DGK Ausschuss Geodäsie (DGK)
der Bayerischen Akademie der Wissenschaften

Reihe C | Dissertationen | Heft Nr. 949

Thomas Roßberg

Deep Learning-Based Estimation of NDVI Data from SAR Backscatter to Enhance Global Vegetation Monitoring

München 2025

Verlag der Bayerischen Akademie der Wissenschaften, München

ISSN 0065-5325 ISBN 978 3 7696 5361-8

DGK Ausschuss Geodäsie (DGK)
der Bayerischen Akademie der Wissenschaften

Reihe C Dissertationen Heft Nr. 949

Deep Learning-Based Estimation of NDVI Data from SAR Backscatter to Enhance Global Vegetation Monitoring

Vollständiger Abdruck der von der Fakultät für Luft- und Raumfahrttechnik
der Universität der Bundeswehr München
zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

angenommenen Dissertation

von

Thomas Roßberg

München 2025

Verlag der Bayerischen Akademie der Wissenschaften, München

ISSN 0065-5325 ISBN 978 3 7696 5361-8

Adresse des Ausschusses Geodäsie (DGK):

Ausschuss Geodäsie (DGK) der Bayerischen Akademie der Wissenschaften
Alfons-Goppel-Straße 11 ● D – 80 539 München
Telefon +49 – 89 – 23 031 1113 ● Telefax +49 – 89 – 23 031 - 1283 / - 1100
e-mail post@dgk.badw.de ● http://www.dgk.badw.de

Gutachter: 1. Univ.-Prof. Dr.-Ing. habil. Michael Schmitt
2. Prof. Juan Manuel Lopez-Sanchez

Die Dissertation wurde am 25.09.2024 bei der Universität der Bundeswehr München eingereicht
und durch die Fakultät für Luft- und Raumfahrttechnik am 03.01.2025 angenommen.
Die mündliche Prüfung fand am 03.02.2025 statt.

Diese Dissertation ist auf dem Server des Ausschusses Geodäsie (DGK)
der Bayerischen Akademie der Wissenschaften, München unter <http://dgk.badw.de/>
sowie auf dem Server der Universität der Bundeswehr, München unter
<https://athene-forschung.unibw.de/151902> elektronisch publiziert.

ISSN 0065-5325 ISBN 978 3 7696 5361-8

sehnsucht auch, das ist belegt
heißt die triebkraft, die uns bewegt
wir wärn nur satte marionetten
wenn wir keine sehnsucht hätten

drum hebt das glas mit mir
und gemeinsam trinken wir
auf die liebe und auf die sehnsucht

Oxo86

Abstract

Optical remote sensing provides invaluable data for monitoring the Earth and its vegetation. However, cloud cover hinders the acquisition of images and leads to data gaps. While radar remote sensing can penetrate clouds, different sensing principles and data characteristics prevent a direct data transfer between the two modalities. This thesis aims to bridge this gap by translating synthetic aperture radar (SAR) data into the most commonly used optical vegetation index, the normalized difference vegetation index (NDVI).

First, the relationship between SAR backscatter and NDVI values is explored as a basis for a potential translation. This is done for three globally distributed agricultural study areas covering a range of environmental conditions. The analysis includes data from several SAR sensors with different frequencies, including C-, S-, X- and L-band data. The investigation reveals a notable relationship between S- and C-band data, but also demonstrates the influence of numerous factors on this relationship, limiting generalization.

Building on the previously established relationship, the next step is to demonstrate the estimation of NDVI images from SAR backscatter data. For this purpose, a U-Net, a deep learning model, is trained. To allow a global application, a comprehensive dataset is created, named SEN12TP, consisting of close temporal pairs of Sentinel-1 and Sentinel-2 images from over 1200 different areas with a balanced distribution considering land cover, climate, and seasonality. The evaluation demonstrates the low error and good spatial detail of the trained U-Net. Further, it is shown that the model is globally applicable, outperforming a region-specific model.

Time series can be generated from the SAR-estimated NDVI images, but their utility would be considerably enhanced by integrating them with the available sparse optical data. Consequently, a flexible approach to fuse remotely sensed time series is presented as the third and final aspect of this dissertation. This approach is based on an RNN. For training purposes, a dataset consisting of 1.5 years of data and regions from the SEN12TP dataset was created. The results demonstrate the successful fusion of SAR-estimated and optical NDVI time series. A low error is achieved for both short and long gaps while allowing for the global application of this method.

Overall, this thesis presents a comprehensive framework to overcome the inherent limitations of optical, cloud-affected vegetation indices. This is achieved by augmenting the indices with information derived from SAR data.

Kurzfassung

Die optische Fernerkundung liefert wertvolle Daten zur Überwachung der Erde und ihrer Vegetation. Bewölkung verhindert jedoch die Aufnahme von Bildern und führt zu Lücken in den Daten. Obwohl die Radarfernerkundung Wolken durchdringen kann, verhindern unterschiedliche Messprinzipien und Bildeigenschaften eine direkte Übertragung zwischen den beiden Modalitäten. Ziel dieser Arbeit ist es, diese Lücke zu schließen, indem synthetic aperture radar (SAR)-Daten in den am häufigsten verwendeten optischen Vegetationsindex, den normalized difference vegetation index (NDVI), übersetzt werden.

Als erster Schritt und Grundlage für eine mögliche Übersetzung wird der Zusammenhang zwischen SAR-Rückstreuung und NDVI-Werten für drei weltweit verteilte landwirtschaftliche Studiengebiete untersucht. Für diese Gebiete weden sowohl optische als auch SAR-Daten (C-, S-, X- und L-Band) analysiert. Die Untersuchung zeigt einen Zusammenhang zwischen dem NDVI und S- sowie C-Band-Daten, verdeutlicht aber auch den Einfluss zahlreicher Faktoren auf diese Beziehung, was eine allgemeingültige Aussage einschränkt.

Aufbauend auf dieser ermittelten Beziehung wird in einem weiteren Schritt die Abschätzung von NDVI-Bildern aus SAR-Rückstreudaten demonstriert. Dazu wird ein U-Net, ein tiefes neuronales Netz, trainiert. Um eine globale Anwendung zu ermöglichen, wird ein umfangreicher Datensatz mit dem Namen SEN12TP erstellt, der aus zeitlich eng gekoppelten SAR- und optischen Bildern von über 1200 Gebieten besteht. Die Evaluierung des trainierten U-Nets zeigte geringe Fehler und eine gute räumliche Auflösung. Es zeigte sich, dass das Modell global anwendbar ist und einem regional angepassten Modell überlegen ist.

Aus den SAR-geschätzten NDVI-Bildern können Zeitreihen generiert werden, deren Nützlichkeit durch die Integration mit den verfügbaren unregelmäßigen optischen Daten zusätzlich verbessert werden könnte. Daher wird als letzter Aspekt dieser Arbeit ein flexibler Ansatz zur Fusion von Satellitenbildzeitreihen vorgestellt. Dieser Ansatz basiert auf einem rekurrenten neuronalen Netzwerk (RNN). Die Auswertung zeigt die erfolgreiche Fusion von SAR-abgeleiteten und optischen NDVI-Zeitreihen. Sowohl für kurze als auch für längere Lücken konnte ein geringer Vorhersagefehler erzielt werden und gleichzeitig die globale Anwendbarkeit dieser Methode gezeigt werden.

Insgesamt stellt diese Arbeit einen zuverlässigen Ansatz zur Verbesserung der durch Bewölkung gestörten optischen Vegetationsindizes aus SAR-Daten dar.

Contents

Acronyms

1 Introduction

1.1 Motivation

The observation of Earth from space is an invaluable tool for understanding the complex systems that characterize our planet. Monitoring the status of vegetation on a global scale is of particular importance, as this has become critical for numerous applications in light of the pressing global challenges humanity currently faces. Climate change, desertification, and food security are just a few examples of these issues. Therefore, understanding the health and dynamics of plant ecosystems, including agricultural crops and natural forests, has never been more vital. Effective vegetation monitoring allows humanity to predict crop yields, assess the impact of deforestation, and estimate carbon sinks and sources, all of which are essential for informed environmental management.

In this context, multispectral optical data serves as a powerful tool, offering detailed insights into plant health and growth patterns. The wavelengths utilized by optical sensors are highly informative about vegetation, and there are well-established techniques for analyzing such data. Furthermore, the data allows for straightforward visual interpretation by humans, and, as a method established at an early stage, long-term archives of optical data are available. One particularly common tool for vegetation monitoring is the normalized difference vegetation index (NDVI). The NDVI combines red and infrared spectral bands to provide an easily interpretable value of the plant health, vegetation state, and its "greenness".

While optical sensors offer numerous advantages for vegetation monitoring, they do have a significant limitation: their operating principle is hindered by cloud cover. This challenge is particularly pronounced in tropical regions, where the monthly mean cloud fraction often exceeds 80 %. However, it is also evident in moderate climates, such as that of Germany, where the mean daily cloud cover remains above 50 % during the summer months (King et al. 2013). Moreover, subtropical regions encounter this issue during the rainy season, when cloud cover is very high just as vegetation growth is reaching its peak.

For monitoring purposes, these data gaps can exceed 30 days across extensive regions of the tropics and subtropics (Flores-Anderson et al. 2023). An illustrative example of these long-term gaps is presented in Figure 1.1, which shows that nearly all images of an area in Burkina Faso in subtropical Africa between June and October are at least partially obscured by clouds. This period corresponds to the rainy season, which

is also the primary growing season (Belesova et al. 2017), highlighting the critical importance of capturing data during this period for effective vegetation and crop research. Thus, relying solely on optical sensors that require clear skies can lead to significant data loss, underscoring the need for alternative monitoring methods.

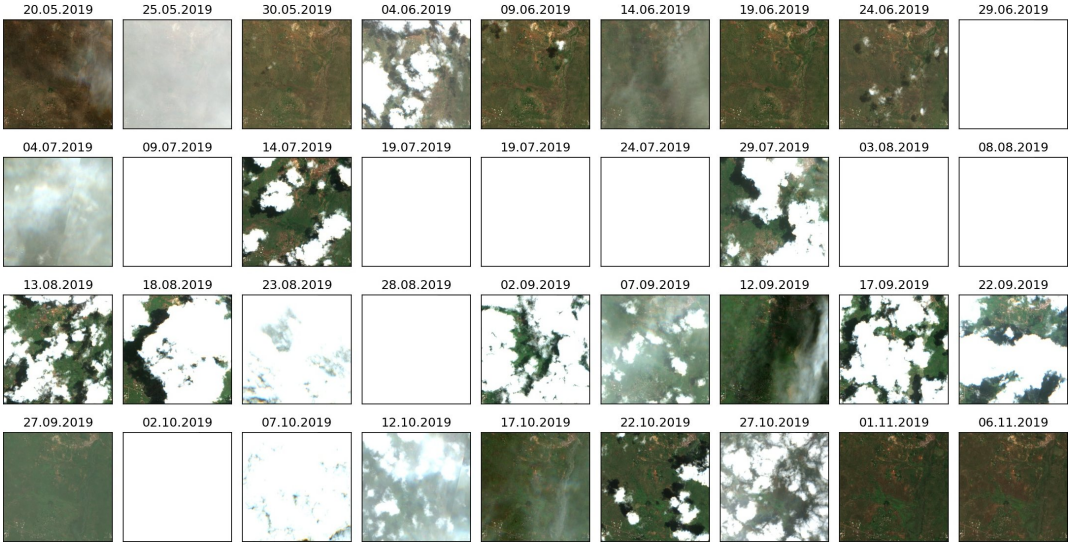


Figure 1.1: Sentinel-2 RGB images of an area next to the city Gaoua in southern Burkina Faso. Clouds prevent the monitoring of the area between June to October, the rainy and single growing season in that region (Belesova et al. 2017). Completely white images are fully covered with clouds.

Synthetic aperture radar (SAR) sensors offer a potential solution to this dilemma. They utilize microwave radiation, which is capable of penetrating cloud cover and are active sensors, enabling the imaging of the Earth's surface during both day and night. While SAR sensors offer great advantages, they use a different part of the electromagnetic spectrum and sense different properties of a surface than optical sensors. This leads to some challenges for their use: relationships and algorithms designed for optical data need to be redeveloped and adjusted to accommodate SAR data. Furthermore, the more complex processing of SAR data and its less intuitive interpretation hinder the adoption and usage by researchers and practitioners. Finally, no long-term data archives of easily comparable SAR data exist limiting the use for long-term studies.

The limitations of both optical and SAR sensors underscore the necessity of a combined approach to leverage their respective strengths while addressing their respective shortcomings. For vegetation monitoring, one way to achieve this is by using the optical NDVI as a proxy of vegetation status and translating SAR data into NDVI data. This integration preserves the usage of the established NDVI while mitigating the impact of cloud-induced data gaps, thereby facilitating more reliable and continuous monitoring of vegetation dynamics.

1.2 Objectives and structure of this thesis

The primary goal of this dissertation is to demonstrate that NDVI data availability can be improved by translating SAR backscatter into NDVI values, ultimately improving vegetation and agricultural monitoring. The first two chapters, Chapter 2 and Chapter 3, provide an overview of the essential knowledge required to understand the proposed approach, as well as related research. Chapter 4 analyses the relationship between SAR backscatter and NDVI values, thereby establishing the foundation of the proposed approach. Having established a relation, NDVI images are estimated using SAR images of a single date (Chapter 5), before using this and combining dense SAR-estimated NDVI time series with the sparser, but accurate, optical NDVI time series. The result are dense, cloud-independent time series (Chapter 6). The thesis is discussed in Chapter 7, followed by a summary and suggestions for future work in Chapter 8.

2 Space-Borne Remote Sensing of Vegetated Surfaces

What you'll learn This chapter provides a brief overview of the fundamental concepts that are necessary to understand the work described in this thesis. It covers remote sensing of the Earth's surface using optical and microwave sensors, the derivation of vegetation indices from this data, and the principle and architecture of deep learning methods.

The Earth's surface is highly dynamic, constantly transforming due to natural processes and human activities. Remote sensing provides a powerful tool to capture and monitor these changes systematically and quantitatively. This chapter aims to provide a comprehensive understanding of the fundamentals of remote sensing, with a particular focus on the principles of optical and synthetic aperture radar (SAR) imaging, as well as the deep learning techniques that will be employed in the subsequent chapters.

Remote sensing encompasses a variety of technologies designed to sense various properties of the Earth's surface. By utilizing different parts of the electromagnetic spectrum, these technologies provide crucial data that can be used to monitor vegetation health, assess land use changes, and evaluate environmental conditions.

The electromagnetic spectrum available for remote sensing spans from visible light to infrared (IR) and thermal waves and additional microwave frequencies. While the sun emits some ultraviolet (UV) light, this cannot be used for remote sensing, as the atmosphere and more specifically its oxygen and ozone, absorb it. Similarly, large parts of the infrared (IR) and long radio waves do not penetrate through the atmosphere. The parts of the electromagnetic spectrum where waves can pass through the atmosphere are called atmospheric windows. They are depicted in Figure 2.1. For these windows, different sensors can capture various features of the Earth's surface. This thesis focuses on images captured in optical light (visible, near-infrared (NIR), and shortwave infrared (SWIR)) with wavelengths from 400 to 2200 nm and radar frequencies (with wavelengths from 1 to 100 cm).

Data is acquired using space-borne platforms which allow global coverage and consistent monitoring. Space-borne sensors can provide data at various spatial, temporal, and spectral resolutions, making it possible to effectively monitor and analyze vegetated surfaces across the globe. For the work in this thesis, imagery with

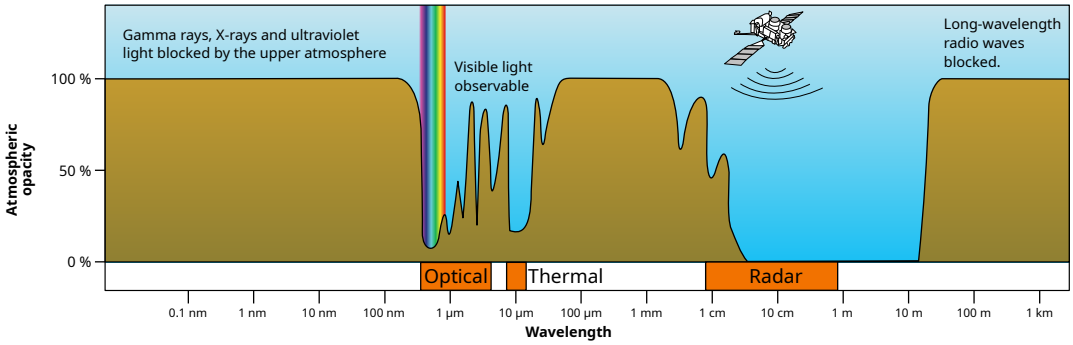


Figure 2.1: Opacity of Earth's atmosphere for electromagnetic waves and the atmospheric windows used for remote sensing. Three areas of the spectrum are used, which are denoted with orange: optical and infrared radiation with wavelengths from 400 to 2000 nm, thermal radiation of the Earth around 10 µm, and microwaves with wavelengths of 1 to 100 cm for radar sensors.

a medium spatial resolution (on the order of tens of meters) and a medium temporal resolution (with revisits every few days) are used.

2.1 Optical Sensors and Data

Optical sensors are fundamental tools in remote sensing, capturing the amount of sunlight reflected by the Earth's surface. Since they are passive sensors they rely on the sun to illuminate the Earth. One important property characterizing them (next to spatial resolution and temporal revisit) is their spectral resolution and their ability to detect light at different wavelengths. Thereby, they can be categorized into panchromatic, multispectral, and hyperspectral sensors with one, tens, and hundreds of spectral bands, respectively.

Optical sensors are designed to operate within a specific portion of the electromagnetic spectrum, which encompasses the range of wavelengths from ultraviolet to infrared light. The visible light with a wavelength from ~400 nm to ~750 nm, encompassing the colors blue, red, yellow, and green, is commonly used. In addition, near infrared (NIR) and short wave infrared (SWIR) light is of significant value for the assessment of vegetation and its moisture, and therefore it is frequently captured by optical sensors. Their wavelengths go from ~750 nm to ~1000 nm and from ~1000 nm to ~2500 nm, respectively. The boundaries between these ranges are gradual and there is some overlap between them. These spectral bands, in conjunction with the opacity of the atmosphere, are illustrated in Figure 2.2.

The sensed signal is dependent on various factors, mainly absorption, scattering, and reflection. As the light has to pass through the atmosphere, the atmospheric

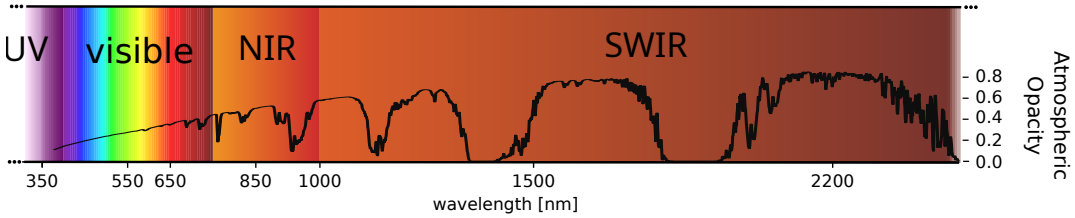


Figure 2.2: Part of the electromagnetic spectrum that is utilized by optical sensors. It shows the wavelengths for ultraviolet, visible, NIR, and SWIR light. The opacity of the atmosphere is included.

window, for which the absorption by atmospheric gases is minimal enough for signal capture from space, is important. Therefore, some wavelengths are unsuitable for space-borne earth observation. On the Earth's surface, each material reflects and absorbs light differently, which is captured in the spectral reflectance curves of each material as shown exemplarily for water, soil, and green vegetation in Figure 2.3. Therefore, having information about the reflectance of a surface for different wavelengths allows to distinguish surface types. Thereby, different spectral bands can capture complementary information about each surface. To illustrate this, an example is given in Figure 2.4: combining different spectral bands allows one to visualize different properties of a surface.

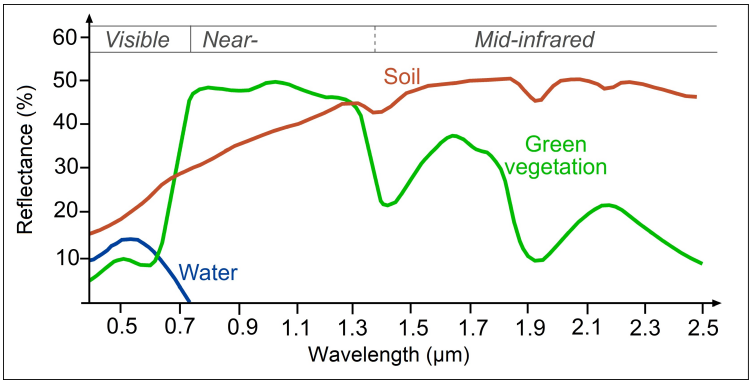


Figure 2.3: Spectral reflectance curves of water, green vegetation, and soil (Flores et al. 2019).

Optical sensors on satellite platforms typically follow sun-synchronous orbits, passing over the same part of the Earth at roughly the same local solar time. This allows a simple comparison of images taken on different days without having to correct for widely varying lighting conditions. The overfly time at the equator is usually before noon, a time chosen to balance good illumination conditions with minimal shadows and decreased cloud and haze built up compared to noon or afternoon.

The raw data captured by optical sensors undergo several processing steps to become

useful for analysis. The first step is to include radiometric corrections to address sensor-specific anomalies, geometric corrections to align the image with geographic coordinates, coregistration of different spectral bands, and conversion from digital numbers (DNs) to reflectance values. This results in Level-1C data also referred to as top of atmosphere (TOA) reflectance values. To acquire surface reflectance values (also called bottom of atmosphere (BOA) reflectances or Level-2A data), additional processing steps are applied. This primarily corrects for the influence of the atmosphere, which accounts for the absorption of water vapor, dust, and other molecules, but can also include corrections for multi-path scattering.

Across one image and for images of different acquisitions, the angle between the surface, sun, and sensor is changing. As most surfaces are anisotropically and not Lambertian, e.g. the reflectance is dependent on the sun-target-sensor geometry, the same surface appears different across one image or for images of different acquisitions. This geometry dependence is described by the bi-directional reflectance distribution function (BRDF) and can be corrected, which results in synthetic nadir BRDF-adjusted reflectance (NBAR) values.

2.2 Microwave Sensors and Data

Radar (radio detection and ranging) sensors rely on microwave radiation instead of visible or infrared light like optical sensors. Microwave radiation has a wavelength considerably longer than optical light and therefore almost no interaction with clouds or rain which allows weather-independent usage. Additionally, they are active sensors that emit electromagnetic waves and capture the returning signals. Therefore, they don't rely on the sun's illumination and allow imaging day and night.

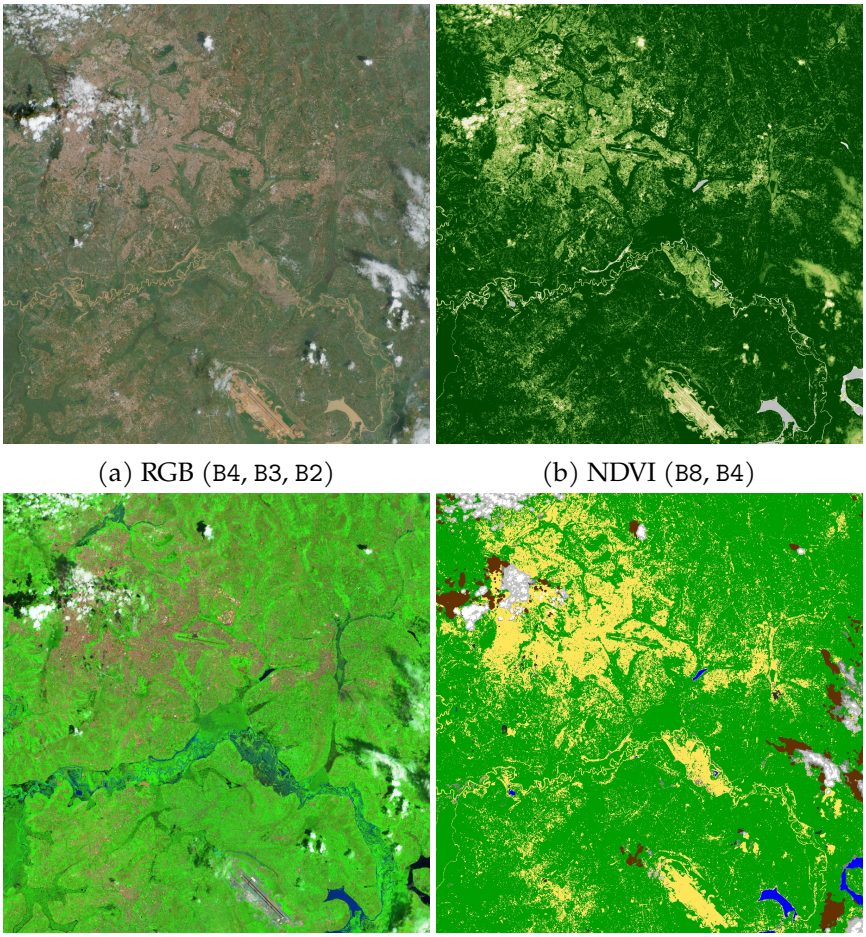
Synthetic aperture radar (SAR) sensors are a special kind of radar sensor, which synthesizes a large antenna by using a moving platform and appropriate signal processing to achieve a high resolution. The resolution in azimuth direction δ_a (also called along-track direction) is only dependent on the antenna aperture in azimuth direction D_a for these systems:

$$\delta_a = \frac{D_a}{2}. \tag{2.1}$$

The slant range resolution δ_r (or across track resolution) is inversely related to the transmitted signal duration τ or directly related to the bandwidth B used for transmission:

$$\delta_r = \frac{c \cdot \tau}{2} = \frac{c}{2B}. \tag{2.2}$$

The relationship between pulse duration and bandwidth can be explained using the Fourier analysis, which shows that a shorter pulse duration τ corresponds to a broader bandwidth B, and vice versa.



(a) RGB (B4, B3, B2) (b) NDVI (B8, B4)

(c) SWIR false color (B12, B8A, B4) (d) Scene classification layer

Figure 2.4: Different visualizations of multispectral optical data of a scene around Kigali (Rwanda) on 2024-01-27. Located in the upper left is Kigali, obscured by some clouds. Sentinel-2 data is used and B3, ... denote the used band. The used data includes a scene classification layer, classifying most of the area as vegetated, or non-vegetated, but also denoting clouds, and their shadows as well as water.

Satellites equipped with SAR sensors are typically in sun-synchronous orbits, positioned along the day-night boundary to ensure that the solar panels are always illuminated, meeting the high power requirements of the sensors. This yields local overpass times in the morning and evening.

Because SAR sensors are side-looking the slant range resolution given in Equation 2.2 is not the same as when the image is projected onto the ground, which is illustrated in Figure 2.5. Instead, the slant range resolution δ_r and the ground (projected) range resolution δ_g are related through simple trigonometry and the incidence angle θ_i.

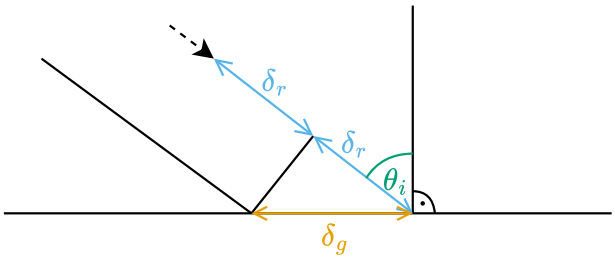


Figure 2.5: Relationship between slant range resolution δ_r and ground projected range resolution δ_g. θ_i denotes the incidence angle.

The ground range resolution is always lower than the slant range resolution and typical incidence angles of 20°, 30°, or 40° result in a 2.9, 2, or 1.6 times higher ground range resolution.

2.2.1 Factors Affecting SAR Backscatter

The radar response of a surface depends on numerous factors that collectively influence its appearance in a SAR image. Key determinants include roughness and dielectric properties of a surface, as these define the predominant interaction of electromagnetic waves with it. The difference in dielectric constants between two mediums dictates whether transmission or reflection, also termed scattering, predominates. A higher difference in dielectric constants results in more energy being reflected at the boundary, with less energy penetrating into the second medium. In the context of remote sensing sensors, where the initial medium is air, the second medium is determining which effect dominates. As such, the second medium's composition is the key factor of the received radar signal. Thereby, water and the water content play a crucial role for most natural surfaces like vegetation or soil, as it strongly influences the dielectric constant of the surface. For instance, almost all signal energy is reflected from water surfaces or wet soils, whereas in contrast for dry soils more signal energy enters the soil and is scattered there. Absorption is another factor that affects radar response; it converts radar energy into thermal energy rather than reflecting it back to the sensor, leading to attenuation of the radar signal.

Surface roughness and orientation also determine the type of scattering that occurs. In general, rough surfaces reflect more energy back to the sensor, while smooth surfaces reflect more energy away from the sensor, which is illustrated in Figure 2.6. This can be observed well for water surfaces: a calm sea leads to virtually no signal detected while for a rough sea the sensor can receive a response.

If there is not a simple surface and the transmitted signal is scattered multiple times before returning to the sensor, two other scattering types can occur. Double bounce occurs when the signal is scattered off two surfaces before returning to the sensor.

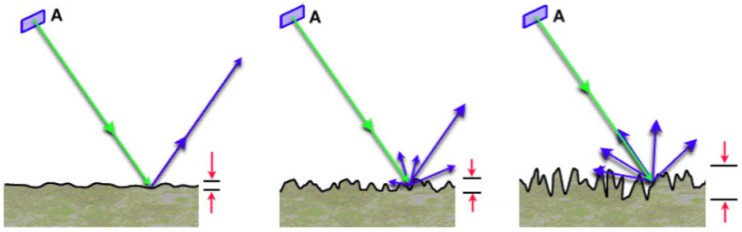


Figure 2.6: Influence of the surface roughness on radar scattering. Smooth surfaces (left) behave like specular mirrors, reflecting electromagnetic waves away from the sensor. Conversely, for intermediately rough (middle) and rough surfaces (right) the sensor can receive a scattered signal. The categorization of a surface as smooth or rough is influenced by the radar wavelength and the incidence angle. Figure from Flores et al. (2019) and modified.

This mainly occurs for man-made buildings like in urban areas, but also for inundated vegetation, where the water surface acts as a mirror and standing vegetation like tree trunks reflects the signal. Another important scattering type is volumetric scattering. This occurs when the radar signal is scattered numerous times in a complex manner, for instance in vegetation. There, not a single medium border is present, but instead a multitude of differently oriented surfaces at different positions. Therefore, radar waves are scattered multiple times at each air–leaf surface before eventually being received by the sensor. This type of scattering redistributes the radar energy in various directions, often resulting in complex signal returns. Additionally, depolarization occurs, where the polarization of the incoming wave is changed to another one. The different types of scattering are shown in Figure 2.7. Other special cases exist, like facade, Bragg, or dihedral corner reflector scattering, but they are in general not relevant, especially for vegetation monitoring.

Another important sensor parameter influencing the radar response is the used radar wavelength. Different frequency ranges of the electromagnetic spectrum are divided into frequency bands, with commonly used ones for SAR systems listed in Table 2.1. The wavelength determines how opaque or transparent objects appear, because many objects and radar waves are similar in size, ranging from centi- to decimeters. For instance, X-band radar waves interact directly with the uppermost leaves of a tree because of the short wavelength of 3 cm, whereas C-band waves already penetrate deeper into a tree and interact with branches. L- or P-band waves interact less with leaves and the tree canopy and more with the tree branches and trunk. This is illustrated in Figure 2.8.

Another factor significantly affecting the received radar response is the polarization of the radar waves, which determines the orientation in which the electromagnetic waves are transmitted and received. SAR sensors predominantly transmit linear polarized waves – either horizontal (H) or vertical (V) – though they can also emit circular

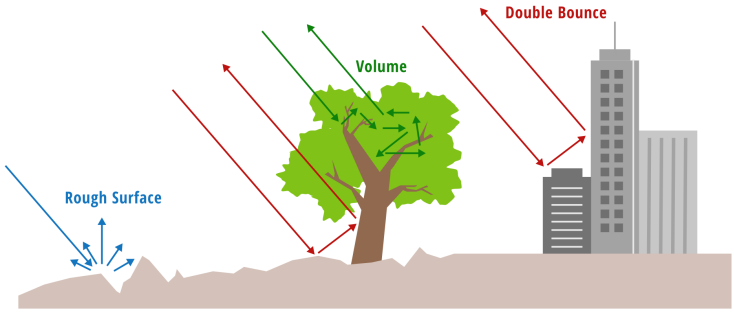


Figure 2.7: Scattering types of radar waves that are received back at the sensor. For non-vegetated surfaces, rough surface scattering occurs, whereas for vegetation and trees volumetric scattering occurs. Double-bounce scattering is dominant for urban areas and man-made structures but also happens for tree trunks. Figure from Flores et al. (2019).

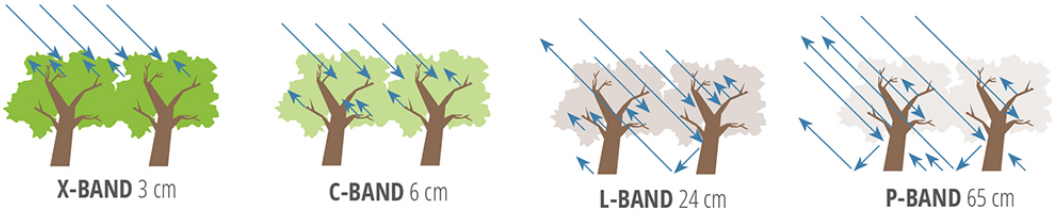


Figure 2.8: Illustration of the penetration depth of radar waves into trees in dependence on the used wavelength (Flores et al. 2019).

polarized waves. During reception, SAR systems can differentiate between the two polarizations, resulting in data that can be classified into four types: VV (vertical transmit, vertical receive), VH (vertical transmit, horizontal receive), HH (horizontal transmit, horizontal receive), and HV (horizontal transmit, vertical receive). When data includes all four polarizations, it is referred to as quad-polarized. Data with only two polarizations, typically featuring one transmit and two receive polarizations such as VV and VH or HH and HV, is labeled dual-polarized.

The interaction of radar waves with different surfaces varies depending on the polarization. Bare surfaces typically maintain polarization, showing little depolarization, whereas vegetated surfaces often exhibit significant depolarization, altering the polarization state of the waves. Additionally, the orientation of vegetation affects the radar response. For instance, narrow-leaf crops such as wheat or rice interact differently with horizontal and vertical waves due to their alignment relative to the wave direction, either perpendicular or in parallel (Liu et al. 2016).

Band	Frequency (GHz)	Wavelength (cm)
P	0.3 – 1	30 – 100
L	1 – 2	15 – 30
S	2 – 4	7.5 – 15
C	4 – 8	3.8 – 7.5
X	8 – 12	2.4 – 3.8

Table 2.1: Commonly used frequency bands of SAR systems (Flores et al. 2019, p. 29). Note that these bands are not rigorously defined and can be slightly different for other disciplines.

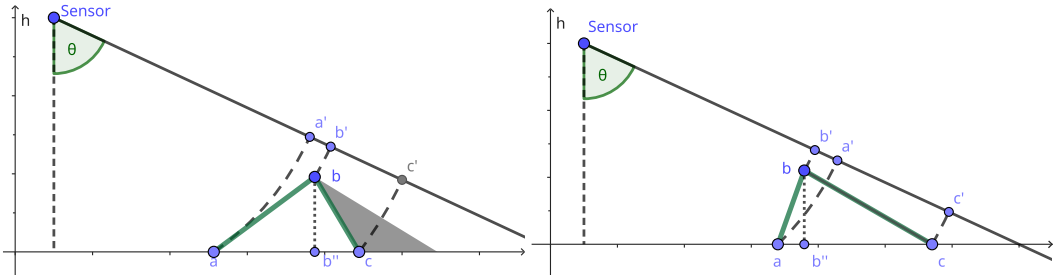
2.2.2 Terrain-Induced Effects

Because SAR sensors are side-looking, the terrain induces three effects: shadow, layover, and foreshortening. These phenomena, while primarily relevant in urban environments, are also prevalent in mountainous regions. A not-so-steep mountain is affected by foreshortening, where sensor-facing slopes appear shorter and slopes facing away from the sensor appear longer in the SAR data. If a mountain slope is steeper, then layover can occur, where the radar signal from the top of it returns to the sensor before the signal from the base. This causes the top to appear 'laid over' towards the sensor and one SAR pixel contains information of two disparate points on the surface. For even steeper terrain, shadowing can occur when radar waves are blocked by the terrain, creating areas with no return signal, similar to how sunlight can cast shadows. All these effects are illustrated in Figure 2.9. Foreshortening and shadows are also visible in Figure 2.13d: the mountain peaks appear very bright on one side but some have areas without any backscatter on the other side. Collectively, these effects can complicate the interpretation of SAR imagery in areas with topography.

2.2.3 Statistics of SAR Backscatter and Speckle

The backscatter of a resolution cell is composed of the contributions from all individual scatterers within it. For instance, in a forested area, these scatterers include individual trees, leaves, and branches. The total signal received at the sensor is the coherent (or complex) sum of all these scatterers' contributions. The phase of the individual scatterers is random, leading to both constructive and destructive interference. This randomness results in a phenomenon known as the random walk, depicted in Figure 2.10, where the variability in the coherent sum of scatterers produces speckle.

Speckle manifests as high variability in radar backscatter among neighboring pixels that represent the same surface. When the resolution cell includes a large number of



(a) Foreshortening and shadow (b) Layover

Figure 2.9: Effects of the terrain on the acquired SAR data due to the side-looking geometry. (a) Areas inclined towards the sensor can lead to foreshortening, where the objects a, b appear closer in the SAR image a', b'. Shadow leads to points not being captured, like c, which does not appear in the SAR image. (b) Layover occurs for steep terrain, where the top b appears before the bottom a in the SAR image. b'' denotes the position of b in an Nadir optical image. θ is the incidence angle.

scatterers, such as in Sentinel-1 acquisitions over natural scenes with soil, rock, or vegetation, fully developed speckle is observed. In this case, the received amplitude at the sensor follows a Rayleigh distribution, while the phase follows a uniform distribution. These distributions are illustrated in Figure 2.11. As the phase in a single acquisition is uniformly distributed, it contains no useful information. For urban scenes or very high-resolution data the initial assumption does not hold, but instead, single scatterers become dominant, leading to partially developed speckle.

When the amplitude values are converted to a logarithmic scale expressed in decibels, the Rayleigh distribution is transformed into a distribution approximately following a normal distribution.

Speckle provides information about the structure within a resolution cell if the underlying backscatter distribution is obtainable. However, the estimation of this distribution is not feasible using a single pixel, which presents a challenge for the analysis of the sub-resolution structure of a pixel. Only by examining the backscatter distribution over an area of the same structure, or a pixel over time, can the distribution be assessed. In practice, it is difficult to determine a priori which areas have similar characteristics. Consequently, speckle complicates data analysis and appears as noise that degrades image quality. To reduce the impact of speckle, multi-look processing or spatial filtering can be employed which spatially averages the data.

2.2.4 Interferometric SAR (InSAR)

Interferometric SAR (InSAR) is a technique to utilize the phase information of two SAR acquisitions taken at slightly different positions. Depending on the position

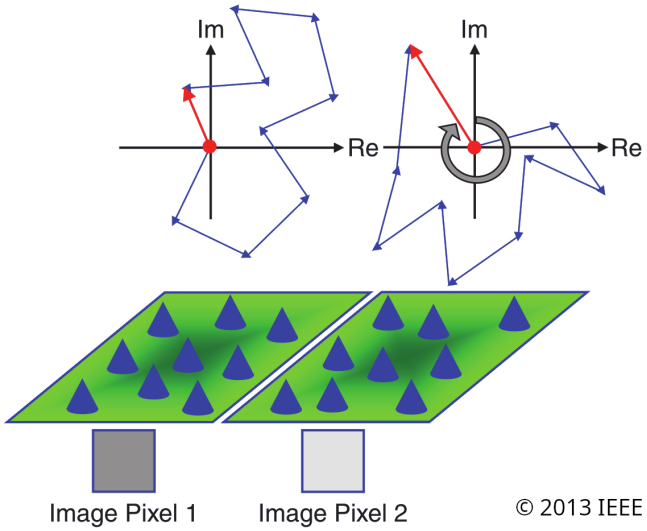
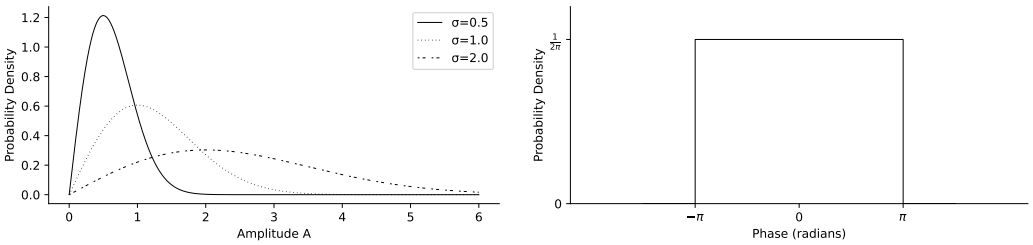


Figure 2.10: Illustration of the formation of speckle (Moreira et al. 2013). The response of a resolution cell is the interference of all the single scatterers within it, which is mathematically expressed as the complex (or coherent) sum of all contributing signals. Each single contribution (blue arrows) has a random phase and the resulting amplitude (red arrow) can therefore be highly variable between pixels of the same surface.

relative to each other, one can differentiate along- and across-track interferometry. For along-track interferometry, the receiving antennas are positioned parallel to the flight track, which allows to sense motion on the ground, for example, river or ocean currents. In contrast, across-track interferometry has the SAR data acquired from two positions perpendicular to the flight path. This allows to retrieve information about the terrain.

Across-track interferometry is based on the fact that the phase difference between two acquisitions $\Delta\phi$ is related to the difference in the distance to the target Δr (Klausing et al. 2000, p. 276). If the position where the images were taken is exactly known, this allows to calculate the height of the captured area. A similar approach to estimate 3D positions or the terrain is possible with optical data, termed (stereo) photogrammetry. In contrast to interferometric SAR, not the phase difference is utilized, but instead the small angular shift for a target between the two images.

There are two options to acquire the data at two positions to allow InSAR processing. Either two antennas are carried by the sensor (single-pass interferometry) or a sensor with a single antenna has to fly the same flight pass two times (repeat-pass interferometry). Single-pass interferometry allows a higher quality of the phase measurements, however at an increased hardware cost and sensor size. Therefore,



(a) (b)

Figure 2.11: Probability density functions of the (a) amplitude and (b) phase of a pixel exhibiting fully developed speckle. σ gives the radar cross-section of the area, indicating how strongly the surfaces scatter the signal back to the sensor. While the amplitude is Rayleigh distributed, the phase is uniformly distributed.

repeat-pass interferometry is often used for space-borne sensors because it requires only a single antenna.

In the context of repeat-pass InSAR, spatial and temporal baselines are critical parameters that affect the accuracy and reliability of the interferometric measurements. Thereby, the spatial baseline refers to the physical separation between the two antenna positions and the temporal baseline to the time interval between the two SAR acquisitions. The selection of the spatial baseline is a trade-off between sensitivity and robustness. A larger spatial baseline increases sensitivity, as the phase change is larger for terrain height changes. However, if the phase changes too quickly, it can lead to unwrapping errors, where the height change between pixels cannot be used to accurately estimate the terrain because the height changes too quickly. The temporal baseline is usually desired to be short. This allows for a higher quality of the measure phase difference because the surface changes less between the acquisitions. However, for space-borne platforms, the temporal baseline is fixed by the orbit. Only sensor constellations allow a selection of the temporal baseline.

In order to achieve high-quality elevation data, it is essential that the positional difference of the antennas results in a change of the phase, and that no other factors influence the measured phase difference. In practice, however, a multitude of additional factors influence the phase, resulting in the introduction of noise into the measured phase difference. To evaluate the quality of the phase measurement, the coherence γ is employed. A coherence value of 1 indicates an optimal phase measurement, whereas a coherence value of 0 indicates a lack of correlation between the signals, rendering them unusable.

The overall coherence γ of a pixel is influenced by several contributing factors. In repeat-pass InSAR systems, temporal coherence γ_{temporal} represents a significant contributor to the loss of coherence. It is caused by changes in the surface between

the two acquisitions. In the case of vegetated surfaces, this is a particularly significant factor, given that the vegetation is subject to fluctuations caused by wind and growth, as well as other changes. An additional source of decorrelation is the thermal noise of the SAR system, which is expressed by the thermal coherence γ_{thermal}. It is determined by the signal-to-noise ratio of a resolution cell; thus, a pixel with high backscatter exhibits less decorrelation. The geometric coherence γ_{geom} accounts for changes due to the slight discrepancy in the viewing angle of the two antennas. However, it can be significantly reduced in practice through the use of a-priori filtering. As radar waves are capable of penetrating into a volume, mainly for vegetation and snow, the phase of the two acquisitions can differ. This also leads to decorrelation of the phase measurement and is expressed by the volume coherence γ_{vol}.

Not all factors of the coherence can be directly measured, therefore γ has to be estimated. In practice, many methods consider small regions and the local neighborhood of a pixel to estimate the coherence. This leads to coherence maps with a resolution several times worse than the original SAR resolution. Recent deep learning approaches aim to mitigate this limitation, providing coherence estimates at the original SAR resolution (Sica et al. 2021; Geara et al. 2024).

While InSAR is a powerful technique, several factors limit its usability for large areas or multiple scenes. These include the complexity of the method, the high computational resources required, and for repeat-pass interferometry, the sensitivity to temporal decorrelation. Additionally, not all missions have an orbit controlled precisely enough to allow repeat-pass InSAR analysis for all acquisitions. Instead, only by chance a suitable spatial baseline is achieved.

An in-depth explanation of InSAR methods can be found in Hanssen (2001).

2.2.5 Processing SAR Data and Backscatter Normalization

Processing SAR data to acquire an image suitable for analysis involves many different steps. Raw data first needs to be focused, which results in a single-look complex (SLC) image in the slant range. Thereby, for each pixel, a complex value is stored allowing to retrieve phase ϕ and amplitude A of the received signal. For the work done in this thesis, only the amplitude A or the intensity $I = A^2$. The intensity is also referred to as detected values.

The radar backscatter β describes the ratio of power received by a surface and the power scattered back to the sensor. As the received energy is dependent on the observed area, the energy is normalized by the area. This allows for comparable values between images. This results in the backscatter *coefficient*. Depending on which area is used for this normalization, three backscatter coefficients β°, σ°, and γ° are formed. The first uses an area in the slant range and results in the radar brightness

or beta naught backscatter β° (Raney et al. 1994). For typical SAR data, the DNs of SAR data refer to the backscatter amplitudes, so β° intensities are calculated as

$$\beta^\circ = \frac{\beta}{A_\beta} = \frac{\mathrm{DN}^2}{A_\beta} \tag{2.3}$$

with A_β being calculated using the azimuth and range resolution. As the slant range geometry is used, this does not require any knowledge about the position of the sensor relative to the surface.

Sigma naught σ° and gamma naught γ° both rely on information about the terrain. The terrain information can either come from the reference ellipsoid or the local terrain. For the flat earth model using the reference ellipsoid, only the sensor position relative to the Earth is required, while a precise DEM is needed when the local terrain should be used. Sigma naught σ° uses the ground area tangent to the terrain, the dashed rectangle A_σ in Figure 2.12:

$$\sigma^\circ = \frac{\beta}{A_\sigma} = \frac{\mathrm{DN}^2}{A_\sigma} = \beta^\circ \sin\theta. \tag{2.4}$$

If the reference area is defined to be in the plane perpendicular to the line of sight from the sensor, the dotted rectangle A_γ in Figure 2.12, gamma naught backscatter γ° is obtained:

$$\gamma^\circ = \frac{\beta}{A_\gamma} = \frac{\mathrm{DN}^2}{A_\gamma} = \beta^\circ \tan\theta. \tag{2.5}$$

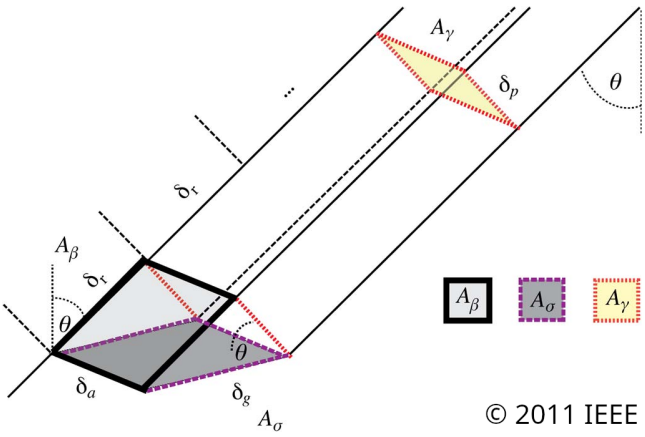


Figure 2.12: Illustration of different alternatives to normalize the backscatter β (Small 2011). $A_\beta, A_\sigma, A_\gamma$ describe the reference areas to form $\beta^\circ, \sigma^\circ$ and γ° backscatter coefficients, respectively. θ denotes the incidence (or off-nadir) angle of the resolution cell. The resolution is given in slant range δ_r, ground range δ_g, and azimuth δ_a.

The side-looking geometry leads to some distortions, as explained in the sections above and illustrated in Figure 2.9. To correct for these, different processing steps have to be applied. Firstly, the targets are displaced due to foreshortening and their location needs to be corrected. This is done during geocoding, where the slant range pixels are projected onto the ground, which is also referred to as (geometric) terrain correction (TC). However, foreshortening and layover also lead to multiple slant range pixels mapped to a single ground range projected one and therefore, appearing too bright (Small 2011). This is corrected using terrain flattening (TF). Data with both TC and TF applied are referred to as radiometric terrain corrected (RTC) data.

A comparison of a SAR scene with different corrections applied (TC and TF) and using different backscatter normalizations ($\beta^\circ, \sigma^\circ, \gamma^\circ$) is depicted in Figure 2.13.

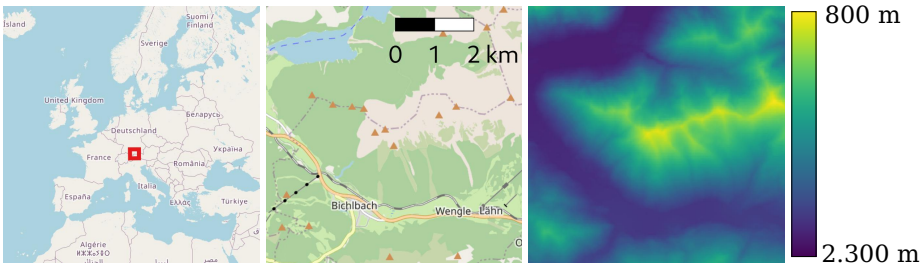
2.3 Vegetation Indices

Vegetation indices (VIs) are a common tool for vegetation monitoring, providing a simplified yet powerful means of assessing and analyzing vegetation health and cover. They condense information from multiple spectral bands into single values, making interpretation and analysis both easier and more effective. VIs leverage the different behaviors of surfaces when exposed to various wavelengths of electromagnetic radiation. For example, healthy vegetation typically reflects more near-infrared light and less visible light, while bare soil or stressed vegetation has a higher reflectance for visible light (cf. Figure 2.3). By combining these spectral responses into a single index, researchers can gain insights into vegetation conditions that are not apparent from individual spectral bands. The most commonly used VI is the normalized difference vegetation index (NDVI), but next to that, a myriad of other VIs exists. VIs are also not limited to optical data, but combinations for SAR data also exist.

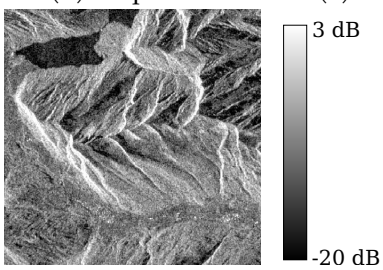
2.3.1 Normalized Difference Vegetation Index (NDVI)

The NDVI is a widely recognized and utilized metric to assess the vegetation state. It is based on the characteristic spectral reflectance of green vegetation: leaf pigments absorb most of the visible light with chlorophyll responsible for a high absorption of red light. In contrast, most of the near-infrared (NIR) light is reflected by green vegetation. The wavelength where the reflectance changes is called the red edge located around 700 nm (Albertz 2009, pp. 17–23) and apparent in Figure 2.3. Now, calculating the ratio of the reflectance of the red and NIR light results in an indicator, of how much chlorophyll a leaf and plant contains, how much photosynthesis it can do, and overall its health. The normalized difference between the two bands is formed as:

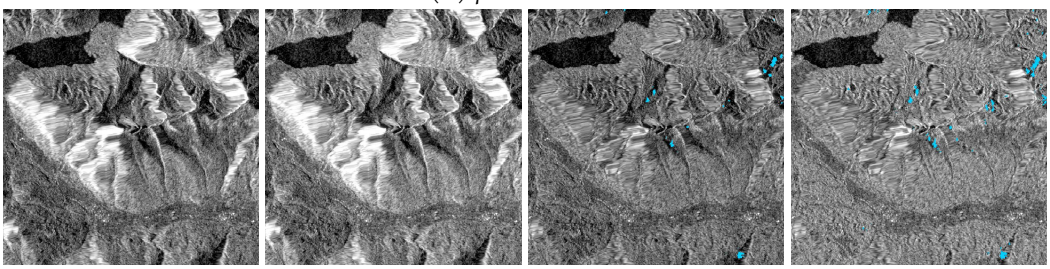
$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}} \tag{2.6}$$



(a) Location (b) Map view (c) Elevation



(d) β°



(e) σ°, TC (f) γ°, TC (g) σ°, TC & TF (h) γ°, TC & TF

Figure 2.13: Comparison of differently processed SAR data of a mountainous scene located in the Alps. (d) Data in slant range geometry with foreshortening clearly visible. (e, f) Geocoded data with terrain correction (TC) applied. (g, h) Radiometric terrain corrected data with TC and TF applied. All images show Sentinel-1 VV data in decibels. Missing pixels due to radar shadow are colored cyan .

Using a normalized difference instead of a plain ratio results in a more robust indicator, which allows comparison for example for different illumination conditions. One of the first to use this equation are Rouse et al. (1974). While the NDVI is a useful indicator of the "greenness" of vegetation, it is not based on any direct physical, biological, or chemical relationship. Still, the NDVI is closely linked to the chlorophyll content of vegetation as well as the amount of vegetation cover and a myriad of studies link the NDVI to plant parameters.

Using Equation 2.6 results in values ranging from -1 to 1. Values less than 0 typically indicate water, values between 0 and 0.25 correspond to bare soil or dry matter, values between 0.25 and 0.4 correspond to soil with some vegetation, and values greater than 0.4 are indicative of vegetated surfaces (Filgueiras et al. 2019). Due to the saturation of NDVI, areas with high NDVI can have very different appearances. Both a green meadow and a tropical forest can have NDVI values close to 1 despite the fact that they are significantly different and have largely different biomass.

The use of a normalized difference contributes to the robustness of the NDVI under varying lighting conditions. However, atmospheric conditions influence it, necessitating atmospheric corrections of the reflectance values before calculating the NDVI to ensure its accuracy (Moravec et al. 2021). Next to that also cloud or terrain shadows influence the NDVI (leading to a decrease) and the sun and looking angle, as it is somewhat sensitive to changes of the bidirectional reflectance distribution function (BRDF), albeit to a lesser extent (Shibayama et al. 1986).

2.3.2 Other Optical Multispectral Indices

Next to the NDVI a myriad of other multispectral indices exist. Some of them aim to improve upon the NDVI, while others have different uses altogether, like for water or vegetation moisture mapping.

The enhanced vegetation index (EVI) is an example of a VI trying to improve upon the NDVI. Its formulation includes blue light next to red and NIR light to reduce the influence of the soil background noise and to enhance sensitivity to dense vegetation. Two other VIs aiming to improve upon the NDVI are the modified chlorophyll absorption in reflectance index (MCARI), which tries to quantify chlorophyll content in vegetation, and the soil-adjusted vegetation index (SAVI), which tries to improve effectiveness in areas with significant soil exposure, such as sparsely vegetated regions.

For an easy detection of water surfaces the normalized difference water index (NDWI) and the modified normalized difference water index (mNDWI) can be used. The NDWI (McFeeters 1996) leverages the reflectances of green (G) and NIR light. Vegetation and soil exhibit a high reflectance in the NIR spectrum, whereas water demonstrates a low NIR reflectance and a high green reflectance. This allows to effectively

distinguish water bodies from vegetated and terrestrial surfaces using the normalized difference of green and NIR reflectances:

$$\text{NDWI} = \frac{G - NIR}{G + NIR}. \tag{2.7}$$

The mNDWI is an enhancement over the NDWI to especially improve discrimination between water and built-up land areas (Xu 2006). It utilized SWIR and NIR reflectances:

$$\text{mNDWI} = \frac{\text{G} - \text{SWIR}}{\text{G} + \text{SWIR}}. \tag{2.8}$$

To monitor the total amount of water in vegetation, the NIR and SWIR bands can be used to form the normalized difference moisture index (NDMI):

$$\text{NDMI} = \frac{NIR - SWIR}{NIR + SWIR}. \tag{2.9}$$

Combining NIR and SWIR reflectances into a VI was done by multiple researchers, therefore this is also known as normalized difference water index Gao (1996), land surface water index (Chandrasekar et al. 2010), leaf water content index (Hunt et al. 1987), or normalized burn ratio (NBR) (López García et al. 1991).

2.3.3 Radar Vegetation Indices

Indices can be formed using SAR data next to using data from optical sensors. Usually, different polarizations of a SAR acquisition are used, simply because data of multiple frequencies is hard to acquire with SAR sensors. Radar indices exploit the different responses of a surface for the different polarizations, similar to how the different spectral bands are exploited for optical VIs.

An alternative approach to radar indices involves the use of polarimetric decompositions of SAR data. These decompositions enable the isolation of different scattering mechanisms, such as volumetric and double-bounce scattering, by leveraging the comprehensive polarimetric information in the SAR signal. While some decompositions can be performed with just two polarizations, usually all four polarizations are required. The requirement for quad-polarized data, along with the significant computational demands for processing, has limited their widespread adoption. A detailed description of these decomposition methods can be found in Lee et al. (2017).

The simplest way to combine polarizations is to calculate the ratio between them, called the cross-ratio (CR). Typically, it is the ratio between cross- and co-polarized data:

$$\text{CR} = \frac{\text{co-pol}}{\text{cross-pol}}. \tag{2.10}$$

With the four polarizations VV, VH, HH, and HV data, this allows to form four different CRs: $^{HH}/_{HV}$, $^{HH}/_{VH}$, $^{VV}/_{HV}$, and $^{VV}/_{VH}$.

The radar vegetation index (RVI) is a more elaborate index combining three polarizations assuming $\sigma_{VV} = \sigma_{HH}$ (Kim et al. 2000):

$$\text{RVI} = \frac{8\sigma_{HV}}{\sigma_{HH} + \sigma_{VV} + 2\sigma_{HV}}. \tag{2.11}$$

RVI values range from 0 to 1 with higher values indicating a higher vegetation cover and water content (Flores et al. 2019).

As commonly used sensors only provide data of two polarizations with one transmit and two receive polarizations, the RVI has to be modified. Thereby it is assumed, that the cross- and the two co-polarized data are identical ($\sigma_{HH} = \sigma_{VV}$ and $\sigma_{VH} = \sigma_{HV}$) For VV/VH (Trudel et al. 2012) or HH/HV data (Nasirzadehdizaji et al. 2019) this results in

$$\text{RVI}_{VV} = \frac{4\sigma_{VH}}{\sigma_{VV} + \sigma_{VH}} \tag{2.12}$$

and

$$\text{RVI}_{HH} = \frac{4\sigma_{HV}}{\sigma_{HH} + \sigma_{HV}}. \tag{2.13}$$

The assumption, that co- and cross-polarized backscatter is the same, only holds for isotropic surfaces like forests or bare soil. However, for croplands it is not necessarily true, as they can have a strong dependence on the orientation and polarization of the incoming radar waves.

All of the indices above use backscatter data in linear scale. Instead of σ° backscatter, also γ° could be utilized.

The presented indices are just a selection of all proposed ones. But while there are many proposed SAR indices for vegetation monitoring, there is little research on their usage and application (Hu et al. 2024).

2.4 Auxiliary Geospatial Data

Next to SAR and optical data other geospatial products describe the surface of the Earth. Of particular interest for this dissertation are climate zones, elevation, and land cover/land use classifications.

The most commonly used climate classification is the Köppen-Geiger classification described for example in Peel et al. (2007). They categorize the Earth according to the vegetation that is growing in that region, which is determined by the average temperature as well as precipitation and their course throughout the year. Köppen-Geiger climate zones are encoded with two or three letters, describing one of the four main classes, a subclass, and potentially a variant. A map of these climate zones is given in Figure 2.14.

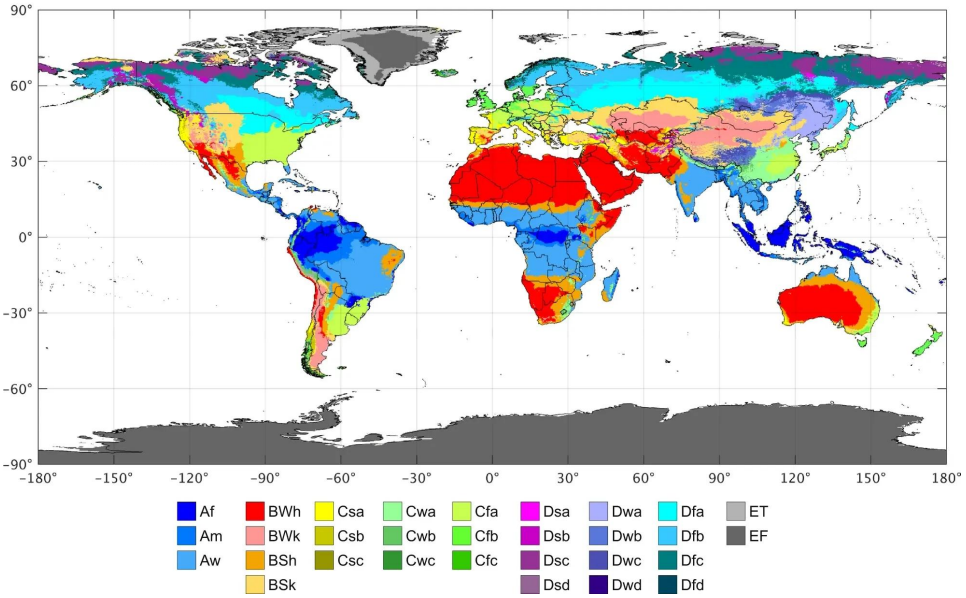


Figure 2.14: Köppen-Geiger climate zones of the Earth (Beck et al. 2018).

The elevation and terrain are represented by digital elevation models (DEMs). Thereby, digital surface models (DSMs) and digital terrain models (DTMs) can be differentiated, which express slightly different heights. DSMs express the surface elevation, which includes buildings, infrastructure, and vegetation whereas DTMs only capture the bare topography. The first (almost) global DEM was acquired during the shuttle radar topography mission (SRTM) and improved versions of this are still in use today. Other commonly used DEMs are the ALOS World 3D DEM (Takaku et al. 2020), the NASADEM (Buckley et al. 2020), or the Copernicus DEM[1]. An example of the data of a DEM is shown in Figure 2.15.

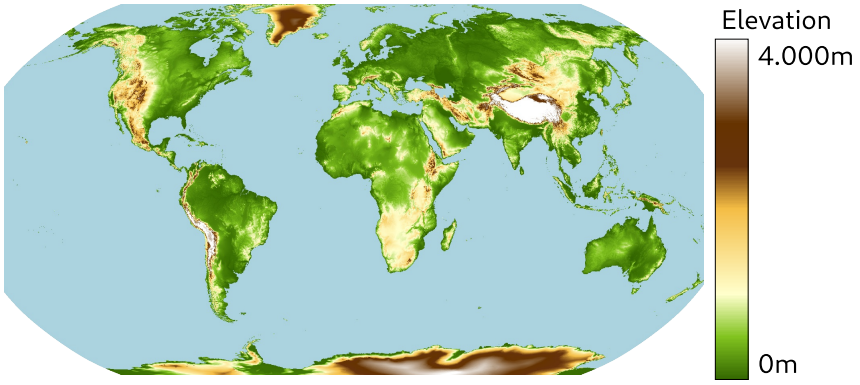


Figure 2.15: Elevation of the Earth of the Copernicus DEM.

[1] https://spacedata.copernicus.eu/collections/copernicus-digital-elevation-model

The surface of the Earth can be classified according to what is covering it and how it is used. Maps capturing this information are called land use/land cover (LULC) maps. As different applications require different classification schemes, a myriad of LULC maps exist. Three examples are the MODIS Land Cover Type map, the Copernicus Global Land Cover Layers, and the ESA WorldCover product. They are based on data from different sensors and therefore have different resolutions, ranging from 500 m and 100 m to 10 m. An example of the classifications of the ESA WorldCover map is given in Figure 2.16.

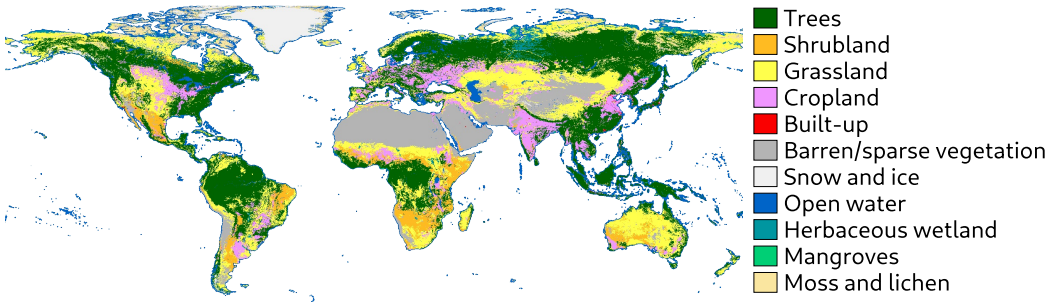


Figure 2.16: Visualization of the land cover classes of the ESA WorldCover v100.

2.5 Machine Learning and Deep Neural Networks

Machine learning (ML) focuses on developing algorithms that can identify patterns in data and make decisions based on those patterns. Essentially, ML is about allowing computers to come to decisions when provided with some inputs without explicitly defining rules to come to correct results. This is achieved by developing methods that can generate these rules themselves when provided with enough examples of input data and expected outputs.

Traditionally, machine learning techniques include models such as linear regression, decision trees, random forests (RFs), and clustering algorithms. All of these approaches can handle only a limited amount of input data, which often requires manually selecting features extracted from the data. For instance, not the raw pixel values of a field can be fed to ML models, but instead only statistical metrics like the mean or the standard deviation. This feature engineering is critical for machine learning models to learn patterns related to the problem at hand.

Artificial neural networks (NNs) are another machine learning method inspired by the structure and function of the human brain. A NN consists of interconnected nodes, or "neurons," that process information in a manner similar to biological neural networks. These nodes are organized into layers: the input layer receives input data, hidden layers transform the input through numerous computations, and the output

layer generates the result. NNs can be both a machine or deep learning method, which is determined by the number of nodes and therefore model parameters. Shallow NNs with one input, one hidden, and one output layer are considered a machine learning technique. In contrast, deep NNs include numerous hidden layers and a greatly increased number of parameters to optimize, thus they are a deep learning approach. The increased number of layers allow deep NNs to learn more complex representations of the data compared to shallow models.

Machine learning and deep learning have two primary distinctions: Firstly, the number of parameters, and secondly, the manner in which features of the data are extracted and learned. Traditional machine learning models with a smaller parameter number rely extensively on manually crafted feature extraction. The required feature engineering necessitates not only domain expertise and a considerable amount of trial and error but it is also a labor-intensive process. In contrast, deep learning models are capable of automatically learning feature representations directly from the raw data due to their highly increased parameter count. This capability is particularly advantageous when working with unstructured data, such as images and text, where patterns and features are complex, multidimensional, and not easy to extract.

Machine learning models are typically optimized by using a set of pairs of input data and expected labels. Training begins with the initialization of the model's parameters, which are often set to random values. The primary objective is to iteratively refine these parameters in order to ensure that the model's predictions are in close alignment with the expected outputs for a given set of inputs. In the case of supervised learning employed in this dissertation, the expected outputs are known and can be used as labels.

The difference between the predicted and actual outputs is quantified using a loss function. The loss function serves as an objective measure of the model's performance, and the ultimate goal is to minimize this loss through iterative adjustments of the model's parameters. Loss functions used for the work in this dissertation will be explained in Section 2.5.4 below.

To update the parameters of a NN, backpropagation is employed. This technique entails calculating the gradient of the loss function with respect to each model parameter, subsequently propagating these gradients from the output layer back to the input layer. These gradients indicate the direction and magnitude of change required to achieve a minimal loss. With these gradients, the parameters can be updated using gradient descent, an optimization method. By applying these steps multiple times, the model's performance is iteratively enhanced.

The architecture of a DL model is a determining factor in the model's ability to process different types of data effectively. The architecture of a NN is determined by a number of factors, including how the neurons of the network are connected, the depth and width of the layers, and the operations performed. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are among the most

commonly used architectures for processing spatial and temporal data, respectively. They are explained in the following two subsections.

A comprehensive examination of deep learning techniques along with their mathematical foundations and practical implementations is presented in Goodfellow et al. (2016).

2.5.1 Convolutional Neural Networks (CNNs) for Image Data

Convolutional neural networks (CNNs) are designed to efficiently process grid-like data such as images. They leverage convolutional operations, whereby the spatial neighborhood of each pixel is considered. Convolutional layers together with pooling layers allow the model to extract high-level features from images. Thereby, convolutional layers apply a set of filters to the input image, creating feature maps highlighting different image characteristics. The spatial dimensions of these feature maps are reduced using pooling layers to increase computational efficiency and learn complex patterns of larger areas. When an image should be classified, then usually fully connected layers are utilized. Other variations of the architecture allow to translate images, useful for instance to create pixel-wise segmentations. Prominent architectures in this domain include the U-Net (Ronneberger et al. 2015) or ResNets (He et al. 2016).

The operation of a CNN can be demonstrated using LeNet, an early and basic CNN designed to categorize images based on the number depicted in them (LeCun et al. 1998). LeNet is comprised of convolutional, max pooling, and fully connected layers, as well as activation functions. The convolutional layers utilize learnable filter kernels to extract features from the input image or a feature map. These filters move across the image, capturing features from all regions. This operation generates as many feature maps as there are filters, with each feature map having the same or smaller size than the initial input image. The output size depends on how the image borders are treated during the convolution process. Then, the extracted feature maps are passed to max-pooling layers, which reduce the spatial dimensions, effectively subsampling the images by selecting the maximum value from each region covered by the filter. This emphasizes the most prominent features while discarding less significant details. Subsequently, a non-linear activation function, such as the sigmoid or ReLU (rectified linear unit), is applied to introduce non-linearity and facilitate the learning of complex features. This process of convolution, followed by pooling and activation, is repeated until the feature maps have become too small to perform further convolutional operations. The two-dimensional feature maps are then flattened into one-dimensional feature vectors, preparing them for the fully connected layers. The fully connected layers use the extracted features of the whole image to classify it and generate the final prediction. An example of such a network and the output of each layer for an example input is shown in Figure 2.17.

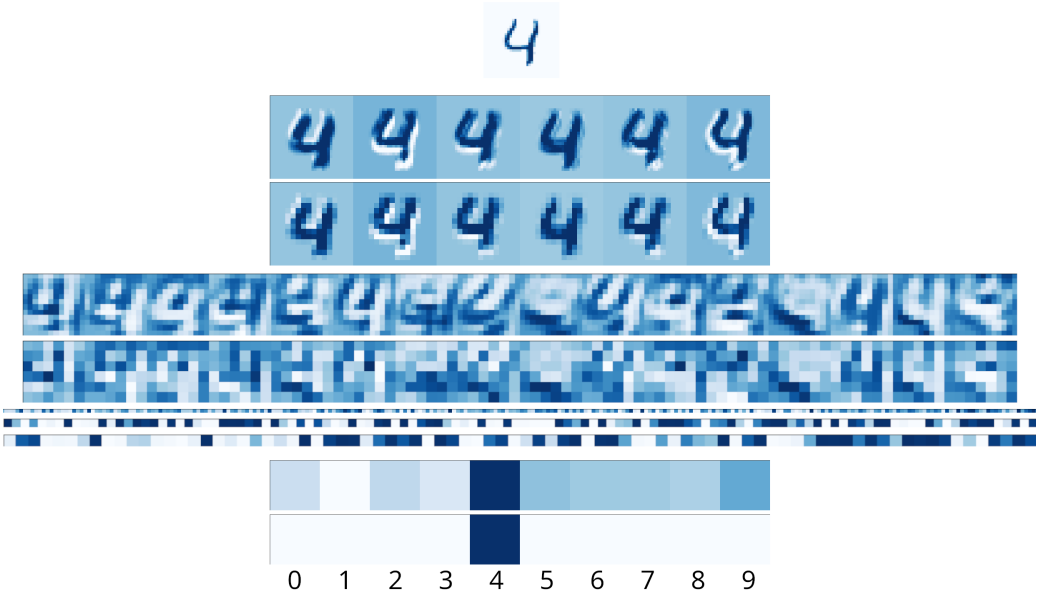


Figure 2.17: Visualization of the output of each layer of a LeNet CNN for an input image example depicting a "4". At the top, the input image is shown and below that the outputs of the convolutional, pooling, and fully connected layers. At the bottom is the model prediction. White to dark blue colors depict low to high values.

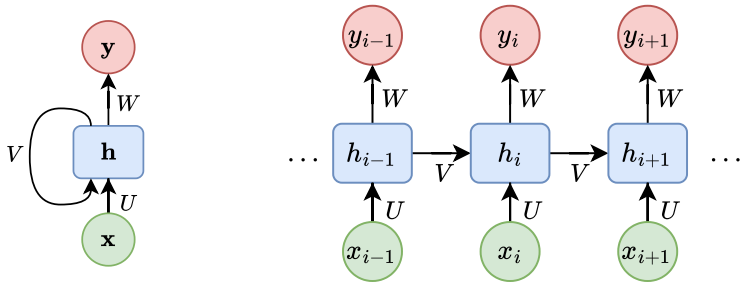
2.5.2 Recurrent Neural Networks (RNNs) for Sequence Data

Recurrent neural networks (RNNs) are a network architecture that can efficiently handle sequential data like time series. To efficiently extract patterns within a sequence they have a memory mechanism, which allows them to retain and combine information about different parts of the sequence.

The architecture of a simple RNN is relatively straightforward. At the core of this architecture is the hidden state vector $\mathbf{h}$, responsible for maintaining a running accumulation of relevant information when processing an input sequence $\mathbf{x}$ to produce an output sequence $\mathbf{y}$. The sequences $\mathbf{x}, \mathbf{y}, \mathbf{h}$ consist of vectors x_i, y_i, h_i for each sequence element. The processing facilitates three weight matrices: U, V, and W. Here, U serves to transform input vectors, V updates the hidden state, and W projects the hidden state to output vectors.

To process a sequence, the hidden state is initialized at first, using a random, zero, or learnable initialization. The model then sequentially processes each element x_i of the sequence. For each timestep i, a new hidden state h_i is calculated by integrating the current input x_i, the preceding hidden state h_{i-1}, and the weight matrices U and V. This integration is typically accompanied by the application of a nonlinear activation function such as sigmoid or tanh. From this new hidden state h_i, the output

vector y_i of the current step is derived using the weight matrix W. This architecture is illustrated in Figure 2.18, which includes a compact visualization and a graphic showing the individual steps after unfolding the sequence.



(a) Compact (b) Unfolded

Figure 2.18: The architecture of an RNN with (a) a compact visualization and (b) an unfolded one over the sequence. The input sequence **x** (green) is used to update the hidden state **h** (blue), with which the output sequence **y** (red) is created. U, V, W are the trainable weight matrices to transform the vectors.

The basic RNN architecture explained above is seldom used in practice due to challenges when training it with long sequences, primarily caused by the vanishing gradient problem. This issue arises when the gradients of the loss function become exceedingly small as they are backpropagated through the network, resulting in minimal updates to the weights associated with early sequence steps which hinders effective learning. To address this limitation, more advanced RNN architectures like long short-term memory (LSTM) networks or gated recurrent units (GRUs) are generally preferred. These architectures incorporate mechanisms to selectively retain or forget information in the hidden state, which results in the gradient not diminishing and more effective learning of long sequences.

RNNs offer great flexibility, as the input and output sequences can have different lengths. For instance, some steps of the output can be ignored, or the hidden state can be advanced without providing new input. This results in different configurations known as one-to-many, many-to-one, and many-to-many, depicted in Figure 2.19. Each of the configurations has use cases and applications suited for it. A one-to-many configuration might be employed to generate a descriptive sequence, such as a description from a singular input, for instance an image or better the feature vector describing the image. The many-to-many configuration is particularly relevant in applications such as machine translation, where a complete input sequence is transformed into a corresponding output sequence, for instance, to translate text from one language to another.

Despite their strengths, RNNs face constraints, particularly due to their inherent sequential processing, which limits parallelization and leads to longer training times.

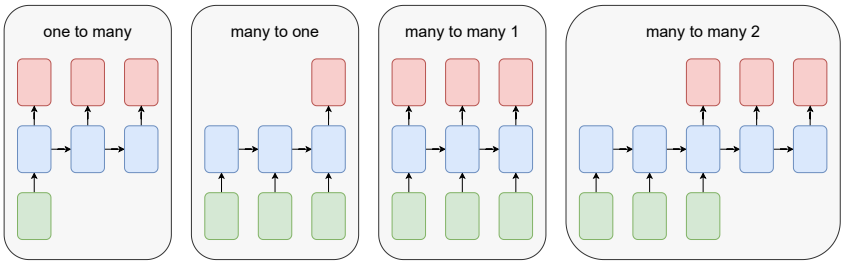


Figure 2.19: Possible configurations of an RNN.

The transformer architecture addresses many of these challenges by allowing for parallel processing of sequence data, significantly speeding up training compared to sequential RNNs. Additionally, transformers use self-attention mechanisms that are more effective at capturing long-range dependencies, further enhancing their performance, but at the cost of increasing memory and compute requirements. Recently, xLSTM, an enhanced LSTM architecture, has been proposed (Beck et al. 2024), promising solutions for some of the problems of RNNs and serving as an alternative to transformers. However, practical experience with xLSTMs is still limited due to their recency.

2.5.3 Data Normalization

Data normalization is an essential preprocessing step in the training of deep neural networks, aimed at enhancing the efficiency and overall performance of the training process. This procedure mitigates the variability in data scales and distributions, which, if not addressed, can result in certain features disproportionately influencing the model's learning, potentially leading to suboptimal or unstable outcomes. For example, when inputs exhibit varying magnitudes, those with higher values can have an undue impact on the output, irrespective of their true significance.

Two normalization techniques are min-max normalization and z-score standardization. Min-max normalization involves rescaling input values x from an initial range $[a, b]$ to a new specified range, often $[0, 1]$, using the transformation:

$$x' = \frac{x - a}{b - a}. \tag{2.14}$$

Conversely, z-standardization adjusts the data to achieve a mean μ of 0 and a standard deviation σ of 1. The formula

$$x' = \frac{x - \mu}{\sigma}. \tag{2.15}$$

is used.

2.5.4 Loss Functions and Evaluation Metrics

Loss functions are essential for training NNs and additionally serve as metrics to evaluate model performance. A variety of loss functions exists, each with slightly different characteristics. They compare the predicted values $\hat{y}$ with the labels y. N denotes the number of values which are compared. Two of the most commonly used losses are the mean absolute error (MAE), also referred to as $\mathcal{L}_1$ loss, and the mean squared error (MSE) (also called $\mathcal{L}_2$ loss):

$$\text{MAE} = \frac{1}{N} \sum_{i=0}^{(N-1)} |y_i - \hat{y}_i| \tag{2.16}$$

$$\text{MSE} = \frac{1}{N} \sum_{i=0}^{(N-1)} (y_i - \hat{y}_i)^2 \tag{2.17}$$

They use the absolute and the squared sum of errors. Using the squared error sum results in a significantly increased influence of large errors and wildly wrong predictions. For the MSE, the unit is the square of the unit of the target variable, for instance, m^2 instead of m for a length measure. Because of this, the root mean squared error (RMSE) takes the root of the MSE to have the same unit as the target variable, which enables an easy interpretation of the magnitude of the errors:

$$\text{RMSE} = \sqrt{\text{MSE}} \tag{2.18}$$

Another measure based on the MSE is the peak signal-to-noise ratio (PSNR):

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right) \tag{2.19}$$

It normalizes the MSE with the maximum value of the data MAX_I, squared for the correct unit, in a logarithmic scale.

A loss for image data is the structural similarity index (SSIM), which assesses the perceptual similarity between two images considering the luminance, contrast, and structure of two images:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{2.20}$$

Here, x and y denote windows of two images, μ_x and μ_y their means, σ_x^2 and σ_y^2 their variances, and σ_{xy} the covariance of them. Factors c_1 and c_2 ensure numerical stability, calculated using L, the dynamic range of pixel values, and small values $k_1 = 0.01$ and $k_2 = 0.03$ using $c_n = (k_n L)^2$. To get the final SSIM of a whole image, the mean over the SSIM of all windows is calculated.

The Pearson correlation coefficient ρ measures the linear relationship between two continuous variables, with its values ranging from -1 to 1. It is calculated using the mean $\mu_y, \mu_{\hat{y}}$ of y and $\hat{y}$:

$$\rho = \frac{\sum_{i=0}^{N-1} \left(\hat{y}_i - \mu_{\hat{y}}\right)\left(y_i - \mu_y\right)}{\sqrt{\sum_{i=0}^{N-1} \left(\hat{y}_i - \mu_{\hat{y}}\right)^2 \sum_{i=0}^{N-1} \left(y_i - \mu_y\right)^2}} \tag{2.21}$$

A coefficient around 0 suggests no linear relationship of the data while -1 or 1 indicates strong negative and positive linear relationships, respectively.

The use and interpretability of the Pearson correlation coefficient ρ is based on several assumptions: both variables should be normally distributed, the relationship between the variables should be linear, and the data should exhibit homoscedasticity. Additionally, the Pearson correlation is highly sensitive to outliers, which can disproportionately affect the correlation value and lead to misleading interpretations. Therefore, excluding outliers can be valuable in analyses. When the assumptions of Pearson's ρ are severely violated, alternative correlation coefficients such as Spearman's rank correlation or Kendall's tau correlation should be used.

The coefficient of determination, R^2, builds upon the Pearson correlation coefficient to quantify the proportion of variance in one variable that is predictable from the other variable. It is the square of Pearson's ρ, thus taking on values between 0 and 1, where a value closer to 1 indicates a stronger relationship of the variables within a linear model. R^2 provides an intuitive measure of the model's fit, representing how well the observed outcomes are replicated by the model. However, like ρ, R^2 assumes linearity and can be misleading in the presence of outliers or non-linear relationships.

3 State of the Art for NDVI Estimation From SAR Data

What you'll learn This chapter reviews the research and current state of the art related to utilizing SAR data for NDVI estimation before detailing the contributions made by this dissertation.

In this chapter, research relevant to this thesis is reviewed. Initially, foundational studies that link SAR data with NDVI values are examined. Subsequently, various SAR-based NDVI estimation methods and approaches to densify NDVI time series are summarized. As machine learning techniques are employed in this work, the datasets used to potentially train these models are also discussed. Finally, the specific contributions of this dissertation to address the identified research gaps are detailed.

3.1 Relationship between SAR Backscatter and VIs

The foundation for estimating vegetation indices (VIs) from synthetic aperture radar (SAR) data is based on the relationship between these data sources. To understand if this relationship exists and whether it can be effectively utilized, we need to delve into the properties influencing VIs and SAR backscatter. The normalized difference vegetation index (NDVI), the most commonly used VI, is affected by the spectral characteristics, chemical properties, and specifically the chlorophyll content of the vegetation (Gitelson et al. 1997). On the other hand, backscattered radar waves are influenced by the physical and electrical properties of the medium, such as roughness, geometry, and water content (Moreira et al. 2013). Given these differing influences, a direct relationship between NDVI and SAR backscatter is not immediately plausible. Therefore, vegetation parameters that might connect both measures are gathered before examining existing research exploring a direct relationship between the data.

3.1.1 Indirect Relationship Through Vegetation and Plant Parameters

Numerous studies have highlighted the ability to estimate various vegetation, surface, and crop parameters using either NDVI values or SAR data, suggesting a potential

connection. For instance, the leaf area index (LAI) has been linked to the NDVI in different environments such as mangroves (Tian et al. 2017) and wheat fields (Duchemin et al. 2006). Similarly, the LAI is also related to SAR backscatter in mangroves (Kovacs et al. 2013) and wheat fields (Vreugdenhil et al. 2018), as well as for barley (Harfenmeister et al. 2019). Another vegetation parameter, crop biomass, has been effectively estimated using both SAR data (Harfenmeister et al. 2019; Khabbazan et al. 2019; Vreugdenhil et al. 2018) as well as NDVI data (Hansen et al. 2003; Chandel et al. 2019). The water content of vegetation is another integral parameter. NDVI values are related to fuel water content (FWC) in meadows (Castro et al. 2014) and equivalent water thickness (EWT) in grass- and shrublands (Jiang et al. 2009). SAR data relates to vegetation water content (VWC) in English Rye Grass (Khabbazan et al. 2019), wheat (Harfenmeister et al. 2019), and rape, corn, and wheat (Vreugdenhil et al. 2018). Lastly, the land cover can be roughly classified using NDVI values (Hansen et al. 2000; Huete 1997; Pettorelli et al. 2005) as well as the radar backscatter (Balzter et al. 2015; Mahdianpari et al. 2017).

3.1.2 Relationship Between Measurements of Optical VIs and Radar Backscatter

A relationship between NDVI and SAR backscatter values of Sentinel-1's C-band data is shown for croplands in Central African (Ruciński et al. 2023), for maize in Spain (Alvarez-Mozos et al. 2021), for barley and maize in France (Veloso et al. 2017), for different crops in Italy (Mastro et al. 2023), for barley, wheat, grassland, and maize in Germany (Holtgrave et al. 2020), and multiple crops in India (Periasamy 2018). The relationship using C-band data of RADARSAT, another SAR satellite, is shown for maize and onions (Moran et al. 2012) or canola in Canada (Jiao et al. 2018) as well as with X-band data for herbaceous crops like maize and alfalfa in Italy (Capodici et al. 2013) and carrots in Austria (Segalini et al. 2014). Yet, not only for croplands this NDVI–SAR backscatter relationship was found, but also for grasslands and meadows (Stendardi et al. 2019; Abdel-Hamid et al. 2020; Wang et al. 2013) and forests (Frison et al. 2018).

Next to the SAR backscatter, the interferometric coherence can also be related to the NDVI. This relationship is investigated in two studies, both finding a high correlation for croplands, especially when a short temporal baseline is used (Villarroya-Carpio et al. 2022; Villarroya-Carpio et al. 2024).

While numerous studies have demonstrated a relationship, they typically investigate only a single study area without comparing it to other regions with different climates or crops. Additionally, these studies usually rely on data from a single sensor with a single frequency, preventing comparisons between different bands to identify the most suitable one. The investigation of a single SAR frequency in each of these studies is driven by two factors: the optimal wavelength, which determines the depth of penetration into the vegetation canopy, and the availability of data at a particular

wavelength. Moreover, most research has focused on C-band and occasionally X-band data, but investigations for S- and L-band data were not carried out. For S-band data, this is mainly due to the limited number of deployed systems using this frequency, with NovaSAR-1 currently being the only one with public data access. However, the anticipated launch of NISAR (the NASA-ISRO Synthetic Aperture Radar), featuring both S- and L-band SAR sensors, will dramatically increase data availability and foster further research.

3.2 Estimation of NDVI Values Using SAR Backscatter Data

Having established a link between NDVI and radar backscatter values, the next logical step is to focus on methodologies aimed at estimating NDVI values using SAR data. A similar task is the reconstruction of cloud-contaminated optical imagery aided by SAR data. But while for that task the reflectance values of all spectral bands have to be estimated, the translation of SAR into NDVI data requires only estimating a single value per pixel. Importantly, NDVI estimation must be precise, whereas cloud removal often aims merely to create images suitable for visual analysis.

3.2.1 Machine Learning Methods

Simple approaches for NDVI estimation from SAR data use machine learning models that operate on individual pixels, disregarding temporal or spatial context. An example model are random forest (RF) regressors, which are for example applied to areas in Brazil (Filgueiras et al. 2019), fields in India (Agrawal et al. 2022), or to forested regions in the Czech Republic (Paluba et al. 2023). Another example is the work by Chen et al. (2023b), who additionally include weather data and apply this method successfully to Alfalfa fields in the United States of America. A slightly different approach is used by Pelta et al. (2022): to fill gaps in the NDVI imagery of a Sentinel-2 tile, a RF model is trained using the SAR-NDVI relationship of this tile's data of the last year. Santos et al. (2022) conduct a comparison for this task between an RF and two other machine learning models for an area in Brazil and find the RF to perform best. Instead of using only a single SAR frequency, Lasko (2022) used C- and L-band simultaneously to successfully estimate NDVI and the normalized difference water index (NDWI) used for water body detection (cf. Equation 2.7).

3.2.2 Deep Learning Methods

Multiple studies employed deep learning methods instead of the simpler pixel- or field-level machine learning approaches. One of the earliest works is done by Scarpa et al. (2018), who use a small convolutional neural network (CNN) for a small area

in Burkina Faso, Africa, with decent results. Instead of a supervised CNN training, self-supervised generative adversarial networks (GANs) can also be utilized, as done for areas in India (Ramathilagam et al. 2023) or using the global SEN12MS dataset (Soglia et al. 2023). The use of another generative deep learning architecture, diffusion models, is demonstrated by Ettari et al. (2024) for SAR-to-NDVI translation, but without a thorough evaluation.

All of the presented machine and deep learning models are only applicable to very limited areas, as their training data was limited to single regions. With the datasets confined to specific regions, capturing limited types of land covers, crops, and climates, these models lack generalizability and cannot be applied globally.

3.3 Creating Dense NDVI Time Series Using SAR Data

To create an NDVI time series without cloud-induced gaps, multiple approaches were developed and evaluated.

Traditional techniques involve simple linear interpolation and filtering methods (Chen et al. 2004), or leveraging multi-year time series data to fill gaps (Li et al. 2021). These methods are efficient for short gaps and long-term trends but struggle in scenarios demanding high temporal resolution, such as dynamic cropland monitoring.

Sensor fusion methods can combine data from multiple optical sensors. When both sensors have comparable resolutions and characteristics, this is straightforward and consists of matching and harmonizing the reflectance values to ensure consistent spectral information resulting in a higher temporal resolution. One prominent example is the *Harmonized Landsat and Sentinel-2* product (Claverie et al. 2018). If however, the spatial resolution is wildly different, more complex spatio-temporal fusion methods have to be applied, combining coarse-resolution frequent imagery with high-resolution, but less frequent imagery (Gao et al. 2006; Wang et al. 2018; Chen et al. 2021). These approaches can work well but fail in regions with persistent cloud cover, where no cloud-free optical imagery can be acquired for weeks.

Instead of using optical time series, data from SAR sensors can be used. They offer a promising alternative due to their cloud-penetrating capabilities and their ability to penetrate into the vegetation, allowing interaction with a larger portion of the vegetation cover. This can be useful, for example, in obtaining information on vegetation biomass where optical sensors can only use data from the topmost vegetation surface. SAR time series have been employed in various studies for crop monitoring (McNairn et al. 2016; Liu et al. 2019; Hosseini et al. 2019) or forest biomass retrieval (Mermoz et al. 2015; Pichierri et al. 2018), yet these approaches have multiple drawbacks. Additional research is needed to relate radar data to plant properties, as existing studies linking them to optical data are not applicable. Furthermore, SAR data requires more

complex processing and more expertise than optical data. Lastly, there is a lack of long-term data archives, while the Landsat archive contains continuous optical data for decades.

To exploit the strengths of both sensor types, optical and SAR ones, recent studies have explored translating SAR data to optical or NDVI data. Notable works include sequence-to-sequence models translating SAR time series to NDVI time series (Zhao et al. 2020; Garioud et al. 2021). However, they are limited to specific regions and don't show, whether they are transferable to other regions. Additionally, the proximity of training and test pixels is likely leading to insufficient understanding of the model's real performance and an overestimated test performance. Furthermore, one work demands a common temporal grid which hinders the easy use for other regions with different revisit times of the used sensors. In addition, data from field polygons must be averaged for the method to work, making field polygons a necessity. Another interesting work translates SAR backscatter to NDVI values of small polygons (< 1 ha) and utilizes the prediction uncertainty for the fusion with the sparse optical NDVI time series (Chen et al. 2023a).

Although the existing research has achieved notable advances in the utilization of both optical and SAR data for the generation of NDVI time series, significant gaps in the literature remain unaddressed. Most notably, there is a lack of methodologies that effectively integrate time series of different modalities and sensors using adaptable deep learning models. Moreover, previous studies have predominantly focused on region-specific applications, thereby failing to develop models that are universally applicable across diverse geographic and environmental conditions.

3.4 Spatio-Temporal Methods

In the previous two sections, the cited studies utilize either single pixel values, spatial context, or temporal context to estimate the NDVI. However, it is also possible to simultaneously exploit both spatial and temporal contexts for more accurate NDVI estimation. This would enable exploiting the full potential of the existing data at the cost of a higher data requirement, more processing power needed to train models, and a more complicated deployment. One of the few studies using both spatial and temporal correlations is done by Mao et al. (2023) to reconstruct cloud-contaminated NDVI imagery. Another study – not targeting the NDVI but instead creating a biomass proxy – exploits the temporal characteristics of Sentinel-1 data and the spatial structure of Sentinel-2 NDVI imagery for the final product (Burger et al. 2024). The little amount of studies show, that further research is needed in this regard.

3.5 Remote Sensing Datasets Including SAR and Optical Data

Multiple large datasets with different characteristics and foci were published containing both optical and SAR imagery. One of the first global, large-scale datasets is SEN12MS (Schmitt et al. 2019b). In that dataset, optical and SAR images are included, originating from the same season for each area. This entails, that they can be acquired up to three months apart from each other. BigEarthNet is a similar dataset, but contains imagery only for Europe and not globally (Sumbul et al. 2019). Three other datasets are aimed at cloud removal tasks: CloudSEN12, SEN12MS-CR, and SEN12MS-CR-TS (Aybar et al. 2022; Meraner et al. 2020; Ebel et al. 2022). They contain either paired imagery of a single or multiple dates, but all of them include clouds, rendering them incompatible with SAR-to-NDVI translation. In conclusion, no global dataset is available containing cloud-free optical imagery and SAR imagery of the same date for each area.

3.6 Contributions of This Thesis

The previous sections highlighted a strong research interest in understanding the relationship between SAR backscatter and NDVI. Additionally, numerous approaches strive to enhance vegetation monitoring and estimate NDVI values from SAR backscatter data. However, several research gaps exist, which this thesis aims to close. The key contributions of this dissertation are:

1. a globally applicable model to estimate NDVI images from SAR imagery,
2. an approach to fuse sparse optical NDVI time series with dense SAR-estimated NDVI time series,
3. the presentation of a global, balanced dataset of paired Sentinel-1 SAR and Sentinel-2 optical data of the same date, called SEN12TP, and
4. the investigation of the SAR–NDVI relationship for S-band SAR data as well as a comparison of this relationship between different SAR frequencies for multiple regions.

The foremost contribution of this thesis is advancing the estimation of NDVI images from SAR data. For this task, a globally applicable model is presented and thoroughly evaluated. This universal model can be applied to all areas of the world without the need for any fine-tuning to adapt it to a specific region. This advancement is made possible by the novel dataset SEN12TP (Sentinel-1 and -2 images, timely paired), which is presented in this thesis. The SEN12TP dataset is unique as it consists of SAR and optical image acquisitions from the same day. Thereby, a sophisticated sampling strategy ensures a balanced and global distribution of the images, enabling an accurate translation from radar to optical data regardless of the location.

The second major contribution is a novel approach for generating dense NDVI time series by integrating sparse optical NDVI time series with dense SAR-estimated NDVI time series using an RNN architecture. This fusion technique leverages the spatial context provided by SAR-derived NDVI values to produce dense NDVI time series. The innovation of this approach is twofold: it demonstrates the flexible fusion of earth observation time series data from diverse sources and highlights the potential for global application.

The final contribution is the analysis of the SAR–VI relationship. This thesis will be the first to link NDVI and S-band backscatter data. Additionally, for three other SAR frequencies the SAR–VI relationship is analyzed and compared for three study areas with different crop types to provide a comprehensive insight into the spatial applicability and performance of SAR-based vegetation monitoring.

4 Exploration of the SAR–NDVI Relationship

What You'll Learn This chapter explores the relationship between SAR backscatter and NDVI values for three diverse agricultural study areas with a focus on S- and C-band SAR data, but also examining L- and X-band data.
Based on This chapter's material originates from the article "Comparing the relationship between NDVI and SAR backscatter across different frequency bands in agricultural areas" published in *Remote Sensing of Environment* (Roßberg et al. 2025).

Translating SAR backscatter to NDVI values presents a promising approach for enhancing vegetation monitoring. This method retains the use of the widely adopted and thoroughly researched NDVI while mitigating data gaps caused by cloud cover through the deployment of weather-independent SAR data. The foundation of this translation lies in establishing a relationship between SAR backscatter and NDVI values. Previous studies have explored this relationship (cf. Section 3.1), but typically within single regions characterized by similar land cover, using only a single SAR frequency. Furthermore, no prior research has investigated this relationship using S-band SAR data. This chapter aims to fill this research gap and compares the SAR–NDVI relationship across different regions, agricultural practices, and data of multiple SAR sensors operating at different frequencies including S-band data.

For this extensive analysis, three study areas are chosen located in Australia, South America, and South East Asia. They vary in the grown crops and agricultural practices and are described in detail in Section 4.1. For all these areas, C- and S-band data from the Sentinel-1 and NovaSAR-1 sensors are acquired and compared with Sentinel-2 derived NDVI values. For the third study area, additionally L- and X-band data from the SAOCOM and Cosmo-SkyMed sensors could be acquired. Using this data, the relationship between SAR and NDVI values is analyzed in Section 4.3 and afterwards discussed in Section 4.4.

4.1 Study Sites and Used Data

4.1.1 Study Areas

To gain a broad understanding of the relationship between SAR backscatter and NDVI values, three different areas with distinct climates and agricultural practices were selected: near the cities of Boort in Australia and Bell Ville in Argentina, and an area in the Mekong River Delta in Vietnam. Their location on the globe is shown in Figure 4.1. The choice of study areas was restricted due to the limited availability of suitable S-band NovaSAR-1 data. Nevertheless, these three regions offer a variety of climates and crop types, providing a robust basis for analysis. Mainly Sentinel-1 C-band and NovaSAR-1 S-band data are used, but for one study area COSMO-SkyMed X-band and SAOCOM L-band data are also used. The study areas are described in detail in the following sections.

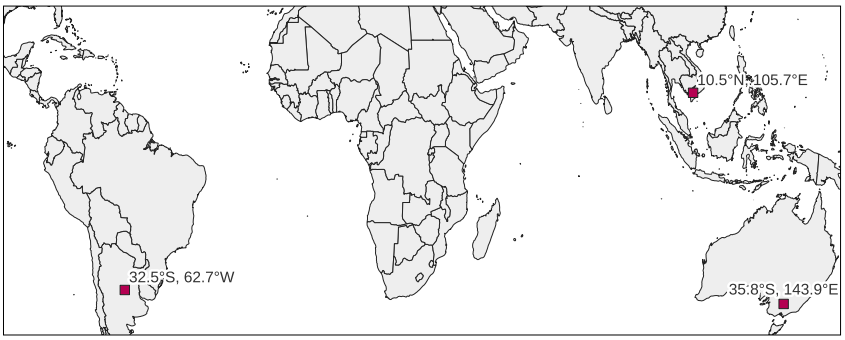


Figure 4.1: Location of the three investigated study regions.

Boort (Australia)

Boort, located in Australia, is characterized by a cold semi-arid climate (BSk of the Köppen-Geiger classification) with an average annual temperature of 15.6 °C and 371 mm of precipitation. The climate features warm summers from December to February with average temperatures ranging from 20 to 22 °C and low precipitation (20 mm/month), and cooler winters from June to August with temperatures around 8 to 9 °C and higher precipitation (40 mm/month) (Karger et al. 2017; Karger et al. 2021).

The region's elevation varies from 60 to 110 meters, with mild elevation changes and slopes of less than 1°. The primary crops include wheat (43 %), barley (25 %), and canola (17 %), as well as pasture (13.2 %) (Australian Bureau of Statistics 2024). Wheat and barley are typically planted from April to June and harvested between

October and January, while canola is planted in April and harvested from mid-October (U.S. Department of Agriculture 2021).

Three NovaSAR-1 acquisitions were selected for this area, two during the winter growing season (2021-08-07 and 2022-06-01) and a third on 2022-01-20 during the drier summer without much green vegetation. All three images were acquired in ScanSAR mode with three polarizations VV, HH, and HV. They have slightly different resolutions of 35 m and 30 m for the summer and winter acquisitions, respectively, and also different incidence angles ranging from 15.1° to 28.5°, as they were acquired from different orbits. For this area, CSIRO provided analysis-ready data with radiometric terrain correction. The optical and Sentinel-1 C-band SAR images are acquired on approximately the same days as the NovaSAR-1 data, with a maximum difference of three days. The exact acquisition dates of all three sensors, together with the incidence angles and, for NovaSAR, the ScanSAR mode resolution used, are detailed in Table 4.1.

NovaSAR			Sentinel-1		Sentinel-2
date	resolution	angle	date	angle	date
2021-08-07	35 m	15°-22°	2021-08-06	34°-38°	2021-08-08
2022-01-20	30 m	24°-29°	2022-01-21	34°-38°	2022-01-20
2022-06-01	35 m	15°-22°	2022-06-02	34°-38°	2022-06-04

Table 4.1: Overview of the used SAR and optical scenes used for the Boort study area in Australia. The incidence angle of each scene is given for both NovaSAR-1 and Sentinel-1. Additionally, the ScanSAR mode resolution of each NovaSAR-1 scene is listed.

Bell Ville (Argentina)

Bell Ville, located in Argentina, experiences a monsoon-influenced humid subtropical climate (Köppen-Geiger classification: Cwa) with an average temperature of 17.4 °C and 952 mm of annual precipitation. The region has mild winters from June to August with temperatures of 10 °C to 12 °C and minimal precipitation (13–17 mm/month), and hot summers from December to February with average temperatures of 23–24 °C and significant precipitation (120–160 mm/month). In the Union Department, whose capital Bell Ville is, mainly soy (46 %), corn (29 %), and wheat (16 %) are cultivated (Dirección de Estimaciones Agrícolas 2023). The agricultural season runs from November to May, with soy planted from mid-November to mid-January, corn from mid-September to November, and wheat from mid-May to July, with corresponding harvest seasons of mid-March to mid-June for soy and corn, and mid-November to mid-January for wheat (U.S. Department of Agriculture 2024). The region is relatively flat, with elevations ranging from 115 to 140 meters.

Six NovaSAR-1 acquisitions were selected for further analysis. They are acquired in the 20m ScanSAR mode with two polarizations, HH and HV, and have incidence angles between 22° and 31°. They are provided by CSIRO in ScanSAR detected (SCD) format. The corresponding Sentinel-1 and -2 images are acquired within three days of the NovaSAR-1 acquisition, with only two exceptions, where Sentinel-1 data are acquired five and six days after NovaSAR-1 and Sentinel-2, respectively. All scenes were acquired between December 2023 and April 2024, the main growing season for this area. The exact acquisition times of all three sensors together with the incidence angle are given in Table 4.2.

NovaSAR		Sentinel-1		Sentinel-2
date	angle	date	angle	date
2023-12-24	27–30°	2023-12-20	35–37°	2023-12-25
2024-01-02	29–31°	2024-01-01	35–37°	2024-01-04
2024-01-19	24–27°	2024-01-13	35–37°	2024-01-19
2024-02-16	29–30°	2024-02-18	35–37°	2024-02-13
2024-03-03	22–24°	2024-03-01	35–37°	2024-03-04
2024-03-25	24–27°	2024-03-25	35–37°	2024-03-24

Table 4.2: Overview of the acquisition dates of the three sensors utilized for the Bell Ville study area in Argentina. The incidence angle is given for the SAR data.

Mekong River Delta (Vietnam)

The third study area is located in the Mekong River Delta in southern Vietnam in the province of Đồng Tháp. It is almost entirely covered with rice fields, as the Mekong Delta produces 56 % of Vietnam's rice (U.S. Department of Agriculture 2017). The terrain is extremely flat, with elevations ranging from 0 to 10 meters above sea level. The climate of the region is predominantly monsoonal (Köppen-Geiger classification As and Aw), with an average annual precipitation of 1600 mm. The rainy season occurs from May to November, with more than 150 mm of rain per month, while the dry season from December to April has less than 50 mm of rain per month. Temperatures remain fairly constant throughout the year, ranging from 26–29 °C. Rice is cultivated up to three times per year, with our study period covering the summer-autumn (July 1 to January 31) and autumn-winter (March 1 to August 31) seasons (Clauss et al. 2018).

Two dates were chosen in March and August 2023. For these dates, SAR data could be acquired at four frequencies: NovaSAR-1 S-band, SAOCOM L-band, Sentinel-1 C-band, and Cosmo-SkyMed Second Generation (CSG) X-band data. Only CSG data could not be acquired on both dates, but only for August 2023. Data from all sensors

are acquired no more than three days apart from the Sentinel-2 optical acquisition. For Sentinel-1, both ascending and descending orbits could be used. Similarly, two SAOCOM acquisitions were available in August 2023.

The NovaSAR-1 data are acquired in 20m resolution dual-polarized ScanSAR mode with HH and HV polarization and are provided by CSIRO in SCD format. The orbit state and position are inaccurate for acquisition on 2023-08-06, resulting in a geolocation error of several kilometers, which was manually corrected after data processing. This leads to small errors in the backscatter, as the angle of incidence is off by 0.2°, as well as errors where the terrain is not flat. Since most of the terrain is very flat, this does not affect the data analysis and results. SAOCOM and CSG data were acquired in stripmap mode with HH and HV polarization, Sentinel-1 in interferometric wide swath (IW) mode. All three SAOCOM acquisitions were not perfectly geolocated but were off by 10–50 m, which was corrected manually.

For all five sensors, the exact acquisition times as well as the incidence angles of the SAR scenes are given in Table 4.3.

NovaSAR-1		Sentinel-1		SAOCOM		CSG		Sentinel-2
date	angle	date	angle	date	angle	date	angle	date
2023-03-05	25–27°	2023-03-05	38–41°	2023-03-03	21–23°			2023-03-06
		2023-03-06	40–43°	2023-03-06	29–31°			
2023-08-08	13–15°	2023-08-08	38–41°	2023-08-05	29–31°	2023-08-06	44–46°	2023-08-08
		2023-08-09	40–43°					

Table 4.3: Overview of the acquisition dates for the Mekong River Delta study area in Vietnam along with the incidence angle for all SAR scenes.

4.1.2 SAR Data

This study utilizes SAR data from multiple sensors with different frequencies and imaging modes.

NovaSAR-1 is a space-borne S-band SAR system launched in 2018 and operated by a consortium of various space agencies and partners (Held et al. 2019). Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO) provides data free of charge on their data portal[1].

Sentinel-1 is a global monitoring mission by the European Space Agency (ESA) utilizing C-band SAR to provide continuous data of the Earth's surface (Torres et al. 2012). The Sentinel-1 data used in this study were captured in IW mode with VV and VH polarizations with a raw slant range resolution of 3 × 22m (Rg × Az). The Level-1 Ground Range Detected (GRD) data were downloaded from the Copernicus

[1] `https://data.novasar.csiro.au/`

Browser data platform[2], which are already detected, multi-looked, and projected to the ground range.

The Argentine SAOCOM 1A and 1B satellites are part of the L-band SAR constellation, which is overseen by the National Space Activities Commission (CONAE). The single-look complex (SLC) data utilized in this study is taken in dual-polarized stripmap mode with a ground resolution of 10m × 5m (Rg × Az) and includes HH and HV polarizations.

The fourth sensor used is from the second-generation COSMO-SkyMed (CSG) satellite constellation, which carries X-band SAR systems and is operated by the Italian Space Agency (ASI). Images are taken in dual-polarized (HH and HV) stripmap mode with a resolution of 3 m × 3 m. The data are provided in the Level-1B format, which is already focused, amplitude-detected, radiometrically equalized, and represented in ground projection.

An overview of the data from all sensors including the imaging mode and data product for each study area is provided in Table 4.4. Upon visualizing an example image of each SAR sensor the resolution differences of the sensors become evident. CSG and SAOCOM demonstrate superior resolution compared to Sentinel-1, which still exhibits a better resolution compared to NovaSAR-1 data. For each sensor, the data over the Mekong River Delta of a single date together with a zoomed-in detail view is displayed in Figure 4.2.

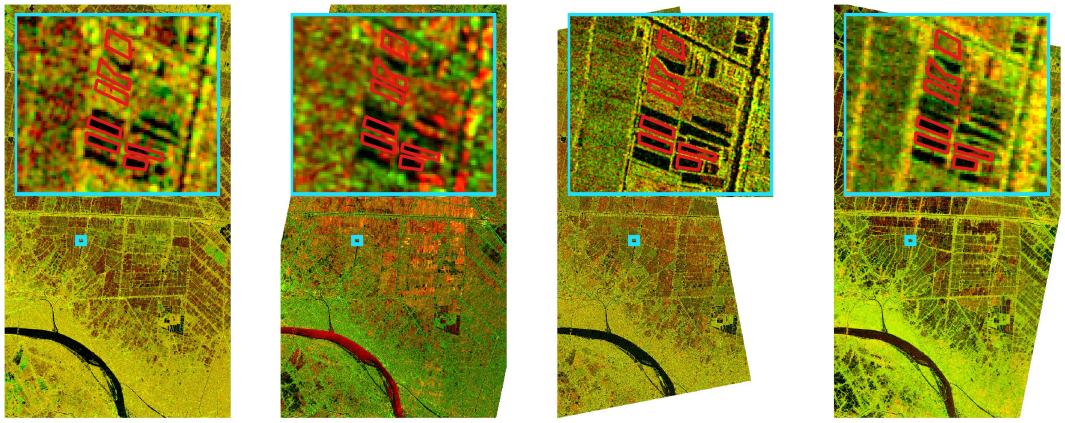
Table 4.4: Overview of SAR sensors utilized, along with their imaging parameters and data availability for each of the three study areas Boort (B), Bell Ville (BV), and Mekong River Delta (MRD).

Sensor	Band	Frequency (GHz)	Wavelength (cm)	Image mode	Polarization	Resolution (Rg × Az) (m × m)	data product	pixel spacing (m × m)	study region B	BV	MRD
NovaSAR-1	S	3.200	9.4 cm	ScanSAR	HH,HV,VV	30×30, 35×35	ARD	22.3×22.3	x	-	-
NovaSAR-1	S	3.200	9.4 cm	ScanSAR	HH,HV	20×20	CSD	10×10	-	x	x
Sentinel-1	C	5.405	5.5 cm	IW	VH,VV	20×22	GRD	10×10	x	x	x
SAOCOM	L	1.275	23.5 cm	Stripmap	HH,HV	3.75×3.23	L1A	2.4×6.95	-	-	x
CSG	X	9.600	3.1 cm	Stripmap	HH,HV	3×3	DGM_B	1.25×1.25	-	-	x

4.1.3 Optical data

The optical data utilized in this study is from Sentinel-2, a global monitoring mission by ESA and provides multispectral imagery with a 10-meter resolution. Captured spectral data includes visible, near-infrared, and shortwave infrared light with wavelengths from 443 to 2200 nm. Surface reflectance data (Level-2A) are used in this study, which are processed using Sen2Cor by ESA. In addition to providing surface

[2] `https://browser.dataspace.copernicus.eu/`



(a) Sentinel-1 2024-08-09 (b) NovaSAR 2023-08-06 (c) CSG 2023-08-06 (d) SAOCOM 2023-08-05

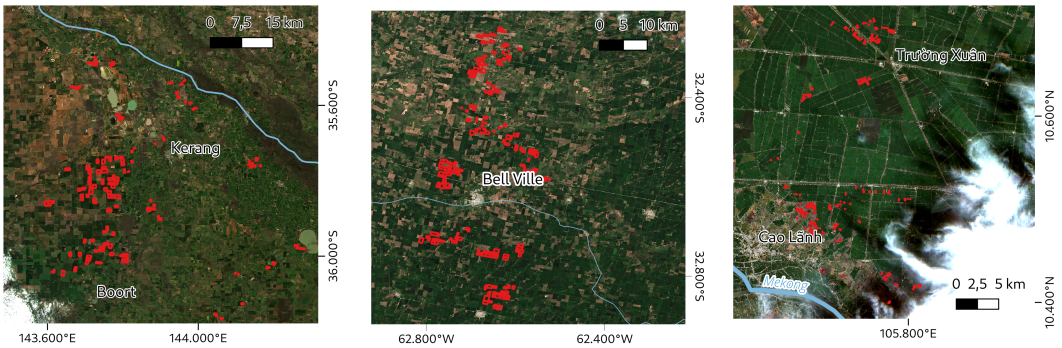
Figure 4.2: Visualization of the SAR data for the Mekong River Delta study area. A zoomed-in detail view is shown to illustrate the different image resolutions. The polarizations of each image are displayed in the red and green channels of the image, and the acquisition date is given for each image.

reflectance data free from atmospheric effects, Sen2Cor also generates a scene classification layer (SCL) scene classification layer (SCL) that annotates, among other things, cirrus and other clouds, their shadows, and water surfaces. An example visualization of the optical data for each study area is given in Figure 4.3.

4.1.4 Field outlines

The outlines of the agricultural fields in all three study regions were delineated manually using data from the Sentinel-2 satellite as well as Google aerial imagery. It was ascertained that all fields exhibit a uniform NDVI across all dates, as well as a homogeneous appearance in the SAR imagery. To avoid mixed pixels at the field boundaries, all polygons are designed to remain entirely within the field edges, with a buffer zone around the edge.

In total, around 200 fields were delineated for each of the three study areas. The Boort study area encompasses 174 fields, the Bell Ville area 204, and the Mekong River Delta region 208. Because of different agricultural practices, the field sizes change substantially between the study regions: Both Boort and Bell Ville feature larger fields with a median size of 16.4 hectares and 30 hectares, respectively. In the case of Boort, one-third of the fields are less than 10 hectares in size, another third are between 10 and 40 hectares, and the remainder are up to 120 hectares in size. In the Bell Ville area, field sizes are more homogeneous, with 60 % of fields measuring between 10 and 40 hectares. In contrast, the fields in the Mekong River Delta are



(a) Boort (Australia), 2021-08-08

(b) Bell Ville (Argentina), 2024-01-19

(c) Mekong River Delta (Vietnam), 2023-03-06

Figure 4.3: Sentinel-2 RGB visualization of the three study areas overlaid with the field outlines.

relatively small, with 85 % of them being smaller than 2 hectares and almost all of them being smaller than 4 hectares. For all three regions, the labeled polygons are shown in Figure 4.3, superimposed on the Sentinel-2 RGB image.

4.2 Data Processing and Analysis

In the following sections, the prerequisites for the data analysis are detailed. This includes SAR and optical data processing as well as a description of the data analysis methods.

4.2.1 SAR Data Processing

The SAR data used in this study is sourced from different sensors and each sensor needs different processing steps to yield data suitable for analysis.

For the NovaSAR-1 data, both SCD and analysis ready data (ARD) products were utilized. To the SCD data, we applied terrain correction and calibration in order to convert the digital numbers (DN) to intensity values. The ARD products that were utilized were already multi-looked, geocoded, and had undergone radiometric terrain flattening to γ° backscatter values. Accordingly, the ARD intensity data were solely transformed to a logarithmic scale. It should be noted that neither the SCD nor the ARD products include any corrections for antenna patterns or scalloping effects.

The processing of Sentinel-1 data was performed in a standard manner: precise orbit application, thermal and border noise removal, calibration, and geocoding.

The processing chain for SAOCOM data included calibration to intensity values, multi-looking with 1 × 3 looks in range × azimuth, conversion from slant range to ground range, and geocoding. CSG data processing involved 3 × 3 multi-looking and geocoding. Sentinel-1, SAOCOM and CSG data are normalized to σ° backscatter coefficients.

For processing the Sentinel toolbox SNAP was employed. Geocoding was performed using the Copernicus 30-meter digital elevation model (DEM). All data were converted to intensity values on a logarithmic scale, expressed in decibels (dB).

Given that all sensors provided both cross- and co-polarized data, the cross-ratio can be calculated by dividing the co-polarized data by the cross-polarized data, with both polarizations expressed in a linear scale (cf. Section 2.3.3). The combination of both polarizations into a single value enables a joint analysis of both polarizations and may potentially enhance the strength of the relationship. As the sensors provide different polarizations, different cross-ratios can be calculated: $^{HH}/_{HV}$, $^{VV}/_{HV}$, or $^{VV}/_{VH}$.

In preliminary experiments, the potential of the radar vegetation index (RVI, see Equations 2.12 and 2.13) was also investigated. However, no substantial improvements could be identified in comparison to single polarizations or the cross-ratio and consequently, the RVI was not employed in this study.

The geolocation accuracy of several SAOCOM and NovaSAR-1 scenes was insufficient, which required a manual correction. As this was done solely on the processed data, minor inaccuracies were introduced during the SAR processing due to the mislocated DEM and potentially erroneous incidence angles. However, these errors are sufficiently small to not affect the results. The geolocation accuracy of NovaSAR-1 is particularly poor when the satellite state vector is based on two-line element (TLE) data, as opposed to the more accurate GPS. The use of ground control points has the potential to enhance the SAR orthorectification, as demonstrated in Joshi et al. (2022), but was not used in this study due to the manual effort required for accurately acquiring them.

4.2.2 Optical Data Processing

The optical data from Sentinel-2 and the derived NDVI are used to investigate the relationship with SAR backscatter data. The optical data comes with a classification map, the scene classification layer (SCL), which contains information about the depicted object or surface of each pixel like clouds or snow. This mask is used to mask cirrus and other clouds as well as their shadows. Consequently, only those pixels classified as either *vegetated* or *non-vegetated* are kept. The SCL mask is upsampled from a resolution of 20 m to 10 m using nearest neighbor upsampling, as the used multi-spectral bands have a resolution of 10 m.

The NDVI is calculated using the red (R) and infrared (NIR) bands, specifically Sentinel-2 bands `B4` and `B8`, using Equation 2.6.

In the Mekong River Delta, the rice paddy fields are flooded at the beginning of the season. As the classification of *water* in the SCL was insufficient in detecting this flooding, we additionally calculate the modified normalized difference water index (mNDWI) (cf. Section 2.3.2). The mNDWI utilizes green and short wave infrared spectral bands, for which we chose Sentinel-2 bands `B3` and `B11`, respectively. It is calculated according to Equation 2.8. As Sentinel-2's SWIR channel `B11` is acquired at a resolution of 20 m, we upsampled it to 10 m resolution using bilinear interpolation. During data analysis, the mNDWI value is used to filter out fields that are potentially flooded, as described in the following Section 4.2.3.

4.2.3 Data Analysis

For this study, a field-wise approach was employed for data analysis to assess the relationship between SAR backscatter and NDVI values. A preliminary pixel-wise analysis did not show a significant relationship between them, which is likely due to the strong influence of speckle. This is also described in Roßberg et al. (2024a), where a pixel-wise analysis yielded inconclusive results.

The delineated field polygons were used to extract the median values from the SAR and optical raster images. Taking the average mitigates the impact of speckle and noise and enhances the reliability of the data. In the case of the Mekong River Delta study area, flooded fields with an average mNDWI above 0.2 were excluded to ensure that only relevant agricultural fields were analyzed.

To quantify the linear relationship between the SAR backscatter and NDVI values, Pearson's ρ correlation coefficient was employed (cf. Section 2.5.4). As Pearson's ρ is sensitive to outliers, extreme values were masked in the analysis. If a field has a mean NDVI or backscatter within the top or bottom 1 % of all fields, it is excluded. The calculation of ρ was performed using the Python package SciPy.

Scatter plots were generated to visualize the relationship between backscatter and NDVI values. An ideal scenario would be a 1:1 mapping where each backscatter value directly corresponds to an NDVI value, which would imply a perfect linear relationship. This visualization aids in assessing the potential for direct translation of backscatter data to NDVI values, which is crucial for evaluating the suitability of SAR data, particularly at the S-band, for vegetation monitoring and NDVI estimation.

4.3 Observed SAR–NDVI-Relationship

The results of each study area are presented in the following subsections.

4.3.1 Boort (Australia)

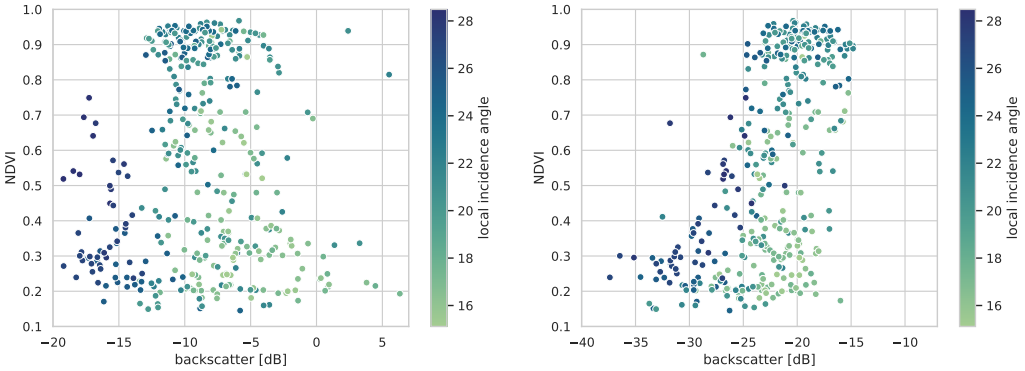
The NDVI values for all fields fall between 0.2 and 1, indicating that the data encompass all vegetative stages, from barren to fully grown fields. Values around 0.25 and 0.9 are more common than values situated between these distribution peaks.

Both NovaSAR-1 and Sentinel-1 backscatter distributions are approximately bell-shaped but differ in magnitude. Co-polarized backscatter values exceed cross-polarized ones by 10 dB.

A positive relationship between backscatter values and NDVI is evident for NovaSAR-1 cross-polarized HV data. No discernible relationship is evident between the backscatter values and the NDVI for both co-polarized HH and VV data. This can be numerically verified by the moderately high ρ of 0.54 for HV, but the negligible low ones of 0.0 and -0.14 for HH and VV, respectively. Using the cross-ratio instead of single polarization does increase the relationship with NDVI values mildly for the $^{VV}/_{HV}$ cross-ratio to $\rho = -0.64$, but is in the same range for the $^{HH}/_{HV}$ cross-ratio with $\rho = -0.55$. Visually, a better relationship with NDVI values of the cross-ratio data compared to the HV polarization can not be clearly attested. The local incidence angle influences the backscatter values with higher backscatter occurring for lower incidence angles. This effect does not occur for the cross-ratio. The scatter plots for all three NovaSAR-1 polarizations relating them to Sentinel-2 NDVI are presented in Figure 4.4.

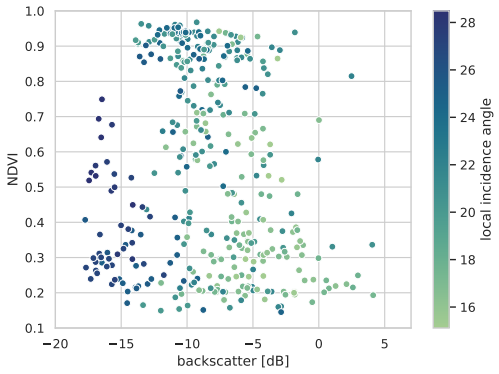
The results for the Sentinel-1 backscatter data are analogous to those observed for the NovaSAR-1 backscatter. Cross-polarized VH backscatter demonstrates a positive correlation with NDVI values, while no discernible relationship is evident in the case of co-polarized VV data. Once more, this yields a moderately high ρ value of 0.52 for VH polarization, but a ρ of zero for VV-polarized data. However, in the scatter plots data from winter and summer is easily distinguishable, because in winter predominantly low NDVI values occur, whereas in summer all NDVI values occur. This seasonality is also apparent in the backscatter distributions, for instance, the VH backscatter values range from -40 to -25 dB for fields in winter with low NDVI, whereas almost all fields in summer have a VH backscatter > -25 dB. This seasonal pattern is not discernible for NovaSAR; this may be attributed to the reduced number of data points, given that on that winter date the NovaSAR-1 acquisition only partially overlaps with the study area.

As with the NovaSAR-1 data, the cross-ratio results in an improvement in the correlation coefficient, ρ, for Sentinel-1 data, with a value of 0.52 for VH polarization and -0.74 for the cross-ratio $^{VV}/_{VH}$. While this effect is not readily apparent in visual inspection, it may be present, as evidenced by the scatterplots in Figure 4.5 that compare NDVI data with VH and VV polarizations and the cross-ratio. However, the overall distribution of the cross-ratio is comparable to that of the VH polarized data mirrored at the y-axis.

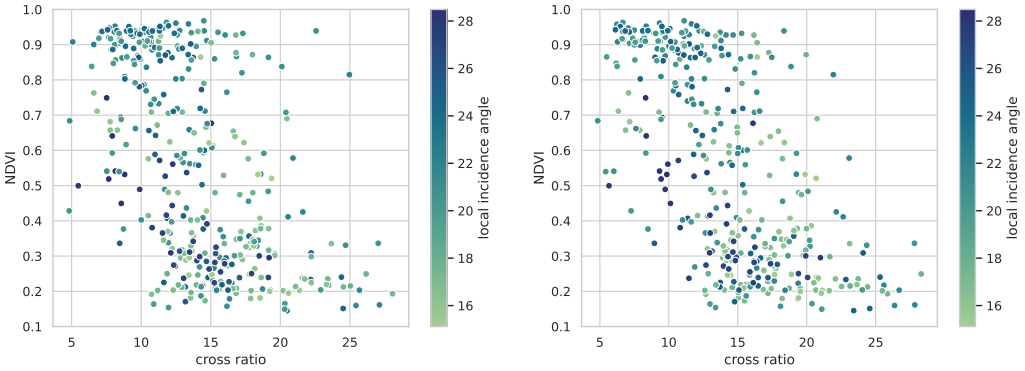


(a) HH polarization

(b) HV polarization



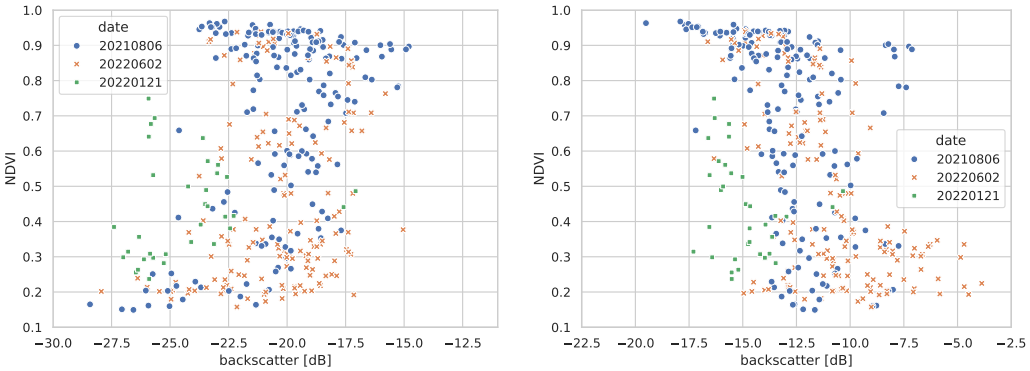
(c) VV polarization



(d) cross-ratio $^{HH}/_{HV}$

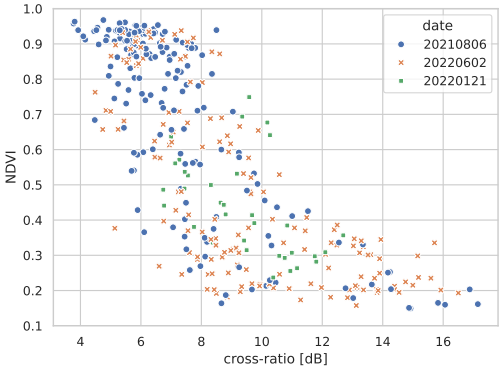
(e) cross-ratio $^{VV}/_{HV}$

Figure 4.4: Comparison of NDVI and NovaSAR-1 data for the Boort study area in Australia. All three polarizations HH, HV, VV, and the two cross-ratios $^{VV}/_{HV}$ and $^{HH}/_{HV}$ are given. The local incidence angle of the SAR data is indicated by different colors.



(a) VH polarization

(b) VV polarization



(c) cross-ratio $^{VV}/_{VH}$

Figure 4.5: Comparison of NDVI and Sentinel-1 data for the two polarizations VV, VH, as well as the cross-ratio $^{VV}/_{VH}$ for the Boort study area in Australia. The date on which the Sentinel-1 data were acquired is indicated by different colors. The incidence angle of all samples is between 34 to 38°.

Sensor	Data	ρ	N
NovaSAR	HV	0.54	(366)
	HH	0.00	(366)
	VV	-0.14	(366)
	$^{HH}/_{HV}$	-0.55	(366)
	$^{VV}/_{HV}$	-0.64	(366)
Sentinel-1	VV	0.00	(455)
	VH	0.52	(455)
	$^{VV}/_{VH}$	-0.74	(457)

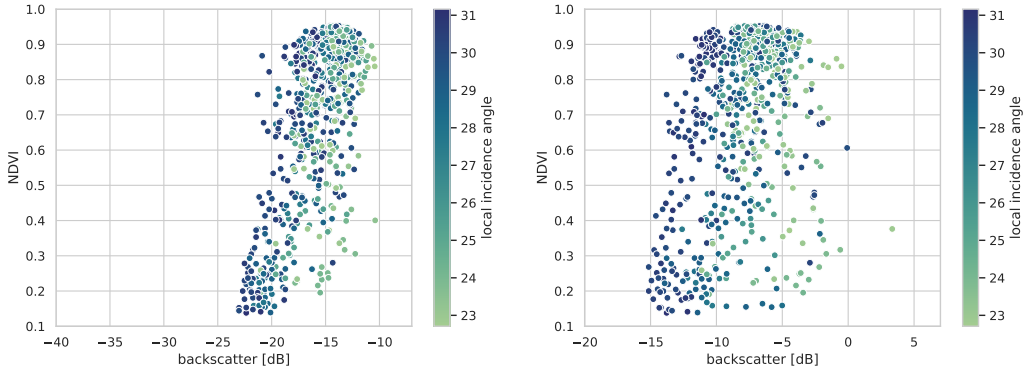
Table 4.5: Relationship between NDVI values and SAR data for the study area Boort (Australia). Next to the polarization, the cross-ratios formed from them is used. ρ denotes the Pearson's correlation coefficient and N the number of samples.

4.3.2 Bell Ville (Argentina)

In general, the findings for this study area are comparable to those of Boort, Australia, presented in the previous Section 4.3.1. For NovaSAR's HV data, a positive relationship is evident, whereby an increase in cross-polarized backscatter corresponds with an increase in NDVI values. Furthermore, the HH S-band data exhibits a weak positive correlation with the NDVI. This is reflected in the correlation coefficients, with $\rho = 0.68$ and $\rho = 0.35$ for HV and HH polarized data, respectively. For the cross-ratio $^{HH}/_{HV}$ ρ falls between these two values, with $\rho = -0.48$. Once more, the incidence angle, ranging from 22° to 31°, exerts an influence on the backscatter, with higher backscatter values occurring for lower incidence angles. This is evident from the scatter plots depicted in Figure 4.6.

The data obtained from the Sentinel-1 satellite appears to exhibit a positive relationship with the NDVI data for both polarizations, similar to the findings observed with the NovaSAR-1 satellite. The relationship with NDVI values is clear for both cross-polarized VH and co-polarized VV data. This is illustrated in Figure 4.7. However, the correlation coefficient for both polarizations is nearly identical with $\rho = 0.62$ and $\rho = 0.68$ for VV and VH polarized data, respectively. The cross-ratio $^{VV}/_{VH}$ appears to be unrelated to NDVI values with $\rho = -0.10$.

The Pearson's correlation coefficient ρ for both NovaSAR-1 and Sentinel-1 data is given in Table 4.6.



(a) HV polarization (b) HH polarization

Figure 4.6: Comparison of NDVI and NovaSAR-1 data for polarizations HH and HV for the Bell Ville study area in Argentina. The local incidence angle is indicated for each data point by its color.

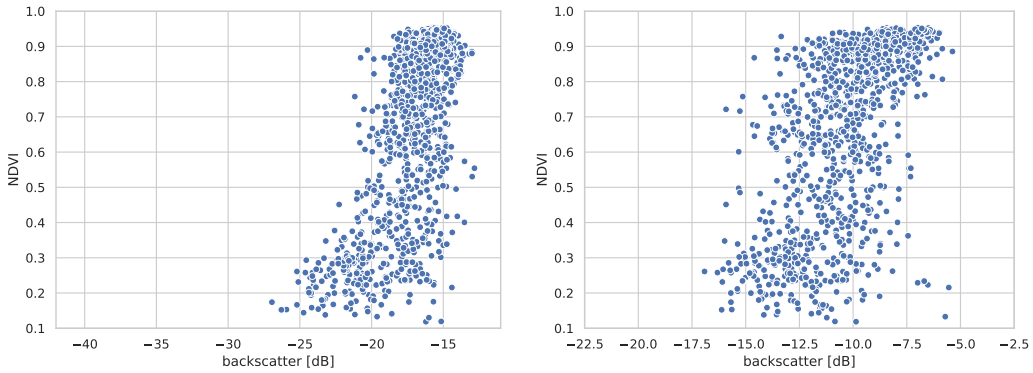
Sensor	Data	ρ	N
NovaSAR	HV	0.68	(706)
	HH	0.35	(703)
	$^{HH}/_{HV}$	-0.48	(704)
Sentinel-1	VV	0.62	(1095)
	VH	0.68	(1096)
	$^{VV}/_{VH}$	-0.10	(1097)

Table 4.6: Relationship between NDVI values and SAR data for the study area Bell Ville (Argentina). Next to the polarizations, the cross-ratios formed from them are used. ρ denotes the Pearson's correlation coefficient and N the number of samples.

4.3.3 Mekong River Delta (Vietnam)

The results of the Mekong River Delta study area differ from those of the previous two regions. The results demonstrate that, in the case of NovaSAR-1 S-band data, the HH polarized data exhibits a positive correlation with NDVI values, rather than the HV polarized data. This can be confirmed by the high correlation coefficient for HH data with $\rho = 0.51$ compared to $\rho = 0.15$ for HV data. As observed in the previous two study regions, the incidence angle influences the backscatter values. The resulting scatter plots, colored by the local incidence angle, are shown in Figure 4.8.

A similar pattern to NovaSAR-1 data is observable for Sentinel-1 data: co-polarized VV data exhibits a high negative relationship with NDVI values with $\rho = -0.59$ but



(a) VH polarization (b) VV polarization

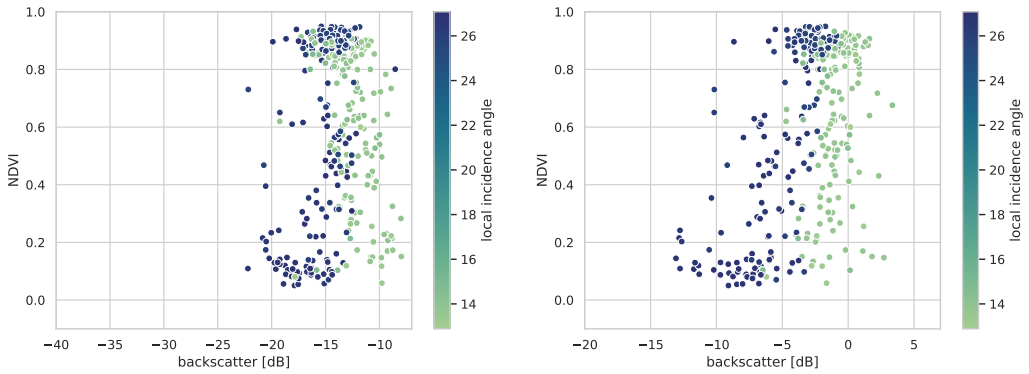
Figure 4.7: Comparison of NDVI and Sentinel-1 data for the two polarizations VV and VH for the Bell Ville study area in Argentina. The incidence angle is 35 to 37°.

cross-polarized data shows no relationship with the NDVI with $\rho = -0.02$. The cross-ratio $^{VV}/_{VH}$ exhibits a similar relationship with the NDVI as the HV polarized data with $\rho = -0.62$. When analyzing the scatter plots in Figure 4.9, the distribution for VH data exhibits two regions. The first region encompasses higher NDVI values above 0.25, where a negative relationship appears to be present. In contrast, a positive relationship is observed for the second region encompassing NDVI values below 0.25. This can be verified by calculating ρ separately for fields with NDVI values above and below 0.25. For vegetated fields (NDVI>0.25) a moderate negative relationship is found with $\rho = -0.43$, whereas for fields with NDVI<0.25 a weak positive relationship is found with $\rho = 0.32$. The relationship between backscatter and NDVI is illustrated in Figure 4.9.

For COSMO-SkyMed (CSG) X-band data, the cross-pol HV data seems to have a negative relationship with the NDVI data with $\rho = -0.59$. In contrast, the co-polarized HH data exhibits no discernible relationship. As CSG data were only available for a single date in August the amount of samples and data diversity is limited which decreases the certainty of the results. Scatterplots of both polarizations are given in Figure 4.10.

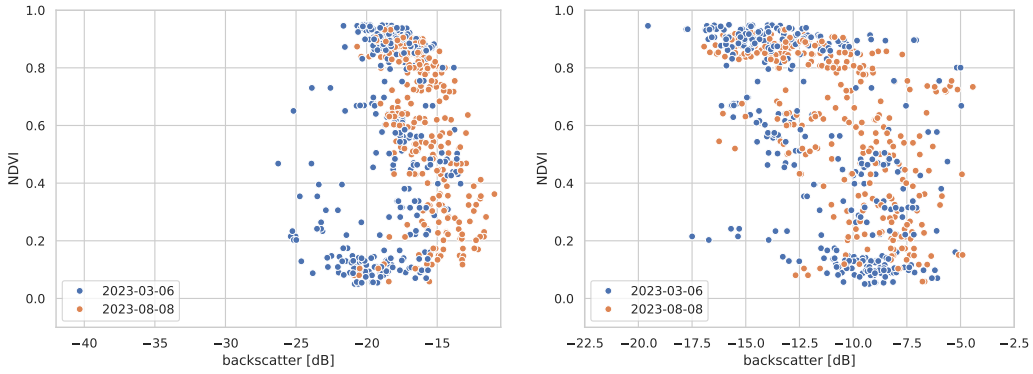
The L-band SAOCOM data represents the fourth investigated SAR frequency and appears to have no discernible relationship with the NDVI data. All backscatter values in both HV and HH polarizations occur for all NDVI values. This is shown in Figure 4.11 and confirmed by correlation coefficients ρ of 0.0 and -0.18 for HH and HV polarization, respectively. Only the cross-ratio $^{HH}/_{HV}$ exhibits a slightly higher ρ of 0.24, indicating a very weak relationship.

An overview of all Pearson's correlation coefficients of the different sensors, their polarizations, and cross-ratios is given in Table 4.7.



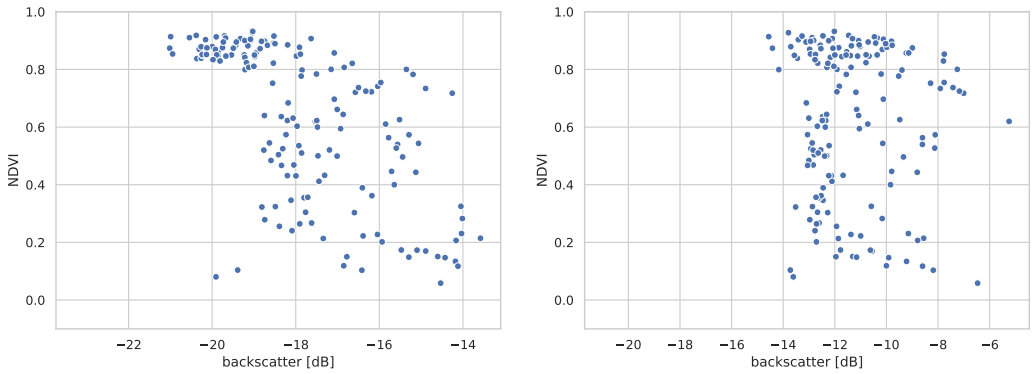
(a) HV polarization (b) HH polarization

Figure 4.8: Relationship between NovaSAR-1 backscatter and NDVI values for the Mekong River Delta study area in Vietnam. The local incidence angle of the NovaSAR-1 data is indicated for each data point by its color.



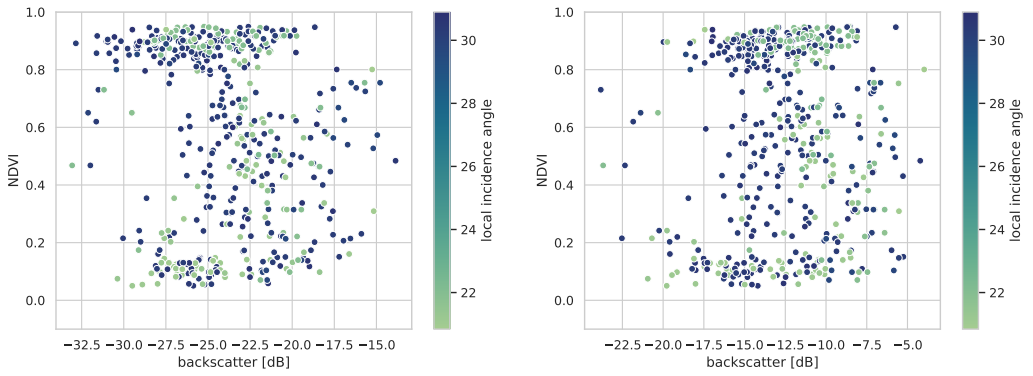
(a) VH polarization (b) VV polarization

Figure 4.9: Relationship between Sentinel-1 backscatter and NDVI values for the Mekong River Delta study area in Vietnam. The Sentinel-1 acquisition data of each sample is indicated by its color.



(a) HV polarization (b) HH polarization

Figure 4.10: Relationship between CSG backscatter and NDVI values for the Mekong River Delta study area in Vietnam.



(a) HV polarization (b) HH polarization

Figure 4.11: Relationship between SAOCOM SAR backscatter and Sentinel-2 NDVI values for the Mekong River Delta study area in Vietnam. The local incidence angle of the SAOCOM data is indicated for each data point by its color.

Sensor	Data	ρ	N
NovaSAR	HV	0.15	313
	HH	0.51	313
	CR $^{HH}/_{HV}$	0.51	313
Sentinel-1	VV	-0.59	633
	VH	-0.02	632
	VH for NDVI < 0.25	0.32	143
	VH for NDVI > 0.25	-0.46	485
	$^{VV}/_{VH}$	-0.62	629
SAOCOM	HH	0.00	478
	HV	-0.18	478
	$^{HH}/_{HV}$	0.24	479
CSG	HH	-0.01	153
	HV	-0.59	152
	$^{HH}/_{HV}$	0.61	152

Table 4.7: Relationship between NDVI values and SAR data for the study area Mekong River Delta (Vietnam). ρ denotes the Pearson's correlation coefficient and N the number of samples. Next to the polarization, the cross-ratio formed from them is used. Additionally, for Sentinel-1 cross-pol VH data, the relationship is given for NDVI values above and below 0.25, because two distinct relationships are apparent in the scatterplot of the data (cf. Figure 4.9).

4.4 Discussion

In the previous section, a relationship between NDVI and C- and S-band SAR backscatter and potentially also with X-band data could be shown. No such relationship was observed in the L-band data.

The results for two of the three areas, Boort (Australia) and Bell Ville (Argentina), indicate a strong correlation between S- and C-band cross-polarized backscatter and NDVI. This is to be expected, as the volumetric scattering of vegetation increases with increasing vegetation volume, which in turn leads to higher cross-polarized responses (Flores et al. 2019). In Bell Ville, not only cross-polarized but also co-polarized S- and C-band data exhibited a notable relationship with the NDVI, which may be due to the different crop types. While broad-leaf crops like soybeans and maize are the predominant crops in Bell Ville, narrow-leaf crops like wheat and barley are the primary crops in Boort (cf. Section 4.1.1). The distinct backscatter characteristics of these two crop types (Macelloni et al. 2001) may be responsible for the different backscatter–NDVI relationships observed. For instance, one study demonstrated that the crop height of corn, a broad-leaf crop, exhibits a good relationship with both HH and HV RADARSAT-2 backscatter. However, for wheat, a narrow-leaf crop, only HV polarized data demonstrated a significant relationship (Liao et al. 2017).

Contrastingly, the third study area comprising rice paddy fields in Vietnam exhibited distinctive backscatter characteristics that set them apart from the other two regions. For these fields, co-polarized HH and VV data were related to NDVI, while cross-polarized HV and VH data were not. It is unclear, why a relationship can be found only for co-polarized and not cross-polarized data, especially given that other researchers have identified a relationship between plant parameters and backscatter for paddy rice. For instance, the cross-polarized backscatter of a ground-based scatterometer has been demonstrated to relate well to the leaf area index (LAI), fresh weight or biomass, and canopy height (Inoue et al. 2002; Jia et al. 2014). Additionally, RADARSAT-2 backscatter and phenology are linked for rice (Lopez-Sanchez et al. 2014). Another particularity of the cross-polarized VH C-band data is the existence of two distinct relationships between C-band VH backscatter and NDVI values, one for NDVI values below and another one for values above 0.25. A possible explanation for this phenomenon is that low NDVI values, which are likely due to the presence of water surfaces without much vegetation cover, result in a low backscatter, which is to be expected given that water acts as a specular surface and mirror, reflecting most of the signal's energy away from the sensor. As more of the surface is covered by rice and not by water, the NDVI as well as the backscatter increase because the surface becomes less specular. As the NDVI increases above 0.25, a negative relationship between NDVI and VH backscatter occurs with a VH backscatter decrease for increasing NDVI values. The reason for this behavior is currently unknown. It is similarly unclear, why these two relationships could not be observed for S-band data. One potential explanation might be attenuation whereby the vertically polarized waves

interact strongly with the vertically oriented rice stalks leading to some attenuation. It remains unclear, why volumetric scattering which leads to a high cross-pol response and is typically high for dense vegetation, is not the dominant factor in this case.

An unknown factor influencing the result might be the availability of optical imagery. In the event that optical data is only available for certain vegetative periods, a comprehensive analysis cannot be conducted. In particular, if the monsoon period, which is often accompanied by heavy cloud cover, coincides with the transplanting of rice and the early growth stages, when the NDVI increases, then no data may be suitable for analysis. For rice, this presents a more significant challenge, as the crop develops from seedlings to mature plants in less than 30 days, and the NDVI increases from approximately zero to nearly one during this period (Onojeghuo et al. 2018). Capturing this rapid change is difficult due to the frequent cloud cover. Consequently, only the later vegetative stages can be readily captured with optical data and analyzed.

The different signs of ρ of S-band HH and C-band VV backscatter indicating positive and negative relationships can be explained by the geometric properties of the rice plants. One strong scattering effect for this area with the inundated surface is double bounce, which typically results in a higher HH response (Flores et al. 2019). Additionally, the vertical orientation of the rice stems contributes to the higher HH response relative to the VV one (Oh et al. 2009).

4.4.1 Data Availability and Quality

One problem of the analysis is the availability of suitable data. For cost reasons, the investigation relied exclusively on archival data. Within these data archives, it was necessary to identify acquisitions with matching areas and dates. While this is easy to do for Sentinel-1 data due to it monitoring the whole Earth regularly, all other sensors only capture data from limited areas, which overlap only by chance. Therefore, four SAR bands could only be used for one study area. If money were not a constraint, one solution to this problem could be to task images. This would further increase the value of the analysis, as a low temporal distance can be assured. At the same time, there is a risk that optical data will be unavailable due to cloud cover, which would thwart the entire analysis.

Another issue was the quality of the data provided. Data from two sensors had an inaccurate geolocalization, and NovaSAR-1 data lacked corrections for antenna patterns and scalloping. While geolocation errors could be manually corrected, this was only applied to the geocoded data, not the unprocessed data, resulting in slightly inaccurate backscatter values. The missing corrections for NovaSAR-1 data were not addressed. While these shortcomings did not severely limit the analysis, they nevertheless complicate the utilization of the data and the manual corrections entail additional work.

4.4.2 L- and X-band Sensitivity to Vegetation Indices

No discernible relationship was identified between L-band backscatter and NDVI values. This is unexpected, given that other studies have demonstrated a high sensitivity of L-band data for LAI and total fresh weight (Inoue et al. 2002), which are in turn related to NDVI (Ryu et al. 2020). In contrast, Paloscia (2002) notes a difference in biomass sensitivity between broad-leaf and narrow-leaf crops such as rice. L-band data proved more effective for broad-leaf crops, while C-band data demonstrated greater suitability for narrow-leaf crops. This finding is consistent with the observations of Hosseini et al. (2017), who noted that RADARSAT C-band data outperformed UAVSAR L-band data in estimating wheat biomass, a crop with a similar plant structure to rice.

The results for the X-band data were inconclusive due to the availability of only one suitable scene. A relationship between cross-polarized HV data and NDVI may exist, but this is not the case for HH co-polarized data. This finding is inconsistent with the results of other research, which demonstrated a relationship between HH X-band data and NDVI values as well as with the leaf area index (LAI) for 30 flooded rice paddy fields with a high ρ of 0.584 (Hirooka et al. 2015).

However, another study investigated wheat and barley, which have a similar plant structure to rice but are not cultivated on flooded fields, and demonstrated a negative relationship between co-polarized HH and VV X-band data and LAI (Fontanelli et al. 2013). Consequently, further research is required to examine the behavior of X-band data in relation to the NDVI and plant parameters.

4.4.3 Effectiveness of the Cross-Ratio

This study also evaluated the effectiveness of the cross-ratio (CR), finding it beneficial only in certain cases, dependent on the specific study area and sensor used. In a few areas, the relationship between CR and NDVI values was comparable or slightly stronger than that observed with a single polarization. If a study would like to use only one feature, calculating the CR would make selecting the optimal polarization unnecessary. However, in some cases, no relationship between CR and NDVI could be demonstrated, whereas with a single polarization a relationship existed. This illustrates, that no universal guidance can be derived from the data, and in each case, the utility of the CR must be reassessed.

4.4.4 Exploring the Relationship Between NDVI and Interferometric Coherence

Although interferometric coherence is a valuable indicator of vegetation state, particularly for short temporal baselines (Villarroya-Carpio et al. 2024), it was not analyzed

here. This is primarily due to limitations in data availability. Of the four investigated SAR sensors, only the full Cosmo-SkyMed constellation, comprising both first and second-generation satellites, can provide interferometric acquisitions with a temporal baseline between one to six days. This could warrant a separate investigation, but similar X-band data of the TerraSAR-X, TanDEM-X, and PAZ sensors has already been investigated (Villarroya-Carpio et al. 2024). The Sentinel-1 mission allows interferometric analysis, but currently only for a temporal baseline of 12 days due to the failure of Sentinel-1B at the end of 2021. Such a long temporal baseline is unsuitable for most vegetation types. Following the anticipated launch of Sentinel-1C at the end of 2024, temporal baselines of six days will once again be available for this C-band sensor. NovaSAR-1 and SAOCOM are unsuitable for interferometric analysis of vegetation. Their orbital tubes are not narrowly controlled to maintain consistent spatial baselines. Only by chance a suitable spatial baseline for InSAR analysis is achieved (Li et al. 2022; Roa et al. 2021). This results in very long temporal baselines, significantly reducing the coherence for vegetated surfaces, rendering their data unsuitable for such an investigation.

4.4.5 Implications for SAR-Based NDVI Estimation

All these results show that the relationship between SAR backscatter and NDVI values is multifaceted and influenced by various factors. Our findings suggest that while a discernible relationship exists between backscatter and NDVI, its complexity is heightened by a number of factors. These include vegetation-related factors such as crop type, plant phenology, plant shape, or leaf arrangement; factors related to the SAR data such as SAR frequency and polarization; and geometric considerations including incidence angle and field orientation. In particular, different crop types, characterized by unique phenological stages, plant shapes, and leaf arrangements, exhibit varying backscatter responses. These findings underscore the challenges in developing a universal model for NDVI retrieval from SAR data due to the complex interplay of these factors. A generalized model would require the simultaneous disentanglement of the influence of all the different factors, which is a challenging task.

In conclusion, while this chapter confirms the relationship between SAR backscatter and NDVI for specific frequency bands and conditions, it also highlights several challenges and gaps that require further investigation to optimize SAR-based vegetation monitoring.

5 Single Image NDVI Estimation from SAR Imagery

What You'll Learn SAR images can be translated to NDVI values by exploiting the relationship between them. This is done here using a U-Net trained with a global and balanced dataset created for this task. The good performance, both numerically and visually, is demonstrated together with the global applicability.
Based on This chapter's material originates from the article "A Globally Applicable Method for NDVI Estimation from Sentinel-1 SAR Backscatter Using a Deep Neural Network and the SEN12TP Dataset" published in the *Journal of Photogrammetry, Remote Sensing and Geoinformation Science* (Roßberg et al. 2023a).

To address the challenge of vegetation monitoring in areas frequently obscured by cloud coverage, synthetic aperture radar (SAR) imagery can be translated into vegetation index (VI) measures, specifically the normalized difference vegetation index (NDVI). This leverages SAR's ability to penetrate cloud cover and provide weather-independent imagery. Therefore, the goal of this chapter is to develop a method for estimating NDVI values from single SAR images globally. To achieve this goal, a large deep learning model is trained, which requires an extensive datasets comprising paired optical and SAR imagery captured over the same areas and at the same days.

5.1 The SEN12TP Dataset

For training a deep neural network a large dataset comprised of SAR and cloud-free optical imagery with a small temporal distance is required. Existing datasets are unsuitable (cf. Section 3.5), because they either do not contain image pairs of the same day, are not filtered for clouds, or do not have a global coverage. Therefore, a new dataset has to be created.

The creation of the SEN12TP called dataset (Sentinel-1 and -2 images, timely paired) is described in the following sections. This description covers the sampling of the images, the remote sensing data incorporated, and the characterization of the resulting

dataset. All steps were implemented using Google Earth Engine (GEE) (Gorelick et al. 2017).

5.1.1 Selection of Image Locations

With the goal of a versatile and globally applicable model in mind, a balanced and geospatially uniformly distributed training dataset has to be created. To capture global vegetation conditions as well as possible, all land cover types, climates, and seasons have to be taken into account. Their distribution should be balanced to ensure an unskewed model training. As a basis for the data sampling, the global land cover product *Copernicus Global Land Service (CGLS) 100m version 3* is chosen (Buchhorn et al. 2021). To account for different climate zones the Köppen-Geiger classification is used (Peel et al. 2007), as it is based on the vegetation found on the ground (cf. Section 2.4). Different seasons are incorporated by selecting dataset samples using the month of each data acquisition.

5.1.2 Selection of Suitable Areas and Images

To select suitable areas and images for the dataset, three main steps are carried out: firstly, candidate points are selected and filtered, then these points are transformed into regions of interest (ROIs), and finally their data is downloaded. An overview of all steps detailed in the next paragraphs is given in the flowchart in Figure 5.1.

Candidate points are selected using the climate zone and land cover class. Both climate zones and land cover classes are simplified as described in Table 5.1 and Table 5.2, respectively. Additionally, the CGLS land cover class *water bodies and open sea* is excluded as well as the polar areas north of 78° N and south of 60° S as they are not vegetated. The land cover product is reprojected to 2 km scale and thereby downsampled to ensure that the surroundings of a sampled point contains pixels of the same class. If the original 100 m resolution would be used to sample candidate points for ROIs there is a chance that only one pixel of the this land cover class would be in the ROI. For example a small park in an urban area: it could be selected as candidate for a forest but in fact most of the surrounding area is urban. However choosing a too large of a scale results in missed land cover types with smaller spatial extends like small villages or forest stands. Therefore the chosen resolution for sampling is a trade-off between including smaller objects and a perfectly balanced pixel distribution.

A stratified sample of points is drawn from the combination of simplified and downsampled land cover classes and climate zones. Neighboring points are filtered to avoid overlapping ROIs. Sentinel-1 and -2 sensing times between 28 March 2017 and 31 December 2020 were retrieved after selecting only cloud-free images and images covering the area. Using the retrieved sensing times, pairs of radar and optical images

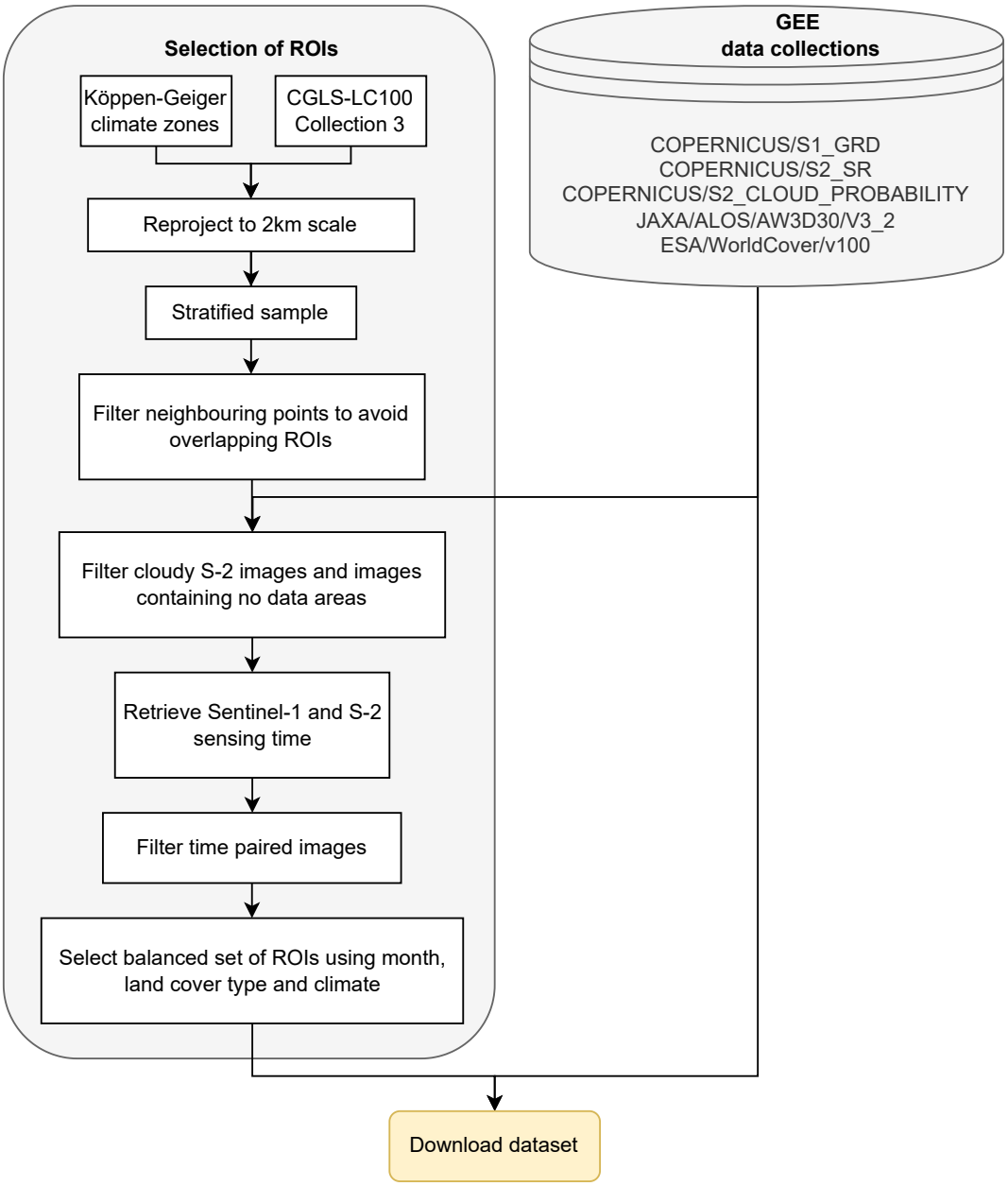


Figure 5.1: Overview of the selection process of the regions of interest (ROI) of the dataset. ROI candidate points are selected using the CGLS landcover and the Köppen-Geiger climate zones and are then filtered. For all final ROIs data of multiple data collections is downloaded.

are selected which were taken within 12 h from each other. This avoids temporal decorrelation between radar and optical values. Finally, image pairs are selected from all suitable ones to ensure a balanced distribution across climate zones, land cover classes, and months of the year.

Around each filtered candidate point a square with an edge length of 20 km is determined, which form the ROIs of the dataset. This leads to a imbalanced land cover distribution, as only the center of a ROI is required to be of the specified land cover. However, choosing ROIs with a smaller size would require a higher number of ROIs when aiming for the same dataset size. As this would become computationally expensive and prohibitively time consuming, we refrained from doing so.

All ROIs are to contain at most 5 % clouds and 5 % no data areas. For cloud filtering, the cloud probabilities calculated using the *s2cloudless* package (Zupanc 2017) were used, which are contained in the GEE data collection `COPERNICUS/S2_CLOUD_PROBABILITY`. From the cloud probability the cloud masks are generated with a cloud probability threshold of 20 %. Stricter filtering would exclude too many ROIs. Cloud shadows are not filtered, because with cloud free imagery, only a marginal amount of pixel could be affected by cloud shadows, cast from clouds outside the ROI.

Table 5.1: Mapping from all to simplified Köppen-Geiger climate classes for sampling of the SEN12TP ROIs. For explanation of the codes of the climate classes refer to Peel et al. (2007).

Simplified class			**Original classes**
	Af	Tropical, rainforest	Af
	Am	Tropical, monsoon	Am
	Aw	Tropical, savannah	Aw
	BW	Arid desert	BWh, BWk
	BS	Arid Steppe	BSh, BSk
	Cs	Temperate, dry summer	Csa Csb Csc
	Cw	Temperate, dry winter	Cwa, Cwb, Cwc
	Cf	Temperate, no dry season	Cfa, Cfb, Cfc
	Ds	Cold, dry summer	Dsa, Dsb, Dsc, Dsd
	Dw	Cold, dry winter	Dwa, Dwb, Dwc, Dwd
	Df(a|b)	Cold, no dry season 1	Dfa, Dfb
	Df(c|d)	Cold, no dry season 2	Dfc, Dfd
	ET	Tundra	ET, ETH
	EF	Frost	EF, EFH

Table 5.2: Mapping of the Copernicus Global Land Cover classes to simplified ones. The classes of the CGLS-LC100 Collection 3 are described in Buchhorn et al. (2021).

Original classes	Simplified class
20	Shrubs
30	Herbaceous vegetation/grassland
40	Cropland
50	Urban and built up
60, 100	Bare, sparse vegetation, moss, lichen
70	Snow and Ice
80, 200	Permanent water bodies and open sea
90	Herbaceous wetlands
111-116, 121-126	Forests

5.1.3 Remote Sensing and Auxiliary Geospatial Data Included in the Dataset

For each ROI of the dataset, different modalities and remote sensing products are retrieved. The main modalities are SAR and optical imagery, as they are needed for the SAR-to-NDVI-translation task. Next to that, elevation data is included, as Scarpa et al. (2018) have shown an improved performance when using it as auxiliary input. Furthermore, land cover data was found as one of the most important input variables by Santos et al. (2022), thus it is also included in the dataset.

The GRD Sentinel-1 product `COPERNICUS/S1_GRD` provided in the GEE data catalog consists of σ° backscatter values in dB. The following preprocessing steps are applied as described in GEE's documentation[1]: application of the orbit file, border and thermal noise removal, calibration, and geocoding. However, due to the side-looking sensor geometry, the terrain leads to geometric artefacts (cf. Section 2.2), which require a more complex correction of the radar backscatter (Small 2011). This could negatively impact the relation to the NDVI, due to different radar backscatter values being sensed for the same ground conditions depending on the terrain and radar geometry. To counter this effect, terrain flattening can be applied (Small 2011). This flattening is done on GEE using the model optimized for volume scattering on the ground as described in Vollrath et al. (2020). Including σ° as well as the terrain flattened γ° backscatter allows a comparison of both backscatter values. The different effects of geocoded and terrain flattened data for both σ° and γ° backscatter coefficients are illustrated in Figure 2.13.

The optical data of the Sentinel-2 Level-2A data collection `COPERNICUS/S2_SR` is used. It contains 12 spectral bands of bottom of atmosphere (BOA) reflectance data where

[1] https://developers.google.com/earth-engine/guides/sentinel1

atmospheric effects are corrected using *sen2cor* (Main-Knorn et al. 2017). Additionally, the cloud probability from `COPERNICUS/S2_CLOUD_PROBABILITY` is added as a thirteenth band.

For elevation data the 30m resolution ALOS World 3D DEM (version 3.2) (Takaku et al. 2020), contained in the `JAXA/ALOS/AW3D30/V3_2` data collection is included. That source of elevation data was chosen because its global coverage and high accuracy (Uuemaa et al. 2020). The ESA WorldCover v100 land cover product is included (Zanaga et al. 2021) as it is currently the highest-resolution land cover product available, boasting a 10 m resolution and a high accuracy. It is contained in the GEE collection `ESA/WorldCover/v100`.

For each ROI Sentinel-1 σ° and γ° radar backscatter, Sentinel-2 BOA reflectances, ALOS World 3D DEM, and ESA WorldCover data is downloaded. Images and bands with a lower resolution than 10 m were upsampled using nearest neighbor upsampling.

5.1.4 Final Dataset

The final SEN12TP dataset consists of timely paired Sentinel-1 and -2 images covering 2319 image pairs and has a size of 221.6 GB. The images cover 484 400 km^2 and 1236 distinct areas, given that from some areas images from several months were selected. The spatial distribution of the ROIs is shown in Figure 5.2 and one sample of the dataset is shown in Figure 5.3 to visualize the different modalities.

The distribution for the different land cover types is shown in Table 5.3. Despite the balanced sampling, the distribution is unbalanced: for instance belong only 3 % of all pixels to *Snow and Ice*, but 26.5 % belong to *Grassland* and 25 % belong to *Trees*. The imbalance stems from the selection of points: the sampling points have a balanced landcover distribution with the surrounding 2 km square area considered. However, the surrounding area of the ROI can be different and in theory only 1 % (2 km × 2 km of 20 km × 20 km) of each ROI is required to be of this land cover class. For vegetation monitoring, *Grassland*, *Trees*, and *Cropland* are the most important land covers types and they are also the most common classes in the dataset. Therefore, the differences in land cover occurrence do not pose a problem for our application. If this dataset is used for other applications this imbalance has to be taken into consideration.

The distribution of daily sensing times of all optical and radar images is shown in Figure 5.4. All sensor platforms have a sun-synchronous orbit, therefore the sensing time does not differ much for each platform. Most images from each platform are taken in a 2 to 4 h range, mostly due to differences between the local time zone and local solar time. Optical images are only acquired during the ascending orbit by the Sentinel-2 platforms at 10:30 mean local solar time, which is a compromise between a good illumination at noon and less cloud coverage before noon (ESA 2015, Sec. 1.5.1). During the descending orbit at night, no images can be acquired.

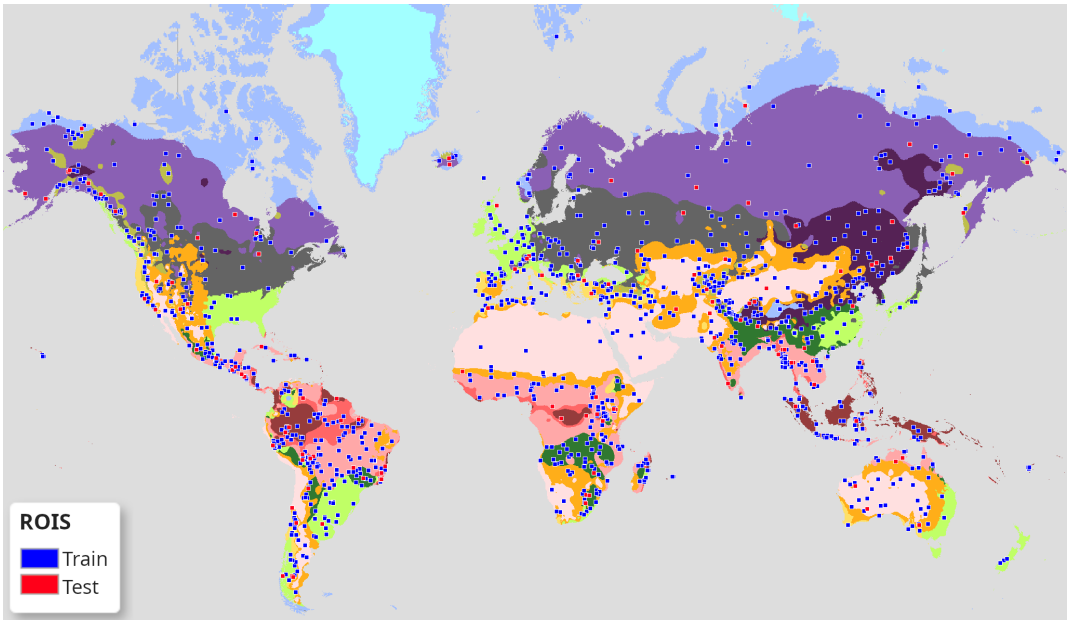


Figure 5.2: Global distribution of the selected ROIs, split into training (blue squares) and test regions (red squares). Underlayed are the simplified Köppen-Geiger climate classes. For the color legend refer to Table 5.1.

In contrast, both orbits can be used for sensing with radar data: the descending orbit takes data in the morning around 7 AM local time whereas the ascending orbit occurs in the evening around 7 PM local time.

This affects the dataset in two ways: Firstly, pairs of optical and radar data cannot have an identical sensing time. The minimum time difference of image pairs is around 4.5 h. Secondly, diurnal weather cycles affect data differently for ascending and descending radar orbit, mainly due to water on the vegetation. This water changes the radar response without changing the vegetation condition and thereby complicates the retrieval of vegetation parameters from radar data. In the tropics, the water stems from precipitation which is more frequent in the afternoon and evening (Kikuchi et al. 2008). In moderate climates, dew can affect the images acquired in the morning (Khabbazan et al. 2019). Therefore, the model has to compensate the effect of water on the vegetation and the imperfectly paired relation of optical and radar data.

The final dataset is published on Zenodo (Roßberg et al. 2023b). The uploaded archive file consist of a directory for each scene and the directory of each scene contains the .tif image files of Sentinel-1, Sentinel-2, DSM, and ESA WorldCover land cover data.

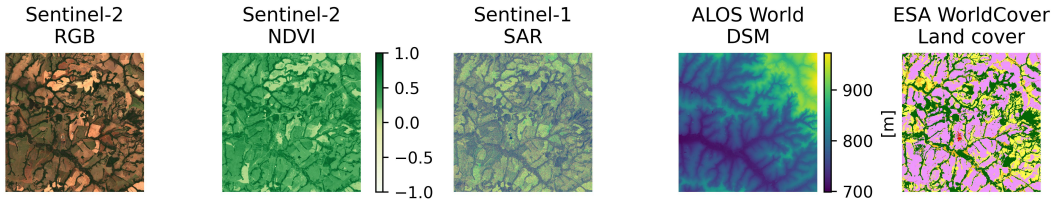


Figure 5.3: One sample of the dataset. Displayed are the Sentinel-2 derived RGB and NDVI image, the Sentinel-1 SAR data as a false-color image (R: $\sigma^\circ_{\mathrm{VV}}$, G: $\sigma^\circ_{\mathrm{VH}}$, B: $\sigma^\circ_{\mathrm{VV}}/\sigma^\circ_{\mathrm{VH}}$), the ALOS World 3D DSM, and the ESA WorldCover v100 land cover map. For a color legend of the ESA WorldCover refer to Figure 2.16. Each image has a size of 2.5 km × 2.5 km.

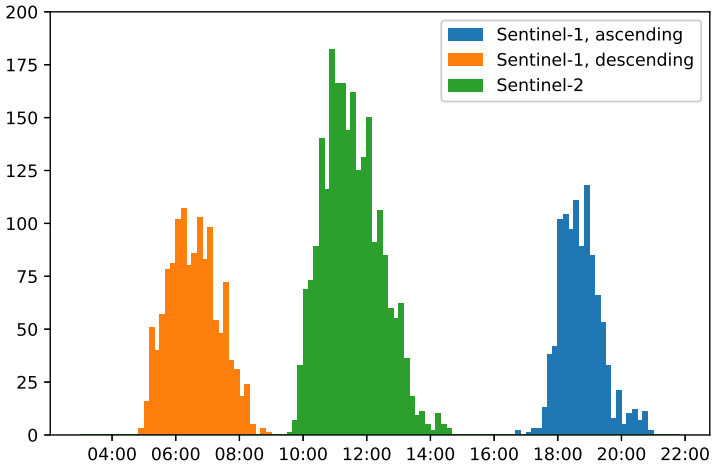


Figure 5.4: Distribution of the sensing times of the optical images, given in the local timezone. All Sentinel-2 images are taken at the ascending orbit.

5.2 NDVI Estimation from SAR

5.2.1 Deep Learning Model

To estimate the NDVI from SAR backscatter data, a slightly adapted U-Net architecture is used (Ronneberger et al. 2015). This is one of the most established architectures for semantic segmentation and pixel-wise regression purposes, and is also commonly used in remote sensing (Yuan et al. 2021). U-Net models are a special convolutional neural network (CNN) (see Section 2.5.1) and consist of two parts: an encoder branch and a decoder branch. The encoder extracts meaningful features, and to do so, reduces the spatial resolution of the input, in our case SAR backscatter values. For this, the spatial neighborhoods of each pixel is considered using convolution kernels.

Table 5.3: Pixel distribution of the ESA WorldCover v100 land cover classes and the percentage of each class for all pixels of the SEN12TP dataset.

Worldcover Class	Pixels	%
Trees	2326.7×10^6	25.1
Shrubland	671.4×10^6	7.2
Grassland	2460.1×10^6	26.5
Cropland	1118.2×10^6	12.0
Built-up	213.2×10^6	2.3
Barren/sparse vegetation	1192.7×10^6	12.8
Snow and ice	236.9×10^6	2.6
Open water	456.1×10^6	4.9
Herbaceous wetland	350.6×10^6	3.8
Mangroves	18.5×10^6	0.2
Moss and lichen	240.3×10^6	2.6
Sum	9284.7×10^6	100.0

The decoder calculates the output NDVI values from the extracted and compressed features. To avoid losing spatial detail in the output, skip connections are added between the encoder and decoder which transfer high-frequency details. The U-Net architecture id depicted in Figure 5.5

The original U-Net was slightly modified for SAR-to-NDVI translation. Originally, two output channels with no activation function were used to create probability maps of background and foreground classes. In contrast, here only one output channel is used, which is passed into a sigmoid activation function. This ensures that the model output $\hat{y}$ is limited to $[0, 1]$. To get the predicted NDVI values y_{NDVI}, this output has to be transformed to [-1, 1] using the equation $y_{\text{NDVI}} = 2(\hat{y} - 0.5)$. Additionally, instead of up-convolution layers, bilinear upsampling with standard convolution layers was used in the decoder. This avoids checkerboard artifacts in the output image (Odena et al. 2016). To train the model, the $\mathcal{L}_1$ loss (mean absolute error, MAE) is utilized, because the amount of spatial detail is better than with the $\mathcal{L}_2$ loss (mean squared error, MSE) (Zhao et al. 2017).

5.2.2 Training Procedure and Data Preprocessing

To input the data into the model it has to be preprocessed and normalized which is done in multiple steps.

First, to ensure a fair evaluation, the 20 km × 20 km ROI images are split up spatially into a training, validation, and test set. This spatial split ensures that the model is

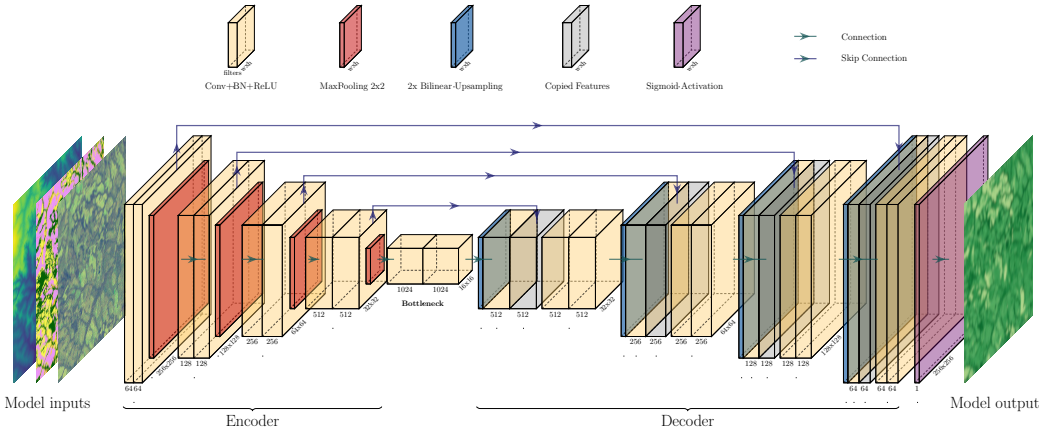


Figure 5.5: U-Net architecture for SAR-to-NDVI image translation based on Ronneberger et al. (2015). Features are extracted from the model inputs in the Encoder and transfromed by the Decoder into the model output. Unlike in the original paper, the feature maps are enlarged by a bilinear upsampling. *Conv*, *BN*, and *ReLU* denote convolutional layers, batch normalization, and ReLU activation functions, respectively. The number of filters a convolutional layer has is shown below each layer, next to it are the width and height w × h.

not trained with images which are also used for evaluation. 80 % of the ROI images are used for training and both validation and test set consist of 10 % of the scenes.

For training, the large ROIs are split up into smaller patches with a size of 256 px × 256 px. A stride of 249 px is chosen to split each 2000 px scene into 8 × 8 tiles. Then, patches containing clouds or areas with missing or invalid data are removed on the fly since some scenes contain small clouds or missing data areas. Filtering bad data areas is done in this two step fashion because creating a perfectly clean dataset with large ROIs is infeasible. Only a small number of tiles is excluded in this step, which shows the high data quality of the dataset: for the train set only 2172 of the 134 080 (1.6 %) tiles of the train set are excluded. After filtering the data it is clipped to remove outliers and anomalous sensor values. Values exceeding the clipping threshold are set to the threshold value. The multi-spectral channels of the Sentinel-2 data, the VV and VH polarizations of the Sentinel-1 data, and the DEM are clipped to $[0, 10\,000]$, $[-25\,\text{dB}, 0\,\text{dB}]$, $[-32.5\,\text{dB}, 0\,\text{dB}]$, and $[-450\,\text{m}, 9000\,\text{m}]$, respectively. The SAR clipping values are the same as in (Meraner et al. 2020), the DEM clipping values are the approximate lowest and highest possible elevation of the earth. The Sentinel-2 values are clipped at 10 000 because this equals a reflectance of 100 %. Finally, all values, including the land cover classification, are normalized using min-max-normalization (cf. Chapter 2).

The NDVI is calculated on the fly from the Sentinel-2 images with the red band `B4`

and the infrared band B8 as in Equation 2.6. As the model output is bounded to $[0, 1]$ by the virtue of the output sigmoid function, the NDVI is transformed to be in that interval for loss calculation.

For model training the Adam optimizer (Kingma et al. 2014) with an initial learning rate of 1×10^{-3} is used. Learning rate scheduling is used and the learning rate multiplied by 0.5 every seven epochs resulting in a learning rate of 5×10^{-4} after seven epochs, 2.5×10^{-4} after 14 epochs and so on. Early stopping is used to avoid overfitting by ending training when the performance does not improve anymore. For each epoch, the MAE of the validation set was calculated. When after a patience period of ten epochs no improvement of the MAE was achieved, training was stopped. The batch size is set to 32 and no data augmentation was used.

All models were implemented using Python 3.7, PyTorch 1.10, and PyTorch Lightning 1.3.5. Training was conducted on two GPUs and took approximately 27 min and 38 min per epoch using an NVIDIA RTX A6000 or an NVIDIA Quadro RTX 8000, respectively.

5.3 Experiments and Results

With the created SEN12TP dataset, the model is trained and evaluated. The performance of the model is evaluated numerically (Section 5.3.1), visually for example regions (Section 5.3.2), in comparison with a similar, but region-specific approach (Section 5.3.3), and by creating radar-derived NDVI time series with a stack of SAR images of an area (Section 5.3.4).

The performance is evaluated using the MAE, the MSE, the peak signal-to-noise ratio (PSNR), and the structural similarity index (SSIM), which are described in Section 2.5.4.

5.3.1 Backscatter and Auxiliary Input Performance

To compare the importance of the radar backscatter and auxiliary input data for NDVI estimation, models trained with different inputs are compared. For this, the predictions of the SEN12TP test set are calculated and compared to the actual optical NDVI. This allows a fair evaluation of the model performance, as these scenes are not used during training or validation and contain completely new and unseen data for the model. For this evaluation the scenes were not split up into smaller patches like during training to avoid border artifacts. A batch size of 1 had to be used, as only one 2000 px × 2000 px large image fit into the GPU memory. Each model was trained five times with different splits of the dataset to get the average model performance.

Table 5.4: Performance of models trained with radar backscatter data and the ALOS World 3D DSM elevation data on the test set. γ° denotes the backscatter with radiometric terrain flattening, σ° the one without. For each model, five models were trained and the average and the standard deviation of the model performances were reported. ↑ denotes that higher values are better, ↓ denotes that lower ones are better.

	MAE ↓		**MSE ↓**		**PSNR ↑**		**SSIM ↑**	
Model inputs	mean	std	mean	std	mean	std	mean	std
γ°_{VV}, γ°_{VH}	0.121	0.001	0.032	0.001	21.0	0.07	0.547	0.005
σ°_{VV}, σ°_{VH}	0.122	0.001	0.032	0.001	20.9	0.08	0.545	0.003
γ°_{VV}, γ°_{VH}, DSM	0.122	0.002	0.033	0.001	20.8	0.12	0.543	0.004
σ°_{VV}, σ°_{VH}, DSM	0.123	0.001	0.034	0.001	20.8	0.12	0.543	0.003
σ°_{VH}	0.141	0.002	0.041	0.001	19.9	0.09	0.507	0.003
σ°_{VV}	0.152	0.003	0.047	0.002	19.3	0.18	0.499	0.006

For the first set of experiments, the performance is evaluated using radar backscatter data without (σ°_{VV}, σ°_{VH}) and with radiometric terrain correction (γ°_{VV}, γ°_{VH}), and the elevation retrieved from the DSM. The elevation data is included here, as it is a requirement for the terrain correction and also globally available free of cost.

All models using both radar polarizations VV and VH achieve a good performance very similar to each other with a low MAE of 0.12. There are only insignificant performance differences between the different models with the exception of the model trained with (γ°_{VV}, γ°_{VH}), which is slightly better as shown in Table 5.4. Using only one radar polarization increases the error, in the case of σ°_{VH} to a MAE of 0.14, in case of σ°_{VV} to 0.15.

For the next set of experiments, auxiliary geospatial data in form of the ESA WorldCover land cover map is included. This decreases the MAE to 0.10. Similar to experiments without the land cover map, adding the DSM or applying radiometric terrain flattening does not improve the performance. Results are given in Table 5.5.

In the last set of experiments no backscatter data is used, only the auxiliary data from the ESA WorldCover and DSM. This is meant to provide a lower-end baseline. The MAE using only the land cover is 0.15 and thereby worse than using only backscatter data or using backscatter and land cover data jointly. The performance is even worse when utilizing only the DSM, for which a MAE of 0.26 is achieved. The evaluation for all metrics is given in Table 5.6.

To have an understanding of the performance of our model for different land covers, we evaluate the performance for each land cover class separately. The model shows similar performance for almost all of the land cover classes, with an MAE between 0.10 to 0.11. Only for *open water* a slightly better performance is achieved (MAE = 0.099),

while for *barren/sparse vegetation* the performance is slightly worse (MAE = 0.115). The full results are shown in Table 5.7.

Table 5.5: Performance of models trained with radar backscatter data and the ESA WorldCover land cover map as auxiliary data. For each input combination five models were trained. ↑ denotes that higher values are better, ↓ denotes that lower ones are better.

	MAE ↓		**MSE ↓**		**PSNR ↑**		**SSIM ↑**	
Model inputs	mean	std	mean	std	mean	std	mean	std
γ°_{VV}, γ°_{VH}, WorldCover	0.102	0.002	0.024	0.000	22.3	0.06	0.609	0.005
σ°_{VV}, σ°_{VH}, WorldCover	0.103	0.003	0.024	0.001	22.3	0.16	0.609	0.009
σ°_{VV}, σ°_{VH}, WorldCover, DSM	0.106	0.004	0.025	0.001	22.1	0.24	0.594	0.011

Table 5.6: Performance of models trained only with auxiliary data but not radar backscatter. For each input combination five models were trained and averaged. ↑ denotes that higher values are better, ↓ denotes that lower ones are better.

	MAE ↓		**MSE ↓**		**PSNR ↑**		**SSIM ↑**	
Model inputs	mean	std	mean	std	mean	std	mean	std
WorldCover, DSM	0.154	0.005	0.049	0.003	19.1	0.25	0.535	0.007
WorldCover	0.155	0.005	0.050	0.004	19.1	0.32	0.533	0.015
DSM	0.263	0.003	0.104	0.002	15.8	0.08	0.373	0.011

5.3.2 NDVI Image Prediction

By using the trained model NDVI images can be predicted for all areas of the world from SAR backscatter. When evaluating the model on the test set of SEN12TP a low error with a good level of spatial details is achieved, as shown in Figure 5.6. Especially for forested, grassland, and cropland areas the NDVI is predicted with high accuracy. Even some fine details like roads or urban areas can be predicted properly. Despite the generally very good results, some remaining limitations of the predicted NDVI images become apparent. One problem is the loss of spatial detail making the predicted NDVI images seem slightly less sharp than their original counterparts. One example of this phenomenon is the delineation between field plots. All delineations between land cover types however are captured well due to the inclusion of the WorldCover land cover map. Water bodies and wetlands also present a problem, as their NDVI values are predicted with quite large errors.

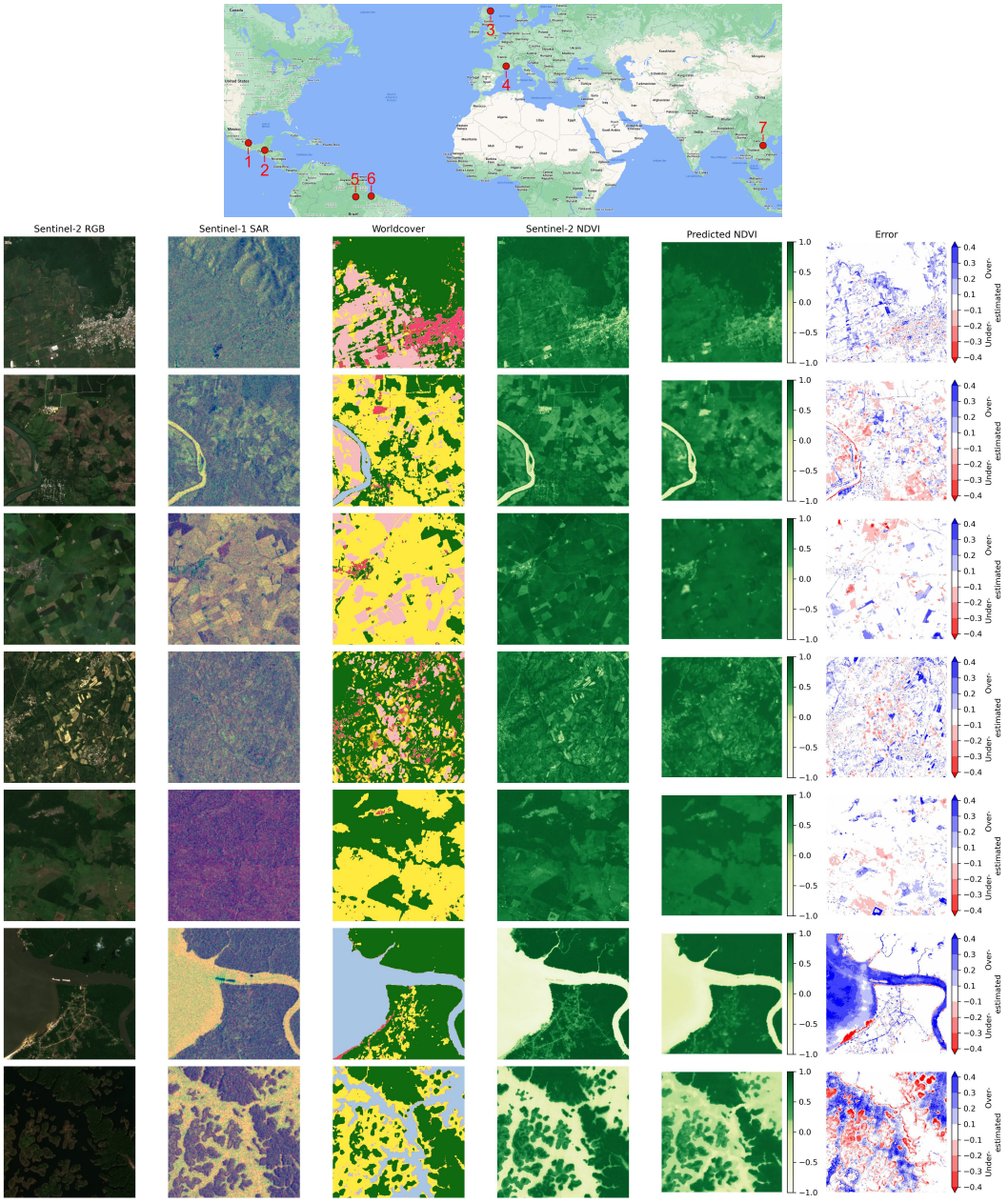


Figure 5.6: Data visualization and comparison of the optical NDVI and the SAR-derived one for images of the hold-out test set.. At the top the location of each image is shown with the number denoting the row below. From left to right are shown the Sentinel-2 RGB, the Sentinel-1 false-color ((R: σ°_{VV}, G: σ°_{VH}, B: $\sigma^\circ_{VV}/\sigma^\circ_{VH}$)), the land cover map, the optical NDVI, the SAR-derived NDVI, and the difference between optical and SAR-derived NDVI.

Table 5.7: Mean absolute error (MAE) per ESA WorldCover class of the predicted NDVI values on the test set. The performance is averaged over five models, each using σ° backscatter data, the WorldCover, and the DSM as input. For each model first the MAE per land cover class is calculated and then the mean is calculated.

Worldcover class	Mean MAE
Open water	0.099
Cropland	0.104
Trees	0.104
Grassland	0.105
Herbaceous wetland	0.105
Mangroves	0.106
Moss and lichen	0.106
Built-up	0.106
Shrubland	0.107
Snow and ice	0.109
Barren / sparse vegetation	0.115

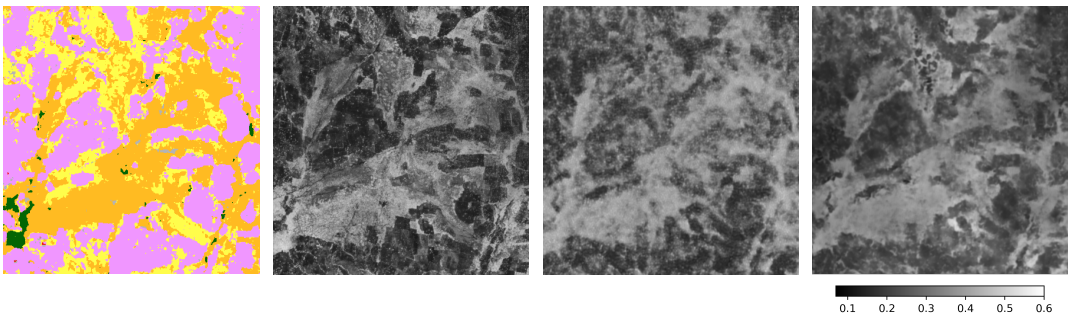
5.3.3 Comparison to a Region-Specific Model

In order to evaluate the performance of the presented approach in comparison to the previously published approach by Scarpa et al. (2018), the two approaches are compared using the area of Scarpa et al. (2018). That study uses data from a small region in the southwest of Burkina Faso with parts of it designated as test region to evaluate model performance. Because the training and test data are next to each other their model is very likely only applicable to that region or similar ones and will likely achieve a suboptimal performance for other regions.

In contrast, the presented model is globally applicable and does not need fine-tuning for specific regions. The closest image the model has seen during training, validation or testing has a distance of more than 150 km from the region used by Scarpa et al. (2018).

To compare both approaches SAR, DSM, and WorldCover data from GEE of the area is downloaded. The data provided by Scarpa et al. (2018) was not used for two reasons: firstly, the same data preprocessing as used for model training had to be ensured. Secondly, there is a slight shift between SAR data and NDVI label pixels in the provided dataset off approximately half a pixel. This shift can be avoided by using the georeferencing of the NDVI images provided by Scarpa et al. (2018). The original NDVI images were used to calculate the evaluation metrics values to ensure a fair evaluation.

A quantitative comparison of our results with the results of Scarpa et al. (2018) is shown in Table 5.8. Our trained model achieves a higher SSIM for all input modalities. This higher level of spatial detail can also be seen in the predicted NDVI images, together with a lower amount of noise as shown in Figure 5.7. In contrast to our high spatial performance, their model achieves a higher Pearson correlation. For the PSNR, our model is better when using only radar backscatter, but their model is better using backscatter and elevation data. The best-performing model of all possible inputs is our model using backscatter, elevation, and land cover data. However, this comparison is not fully fair, as Scarpa et al. (2018) did not evaluate the effectiveness of using land cover information. The results still demonstrate that our model is able to achieve a similar or even better performance than a highly area-specific model trained for a small region.



(a) ESA WorldCover (b) Optical NDVI (c) Prediction from Scarpa et al. (2018) (d) Prediction of the presented approach

Figure 5.7: Visual comparison of the optical and radar derived NDVI for the data of Scarpa et al. (2018), located in the Houet province, Burkina Faso, Africa. The optical NDVI is from 4 June 2016. Both models use the Sentinel-1 radar backscatter from 30 May 2016 together with the elevation data for NDVI estimation. Figures 5.7b and 5.7c are taken from Scarpa et al. (2018). The test site mainly consists Cropland (■) Grassland (■), and Shrubland (■). For a full legend of the WorldCover classes see Figure 2.16.

5.3.4 Naive Densification of NDVI Time Series

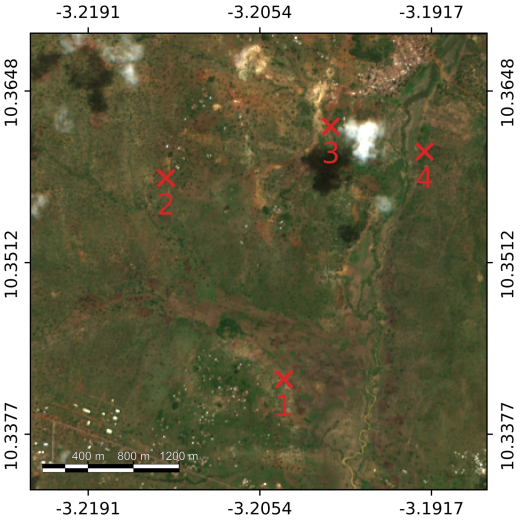
For many applications time series of vegetated surfaces are more valuable than an image of a single date. Therefore, the NDVI was predicted for all Sentinel-1 images of the area used in Figure 1.1 for the year 2019. As model input the σ° backscatter, DSM, and WorldCover data was used. The NDVI was predicted using an ensemble of the five models which enables to calculate the mean and variance of the predicted NDVI values. Together with the NDVI retrieved from all Sentinel-2 images multiple time series were created for several example points with different land covers.

The time series show a high correspondence between the mean predicted NDVI values and the optical retrieved values from cloud-free images as can be seen in Figure 5.8. The increase of the NDVI during the rainy season between June and October is captured by the SAR-derived NDVI predictions. In contrast to the optical values, a higher temporal resolution is achieved, because all images can be used and not only cloud-free ones.

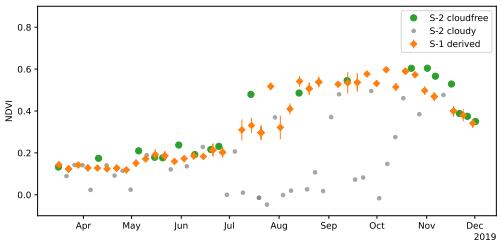
The standard deviation is small for most of the predicted pixels, showing a high agreement and certainty between the models. One outlier of the mean predicted NDVI at the beginning of August is notable. This outlier is observable for all four examples and most pronounced for *Shrubland* where it coincides with a large standard deviation.

Table 5.8: Performance comparison of our model to the model of Scarpa et al. (2018). The average performance of our models trained with five different data splits is averaged (mean) and the standard deviation is stated. Different data was used for the models: SAR (σ°_{VV}, σ°_{VH}), SAR+ (σ°_{VV}, σ°_{VH}, DSM), SARW (σ°_{VV}, σ°_{VH}, WorldCover), and SAR+W (σ°_{VV}, σ°_{VH}, DSM, WorldCover). The best performance comparing our model to the model of Scarpa et al. (2018) using the same input data is *emphasized*. The best overall performance for each metric is marked in **bold**. ↑ denotes that higher values are better, ↓ denotes that lower ones are better.

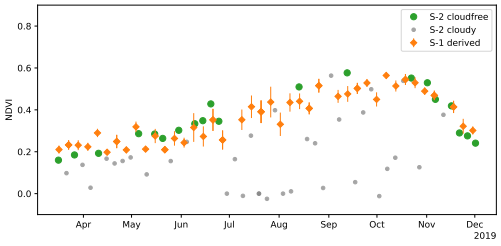
		Pearson ↑		**PSNR ↑**		**SSIM ↑**		**MAE ↓**	
Input data	**Model**	mean	std	mean	std	mean	std	mean	std
SAR	Ours	0.5785	0.0150	*18.41*	0.24	*0.5684*	0.0018	0.1010	0.0028
SAR	Scarpa et al. 2018	*0.6118*	—	16.83	—	0.3942	—	—	—
SAR+	Ours	0.5415	0.0414	17.53	0.97	*0.5563*	0.0075	0.1165	0.0169
SAR+	Scarpa et al. 2018	*0.6207*	—	*18.27*	—	0.4218	—	—	—
SARW	Ours	0.6139	0.0373	17.82	0.90	0.5706	0.0105	0.1086	0.0134
SAR+W	Ours	**0.6370**	0.0360	**18.72**	0.37	**0.5786**	0.0062	**0.1001**	0.0061



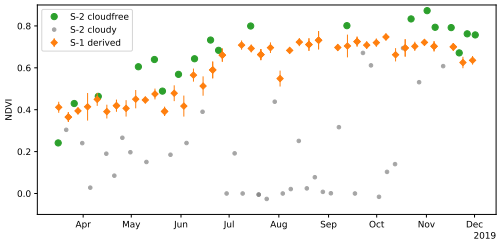
(a) Location of the points with the underlayed Sentinel-2 RGB image of 9. June 2019.



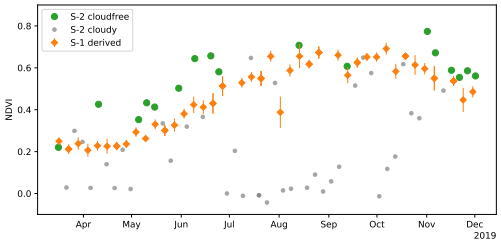
(b) Point 1, land cover *Cropland*



(c) Point 2, land cover *Grassland*



(d) Point 3, land cover *Trees*



(e) Point 4, land cover *Shrubland*

Figure 5.8: NDVI time series for different points of the area in Figure 1.1. Shown are NDVI time series derived with our model and calculated from Sentinel-2 optical data of four image pixels with different land cover classes. Five model predictions were created using the SAR data σ°, the DSM, and the WorldCover as input. The mean of the predicted values is denoted with a diamond, the bar shows the standard deviation of the models. The land cover information is taken from the ESA WorldCover v100 map. Cloud covered values are de-emphasized.

5.4 Discussion

5.4.1 General Validity of the Model

The results summarized in Tables 5.4 and 5.5 and illustrated in Figure 5.6 show the general high performance of our approach, unbiased by the land cover type (cf. Table 5.7). The successful prediction of the NDVI from SAR data demonstrates, that the relationship of SAR backscatter and NDVI values can be exploited despite the different sensing principles and wavelengths used. This is in accordance with the research literature which could also relate backscatter with biophysical parameters and NDVI values (cf. section 3.1).

One problem of the predicted NDVI maps is the spatial resolution: the predicted NDVI is blurrier compared to the Sentinel-2 derived NDVI as visible in Figure 5.6. This is caused by the lower resolution of the SAR input data. The used Sentinel-1 data has a spatial resolution of 20 m × 22 m (IW mode, ground projected, multi-looked data) (Bourbigot et al. 2016), whereas the Sentinel-2 data has a spatial resolution of 10 m (Drusch et al. 2012), even though both products are provided with a 10 m pixel spacing. Another factor deteriorating the spatial detail is the speckle noise of the radar which makes it hard to predict small objects. Combining these two factors results in a blurrier NDVI of the prediction compared to the optical retrieved NDVI and a loss of fine spatial detail.

Another problem in the predicted NDVI maps are wetlands and water surfaces. In these areas the predicted NDVI is imprecise (cf. Figure 5.6). One possible reason for this might be the low amount of water pixels in the dataset with 4.9 % (cf. Table 5.3). The interaction of radar waves with water might also be another cause. Firstly, radar waves do not penetrate the water but are reflected at the water surface. Secondly, the signal energy is reflected away from the sensor so that only very little energy is received back at the sensor and the weak received signal does not contain any information about what is below the water surface. In contrast, optical light penetrates into the water to some extent and is also reflected back to the sensor by the plants, algae, bacteria, and other particles on the water surface and in the water. Therefore, optical sensors can be used to retrieve meaningful information about the water composition, and the amount of chlorophyll containing matter. This drawback of our approach does not impose many restrictions on monitoring vegetation on land which are mainly agricultural or forested areas. To avoid an erroneous retrieval of NDVI values all water bodies can be masked using existing water body layers such as the *MODIS/Terra Land Water Mask* (Carroll et al. 2017). The only problem could be paddy fields, which are similar to wetland areas, as rice is an important crop grown on paddy fields. However, other studies showed the feasibility to monitor paddy rice fields using SAR data (Inoue et al. 2014). Therefore, our model and dataset could be adapted and improved for this special case to allow the retrieval of NDVI values from these areas as well.

5.4.2 Influence of Auxiliary Model Inputs

The choice of the model inputs changes how accurate the NDVI can be derived as shown in Tables 5.4 to 5.6. The two polarizations of the radar backscatter are the most important input data as they achieve the highest performance when used as sole modality. This good performance can be increased when the land cover map is also fed to the model as an auxiliary input. Using the two SAR polarizations separately shows a higher relative importance of the VH polarization compared to the VV polarization. The performance of using solely $\sigma^{\circ}_{\mathrm{VV}}$ data is on par with using only the WorldCover as input, however both achieve only a substandard performance. The low performance when using the land cover information without the radar backscatter shows that it is only supporting the model, but not being the main information source of the model.

The improved model performance when using the ESA WorldCover can have multiple reasons. One reason could be the influence of prior knowledge of the underlying land cover. Different relationships between SAR backscatter and NDVI exist for different land cover types, as it is an important feature for NDVI estimation (Santos et al. 2022). For example, high intensities in urban environments will certainly propagate into lower NDVI values than high intensities in croplands. If the model does not have to learn the distinction between the different land cover classes itself but receives prior knowledge about them, this leaves more model capacity for the actual implicit regression models. Another reason could be the high resolution and sharp edges included in the land cover map, which helps to add spatial detail, at least at the edges of different land covers.

Adding the DSM as model input does not improve model performance. The radar backscatter is influenced by the elevation and the terrain due to the side-looking nature of the sensor. Two identical areas with the same vegetation and land cover will have a different backscatter if the terrain is different (e.g. slope and aspect) (Small 2011). This weakens the relationship between radar backscatter and NDVI values and should reduce the model performance. However, a similar performance is achieved independent of whether these terrain effects are corrected for or not, and adding the DSM as model input does not have an effect on model performance either. The reason for this might be that the model is able to learn the influence of the terrain on the backscattered signal and correct it. This suggests that terrain flattening as preprocessing step or including the DSM as model input might be unnecessary when deep neural networks are used. Additionally, the elevation correlates with the NDVI only very little, as shown in Table 5.6 by the poor performance when using only the DSM as model input.

5.4.3 Creation of Dense NDVI Time Series

As illustrated in Figure 5.8, the model prediction appears to be a valuable tool for addressing gaps in NDVI time series data due to cloud coverage. The SAR data of a single data is used to estimate the NDVI image. This utilized only the spatial context of the data, but neglects any information contained in temporal patters. This approach employs only the spatial context of the data, thereby neglecting any information that may be derived from temporal context. An alternative approach to obtaining a dense NDVI time series is to first extract a time series of SAR values and then estimate the NDVI time series, as has been demonstrated by Zhao et al. (2020). A comparison of these two approaches reveals that the first is more flexible, as it requires only a single SAR image, rather than a complete time series. Consequently, the presented method facilitates real-time applications, enabling the prediction of NDVI for the most recent image.

6 Fusion of SAR-Derived and Optical NDVI Time Series

What You'll Learn This chapter details how sparse optical NDVI time series can be fused with denser SAR-estimated NDVI time series. For this, a recurrent neural network is trained with a dataset specifically created for this task.
Based on This chapter's material originates from the article "Dense NDVI Time Series by Fusion of Optical and SAR-Derived Data" and published in the *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (Roßberg et al. 2024b).

As shown in the previous Chapter 5, NDVI imagery can be derived from SAR images. To simplify the model and its training, the approach only regarded the spatial context and used a single image, but neglected the rich information contained in the temporal context, which limits performance and effectiveness. Additionally, available cloud free optical imagery is neglected. To address these shortcomings and to create dense, accurate NDVI time series, this chapter presents an approach to fuse the sparse optical NDVI time series with the dense, but less accurate SAR-derived NDVI time series.

6.1 Creation of the Time Series Dataset

The proposed time series fusion approach integrates optical and SAR-derived NDVI data, necessitating the creation of a comprehensive dataset for training and evaluation. Data were selected from the Sentinel-1 SAR and Sentinel-2 optical sensors, motivated by their global coverage, high revisit frequency (at the equator 5 days for Sentinel-2 and 6 days for Sentinel-1), and free data access. The regions of interest (ROIs) are derived from the 1206 globally distributed ROIs of the SEN12TP dataset (cf. Section 5.1). These ROIs are chosen for their balanced representation in terms of land cover, climate, and global distribution and their locations are displayed in Figure 5.2.

The SEN12TP dataset with its ROIs each measuring 20 km × 20 km, amounts to a size of 222 GB. Extending this to include images at many different dates from each region would significantly increase the data volume. To limit the required storage, only the

central area of each ROI was utilized. For these areas, all available imagery between 1 September 2019 and 30 March 2021 is downloaded to form the image time series. The ROIs have a size of 2.56 km × 2.56 km corresponding to 256 px × 256 px at 10 m resolution.

Overall, 157 050 SAR and 125 860 optical images were downloaded, encompassing 252 GB. Subsequent sections detail the data and necessary preprocessing steps for both optical NDVI and SAR-derived NDVI values. For preprocessing and image download Google Earth Engine (GEE) (Gorelick et al. 2017) was used.

6.1.1 Optical Data and its Processing

As the source of optical data and to calculate the NDVI, data of the Sentinel-2 multispectral sensors is used, which is masked for clouds and cloud shadows and coreferenced onto each other. Thereby, atmospherically corrected Level-2A of the GEE collection `COPERNICUS/S2_SR_HARMONIZED` is used. To mask all clouds, the cloud probabilities contained in the GEE collection `COPERNICUS/S2_CLOUD_PROBABILITY` created using the Sentinel Hub's cloud detector (Zupanc 2017) are used with a threshold set to 40 %. To clean the borders of the detected clouds, the cloud probabilities are first convolved with a circle-shaped filter with a radius of 40 m, thresholded and dilated again with 20 m large circular filter kernel. These parameters are taken from the s2cloudless example script (Aleksandrov et al. 2023).

Cloud shadows were masked using a geometrical method by projecting the previously calculated cloud masks onto the ground (Schmitt et al. 2019a). The alternative approach, which is not employed in this study, is the spectral method that utilizes the different multispectral bands for shadow detection. Potential cloud shadows are all cloud mask pixels projected along the sun's azimuth angle for 2 km. They are intersected with dark areas, which are all areas with a reflectance sum of the bands `B8`, `B11`, and `B12` smaller than 0.3 which are not water according to the scene classification layer (SCL) band. Again, to clean up the mask, a morphological opening is applied with an erosion (40 m circle) followed by a dilation (100 m circle) (Schmitt et al. 2019a). The bands `B2`, `B3`, `B4`, `B8`, `B11`, and `B12` together with the cloud and cloud shadow masks are downloaded for all optical images which are at least 10 % cloud and cloud shadow-free.

Upon download of the images, the need for coregistering each optical image stack became apparent, as the geolocalization of the Sentinel-2 imagery is not pixel perfect, something also noted by Requena-Mesa et al. (2021). We use a method similar to the one in *eolearn* Python package (Peressutti 2023): the template image upon which each image is registered on is the gradient of the temporal mean of all images after cloud masking and conversion to grayscale. To find the best matching position, a translation-only motion model with the enhanced correlation coefficient is used.

Finally the NDVI is calculated using Sentinel-2's red and infrared bands `B8` and `B4` according to Equation 2.6.

6.1.2 SAR Data

The SAR data as the second modality of the approach is acquired by the Sentinel-1 sensors and sourced from the GEE collection `COPERNICUS/S1_GRD`. From this collection, data acquired in interferometric wide swath (IW) mode with two polarizations VV and VH is used. For this ground range detected (GRD) data, geometric terrain correction is applied and the data is transformed to a logarithmic scale. This yields sigma naught backscatter values σ°, expressed in decibels (dB), to be used in the method.

6.1.3 SAR-Estimated NDVI Images

To estimate NDVI values from SAR data, we use the model described in the previous Chapter 5. To create the SAR-estimated NDVI images, an ensemble of five models is used. Taking the mean from the outputs of five models ensures superior model performance and reliability. In the end, an SAR-estimated NDVI image is created for each location and SAR acquisition.

6.1.4 Time Series Extraction and Final Dataset

Time series are extracted from the optical, SAR, and SAR-derived NDVI images. Only every fifth column and every fifth row is used forming a regular 5 px grid. This reduces the data volume to a manageable level and additionally reduces unneeded redundancy due to the high similarity between the time series of neighboring pixels. Next to the image data, the acquisition dates are extracted for each time step. Overall 330 GB of image data is used (consisting of 72 GB optical, 181 GB SAR, and 78 GB SAR-estimated NDVI imagery) resulting in 124 GB of extracted time series data.

The whole process of the creation of the dataset is depicted in Figure 6.1.

Two splits of the dataset are created to assess spatial and temporal generalization performance. The spatial split dataset contains the same train, validation, and test scenes as those found in the SEN12TP dataset, while the temporal split dataset uses the first twelve months of all scenes for training and the remaining six months of all scenes for testing. This allows performance evaluation on unseen data from different scenes and future dates.

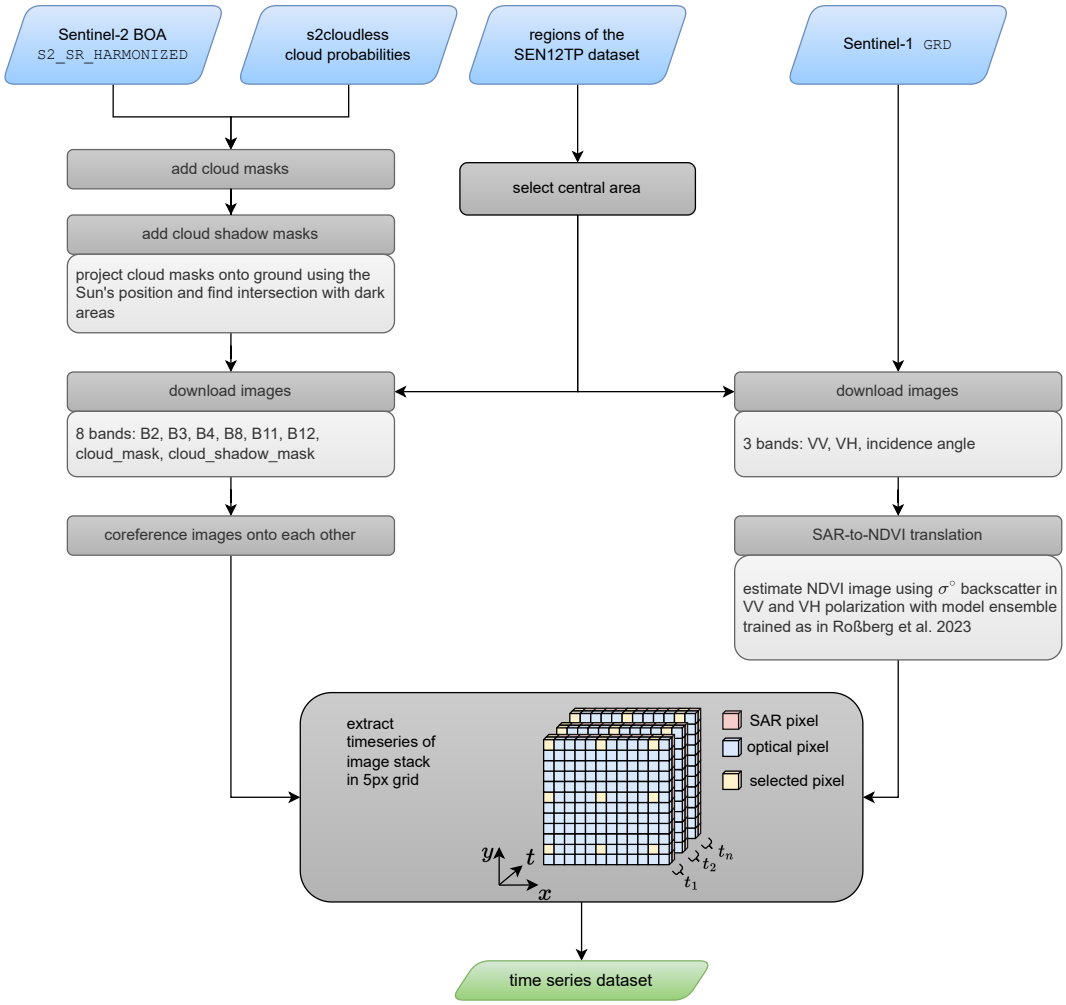


Figure 6.1: Overview of the creation of the time series used for training and testing with the required preprocessing steps. Sentinel-1 and -2 image series are processed and retrieved from Google Earth Engine (GEE) for the regions of the SEN12TP dataset. From the image stacks, pixel-wise time series are retrieved in a regular 5 px grid to form the dataset.

6.2 RNN-Based Time Series Fusion

To fuse the optical and SAR-derived NDVI time series a gated recurrent unit (GRU) is employed (Cho et al. 2014), a common recurrent neural network (RNN) variant. RNNs were selected because of their suitability for handling variable sequence lengths and accommodating missing data (Weerakody et al. 2021). Utilizing a many-to-many configuration (see Figure 2.19) the chosen GRU model generates an output prediction for each step of the input sequence.

The model input is first transformed using a single fully connected layer with 128 neurons and ReLU activation with shared weights for each time step. Then, the GRU consisting of three bidirectional layers, each with a hidden size of 256 and a dropout layer with a dropout rate of 0.3 fuses the time series while learning temporal patterns. Finally, the GRU outputs are fed to two consecutive linear layers, again with shared weights for each time step. The first linear layer has 128 neurons and a ReLU activation function, while the second layer, which serves as the final output, uses a sigmoid activation function. The model is depicted in Figure 6.2. The hidden state of the GRU is initialized for each batch with random weights taken from a normal distribution. The model is trained using the Adam optimizer (Kingma et al. 2014) with a learning rate of 5×10^{-4}, a batch size of 128, and the mean squared error (MSE) loss.

The extracted time series (cf. Section 6.1) undergo further processing before being passed to the RNN model. The optical NDVI time series serve as both the label and the model input. For each time series, a subset of values is randomly selected as the label, while the remaining values are used as input. We use 66 % = $^2/_3$ of the values as labels and the remaining ones as model inputs to facilitate model learning. This simulates frequent cloud cover and showed a good performance in preliminary tests compared to using only 33 % or 50 % of the values as labels. For each training epoch, the selection as label or model input is different to avoid overfitting and to make the model more robust.

From the 18-month long time series sequences with a length between 1 and 6 months were extracted to train a versatile model which can be used for a variety of use cases. The extracted sequence length is a balance between data demands and practicality: Extremely short sequences are avoided as they lack sufficient usable data, while excessively long sequences demand impractical data resources impeding easy applicability when deploying the model.

From these sequences, all steps that contain no values for all input features or the label are removed. This leads to shorter sequence lengths, which improved fusion performance.

All model inputs as well as the optical NDVI labels are normalized. Model inputs undergo Z-standardization using the mean and standard deviation from the training

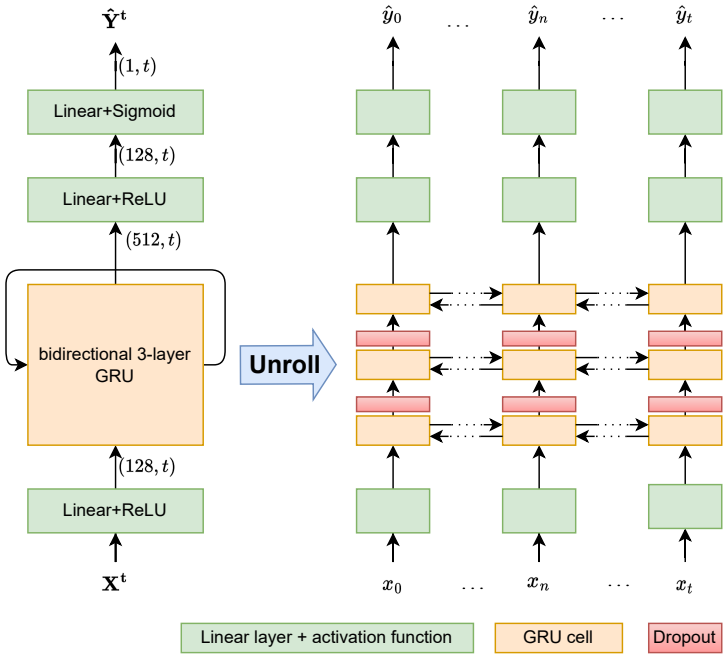


Figure 6.2: Overview of the RNN model architecture. To the right is the unrolled version. $\mathbf{X}^t$ denotes the input time series, $\hat{\mathbf{Y}^t}$ the predicted or rather the fused time series. The numbers denote the feature dimensions with the variable sequence length t.

set. The labels are clipped to the range $[-1, 1]$ before they undergo min-max normalization to ensure that their values fall within the interval $[0.2, 0.8]$, which corresponds to the linear region of the sigmoid activation function. Both Z-standardization and Min-Max normalization are explained in Section 2.5.3. The day of the year (DOY) is not put directly into the model, but sin(DOY) and cos(DOY) are used to avoid sudden jumps at the end of the year and to capture cyclic patterns of the time series more efficiently.

6.3 Results

To evaluate the fusion approach, a numerical assessment is conducted on the globally distributed test scenes (Section 6.3.1). Subsequently, the resulting fused time series are presented in detail for two example areas (Section 6.3.2).

6.3.1 Quantitative Results

To show the good performance of our approach we apply the trained fusion model on the test set data and calculate the error between optical NDVI and the fused NDVI. The 124 globally distributed test regions of the SEN12TP dataset are used as depicted in Figure 5.2). As a baseline to our model, we use linear interpolation of the optical data as well as a histogram-based gradient boosting regression tree (hGBRT), a machine learning model able to handle missing data.

The fusion approach *NDVI-Fusion* achieves a very low error, with an mean absolute error (MAE) of 0.0478. Using the fusion model only with the optical NDVI time series or the optical NDVI together with the σ° SAR backscatter results in a slightly higher MAE of 0.0513 and 0.0517, respectively. For the two baselines, a higher MAE of 0.0675 of the hGBRT model, and a slightly higher MAE of 0.0482 using linear interpolation is achieved. The relative order of the MAE of the compared models is the same for the RMSE and R^2 score as shown in Table 6.1.

Table 6.1: Comparison of the performance of the different models using the spatially split test set data.

Method	MAE	RMSE	R^2
linear interpolation	0.0482	0.0843	0.9190
hGBRT baseline	0.0675	0.1081	0.8680
Optical-Only	0.0513	0.0882	0.9115
NDVI-Backscatter	0.0517	0.0874	0.9130
NDVI-Fusion	0.0478	0.0806	0.9261

To demonstrate that the fusion approach can generalize from one year to another and therefore, be used in a real-world scenario, the model is trained using the temporally split dataset. For this, model training is conducted using data from all areas of the first year and before being tested using data from the remaining 6 months. This ensures that all testing data follows the training data temporally, thereby simulating a real-world scenario where the model, trained on current data, is subsequently applied to future data. Compared to the model trained with spatially split data we see a slight performance decrease, the MAE increases from 0.0478 to 0.0498 as illustrated in Table 6.2.

The distribution of gap lengths in the dataset is unbalanced: most gaps are rather short with 70 % of the gaps being shorter than six days as shown in Figure 6.3. The peaks for gap lengths with a duration as multiple of five days are due to the five-day revisit of the Sentinel-2 constellation (ESA 2015). Because of this imbalanced distribution and the importance of filling especially longer gaps, the predicted error is determined in conjunction with the length of the optical gap or rather how near the

Table 6.2: Performance of NDVI-Fusion model on temporally split data. For comparison, the performance using the spatial split from Table 6.1 is reported.

data split	MAE	RMSE	R^2
spatial	0.0478	0.0806	0.9261
temporal	0.0498	0.0855	0.9017

next optical NDVI value is. For this analysis, the gap length for each optical NDVI used as a label is calculated by determining the minimum distance to the nearest optical NDVI used as a model input, either preceding or following it. The hGBRT model performance is not included here, because the aggregation step from daily to weekly NDVI values only allows rather broad gap length classes: the weekly values can have a distance of 1 to 6 days gaps between them and having no value for a week could signify an NDVI data gap from 7 up to 20 days.

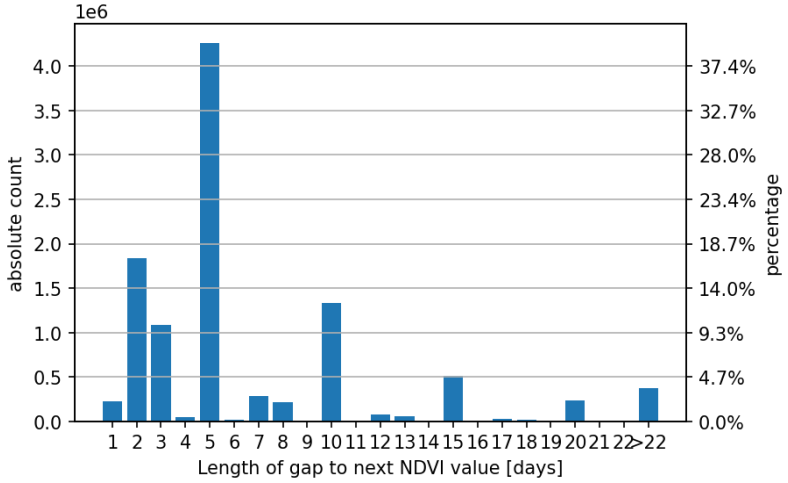


Figure 6.3: Distribution of the gap lengths of the optical NDVI time series in the test set used for performance calculation. Gaps with a length of 5, 10, 15, and 20 days are more common because of the 5-day revisit of the Sentinel-2 constellation.

The evaluation shows, that the error of all models increases with an increasing gap length. The increase however is different for the different models: linear interpolation is becoming significantly worse for long gaps ($\geq$ 20 days long), whereas the fusion approach only shows a mild increase as listed in Table 6.3.

Table 6.3: Comparison of the model's performances for differently long gaps in the optical data for the spatially split test set data. The performance of the NDVI-Fusion model is also reported for the temporally split test data.

Metric	Method	gap length [days] < 5	5–9	10–14	15–19	≥ 20
MAE	Interpolation	0.035	0.040	0.047	0.056	0.088
	Optical-Only	0.034	0.040	0.048	0.058	0.091
	NDVI-Fusion	0.034	0.040	0.046	0.054	0.076
	temporal split	0.035	0.042	0.050	0.060	0.082
RMSE	Interpolation	0.063	0.070	0.080	0.091	0.137
	Optical-Only	0.059	0.067	0.080	0.092	0.139
	NDVI-Fusion	0.059	0.066	0.077	0.087	0.116
	temporal split	0.062	0.072	0.084	0.097	0.126
R^2	Interpolation	0.949	0.941	0.924	0.906	0.809
	Optical-Only	0.955	0.945	0.925	0.904	0.798
	NDVI-Fusion	0.955	0.946	0.930	0.914	0.860
	temporal split	0.919	0.919	0.900	0.876	0.823

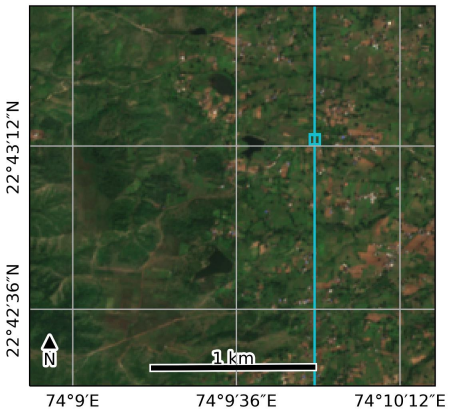
6.3.2 Qualitative Results

To demonstrate the fusion approach visually and present qualitative results, two example areas of the 124 areas of the test set are selected. They are chosen to demonstrate the lower and upper performance bounds: one area located in the south of Vietnam has one of the worst test performances and highest errors, whereas the other area has a very low error and is located in India. In the subsequent text, these are referred to as the *Vietnam* and *India* example areas, respectively. The location of both areas, along with an RGB image, is given in Figure 6.4.

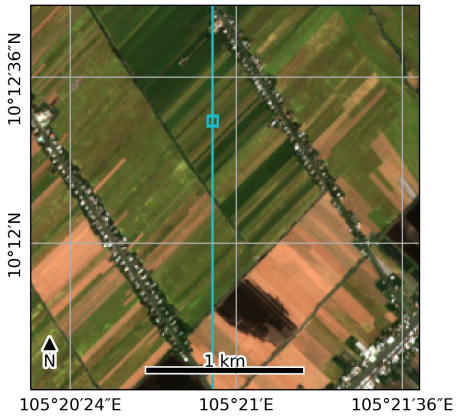
All plots are created by masking all optical NDVI values in a sliding window approach: going over each NDVI value of the time series, the NDVI is masked, the fusion model is run, and the prediction of this date is saved. This ensures that for each predicted value the optical NDVI of that date was not seen by the model, but all the others are.

Analyzing the time series of a single pixel, whose location is displayed as cyan square in Figure 6.4, shows that the SAR-derived NDVI values have a similar behavior, but by applying the fusion approach and augmenting the SAR-derived with the optical NDVI values results in a very high agreement between optical NDVI values of a date and the prediction at that day. The complex and quickly-changing NDVI patterns of

(a) Location of the two example areas.



(b) *India*, 2019-10-12



(c) *Vietnam*, 2020-02-25

Figure 6.4: Location of two example areas used for demonstrating our time series fusion approach marked red on a world map (top) and Sentinel-2 true-color images of these two locations (bottom). The vertical line and the box in cyan denote which data is used to display exemplary time series.

the *Vietnam* example are modeled very well with low errors, even though the optical NDVI values are often not dense and have long gaps. The *India* example area has very dense optical values most of the years, only around August to October does the NDVI increase and becomes rather sparse. This is very well reflected in the fused NDVI. Both examples are shown in Figure 6.5.

To visualize the performance not only on a single pixel, but on multiple pixels at once, the NDVI of a whole image row can also be plotted (denoted as a cyan vertical line in Figure 6.4). The optical data has many gaps due to clouds and cloud shadows, which also result in gaps in the calculated error. In contrast, the fused NDVI is dense

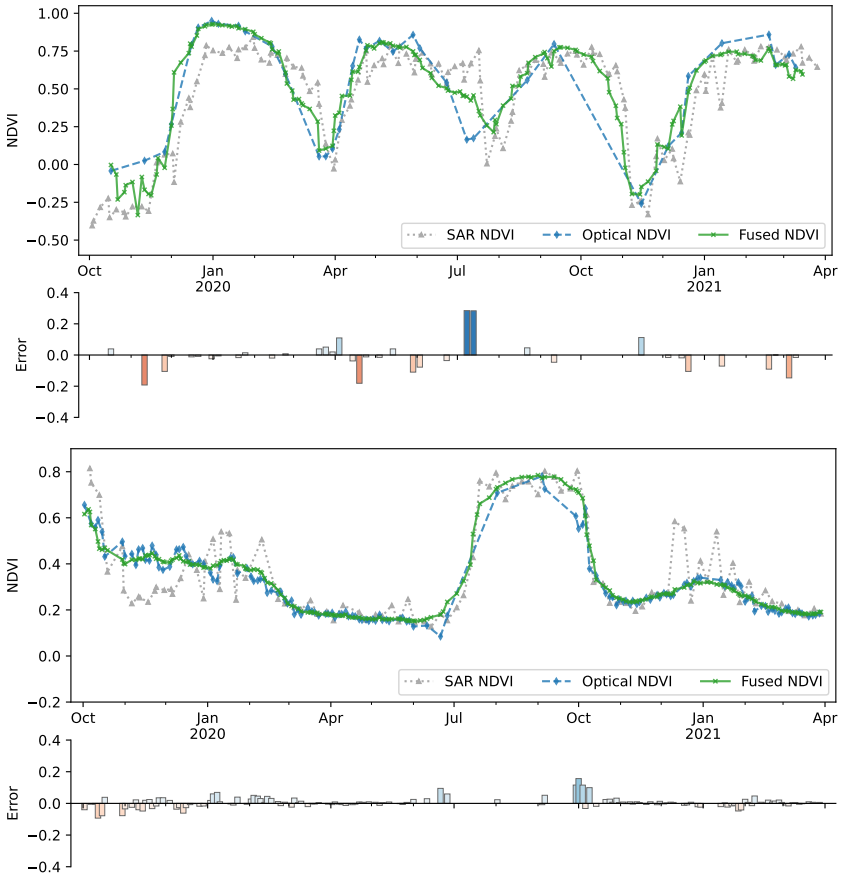
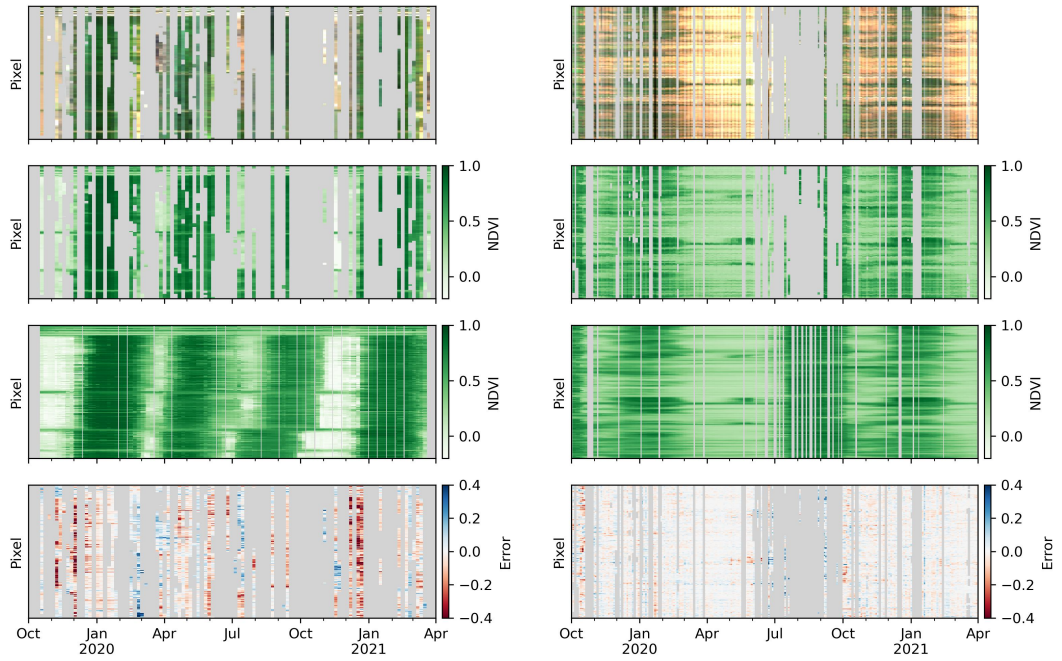


Figure 6.5: Comparison of the NDVI time series and associated errors for optical, SAR-derived, and fused NDVI data for a cropland pixel within the *Vietnam* (top) and *India* (bottom) example areas. The location of the used pixels is given in Figure 6.4. The fused NDVI time series (green) is dense and is closely aligned with the optical NDVI measurements, illustrating the effectiveness of the fusion approach. In contrast, the optical NDVI time series (blue) suffers from irregular acquisition intervals due to cloud cover. Meanwhile, the SAR-estimated NDVI (gray) is consistent in temporal coverage but displays a higher level of uncertainty and noise.

without long gaps and results in a low error for almost all of the dates. For the *India* example area the NDVI is sufficiently dense for most of the year but between June and October, almost no cloud-free pixels could be acquired. Our fusion approach can estimate reasonable NDVI values and achieves a low error for the few cloud-free optical dates. Only for the *Vietnam* example area a few dates in December have a higher error which coincides with a drastic increase of the NDVI at these dates. Visualizations for both areas are depicted in Figure 6.6.



(a) *Vietnam* example area (b) *India* example area

Figure 6.6: Comparison of the NDVI of one image row over time for the example areas. The RGB (top row) and the NDVI derived from optical data (second row) have many gaps due to cloud coverage. In contrast, the fused NDVI (third row) is almost gap-free and closely aligned with the optical NDVI as evidenced by the low error (bottom row). The location of the row is shown in Fig 6.4. Missing data due to missing image retrievals or masked clouds and cloud shadows is displayed in gray.

Lastly, the fusion approach is compared for a series of images. For 15 partly cloud-free images of the *Vietnam* example area between December 2019 and February 2020, the fused NDVI is calculated and compared to the optical one. For almost all scenes a low error is achieved while filling all the cloud-induced gaps. This is displayed in Figure 6.7.

All these figures show the high performance of the approach and the high accordance with the optical NDVI.

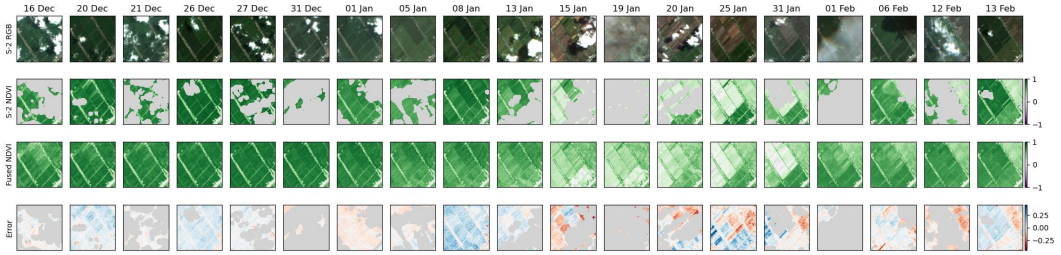


Figure 6.7: Overview over the RGB and NDVI images between December 2019 and February 2020 for the *Vietnam* example area. The top row shows the optical images, and the second row the NDVI images masked for clouds and cloud shadows. The fused images (third row) are only shown for all dates, where S-2 data was captured. The bottom row shows the error between optical and fused NDVI images. Gray areas are masked because of detected clouds or cloud shadows.

6.4 Discussion

An approach to fuse time series of different modalities was presented. This approach integrates sparse but accurate optical NDVI time series with denser, albeit less precise, SAR-estimated NDVI time series, effectively addressing gaps caused by clouds and cloud shadows. Both quantitatively as well as qualitatively the high performance and low error of the fused NDVI time series could be shown. Gaps can be filled and the resulting time series have a high similarity with the optical values.

The findings reveal that pre-fusion NDVI estimates from individual SAR scenes are preferable to the direct use of SAR backscatter (cf. Table 6.1). This efficacy stems primarily from two factors: First, the SAR-to-NDVI translation leverages the spatial neighborhood of the SAR data, enhancing NDVI estimation accuracy by considering each pixel's spatial context. Secondly, SAR-derived NDVI values are less impacted by

speckle noise, a prevalent issue in SAR backscatter time series, resulting in a cleaner dataset for fusion.

As shown numerically in Table 6.3, a low error is achieved for short as well as for long gaps in the optical data. Nonetheless, there are instances showing limitations of the approach. Firstly, our model struggles to accurately predict abrupt changes in the time series due to agricultural practices or vegetation burns. This can be evidenced in the *India* example area, where higher errors were observed during rapid NDVI fluctuations in July and October 2020, as depicted in Figure 6.5. Similarly, for the *Vietnam* example area, on multiple occasions a sudden change results in an increased error, for example in April or July 2020.

These inaccuracies can be attributed to two main reasons: Firstly, the necessity for high temporal precision, where even slight timing discrepancies can lead to significant errors, and secondly, a lack of data with rapid vegetation transitions. As most of the vegetation is rather stable, only very little optical data is available with sudden changes in the NDVI. To mitigate this, we could expand our dataset to include more instances with high and quick vegetation and NDVI changes. However, it might be computationally expensive to acquire a sufficient amount of suitable time series, as there has to be a rapid vegetation change and coinciding available optical data. Another option could be to oversample time series with frequent changes. This however would likely result in an oversampling of agricultural areas and seasonal vegetation, resulting in an higher imbalance of the training data.

Another case in which the method has a lower accuracy is in flooded areas or for water bodies. This is due to the interaction of microwave signals with water surfaces. The reflection of microwaves off water results in minimal backscatter signals and consequently, a lack of usable information for NDVI translation in the received signal. Without information about the Earth's surface in the SAR data, the translation to NDVI values is not accurately possible, which results in an increased error of the SAR-derived NDVI time series and therefore also an increased error in the fused NDVI series in such environments. This phenomenon likely explains the elevated errors observed on 25 January in Figure 6.7, as the rice fields in the studied area undergo regular flooding (Nguyen et al. 2022).

It is also acknowledged that the optical NDVI, the standard of truth, has some inherent uncertainty and inaccuracy. These uncertainty arises from factors such as atmospheric correction methodologies and algorithms as well as differences in the topographic, solar, and viewing angles. No corrections for differing sun–surface–sensor angles were applied, to reduce the complexity in preprocessing and remain close to the ESA-provided data. Moreover, the cloud and shadow masking, while effective, is not infallible, with occurrences of both false positives and false negatives. A manual annotation of clouds is infeasible, due to the sheer volume of images. However, employing a more sophisticated temporal cloud detection algorithm like in Zhu et al. (2014) could improve cloud detection accuracy, albeit at the cost of increased computational demand and more complex data processing.

7 Discussion

A central objective of research is to engage in a critical discourse and analysis of one's own contributions. Accordingly, the following chapter presents a critical review of the dissertation's work.

This dissertation presented an approach to mitigate the impact of clouds on the availability and quality of NDVI data through the deployment of two deep learning models. The first model translated a SAR image into an NDVI image, while the second model fused dense time series of the SAR-estimated NDVI values with the sparse, cloud-affected optical time series.

General Validity As demonstrated in this dissertation, NDVI data can be successfully enhanced by leveraging SAR data and deep learning models. Successfully achieving this requires the synergy of both models – first translating single-scene SAR data to NDVI and then fusing the time series. While the initial SAR-to-NDVI translation provides a low error and a good amount of spatial detail considering the low resolution of Sentinel-1 data, the predicted NDVI values are not fully sufficient for all downstream applications. The needed high quality is achieved by the subsequent time series fusion, integrating the sparser optical data. This enables the prediction performance to reach a level suitable for practical applications.

Computational Efficiency The use of two deep learning models makes computational efficiency an important consideration. Although the first model, a CNN, is highly efficient in utilizing available GPU resources, the second model, an RNN, is not as readily scalable. The RNN processes data sequentially, which inherently limits the extent to which it can be parallelized. This, in turn, constrains the utilization of GPUs. Consequently, the complete workflow, encompassing data download, SAR-to-NDVI model execution, and RNN-based time series fusion for each pixel, is both computationally demanding and time-consuming.

Benefit of Auxiliary Geospatial Data Incorporating auxiliary geospatial data, such as elevation information and land cover maps, can significantly influence a model's performance, either enhancing or diminishing it based on the specific remote sensing task and the timeliness of the data used. In the case of SAR-to-NDVI-translation, the integration of land cover data has been shown to improve model performance

(cf. Table 5.5). However, it was not investigated if the improved performance also translates into an improvement of downstream applications using the estimated NDVI data.

For certain downstream applications, leveraging the more accurate NDVI data produced by including the land cover information may paradoxically lead to suboptimal outcomes. An illustrative example is areas impacted by forest clearance and deploying the estimated NDVI for the detection of deforestation. When the SAR-to-NDVI model is given both backscatter and land cover data to estimate the NDVI, it may receive contradictory signals; the land cover data might still indicate the presence of forest, while backscatter data suggest that changes have occurred. Such inconsistencies may result in erroneous NDVI predictions. In contrast, restricting the model to backscatter data alone eliminates these contradictions, potentially enhancing the accuracy of the predicted NDVI and the subsequent deforestation detection for this particular scenario.

This example underscores a critical insight: that not all available data should be utilized indiscriminately in the hope of improving performance. Instead, the suitability of the data should be carefully considered in relation to the specific application.

Evaluation of Terrain-Flattened Backscatter as Model Input The SAR-to-NDVI translation performance was compared for different input data. One of these comparisons was whether the SAR data passed into the deep learning model should have terrain flattening applied or if it is sufficient to use only geocoded data with the backscatter still influenced by the terrain. The conclusion was, that this is not necessary, as the performance of using terrain flattened γ° backscatter is not significantly better than σ° backscatter without terrain flattening applied.

However, during the finalization of this thesis, two issues emerged. Firstly, no separate evaluation for flat and for hilly terrain was carried out to support the claim, that terrain flattening is not required. This differentiation is crucial because terrain flattening is not necessary for flat regions, but might provide a performance boost in hilly areas. Associated with that, the distribution of terrain slopes of the SEN12TP dataset was not examined to assess how balanced the dataset is in that regard.

Secondly, comparing terrain flattened with non-terrain flattened data poses challenges since pixels in the radar shadow are handled differently. Geocoding assigns backscatter values to pixels located within radar shadow regions; however, no actual backscatter from the surface can be sensed in these areas, and the assigned low backscatter values are only noise. In contrast, the applied terrain flattening method sets pixels in the radar shadow as no-data pixels. These no-data areas cannot be processed by the neural network. Therefore, it is necessary to either substitute the pixel's value with a predetermined value or exclude the entire patch containing no-data pixels from the training process. In the presented approach the second option was chosen and any image patches containing no-data pixels were discarded.

This results in slightly different patches seen by the model during training, which could affect the model's predictions. Consequently, the comparability between terrain-flattened data and data without that processing is reduced.

Prediction Uncertainty Due to Insufficient Data Uncertainty estimation provides valuable insight into machine learning models and can be used, for example, to assess the reliability of model outputs. Uncertainty is generally categorized into two types: aleatoric and epistemic. Aleatoric uncertainty arises from data-related issues, such as noise or not enough information in the input to enable an accurate prediction. In contrast, epistemic uncertainty relates to limitations of the modeling approach, such as the chosen model architecture and the available training data. Together, these uncertainties can provide insight into the reliability of the predictions and the model's confidence in its output.

In a recent study by Baumann, Roßberg, and Schmitt (2023), a U-Net architecture, called MIMO U-Net, was developed to efficiently estimate both aleatoric and epistemic uncertainties on a per-pixel basis. An example application on which the architecture was evaluated was SAR-to-NDVI translation. For this application, the aleatoric uncertainty was reported to be higher ($\sigma = 0.187$ with σ denoting the estimated standard deviation) than the epistemic uncertainty ($\sigma = 0.049$). The elevated aleatoric uncertainty suggests that data from a single Sentinel-1 scene does not contain enough information for a highly accurate NDVI estimation. Consequently, incorporating additional data as input could improve the accuracy of the translation. Factors contributing to the insufficient prediction and high uncertainty include issues such as speckle and the low resolution of the Sentinel-1 SAR data. In addition, it's important to recognize the inherent limitations of SAR-to-NDVI translation due to fundamental differences in the sensing principles of SAR and optical sensors.

8 Conclusions and Perspectives

8.1 Summary and Conclusion

Cloud coverage poses a significant problem for vegetation monitoring based on optical data and prevents continuous observation for instance of crops, meadows, or forests. This thesis has aimed to mitigate the impact of clouds on the quality and availability of optical NDVI data which enables continuous vegetation monitoring despite cloud cover. This has been achieved by leveraging SAR data, able to penetrate through clouds, and deep learning models, which can learn complex patterns in the data. Overall, three contributions were presented.

Firstly, a globally applicable method to translate SAR imagery into NDVI values is proposed. This method is based on a deep learning model and can accurately predict the NDVI image of a scene when given the SAR backscatter data of that scene. To train this model, a novel dataset, named SEN12TP, was created. This dataset consists of Sentinel-1 SAR and Sentinel-2 optical imagery, which is timely paired so that images of the same scene are acquired on the same day. To ensure optimal performance regardless of the scene, a diverse set of images was acquired, taking into account land cover types, climate zones, and months. This results in a dataset that captures all states of the Earth's surface and vegetation, allowing models trained with it to be applied universally.

Secondly, driven by the desire to create dense and accurate NDVI time series, an approach was developed to fuse the existing, albeit sparser, optical NDVI time series with time series of SAR-estimated NDVI values. This method is based on a recurrent neural network (RNN) architecture which is capable of handling missing data and variable sequence lengths. The RNN is trained using a dataset created based on the same locations as the SEN12TP dataset. Smaller image footprints than those of SEN12TP were used to maintain a manageable dataset size while including data of one and a half years for each location. The trained model demonstrated a consistently good accuracy, even for cases with extended data gaps in the time series. Additionally, it generalized well across different regions and seasons allowing for global applicability. Visual assessments confirm a high similarity to optical values, supporting the reliability of denser NDVI time series for downstream applications despite challenges posed by cloud cover.

Finally, a thorough analysis of the relationship between NDVI values and SAR backscatter data was conducted. This provides a foundation for estimating NDVI

values from SAR data. Unlike many existing studies, the investigation did not focus on a single SAR frequency or a single study area. Instead, multi-frequency SAR data (X-, C-, S-, and L-band) with multiple polarizations was employed and the analysis was carried out for three diverse study regions with different climate zones, crops, and agricultural practices. The relationship between NDVI and SAR backscatter has been demonstrated; however, no universal link has been identified. Rather, a multitude of factors, including crop type, acquisition geometry, SAR frequency, and SAR polarization, influence the relationship between the two variables.

8.2 Future Work

The work presented in this thesis has identified several promising avenues for future research.

Model Evaluation with Data of Other Sensors The use of different SAR sensors and frequencies can provide additional insight and potentially increase the robustness and generalizability of the models. The upcoming NASA-ISRO Synthetic Aperture Radar (NISAR) mission, which will provide free access to S- and L-band data, offers significant potential for enhancing vegetation monitoring applications. The NDVI estimation using S- or L-band images could be thoroughly evaluated with this sensor's data. In addition, the integration of SAR data from sensors using different frequencies could be tested. In particular, the combination of NISAR with Sentinel-1 C-band data could significantly improve the accuracy of NDVI estimates.

An increase in NDVI estimation accuracy might also be achieved by combining Sentinel-1 data with data from existing SAR sensors other than NISAR. However, this is costly, as data from most other SAR sensors are either not freely available or freely available only in limited quantities for research purposes. This limitation hinders the ability to train a large model from scratch.

Incorporation of Additional Input Data A potential area for future investigation is the integration of additional remote sensing data into the SAR-to-NDVI model, complementing the SAR image. This is recommended because providing a single SAR scene to the model leads to a considerable aleatoric uncertainty, thereby suggesting the necessity for additional data for an accurate forecast (Baumann et al. 2023).

One potential approach to achieve this is the combination of an outdated but cloud-free optical image with the current SAR image which may enhance structural details. However, it is essential to properly implement and test this approach, as a naive implementation of simply adding the image could potentially yield suboptimal results. In such a scenario, the model might be inclined to select the more straightforward

option of predicting the outdated NDVI of the old image. It may therefore be advisable to input only a panchromatic image or just the edges of the optical image, expressed as the gradient, as structural cues. Another challenge of this approach may be the availability of cloud-free data for heavily clouded regions like the tropics. Using a cloud-free mosaic might help to overcome this issue; however, typically mosaics cannot be generated without artifacts, which the model would have to learn to ignore.

Combined Spatio-Temporal Approach Having presented an approach in this thesis that is split up into a spatial (chapter 5) and a temporal model (chapter 6), it might be worthwhile to implement and evaluate a combined approach with a spatio-temporal model. This could be achieved with a transformer architecture as it allows for a flexible incorporation of both temporal and spatial information. This work could also be combined with the development of a satellite image series foundation model, at least for vegetation. Several optical and SAR sensors could be integrated and the temporal and spatial correlations could be fruitfully exploited. A drawback of such a generalist approach is the need for a very large dataset, comprising terabytes of satellite data and substantial computational resources. This limits not only who can train such a model, but also who can actually use it, since not all research institutions or potential end-users have access to GPUs, let alone high-performance ones with sufficient memory.

Estimation of the Prediction Uncertainty Baumann, Roßberg and Schmitt (2023) presented a novel model architecture to jointly estimate aleatoric and epistemic prediction uncertainties. An example application of this model was the SAR-to-NDVI translation, where a good performance was achieved. However, its evaluation focused on the uncertainty of the model and less on the estimated NDVI values, which would warrant a more elaborate analysis. Having a measure of the uncertainty of the predicted NDVI values is not necessary for most applications that use the resulting NDVI values, but it could still increase confidence in the predicted values. It might also be worthwhile to integrate the estimated uncertainties into the fusion of the time series (chapter 6).

Transfer to Additional Vegetation Indices This thesis focused on the NDVI as the most widely used vegetation index. However, other indices exist, such as the normalized difference moisture index (NDMI), which is an indicator of vegetation moisture. Since vegetation moisture also affects radar backscatter, it may be feasible to translate SAR imagery into moisture indices. Investigating this could provide a more comprehensive understanding of vegetation health, as the NDMI provides additional insights that the NDVI alone cannot capture. Future studies could investigate the relationship between SAR backscatter and the NDMI or other vegetation

indices, potentially leading to more robust and versatile vegetation monitoring models. This could also lead to a multi-task learning (MTL) model that predicts multiple vegetation indices simultaneously.

Inclusion of Weather Data An intriguing aspect would be to include weather data in the time series fusion. This addition would allow the model to learn the effects of weather on both plant growth and the sensed data. If successfully trained, the model would be able to disentangle the effects of vegetation growth, plant water content, and soil moisture on the received signal. This could further improve performance. In addition, the model could gain the ability to predict vegetation conditions under various future weather scenarios. This would work by providing only simulated weather data without providing optical or SAR data. However, training such a model poses significant challenges. It would require an extensive dataset and a large model capable of capturing the underlying complex patterns, necessitating significant computing power.

Exploring Downstream Applications Finally, this dissertation did not investigate the practical application of the estimated NDVI images and dense NDVI time series for subsequent uses. A variety of applications rely on NDVI data, including precision agriculture, drought monitoring, and climate impact assessment. Demonstrating an improvement in these applications using the enhanced NDVI data would not only validate the methods developed in this research but also demonstrate their practical utility. Future work could include the implementation of case studies and validation exercises in a variety of application domains to demonstrate the utility and effectiveness of the enhanced NDVI data.

Publications

The following publications were created during the period in which this dissertation was being prepared, and this thesis is based on them.

Peer-Reviewed Journal Articles

Roßberg T, Schmitt M (2023). "A Globally Applicable Method for NDVI Estimation from Sentinel-1 SAR Backscatter Using a Deep Neural Network and the SEN12TP Dataset". *Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 91 (3), pp. 171–188.

Roßberg T, Schmitt M (2024). "Dense NDVI Time Series by Fusion of Optical and SAR-Derived Data". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17, pp. 7748–7758.

Roßberg T, Schmitt M (2025). "Comparing the relationship between NDVI and SAR backscatter across different frequency bands in agricultural areas". *Remote Sensing of Environment* 319, Art. no. 114612.

Conference Papers

Roßberg T, Schmitt M (2022). "Estimating NDVI from Sentinel-1 SAR Data Using Deep Learning". *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 1412–1415.

Roßberg T, Schmitt M (2022). "Towards a Global Model for NDVI Estimation from Sentinel-1 SAR Backscatter". *Proceedings of 14th European Conference on Synthetic Aperture Radar*, pp. 245–248.

Roßberg T, Schmitt M (2023). "Temporal Upsampling of NDVI Time Series by RNN-Based Fusion of Sparse Optical and Dense SAR-Derived NDVI Data". *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 5990–5993.

Baumann A, **Roßberg T**, Schmitt M (2023). "Probabilistic MIMO U-Net: Efficient and Accurate Uncertainty Estimation for Pixel-Wise Regression". *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 4498–4506.

Roßberg T, Schmitt M (2024). "Comparing SAR at Multiple Frequencies with Optical Vegetation Indices for an Area in Vietnam". *Proceedings of 15th European Conference on Synthetic Aperture Radar*, 650–653.

List of Figures

List of Tables

Bibliography

Abdel-Hamid A, Dubovyk O, Graw V, and Greve K (2020). "Assessing the impact of drought stress on grasslands using multi-temporal SAR data of Sentinel-1: A case study in Eastern Cape, South Africa". *European Journal of Remote Sensing* 53.sup2, pp. 3–16.

Agrawal R, Mohite JD, Sawant SA, Pandit A, and Pappula S (2022). "Estimation of NDVI for cloudy pixels using machine learning". *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLIII-B3-2022, pp. 813–818.

Albertz J (2009). *Einführung in die Fernerkundung. Grundlagen der Interpretation von Luft- und Satellitenbildern*. 4th ed. WBG (Wissenschaftliche Buchgesellschaft).

Aleksandrov M, Zupanc A, Lubej M, jgersak, and Lukšič Ž (2023). *Sentinel Hub's Sentinel-2 Cloud Detector*. URL: `https://github.com/sentinel-hub/sentinel2-cloud-detector/blob/master/examples/sentinel2-cloud-detector-example.ipynb` (visited on 11/30/2023).

Alvarez-Mozos J, Villanueva J, Arias M, and Gonzalez-Audicana M (2021). "Correlation between NDVI and Sentinel-1 derived features for maize". *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 6773–6776.

Australian Bureau of Statistics (2024). *Australian Agriculture: Broadacre Crops*. URL: `https://www.abs.gov.au/statistics/industry/agriculture/australian-agriculture-broadacre-crops/2022-23` (visited on 06/27/2024).

Aybar C, Ysuhuaylas L, Loja J, Gonzales K, Herrera F, Bautista L, Yali R, Flores A, Diaz L, Cuenca N, Espinoza W, Prudencio F, Llactayo V, Montero D, Sudmanns M, Tiede D, Mateo-García G, and Gómez-Chova L (2022). "CloudSEN12, a global dataset for semantic understanding of cloud and cloud shadow in Sentinel-2". *Scientific Data* 9.1, Art. no. 782.

Balzter H, Cole B, Thiel C, and Schmullius C (2015). "Mapping CORINE land cover from Sentinel-1A SAR and SRTM digital elevation model data using random forests". *Remote Sensing* 7.11, pp. 14876–14898.

Baumann A, Roßberg T, and Schmitt M (2023). "Probabilistic MIMO U-Net: Efficient and accurate uncertainty estimation for pixel-wise regression". *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 4498–4506.

Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, and Wood EF (2018). "Present and future Köppen-Geiger climate classification maps at 1-km resolution". *Scientific Data* 5.1, Art. no. 180214.

Beck M, Pöppel K, Spanring M, Auer A, Prudnikova O, Kopp M, Klambauer G, Brandstetter J, and Hochreiter S (2024). "xLSTM: Extended long short-term memory". arXiv Preprint: `https://arxiv.org/abs/2405.04517`.

Belesova K, Gasparrini A, Sié A, Sauerborn R, and Wilkinson P (2017). "Household cereal crop harvest and children's nutritional status in rural Burkina Faso". *Environmental Health* 16, Art. no. 65.

Bourbigot M, Johnsen H, Piantanida R, and Hajduch G (2016). *Sentinel-1 Product Definition*. Doc. Nr: S1-RS-MDA-52-7440.

Buchhorn M, Smets B, Bertels L, Roo BD, Lesiv M, Tsendbazar NE, Li L, and Tarko A (2021). *Copernicus Global Land Service: Land Cover 100m: version 3 Globe 2015-2019: Product User Manual*. Version Dataset v3.0, doc issue 3.4.

Buckley SM, Agram PS, Belz JE, Crippen RE, Gurrola EM, Hensley S, Kobrick M, Lavalle M, Martin JM, Neumann M, Nguyen QD, Rosen PA, Shimada JG, Simard M, and Tung WW (2020). *NASADEM*. Tech. rep. NASA JPL.

Burger R, Aouizerats B, Besten N den, Guillevic P, Catarino F, Horst T van der, Jackson D, Koopmans R, Ridderikhoff M, Robson G, Zajdband A, and Jeu R de (2024). "The biomass proxy: Unlocking global agricultural monitoring through fusion of Sentinel-1 and Sentinel-2". *Remote Sensing* 16.5, Art. no. 835.

Capodici F, D'Urso G, and Maltese A (2013). "Investigating the relationship between X-band SAR data from COSMO-SkyMed satellite and NDVI for LAI detection". *Remote Sensing* 5.3, pp. 1389–1404.

Carroll M, DiMiceli C, Wooten M, Hubbard A, Sohlberg R, and Townshend J (2017). *MOD44W MODIS/Terra Land Water Mask Derived from MODIS and SRTM L3 Global 250m SIN Grid V006*. Version 6.

Castro M, Parra JC, Morales LJ, and Salas C (2014). "Establishment of empirical relations between fuel moisture content and the normalised difference vegetation index". *Journal of Soil Science and Plant Nutrition* 14.3, pp. 670–675.

Chandel NS, Tiwari PS, Singh KP, Jat D, Gaikwad BB, Tripathi H, and Golhani K (2019). "Yield prediction in wheat (Triticum aestivum L.) using spectral reflectance indices". *Current Science* 116.2, pp. 272–278.

Chandrasekar K, Sesha Sai MVR, Roy PS, and Dwevedi RS (2010). "Land surface water index (LSWI) response to rainfall and NDVI using the MODIS vegetation index product". *International Journal of Remote Sensing* 31.15, pp. 3987–4005.

Chen D, Hu H, Liao C, Ye J, Bao W, Mo J, Wu Y, Dong T, Fan H, and Pei J (2023a). "Crop NDVI time series construction by fusing Sentinel-1, Sentinel-2, and environmental data with an ensemble-based framework". *Computers and Electronics in Agriculture* 215, Art. no. 108388.

Chen J and Zhang Z (2023b). "An improved fusion of Landsat-7/8, Sentinel-2, and Sentinel-1 data for monitoring alfalfa: Implications for crop remote sensing". *International Journal of Applied Earth Observation and Geoinformation* 124, Art. no. 103533.

Chen J, Jönsson P, Tamura M, Gu Z, Matsushita B, and Eklundh L (2004). "A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter". *Remote Sensing of Environment* 91.3–4, pp. 332–344.

Chen N, Tsendbazar NE, Hamunyela E, Verbesselt J, and Herold M (2021). "Sub-annual tropical forest disturbance monitoring using harmonized Landsat and Sentinel-2 data". *International Journal of Applied Earth Observation and Geoinformation* 102, Art. no. 102386.

Cho K, Merriënboer B van, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, and Bengio Y (2014). "Learning phrase representations using RNN encoder–decoder for statistical machine translation". *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1724–1734.

Clauss K, Ottinger M, Leinenkugel P, and Kuenzer C (2018). "Estimating rice production in the Mekong Delta, Vietnam, utilizing time series of Sentinel-1 SAR data". *International Journal of Applied Earth Observation and Geoinformation* 73, pp. 574–585.

Claverie M, Ju J, Masek JG, Dungan JL, Vermote EF, Roger JC, Skakun SV, and Justice C (2018). "The harmonized Landsat and Sentinel-2 surface reflectance data set". *Remote Sensing of Environment* 219, pp. 145–161.

Dirección de Estimaciones Agrícolas (2023). *Datos Bioeconomía - Agricultura - Estimaciones agrícolas*. URL: `https://datos.magyp.gob.ar/dataset/estimaciones-agricolas` (visited on 07/01/2024).

Drusch M, Bello UD, Carlier S, Colin O, Fernandez V, Gascon F, Hoersch B, Isola C, Laberinti P, Martimort P, Meygret A, Spoto F, Sy O, Marchese F, and Bargellini P (2012). "Sentinel-2: ESA's optical high-resolution mission for GMES operational services". *Remote Sensing of Environment* 120, pp. 25–36.

Duchemin B, Hadria R, Erraki S, Boulet G, Maisongrande P, Chehbouni A, Escadafal R, Ezzahar J, Hoedjes J, Kharrou M, Khabba S, Mougenot B, Olioso A, Rodriguez JC, and Simonneaux V (2006). "Monitoring wheat phenology and irrigation in Central Morocco: On the use of relationships between evapotranspiration, crops coefficients, leaf area index and remotely-sensed vegetation indices". *Agricultural Water Management* 79.1, pp. 1–27.

Ebel P, Xu Y, Schmitt M, and Zhu XX (2022). "SEN12MS-CR-TS: A remote sensing data set for multi-modal multi-temporal cloud removal". *IEEE Transactions on Geoscience and Remote Sensing* 60, Art. no. 5222414.

Ettari A, Nappa A, Quartulli M, Azpiroz I, and Longo G (2024). "Adaption of diffusion models for remote sensing imagery". *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 8251–8254.

European Space Agency (2015). *Sentinel-2 User Handbook*.

Filgueiras R, Mantovani EC, Althoff D, Filho EIF, and Cunha FF da (2019). "Crop NDVI monitoring based on Sentinel 1". *Remote Sensing* 11.12, Art. no. 1441.

Flores A, Herndon K, Thapa R, and Cherrington E (2019). *Synthetic Aperture Radar (SAR) Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation*. NASA.

Flores-Anderson AI, Cardille J, Azad K, Cherrington E, Zhang Y, and Wilson S (2023). "Spatial and temporal availability of cloud-free optical observations in the tropics to monitor deforestation". *Scientific Data* 10.1, Art. no. 550.

Fontanelli G, Paloscia S, Zribi M, and Chahbi A (2013). "Sensitivity analysis of X-band SAR to wheat and barley leaf area index in the Merguellil Basin". *Remote Sensing Letters* 4.11, pp. 1107–1116.

Frison PL, Fruneau B, Kmiha S, Soudani K, Dufrêne E, Toan TL, Koleck T, Villard L, Mougin E, and Rudant JP (2018). "Potential of Sentinel-1 data for monitoring temperate mixed forest phenology". *Remote Sensing* 10.12, Art. no. 2049.

Gao BC (1996). "NDWI — a normalized difference water index for remote sensing of vegetation liquid water from space". *Remote Sensing of Environment* 58.3, pp. 257–266.

Gao F, Masek J, Schwaller M, and Hall F (2006). "On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance". *IEEE Transactions on Geoscience and Remote Sensing* 44.8, pp. 2207–2218.

Garioud A, Valero S, Giordano S, and Mallet C (2021). "Recurrent-based regression of Sentinel time series for continuous vegetation monitoring". *Remote Sensing of Environment* 263, Art. no. 112419.

Geara C, Gelas C, De Vitry L, Colin E, and Tupin F (2024). "InSAR2InSAR: A self-supervised method for InSAR parameters estimation". *Proceedings of 32nd European Signal Processing Conference*, pp. 651–655.

Gitelson AA and Merzlyak MN (1997). "Remote estimation of chlorophyll content in higher plant leaves". *International Journal of Remote Sensing* 18.12, pp. 2691–2697.

Goodfellow I, Bengio Y, and Courville A (2016). *Deep learning*. The MIT Press.

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, and Moore R (2017). "Google Earth Engine: Planetary-scale geospatial analysis for everyone". *Remote Sensing of Environment* 202, pp. 18–27.

Hansen MC, Defries RS, Townshend JRG, and Sohlberg R (2000). "Global land cover classification at 1 km spatial resolution using a classification tree approach". *International Journal of Remote Sensing* 21.6-7, pp. 1331–1364.

Hansen P and Schjoerring J (2003). "Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression". *Remote Sensing of Environment* 86.4, pp. 542–553.

Hanssen RF (2001). *Radar interferometry. Data interpretation and error analysis*. Vol. 2. Kluwer Academic Publishers.

Harfenmeister K, Spengler D, and Weltzien C (2019). "Analyzing temporal and spatial characteristics of crop parameters using Sentinel-1 backscatter data". *Remote Sensing* 11.13, Art. no. 1569.

He K, Zhang X, Ren S, and Sun J (2016). "Deep residual learning for image recognition". *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE.

Held A, Zhou ZS, Ticehurst C, Rosenqvist A, Parker A, and Brindle L (2019). "Advancing Australia's imaging radar capability under the Novasar-1 partnership". *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 8370–8373.

Hirooka Y, Homma K, Maki M, and Sekiguchi K (2015). "Applicability of synthetic aperture radar (SAR) to evaluate leaf area index (LAI) and its growth rate of rice in farmers' fields in Lao PDR". *Field Crops Research* 176, pp. 119–122.

Holtgrave AK, Röder N, Ackermann A, Erasmi S, and Kleinschmit B (2020). "Comparing Sentinel-1 and -2 data and indices for agricultural land use monitoring". *Remote Sensing* 12.18, Art. no. 2919.

Hosseini M and McNairn H (2017). "Using multi-polarization C- and L-band synthetic aperture radar to estimate biomass and soil moisture of wheat fields". *International Journal of Applied Earth Observation and Geoinformation* 58, pp. 50–64.

Hosseini M, McNairn H, Mitchell S, Robertson LD, Davidson A, and Homayouni S (2019). "Synthetic aperture radar and optical satellite data for estimating the biomass of corn". *International Journal of Applied Earth Observation and Geoinformation* 83, Art. no. 101933.

Hu X, Li L, Huang J, Zeng Y, Zhang S, Su Y, Hong Y, and Hong Z (2024). "Radar vegetation indices for monitoring surface vegetation: Developments, challenges, and trends". *Science of the Total Environment* 945, Art. no. 173974.

Huete A (1997). "A comparison of vegetation indices over a global set of TM images for EOS-MODIS". *Remote Sensing of Environment* 59.3, pp. 440–451.

Hunt REJ, Rock BN, and Nobel PS (1987). "Measurement of leaf relative water content by infrared reflectance". *Remote Sensing of Environment* 22.3, pp. 429–435.

Inoue Y, Kurosu T, Maeno H, Uratsuka S, Kozu T, Dabrowska-Zielinska K, and Qi J (2002). "Season-long daily measurements of multifrequency (Ka, Ku, X, C, and L) and full-polarization backscatter signatures over paddy rice field and their relationship with biological variables". *Remote Sensing of Environment* 81.2–3, pp. 194–204.

Inoue Y, Sakaiya E, and Wang C (2014). "Capability of C-band backscattering coefficients from high-resolution satellite SAR sensors to assess biophysical variables in paddy rice". *Remote Sensing of Environment* 140, pp. 257–266.

Jia M, Tong L, Zhang Y, Chen Y, and Chopra R (2014). "Multi-temporal radar backscattering measurements and modelling of rice fields using a multi-frequency (L, S, C, and X) scatterometer". *International Journal of Remote Sensing* 35.4, pp. 1253–1271.

Jiang Z, Li L, and Ustin SL (2009). "Estimation of canopy water content with MODIS spectral index". *Proceedings of SPIE*. Vol. 7454, Art. no. 74540V.

Jiao X, McNairn H, Hosseini M, and Homayouni S (2018). "Characterization of canola canopies using optical and SAR imagery". *Remote Sensing and Modeling of Ecosystems for Sustainability XV*, Art. no. 1076704.

Joshi A, Agrawal S, and Chauhan P (2022). "Geolocation accuracy improvement for NovaSAR-1 imagery acquired through TLE orbit". *Geocarto International* 37.26, pp. 12414–12435.

Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP, and Kessler M (2017). "Climatologies at high resolution for the earth's land surface areas". *Scientific Data* 4.1, Art. no. 170122.

Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP, and Kessler M (2021). *CHELSA*. Climatologies at high resolution for the earth's land surface areas. Version 2.1.

Khabbazan S, Vermunt P, Steele-Dunne S, Arntz LR, Marinetti C, Valk D van der, Iannini L, Molijn R, Westerdijk K, and Sande C van der (2019). "Crop monitoring using Sentinel-1 data: A case study from the Netherlands". *Remote Sensing* 11.16, Art. no. 1887.

Kikuchi K and Wang B (2008). "Diurnal precipitation regimes in the global tropics". *Journal of Climate* 21.11, pp. 2680–2696.

Kim Y and Zyl J van (2000). "On the relationship between polarimetric parameters". *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*. Vol. 3. IEEE, pp. 1298–1300.

King MD, Platnick S, Menzel WP, Ackerman SA, and Hubanks PA (2013). "Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua satellites". *IEEE Transactions on Geoscience and Remote Sensing* 51.7, pp. 3826–3852.

Kingma DP and Ba J (2014). "Adam: A method for stochastic optimization". arXiv Preprint: `http://arxiv.org/pdf/1412.6980v9`.

Klausing H and Holpp W (2000). *Radar mit realer und synthetischer Apertur. Konzeption und Realisierung*. Oldenbourg.

Kovacs JM, Jiao X, Flores-de-Santiago F, Zhang C, and Flores-Verdugo F (2013). "Assessing relationships between Radarsat-2 C-band and structural parameters of a degraded mangrove forest". *International Journal of Remote Sensing* 34.20, pp. 7002–7019.

Lasko K (2022). "Gap filling cloudy Sentinel-2 NDVI and NDWI pixels with multi-frequency denoised C-Band and L-Band synthetic aperture radar (SAR), texture, and shallow learning techniques". *Remote Sensing* 14.17, Art. no. 4221.

LeCun Y, Bottou L, Bengio Y, and Haffner P (1998). "Gradient-based learning applied to document recognition". *Proceedings of the IEEE* 86.11, pp. 2278–2324.

Lee JS and Pottier E (2017). *Polarimetric Radar Imaging. From Basics to Applications*. Taylor & Francis Group.

Li S, Xu L, Jing Y, Yin H, Li X, and Guan X (2021). "High-quality vegetation index product generation: A review of NDVI time series reconstruction techniques". *International Journal of Applied Earth Observation and Geoinformation* 105, Art. no. 102640.

Li Y, Chen X, Chen Z, and Hu C (2022). "S-band spaceborne SAR interferometric coherence analysis: A study case in Peth, Australia". *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 5812–5815.

Liao C, Wang J, Shang J, Huang X, Liu J, and Huffman T (2017). "Sensitivity study of Radarsat-2 polarimetric SAR to crop height and fractional vegetation cover of corn and wheat". *International Journal of Remote Sensing* 39.5, pp. 1475–1490.

Liu Ca, Chen Zx, Shao Y, Chen Js, Hasi T, and Pan Hz (2019). "Research advances of SAR remote sensing for agriculture applications: A review". *Journal of Integrative Agriculture* 18.3, pp. 506–525.

Liu Y, Chen KS, Xu P, and Li ZL (2016). "Modeling and characteristics of microwave backscattering from rice canopy over growth stages". *IEEE Transactions on Geoscience and Remote Sensing* 54.11, pp. 6757–6770.

López García MJ and Caselles V (1991). "Mapping burns and natural reforestation using thematic mapper data". *Geocarto International* 6.1, pp. 31–37.

Lopez-Sanchez JM, Vicente-Guijalba F, Ballester-Berman JD, and Cloude SR (2014). "Polarimetric response of rice fields at C-band: Analysis and phenology retrieval". *IEEE Transactions on Geoscience and Remote Sensing* 52.5, pp. 2977–2993.

Macelloni G, Paloscia S, Pampaloni P, Marliani F, and Gai M (2001). "The relationship between the backscattering coefficient and the biomass of narrow and broad leaf crops". *IEEE Transactions on Geoscience and Remote Sensing* 39.4, pp. 873–884.

Mahdianpari M, Salehi B, Mohammadimanesh F, and Motagh M (2017). "Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSAR-X imagery". *ISPRS Journal of Photogrammetry and Remote Sensing* 130, pp. 13–31.

Main-Knorn M, Pflug B, Louis J, Debaecker V, Müller-Wilm U, and Gascon F (2017). "Sen2Cor for Sentinel-2". *Image and Signal Processing for Remote Sensing XXIII*. Vol. 10427, Art. no. 1042704.

Mao Y, Van Niel TG, and McVicar TR (2023). "Reconstructing cloud-contaminated NDVI images with SAR-Optical fusion using spatio-temporal partitioning and multiple linear regression". *ISPRS Journal of Photogrammetry and Remote Sensing* 198, pp. 115–139.

Mastro P, Peppo MD, Crema A, Boschetti M, and Pepe A (2023). "Statistical characterization and exploitation of synthetic aperture radar vegetation indexes for the generation of leaf area index time series". *International Journal of Applied Earth Observation and Geoinformation* 124, Art. no. 103498.

McFeeters SK (1996). "The use of the normalized difference water index (NDWI) in the delineation of open water features". *International Journal of Remote Sensing* 17.7, pp. 1425–1432.

McNairn H and Shang J (2016). "A review of multitemporal synthetic aperture radar (SAR) for crop monitoring". *Multitemporal Remote Sensing. Methods and Applications*. Vol. 20, pp. 317–340.

Meraner A, Ebel P, Zhu XX, and Schmitt M (2020). "Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion". *ISPRS Journal of Photogrammetry and Remote Sensing* 166, pp. 333–346.

Mermoz S, Réjou-Méchain M, Villard L, Le Toan T, Rossi V, and Gourlet-Fleury S (2015). "Decrease of L-band SAR backscatter with biomass of dense forests". *Remote Sensing of Environment* 159, pp. 307–317.

Moran MS, Alonso L, Moreno JF, Mateo MPC, Cruz DF de la, and Montoro A (2012). "A RADARSAT-2 quad-polarized time series for monitoring crop and soil conditions in Barrax, Spain". *IEEE Transactions on Geoscience and Remote Sensing* 50.4, pp. 1057–1070.

Moravec D, Komárek J, López-Cuervo Medina S, and Molina I (2021). "Effect of atmospheric corrections on NDVI: Intercomparability of Landsat 8, Sentinel-2, and UAV sensors". *Remote Sensing* 13.18, Art. no. 3550.

Moreira A, Prats-Iraola P, Younis M, Krieger G, Hajnsek I, and Papathanassiou KP (2013). "A tutorial on synthetic aperture radar". *IEEE Geoscience and Remote Sensing Magazine* 1.1, pp. 6–43.

Nasirzadehdizaji R, Balik Sanli F, Abdikan S, Cakir Z, Sekertekin A, and Ustuner M (2019). "Sensitivity analysis of multi-temporal Sentinel-1 SAR parameters to crop height and canopy coverage". *Applied Sciences* 9.4, Art. no. 655.

Nguyen VH, Stuart AM, Nguyen TMP, Pham TMH, Nguyen NPT, Pame ARP, Sander BO, Gummert M, and Singleton GR (2022). "An assessment of irrigated rice cultivation with different crop establishment practices in Vietnam". *Scientific Reports* 12.1, Art. no. 401.

Odena A, Dumoulin V, and Olah C (2016). "Deconvolution and checkerboard artifacts". *Distill*.

Oh Y, Hong SY, Kim Y, Hong JY, and Kim YH (2009). "Polarimetric backscattering coefficients of flooded rice fields at L- and C-Bands: Measurements, modeling, and data analysis". *IEEE Transactions on Geoscience and Remote Sensing* 47.8, pp. 2714–2721.

Onojeghuo AO, Blackburn GA, Wang Q, Atkinson PM, Kindred D, and Miao Y (2018). "Rice crop phenology mapping at high spatial and temporal resolution using downscaled MODIS time-series". *GIScience & Remote Sensing* 55.5, pp. 659–677.

Paloscia S (2002). "A summary of experimental results to assess the contribution of SAR for mapping vegetation biomass and soil moisture". *Canadian Journal of Remote Sensing* 28.2, pp. 246–261.

Paluba D, Saux BL, Sarti F, and Stych P (2023). "Estimating optical vegetation indices with Sentinel-1 SAR data and AutoML". arXiv Preprint: `https://arxiv.org/abs/2311.07537`.

Peel MC, Finlayson BL, and McMahon TA (2007). "Updated world map of the Köppen-Geiger climate classification". *Hydrology and Earth System Sciences* 11.5, pp. 1633–1644.

Pelta R, Beeri O, Tarshish R, and Shilo T (2022). "Sentinel-1 to NDVI for agricultural fields using hyperlocal dynamic machine learning approach". *Remote Sensing* 14.11, Art. no. 2600.

Peressutti D (2023). *How to Co-Register Temporal Stacks of Satellite Images*. EOResearch. URL: `https://medium.com/sentinel-hub/how-to-co-register-temporal-stacks-of-satellite-images-5167713b3e0b` (visited on 08/16/2023).

Periasamy S (2018). "Significance of dual polarimetric synthetic aperture radar in biomass retrieval: An attempt on Sentinel-1". *Remote Sensing of Environment* 217, pp. 537–549.

Pettorelli N, Vik JO, Mysterud A, Gaillard JM, Tucker CJ, and Stenseth NC (2005). "Using the satellite-derived NDVI to assess ecological responses to environmental change". *Trends in Ecology & Evolution* 20.9, pp. 503–510.

Pichierri M, Hajnsek I, Zwieback S, and Rabus B (2018). "On the potential of polarimetric SAR interferometry to characterize the biomass, moisture and structure of agricultural crops at L-, C- and X-bands". *Remote Sensing of Environment* 204, pp. 596–616.

Ramathilagam AB, Natarajan S, and Kumar A (2023). "SAR2NDVI: A pix2pix generative adversarial network for reconstructing field-level normalized difference vegetation index time series using Sentinel-1 synthetic aperture radar data". *Journal of Applied Remote Sensing* 17.2, Art. no. 024514.

Raney R, Freeman T, Hawkins R, and Bamler R (1994). "A plea for radar brightness". *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 1090–1092.

Requena-Mesa C, Benson V, Reichstein M, Runge J, and Denzler J (2021). "EarthNet2021: A large-scale dataset and challenge for Earth surface forecasting as a guided video prediction task". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1132–1142.

Roa Y, Rosell P, Solarte A, Euillades L, Carballo F, García S, and Euillades P (2021). "First assessment of the interferometric capabilities of SAOCOM-1A: New results over the Domuyo Volcano, Neuquén Argentina". *Journal of South American Earth Sciences* 106, Art. no. 102882.

Ronneberger O, Fischer P, and Brox T (2015). "U-Net: Convolutional networks for biomedical image segmentation". *Medical Image Computing and Computer-Assisted Intervention*. Vol. 9351, pp. 234–241.

Roßberg T and Schmitt M (2023a). "A globally applicable method for NDVI estimation from Sentinel-1 SAR backscatter using a deep neural network and the SEN12TP dataset". *Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 91 (3), pp. 171–188.

— (2023b). *SEN12TP - Sentinel-1 and -2 images, timely paired*. Dataset available at `https://zenodo.org/records/7342060`.

— (2024a). "Comparing SAR at multiple frequencies with optical vegetation indices for an area in Vietnam". *Proceedings of 15th European Conference on Synthetic Aperture Radar*, pp. 650–653.

— (2024b). "Dense NDVI time series by fusion of optical and SAR-derived data". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17, pp. 7748–7758.

— (2025). "Comparing the relationship between NDVI and SAR backscatter across different frequency bands in agricultural areas". *Remote Sensing of Environment* 319, Art. no. 114612.

Rouse JW, Haas RH, Schell JA, and Deering DW (1974). "Monitoring vegetation systems in the great plains with ERTS". *Proceedings of the Third Earth Resources Technology Satellite-1 Symposium*, pp. 309–317.

Ruciński M, Foks-Ryznar A, Mayos LP, Woźniak E, Domingo-Marimon C, Jenerowicz-Sanikowska M, Krupiński M, Gromny E, and Aleksandrowicz S (2023). "The multi-temporal relationship between Sentinel-1 SAR features and Sentinel-2

NDVI for different land use / land cover classes in Central Africa". *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 325–328.

Ryu JH, Jeong H, and Cho J (2020). "Performances of vegetation indices on paddy rice at elevated air temperature, heat stress, and herbicide damage". *Remote Sensing* 12.16, Art. no. 2654.

Santos EP dos, Silva DD da, Amaral CH do, Fernandes-Filho EI, and Dias RLS (2022). "A machine learning approach to reconstruct cloudy affected vegetation indices imagery via data fusion from Sentinel-1 and Landsat 8". *Computers and Electronics in Agriculture* 194, Art. no. 106753.

Scarpa G, Gargiulo M, Mazza A, and Gaetano R (2018). "A CNN-based fusion method for feature extraction from Sentinel data". *Remote Sensing* 10.2, Art. no. 236.

Schmitt M, Hughes LH, Qiu C, and Zhu XX (2019a). "Aggregating cloud-free Sentinel-2 images with Google Earth Engine". *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-2/W7, pp. 145–152.

Schmitt M, Hughes LH, Qiu C, and Zhu XX (2019b). "SEN12MS – A curated dataset of georeferenced multi-spectral Sentinel-1/2 imagery for deep learning and data fusion". *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-2/W7, pp. 153–160.

Segalini F, Guarini R, Notarnicola C, Vuolo F, and Dini L (2014). "COSMO-SkyMed® imagery for crops characterization". *Proceedings of 34th EARSeL Symposium*, pp. 7.1–7.11.

Shibayama M, Wiegand CL, and Richardson AJ (1986). "Diurnal patterns of bidirectional vegetation indices for wheat canopies". *International Journal of Remote Sensing* 7.2, pp. 233–246.

Sica F, Gobbi G, Rizzoli P, and Bruzzone L (2021). "Φ-net: Deep residual learning for InSAR parameters estimation". *IEEE Transactions on Geoscience and Remote Sensing* 59.5, pp. 3917–3941.

Small D (2011). "Flattening gamma: Radiometric terrain correction for SAR imagery". *IEEE Transactions on Geoscience and Remote Sensing* 49.8, pp. 3081–3093.

Soglia P, Gonzalez P, Azpiroz I, Otamendi U, Quartulli M, and Salata S (2023). "Estimating NDVI from SAR images using conditional generative adversarial networks". *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 5222–5225.

Stendardi L, Karlsen S, Niedrist G, Gerdol R, Zebisch M, Rossi M, and Notarnicola C (2019). "Exploiting time series of Sentinel-1 and Sentinel-2 imagery to detect meadow phenology in mountain regions". *Remote Sensing* 11.5, Art. no. 542.

Sumbul G, Charfuelan M, Demir B, and Markl V (2019). "BigEarthNet: A large-scale benchmark archive for remote sensing image understanding". *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 5901–5904.

Takaku J, Tadono T, Doutsu M, Ohgushi F, and Kai H (2020). "Updates of 'AW3D30' ALOS global digital surface model with other open access datasets". *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLIII-B4-2020, pp. 183–189.

Tian J, Wang L, Li X, Gong H, Shi C, Zhong R, and Liu X (2017). "Comparison of UAV and WorldView-2 imagery for mapping leaf area index of mangrove forest". *International Journal of Applied Earth Observation and Geoinformation* 61, pp. 22–31.

Torres R, Snoeij P, Geudtner D, Bibby D, Davidson M, Attema E, Potin P, Rommen B, Floury N, Brown M, Traver IN, Deghaye P, Duesmann B, Rosich B, Miranda N, Bruno C, L'Abbate M, Croci R, Pietropaolo A, Huchler M, and Rostan F (2012). "GMES Sentinel-1 mission". *Remote Sensing of Environment* 120, pp. 9–24.

Trudel M, Charbonneau F, and Leconte R (2012). "Using RADARSAT-2 polarimetric and ENVISAT-ASAR dual-polarization data for estimating soil moisture over agricultural fields". *Canadian Journal of Remote Sensing* 38.4, pp. 514–527.

U.S. Department of Agriculture, Foreign Agricultural Service, International Production Assessment Division (2017). *South East Asia - Crop Production Maps*. URL: `https://ipad.fas.usda.gov/rssiws/al/crop_production_maps/seasia/Vietnam_Total_Rice.png` (visited on 02/05/2024).

— (2021). *Crop Calendars for Australia*. URL: `https://ipad.fas.usda.gov/rssiws/al/crop_calendar/as.aspx` (visited on 07/25/2024).

— (2024). *Crop Calendars for Southern South America*. URL: `https://ipad.fas.usda.gov/rssiws/al/crop_calendar/ssa.aspx` (visited on 07/25/2024).

Uuemaa E, Ahi S, Montibeller B, Muru M, and Kmoch A (2020). "Vertical accuracy of freely available global digital elevation models (ASTER, AW3d30, MERIT, TanDEM-x, SRTM, and NASADEM)". *Remote Sensing* 12.21, Art. no. 3482.

Veloso A, Mermoz S, Bouvet A, Toan TL, Planells M, Dejoux JF, and Ceschia E (2017). "Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications". *Remote Sensing of Environment* 199, pp. 415–426.

Villarroya-Carpio A and Lopez-Sanchez JM (2024). "Influence of temporal baseline on X-band repeat-pass coherence for crop monitoring". *Proceedings of 15th European Conference on Synthetic Aperture Radar*, pp. 678–683.

Villarroya-Carpio A, Lopez-Sanchez JM, and Engdahl ME (2022). "Sentinel-1 interferometric coherence as a vegetation index for agriculture". *Remote Sensing of Environment* 280, Art. no. 113208.

Vollrath A, Mullissa A, and Reiche J (2020). "Angular-based radiometric slope correction for Sentinel-1 on Google Earth Engine". *Remote Sensing* 12.11, Art. no. 1867.

Vreugdenhil M, Wagner W, Bauer-Marschallinger B, Pfeil I, Teubner I, Rüdiger C, and Strauss P (2018). "Sensitivity of Sentinel-1 backscatter to vegetation dynamics: An Austrian case study". *Remote Sensing* 10.9, Art. no. 1396.

Wang Q and Atkinson PM (2018). "Spatio-temporal fusion for daily Sentinel-2 images". *Remote Sensing of Environment* 204, pp. 31–42.

Wang X, Ge L, and Li X (2013). "Pasture monitoring using SAR with COSMO-SkyMed, ENVISAT ASAR, and ALOS PALSAR in Otway, Australia". *Remote Sensing* 5.7, pp. 3611–3636.

Weerakody PB, Wong KW, Wang G, and Ela W (2021). "A review of irregular time series data handling with gated recurrent neural networks". *Neurocomputing* 441, pp. 161–178.

Xu H (2006). "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery". *International Journal of Remote Sensing* 27.14, pp. 3025–3033.

Yuan X, Shi J, and Gu L (2021). "A review of deep learning methods for semantic segmentation of remote sensing imagery". *Expert Systems with Applications* 169, Art. no. 114417.

Zanaga D, Kerchove RVD, De Keersmaecker W, Souverijns N, Brockmann C, Quast R, Wevers J, Grosu A, Paccini A, Vergnaud S, Cartus O, Santoro M, Fritz S, Georgieva I, Lesiv M, Carter S, Herold M, Li L, Tsendbazar NE, Ramoino F, and Arino O (2021). *ESA WorldCover 10 m 2020 v100*.

Zhao H, Gallo O, Frosio I, and Kautz J (2017). "Loss functions for image restoration with neural networks". *IEEE Transactions on Computational Imaging* 3.1, pp. 47–57.

Zhao W, Qu Y, Chen J, and Yuan Z (2020). "Deeply synergistic optical and SAR time series for crop dynamic monitoring". *Remote Sensing of Environment* 247, Art. no. 111952.

Zhu Z and Woodcock CE (2014). "Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change". *Remote Sensing of Environment* 152, pp. 217–234.

Zupanc A (2017). *Improving Cloud Detection with Machine Learning*. URL: `https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13` (visited on 02/06/2023).

Acknowledgments

First and foremost, I would like to express my deep gratitude to my doctoral advisor, Prof. Michael Schmitt, for his invaluable supervision and guidance throughout the course of my research. His insightful feedback and support have been instrumental in the completion of this work. I also want to thank him together with Prof. Juan Manuel Lopez-Sanchez for being examiners of this dissertation and Prof. Roger Förstner for being head of the examination comitee.

I am also incredibly grateful to my colleagues, Max, Burak 🙃, Michael, Jonathan, Mojgan, Paolo, Manuel, Anton, and Deepika. Thank you for your camaraderie and friendship, the fun times in the office, during the barbecue, or at conferences, and all your help.

To my girlfriend: Thank you for your love, support, and understanding. Thank you from the bottom of my heart.

I am also profoundly thankful for the love, support, and encouragement from my parents, sisters, and brother.

Data Acknowledgments

I thank ESA for providing Sentinel-1 and Sentinel-2 data, as well as ESA and e-GEOS for access to COSMO-SkyMed satellite images under the TPM scheme.
I gratefully acknowledge the use of SAOCOM® products, courtesy of CONAE - 2023. All rights reserved. The products were processed under a CONAE license.
This research utilized NovaSAR-1™ Level 1 Data © SSTL 2023. The data, generated by the NovaSAR-1™ satellite owned and operated by Surrey Satellite Technology Limited, was provided by CSIRO as the authorized licensor of SSTL Data.
I would like to express my gratitude to CSIRO and, in particular, Zheng-Shu Zhou, for providing me with comprehensive responses to my queries about NovaSAR-1 data and its processing.