDRISI; and eCognition (especially for segmentation). ArcMap is primarily a GIS software, but due to earlier collaboration with Erdas it has some limited image processing capabilities. QGIS is also increasing its capabilities to include image processing routines. Several commercial software have the ability to write and implement code, such as ENVI's IDL interface. The more commonly available open source software are provided through R and OSGeo4W, although the number of these available are foreseen to increase over the next decade. Finally, those with a knowledge of a programming language, such as C++, can also write their own code to use for image processing.

Statistically, common classification methods, (for example maximum likelihood classification, which as you read earlier is actually a form of Discriminant Analysis) can be found in statistical packages (like SAS, MINITAB, SPSS etc), instead of an image processing system. While this isn't necessary, because most image processing systems (like ERDAS Imagine) have classification routines, it is possible. These days it is becoming more common to use the statistical program R for image classification activities.

10. ACCURACY ASSESSMENT

10.1. Accuracy assessment of remote sensing data products

Accuracy assessment of the map product is often an important element for the users of the data. It is sometimes not carried out, however, for large area projects, due to the lack of reference data, or limitations in project time or funding. An independent and objectively collected evaluation data set is essential to an unbiased assessment of the map product. Accuracy assessment numbers give the user insight into the average accuracy of the map, the accuracy of different classes/categories, and why errors may be happening.

10.1.1. The need for probability sampling

Data for accuracy assessment should be collected with a method that is much more accurate than the product to be evaluated. In this case field visits (with accurate GPS positions), or interpretation of aerial photos are most often used for accuracy assessment.

Accuracy assessments should be done with a sample that is objectively selected according to the rules of probability sampling. This means that a random sample of evaluation points is perfectly valid, and also a systematic sample is quite OK from most aspects. It is also allowed to use a stratified approach, where a different number of evaluation points is used for different strata. However, what is not allowed in probability sampling, is to sample reference plots in a *subjective* way, such as “where it is convenient”, or to “move” plots that happen to be located in less favorable places. If there is a need to avoid certain types of plot locations, rules for this have to be created in advance (for example after a pilot survey), and the accuracy assessment figures are then only valid with the restrictions given by those rules. An important characteristic of validation data is that it should have high quality measurements (position and inventory measurement).

10.2. Assessing thematic class accuracy

The number of accuracy assessment plots needed is often an issue, and the answer to that question often needs a statistical analysis that is not part of this course. However, there is a rule of thumb, where 50 samples per class are recommended (Congalton 1991).

10.2.1. The error matrix

The most usual way to analyze and present an accuracy evaluation is to construct an *error matrix*, also called a *confusion matrix* or *contingency*

table. The error matrix is a table with a row for each class in the classification, and a column for each category according to the field (or photo) evaluation. A simple example of an error matrix is given below:

	Field data				
Data according to classification:	Coniferous forest	Deciduous forest	Non- forest	**Total**	**User's Accuracy**
Coniferous forest	**30**	2	20	52	58%
Deciduous forest	3	**10**	2	15	67%
Non-forest	4	5	**30**	39	77%
Total	37	17	52	106	66%
Producer's Accuracy	81%	59%	58%		
Overall Accuracy	66%				

The *"producer's accuracy"* = how large a *proportion of the field data* for a given class was correctly classified:
Coniferous forest: 30 / 37 = 81 %
Deciduous forest: 10 / 17 = 59 %
Non-forest: 30 / 52 = 58 %

The "*user's accuracy*" = how large a *proportion of the classified data* for a given class had the correct label according to the field data:
Coniferous forest: 30 / 52 = 58 %
Deciduous forest: 10 / 15 = 67 %
Non-forest: 30 / 39 = 77 %

The ratio between Producer's and User's Accuracy for a single class **c**an reveal whether a class has over-classified or under-classified, according to the reference data. For example, there were many more instances of coniferous forest in the classification than in the reference data. Coniferous forest is "over-classified".

The most used measure of classification accuracy is *overall accuracy*, which means the total number of correctly classified pixels, divided with the total number of pixels, in the above table, the overall accuracy is estimated as:

(30 + 10 + 30) / 106 = 66 %

In general, an overall accuracy of 80% for about 8 land cover classes is rather reasonable.

The *Kappa coefficient* ($\hat{k}$) or *k-hat* is a measure where the overall accuracy have been reduced with the accuracy that only depends on the chance agreement between the classification and the field evaluation. The measure has a value between 0 and 1, and is always lower than the overall accuracy. The detailed formula is given in Lillesand et al. The general formula is:

k-hat = (observed accuracy – chance agreement) / (1 - chance agreement)

10.3. Assessing continuous estimate accuracy

For continuous estimates RMSE is a common way to assess error.

The RMSE can be calculated as:

$$RMSE = \sqrt{\frac{\left(\sum_{i=1}^{n} (\Delta_i)^2\right)}{n}}$$

where $\Delta_i = (\hat{v}_i - v_i)$

Δ_i is differences between estimated parameter ($\hat{v}_i$) and the field measurements (or "truth") of the same parameter (v_i) for the *i*th observation (in our previous lab each observation was each stand in the database), and *n* is the number of observations.

Relative RMSE is comparing the error of the estimated parameter as a percent to the mean value of the field measured parameter. For example, if the RMSE of the estimate was 10 and the field measured parameter's mean was 100, then the relative RMSE would be 10%.

Bias can also be calculated as $\frac{\sum (\hat{v}_i - \bar{v})}{n}$

10.4. Cross-validation

Cross-validation is a method used when independent data are not available to conduct a separate accuracy assessment, because the data have been used for training or model creation. Leave-one-out cross-validation removes one of the plots from the reference data set, creates the model with the remaining data, and assesses the accuracy for the single plot not used. The procedure then goes through the entire data set in this manner, and the results are compiled. Cross-validation gives one a good idea of model-performance, and is not a "true" assessment of the result, but can be used as an indication when there is a limited reference data set to work with.

10.5. Considerations about the collection of field data

There will be geometry errors between the field data and the image data, which will reduce the apparent classification accuracy if a "pixel wise" evaluation is done. Furthermore, it is often difficult to obtain a correct class label for just a point in field, at least without a lot of time consuming measurements. To judge a class label for a bit larger area is often easier. A classified map itself is also more accurate if a mean value for a segment is

used. These are some arguments about why it might be better to evaluate accuracy for groups of pixels (e.g. segments with the size of stands) instead of single pixels. Furthermore, a judgment in field might often be near the class definition between two different classes. There are also ways to compute an error matrix where classes that are very similar can be considered less wrong than classes that are very different, which is called "fuzzy accuracy assessment".

Other tricky topics are how to get accuracy assessment plots for change detection? Some possible data sources are serial photos from past time points, long term inventory series, or you might simulate a change.

10.6. Further reading

Stehman and Czaplewski (2003) defined four criteria that should be met: 1) probability sampling, 2) adequate sample sizes with which to estimate user's accuracies with acceptable level of precision, 3) cost efficiency must be considered, and 4) spatial distribution of samples must be representative across the area of interest.

Stehman and Czaplewski (1998) have established three basic elements to consider in the design of an accuracy assessment plan: the sampling design, the response design, and the estimation and analysis protocol. The sampling unit may be a pixel, fixed-area plot or polygon, although the optimal unit depends on the application. Stehman *et al.* (2000) favor pixel-based evaluation units, as larger units render the results non-site specific. Polygon assessments also tend to lead to conservative estimates of classification accuracy (Verbyla & Hammond, 1995). Class homogeneity within the accuracy assessment unit is appealing, but not necessary, and if intentionally included, may bias the assessment of the map accuracy. A design-based sample with known inclusion properties is best, but the distances between plots should be large enough that potential spatial auto-correlation effects do not influence the result. Definitions constituting correct and incorrect responses should be established (e.g., if polygon accuracy assessment units are used, the rule may be that a "correct" classification requires a majority of the classified pixels to be correctly labeled as the dominant class, or the rule may be that the two most dominant classes must both be classified).

As with training data, questions regarding the sample size and sampling scheme of the accuracy assessment data need to be addressed. Stehman (2001) suggests that a sample size of 100 samples per class assures the population is estimated adequately. Congalton and Green (2009) suggest a minimum of 50 samples per class. To capture the necessary number of samples for rare classes, a stratified sample may be useful (Stehman, 2001).

The quality of evaluation data is of much importance, and a quality check of the evaluation data should be carried out before use.

Accuracy assessment is often presented in an "error matrix", with errors of commission (called "user's accuracy") and omission (called "producer's accuracy") for each class in the map, as well as a measure of overall accuracy. The kappa statistic, or k-hat, is also a measure of overall accuracy, and is intended to account for the chance of random agreement (Congalton and Green, 2009). A desired overall map accuracy of 85% is often given as a benchmark, but may not be realistic to achieve (Wulder *et al.*, 2006). Fuzzy accuracy assessment (Gopal & Woodcock, 1994; Foody, 2002) can be a useful measure of portraying different types of errors that may be more or less acceptable.

It is slightly more common that inventory data are used for accuracy assessment than for training (e.g., Riemann *et al.*, 2010). Wulder *et al.* (2006) encountered difficulties when applying 2 ha polygon-based inventory data due to differences between the raster and vector data, particularly because the polygon interpretation included heterogeneous cover. When purpose-collected video data were later photo-interpreted for accuracy assessment, the uncertainty in the photo-interpretation and the lack of a probability-based sample were drawbacks (Wulder *et al.*, 2007). One of the primary requests emerging after Canada's EOSD land cover mapping project was for improved collection strategies of calibration (training) and validation (accuracy assessment) data (Wulder *et al.*, 2008).