FACULTY
OF SCIENCE

Using forest reflectance modelling to estimate albedo for narrow-view satellites

by

Nick Bergmans nick@fambergmans.net

Thesis submitted for the degree of Master of Science
Project duration: 4 months
Supervised by:

Prof. Lars Eklundh and Dr. Hongxiao Jin
Department of Physical Geography and Ecosystem Science

Dr. Johan Friberg
Department of Physics
Division of Nuclear Physics

October 2023

Abstract

Surface albedo estimations from remote sensing are invaluable for energy balance descriptions in climate research. The aim of this work is to evaluate the possibility of using Sentinel-2 satellite data to get the total albedo of a small, vegetated area. In this thesis, a simple machine learning model is created to convert simulated Sentinel-2 reflectance measurements to surface albedo.

The Forest Reflectance and Transmission model is used to simulate the reflectance behaviour of a pine forest and a field vegetation stand. The angular behaviour was analysed, and it was found that the directional reflectances at near-nadir view angles are typically lower than the hemispherical reflectance, though this effect also depends on the solar angle. As a result, it is important to include angular modelling in the albedo estimations. Next, a direct- estimation and narrow-to-broadband linear regression model were trained with FRT simulated reflectance data. The ability to predict albedo from the FRT data was high, with $r^2 > 0.94$ for the direct estimation and $r^2 > 0.999$ for the narrow-to-broadband model. Finally, the models for the pine forest and field were compared and found to be distinctly different.

In conclusion, it is important to include angular modelling in albedo estimations. The regression models presented in this thesis perform well for the simulated vegetation stands. Moreover, it is valuable to train separate models for different land use classes. With some further improvements, the regression models for Sentinel-2 data have potential for accurate evaluation of surface albedo at a fine spatial resolution.

Acknowledgements

First and foremost, I would like to thank all my supervisors for their guidance in this project; I want to thank Johan Friberg for always having time to talk and giving advice, and I would like to thank Lars Eklundh and Hongxiao Jin for introducing me to their research field, for always thinking along with me, and for giving me valuable feedback.

I also would like to Andres Kuusk for being available to explain his model, and for taking my questions seriously.

In addition, I want to thank my office roommates for always being available for a chat and keeping the spirits high.

Lastly I want to thank my parents, my girlfriend, my friends and, not to forget, the family dog, for their unwavering support during the more challenging times.

Contents

List of Abbreviations

ASTM	American Society for Testing and Materials
BOA	bottom of atmosphere
BRDF	bidirectional reflectance distribution function
BHR	bihemispherical reflectance
BSA	black-sky albedo
DHR	directional-hemispherical reflectance
ESA	European Space Agency
FOV	field of view
FRT	Forest Reflectance and Transmittance
HDR	hemispherical-directional reflectance
MISR	Multi-angle Imaging SpectroRadiometer
MODIS	Moderate Resolution Imaging Spectroradiometer
MSE	mean squared error
MSI	MultiSpectral Instrument
NASA	National Aeronautics and Space Administration
NIR	near-infrared
NREL	National Renewable Energy Laboratory
S2A	Sentinel-2A
S2B	Sentinel-2B
SWIR	shortwave infrared
TOA	top of atmosphere
VIIRS	Visible Infrared Imaging Radiometer Suite
VNIR	visible and near-infrared
WSA	white-sky albedo

1

Introduction

1.1 Research problem

Remote sensing and satellite data have been invaluable for monitoring the Earth in recent years [1, 2]. Surface albedo is a quantification of the reflectivity of the surface for solar radiation. This reflectivity depends on the land cover. It is an important factor in Earth system modelling and climate change studies. A higher albedo means that more light is reflected by the earth's surface. As a result, areas with high albedo experience less warming of the surface and the surrounding air. A high global surface reflectance thus decreases the energy that is used to heat the Earth system [3]. In the interest of decreasing the accelerated global warming then, a high global albedo is desired.

In response to the warming climate, vegetation in the arctics has increased productivity. As a result, snow in high latitude areas is masked by protruding plants, and the albedo of the regions is decreased. This decline of albedo causes a feedback loop that can cause warming in these areas to be amplified [4, 5]. Additional positive feedback comes through changes in evapotranspiration in the tropics, the carbon balance and ocean albedo [1, 6].

The vegetation plays a critical role in these processes, since land usage impacts the albedo directly. Choosing to change the land usage or vegetation is a way to influence the albedo. Before doing this, the effects on the ecosystem should be studied thoroughly [7]. Also, it is critical to know the effects on the surface albedo of changing the vegetation.

We are interested in surface albedo estimates both on a large scale and more local scale. A common way of gathering surface albedo is through Earth observation satellites, both on a large [8] and a local scale [9] scale, by a variety of satellite systems.

Satellites measure the surface reflectance under a specific view angle. Because of anisotropies in almost any natural surface, the reflectance of a typical surface depends on the angle of the incoming light, as well as the angle under which the reflected radiation is measured [10, 11]. To make an accurate estimate of the total reflected radiation, the geometrical distribution needs to be considered. A conversion has to be applied to obtain the reflectance in all directions from the directional measurements.

Besides this geometrical consideration, the narrowband satellite measurements also need to be converted to a broadband albedo. The satellite sensors measure the reflectances for specific wavelength bands, but for energy balance considerations, one is interested in the reflection over the whole wavelength spectrum. As not all wavelengths are reflected equally, a transformation has to be applied to get an estimate for the total albedo.

1.2 Aim

The aim of this work is to evaluate the possibility of using Sentinel-2 satellite data to obtain the total albedo of a small, vegetated area. This is realised by creating a simple machine learning model to convert the Sentinel-2 reflectances to surface albedo. Instead of using actual data, the model will be developed based on reflectance modelling of the surface through the Forest Transmittance and Reflectance (FRT) model by Kuusk and Nilson [12, 13]. The focus is on developing a model for specific locations and situations, and evaluating its uncertainties.

Models to get surface albedo from satellite data have already been developed for various satellite systems. An overview of research on this topic until 2015 is given by Qu et al. [14]. Most notable is the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument of the Terra and Aqua NASA satellites [15]. A model for converting its directional, narrowband reflection measurements to a hemispherical, broadband albedo was developed by Strahler et al. [11]. A more recent example is the albedo estimation for the Landsat system [16]. An overview of the estimation methods for Landsats various generations is given by He et al. [17].

A downside of these remote sensing systems mentioned previously is their spatial resolution. Especially the MODIS instrument has a relatively coarse spatial resolution. The MODIS wavelength bands with the finest spatial resolution are able to discern areas of 250 m [15]. Landsat has a spatial resolution of 30 m in its multispectral bands, 15 m in the panchromatic band [16].

For the analysis of patchy landscape, with many smaller areas of different vegetation, a fine spatial resolution is required for better measurement accuracy [18]. The Sentinel-2 satellite constellation provides measurements with high spatial resolution in thirteen sensor bands. Four of the bands have a 10-metre spatial resolution, six have a 20 meter and three have a 60-metre spatial resolution [19].

Preliminary results from Li et al. [20] show that improving the spatial resolution of the albedo data has a positive influence on the accuracy of the albedo measurements in smaller areas. Their results from the Sentinel-2A satellite were compared with those of MODIS and the Visible Infrared Imaging Radiometer Suite (VIIRS). When setting the satellite results against local surface albedo measurements with a pyranometer, the error in finer resolution albedo products was decreased compared to those of courser resolution albedo products.

The area that pyranometers measure is in the range of several meters to several scores of meters [10]. As a result, there is a mismatch between the courser albedo products from for example MODIS and the measurements used to validate these products, making validation difficult. The finer resolution albedo products that Sentinel-2 could possibly provide, are better able to resolve the same spatial details, making them valuable for validation.

The albedo estimation models developed for a satellite system, such as MODIS and Landsat, are not simply transferrable to other systems. One reason for this is a differing range of angles under which surfaces are scanned [21]. Additionally, the wavelengths of the sensor bands do not generally coincide.

Specifically for Sentinel-2, algorithms for albedo estimation have been proposed by Li et al. [20] and Bonafoni [22]. The model from Li et al. is a regression model, trained with radiative transfer simulations of various surface types. Bonafoni uses a theoretical weighting of the sensor bands based on incoming solar irradiance. Both these models are

developed for a broad range of cases. While Li considers snow-covered surfaces separately from snow-free surfaces, neither work makes a distinction in the albedo estimations for different vegetation types. Since different vegetation types have different spectral signatures, it may lead to more accurate albedo predictions when a conversion model is tailored for a specific vegetation type. The existing models also don't take the solar- and view angles into account.

1.3 Overview

In this thesis, we aspire to evaluate our surface albedo estimations from simulated Sentinel-2 satellite reflectance data. First, some background information on remote sensing and the reflectance of vegetation is given. Then, the methods used in this research are explained.

The research problem is split into four parts. The surface reflectance is modelled with the Forest Transmittance and Reflectance (FRT) model by Kuusk and Nilson [12,13], and its results are used for calculations and analysis in every part.

1. First, the behaviour of the reflectance is investigated. The variation of the monochromatic reflectance for different view- and solar angles is described. Additionally, the reflectance at different wavelengths is analysed. This part of the analysis is done to learn the impact of the geometry and the wavelengths on the albedo.

2. The second aim is to evaluate the legitimacy of the Lambertian approximation. In this approximation, the light reflected by a surface is completely diffuse. If this is a valid assumption, no angular modelling is required for albedo estimations.

3. Next, two simple machine learning models are created to predict the surface albedo from Sentinel-2 reflectances. Instead of using satellite data, the models are developed based on theoretically modelled reflectances from the FRT model for a specific type of vegetation. The models are valid under the assumption of a constant solar angle, as well as under fixed atmospheric conditions. The predictions from the regression models are compared to each other, as well as to predictions from the models by Li et al. and Bonafoni.

4. As the last sub-aim, the impact of different vegetation types on the albedo estimation is investigated. Specifically, two types of vegetation are considered: a pine forest and an open field.

The results from these four sub aims are presented and discussed. Finally, the conclusions and the outlook of this work are given.

2

Theory

2.1 Principles of remote sensing

Remote sensing is gathering information about something from a distance. In most applications, this information gathering is the measuring of physical characteristics. Remote sensing relies on the detection of emitted and reflected electromagnetic radiation [23]. Mostly, it is used to monitor the Earth's surface, and this is what this thesis will focus on as well. However, in principle remote sensing systems can be, and are, applied to observe many other targets.

A remote sensing system consists of several components working together, with first and foremost the data acquisition system. This includes the sensor and the platform on which the sensor is mounted. This platform can be on the Earth's surface, but is frequently airborne or spaceborne. Common examples of such platforms respectively are towers, aircrafts or satellites [10].

Remote sensing systems can be divided into active and passive, based on their working principles. A passive sensor will detect either the natural radiation emitted by the observed object, or the radiation reflected by the object. On the other hand, an active sensor will send out its own electromagnetic radiation towards the object and will then detect the reflected or backscattered radiation.

It is useful to understand some basic terms and concepts of satellites in general. After this, the specifications of the Sentinel-2 satellites can be discussed further to understand the motivation of using its data for estimating the albedo of different vegetation types.

2.1.1 Satellites

Satellites are frequently used in remote sensing of the Earth system. Their purposes can vary greatly, but a few common uses for satellites are geopositioning, weather- and climate monitoring and assessing land use.

Satellite orbits

There are many satellites going around the Earth at the moment. Many of these are constantly gathering information about the part of the surface within their view. The type of satellite and its orbit depend on the purpose of the satellite. One commonly used orbit is the geostationary orbit. In this orbit, the satellite stays over one specific longitude at the equation. These satellites need to fly at an altitude of 35786 km or very close to this.

Another common orbit type is the polar orbit. Satellites in this orbit fly over the poles. They can cross from one pole to the other at different longitudes every new orbit. In this way, these satellites in polar orbit can observe the entire surface of the earth. Most polar orbiting satellites are also sun-synchronous, meaning that they always pass over a specific spot on the surface at the same local time. These satellites have an altitude around 800 km, which is lower than that of the geostationary satellites.

Resolution

Considering the resolution of satellites, often the spatial resolution comes to mind. Depending on applications, different types of resolution are also important.

The spatial resolution is given by the land area covered by one "pixel" of the data. The finer the spatial resolution, the more details of the land cover can be discerned. The spectral resolution describes which radiation wavelengths are detected, and how wide the detection bands are. The radiometric resolution is determined by the data width of a measurement. A higher radiometric resolution means that the sensor readings will be stored with a higher number of digits. With sufficiently accurate sensors, this results in a higher precision of the measured quantity. Another factor in the radiometric resolution is the Signal to Noise Ratio (SNR), which is preferably high [19].

Finally, the temporal resolution describes how often the satellite repeats its measurements in the same location. This can be important if one wants to monitor changes over time. For geostationary satellites this might be arbitrarily high, since it is stationary about one part of the land. However, for satellites that pass over a target once and then have to fly a whole orbit in order to return and make the next measurement, this is not so.

2.1.2 Sentinel-2

In this thesis, the measurements from Sentinel-2, specifically those above the south of Sweden, are simulated. The purpose of this is developing a model to estimate albedo based on Sentinel-2 data. The Sentinel-2 satellite constellation from the European Space Agency is a system that provides data at fine spatial resolution. It is currently made up by 2 satellites, Sentinel-2A(S2A) and Sentinel-2B(S2B). They move in the same polar, sun-synchronous orbit, at an altitude of 786 kilometres. They have a phase difference of 180°, meaning that the two satellites are located on opposite sides of the orbit. Because of this, the passover frequency of the combined system is doubled [19].

MultiSpectral Instrument

The measurement system responsible for the light intensity detection is the MultiSpectral Instrument (MSI) developed by Astrium SAS (now Airbus Defense and Space). The incoming light beam is split using a beam-splitter, to separately focus the Visible and Near-Infra-Red (VNIR) and the Short Wave Infra-Red (SWIR) bands.

The Field of View (FOV) of the MSI instrument is 20.6°. This is achieved by having 12 detectors placed in the focal plane in a staggered formation. This FOV means that the measurements are performed between a zenith angle of -11.3° and 11.3°, perpendicular to the propagation direction of the satellite. Because of this, the instrument is able to scan a 290 km wide swatch on the surface of the Earth. With this so-called push broom sensor design, the satellite does not have to vary the scanning direction of the sensor while

flying over, as is required with a whisk broom sensor design. Because of this, the time for scanning one pixel is increased, and the sensor obtains a stronger intensity signal [19, 24].

Sentinel-2 resolution

The MSI on Sentinel-2 has 13 different spectral bands for which the light intensity is measured. The wavelengths and the spatial resolution of the sensor bands of the Sentinel-2 satellites are given in Figure 2.1. The different bands have 3 possible spatial resolutions. Sensor bands 2, 3, 4 and 8, all located in the VNIR have the finest spatial resolution, of 10 m. Of these, bands 2, 3 and 4 lie in the visible spectrum and 8 in the NIR. Together with bands 11 and 12 in the SWIR, which have a spatial resolution of 20 m, they measure a large part of the solar spectrum. Additionally, there are four other bands with a special resolution of 20 m, located in the VNIR, and three bands with a spatial resolution of 60 m. These last three bands are mainly used for cloud detection (band 10) and atmospheric corrections [19]. The exact wavelength ranges of the bands vary up to several nanometres. For the exact wavelength specifications of the bands for the two satellites, see Appendix A.

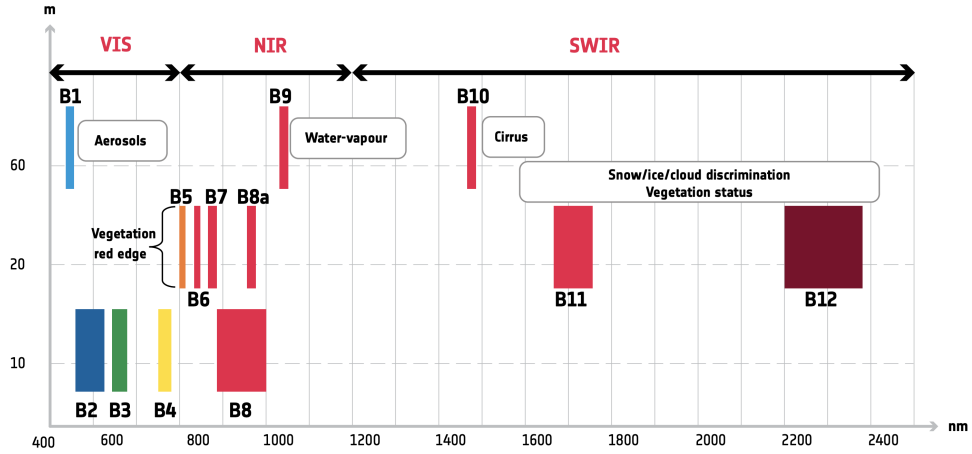

Figure 2.1: The spatial resolution and wavelength domain of the 13 Sentinel-2 bands. Image acquired from ESA [25]

.

The radiometric resolution of the MSI is 12-bit, meaning that 4096 different light intensity values can be registered by the sensors. The SNR varies for the different sensors [19]. For SNR specifications, again refer to Appendix A. When combining the 2A and 2B satellites, the maximum temporal resolution is 5 days.

Sentinel-2 data

The data from the two Sentinel-2 satellites can be obtained at several processing levels [19]. In the Level-0 processing, which happens in real-time, the data is consolidated and metadata is appended. In Level-1 data processing, the data is decompressed and radiometric and geometric corrections are applied. The Level-1 processing is split up in Level-1A, 1B and 1C. At level 1C the data has undergone spatial registration and orthorectification. Briefly, spatial registration involves aligning the pixels in the reflectance data with the correct locations on the ground. Orthorectification is done to correct the

distortion from sensor tilt and elevation differences[1] [26].

At Level-1C, the data is ready for handling by the average user. The reflectances for each sensor band at the Top Of the Atmosphere (TOA) are given, and the data is registered on a geographic coordinate system. The data is supplied with metadata such as the time and the angles at which the measurements were made.

Further processing is done at Level-2A. This processing mainly contains atmospheric corrections and gives the user information about the atmospheric conditions at the time and location of the data acquisition. The main output here are the reflectances at the bottom of the atmosphere (BOA) and maps with the optical depth and water vapour.

The data at level 1B, 1C and 2A is publicly accessible, where 1B is only for expert users. The data can be freely accessed via the Copernicus Data Space Ecosystem [27], or various other satellite data hubs.

2.2 Reflectance

Reflectance can roughly be described as the ratio of the reflected radiant flux and the incoming radiant flux on a surface. When talking about the reflectance of a natural surface such as a forest, it is important to keep in mind that this is not a constant value, but that the reflectance is anisotropic [28]. The reflectance is dependent on the angle under which the light comes in, as well as the outgoing angle. The flux of reflected energy is, generally speaking, not equally divided over the whole hemisphere. Similarly, the reflectance of a surface depends on the wavelength of the light.

Relevant geometry

The location of the satellite and the sun relative to each other and the surface of the Earth are an important factor in reflectance measurements. This section provides a brief introduction on relevant geometrical terms.

In Figure 2.2, the situation is visualised schematically. Here, the yellow dot is a representation of the sun[2], and the angle under which its beam reaches the Earth's surface. The green dot similarly shows the location of the satellite, and the grey area is the tangent plane to the surface of the Earth at the measured location (0,0,0).

To describe the location of either object in a fixed reference system, we make use of the zenith angles θ and azimuth angles ϕ. If the zenith angle $\theta = 0°$, the angle is said to be nadir, pointing directly downward. The angles corresponding to the sun's location are the solar angles and are denoted with the subscript 's'[3]. The angles corresponding to the satellite are supplied with the subscript 'v', for view angle.

When considering a point at the Earth's surface, the zenith angles are generally clear, since our reference system has the z-axis pointing up orthogonally from the surface of the Earth. For some purposes, only the relative azimuthal angle ϕ_r, between the sun and the satellite is considered. For describing the angular dependence of the reflectance, this is sufficient under the assumption that the surface within the spatial resolution of the satellite does not present any extreme directionality.

[1] For more details on types of data corrections in remote sensing, the reader is referred to the book on remote sensing and image processing by Schowengerdt [26]

[2] Since the sun is not so close to the Earth surface, the radial component of its location is not to scale. The dots in the illustration are meant to visualise the angular component of the light beams.

[3] Take care that this subscript should not be taken as being the 's' from satellite.

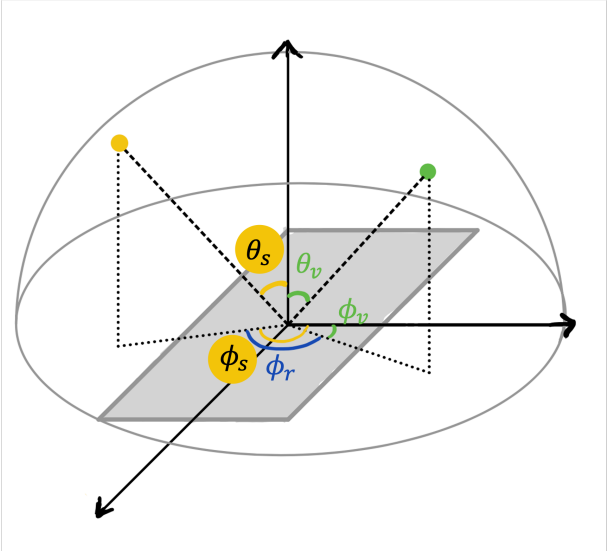

Figure 2.2: The relevant angles used to describe the geometry of the satellite and the sun. The yellow dot represents the sun angle and the green dot represents the satellite angle.

2.2.1 Spectral irradiance

When measuring the surface reflection on the Earth, the incident radiant flux is the solar irradiance. The TOA solar irradiance spectrum is shown in Figure 2.3, and compared to the spectrum of a black-body at $T = 6000K$ [29].

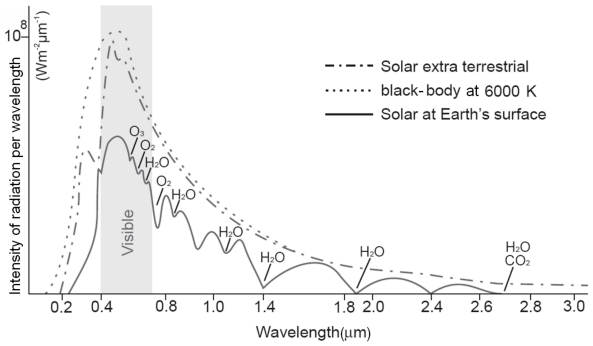

Figure 2.3: Solar irradiance spectrum from a black-body, at the top of the atmosphere (TOA) and at the Earth's surface (BOA). Acquired from Tempfli et al. [29].

The BOA irradiance is also shown. The intensity of this spectrum is decreased because of interaction with the atmosphere. The molecules, particulates and aerosols cause scattering of the incoming (and also the reflected) light. This effect is more prevalent for the shorter wavelengths in the spectrum [30]. Additionally, the electromagnetic radiation is partly absorbed by various molecules in the atmosphere. Some of the molecules with the highest absorption are ozone (O_3), water vapour (H_2O) and carbon dioxide (CO_2) [29]. In Figure 2.4, the transmission coefficient for the solar wavelength spectrum is given. Big dips can be seen corresponding to absorption by certain molecules. In between these dips, the transmission is high. These high transmission ranges are referred to as atmospheric transmission windows, and typically sensors detect radiation in these wavelength ranges [31].

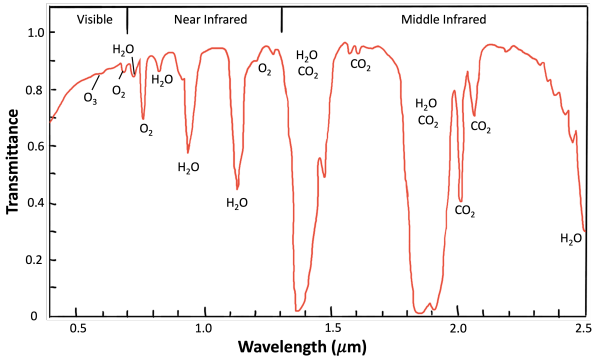

Figure 2.4: Transmittance spectrum through the atmosphere. Atmospheric windows of high transmittance can be seen in between low transmittance regions caused by molecular absorption. Acquired from Smith [32].

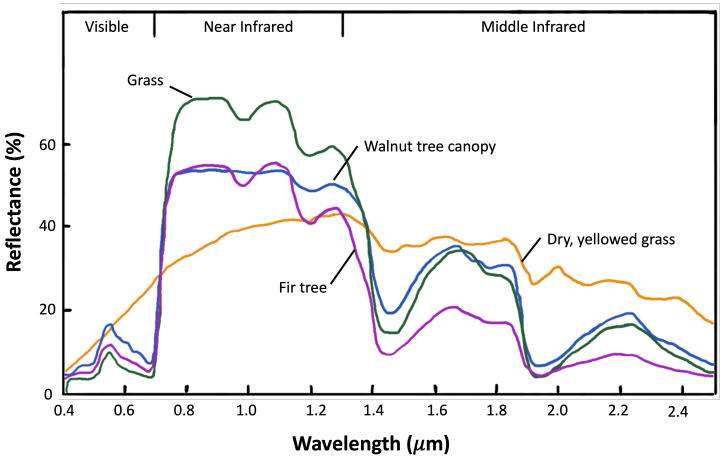

Figure 2.5: Reflectance spectrum for different types of vegetation. Acquired from Smith [32]

2.2.2 Surface reflection

Different surface covers absorb and reflect light of different wavelengths. Their reflectance spectrum depends on three main factors [30]. These are the biochemical composition, the structure and the moisture content of the surface. A combination of these factors gives every material its own spectral signature. This signature is the reflectance factor plotted over the wavelength. In Figure 2.5 the spectral signature of various types of surface cover is shown. Some wavelengths with high atmospheric absorption can be recognised in this spectrum when comparing Figures 2.4 and 2.5. They can be seen as drops in the reflectance rate, such as the ones just before 1000 nm or around 1400 nm.

Because different vegetation types display a distinct difference in their spectral signatures, the reflection spectra of land surfaces can be used to classify land cover or monitor vegetation health through remote sensing [33].

Different factors influence different parts of the spectrum. The visible wavelengths are impacted by, for example, concentrations of pigment, chlorophyll and iron. Examples of factors that influence the near-infrared (NIR) wavelengths are cell structure and canopy structure. Higher wavelengths are impacted among others by moisture content and plant stress.

In the reflectance of vegetation, one typically sees a steep increase in the reflectance spectrum when increasing the wavelength from the red band to the NIR range. This range of wavelengths with rapid change is called the red edge and is useful in determining

chlorophyll and nitrogen concentration in the vegetation, as well as plant stress[4] [34, 35].

Another characteristic of the surface reflectance is the hotspot effect. When looking at the geometry of the reflection, the principal (solar) plane is defined as the plane that contains both the surface normal and the sun. If the satellite is located somewhere in this plane, $\phi_r = 0°$. In this plane, around $\theta_v = -\theta_s$, the reflectance has a peak. The hotspot is this angle where the reflectance is increased.

2.2.3 Bidirectional Reflectance Distribution Function

For a completely smooth surface, the light of a single incoming ray will be reflected in only one direction. Merely measuring the reflected light in this view angle will then give the total reflected radiant flux.

On the other hand, for a completely diffuse surface, the light of a single incoming ray is instead reflected equally in every direction. Such an ideally diffuse surface is called a Lambertian surface. For a Lambertian surface it is sufficient to measure the intensity of the reflected light in one direction. Lambert's cosine law [36] then states that the measured intensity is only dependent on the cosine of the zenith angle.

$$I = I_0 \cos\theta \tag{2.1}$$

Here, I is the measured intensity at zenith angle θ and I_0 is the intensity at $\theta = 0$. The factor $\cos\theta$ is a result from the decreased solid angle when an area is viewed from a non-nadir zenith angle (Figure 2.6). In this case then, the reflectance only needs to be measured at one angle to get the hemispherical reflectance.

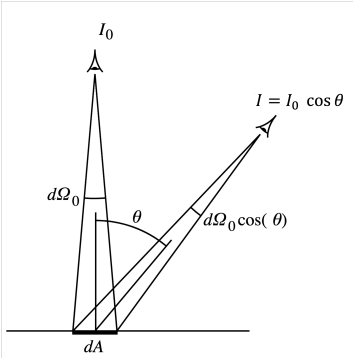

Figure 2.6: Lambert's cosine law. The measured intensity depends only on $\cos\theta$ *for a completely diffuse surface.*

However, for a more natural surface, this is not possible. For such surfaces, the reflectance in every view angle description is described by the bidirectional reflectance distribution function (BRDF). The BRDF is the fraction of the incoming energy from the incident angle that is reflected in the view angle, and is mathematically defined as [10]

$$f_r(\theta_i, \phi_i; \theta_r, \phi_r; \lambda) = \frac{dL_r(\theta_i, \phi_i; \theta_r, \phi_r; \lambda)}{dE_i(\theta_i, \phi_i; \lambda)}. \tag{2.2}$$

Here, dE_i is the increment of spectral irradiance at the angle (θ_i, ϕ_i). It is given per unit solid angle and per unit area surface area. dL_r is the increment of reflected spectral

[4] Since this small area of the wavelength spectrum is of interest for monitoring vegetation, satellites often have sensor bands specifically in this range. Sentinel-2 in particular has three 20 m resolution wavelength bands in this range (see Figure 2.1)

radiance at (θ_r, ϕ_r). The angles (θ_i, ϕ_i) and (θ_r, ϕ_r) correspond to the solar and view angles (ϕ_s, θ_s) and (ϕ_v, θ_v) defined previously. All the terms in this expression are dependent on the wavelength λ, but this is often omitted as an explicit argument in the expression.

The BRDF is determined only by the surface properties of the measurement location, like the land surface structure and the spatial distribution of vegetation. Additionally, the BRDF is influenced by the optical properties of the surface components and by shadows [11]. The measurement conditions do not play a role in the definition of the function.

For an ideal surface, all the incoming energy is divided equally over the hemisphere upon reflection. The fraction of the light reflected in direction (θ_r, ϕ_r), so the BRDF, is then just the constant $\frac{1}{\pi}$.

In practice, various physicals quantities based on the BRDF are used to describe the reflectivity of natural terrain. The reflectance, ρ, is defined as the ratio of the reflected radiant flux to the irradiance and has a value between 0 and 1. Strictly, this is a monochromatic quantity, but in practice it is measured for some (small) wavelength intervals.

Albedo is a quantity very similar to reflectance. It is defined as the ratio between the reflected and the incoming energy. It differs from the reflectance in that it is usually taken over a wavelength range, instead of it being a monochromatic value. Commonly used terms are for example visible light albedo or short wave albedo.

Relative physical quantities

Additionally, there exist physical quantities that describe the reflectance under a specific geometry of both the incident and the observation rays [10, 37]. In Figure 2.7, some possible cases are visualised.

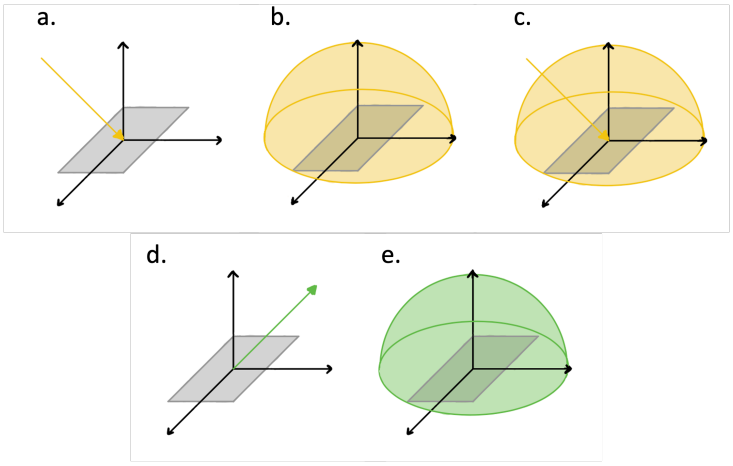

Figure 2.7: Different geometries of incoming (yellow) and reflected (green) radiance. ***a.*** *Directionally incoming flux.* ***b.*** *Diffuse incoming flux.* ***c.*** *Hemispherical incoming flux. A combination of directional and diffuse flux.* ***d.*** *Directionally reflected flux.* ***e.*** *Hemispherical reflected flux.*

The incident light (yellow in Figure 2.7) can be taken as either completely directional(2.7a), completely diffuse(2.7b), or a combination of this (2.7c). The reflected light (green) can either be measured in a specific direction (2.7d) or over the whole hemisphere (2.7e).

The term directional light is used to designate a parallel beam coming from (or going in) a single direction. From a technical standpoint, reflectance measurements will never be completely directional, since a non-zero area is required to do the detection. Similarly,

natural light sources are also not point sources. In these practical situations, directional radiance is replaced by a conical radiance [37]. For measuring the outgoing radiance, a viewing solid angle is specified instead of an infinitesimal one. This solid angle corresponds to the instantaneous FOV of the sensor[5], and should be considered in case it is large [10].

Next, the most important reflectances are described.

Hemispherical-directional reflectance

The hemispherical-directional reflectance (HDR) is the ratio of the reflected radiant flux in direction (θ_r, ϕ_r) to the incoming radiant flux. Here, the incoming flux is considered to be that of the whole hemisphere, both direct and diffuse (Figure 2.7c and 2.7e). The reflectances that are given by the Sentinel-2 satellite data are expressed as HDRs.

The definition of this value in terms of the BRDF is given by Nicodemus, Richmond and Hsia [39], who have introduced the conventions for the various reflectances and reflectance factors.

$$\text{HDR} = d\rho(\theta_i, \theta_r, 2\pi; \theta_r, \phi_r) = \frac{d\Omega_r}{\pi} \int f_r(\theta_i, \phi_i; \theta_r, \phi_r) d\Omega_i \tag{2.3}$$

$$= \frac{d\Omega_r}{\pi} \int_0^{2\pi} \int_0^{\frac{\pi}{2}} f_r(\theta_i, \phi_i; \theta_r, \phi_r) \cos\theta_i \sin\theta_i d\theta_i d\phi_i \tag{2.4}$$

Here $d\Omega$ is the projected solid angle increment, $d\Omega = \cos\theta \sin\theta d\theta d\phi$. For a hemispherical surface, $\Omega = \int d\Omega = \pi$, in units of steradian. Note that here the $\sin\theta_i$ comes from the Jacobian for spherical integration, while $\cos\theta_i$ is the same factor as was seen in Lambert's cosine law (Figure 2.6).

Directional-hemispherical reflectance

The directional-hemispherical reflectance (DHR) gives the ratio of reflected radiant flux over the hemisphere to the directionally incoming radiant flux. It is obtained by integrating the BRDF over all viewing angles.

$$\text{DHR} = \rho(\theta_i, \phi_i; 2\pi) = \frac{d\Phi_r(\theta_i, \phi_i; 2\pi)}{d\Phi_i(\theta_i, \phi_i)} \tag{2.5}$$

$$= \int_0^{2\pi} \int_0^{\pi/2} f_r(\theta_i, \phi_i; \theta_r, \phi_r) \sin\theta_r \cos\theta_r d\theta_r d\phi_r \tag{2.6}$$

The DHR describes the reflection for the situation when combining Figure 2.7a and 2.7e. It corresponds to the monochromatic black-sky albedo (BSA). BSA represents the reflection for the case without particles in the atmosphere. In this case, the incoming sunlight is not scattered and will just travel in a straight path. Looking up at the sky, every direction would be dark, except for the one direction where the light is coming from.

Diffuse hemispherical-hemispherical reflectance

For the diffuse bihemispherical reflectance (BHR_diff), the incoming light is considered to be completely diffuse, coming equally from every direction. The BHR_diff is the fraction of the diffuse incoming flux that is reflected over the hemisphere. It is obtained by integrating the DHR over all incident angles.

[5]The instantaneous FOV differs from the FOV of a satellite, in that it concerns the FOV of the individual sensors within the measuring instrument. It is also considered a measure for spatial resolution [38].

$$\text{BHR\_diff} = \rho(2\pi; 2\pi) = \frac{d\Phi_r(2\pi; 2\pi)}{d\Phi_i(2\pi)} \tag{2.7}$$

$$= \int_0^{2\pi} \int_0^{\pi/2} \rho(\theta_i, \phi_i; 2\pi) \sin\theta_i \cos\theta_i d\theta_i d\phi_i \tag{2.8}$$

This quantity describes the reflection for the situation when combining Figure 2.7b and 2.7e. It corresponds to the monochromatic white-sky albedo (WSA). WSA represents the case where all wavelengths of the incoming light are maximally scattered. From the ground it would appear that the light is coming equally from every direction, and the sky would look white.

2.2.4 Albedo

To approximate the realistic illumination, often a combination of BSA and WSA is taken [10]. The so-called monochromatic blue-sky albedo, or bihemispherical reflectance (BHR) can be expressed as

$$\text{BHR} = d\text{DHR}(\theta_s, \phi_s; 2\pi) + (1-d)\text{BHR\_diff}(2\pi; 2\pi) \tag{2.9}$$

Here, d is the fractional amount of the direct radiant flux and is a factor used to describe the atmospheric conditions [37].

$$d = \frac{I}{I+D} = \frac{I}{L_i}, \tag{2.10}$$

where I is the direct downwelling flux, D is the diffuse downwelling flux and L_i is the total downwelling flux. The downwelling flux is, in general, dependent on wavelength and sun angle, so d is also for a single wavelength and angle.

Alternatively, the BHR can be expressed in terms of the BRDF [39].

$$\text{BHR} = \rho(2\pi; 2\pi) = \frac{1}{\pi} \int \int f_r(\theta_i, \phi_i; \theta_r, \phi_r) d\Omega_r d\Omega_i \tag{2.11}$$

In this expression, we can use the definition for the HDR (equation 2.3) to express the BHR as an integral of the hemispherical-directional reflectance.

$$\rho(2\pi; 2\pi) = \int \left[\frac{1}{\pi} \int f_r(\theta_i, \phi_i; \theta_r, \phi_r) d\Omega_i \right] d\Omega_r \tag{2.12}$$

$$= \int d\rho(\theta_i, \phi_i, 2\pi; \theta_r, \phi_r) \tag{2.13}$$

This last expression shows that the bihemispherical reflectance can also be obtained from integrating the hemispherical-directional reflectance factor.

The above expressions for the BHR give the monochromatic reflectance. In practice, the albedo is desired for certain wavelength ranges. This albedo can be acquired by integrating over the wavelength spectrum, weighted by the incoming radiation flux [10].

$$A(\theta_s, \Lambda) = \frac{\int_\Lambda L_i(\theta_s, \lambda)\alpha(\theta_s, \lambda) d\lambda}{\int_\Lambda L_i(\theta_s, \lambda) d\lambda} \tag{2.14}$$

Here $A(\theta_s, \Lambda)$ is the albedo for a wavelength range Λ, a sun zenith angle θ_s. $L_i(\theta_s, \lambda)$ is the incident radiant flux under sun angle θ_s and of wavelength λ and $\alpha(\theta_s, \lambda)$ is the corresponding monochromatic albedo. For blue-sky albedo, $\alpha = \text{BHR}$.

When considering the albedo over a short range of wavelengths, such as the reflectance measured by a single sensor band in a satellite, we speak of narrowband albedo. If a larger wavelength range is considered, such as the SWIR or the whole solar spectrum, we speak of broadband albedo.

2.2.5 FRT model

In this thesis, the reflectance of the vegetation is modelled with the Forest Reflectance and Transmittance (FRT) model. This model was developed by Andres Kuusk and Tiit Nilson in 2000. [12, 13]. It includes several steps to describe the complex process of light absorption by, reflection on and transmission through vegetation. The model has been tested through a comparison with empirical data [13, 40]. It has also previously been used in research comparing modelled and observed reflectance data [41, 42].

Model components

The FRT model includes a geometrical model of the trees in different vegetation layers. The tree crowns are described as either ellipsoids, cones or a combination of cones and cylinders. The tree crowns have a uniform distribution of branches and leaves. The ground vegetation consists of a homogeneous layer.

To obtain the reflectance and transmission coefficients of the leafs, the PROSPECT model by Jacquemoud and Baret [43] or the LIBERTY model by Dawson et al. [44] can be used. These are both radiative transfer models for the optical properties of leaves. The absorption of a leaf depends on its chemical composition, such as chlorophyll concentration, and its internal structure.

After entering the atmosphere, the incoming light is usually not reflected once on a leaf and then sent outwards in a straight trajectory. Instead, the light is often scattered, transmitted and reflected multiple times by particles in the atmosphere and the multiple layers in the vegetation. Both the first order, single scattering, and the multiple scattering are considered in the FRT model.

The FRT model includes radiative transfer modeling. The 6S atmospheric radiative transfer model by Vermote et al. [45] is used to calculate the incident diffuse and direct fluxes I and Q. Parameters used by the 6S model to calculate the diffuse part of the radiation include the percentage of aerosol components and horizontal visibility.

FRT model output

To run the FRT model, an input file is required in which specific parameters about the different vegetation layers in the tree stand are specified. Additionally, parameters about the atmospheric conditions, the solar- and view angles and the wavelength are included in this file. An example input file is given in Appendix B.1. For more specifics on the FRT model, the reader is referred to the FRT User Guide [46].

The output of the model is the hemispherical-directional reflectance (HDR), where the incoming light is a combination of direct and diffuse sunlight. The output is dependent on the zenith angles of the sun and the satellite, θ_s and θ_v, as well as the azimuthal angle

between the two, ϕ_r, and the wavelength λ. The range for which the model runs is limited to wavelengths of 400 nm to 2400 nm.

The reflectance that is given by the model is calculated at the top of the canopy (TOC) of the vegetation. No atmospheric corrections have been applied to the outgoing radiant fluxes. The results are thus equivalent to the bottom of atmosphere bihemispherical reflectance (BHR).

2.3 Estimation of albedo

The estimation of surface albedo from satellite measurements can roughly be done in two different ways [14]. The reflectances obtained from the satellite sensors are directional and narrowband, and need to be converted to a hemispherical and broadband reflectance.

The first way to do this is in steps, following the diagram in Figure 2.8. If the data obtained from the satellite is the TOA reflectance, it is first converted to the surface reflectance by applying atmospheric corrections. The data at that point is still only in the view direction and needs to be converted to a hemispherical reflectance in the next step. In the following, this step will be referred to as BRDF (angular) modelling or directional-to-hemispherical conversion. Finally, the broadband albedo is obtained from the collection of narrowband albedos. This stepwise processing method has been the traditional way of obtaining the albedo [14].

Alternatively, the conversion can also be done by a so-called direct estimation model. Here, the BRDF modelling and the narrow-to-broadband conversion are combined in one step. The BOA data is directly used to estimate the broadband albedo. The advantage of this type of model compared to the step by step processing is that the errors of the different models are not accumulated over the steps [47]. Examples of this type of model are the one for MODIS developed by Liang, Strahler and Walthall [47, 48], and the one for Landsat presented by Zhang et al. [49].

Both the models developed to perform the individual conversion steps and the direct-estimation models are usually dependent on the satellite system. One of the reasons for this is that various sensors measure the TOA reflectance in a different field of view. Similarly, the spectral bands vary between satellite systems.

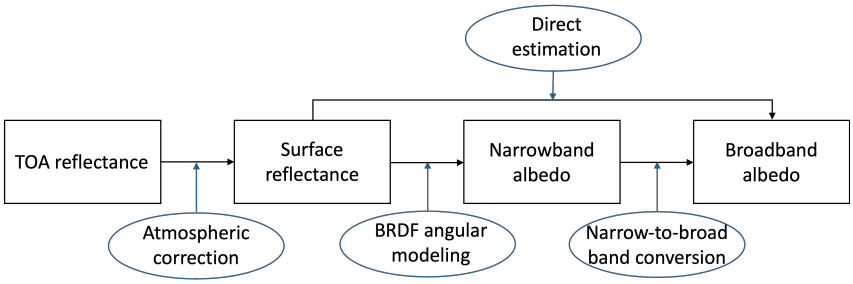

Figure 2.8: Processing of satellite measurement at the top of the atmosphere (TOA) to obtain broadband albedo.

3

Methods

The goal of this research is to create a machine learning model that can be used to predict surface albedo for a specific type of vegetation. The model takes the bottom of the atmosphere (BOA) Sentinel-2 reflectance data for several MSI sensor bands as input. The model is trained for two different vegetation types and these cases are compared to each other. The work needed to achieve this can be split in three parts:

1. Analysing the reflectance of a vegetation stand for different angles and wavelengths.
2. Calculating the surface albedo and simulated satellite data to use as training data.
3. Training and validation of the regression model.

The linear regression model takes the reflectance coefficients for every MSI sensor band as input parameters. These reflectances are measured under one specific view zenith angle θ_v and as such give the reflectance coefficient for one specific angle, and a narrow wavelength band. The desired output is the total reflectance at the Earth's surface.

3.1 Scope and assumptions

In developing the model, some assumptions and simplifications have to be made. The model is developed with a focus on the area of south Sweden with regard to the satellite specifications. Additionally, only two specific types of vegetation are considered. Though these simplifications are necessary to keep the research manageable, the reader should keep them in mind.

The first simplification is that the regression model is trained specifically for one solar zenith angle, $\theta_s = 59°$ This angle was chosen because throughout the year, it is the average solar angle at 10:30 in the morning, in south Sweden. The time 10:30 was chosen, as this is the approximate time Sentinel-2 passes over this area.

The relative azimuthal angle was fixed at $\phi_r = 65°$ for similar reasons. Inspection of Sentinel-2 data for the south of Sweden showed that a typical ϕ_r is somewhere in the range of 50-70°. The reason for ϕ_r not being close to zero is the interference of the hotspot and the satellite shadow in the principal plane [10].

Only one specific case of atmospheric conditions is considered in the simulations. In this case, the sky conditions are clear, with a high visibility. Such conditions are desirable for accurate surface reflectance measurements. The input parameters for the 6S atmospheric radiative transfer modelling are given in Appendix B.2.

3.2 Calculating reflectances

The backbone of this research project is the FRT model introduced in the previous section. It is a Fortran computer program. We wrote a program in Python 3.7 to do the data analysis and model training. A function supplies the FRT model with an input file and executes it.

3.2.1 FRT model

In the FRT input file, all information about the vegetation stand and the atmospheric conditions is supplied. For the purpose of the initial analysis, these parameters were kept according to the example file from the FRT user guide. This file describes the Järvelsja Pine stand in Estonia, and the parameters were validated through empirical data [13]. The input file can be found in Appendix B.1. The input file also contains information on the geometry in the form of θ_s, ϕ_r and θ_v. The last input parameter is the wavelength for which the reflectance factor will be calculated. This output is the hemispherical-directional reflectance (HDR) and is also denoted as $d\rho(\theta_s, \phi_r, \theta_v, \lambda)$.

The model is able to take wavelengths in the 400 nm - 2400 nm range. In Figure 3.1 the ASTM Standard Extraterrestrial Spectrum Reference is plotted. This is a zero airmass solar reference spectrum developed in the year 2000 by the National Renewable Energy Laboratory [50]. It gives the TOA incoming solar power, per area, per wavelength unit and is given in units of $W/m^2/\mu m$.

The FRT wavelength range corresponds to the shaded area in Figure 3.1. Applying a trapezoidal numeric integration on the ASTM data tells us that the FRT wavelength range covers 89% of the incoming TOA solar irradiance. For the transmitted and reflected spectrum, this percentage is somewhat different. Ozone, for instance, has a high absorbance in the ultraviolet part of the spectrum (220-320 nm) [10] and decreases the irradiance that actually reaches the surface of the Earth. According to Francois et al. [51], the leaf reflectance is negligible for wavelengths over 2500 nm.

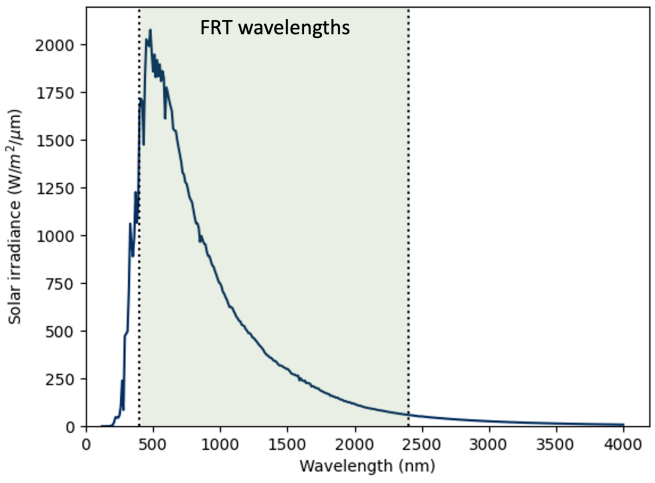

Figure 3.1: The top of atmosphere (TOA) solar irradiance spectrum. The shaded area indicates the wavelengths included in the FRT model. Data: ASTM Standard Extraterrestrial Spectrum Reference [50]

Another limitation in the FRT model appears for the high view zenith angles. Any θ_v below 90° is allowed as input for the model, but angles close to 90° are not recommended

[46]. The model will be less accurate for the higher zenith angles, though the magnitude of the inaccuracies is unknown.

3.2.2 Angular dependence

We first use the FRT model to look at the reflectance coefficient in different view directions. To evaluate the angular dependence of the reflectance, the model was run for a wavelength of 830 nm. This wavelength corresponds to the central wavelength of MSI band 8, which detects the surface reflectance in the NIR part of the wavelength spectrum [52]. While the solar zenith angle and the relative azimuthal angle are fixed, the reflectance is calculated for the view zenith angle of $-90° < \theta_v < 90°$. Two different solar zenith angles $\theta_s = 37°, 59°$, as well as 4 different relative azimuthal angles, are evaluated to study the angular dependence of the HDR.

3.2.3 Wavelength dependence

Next, the FRT model is used to calculate the HDR for all view angles ($-90° < \theta_v < 90°$, $0° \leq \phi_r \leq 180°$) and wavelengths (400nm $\leq \lambda \leq$ 2400nm). The bi-hemispherical reflectance (BHR) is calculated by integrating the HDR over all angles in the hemisphere, as per equation 2.13:

$$\alpha(\theta_s, \lambda) = \text{BHR} = \int d\rho(\theta_i, \phi_i, 2\pi; \theta_r, \phi_r)$$

For the whole wavelength spectrum, the reflectance factor is integrated in this manner. The result is the reflectance spectrum of the pine forest. This spectrum is calculated for $\theta_s = 37°, 59°$.

3.2.4 Broadband hemispherical albedo

From the reflectance spectrum, the broadband hemispherical albedo can be calculated, using equation 2.14:

$$A(\theta_s, \Lambda) = \frac{\int_\Lambda L_i(\theta_s, \lambda)\alpha(\theta_s, \lambda)d\lambda}{\int_\Lambda L_i(\theta_s, \lambda)d\lambda}.$$

The wavelength range we integrate over is 400-2400 nm, and the irradiance L_i used is the Standard Extraterrestial Spectrum Reference in Figure 3.1. The result is the simulated surface albedo for solar angle θ_s and the simulated vegetation stand.

3.2.5 Integration steps

The integrals mentioned in the previous sections are performed numerically, with use of the trapezoidal rule. The impact of the step size on the albedo estimation was evaluated for the different integrals. We do this in order to pick a balanced step size, with the purpose of keeping computing time and data storage manageable, without impacting the precision of the results.

Here, the integration over the zenith angle is from 0° to 90° with an integration step of 1°. This step size was also decreased to 0.5°, and the difference was evaluated. An increment of 1° gave a hemispherical broadband albedo of $A = 0.01511496$, an increment

of 0.5° gave $A = 0.01511440$, which is a relative difference of 10^{-5}. Considering all other inaccuracies, we may conclude from this that taking steps of 1° is sufficient.

For the integration over the azimuth angle, steps of 5° were used to integrate from 0° tot 360°. Taking steps of 5° instead of 1° was evaluated to impact the total reflectance by 0.03% in the case of the Järvelsja Pine stand with a tree density of 1.0 $trees/m^2$. The numeric integration over the wavelength spectrum is done from 400 nm to 2400 nm in steps of 20 nm. Increasing the wavelength step size from 10 nm to 20 nm on top of increasing the angular step to 5° impacts the final results by 0.5%. The steps of 5° and 20 nm were deemed sufficient, considering that other inaccuracies, like the FRT model performance at high view zenith angles, would outweigh these errors of less than 1%.

Figure 3.2 shows a comparison of the wavelength spectrum for some considered step sizes. The graphs in this figure are completely overlapping, indicating that the results are virtually the same for the step sizes considered.

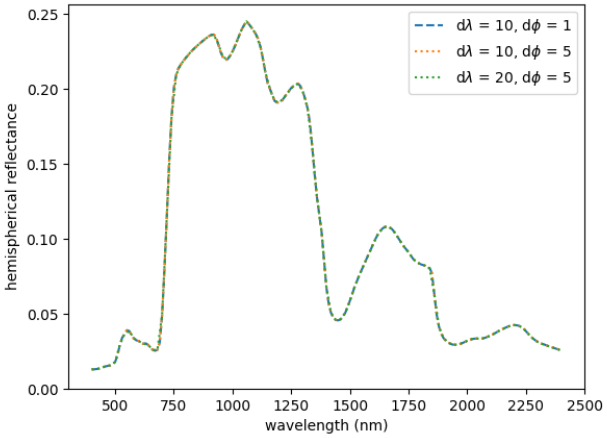

Figure 3.2: Reflectance spectrum with different integration steps for the Järvelsja Pine with tree density = 1.0 trees/m^2.

3.3 Lambertian surface approximation

The measurement from the satellite is a hemispherical-directional albedo. To get the hemispherical-hemispherical albedo, one ideally wants to integrate over all possible view angles, as is described in the previous section. This is not possible for estimating the albedo from satellite data, because the satellite measures the reflectance only for one view angle per time. For Sentinel-2 data, this angle is within the narrow 22.6° FOV.

To get an estimate for the hemispherical-hemispherical albedo at each wavelength, the Lambertian approximation is used. In this approximation, the HDR is assumed to be equal in every direction. In this case, the bihemispherical reflectance is simply the HDR in the view direction.

The reflectance spectra obtained with and without the Lambertian approximation are compared, to evaluate the impact of making this assumption. Additionally, the broadband albedo obtained for these cases by integrating the spectra is calculated. The Lambertian approximation is evaluated for five different θ_v, to span the range of Sentinel-2's FOV. This process is performed for both $\theta_s = 37°$ and 59°, to evaluate validity of the Lambertian approximation for different solar angles.

3.4 Albedo estimation from Sentinel-2

The method for estimating the albedo from Sentinel-2 data is through setting up and training a linear regression model to do direct estimation from the bottom of atmosphere (BOA) satellite data. The regression model takes the equation

$$A = \sum_{B=1}^{N} w_B \rho_B(\Lambda_B) + w_0. \tag{3.1}$$

The broadband albedo A is calculated by taking the sum over the selected sensor bands of the MSI instrument. The narrowband reflectance for every band is calculated over the wavelength range Λ_B. w_B are the regression coefficients applied to every band after training of the model. The constant w_0 is trained similarly.

The wavelength bands considered in the model are bands 2, 3, 4 and 8, 11 and 12. The bands are illustrated in Figure 2.1 and the wavelengths are given in Appendix A. The first four are the 10-meter spatial resolution bands located in the VNIR. Bands 11 and 12 (20-meter resolution) are added to describe the SWIR part of the spectrum.

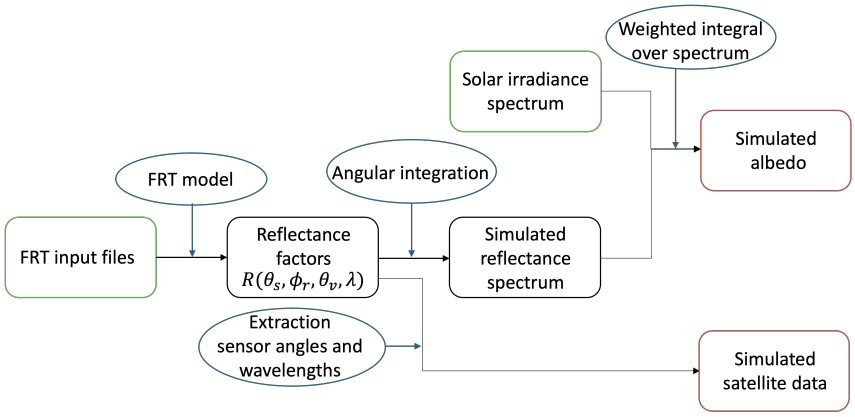

Figure 3.3: Process diagram of creating the training data for the regression model.

3.4.1 Training data

For training the regression model, a training dataset is needed. This training dataset has to consist of input- and output data for the regression. The input data is a set of narrowband reflectances that are measured by the individual Sentinel-2 detectors. The output data is the corresponding hemispherical broadband albedo.

Many data points are needed to train the regression model. The process of creating the training data for one data point is described in Figure 3.3. First, the FRT model is run to obtain the BOA reflectance factor for every view angle and wavelength, with the step sizes indicated before. Throughout all model runs, θ_s is fixed to 59°.

The level-2A data product from Sentinel-2 is the reflectance factor at the bottom of the atmosphere, for angles (θ_v, ϕ_v). These angles and the solar angles are specified in the metadata, so that all relevant angles can be determined. No actual Sentinel-2 data was used in training the model. We are working with a theoretical model to estimate the surface albedo, and the input data used to train the model is the simulated satellite data.

To get this simulated satellite data, the reflectance factor for the view angle of the satellite is extracted from the total set. The relative azimuthal angle was continuously taken to be $\phi_r = 65°$. The average reflectance was determined for the wavelengths within the sensor band.

The hemispherical albedo is calculated through the same process as described before. The reflectance factor is integrated over the hemisphere and subsequently over the wavelength spectrum, weighted by the solar irradiance.

Varying input parameters

The different data points in the training data are obtained by running the FRT model for a set of different conditions. No general rule for the required amount of training data to train regression models exists. However, a lower bound, specifically for linear regression models, seems to be 10 times the amount of trainable coefficients [53]. Often more data than this is needed in practical situations. As the model has 7 weights, it requires upwards of 70 data points. In order to get this training data, the FRT model is run for different 'cases'.

These cases are created by changing some parameters in the FRT input file. The input file contains many parameters describing the physical and chemical properties of the trees in the stand. A more detailed description of the parameter selection is given in Appendix C.

The four chosen parameters are varied more closely around the original values in the input file for the Järvelsja Pine stand. This is done in order to stay within a scenario of a physical forest stand, whilst creating enough variation in the reflectance to train the regression model. The ranges of the varied parameters are shown in Table 3.1.

Table 3.1: Variation of input parameters within training data for pine forest

Parameter	Minimum	Maximum
Stand density ($trees/m^2$)	0.6	1.2
Crown radius (m)	0.5	3
Leaf weight per area (g/m^2)	140	170
Chlorophyll per leaf weight (%)	0.4	0.7
View zenith angle (°)	-11	11

In addition to varying the FRT input files, a second way to obtain more training data is used. To reflect that the satellite measurements happen at any view angle within the Sentinel-2 FOV, θ_v is varied over five values within this range. Since the hemispherical albedo is not dependent on view angle, the output data does not change for these different angles. However, the reflectance measured by the MSI does depend on view angle, and as such the input data is different. The result is five data points for FRT input file that have different simulated satellite data, but the same simulated albedo.

Eventually, all cases are given a number ID and the model is run to compute the simulated satellite data and the simulated albedo for every case. For the pine forest stand, a total number of 720 cases are evaluated.

3.4.2 Regression model

The actual regression training is done following the steps in the diagram in Figure 3.4. The training data is merged so that the simulated satellite data and albedo for the same case are linked. Then, the data is split at an approximate 80/20 ratio. 80% of the data is used to train the model, while 20% remains to test the model.

As a future step, the regression coefficients obtained could then be used to obtain the surface albedo from the Sentinel-2 BOA reflectances. Since this has not yet been done in this work, this part appears dashed in Figure 3.4.

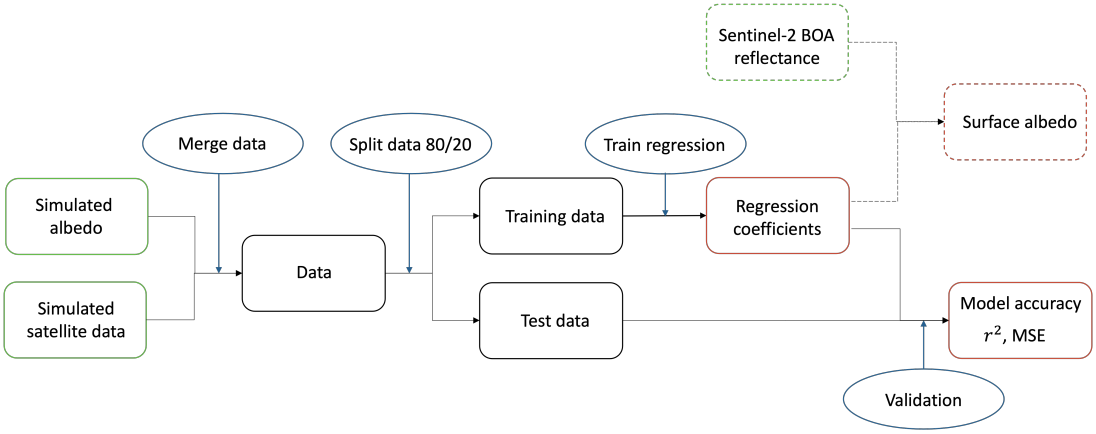

Figure 3.4: Process of training the regression model. The dashed part of the schema indicates future work.

Validation

The testing of the model is performed using a 5-fold randomized cross-validation. 20% of the data is randomly selected and saved for testing, while the remaining data is used to train the model. The coefficient of determination r^2 and the mean squared error (MSE) are determined for this split. This is then done for a total of 5 different random splits, and the mean and standard deviation of the r^2 and MSE values are given[1].

Narrow-to-broadband model

After developing the direct estimation model, it is compared to a narrow-to-broadband model. This narrow-to-broadband model omits the BRDF angular modelling step in the albedo estimation (Figure 2.8). The model is also a linear regression model trained and tested similarly to the direct-estimation model. The difference is that the input parameters are not the directional reflectances obtained from satellite measurements, but instead the simulated bihemispherical reflectance. The regression coefficients from this model represent the narrow-to-broadband conversion.

If the Lambertian approximation is a reasonable assumption to make, angular modelling is not necessary. In this case, it is sufficient to use the narrow-to-broadband model on the satellite data directly. However, in this thesis, the narrow-to-broadband model is trained and tested solely on bihemispherical reflectance data. Still, if the two models are similar, this would indicate the Lambertian approximation to be valid.

3.4.3 Comparison to literature

The model is also compared to two models for Sentinel-2 albedo estimations from literature. The two sets of regression coefficients obtained from research by Bonafoni [22] and Li et al. [20] are for the narrow-to-broadband conversion specifically.

The parameters obtained from Bonafoni are obtained by a theoretically based weighting of the sensor bands. The same sensor bands are used by Bonafoni and in this thesis. The solar spectrum is divided up in wavelength ranges surrounding the MSI bands. The bands are weighted based on the solar irradiance within their wavelength range. Applying this model to Sentinel-2 data requires the assumption of a Lambertian surface.

[1]For more information about cross-validation, the reader is referred to *Pattern Recognition and Machine Learning* by Bishop [54].

Li et al. presents a trained regression model based on radiative transfer simulations and spectral reflectances for a whole range of different surfaces. Distinction of surface type is only made between snow-covered and snow-free surface. This paper uses sensor band 8A instead of sensor band 8, but otherwise the same MSI bands are used.

3.5 Comparison of vegetation

Finally, a second regression model is trained for a different vegetation type. The results for the pine forest stand are compared to a vegetation with very little trees. This 'field' has a density in the upper layer of 0.0001 trees per square meter. To create the training data set, the Leaf Area Index (LAI), the leaf weight per area and the chlorophyll concentration in the ground vegetation are varied. Table 3.2 gives the bounds of these parameters in the training data. Including the varying θ_v, 235 data points are obtained to train the regression model for the field stand.

Table 3.2: Variation of input parameters within training data for field

Parameter	Minimum	Maximum
Leaf Area Index	0.1	0.7
Leaf weight per area (g/m^2)	70	100
Chlorophyll per leaf weight (%)	0.3	0.5
Zenith angle (°)	-11	11

For the field stand, all the same steps are executed as for the pine stand. Finally, the regression models trained for the pine forest and the field are applied to a data set of the other type. From this, different characteristics of the two models can be evaluated.

4

Results

4.1 FRT model results

We are interested in the nature of the output of the FRT model and its dependence on several parameters. In this section, the reflectance is considered for different angles and wavelengths. In all these cases, the reflectance is evaluated using the FRT model with the Järvelsja Pine stand input file.

4.1.1 Angular dependence

The first important part of the geometry that influences the reflectance, is the view zenith angle θ_v. To evaluate its influence, the reflectance was plotted as a function of θ_v in Figure 4.1. This was done for a solar zenith angle of $\theta_s = 37°$, as well as $\theta_s = 59°$. The wavelength is set to $\lambda = 830$ nm. Also, the simulations are run in the principle solar plane, so $\phi_r = 0°$.

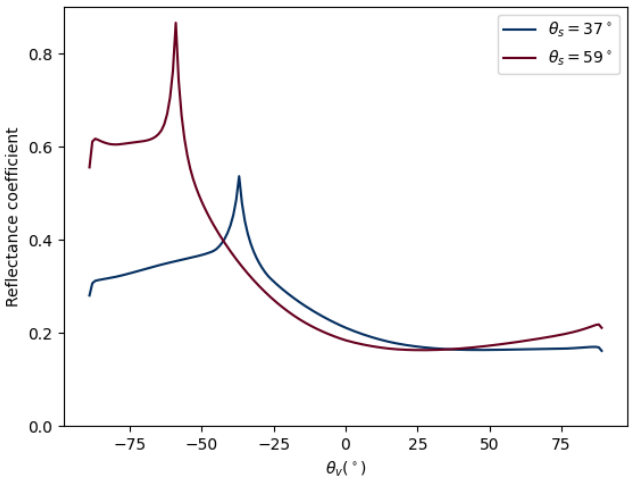

Figure 4.1: The change in the reflectance coefficient for different view zenith angles in the solar principal plane. The angular spectrum is shown for solar zenith angles of 37° and 59°.

In the graph we see that the reflectance has a peak at the zenith angle opposite to the solar zenith angle. This is the location of the hotspot. Towards a zenith angle of 0°, the reflectance slopes down quickly. If θ_v is on the same side of the surface as the sun, the reflectance remains low. At view angles close to 90° or −90°, we see a kink, with the reflectance dropping quickly as the detector approaches the surface.

Comparing the two different θ_s, we see that besides the displacement of the peak to $-\theta_s$, the peak height also differs. For $\theta_s = 59°$, the maximum reflectance is higher. This graph also has a higher reflectance on average, indicating that the reflectance is higher if the sun is lower on the horizon.

The relative azimuthal angle between the sun and the satellite also influences the reflectance. Figure 4.2 shows the reflectance for different relative azimuthal angles ϕ_r. Here, $\theta_s = 59°$ and $\lambda = 830$ nm.

We see that in the principal solar plane ($\phi_r = 0°$), the reflectance is highest. Also, the hotspot is clearly recognisable here. For higher relative angles, the overall reflectance decreases and the hotspot becomes less recognisable. For $\phi_r = 90°$, the shape is symmetrical, and the lowest reflectance occurs for $\theta_v = 0°$.

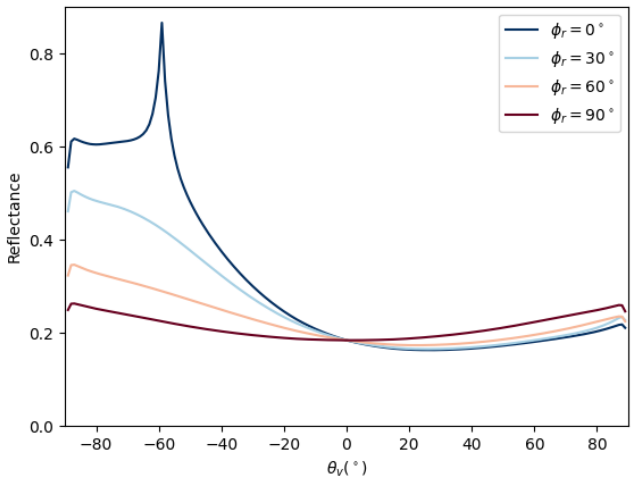

Figure 4.2: The change in the reflectance coefficient for different view zenith angles, with $\theta_s = 59°$. The angular spectrum is shown for four relative azimuthal angles.

4.1.2 Wavelength dependence

The bihemispherical reflectance (BHR) is obtained by integrating over the view angles. Figure 4.3 shows the bihemispherical reflectance spectrum of the Järvelsja Pine stand, with $\theta_s = 59°$. The purple line gives the reflectance obtained from the FRT data.

At the low wavelength end of the spectrum, starting with violet light, there is a low reflectance. Within the visible light spectrum (380-750 nm), the graph has a small peak around the green-yellow wavelengths (500-600 nm) [55]. At the edge of the visible spectrum and the Near Infrared (NIR), the reflectance increases rapidly. Aside from two dips, the reflectance stays high until a wavelength of around 1400 nm. There is another increase in reflectance between 1500 nm and 1900 nm.

In the same graph, the reflected radiance is plotted. This is a multiplication of the reflectance and the incoming solar irradiance, and indicates how much each wavelength contributes to the total reflected energy.

Comparing the reflected radiance with the reflectance, it can be seen that the impact of the visible light is increased. Also, the near-infrared radiation with the shortest wavelengths is very high. The reflected radiance decreases for longer wavelengths. At the high wavelength edge of the calculated wavelength spectrum, the reflected radiance approaches zero. However, it is noticeable that at 400 nm the radiance is not yet near zero.

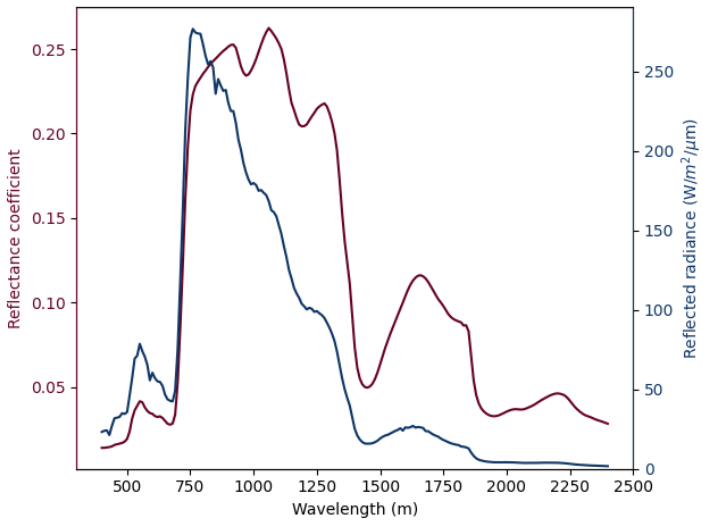

Figure 4.3: ***Purple line, left axis:*** *The reflectance coefficient for different wavelengths in the solar spectrum.* ***Blue line, right axis:*** *The reflected radiance, calculated using the ASTM Standard Extraterrestrial Spectrum Reference [50].*

4.1.3 Satellite measurements

The satellite data at level 2A contains BOA reflectances for all its sensor bands separately [19]. These are measured under a view zenith angle in the field of view of $-11.3°$ to $11.3°$. The satellite reflectance (HDR) measurements at nadir, $\theta_v = 0°$ are simulated from the FRT data. For each sensor band, the simulated measurements are shown as bars in the spectrum in Figure 4.4a. The bihemispherical reflectance spectrum for this case is also plotted.

In Figure 4.4b, the narrowband albedo is shown for the satellite wavelength bands. Here it is calculated from the bihemispherical reflectance (BHR) in the respective bands, instead of from the HDR at nadir. Comparing the two figures shows that the simulated satellite reflectance measurements at nadir are lower than the actual narrowband albedo for the sensor bands.

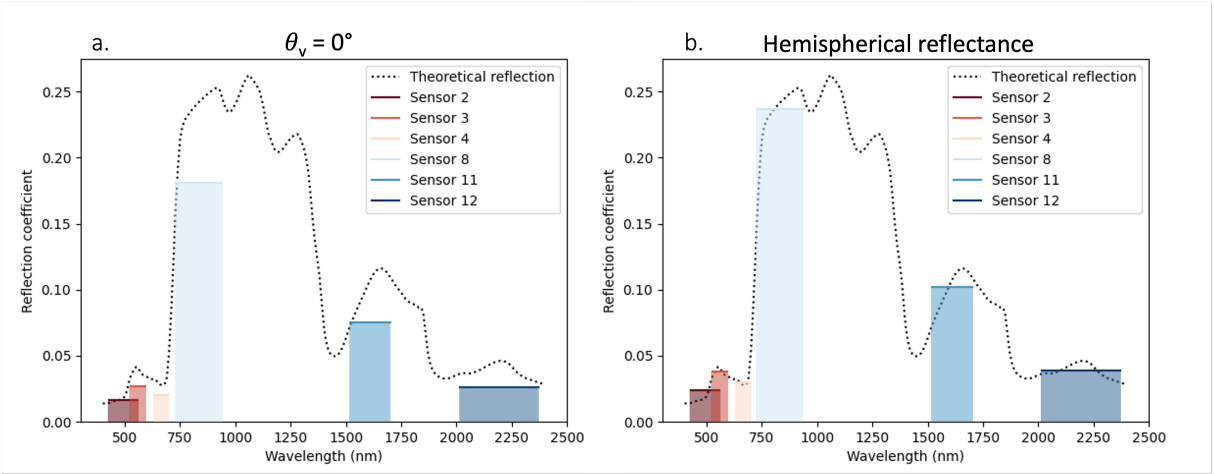

Figure 4.4: ***a.*** *Comparison of the theoretical reflectance spectrum and the estimated satellite measurements. The theoretical reflectance uses integration over the hemisphere. The satellite measurements assume a nadir view zenith angle.* ***b.*** *The sensor measurements if they would measure the hemispherical reflectance.*

4.2 Lambertian approximation

To approximate the bihemispherical reflectance from the satellite measurements, it is possible to use the Lambertian approximation, where the reflectance is equal in all directions. Here we evaluate the validity of this method.

The view angle can vary anywhere between $\pm 11.3°$, which is a narrow, near-nadir range. A selection of 5 angles within this range is made. For every case, the reflectance is evaluated over the wavelength spectrum and plotted in Figure 4.5. In this figure, also the theoretical reflection coefficient, obtained from integrating over all view angles on the hemisphere, is plotted. This is done for both $\theta_s = 37°$ and $\theta_s = 59°$ and can be found in Figures 4.5a and 4.5b, respectively.

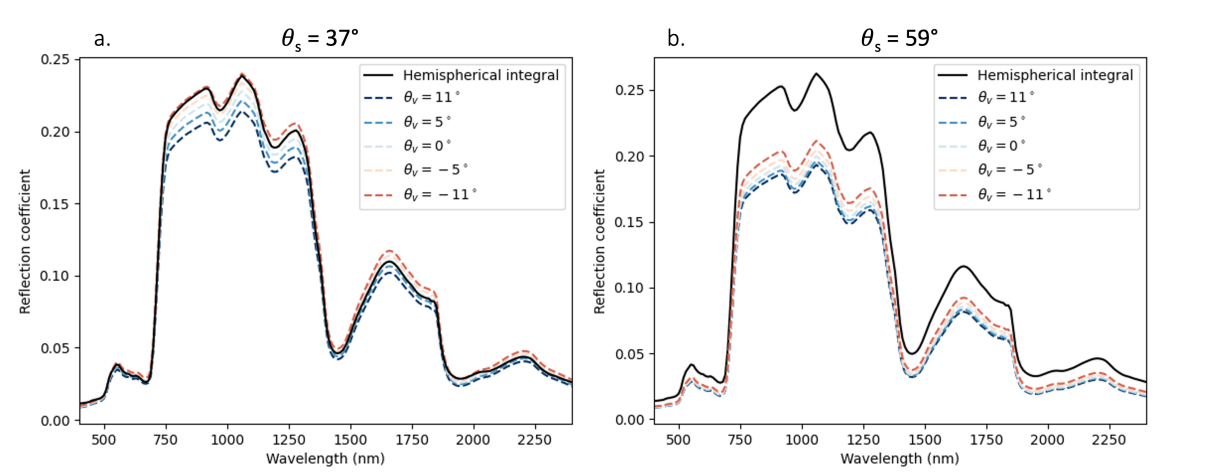

Figure 4.5: Comparison of reflectance over the solar wavelength spectrum between hemispherical integration of FRT results and taking the Lambertian approximation for a solar zenith angle of ***a.*** *$\theta_s = 37°$ and* ***b.*** *$\theta_s = 59°$. The view zenith angle is varied between -11° and 11°*

The first plot shows that under the Lambertian approximation, the spectrum keeps roughly the same shape. For $\theta_s = 37°$, the approximation at higher view zenith angles, $\theta_v = 5°$ and $\theta_v = 11°$ seems fairly accurate. For lower θ_v, the approximation gives gradually lower reflection coefficients. It also shows that between wavelengths of 750 nm and 1350 nm, the Lambertian approximation estimates the reflectance to be lower than the hemispherical reflectance for almost all θ_v. For the lowest wavelengths, below 500 nm, this also seems to be the case. However, for the other wavelengths, this shift is not so clear. Here, the hemispherical spectrum seems to lie in the middle of the set of Lambertian spectra.

Comparingly, Figure 4.5b shows a much bigger difference between the simulated hemispherical reflectance and the approximated ones. All the spectra from the Lambertian approximations lie under the hemispherical spectrum, over the whole wavelength range. The difference is also larger. The theoretical albedo is higher for $\theta_s = 59°$ than for $\theta_s = 37°$. In contrast, the reflectances based on the Lambertian approximation are lower for the θ_s.

To describe how far the Lambertian approximations deviate from the simulated hemispherical albedo, the relative error is calculated for $\theta_s = 37°$ and $\theta_s = 59°$. The differences are presented in respectively Figure 4.6a and Figure4.6b. Here the relative error is understood to be

$$\Delta = \frac{\rho_{sim} - \rho_{Lam}}{\rho_{sim}} * 100\%. \tag{4.1}$$

The error for the lower solar zenith angle is up to 20% of the bihemispherical re-

flectance. For $\theta_s = 59°$, the error is higher. When taking $\theta_v = -11°$, the error varies around the 25%, while $\theta_v = 11°$ has an error up to 40%. In every graph we see that the relative error is lower for the parts of the wavelength spectrum where the reflectance is high. The absolute error in this part of the spectrum is still larger.

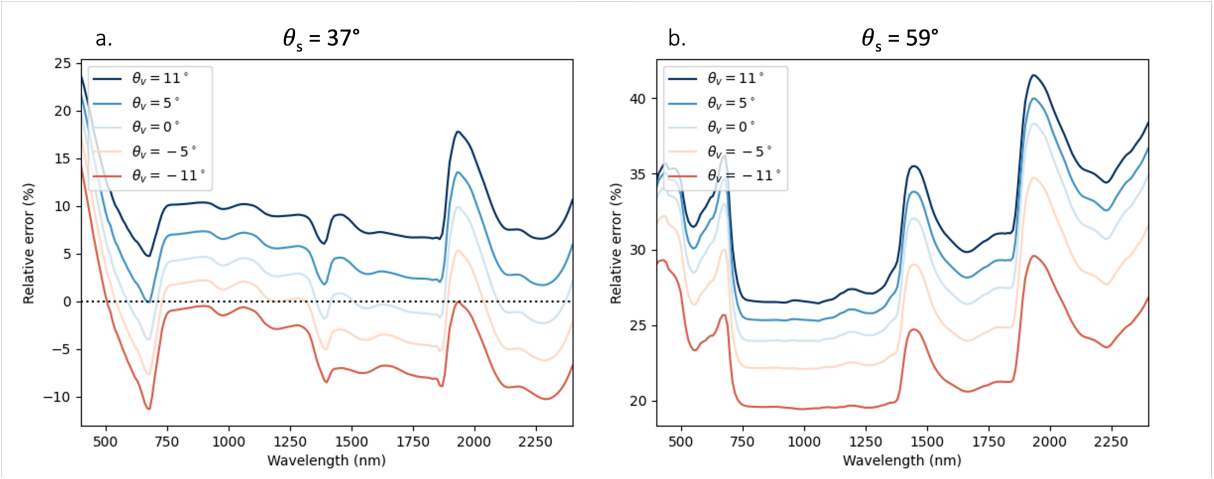

Figure 4.6: The relative error of the Lambertian approximation for a solar zenith angle of ***a.*** $\theta_s = 37°$ *and* ***b.*** $\theta_s = 59°$*. The view zenith angle is varied between -11° and 11°.*

The relative errors shown in Figure 4.6 give the inaccuracy of the Lambertian approximation, at each wavelength within the spectrum. To find the impact on the surface albedo, the broadband albedo is computed both with and without taking the Lambertian approximation.

The simulated broadband albedo for $\theta_s = 37°$ is $A = 0.1050$. For $\theta_s = 59°$, the surface albedo is $A = 0.1145$. We notice a slightly higher reflectance for the lower sun position.

In Table 4.1, the percentage error in the broadband hemispherical albedo is given for the different Lambertian approximations. It shows that for $\theta_s = 37°$, taking the Lambertian approximation gives a 0-10% error in the albedo estimate. For $\theta_s = 59°$, the error lies in the range of 20-30%. It must be noted that in all cases, except for $\theta_s = 37°$ and $\theta_v = -11°$, the Lambertian approximation results in an albedo estimate lower than the simulated value.

Table 4.1: Error percentage in the hemispherical broadband albedo for different Lambertian approximations.

θ_v	$-11°$	$-5°$	$0°$	$5°$	$11°$
Error (%) for $\theta_s = 37°$	2.3	0.6	3.3	6.2	9.6
Error (%) for $\theta_s = 59°$	20.4	23.3	25.3	26.7	28.0

4.3 Linear regression models

4.3.1 Direct estimation

The surface albedo is estimated with a linear regression model. This regression model obeys equation 3.1

$$A = \sum_{B=1}^{N} w_B \rho_B(\Lambda_B) + w_0.$$

The direct estimation model is trained with the FRT data from the pine forest stands. A weight is obtained for the reflectance in every wavelength band, as well as a constant

w_0. The set of regression coefficients obtained for this pine forest model can be found in the w_p row in Table 4.2.

It must be noted that the data was split into training and test data randomly. Different splits of training and test data give somewhat varying weights. All the tables and figures with regression results contain the weights and plots obtained with an arbitrary selection of training and test data. Appendix D shows some examples of coefficients obtained for different training data selection.

The coefficients w_p vary a lot in size. The weights for wavelength band 2, 3 and 4 are large, while band 8 has a relatively small coefficient. It is also noteworthy that several coefficients are negative.

Table 4.2: The direct estimation regression coefficients for the different wavelength bands. The weights for a pine forest stand (w_p) and a field are given (w_f).

	Band 2	Band 3	Band 4	Band 8	Band 11	Band 12	Constant
w_p	72.3105	-29.2303	-13.5841	0.3903	2.3373	-5.1862	-0.1086
w_f	0.2168	0.2455	0.2285	0.3169	-0.3479	0.9778	0.0686

To investigate how good the model is in predicting the albedo, the regression predictions are compared to the simulated surface albedo for the test dataset. This performance analysis method is based on the MODIS model analysis performed by Liang, Strahler and Walthall [48]. The predicted albedo is obtained by applying the regression weights to narrowband reflectances of each sensor in the test data. The simulated albedo is calculated from the FRT simulations and the full integrals over the hemisphere and wavelength spectrum.

Figure 4.7a shows how the predicted albedo compares to the simulated albedo. The line $y = x$ is plotted to show the ideal predictions. A linear fit with a slope of 0.956 is plotted to show the trend in the predictions. The graph also shows a larger spread in the predicted values for the test data with higher simulated albedo.

Again note that the plot and values are for a single split of training and test data. Calculating the same model with different training and test sets, the graph also shows somewhat different behaviour.

A behaviour that is recurring throughout the figures, is that we see a stack of vertical data points with the same simulated albedo. Even though their simulated albedo is exactly the same, the predicted albedo varies, especially at the higher simulated albedos. These different predicted albedos correspond to the different θ_v in the training data.

Using cross-validation, the coefficient of determination r^2 is calculated for different splits of training and test data. The mean r^2 obtained in this way is $r^2 = 0.946$ with a standard deviation of $\sigma_{r^2} = 0.00456$. Similarly, the average Mean Squared Error (MSE) in the albedo predictions is determined to be MSE $= 2.44 * 10^{-5}$.

4.3.2 Narrow-to-broadband

The direct estimation regression model contains both the directional-to-hemispherical and the narrow-to-broadband behaviour. Now we build a regression model that only describes the narrow-to-broadband conversion. The input values in the training data are altered to the bihemspherical reflectance (BHR) for the wavelength bands. Note that this is a theoretical approach to isolate one part of the model, and this is not a practical thing to do, as the satellite data remains directional. This model can only be directly applied to the satellite data if the Lambertian approximation is valid.

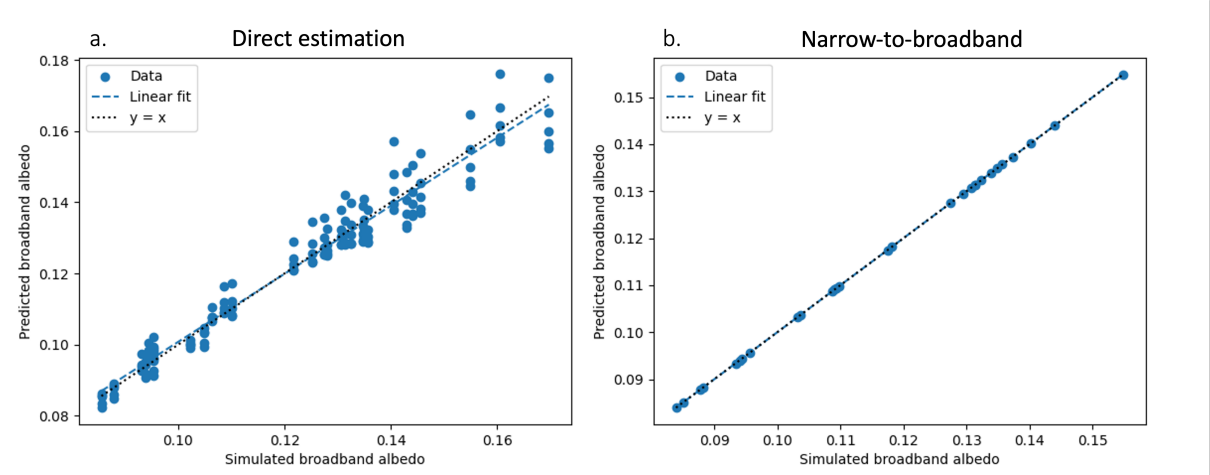

Figure 4.7: The relation between the predicted broadband albedo values and the simulated broadband albedo for the test set. The trend of the predictions are plotted by a linear fit, and the line $y = x$ shows the ideal predictions. ***a.*** *Direct estimation predictions.* ***b.*** *Narrow-to-broadband predictions.*

The narrow-to-broadband regression coefficients w_{phem} are found in Table 4.3. Again, this is for a single split of training and test data. Comparing the coefficients in this case to the regression coefficients for the direct estimation model, it stands out that their absolute values are much lower. The coefficients are also closer together and fewer are negative.

Table 4.3: The narrow-to-broadband regression coefficients for the different wavelength bands. The weights are given for a pine forest (w_{phem}) and a field (w_{fhem}). The weights as presented by Bonafoni [22] and Li et al. [20] are given by w_B and w_L, respectively.

	Band 2	Band 3	Band 4	Band 8	Band 8A	Band 11	Band 12	Constant
w_{phem}	-0.2464	0.3396	0.1427	0.3795	-	0.0831	0.1351	0.00002
w_B	0.2266	0.1236	0.1573	0.3417	-	0.1170	0.0338	-
w_L	0.2688	0.0362	0.1501	-	0.3045	0.1644	0.0356	-0.0049
w_{fhem}	0.1097	0.1155	0.2322	0.3226	-	0.3324	-0.2014	-0.0074

The predicted albedo from the narrow-to-broadband model is plotted against the simulated albedo. The graph in Figure 4.7b shows the result for one split of training and test data. The predicted values are close to the simulated values, with the data almost lying on the $y = x$ line. The slope of the fit through the predicted points is 1.001.

Comparing the spread of the data to that in Figure 4.7a, the vertical spread caused by the different zenith angles is not visible.

Cross-validation is also used for the narrow-to-broadband model. The mean coefficient of determination r^2 is 0.999996 with a standard deviation of $2.55{*}10^{-6}$. The average Mean Squared Error (MSE) is $1.66 * 10^{-9}$. The predictions of only the narrow-to-broadband component of the model are better than those from the direct estimation model.

4.3.3 Comparison to literature

The performance of the regression models is compared to that of models found in earlier research. This is done by creating predictions for the dataset created with the FRT model. In Figure 4.8, surface albedos are predicted for vegetation stands in the test data, using the weights from Bonafoni [22] and Li et al. [20]. They are compared to the simulated FRT surface albedos for the corresponding stands.

The Bonafoni and Li model coefficients are given in Table 4.3, as w_B and w_L, respectively. Note here that the model created by Li uses MSI band 8A instead of band 8, but in the predictions the weight has been applied to the band 8 reflectance. The central wavelength of band 8A is shifted up by about 30 nm, and the bandwidth is much narrower. Refer to Figure 2.1 and Appendix A for a full comparison between the two bands.

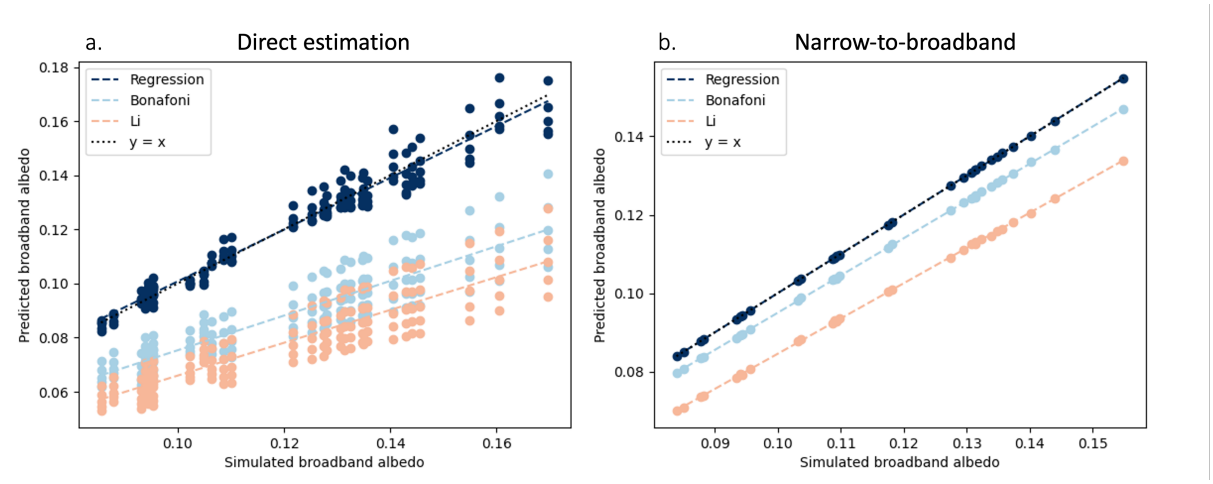

*Figure 4.8: The relation between the predicted broadband albedo values and the simulated broadband albedo. Predictions of the models are compared to those with model coefficients from Bonafoni [22] and Li [20]. **a.** The direct estimation model is trained and tested with FRT data from varying θ_v. **b.** The narrow-to-broadband model is trained and tested with bihemispherical reflectances from integrated FRT data.*

The regression model used in Figure 4.8a is the direct estimation model, including simulated satellite measurements at different θ_v. In Figure 4.8b, the regression model for narrow-to-broadband conversion is used.

The coefficients from literature describe the Sentinel-2 narrow-to-broadband conversion for a general surface. They do not include the BRDF conversion[3] [20, 22]. As such, the models from literature are comparable to the narrow-to-broadband regression model.

The direct estimation model can be compared to a combination of the Lambertian approximation for the angular step and the narrow-to-broadband conversions from literature. This comparison is included to visualise the difference between the two regression models.

The coefficients from literature have generally the same order of magnitude, compared to each other. Comparing the coefficients from earlier work to those of w_{phem} shows both differences and similarities. The biggest difference is the fact that the parameter for MSI band 2 is negative, though it does have the same order of magnitude. Band 12 from w_{phem} is also distinctively larger than its counterparts.

In Figure 4.8 we see that both models from literature estimate the albedo considerably lower than its simulated value. The Li model, has even lower predictions than the Bonafoni model. Also, the difference between predicted and simulated albedo is bigger in 4.8a, where the angular step of the conversion is included. Another observation is that the vertical spread of the different zenith angles in Figure 4.8a is more pronounced for the Bonafoni and Li models than for the trained regression model.

The results from the narrow-to-broadband regression model are more comparable to

[3]Li et al. [20] did evaluate the performance of applying MODIS parameters for the BRDF characterisation, but did not present any parameters to apply for Sentinel-2 data. As such it was chosen to only use the narrow-to-broadband conversion presented in their work.

the Bonafoni model than the Li model. The slopes of the fits in Figure 4.8b are 0.95 and 0.90 for Bonafoni and Li model respectively, while the trained regression model has a slope of 1.00. The difference between the predictions of the three models becomes higher for tree stands with a larger simulated albedo.

4.4 Comparison with different vegetation

4.4.1 Field model

In the previous section, the regression model that was trained using pine forest tree stands was used. In this section, the results for a similar model, but trained using reflectances of field stands, are presented. This is done for both the direct estimation and the narrow-to-broadband model. The coefficients obtained for this type of vegetation can be found as w_f in Table 4.2 and w_{fhem} in Table 4.3.

The weights for the direct estimation field model are all of similar order of magnitude. Compared to the pine forest model, the coefficients are closer together. There are also less negative weights for the field model. Wavelength band 12 has a larger coefficient than the other bands.

The weights for the narrow-to-broadband field model are also of a similar magnitude. Wavelength band 12 has the only negative coefficient. Compared to the other coefficients in the table, the band 4 weight stands out as being somewhat higher, whilst the band 8 coefficient is very similar to that of the other models. The weights in w_f are seemingly slightly higher in magnitude than those in w_{fhem}.

Cross-validation is used to evaluate the performance of the model. The mean coefficient of determination of the direct estimation field model is $r^2 = 0.966$, with a standard deviation of 0.00313. The average MSE is $1.30 * 10^{-6}$. For the narrow-to-broadband field model, these values are $r^2 = 0.99968$, with $\sigma_{r^2} = 0.000249$ and MSE $= 4.07 * 10^{-9}$. The narrow-to-broadband model is better at predicting the albedo than the direct estimation model.

The model albedo predictions are compared to the simulated albedos in Figure 4.9. Figure 4.9a shows the direct estimation estimates and Figure 4.9b shows the narrow-to-broadband results.

Comparing the simulated surface albedo values for the open field data to those of the pine forest, it is clear that the reflectance of the field is generally higher than that of the pine forest. The range of different reflectances is also smaller for the field. In both figures, the data points for the trained regression model do not vary far from the $y = x$ line.

Just as for the pine data, the models from literature predict a lower albedo than the simulated values for the field data. We also see that the Bonafoni model estimates the albedo higher than the Li model.

4.4.2 Pine vs. field model

The performance of the direct estimation pine forest and field regression models on different situations are shown in Figure 4.10. The models trained for one specific vegetation type. Their weights are then used to obtain albedo predictions for a different type of vegetation.

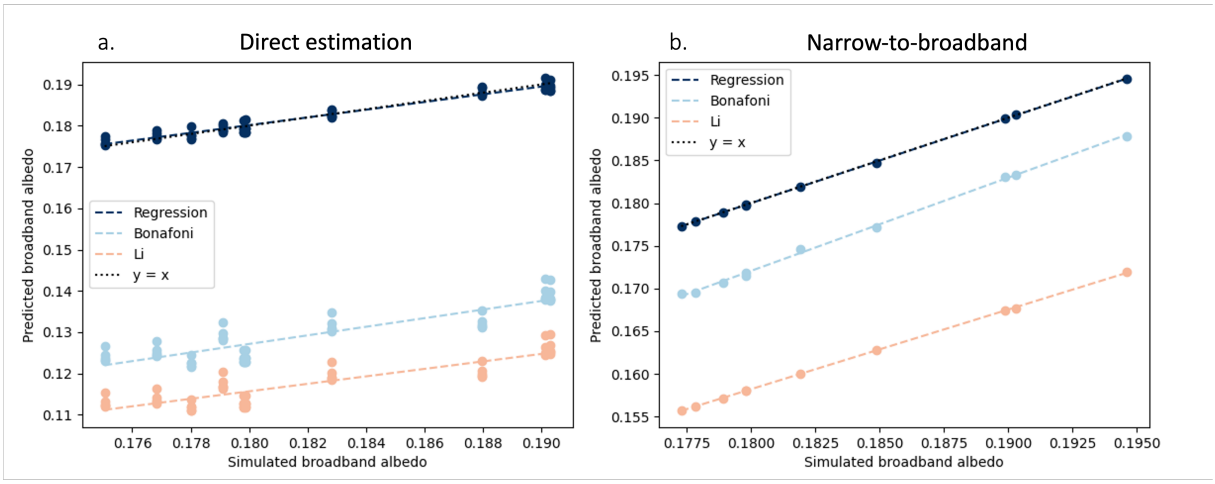

Figure 4.9: Albedo prediction for the open field regression model, compared to the models from literature. ***a.*** *The direct estimation models.* ***b.*** *The narrow-to-broadband models.*

Plot 4.10**a** shows the predictions of the two regression models for a test dataset with only pine forest reflectances. The predictions made with the field parameters w_f are generally too high. The graph shows that this overestimation is higher for lower simulated albedo cases. The vertical spread from different θ_v also seems slightly bigger.

Plot 4.10**b**, on the other hand, shows the predictions of the two models on the open field test data. If the simulated albedo is on the low end for a field, the predictions made by the pine forest model are too low. If the simulated albedo is high on the other hand, the pine model estimates are too high. The vertical spread for different θ_v is larger for the pine model predictions than for the field model predictions.

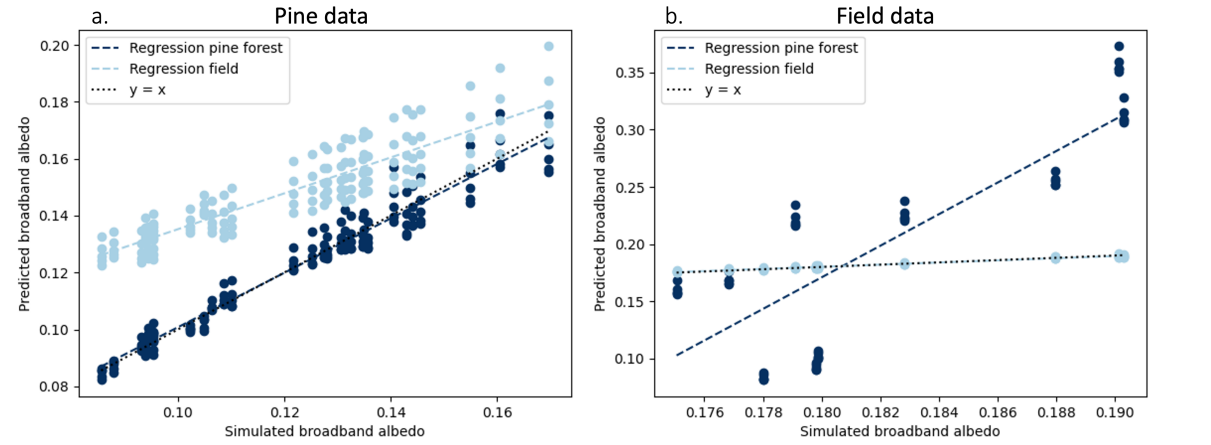

Figure 4.10: Comparison of the pine forest regression model and the field regression model predictions for ***a.*** *Pine stand reflectance data and* ***b.*** *Open field reflectance data.*

5

Discussion

5.1 FRT model

In the angular plots a kink was observed near the extreme view zenith angles. A likely cause for this behaviour is the inaccuracy of the FRT model for high θ_v. This behaviour may have an impact on the bihemispherical albedo calculations. However, in practical situations it is also difficult to determine the reflectance of forests at these angles.

The surface albedo is higher for the high θ_s case, and the reflectances change more quickly with varying view angle. In earlier work, the higher angular dependence of the reflectivity was observed for low sun altitudes [56]. This indicates the importance of the sun zenith angle in BRDF modelling. To evaluate the angular behaviour

Looking at the wavelength spectrum in Figure 4.3, two dips can be seen in the 700-1400 nm range. If we compare the wavelengths of these dips to the transmission spectrum in Figure 2.4, they are most likely caused by atmospheric absorbance, specifically that of water molecules.

The reflectance spectrum also shows that the reflected radiance is not zero for at the lower edge of the plotted spectrum. Since the FRT model does not run for wavelengths below 400 nm, it is not possible to take this contribution to the albedo into account. However, most of the incident and reflected light is absorbed by ozone in the atmosphere [10].

The simulated satellite measurements illustrated in Figure 4.4 show that the measured hemispherical-directional reflectance is generally lower than the bihemispherical reflectance (BHR) that should be considered in albedo calculations. This difference can be explained by the angular dependence of the reflectance. As illustrated in Figure 4.1, the reflectance coefficient is highest at the hotspot. Around nadir however, where the satellite measures the reflectance, the reflectance is much smaller. Typically then, the satellite measurements give a reflectance lower than the BHR.

5.2 Lambertian approximation

For the Lambertian approximation we saw that the estimated reflectance spectrum was generally lower than the BHR spectrum. Within the Sentinel-2 FOV, we also observe that a lower θ_v, results in a higher reflectance. When looking at the angular dependence of the reflectance, we can conclude that this is caused by the hotspot effect. For the lower view zenith angles, the satellite is closer to the hotspot, so the detected reflectance is higher.

The difference between the two solar angles can also be explained by angular dependence of the reflectance. Since the satellite measures at near-nadir angles, taking the Lambertian approximation will give an approximated BHR, that is lower than the actual BHR. The effect is more extreme in the case of high solar zenith angles. This is because for a higher θ_s, the hotspot is located further from nadir and the reflectance at low θ_v is smaller.

The results indicate that the validity of taking a Lambertian approximation depends on the solar- and the view zenith angle. From the relative errors in Table 4.1, it seems that the approximation is not that good in many situations, especially for higher θ_s. With extensive FRT modelling of additional geometries, it is possible to evaluate when the approximation can be applied.

5.3 Linear regression models

5.3.1 Direct estimation model

Because a regression model is purely mathematically based, it is difficult to draw intuitive physical conclusions from the regression weights. In principle, a high regression weight indicates that the corresponding sensor band is important in the computation of surface albedo. However, the bands are not independent, and this intercorrelation can cause some weights to be overly high, whilst others are too low or even negative in order to compensate [57].

The regression weights for the pine forest stand contain several negative values. This can indicate that the albedo contributions from these bands are also included in other bands [57]. This behaviour of the regression weights makes it difficult to draw conclusions on the importance of the individual sensor bands in the albedo predictions.

The higher prediction spread observed for higher simulated albedo in Figure 4.7 is possibly caused by the training data. There are more training data points in the 0.08-0.11 albedo range than in the 0.15-0.18 range. As a result, the model might have picked up a bias to cater more towards these lower albedo pine stands. If the realistic distribution of surface albedo is similar to that of the training set, this is a positive result. If the training set does not represent the realistic distribution of vegetation stands well, this can have a negative influence on the model's prediction accuracy. In this thesis, the realistic distribution was not considered, so this might be a valuable modification.

5.3.2 Narrow-to-broadband model

The weights for the narrow-to-broadband regression model are mostly lower than for the direct estimation model. In part, this can be explained by the input for the models. The HDR measured by the satellite at near-nadir angles is generally lower than the BHR taken as input for the narrow-to-broadband model. Since they still need to predict the same surface albedo, they are expectedly somewhat higher. For w_f and w_{fhem}, this could be a reasonable explanation for some of the differences. It is difficult to compare w_p and w_{phem} because of the extreme values for some bands in w_p. The cause for the behaviour of the w_p is uncertain. A tentative possibility is that the angular modelling of the pine forest reflection causes a big correlation between some sensor bands, leaving band 2 to make the prevalent contribution to the albedo.

5.3.3 Comparison to literature

In Figure 4.8a., the spread between the data from the different zenith angles is more apparent for the predictions made by the models from Li and Bonafoni. We suggest that the training of the model using different zenith angles results in a decreased sensitivity to the variations caused by measurements at different θ_v. This is the difference between the direct estimation model, which contains some angular modelling, and the Lambertian approximation combined with the narrow-to-broadband models from Bonafoni and Li. These results again indicate that some type of BRDF modelling is desirable.

Overall, the models from literature do not match very well with the trained regression model. The reason for differences lie in the way the models are developed. Li et al. also trained a regression model. They used a different set of training data, which was obtained from radiative transfer simulations of all surface types, excluding snow. The data used by Li et. al also consists of a large collection of different solar- and view angles, and atmospheric conditions. This is very different from the model presented in this thesis, which was trained for a specific set of angles, atmospheric conditions and vegetation type.

Bonafoni determines its weights by computing the fraction of the solar radiation at the surface for the surrounding wavelengths of each MSI band. This too is applicable for a very general case. It also does not take any reflectance and absorption properties of the atmosphere or the surface into account.

5.4 Comparison with different vegetation

The range of albedos considered for the field was narrow in comparison to that of the pine forest. This could be attributed to the composition of the training data. For the field, less training data was acquired, and input parameters were varied less. For a better comparison, it would be useful to extend the field stand training data.

The results show that the predictions for the pine forest made with the field regression model are all too high (Figure 4.10a). It is possible that the fraction of the total reflected radiance that falls within the included MSI sensor bands is larger for pine forests than for fields. This could be a result from differences in the spectral signatures. If this were the case however, one would expect that the predictions by the pine forest model in Figure 4.10b would consistently underestimate the albedo.

5.5 General discussion & threats to validity

5.5.1 Training and test data

The regression coefficients obtained showed a noticeable variance with different selections of training and test data. While the variance in the coefficient of determination for these specific cases was not very high, it still indicates that the model is dependent on its training data. This may be improved by creating a bigger set of training and test data.

In general, the amount, and quality of training data is a concern when considering machine learning models. A model is only as good as the data used to train it, and this needs to be kept in mind when drawing conclusions from the presented results. Especially for the open field model, the training data set used is perhaps not extensive enough.

In the plots in sections 4.3 and 4.4 we can see the data is clustered in little groups with similar albedo. Between the groups, few data points from the test set appear. This

is most likely a result of the FRT input parameter selection.

Obtaining more training data is the main way to improve the model. For this project, a balance had to be sought between amount of data on the one hand and runtime and storage space constraints on the other hand. For reference, the amount of uncompressed data stored over the course of this thesis comes to approximately 100GB. Future research may be better suited to handle higher amounts of data and longer computational times.

A second improvement to the training data is to alter its composition. For maintainability, only a specific number of important stand parameters in the FRT input file were changed. Varying additional parameters could lead to a more continuous data set, as opposed to the clustering apparent now. It would also be good to vary the parameters in a way that better reflects the realistic stand distribution.

The use of the FRT model to create the training data also influences the results, as it is a simplification of reality. Additionally, the FRT model displays some inaccuracies for high view angles.

5.5.2 The regression model

Throughout this research the assumption $\theta_s = 59°$ was used, to describe the average solar angle when Sentinel-2 passes over southern Sweden. Additionally, $\theta_r = 65°$ was taken. However, it was also seen that solar angle has a significant impact on both the total albedo and the reflection detected by the satellite. Similarly, the atmospheric conditions have not been varied in the training of the model. Varying the contributions of direct and diffuse light could have an interesting impact on the reflectance.

Since the regression model used is a linear regression model, only linear relations can be described by the model [54]. Complicated systems, such as the reflectance of forest may not be adequately described by a linear model. An indication towards this is given by Liang et al. [48]. They show that for the MODIS and Multiangle Imaging Spectroradiometer (MISR), second and third order polynomial regression models are more successful in predicting the total land surface albedo. The performance of a neural network was even better. Continuation of this research by training higher order regression models, or alternative machine learning models is recommended.

5.5.3 Validation

The validation of the regression models is done through 5-fold cross-validation. The coefficients of determination obtained in this way indicate that the model performs reasonably well in predicting the test data. However, this does not necessarily mean the model will perform well for practical situations, since the training data and test data are theoretical.

The second type of validation was done through a comparison with models from literature. The results from the Bonafoni and Li models differed significantly from our model results. However, it is difficult to draw conclusions from this, as the models from literature are developed for a more general case than the model presented in this thesis.

A useful way to validate the model further would be to use Sentinel-2 satellite data in combination with surface albedo measurements from pyranometers at the surface. For such validation, one needs to consider using the right vegetation type, solar- and view angles, and atmospheric conditions. Doing this validation could give a better indication about the model performance in reality, and is strongly recommended for future work.

6

Conclusion & Outlook

One way to strive against climate change is to impact the surface albedo. This can be done by altering land usage to maximise surface reflectance, but extensive knowledge of the surface albedo for different land use is required. Throughout this thesis, the surface albedo estimations from Sentinel-2 satellite data was investigated, using satellite data and surface albedo simulated with the FRT model.

1. An analysis of the angular behaviour has shown distinctly different reflectance patterns for different solar angles. It was found that the hemispherical-directional reflectances (HDR) at the near-nadir view angles within the satellite FOV are generally lower than the bihemispherical reflectance (BHR).

2. In section 4.2, we determined that taking the Lambertian approximation as opposed to integrating the HDR over the hemisphere has a significant impact on the surface albedo calculations. For higher solar zenith angle θ_s, this error is larger. If $\theta_s = 37°$, the error in the albedo ranges from 0-10%, depending on the view angle θ_v within the range from $-11°$ to $11°$. For $\theta_s = 59°$, this error has increased to 20-30% in the same range of θ_v. The error is smallest for the lower view angle and biggest for the higher θ_v. This effect is important if one considers using a narrow-to-broadband model like that presented by Bonafoni, without performing any BRDF modelling.

3. A direct estimation regression model to predict surface albedo was successfully trained for a pine forest and a field. The resulting models were evaluated using cross-validation. The pine forest model has a mean coefficient of determination of $r^2 = 0.946$ and the field model has $r^2 = 0.966$. These values indicate a good albedo prediction for the FRT simulated vegetation stands.

   The narrow-to-broadband regression model trained for the same vegetation stands showed a higher prediction accuracy than the direct estimation model. The coefficients of determination were $r^2 = 0.999996$ and $r^2 = 0.99968$, for the pine forest and the field respectively.

   The direct estimation model developed includes both angular modelling and narrow-to-broadband conversion in one step. While this leads to less propagation of errors, the regression coefficients obtained are less intuitive.

   A recommendation for further research is to develop a directional-to-hemispherical conversion model as well, solely doing the angular modelling step of the albedo estimation. With such a model, the direct estimation method can quantitatively be

compared to the stepwise albedo estimation methods. Another possibility would be to use BRDF models developed for different satellites, like MODIS, and evaluate the accuracy of applying those.

The model predictions were compared to the predictions of models from literature. Specifically for the FRT simulated vegetation stands, both the models proposed by Li et al. [20] and by Bonafoni [22] estimated the albedo lower than the simulated albedo. Several factors may contribute to the difference between the model presented here and those from earlier works.

The models in this thesis have been designed for a specific scope, which contains defined vegetation types, solar angles and atmospheric conditions. The models from literature are developed for a more general case. Hence, the different levels of specificity make comparison difficult.

4. Section 4.4 shows a comparison between the pine forest and field model. The characteristics of the two vegetation types that can be described by a linear model indeed seem to differ. An interesting next step would be to add more types of vegetation, such as deciduous or mixed forests. Additionally, other land cover types, such as snow, bare ground or water, are valuable to model.

We conclude that the model presented in this thesis is a suitable way to compare the reflectance patterns of different locations and situations. The main recommendation for future work is to validate the model with the use of Sentinel-2 and pyranometer data. Comparing this theory based model to practical data may give valuable insights. Moreover, this type of validation is needed before it can be used for actual albedo predictions in practical situations.

A further recommendation is to investigate the impact of sky conditions. This model may be a convenient way to explore the impact of different aerosol concentration and visibility on the surface albedo. Additionally, the performance of the model may be improved by increasing the amount of training data.

To summarize, we found that it is important to include angular modelling in albedo estimations. When analysing a smaller surface area, for which Sentinel-2 data is suitable, the results were highly dependent on the angles.

The regression models presented in this thesis perform well for the simulated vegetation stands. Moreover, it is valuable to train separate models for different land use classes. Both the comparison to literature and the comparison of the two vegetation types give an indication towards this.

With the results obtained here, a step is taken towards better surface albedo estimations on a smaller scale. The regression models require further development before practical use, but have shown potential for the evaluation of surface albedo at a fine spatial resolution. The Sentinel-2 satellite system is especially suitable for acquiring surface albedo on a smaller scale, because of its fine spatial resolution.

As development of new satellite missions is continuously ongoing, the measurement instruments that provide additional or more detailed spectral data will also appear. Combined with continued development in albedo estimation algorithms across satellite systems, our ability to understand the surface energy balance will improve.

Bibliography

[1] Gordon B. Bonan. Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. *Science*, 320(in Review):1444, 2008.

[2] Shunlin Liang, X. Zhang, Kaicun Wang, X. Zhang, and Martin Wild. Review on Estimation of Land Surface Radiation and Energy Budgets From Ground Measurement, Remote Sensing and Model Simulations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3(3):225–240, 2010.

[3] Darrel Hess and Dennis Tasa. *McKnight's Physical Geography: A Landscape Appreciation*. Pearson, pearson ne edition, 2014.

[4] Paul A. Miller and Benjamin Smith. Modelling tundra vegetation response to recent arctic warming. *Ambio*, 41(SUPPL.3):281–291, 2012.

[5] C. J.W. Bonfils, T. J. Phillips, D. M. Lawrence, P. Cameron-Smith, W. J. Riley, and Z. M. Subin. On the influence of shrub height and expansion on northern high latitude climate. *Environmental Research Letters*, 7(1), 2012.

[6] Abigail L Swann, Inez Y Fung, Samuel Levis, Gordon B Bonan, and Scott C Doney. Changes in Arctic vegetation amplify high-latitude warming through the greenhouse effect. 107(4):1295–1300, 2010.

[7] Gaétan Pique, Dominique Carrer, Emanuele Lugato, Rémy Fieuzal, Raphaël Garisoain, and Eric Ceschia. About the Assessment of Cover Crop Albedo Potential Cooling Effect: Risk of the Darkening Feedback Loop Effects. *Remote Sensing*, 15(13), 2023.

[8] David Parastatidis and Chrysoulakis Nektarios. RSLab Global Land Surface Albedo product from MODIS. *2023 Joint Urban Remote Sensing Event, JURSE 2023*, pages 1–4, 2023.

[9] Miguel O. Roman, Charles K. Gatebe, Yanmin Shuai, Zhuosen Wang, Feng Gao, Jeffrey G. Masek, Tao He, Shunlin Liang, and Crystal B. Schaaf. Use of in situ and airborne multiangle data to assess MODIS- and landsat-based estimates of directional reflectance and albedo. *IEEE Transactions on Geoscience and Remote Sensing*, 51(3):1393–1404, 2013.

[10] Shunlin Liang and Jindi Wang. *Advanced Remote Sensing: Terrestrial Information Extraction and Application*. 2nd edition, 2020.

[11] A. H. Strahler and J. P. Muller. MODIS BRDF Albedo Product : Algorithm Theoretical Basis Document. *MODIS Product ID: MOD43*, Version 5.(April):1–53, 1999.

[12] Andres Kuusk and Tiit Nilson. A directional multispectral forest reflectance model. *Remote Sensing of Environment*, 72(2):244–252, 2000.

[13] Andres Kuusk, Joel Kuusk, and Mait Lang. Modeling directional forest reflectance with the hybrid type forest reflectance model FRT. *Remote Sensing of Environment*, 149:196–204, 2014.

[14] Ying Qu, Shunlin Liang, Qiang Liu, Tao He, Suhong Liu, and Xiaowen Li. Mapping surface broadband albedo from satellite observations: A review of literatures on algorithms and products. *Remote Sensing*, 7(1):990–1020, 2015.

[15] Christopher O. Justice, Eric Vermote, John R.G. Townshend, Ruth Defries, David P. Roy, Dorothy K. Hall, Vincent V. Salomonson, Jeffrey L. Privette, George Riggs, Alan Strahler, Wolfgang Lucht, Ranga B. Myneni, Yuri Knyazikhin, Steve W. Running, Rama R. Nemani, Zhengming Wan, Alfredo R. Huete, Wim Van Leeuwen, Robert E. Wolfe, Louis Giglio, Jan Peter Muller, Philip Lewis, and Michael J. Barnsley. The moderate resolution imaging spectroradiometer (MODIS): Land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing*, 36(4):1228–1249, 1998.

[16] US Geological Survey and NASA. Landsat 9 Data Users Handbook. Technical Report February, USGS, NASA, 2022.

[17] Tao He, Shunlin Liang, Dongdong Wang, Yunfeng Cao, Feng Gao, Yunyue Yu, and Min Feng. Evaluating land surface albedo estimation from Landsat MSS, TM, ETM +, and OLI data based on the unified direct estimation approach. *Remote Sensing of Environment*, 204(October 2017):181–196, 2018.

[18] Yuanyuan Li, Qingsheng Liu, Gaohuan Liu, and Chong Huang. Detect quasi-circular vegetation community patches using images of different spatial resolutions. In *Proceedings of the 2013 6th International Congress on Image and Signal Processing, CISP 2013*, pages 824–829. IEEE, 2013.

[19] European Space Agency. Sentinel-2 User Handbook. Technical Report 1, 2015.

[20] Zhan Li, Angela Erb, Qingsong Sun, Yan Liu, Yanmin Shuai, Zhuosen Wang, Peter Boucher, and Crystal Schaaf. Preliminary assessment of 20-m surface albedo retrievals from sentinel-2A surface reflectance and MODIS/VIIRS surface anisotropy measures. *Remote Sensing of Environment*, 217(August):352–365, 2018.

[21] David P. Roy, Jian Li, Hankui K. Zhang, Lin Yan, Haiyan Huang, and Zhongbin Li. Examination of Sentinel-2A multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance. *Remote Sensing of Environment*, 199:25–38, 2017.

[22] Stefania Bonafoni and Aliihsan Sekertekin. Albedo Retrieval from Sentinel-2 by New Narrow-to-Broadband Conversion Coefficients. *IEEE Geoscience and Remote Sensing Letters*, 17(9):1618–1622, 2020.

[23] James B. Campbell and Randolph H. Wynne. *Introduction of Remote Sensing*, volume 3. 2011.

[24] Peg Shippert. Push Broom and Whisk Broom Sensors, 2013.

[25] European Space Agency. Colour vision for Copernicus: The story of Sentinel-2. *ESA bulletin*, (161):3–9, 2015.

[26] Robert A. Schowengerdt. *Remote Sensing: Models and Methods for Image Processing*. Academic Press, 3rd edition, 2006.

[27] Copernicus. Copernicus Data Space Ecosystem, 2023.

[28] W Wanner, A H Strahler, B Hu, P Lewis, X Li, and C L Barker. Ptoa- P '. 102:143–161, 1997.

[29] Klaus Tempfli, Norman Kerle, Gerrit C Huurneman, Lucas L F Janssen, Wim H Bakker, Wim Feringa, Ambro S M Gieske, Ben G H Gorte, Karl A Grabmaier, Chris A Hecker, John A Horn, Gerrit C Huurneman, Lucas L F Janssen, Norman Kerle, Freek D Van Der Meer, Gabriel N Parodi, Christine Pohl, Colin V Reeves, Frank J Van Ruitenbeek, Ernst M Schetselaar, Klaus Tempfli, Michael J C Weir, Eduard Westinga, and Tsehaie Woldai. *Principles of Remote Sensing: An introductory textbook*. The International Institute for Geo-Information Science and Earth Observation (ITC), fourth edition, 2009.

[30] A.R. Huete. 11 - Remote Sensing for Environmental Monitoring. In Janick F. Artiola, Mark L. Brusseau, and Ian L. Pepper, editors, *Environmental Monitoring and Characterization*, chapter 11 - Remot, pages 183–206. Elsevier, Inc., 2004.

[31] Janardhanan Sundaresan, K M Santosh, Andrea Déri, Rob Roggema, and Ramesh Singh. *Geospatial Technologies and Climate Change*. Springer International, 2013.

[32] Randall B. Smith and TNTmips. Introduction to Hyperspectral Imaging. Technical report, 2012.

[33] Burhan U. Choudhury, Rumi Narzari, Md Zafar, Naseeb Singh, Vinay Kumar Mishra, and Mathyam Prabhakar. Spectral library of crops and discrimination of major vegetables grown in the eastern Himalayan ecosystem: A proximal hyperspectral remote sensing approach. *Ecological Informatics*, 77(August):102263, 2023.

[34] J. G.P.W. Clevers and A. A. Gitelson. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on sentinel-2 and-3. *International Journal of Applied Earth Observation and Geoinformation*, 23(1):344–351, 2013.

[35] D. N.H. Horler, M. Dockray, J. Barber, and A. R. Barringer. Red edge measurements for remotely sensing plant chlorophyll content. *Advances in Space Research*, 3(2):273–277, 1983.

[36] Warren J. Smith. *Modern Optical Engineering: The Design of Optical Systems*. McGraw-Hill, third edition, 2000.

[37] G. Schaepman-Strub, M. E. Schaepman, T. H. Painter, S. Dangel, and J. V. Martonchik. Reflectance quantities in optical remote sensing-definitions and case studies. *Remote Sensing of Environment*, 103(1):27–42, 2006.

[38] Julien Radoux, Guillaume Chomé, Damien Christophe Jacques, François Waldner, Nicolas Bellemans, Nicolas Matton, Céline Lamarche, Raphaël D'Andrimont, and Pierre Defourny. Sentinel-2's potential for sub-pixel landscape feature detection. *Remote Sensing*, 8(6), 2016.

[39] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis. Geometrical Considerations and Nomenclature for Reflectance. *Natl Bur Stand (US) Monogr*, (160):1–52, 1977.

[40] Andres Kuusk, Tiit Nilson, Markko Paas, Mait Lang, and Joel Kuusk. Validation of the forest radiative transfer model FRT. *Remote Sensing of Environment*, 112(1):51–58, 2008.

[41] Lars Eklundh, Lars Harrie, and Andres Kuusk. Investigating relationships between landsat ETM+ sensor data and leaf area index in a boreal conifer forest. *Remote Sensing of Environment*, 78(3):239–251, 2001.

[42] Lars Eklundh, Per Jönsson, and Andres Kuusk. Investigating modelled and observed Terra/MODIS 500-m reflectance data for viewing and illumination effects. *Advances in Space Research*, 39(1):119–124, 2007.

[43] S. Jacquemoud and F. Baret. PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment*, 34(2):75–91, 1990.

[44] Terence P. Dawson, Paul J. Curran, and Stephen E. Plummer. LIBERTY - Modeling the effects of Leaf Biochemical Concentration on Reflectance Spectra. *Remote Sensing of Environment*, 65(1):50–60, 1998.

[45] Eric F. Vermote, Didier Tanré, Jean Luc Deuzé, Maurice Herman, and Jean Jacques Morcrette. Second simulation of the satellite signal in the solar spectrum, 6s: an overview. *IEEE Transactions on Geoscience and Remote Sensing*, 35(3):675–686, 1997.

[46] Andres Kuusk and Tiit Nilson. Forest reflectance and transmittance FRT user guide. Technical report, 2013.

[47] Shunlin Liang. A direct algorithm for estimating land surface broadband albedos from MODIS imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 41(1):136–145, 2003.

[48] Shunlin Liang, Alan H. Strahler, and Charles Walthall. Retrieval of land surface albedo from satellite observations: A simulation study. *Journal of Applied Meteorology*, 38(6):712–725, 1999.

[49] Xiaoning Zhang, Jing Guo, Rui Xie, Ziti Jiao, Yadong Dong, Tao He, Anxin DIng, Siyang Yin, Hu Zhang, Lei Cui, and Yaxuan Chang. Development of the Direct-Estimation Albedo Algorithm for Snow-Free Landsat TM Albedo Retrievals Using Field Flux Measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 58(3):1550–1567, 2020.

[50] National Renewable Energy Laboratory. 2000 ASTM Standard Extraterrestrial Spectrum Reference E-490-00, 2000.

[51] Christophe François, Catherine Ottlé, Albert Olioso, Laurent Prévot, Nadine Bruguier, and Yannick Ducros. Conversion of 400-1100 nm vegetation albedo measurements into total shortwave broadband albedo using a canopy radiative transfer model. *Agronomie*, 22(6):611–618, 2002.

[52] European Space Agency. MultiSpectral Instrument (MSI) Overview, 2023.

[53] AI Forum. How much training data is needed for Machine Learning?, 2021.

[54] Christopher M. Bishop. *Pattern Recognition and Machine Learning*, volume 23. Springer, 2006.

[55] Cecie Starr, Christine A. Evers, and Lisa Starr. *Biology: Concepts and Applications*. Thomson Brooks/Cole, sixth edition, 2006.

[56] Dwight D. Egbert and Fawwaz T. Ulaby. Effects of Angles on Reflectivity. *Photogrammetric Engineering & Remote Sensing*, June:556–564, 1972.

[57] Christoph Molnar. *Interpretable Machine Learning: A Guide For Making Black Box Models Explainable*. Independently published, 2022.

[58] European Space Agency. Sentinel-2: Spectral Resolution, 2023.

Appendix A

Sentinel-2 supplementary specifications

Band number	S2A		S2B		L_{ref} (reference radiance) (W m−2 sr−1 µm−1)	SNR @ Lref
	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)		
2	492.7	65	492.3	65	128	154
3	559.8	35	558.9	35	128	168
4	664.6	30	664.9	31	108	142
8	832.8	105	832.9	104	103	174

Figure A.1: The 10 m spatial resolution bands. From ESA [58]

.

Band number	S2A		S2B		L_{ref} (reference radiance) (W m−2 sr−1 µm−1)	SNR @ Lref
	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)		
5	704.1	14	703.8	15	74.5	117
6	740.5	14	739.1	13	68	89
7	782.8	19	779.7	19	67	105
8a	864.7	21	864.0	21	52.5	72
11	1613.7	90	1610.4	94	4	100
12	2202.4	174	2185.7	184	1.5	100

Figure A.2: The 20 m spatial resolution bands. From ESA [58]

.

Band number	S2A		S2B		L_{ref} (reference radiance) (W m−2 sr−1 µm−1)	SNR @ Lref
	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)		
5	704.1	14	703.8	15	74.5	117
6	740.5	14	739.1	13	68	89
7	782.8	19	779.7	19	67	105
8a	864.7	21	864.0	21	52.5	72
11	1613.7	90	1610.4	94	4	100
12	2202.4	174	2185.7	184	1.5	100

Figure A.3: The 60 m spatial resolution bands. From ESA [58]

.

Appendix B

FRT model

B.1 Input file

```
'Järvselja Pine'.                                  : data set name
124                                                : stand age
1                                                  : # size classes
*** files of refractive index and other tree classes:
'refrind.dat' 'bepe'
x0          xmin         xmax         dx                                                    i
'pine'                                             : species
t_elli                                             : crown form
.1115       .0001        .08          .02          : stand density, m-2                     1
15.9        10.          25.          5.           : tree height, m                         2
4.2         .5           10.          9.           : crown l,m ell V con                    3
0           .5           10.          1.           :      cylinder                          4
1.5         .2           5.           .3           : crown radius, m                        5
18          2.           25.          5.           : trunk diameter, cm                     6
2.67        1.1          3.           8.           : m - total dry leaf weight, kg/tree     7
160         30.          180.         60.          : SLW - leaf weight per area, gm-2       8
3.99        .0           4.5          .5           : eln3 - -ln(1 - eps)                    9
53.57       0.           90.          20.          : thm3 - modal leaf angle                10
.1          .05          .6           .2           : shoot length, m                        11
.3123       .01          1.           .05          : BAI/LAI                                12
1.69        .6           2.8          .05          : tree       distr.       param          13
.4          .1           .6           .05          : H-G assymmetry (phase function)        14
.6                                                 : shoot      shading      coef
'prospect'                                         : leaf optics model
4                                                  : # of leaf components
240.        50.          320.         50.   'waterb.dat'       : c2, % of SLW, comp.1       15
.5540       .3           1.           .2    'chlorp3.dat'      : c1, % of SLW, comp.2       16
97.11       94.          99.8         20.   'drymatter.dat'    : c3, % of SLW, comp.3       17
18.91       0.           40.0         20.          'base.dat'   : c4, % of SLW, comp.4       18
1.6016      1.6          2.8          .5           : leaf str. param. - PROSPECT N           27
.9          .6           1.2          .2           : refraction index ratio                 28
'pine_branch_1.dat'                                : file of branch reflectance
'pinetr1.dat'                                      : file of trunk reflectance
***  Ground vegetation  ***
.208        .01          6.           .3           : LAI2_ground, upper layer               29
.15         .02          .4           .05          : sl2 HS-parameter                       30
1.0         .4           1.           .2           : clmp2 - foliage clumping parameter     31
1.2         0.           2.           .3           : szz - vertical regularity              32
3.99        .0           4.5          .5           : eln2 - -ln(1 - eps)                    33
53.57       0.           90.          20.          : thm2 - modal leaf angle                34
.991        .6           1.3          .2           : n_ratio2                               35
81.7        80.          180.         30.          : SLW2(g/m^2)                            36
'prospect'                                         : leaf optics model, upper layer
4                                                  : # of leaf components
139.082     130.         320.         50.   'waterb.dat'       : c1, % of SLW, comp.1       37
.36         .3           .8           .2.   'chlorp3.dat'      : c2, % of SLW, comp.2       38
99.52       94.          99.8         20.   'drymatter.dat'    : c3, % of SLW, comp.3       39
.10         .0002        4.           .1.   'brownpigm.dat'     : c4, % of SLW, comp.4       40
1.315       1.           2.8          .2           : N2 (PROSPECT)                          49
1.064       .01          1.           .3           : LAI1_ground, lower layer
            50
.15         .02          .4           .05          : sl1 - HS-parameter                     51
1.          .4           1.           .2           : clmp1 - foliage clumping parameter     52
3.0         .0           4.5          .5           : eln1 - -ln(1 - eps)                    53
75.469      0.           90.          20.          : thm1 - modal leaf angle                54
1.224       .6           1.3          .2           : n_ratio1
55
78.54       80.          180.         30.          : SLW1(g/m^2)
            56
```

```
'prospect'                                          : leaf optics model, lower layer
5                                                   : # of leaf components
134.236      130.         320.         50.   'waterb.dat'       : c1, % of SLW, comp.1                57
.425         .3           .8           .2   'chlorp3.dat'       : c2, % of SLW, comp.2                58
0.733        .3           .8           .2   'anthocyanins.dat'  : c3, % of SLW, comp.3                59
98.343       94.          99.8         20.   'drymatter.dat'    : c4, % of SLW, comp.4                60
.50          .0002        4.           .1   'cellp3.dat'        : c5, % of SLW, comp.5                61
1.0053       1.0          2.5          .2          : N1 (PROSPECT)                                     69
'price.dat'  45.                                    : file of Price' vectors, th*
.217         .05          .4           .07         : s1 - soil parameters                              70
-.05         -.1          .1           .02         : s2                                                71
.0           -.05         .05          .02         : s3                                                72
.0           -.04         .04          .02         : s4                                                73
4            .80          .17    .0    .03         : iaer, c(n) - aerosol data (6S)
0.           .09                                    : v, tau_aer(550) - visibility (6S)
2      : *ijob*: 0-single, 1-spectrum, 2-ad, 3-n_sun, 4,5,6-inversion (4-relat., 5-abs. differences, 6-BRF)
1       1       -1.                  : # of Sun angles, spectral bands/BRF values, spectrum step
59.                                  : Sun zeniths
830                                  : spectral bands
0.      1.      65                   : view nadir angle, its increment, and view azimuth angle

****|****|****|****|****|****|****|****|****|****|****|****|****|****|****|****|
```

B.2 Atmospheric parameters

Table B.1: Atmospheric parameters used for the 6S radiative transfer model

Parameter	Value
Dust-like aerosol component (%)	0.80
Oceanic aerosol component (%)	0.17
Water-soluble aerosol component (%)	0.0
Soot aerosol component (%)	0.03
Optical thickness τ_{aer}^{550}	0.09

Appendix C

Training data parameters

The FRT model is run for many different cases in order to get enough training data. To create these cases, several parameters in the FRT input file are varied.

First it is decided which parameters have a big influence on the result, and should thus be varied over. To do this, a selection of them are varied between certain bounds. These bounds are based on some indicated minima and maxima given in the user manual of the FRT model [46].

While changing individual parameters, the model is run and the reflectances are plotted. From the figures, four of the more impactful parameters were chosen to be varied to create the training data. In Figure C.1 we see the resulting figures for these four parameters.

Three of the resulting factors were physical. These were the stand density, the crown radius and the leaf weight per area. Besides this one chemical parameters was chosen, which was the chlorophyll percentage of the leaf weight.

Notable from the chlorophyll graph, Figure C.1d, is that this parameter mainly impacts the visible part of the spectrum. A higher chlorophyll concentration leads to more absorption of red light, causing a greener appearance.

Finally, to create the training data, the four chosen parameters are varied more closely around the original values in the input file for the Järvelsja Pine stand. This is done in order to stay within a scenario of a physical forest stand, whilst creating enough variation in the reflectance to train the regression model. The ranges of the varied parameters are shown in Table 3.1.

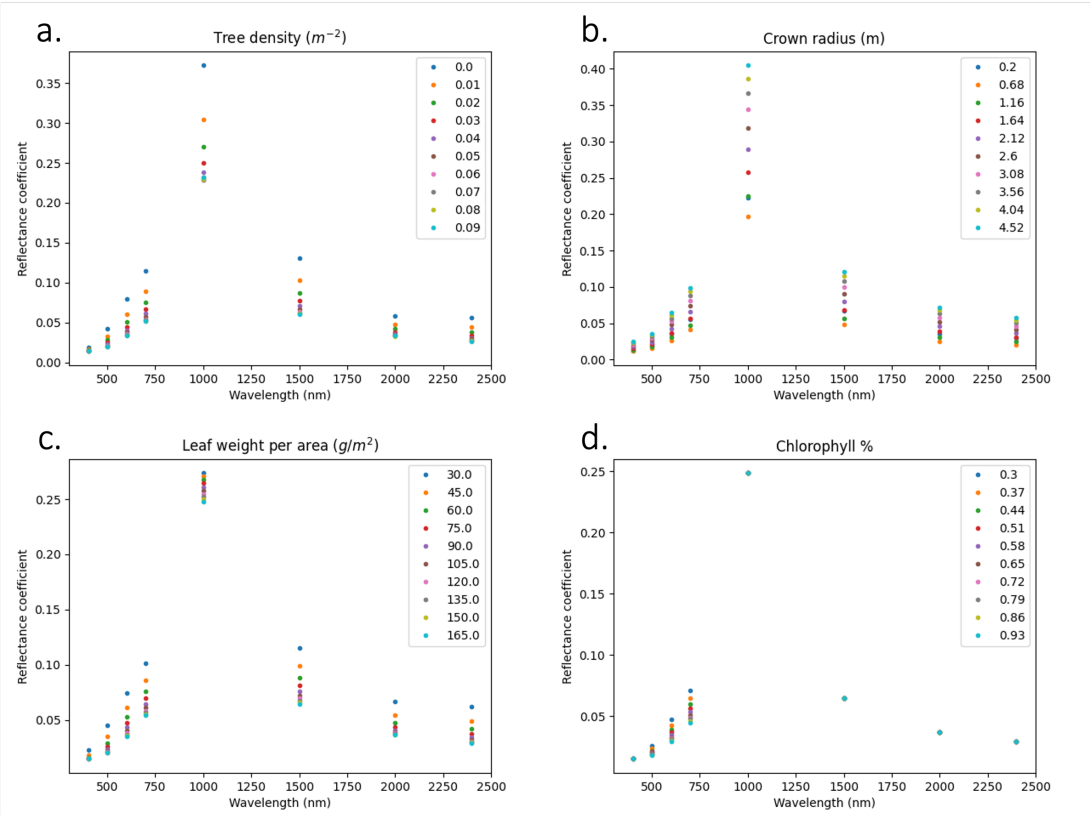

Figure C.1: Reflectance spectrum with different tree stand input parameters.

Appendix D

Additional regression results

D.0.1 Direct estimation model

Table D.1: Regression coefficients w_p

ID	Band 2	Band 3	Band 4	Band 8	Band 11	Band 12	Constant
1	77.7521	-31.4720	-14.4436	0.3492	2.6958	-5.9654	-0.1183
2	75.4187	-30.4233	-14.3491	0.3656	2.5085	-5.4414	-0.1129
3	76.8428	-31.1393	-14.2788	0.3756	2.7608	-6.1590	-0.1203
4	72.3105	-29.2303	-13.5841	0.3903	2.3372	-5.1862	-0.1085

Table D.2: Regression coefficients w_f

ID	Band 2	Band 3	Band 4	Band 8	Band 11	Band 12	Constant
1	0.2168	0.2455	0.2284	0.3168	-0.3478	0.9777	0.0686
2	-0.0153	0.3592	0.2452	0.3351	-0.4064	1.0600	0.0682
3	0.2516	0.2162	0.2434	0.3292	-0.3857	1.0005	0.0695

D.0.2 Narrow-to-broadband model

Table D.3: Regression coefficients w_{phem}

ID	Band 2	Band 3	Band 4	Band 8	Band 11	Band 12	Constant
1	-0.3106	0.3639	0.1593	0.3795	0.0853	0.1296	0.00002
2	-0.2987	0.3567	0.1582	0.3790	0.0894	0.1234	0.000001
3	-0.2976	0.3612	0.1485	0.3802	0.0802	0.1414	0.00004
4	-0.2463	0.3395	0.1426	0.3794	0.0830	0.1350	-0.00002

Table D.4: Regression coefficients w_{fhem}

ID	Band 2	Band 3	Band 4	Band 8	Band 11	Band 12	Constant
1	0.1096	0.1155	0.2322	0.3225	0.3323	-0.2013	-0.0073
2	0.0286	0.1559	0.2460	0.3256	0.3233	-0.1898	-0.0079
3	0.1095	0.1171	0.2290	0.3223	0.3251	-0.1904	-0.0067