UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingenieros Informáticos

Multisource Spatio-Temporal-Spectral Remote Sensing Data Fusion

D O C T O R A L T H E S I S

Submitted for the degree of Doctor by:

Meryeme Boumahdi
MSc in Computer Systems & Networks

Madrid, 2025

UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingenieros Informáticos

Doctoral Degree in Software, Systems and Computing

Multisource Spatio-Temporal-Spectral Remote Sensing Data Fusion

D O C T O R A L T H E S I S

Submitted for the degree of Doctor by:

Meryeme Boumahdi
MSc in Computer Systems & Networks

Under the supervision of:
Consuelo Gonzalo Martín Ph.D
Angel Mario Garcia Pedrero Ph.D

Madrid, 2025

Title: Multisource Spatio-Temporal-Spectral Remote Sensing Data Fusion

Author: Meryeme Boumahdi

Doctoral Programme: Software, Systems and Computing

Thesis Supervision:

Dr. Consuelo Gonzalo Martín, Catedrática de Universidad, Escuela Técnica Superior de Ingenieros Informáticos

Dr. Ángel Mario García Pedrero, Profesor Contratado Doctor, Escuela Técnica Superior de Ingenieros Informáticos

External Reviewers:

Thesis Defense Committee:

Thesis Defense Date:

In loving memory of my father Abdelouahed who gave the little he had so I would have the opportunity of an education. I miss you profoundly, but your love echoes in every corner of my life – my dad, my hero, my eternal inspiration.

Acknowledgement

I always thought that writing the acknowledgments would be the easiest part of this thesis. It has turned out to be the hardest part to write. Because during these five years, I've met exceptional people, and there are simply no words that feel big enough, deep enough, or kind enough to thank them the way they deserve. So, to all of you who have been part of this chapter of my life, thank you. I carry pieces of you with me on every page of this thesis.

I would like to start by thanking my supervisor *Consuelo Gonzalo Martin*. From the beginning, you believed in me. You gave me the opportunity to embark on this PhD journey, and that trust has meant everything to me. I couldn't have wished for a better supervisor or a better human being to guide me through these years. These have been intense years filled with hard work, delays, uncertainties, and results that didn't always meet our expectations. But you always encourage me to grow and give me space to do it. And when it came to paperwork, you were always the first to sign, even at odd hours. But more than that, thank you for being there for me as a person. I will never forget the warmth of your hug when I lost my father. It wasn't just a gesture, it was a reminder that I wasn't alone. That human connection meant more to me than any words could express. Over time, you became more than a supervisor, you became someone I deeply admire, someone I could trust and lean on. You've shown me that it's possible to be gentle and kind, and still take your work seriously, that guidance can be given with both firmness and warmth, and that success does not have to come at the expense of humanity. Even when things did not go well, you always carefully chose your words, never letting feedback become discouragement.

Angel Mario Garcia Pedrero, it was a privilege to work with you. Thank you for your time, your guidance and your sharp eye in helping me correct and reshape this work. Because of you, this thesis went further than I could have ever imagined. I still remember our very first meetings, when you asked me questions about the research and I completely panicked. But instead of judging me, you helped me rebuild my thinking, clarify my ideas, and slowly you taught me how to find my voice. Thanks to you, today I can stand up and discuss my work with confidence. You always knew when to push me gently out of my comfort zone and when to be supportive and understanding. Over the years, you were always there. I could count on you. And more times than I can remember. Beyond all of that, your work ethic has been a constant source of inspiration. The dedication and passion you bring to your work has pushed me to give more, to try harder. Thank you, *Ángel*, for believing in me, for guiding me, and for being such a steady, generous presence throughout this journey. I can proudly say that now you are family.

I can't forget *Mario Lillo-Saavedra*. Thank you for welcoming me not only into your research group, but into your country, your home, and your world. We had met before, during your visits to Madrid, but during my stay in Chile, I got to know you on a deeper level as a mentor, a creative mind, a supportive colleague, a funny person. Your ideas, your encouragement, and your thoughtful feedback gave me a new direction in my work. Because of you, I had one of the most fulfilling and memorable experiences of my PhD journey. But what truly made that time in Chile unforgettable was the warmth with which you and your family welcomed me. Thank you for opening your home to me, for introducing me to your beautiful family,

and for letting me feel like I belong. *Ximena*, thank you for your endless kindness, you were my guide in Chillán, always ready to help, and *Florencia*, even with your shy smile, you offered to translate Spanish. That simple act of kindness meant so much. I truly thank you for everything, for collaboration, conversations, culture, and care. With you, I got to discover not only a beautiful research community, but also a beautiful country and a friend for life.

To my labmates and colleagues, those who are still there and those who have already moved on to a new chapter of their life, thank you. You were the people who turned long days into shared experiences and moments that helped make this journey lighter. *Mahdi*, thank you for always being the one who brings people together to organize gatherings outside the CTB. To everyone I met at the CTB, thank you for the camaraderie, the hallway chats, the help with last-minute technical issues or paperwork.

Among all those people, there's someone I cannot forget *Michela*. The connection we built went far beyond the workplace and far beyond borders. You are one of the most brilliant, kind, and open-hearted people I've met during this journey. You showed me that real friendship doesn't need to come from years of knowing someone, it can grow fast, deeply, and across any language, cultural or religious difference. Thank you for all the funny conversations we had about our (very questionable!) cooking skills, for introducing me to your big and small family, and for showing me the beauty of connection without boundaries.

To my dear friends in Madrid, thank you for becoming such a big part of my life and for turning a city into a home. *Souad* and *Hasnae* for those endless walks that somehow stretched over 15 kilometers because we simply couldn't stop talking. With you, I found a small piece of Morocco in Madrid, where I could speak my language, and you remind me of how are incredible women from our country. *Dionisio*, *Gina* and *Pili*, thank you for opening the doors of your home and your hearts to me. We celebrated so many moments together, New Year dinners filled with laughter, small gatherings that turned into unforgettable memories, and countless nights where I always felt like part of the family.

To my dear friends in Morocco *Halima*, *Hamza*, *Mohamed*, *Riane*. Thank you for proving that true friendship knows no distance. Even with thousands of kilometers between us. You celebrated my achievements as if they were your own and offered comfort in moments when I felt overwhelmed or alone. Despite the time zones, the busy lives, and the physical distance, you made sure that I always felt connected to you, to my home, and to the person I was before this journey began.

None of this would have happened without my parents. *My Father*, my number one fan, always. You were endlessly proud of me for defending me unconditionally, believing in. Thank you for giving me all the love that a father could give, for every sacrifice you made so I could live a better life. You uprooted everything, your home, your friends, your work, to move with me to another city just so I could study in comfort. You never said no to anything I needed. I still remember my very first day of school, when you bought me the best and most expensive backpack and told me: "The smartest girl in town deserves the best." Thank you for giving me the freedom to choose my own path, even when others around us didn't understand or approve. You never cared about what society expected, only what was best

for me. Thank you for raising me with strength, with love, and with the kind of values that carried me through every step of this journey. You always dreamed of the day that I would earn my Ph.D. I just wish you had stayed a little longer just a bit more. Death stole that moment and took you too soon. But you have never truly left me.

Mom, your love has been the constant thread that holds everything together, quiet, powerful, and unconditional. You gave without limits your time, your energy, your prayers, your patience. You carried my worries as if they were your own and celebrated every small victory with the joy of a mother who always believed her child was capable of the impossible. Even when I doubted myself, you never did. Thank you for your quiet sacrifices, the ones I didn't always see at the time. For making sure I was always taken care of, for creating a home filled with safety and love, and for giving me the kind of love that never asked for anything in return. You are the definition of unconditional love.

To the rest of my family, my cousins, uncles, and aunties, thank you for always being there, near or far. Your words of encouragement, your pride in my journey and the warmth you've given me in so many forms have all been part of this achievement. I felt your love even through distance and carried it with me through the hardest and most beautiful moments of these years. And to *Rabia*, I can proudly call my sister. In a time when I lost so much, you stood firmly on my side, offering everything I needed without ever asking for anything in return. Thank you for being a constant when everything else felt uncertain and for lifting me up without needing words. To *Khalifa*, you were my father's best friend, and when he left this world, you could have stepped back, no one would have blamed you. But instead, you chose to stay. You continued to support me when it would have been easier to leave. And for that, you will always be my big brother.

Last but not least, I would like to express my deepest gratitude to *Mujeres por África Foundation and GMV*. Without your support, this journey would have been a thousand times more difficult. You gave me the freedom to focus on my research without the constant weight of financial worry and that is a gift that I will never forget. But your support went far beyond the funding. You opened doors that helped shape the researcher and woman I've become. To *Beatriz* and *Eva*, thank you for always staying in contact, for your guidance, your encouragement, and for showing me what true empowerment looks like in action. You reminded me that lifting women up is not just a slogan, it's a daily practice, and you live it.

Abstract

The generation of high-quality satellite image time series is fundamental for understanding dynamic Earth surface processes and supporting a range of applications including environmental monitoring, vegetation phenology, land cover change detection. However, creating temporally dense, spatially detailed, and spectrally consistent remote sensing time series remains a substantial challenge. These limitations are particularly evident in cloud-prone and topographically complex regions, where persistent cloud cover, spectral discrepancies between sensors, and temporal gaps undermine the usability of optical imagery.

This thesis proposes a comprehensive, multimodal and multisource fusion framework that integrates data from complementary Earth observation sensors to address the compounded limitations of existing remote sensing time series generation methods. The framework is built around three critical challenges: sensor-specific spectral misalignment, inconsistent temporal coverage, and persistent cloud cover. It integrates optical, radar, and topographic information to produce spatially rich, spectrally coherent, and temporally continuous time series.

The first part of the framework focuses on spectral alignment. Differences in spectral characteristics across sensors often result in distortions in fused imagery. This work introduces a spectral alignment strategy that reduces spectral inconsistencies by adjusting the reflectance between sensors. This step is essential to preserve the physical meaning of the data and ensure the reliability of downstream applications that depend on multisensor fusion.

Cloud contamination is also addressed as one of the most persistent obstacles in optical images. A deep learning model is proposed that combines multimodal inputs to restore cloud-obscured areas. This approach is particularly effective in mountainous regions, where frequent cloud cover, snow, and terrain-induced shadows create complex conditions for cloud removal. By leveraging the all-weather capabilities of radar and the structural cues from terrain models, the framework enhances both the availability and quality of cloud-free observations.

The final component of the framework is a data fusion strategy that integrates multi-temporal and multi-resolution imagery to generate temporally dense time series. This fusion is performed while preserving spatial and spectral detail, enabling consistent monitoring even in cases of sparse optical acquisitions. The framework adapts to data availability and landscape variability, ensuring reliable performance under a range of environmental conditions.

In addition to methodological contributions, this thesis introduces a publicly available benchmark dataset designed specifically for spatiotemporal fusion between Sentinel-2 and Sentinel-3 imagery. This dataset spans environmentally diverse locations and includes wide range of spectral bands and temporally matched observations, providing the community with a valuable resource for evaluating and comparing data fusion methods using European satellite missions.

This research advances the state of the art in remote sensing by delivering an adaptable, integrated fusion framework that addresses the core limitations of existing time series reconstruction approaches. The proposed methods enhance the usability of multisensor satellite data and open new possibilities for operational Earth observation in regions where conventional techniques fail. The contributions of this thesis have broad implications for the future

of remote sensing, particularly in supporting scalable and accurate monitoring of dynamic landscapes under real-world observational constraints.

Resumen

La generación de series temporales de imágenes satelitales de alta calidad es fundamental para comprender los procesos dinámicos de la superficie terrestre y para apoyar una amplia gama de aplicaciones, como el monitoreo ambiental, la fenología de la vegetación y la detección de cambios en la cobertura del suelo. Sin embargo, la creación de series temporales con alta frecuencia temporal, alta resolución espacial y consistencia espectral, sigue siendo un desafío significativo, especialmente en regiones montañosas o con alta nubosidad, donde la cobertura persistente de nubes, las discrepancias espectrales entre sensores y las discontinuidades temporales limitan la fiabilidad de los datos ópticos.

Esta tesis propone un marco integral de fusión multimodal y multifuente, que integra datos complementarios de observación terrestre para abordar de forma conjunta estas limitaciones. El marco se enfoca en tres desafíos principales: la desalineación espectral entre sensores, la cobertura temporal inconsistente, y la presencia persistente de nubes. En ella, se explotan las fortalezas de los datos ópticos, radar y topográficos para generar secuencias de imágenes que son ricas espacialmente, coherentes espectralmente y continuas en el tiempo.

El primer componente del marco se centra en la alineación espectral. Las diferencias en las características espectrales entre sensores pueden introducir inconsistencias en los productos fusionados. Se propone una estrategia de ajuste de reflectancia para reducir estas discrepancias y alinear las respuestas espectrales entre sensores. Este paso es esencial para preservar el significado físico de los valores de reflectancia, y así garantizar la fiabilidad de aplicaciones posteriores como la estimación de índices de vegetación o la clasificación del uso del suelo.

El segundo componente aborda la contaminación por nubes, uno de los principales obstáculos para generar series temporales de alta calidad. Se introduce un método de reconstrucción basado en aprendizaje profundo que combina entradas multimodales para restaurar las regiones ocultas por nubes. Este enfoque resulta especialmente eficaz en zonas montañosas, donde la nubosidad frecuente, la presencia de nieve y las sombras provocadas por el relieve complican los métodos tradicionales de eliminación de nubes. Al incorporar las capacidades todo-tiempo del radar y la información estructural derivada de modelos topográficos, el marco mejora significativamente la disponibilidad y calidad de las imágenes reconstruidas sin nubes.

El tercer componente es una estrategia de fusión espaciotemporal que integra datos multitemporales y multirresolución para generar series temporales densas y de alta resolución. Este proceso preserva tanto el detalle espacial como la fidelidad espectral, permitiendo observaciones consistentes incluso cuando las adquisiciones ópticas son limitadas. El marco se adapta a la variabilidad del paisaje y a la disponibilidad de sensores, garantizando un rendimiento robusto bajo condiciones reales.

Además de estas contribuciones metodológicas, la tesis presenta un conjunto de datos de referencia públicamente disponible, diseñado específicamente para la fusión espaciotemporal entre imágenes de Sentinel-2 y Sentinel-3. Este conjunto cubre ubicaciones ambientalmente diversas y ofrece bandas espectrales armonizadas junto con observaciones temporalmente coincidentes, proporcionando a la comunidad una herramienta valiosa para la evaluación

comparativa de métodos de fusión basados en satélites europeos.

Esta investigación impulsa el estado del arte en la teledetección al presentar un marco de fusión adaptable e integrado que aborda las limitaciones clave en la reconstrucción de series temporales. El enfoque propuesto mejora la usabilidad de los datos multifuente y permite una observación terrestre más precisa, escalable y consistente, especialmente en regiones dinámicas o difíciles de observar con métodos convencionales.

Table of Contents

List of Figures

List of Tables

Abbreviations and acronyms

BOA Bottom-Of-Atmosphere

CNN Convolutional Neural Networks

CS Component Substitution

CRT-UNet Cloud Removal with Topographic information UNet

DEM Digital Elevation Model

DL Deep Learning

DN Digital Number

ESA European Space Agency

FSDAF Flexible Spatiotemporal Data Fusion

GEE Google Earth Engine

GAN Generative Adversarial Networks

HSLS High Spatial Low Spectral

HSHS High Spatial High Spectral

HSLT High Spatial Low Temporal

iCOR Image Correction for Atmospheric Effects

LSHS Low Spatial High Spectral

LSHT Low Spatial High Temporal

MAJA MACCS-ATCOR Joint Algorithm

ML Machine Learning

MMT Multisensor Multiresolution Technique

MRA Multiresolution Analysis

MSI MultiSpectral Instrument

MSTBA Multispectral Temporal Band Adjustment

MSI Multispectral Image

OLCI Ocean and Land Colour Instrument

PSNR Peak Signal to Noise Ratio

ReLU Rectified Linear Unit

RM Regression Model

RMSE Root Mean Square Error

SAM Spectral Angle Mapper

SAR Synthetic Aperture Radar

S1 Sentinel-1

S2 Sentinel-2

S2A Sentinel-2A

S2B Sentinel-2B

S3 Sentinel-3

SRTM Shuttle Radar Topographic Mission

SRF Spectral Response Function

SSIM Structural Similarity Index

STARFM Spatial and Temporal Adaptive Reflectance Fusion

STF Spatiotemporal Fusion

TOA Top-of-Atmosphere

UPM Universidad Politécnica de Madrid

VH Vertical-Horizontal

VV Vertical-Vertical

Chapter 1

Introduction

Monitoring the Earth's surface is a critical endeavor in understanding and managing the complex interplay between natural ecosystems and human activities. Remote sensing, with its ability to capture large-scale, continuous observations of the planet, plays an indispensable role in this effort. By providing spatially explicit data across various temporal scales, remote sensing has revolutionized our ability to do environmental monitoring, disaster management and risk assessment. It enables insights into processes that shape the Earth's surface, from deforestation and urbanization to natural disasters and seasonal cycles. The capacity to observe these changes is pivotal not only for scientific understanding, but also to inform sustainable development policies, disaster response strategies, and climate adaptation measures (Pettorelli et al., 2014; Turner et al., 2015).

International organizations such as the Intergovernmental Panel on Climate Change (IPCC) have emphasized the need for robust Earth observation systems to track land use and land cover changes, which are essential for understanding carbon dynamics and mitigating climate change (Nkonya et al., 2019). Similarly, the Sustainable Development Goals (SDGs) of the United Nations underscore the importance of monitoring land cover and vegetation to ensure food security, manage water resources, and protect biodiversity (Nationen, 2015). The critical importance of such monitoring has driven investments in satellite programs such as the European Space Agency's (ESA) Copernicus initiative, which provides high-resolution, open-access data through sensors with different characteristics. Figure 1.1 illustrates the satellites within the Copernicus initiative, including both operational and planned missions.

High-resolution sensors have longer revisit times, while frequent observations come at a coarser scale, and persistent cloud cover further disrupts optical data, creating gaps that hinder accurate monitoring of vegetation and land cover dynamics. These data gaps hinder the capability to accurately capture temporal changes, such as seasonal vegetation cycles, the aftermath of natural disasters, or the impacts of human activities like deforestation and agricultural expansion.

High quality time series characterized by high temporal frequency for continuous observation of dynamic environmental changes and high spatial resolution for detailed representation of land cover features are essential for detecting both rapid and gradual changes across diverse

Figure 1.1: Copernicus Earth observation missions developed and planed by ESA. [Credits: ESA]

ecosystems. Studies by (Pettorelli et al., 2018; Turner et al., 2015) have highlighted the importance of such datasets in advancing our understanding of ecosystem processes, from tracking phenological shifts due to climate change to assessing land degradation in vulnerable regions. Despite significant advancements, achieving high-quality time series data remains a persistent challenge, particularly in regions characterized by frequent cloud cover and complex topography.

Advances in remote sensing technology have paved the way for innovative approaches to bridge these gaps. The development of multi-sensor satellite systems offers a promising solution by combining the unique strengths of different sensors. Optical satellites, such as Sentinel-2 (S2), currently provide detailed spatial and spectral information critical for distinguishing between land cover types, while radar sensors like Sentinel-1 (S1) can penetrate clouds, offering all-weather capabilities. Sentinel-3 (S3) complements these by providing frequent revisit times, ensuring temporal continuity. However, integrating these diverse data sources into cohesive and high-quality time series remains a complex task that requires sophisticated methods for data harmonization, cloud removal, and fusion.

This thesis deals with the development of methodologies to address these challenges, with the goal of generating temporally dense and high-quality time series for monitoring vegetation and land cover changes. These datasets are not only vital for capturing the spatial and temporal variability of land cover dynamics but also for supporting critical applications in agriculture, forestry, hydrology, and conservation. The ability to monitor these changes with precision and consistency has profound implications for policy-making and resource management, particularly in the context of global challenges such as climate change and biodiversity loss.

1.1 Importance of Data Fusion in Remote Sensing

The inherent limitations of individual satellite sensors, including trade-offs between spatial, temporal, and spectral resolutions, often constrain their capacity to fully capture Earth dynamics and land surface changes. Data fusion has emerged as a powerful solution, leveraging the complementary strengths of multiple sensors to overcome these limitations and produce enhanced datasets that better meet the needs of diverse applications.

One of the key benefits of data fusion lies in its ability to improve spatial, temporal, and spectral resolutions simultaneously. By combining data from spatial high-resolution sensors, such as S2 and Landsat, with frequent observations from coarser sensors like S3 or MODIS, researchers have generated temporally dense datasets that retain critical spatial detail. Such datasets enable the monitoring of vegetation health, flood risk, wildfire detection, water surface dynamics and urban planning (Belgiu & Csillik, 2018; F. Gao et al., 2006). For example, integrating Landsat's fine spatial resolution with MODIS's frequent revisits has facilitated the creation of high-quality time series that are indispensable for detecting transient phenomena like crop stress or pest outbreaks (F. Gao et al., 2017).

Data fusion also plays a pivotal role in mitigating the effects of persistent cloud cover, one of the primary challenges in optical remote sensing. In regions with frequent atmospheric disturbances, such as tropical rainforests, mountainous landscapes, and high-latitude snow-covered areas, clouds can obscure the surface and render optical imagery unusable. By integrating optical data with radar observations from sensors like S1, which can penetrate clouds, data fusion enables reconstruction of cloud-free imagery. This approach has been successfully applied to generate gap-free datasets for vegetation monitoring, snow cover mapping, and land use assessments in areas prone to persistent cloud cover (Reiche et al., 2016; Salcedo & Cogliati, 2014).

Applications of optical-radar fusion extend beyond cloud removal. In snow-covered regions, such as the Andes, Alps, or Himalayas, distinguishing between snow and clouds is a significant challenge due to their similar spectral properties. Radar data, with its sensitivity to surface structure and moisture content, provides valuable complementary information that helps differentiate between these features. Studies combining S1 and S2 data have shown improved accuracy in snow extent mapping, particularly in mountainous regions with complex terrain and dynamic snow cover (Solberg et al., 2008). This capability supports hydrological modeling and water resource management by providing reliable information on snowpack dynamics.

Vegetation monitoring is another domain where data fusion has proven transformative. By integrating optical and radar data, researchers have achieved enhanced detection of vegetation growth, stress, and phenological changes. Optical sensors excel at capturing spectral information related to chlorophyll content and leaf structure, whereas radar sensors provide insights into canopy structure and biomass, even under cloudy conditions. Such integrated datasets enable more comprehensive monitoring of forests, agricultural lands, and wetlands, contributing to improved management of natural resources and mitigation of climate change impacts (Joshi et al., 2016; Z. Lu et al., 2010).

Beyond snow and vegetation monitoring, the benefits of data fusion extend to disaster

management and land cover classification. In flood-prone regions, for instance, combining radar data with optical imagery improves the delineation of inundated areas, even in cloudy conditions, enabling more accurate assessments of flood extent and damage (Klemas, 2015; Munawar et al., 2022). Similarly, fused datasets have been used to monitor deforestation and urban expansion, where the integration of radar's structural information with optical spectral detail provides a nuanced understanding of land cover changes (Solórzano et al., 2023).

While the advantages of data fusion are substantial, its successful implementation requires addressing challenges such as sensor misalignments, radiometric inconsistencies, and computational demands. Advanced algorithms and machine learning-based approaches have been developed to harmonize datasets from different sensors and ensure seamless integration (Emelyanova et al., 2013). These methods enable the generation of high-quality, multi-sensor datasets that are robust and reliable, facilitating their application across diverse environmental and societal challenges.

1.2 Motivation and Problem Statement

This thesis tackles the challenges involved in generating high-quality, temporally dense remote sensing time series using data fusion methods. While data fusion has become an essential tool for enhancing the spatial, temporal, and spectral resolutions of remote sensing data, several persistent problems limit its effectiveness, particularly when applied to cloud-prone regions, complex terrains, and multispectral imagery. These challenges hinder the potential of data fusion to fully support critical applications such as vegetation and land cover monitoring, snow cover assessment, and other environmental analyses.

One of the fundamental challenges in remote sensing data fusion is data heterogeneity, arising from the diverse characteristics and sources of datasets. These sources, ranging from satellites to drones and ground-based sensors, differ in spatial resolution, spectral bands, and temporal frequency (H. Wang et al., 2018). While this diversity provides flexibility in selecting data for specific applications, it also demands rigorous preprocessing, calibration, and normalization to ensure compatibility. For example, spatial differences require resolution harmonization, spectral differences necessitate adjustments across bands, and temporal variations demand synchronization for consistent analysis. This inherent complexity underscores the critical need for robust methodologies capable of handling such heterogeneity efficiently.

Another significant challenge lies in ensuring that spectral information remains consistent during the fusion process. Differences in sensor design, calibration methods, and illumination geometries can lead to discrepancies in reflectance values for the same object across different datasets (Jiang et al., 2022; Paolini et al., 2006). Additionally, atmospheric variations at the time of acquisition further contribute to these spectral differences, complicating the integration of multi-sensor data. These inconsistencies can distort the geometric and spectral integrity of features in fused images, leading to spectral inaccuracies and reduced reliability of the time series (Buntilov & Bretschneider, 2008).

Moreover, a significant issue in integrating images from different dates lies in the common practice of processing spectral bands independently, effectively treating them as isolated

entities. This approach neglects the inherent interdependencies among spectral bands, which often contain complementary information that could enhance the reconstruction of spectrally consistent images. Ignoring these cross-band relationships can lead to distortions in the spectral integrity of the fused time series, particularly in complex landscapes where the interplay between vegetation, snow, and atmospheric conditions varies across bands. As a result, the generated time series may lack the precision and robustness required for detailed environmental monitoring, underscoring the need for fusion frameworks that explicitly account for spectral coherence across bands (S. Liu et al., 2022).

Data quality and cloud cover further complicate the process of generating time series. Persistent cloud cover significantly reduces the availability of cloud-free imagery, creating gaps in the time series and limiting its utility for applications such as vegetation monitoring, land cover analysis, and snow cover assessment (Mao et al., 2022). Regions with frequent clouds, such as mountainous regions, are characterized by abrupt elevation changes, and frequent shadowing caused by complex terrain. Existing cloud removal models are often optimized for flatter landscapes and fail to differentiate between clouds, snow, and shadows effectively in such environments. This inadequacy results in errors that compromise the accuracy and utility of reconstructed time series in regions where reliable data is most needed.

In addition to these problems, there is a lack of cohesive frameworks that integrate multiple data fusion techniques to address the diverse limitations of individual methods. Current fusion approaches tend to focus narrowly on improving either spatial, temporal, or spectral resolutions, without leveraging their combined potential. For instance, fusion methods that improve spatial and temporal resolution prioritize temporal consistency but often overlook spectral details, while spatio-spectral fusion techniques may fail to account for temporal gaps or variability. The fragmented nature of these methodologies results in datasets that are less than optimal for monitoring dynamic environmental processes. A comprehensive framework that harmonizes multiple fusion techniques is essential to producing robust, high-quality datasets suitable for a wide range of applications.

1.3 Research Objectives and Contributions

The primary objective of this research is to design, develop and implement a comprehensive framework for generating high-quality time series of remote sensing images for Earth observation applications. This framework aims to address critical challenges posed by cloud cover, data heterogeneity, and spectral inconsistencies, leveraging multi-sensor fusion techniques to enhance spatial, temporal, and spectral resolutions. This thesis contributes to advancing remote sensing methodologies by integrating innovative approaches for data harmonization, cloud removal, and data fusion, with a particular focus on their application in environmentally and climatically diverse regions.

The specific research objectives of this thesis are as follows:

1. **Design and developement of a multi-sensor data harmonization pipeline**: This involves developing a preprocessing approach to align data from S1, S2, and S3, ensuring spatial, spectral consistency. The aim is to address data heterogeneity

challenges by aligning datasets with varying resolutions, sensor characteristics, and temporal frequencies.

2. **Development of a cloud removal model for complex terrains**: A deep learning-based approach is proposed that integrates optical and radar data with topographic information to reconstruct cloud-free imagery. This objective focuses on overcoming the limitations of existing cloud removal techniques, particularly in mountainous areas where cloud cover, snow, and shadows create significant challenges.
3. **Demonstration of the practical utility of the proposed framework**: The final objective is to validate the effectiveness of the proposed techniques by generating reliable S2 time series and evaluating their performance in real-world conditions. This includes testing in dynamic and challenging environments where factors such as cloud cover, terrain complexity, and land cover variability are present.

1.4 Structure of the Thesis

This thesis consists of total six chapters including the introduction chapter. Each one contributing to the overarching goal of generating high-quality time series of remote sensing images. The chapters are outlined as follows:

Chapter 2 reviews the state of the art in data fusion for remote sensing. It provides an overview of remote sensing and Earth observation concepts, introduces key principles, and explores various types of remote sensing imagery. The chapter also addresses challenges in obtaining remote sensing time series and introduces image fusion as a solution. It categorizes fusion techniques and discusses their objectives, methodologies, and challenges with a focus on spatial data integration techniques and their applications. This chapter provides a comprehensive analysis of existing methods, identifying their strengths, limitations, and relevance to the research objectives.

Chapter 3 investigates the fusion of images with different spectral resolutions for time series. The omission of spectral differences can lead to information loss and inaccuracies in the fusion outcome. By examining these spectral discrepancies, we demonstrate how incorporating more bands in the adjustments improves the quality of fused images, ensuring more accurate and reliable results. Additionally, by addressing the issue of overlapping spectral bands from different sensors, we demonstrate the benefits of incorporating previously unused spectral information in spatiotemporal fusion. This highlights the importance of spectral harmonization in multi-sensor integration.

Chapter 4 focuses on a novel cloud removal model that addresses challenges posed by persistent cloud cover, particularly in mountainous regions. By integrating optical and radar data with topographic information, the model effectively reconstructs cloud-free images while preserving spatial and spectral details. Our investigation underscores the significance of incorporating additional topographic data as inputs, particularly in regions characterized by mountainous terrain. Through a comprehensive analysis, we demonstrate how leveraging such data improves the spatial precision and overall efficacy of the model, thereby offering valuable insights for practical applications.

Chapter 5 explores the application of fusion techniques to generate high-quality time series for vegetation and land cover monitoring. Building on the spectral harmonization and cloud removal methodologies introduced earlier, this chapter employs those fusion models to combine multi-sensor data, enhancing temporal resolution while retaining spatial and spectral quality. The effectiveness of the approach is demonstrated through case studies in different mountainous regions, showcasing its potential for real-world environmental monitoring.

The final chapter concludes the thesis by summarizing the key findings and contributions of the research. It reflects on the effectiveness of the proposed methodologies in overcoming challenges associated with cloud cover, data heterogeneity, and spectral variability. Additionally, it discusses the broader implications of the research, highlighting its relevance to remote sensing, environmental monitoring, and sustainable resource management. Finally, it outlines future directions for advancing data fusion techniques, including the integration of emerging sensor technologies, the development of more adaptive fusion models, and the expansion of applications to other dynamic environmental processes.

Chapter 2

Background and Literature Review

The Sentinel satellite constellation, operated by the ESA under the Copernicus program, is designed to provide complementary Earth observation data across multiple spatial, spectral, and temporal scales. In particular, ESA's S1, S2, and S3 satellites form a cornerstone of modern Earth observation, providing critical data for applications, ranging from environmental monitoring to disaster management (Phiri et al., 2020). S1 supplies radar data capable of penetrating clouds and providing high-frequency surface information (X. Zhang et al., 2022), while S2 offers detailed optical images, ideal for land cover and vegetation analysis (Phiri et al., 2020). S3, designed for large-scale environmental monitoring, focuses on ocean and land surface observation with moderate spatial resolution and high temporal frequency (Toming et al., 2017). Each of these satellite's sensors excels in specific areas but also faces inherent limitations due to differences in spatial, spectral, and temporal characteristics.

To fully utilize the potential of these diverse sensors, data fusion techniques have emerged as essential tools, allowing the integration of complementary datasets from multiple sources. By integrating data from multiple sensors with complementary characteristics, data fusion techniques improve remote sensing applications by enhancing spatial resolution, increasing temporal frequency, and ensuring spectral consistency across different observations (Ghassemian, 2016). Combining optical and radar datasets enables continuous monitoring, even in the presence of data gaps caused by atmospheric interference, sensor limitations, or acquisition constraints. Additionally multisensor spectral alignment helps reduce discrepancies arising from different sensor characteristics. However, harmonizing data from different sensors especially those with varying spectral and spatial resolutions, (such as S2 and S3) remains a complex challenge in remote sensing. This chapter provides a comprehensive review of the Sentinels mission and it's sensors systems, outlining the individual capabilities and limitations of S1, S2, and S3. It then explores the various approaches to data fusion, including spectral, spatial, temporal, and multimodal techniques, with a focus on their application in remote sensing.

2.1 Overview of Sentinel 1, 2, and 3 Sensors

Sentinel missions, S1, S2, and S3 play a crucial role in multisensor fusion, as they offer distinct yet overlapping capabilities that, when integrated, enhance time series reconstruction and environmental monitoring. Each of these satellites employs different sensor modalities to capture surface information (Campbell & Wynne, 2011). S2 and S3 carry passive optical sensors, which rely on sunlight to measure reflectance across multiple spectral bands, making them ideal for vegetation analysis, land cover mapping, and water quality assessment. However, these optical sensors are highly susceptible to cloud cover, leading to gaps in time series data (Khanal et al., 2020). In contrast, S1 operates as an active radar mission, using Synthetic Aperture Radar (SAR) to penetrate clouds and acquire imagery under all weather conditions, ensuring continuous monitoring (Potin et al., 2012; Shang et al., 2020). While SAR provides valuable structural information, it lacks spectral detail, making it less effective for land cover classification without auxiliary optical data. These fundamental differences in sensor characteristics present challenges when attempting to integrate their data into a unified time series. S3 offers higher temporal resolution than S2, but its spectral response differs, requiring spectral harmonization. S1 provides consistent observations but lacks spectral information, necessitating multimodal fusion techniques. The sections below provide a detailed overview of each Sentinel mission, highlighting their individual capabilities, limitations, and the challenges associated with integrating their datasets for remote sensing applications.

2.1.1 Sentinel-1

S1 is part of the ESA's Copernicus initiative, designed to provide high resolution radar imagery for global environmental monitoring. The S1 mission consists of two satellites, Sentinel-1A and Sentinel-1B, launched in 2014 and 2016, respectively (Potin et al., 2019). Operating in a near-polar, sun-synchronous orbit at approximately 693 km altitude, these satellites provide global coverage with a six-day revisit time. This revisit frequency is achieved by phasing the two satellites 180° apart, ensuring regular and comprehensive data acquisition. Equipped with a C-band SAR, S1 captures high resolution imagery under all weather conditions, day or night. This radar system is particularly useful for applications where optical sensors might fail due to cloud cover or low-light conditions. S1 operates in four imaging modes, with resolutions as fine as 5 meters and coverage areas as large as 400 km. Its dual-polarisation capability allows for the collection of more detailed information by transmitting and receiving radar signals in different polarizations, enhancing the depth and quality of the data captured (Geudtner et al., 2014).

In addition to its flexible spatial resolution, S1's six-day revisit cycle ensures frequent data collection, making it well-suited for applications that require regular updates, such as disaster management and land-use monitoring. The combination of high spatial resolution and short revisit times enables timely and reliable monitoring of rapid changes on the Earth's surface (Potin et al., 2016). S1 data is widely used in a range of applications, including sea ice monitoring (Boulze et al., 2020), flood and earthquake response (C.-H. Lu et al., 2018), land subsidence tracking, deforestation analysis, and maritime safety (Hakim et al., 2020). The open availability of its data also supports a wide range of scientific research and operational services.

2.1.2 Sentinel-2

S2 consists of two identical satellites, Sentinel-2A (S2A) and Sentinel-2B (S2B), launched on June 23, 2015, and March 7, 2017, respectively. These satellites operate in a sun-synchronous orbit at a mean altitude of 786 *km* and are phased 180° apart, providing a revisit frequency of five days at the equator. Their orbit configuration, with a local solar time of 10:30 AM at the descending node, is optimized to balance solar illumination and reduce cloud interference (Spoto et al., 2012). Equipped with the MultiSpectral Instrument (MSI), S2 captures data in 13 spectral bands, ranging from visible and near-infrared to shortwave infrared (Table 2.1). The spatial resolution varies with four bands at 10 meters, six bands at 20 meters, and three bands at 60 meters, facilitating a wide range of applications (Drusch et al., 2012).

S2 data supports critical environmental monitoring tasks such as agriculture, forestry (Persson et al., 2018), land use (Phiri et al., 2020), and emergency management (Caballero et al., 2019), enabling assessments of crop health (Ghosh et al., 2018), deforestation monitoring (Torres et al., 2021), land cover mapping, and disaster response (Malinowski et al., 2020). The alignment of S2's overpass time with those of Landsat and SPOT-5 permits the integration of its data with existing and historical missions, contributing to long-term time series data collection. The openly accessible data from S2 fosters extensive scientific research and practical applications globally, making it an invaluable asset for sustainable development.

Table 2.1: Sentinel-2 bands Spectral and Spatial resolutions

Band	Bandwidth (nm)	Description	Spatial Resolution	Application
B01	433 – 453	Ultra Blue (Coastal and Aerosol)	60 m	Atmospheric correction, coastal and aerosol studies
B02	457.5 – 522.5	Blue	10 m	Soil and vegetation discrimination
B03	542.5 – 577.5	Green		Vegetation monitoring, water bodies
B04	650 – 680	Red		Chlorophyll absorption, vegetation monitoring
B05	698 – 712	Red Edge	20 m	Vegetation stress, chlorophyll content
B06	733 – 747			Vegetation structure, chlorophyll content
B07	774.5 – 791.5			Vegetation monitoring, biomass estimation
B08	789.5 – 894.5	Near Infrared (NIR)	10 m	Biomass, vegetation structure
B8A	854.5 – 875.5	Narrow NIR	20 m	Water column penetration, vegetation
B09	930.5 – 949.5	Water Vapour	60 m	Water vapor detection
B10	1360.5 – 1389.5	SWIR – Cirrus		Cirrus cloud detection
B11	1565 – 1655	SWIR 1	20 m	Moisture content, soil and vegetation monitoring
B12	2103 – 2277	SWIR 2		Soil moisture, snow/cloud differentiation

2.1.3 Sentinel-3

Sentinel-3 (S3) consists of two identical satellites, Sentinel-3A and Sentinel-3B, launched in 2016 and 2018, respectively. They operate in a sun-synchronous orbit at an altitude of approximately 814 km, phased 140° apart to ensure global coverage with frequent revisits (Donlon et al., 2012). S3 is equipped with several advanced instruments that monitor both oceanographic and land parameters. One of the key instruments, the Ocean and Land Colour Instrument (OLCI), captures data in 21 spectral bands with a spatial resolution of 300 m over all surfaces (Table 2.2), designed to reduce sun-glint and improve the accuracy of observations. The Sea and Land Surface Temperature Radiometer (SLSTR), used for measuring sea and land surface temperatures, operates with two different spatial resolutions: 500 m for visible and shortwave infrared channels and 1 km for thermal infrared channels. This allows for highly accurate global measurements of sea surface temperatures, with an accuracy better than 0.3 $kelvin$ (Donlon et al., 2012).

The S3 combination of instruments provides essential data by measuring the energy reflected from the Earth's surface. This is used to monitor sea surface height (J. Yang et al., 2020), ocean color (Toming et al., 2017), land surface temperatures, and vegetation health (X. Hu et al., 2019). Its data is crucial for applications such as tracking sea level rise, monitoring marine biological productivity (Lapucci et al., 2023), detecting oil spills, and assessing land cover changes (Reyes-Muñoz et al., 2022). The availability of this data supports a wide range of scientific research and operational applications, contributing to long-term environmental monitoring and sustainable resource management.

2.2 Trade-offs and Limitations in Remote Sensing: The Need for Data Fusion

One of the enduring challenges in remote sensing is managing the trade-offs between spatial, spectral, and temporal resolution. Enhancing one of these dimensions often comes at the cost of another (Lillo-Saavedra & Gonzalo, 2006). For instance, increasing spatial resolution improves the detection of fine-scale features such as individual buildings or small agricultural plots, but it reduces swath coverage and may limit the sensor's spectral richness, which is essential for distinguishing surface materials like vegetation, soil, and water. Likewise, improving temporal resolution such as with S1, which revisits the same location every six days supports near real-time monitoring but typically comes with limited spectral capabilities due to its radar-based design (Y. Zhang & Jiang, 2014). In contrast, optical sensors offer richer spectral data but are affected by atmospheric conditions and revisit intervals. These trade-offs highlight the complexity of sensor selection for time series analysis, where spatial, spectral, and temporal requirements must be carefully balanced according to the objectives of the study.

In addition to these trade-offs, remote sensing data acquisition faces limitations caused by environmental and technical factors, particularly for optical sensors. Atmospheric conditions, such as cloud cover, haze, and aerosols, play a significant role in altering the quality and accuracy of the data collected (Prudente et al., 2020). Because of its reliance on sunlight as

Table 2.2: Sentinel-3 OLCI spectral bands

Band	Bandwidth (nm)	Spatial Resolution	Application
Oa1	392.5 - 407.5	300 m	Aerosol correction, improved water constituent retrieval
Oa2	407.5 - 417.5		Yellow substance and detrital pigments (Turbidity)
Oa3	437.5 - 447.5		Chl absorption max biogeochemistry, vegetation
Oa4	437 - 447		High Chl, other pigments
Oa5	505 - 515		Chl, sediment, turbidity, red tide
Oa6	555 - 565		Chlorophyll reference (Chl minimum)
Oa7	615 - 625		Sediment loading
Oa8	660 - 670		Chl (2nd Chl abs. max.), sediment, yellow substance/vegetation
Oa9	670 - 677.5		For improved fluorescence retrieval and to better account for smile together with the bands 665 and 680 nm
Oa10	677.5 - 685		Chl fluorescence peak, red edge
Oa11	703.7 - 713.7		Chl fluorescence baseline, red edge transition
Oa12	716.2 - 791.2		O2 absorption/clouds, vegetation
Oa13	760 - 762.5		O2 absorption band/aerosol correction
Oa14	762.5 - 766.2		Atmospheric correction
Oa15	766.2 - 768.7		O2A used for cloud top pressure, fluorescence over land
Oa16	771.2 - 786.25		Atmospheric correction/aerosol correction
Oa17	855 - 875		Atmospheric correction, aerosol correction, clouds, pixel co-registration
Oa18	880 - 890		Water vapour absorption reference band. Common reference band with SLST instrument. Vegetation monitoring
Oa19	895 - 905		Water vapour absorption/vegetation monitoring
Oa20	930 - 950		Water vapour absorption, atmospheric/aerosol correction
Oa21	1000 - 1040		Atmospheric/aerosol correction

their primary source, they are susceptible to disruptions caused by cloud cover, which can block or distort the electromagnetic signals traveling between the Earth's surface and the sensor. This limitation is particularly severe in regions with frequent cloud cover as high elevated locations and mountainous areas, where optical sensors like S2 may miss critical data, creating gaps in time series monitoring. Moreover, clouds, haze, and fine particulate matter can reduce image contrast and clarity, making it difficult to extract detailed information from the imagery.

Atmospheric interference also affects the spectral accuracy of optical sensors. Scattering, absorption, and atmospheric turbulence can distort the reflected signals received by the sensor,

introducing errors in surface reflectance measurements. To address these issues, atmospheric correction processes among other preprocessing techniques are applied to optical imagery to remove or minimize the effects of atmospheric interference (Sola et al., 2018). Techniques like radiative transfer models and calibration are used to correct for the impact of atmospheric particles and gases, but these processes are complex and may not fully eliminate the distortions in every case.

The limitations of individual sensors in remote sensing, particularly the trade-offs between spatial, spectral, and temporal resolutions, emphasize the need for integrating data from multiple sources. These trade-offs, along with challenges posed by atmospheric conditions, make it difficult for any single sensor to meet the diverse requirements of different applications. To address these limitations, data fusion has emerged as a key technique that combines information from various sensors, each with distinct capabilities, to improve the overall quality and reliability of the data. Data fusion allows for the integration of complementary datasets from sensors with different strengths, such as higher spatial resolution from one sensor and better temporal or spectral coverage from another (Schmitt & Zhu, 2016; J. Zhang, 2010). This approach enhances the completeness and accuracy of the information collected, making it more suitable for complex applications that require detailed and continuous monitoring of Earth's surface.

2.3 Image Fusion

As mentioned before, data fusion is an essential solution in remote sensing, helping to overcome the limitations of individual sensors. These techniques can be broadly categorized based on the types of sensors involved and the specific resolutions they aim to enhance. A primary distinction is made between unimodal fusion and multimodal fusion. Unimodal fusion, also known as optical sensor fusion, involves the integration of data from multiple optical sensors. This approach can be further subdivided into three types, each targeting specific resolution improvements. Spatiospectral fusion enhances spectral information while preserving spatial details, facilitating the differentiation of various surface materials or vegetation types. Spatiotemporal fusion addresses differences in spatial and temporal resolutions to generate datasets with both high spatial detail and frequent temporal updates. An example is the fusion of data from sensors like Landsat and MODIS to create a time series with the spatial resolution of Landsat and the temporal resolution of MODIS. Spatio-temporal-spectral fusion integrates spatial, temporal, and spectral dimensions, providing detailed, frequent, and spectrally rich data essential for complex environmental monitoring tasks (Ghamisi et al., 2019). In contrast, multimodal fusion involves combining data from different types of sensors, such as integrating optical imagery with radar data. This approach leverages the complementary strengths of each sensor type to enrich the overall information quality (Y. Wang et al., 2024).

2.3.1 Optical Sensor Fusion

Optical sensor fusion approach allow enhancing spatial, spectral, and temporal resolutions, leading to more detailed and accurate data regarding one or more resolutions. As shown in Figure 2.1 the main categories of optical sensor fusion can be grouped based on the specific

technique used. This section presents the methodologies and advantages of optical sensor fusion, highlighting its critical role in advancing remote sensing capabilities.

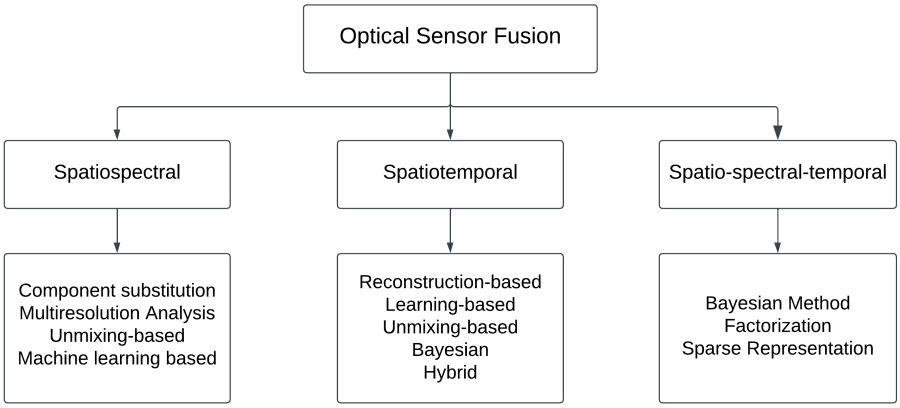


Figure 2.1: Categories of optical sensor fusion approaches

Spatiospectral Fusion

Spatiospectral fusion methods have been extensively explored in remote sensing to integrate the low spatial resolution with high spectral resolution (LSHS) with high spatial resolution and low spectral resolution (HSLS) to generate high spatial resolution and high spectral resolution (HSHS) (H. Song et al., 2014). These methods aim to preserve both spectral fidelity and spatial details, enhancing the quality of remote sensing data for various applications. Figure 2.2 illustrates the principle of spatiospectral fusion. The approaches can be categorized into four main groups: Component Substitution (CS) methods, Multiresolution Analysis (MRA) methods, Unmixing-based methods, and Machine Learning-based methods as previously mentioned in Figure 2.1. Each of these categories employs distinct strategies for fusion, with varying trade-offs in computational complexity, spectral preservation, and spatial enhancement.

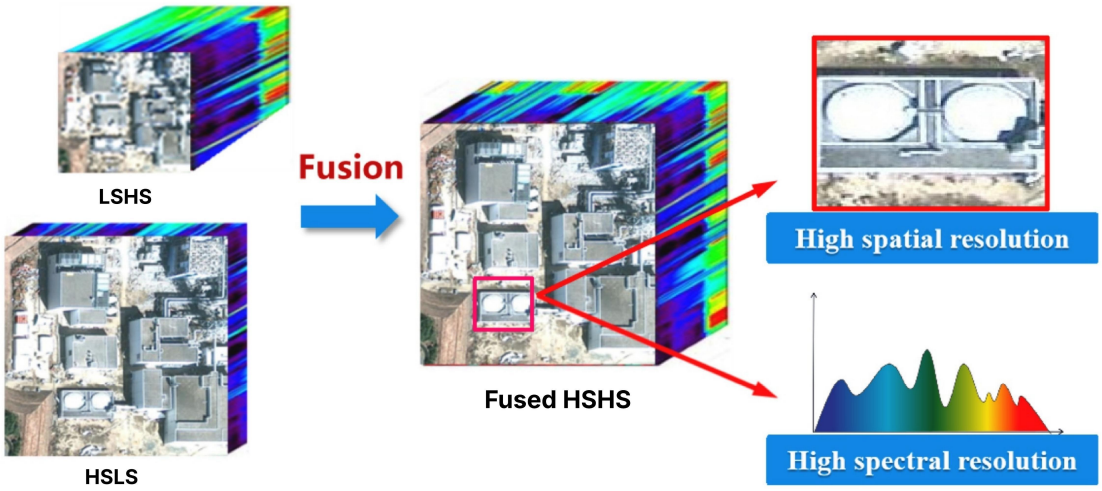


Figure 2.2: The input and output of spatiospectral data fusion. Inspired by (Dian et al., 2021)

CS methods operate by first upsampling LSHS to match the spatial resolution of a HSLS and then transforming the spectral components to separate spatial and spectral information (L.-J. Deng et al., 2019; J. Ma et al., 2020; Meng et al., 2018). The high-resolution spatial details from the HSLS image are then injected into the transformed data, followed by an inverse transformation to reconstruct the fused image. Popular CS techniques include Intensity-Hue-Saturation (IHS) transformation (Carper et al., 1990), Principal Component Analysis (PCA) (Kwarteng & Chavez, 1989), and Gram–Schmidt (GS) transformation (Laben & Brower, 2000). These methods are computationally efficient and widely used but often introduce spectral distortions, particularly in regions with significant spectral variability.

Multiresolution Analysis (MRA) methods employ decomposition techniques to extract spatial details at multiple scales from HSLS image and integrate them into the LSHS image, while maintaining spectral consistency. These methods rely on mathematical transformations such as Wavelet transform (Mallat, 1989), Laplacian pyramid (LP) decomposition (Burt & Adelson, 1987), and Curvelet transform, which decompose images into different frequency components (Starck et al., 2007). High-frequency spatial details extracted from the HSLS image are then injected into the LSHS to enhance spatial resolution. Authors in (Otazu et al., 2005) used Spectral Response Function (SRF) to calculate the amount of information to be injected. MRA methods are advantageous for preserving spectral information, but may struggle with artifacts in regions with complex textures or mixed spectral responses.

Unmixing-based methods approach the fusion problem by modeling multispectral pixels as a mixture of pure spectral signatures (endmembers). These methods use spectral unmixing techniques, such as sparse matrix factorization (SMF) and low-rank matrix factorization, to estimate the contributions of different endmembers to each pixel (Kawakami et al., 2011). Techniques like Vertex Component Analysis (VCA) (Nascimento & Dias, 2005) and Maximum a Posteriori (MAP) estimation help extract spectral information from low-resolution HSI and use it to reconstruct high-resolution data (Simoes et al., 2014). These methods are particularly effective in applications requiring spectral consistency but may be computationally expensive when dealing with large datasets.

Machine Learning-based methods, including deep learning approaches, leverage data-driven models to learn spatial and spectral features directly from training datasets. Convolutional Neural Networks (CNNs) have gained popularity in LSHS and HSLS fusion due to their ability to capture complex spatial patterns and spectral relationships. Approaches such as PanNet architecture (J. Yang et al., 2017), multi-scale CNNs (Han et al., 2019), and two-branch CNNs separate spatial and spectral learning pathways to optimize fusion quality (J. Yang et al., 2018). Although machine learning-based methods achieve state-of-the-art results, they require large amounts of training data and substantial computational resources, limiting their applicability in cases with limited labeled datasets.

Even though spatiospectral fusion methods have contributed significantly to the optical sensor fusion field, they were not the focus of this thesis, as they typically require the simultaneous acquisition of LSHS and HSLS data to improve spectral quality. This constraint makes them less suitable for applications requiring the generation of high quality time series.

Spatiotemporal Fusion

Spatiotemporal fusion (STF) is a remote sensing technique that aims to generate imagery with both high spatial and high temporal resolution (HSHT) by integrating data from multiple sensors with complementary characteristics. Typically, one sensor provides high spatial but low temporal resolution (HSLT), such as S2 or Landsat, while another offers low spatial but high temporal resolution (LSHT), such as MODIS or S3. By leveraging the strengths of each, STF addresses the limitations of individual sensors that cannot capture fine spatial details frequently over time (Zhu et al., 2018). Figure 2.3 illustrates the principle of STF. The central principle of STF is to estimate fine resolution observations at unobserved dates by learning patterns from observed combinations of coarse and fine resolution images. This enables the construction of continuous and high resolution time series critical for monitoring dynamic processes such as vegetation growth (Son et al., 2016), land use changes (Senf et al., 2015), and natural disasters (F. Zhang et al., 2014).

A key requirement for effective STF is that the spectral bands used from different sensors must correspond to the same or closely aligned spectral ranges. Attempting to fuse bands with fundamentally different spectral intervals would result in physically inconsistent outputs and introduce spectral distortions in the fused imagery. Therefore, proper spectral band selection is a necessary preprocessing step before any fusion can be reliably performed.

Over the past two decades, a wide range of STF methods have been developed, which can be broadly categorized into weight-based (also known as reconstruction-based), unmixing-based, hybrid, and learning-based approaches. Each category differs in its theoretical assumptions, algorithmic structure. The remainder of this section introduces these categories, describing how each performs the fusion, their respective advantages and limitations, and the rationale for focusing on specific methods in this thesis.

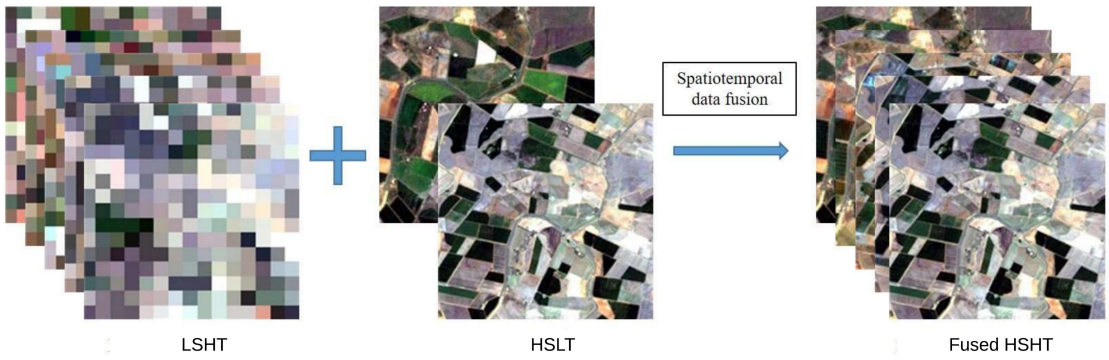


Figure 2.3: The input and output of spatiotemporal data fusion. Inspired by (Zhu et al., 2018)

Unmixing-based methods rely on the principle that coarse-resolution pixels are composed of mixtures of finer land cover components. These models typically classify fine-resolution images to identify land cover types, compute their proportions within coarse pixels, and use this information to disaggregate or "unmix" the coarse pixels into fine-scale predictions. Notable examples include the Multisensor Multiresolution Technique (MMT) (Zhukov et al., 1999), STDFA (M. Wu et al., 2012), and RSpatialU (Y. Xu et al., 2015). While effective in certain applications, such as land cover fraction estimation, these methods require accurate and stable

classification maps, and often struggle in complex or dynamic landscapes due to intra-class spectral variability and rigid assumptions about class consistency. These limitations make them less suitable for high-resolution time series generation in heterogeneous and cloud-prone regions.

Bayesian-based methods approach the fusion from a probabilistic perspective, modeling uncertainty and variability in the relationships between coarse and fine-resolution data. These methods often rely on prior knowledge, such as multi-temporal land cover maps or historical reflectance trends, to constrain the fusion process. For example, the STRUM model (Gevaert & García-Haro, 2015) integrates Bayesian-constrained unmixing with weighted functions to reconstruct fine-resolution reflectance from coarse inputs. Similarly, the STIMFM model (X. Li et al., 2017) combines Bayesian theory with linear mixing to fuse coarse images and high-resolution land cover maps. The strength of Bayesian approaches lies in their ability to incorporate prior distributions and handle uncertainty systematically. However, they depend heavily on the quality and availability of auxiliary data such as prior land cover or class-specific reflectance statistics. Their performance also degrades when the underlying assumptions about class stability or prior distributions are violated, making them less suitable for dynamic environments with frequent land cover changes or limited reference information.

Machine learning (ML) and deep learning (DL) methods have emerged as powerful alternatives in recent years due to their ability to model complex, nonlinear relationships directly from the data. These data-driven approaches use techniques such as random forests (Hutengs & Vohland, 2016), CNNs (H. Song et al., 2018), generative adversarial networks (GANs) (H. Zhang et al., 2020), and transformers (W. Li et al., 2021) to learn mappings between coarse and fine-resolution observations. By bypassing handcrafted rules and weights, ML/DL models can capture subtle spatial and temporal patterns across diverse landscapes. However, this flexibility comes at a cost. These methods typically require large volumes of well-aligned training data and are computationally intensive, making them less practical in scenarios with limited data availability or high cloud contamination. Additionally, their black-box nature often reduces interpretability, which can be a drawback in applications where explainability and physical consistency are important.

While those categories offer valuable contributions to the field, their limitations in terms of data requirements, computational cost, and limited interpretability make them less suited for the goals and constraints of this thesis. An additional and critical constraint was the availability of openly published and reproducible implementations. Many recent state-of-the-art models lack publicly accessible code, making them difficult to adapt, validate, or compare fairly within a unified framework. Given these considerations, we focuse on two well-established and accessible categories: Weight-Based (Reconstruction-Based) and Hybrid methods. These approaches offer a balance between accuracy, flexibility, and reproducibility. They are also supported by published implementations or are straightforward to re-implement based on existing literature. The following provide a detailed overview of these two categories, outlining their methodological foundations, representative models, and the rationale for their use in this thesis.

Reconstructed-based also known as Weight-based spatiotemporal fusion methods estimate fine resolution pixel values by combining information from multiple input images using a

weighted function. These methods have gained popularity due to their flexibility and ease of implementation. The first and most widely known model in this category is the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) (F. Gao et al., 2006). It is designed to fuse HSLT (such as Landsat) with LSHT imagery (such as MODIS). The core assumption of STARFM is that the surface reflectance observed by both sensors at the same location is similar, apart from a systematic error that varies with pixel characteristics. This error, which remains consistent over short periods, can be calculated from a base pair of Landsat and MODIS images acquired on the same date t_1. Once the error is determined, it can be applied to predict Landsat-like imagery for other dates using MODIS data t_2. The algorithm works by reprojecting MODIS data to match the Landsat grid, then using a moving window to identify spectrally similar pixels in the Landsat image. A weight for each pixel is calculated based on three factors: (i) spectral difference between the Landsat-MODIS pair; (ii) temporal difference between MODIS images; and (iii) spatial distance between neighboring pixels. The surface reflectance of the central pixel is predicted by combining weighted information from both datasets. The algorithm requires at least one pair of fine resolution (Landsat) and coarse resolution (MODIS) data on a prior or posterior date and one coarse resolution image for the prediction date. While STARFM works well in homogeneous landscapes, it faces limitations in areas with mixed land cover (Zhu et al., 2010). To address this, (Zhu et al., 2010) introduced ESTARFM (Enhanced STARFM) to improve performance in heterogeneous environments. Unlike STARFM, ESTARFM assumes the changes are proportional, introducing a conversion coefficient to better adjust for differences between fine and coarse resolution reflectances. This refinement allows for greater accuracy, particularly in mixed-pixel environments, where land cover types vary significantly. STAARCH (Spatial Temporal Adaptive Algorithm for mapping Reflectance Change) further enhances this approach by detecting land cover change points from dense time series coarse images, making it suitable for dynamic landscapes where land-use changes are frequent (Hilker et al., 2009). Additional improvements have been made through the incorporation of more advanced techniques. RWSTFM (Rigorously-Weighted Spatiotemporal Fusion Model) applies geostatistical techniques such as ordinary kriging to calculate weights for neighboring pixels, improving prediction accuracy through spatial autocorrelation (J. Wang & Huang, 2017).

To deal with strong seasonal changes, the Fit-FC (Q. Wang & Atkinson, 2018) conducted a fusion between S2 and S3 images handling seasonal and phenological variability in vegetation and agricultural monitoring. The method is based on three main steps and requires one pair of fine and coarse resolution images at the base date and one coarse resolution image on the target date. First, a local regression model (RM) is fitted between the coarse images at two different times to predict the fine resolution image. Since this leads to blocky artifacts due to coarse regression, spatial filtering (SF) is applied next, smoothing artifacts by using spectrally similar neighboring pixels ensuring that spatially close pixels with similar spectral properties contribute more to the final prediction. The final step is residual compensation (RC), which adjusts the prediction by downscaling the residual errors between the predicted and observed coarse images using bicubic interpolation. These residuals are refined using spectrally similar pixels to preserve spatial and spectral details. These updated residuals are added back to the filtered image, improving both the spatial and spectral accuracy of the final high resolution prediction. Other methods focused on improving model accuracy by accounting for the

physical properties of sensors and environmental factors. For example, the Semi-Physical Fusion Approach leverages MODIS bidirectional reflectance function (BRDF)/albedo products to predict Landsat reflectance, enhancing the ability to predict surface changes with greater precision (Roy et al., 2008).

Hybrid-based methods integrate multiple strategies, such as unmixing, Bayesian theory, and weight functions, to improve performance in complex environments. One example is the spatial and temporal reflectance unmixing model (STRUM), which applies Bayesian-constrained unmixing on the change map derived from coarse resolution images. After determining the reflectance changes for each class, STRUM uses weighted functions similar to STARFM to create a fused image using moving windows (Gevaert & García-Haro, 2015). An enhanced version of STRUM replaces the classification map with abundance images derived from prior fine resolution images, while also introducing sensor difference adjustments to further improve accuracy (J. Ma et al., 2018). Another hybrid method improves STARFM by incorporating unmixing-based strategies to first generate abundance images from coarse-resolution data, which are then fed into the STARFM algorithm to improve prediction accuracy (Xie et al., 2016).

The FSDAF (Flexible Spatiotemporal Data Fusion) model, developed by (Zhu et al., 2016), offers a more flexible approach for high resolution image prediction, especially in regions with complex and varying land cover. FSDAF is a prominent hybrid approach that combines linear unmixing, spatial interpolation, and weighted function-based methods. It requires one pair of fine and coarse resolution images at the base date, a coarse resolution image at the target date, and a land cover map. The model estimates temporal changes in land cover using unmixing and generates spatial predictions using thin plate spline (TPS) interpolation. By capturing both temporal changes and local spatial variability, FSDAF improves prediction accuracy. The residuals between temporal and spatial predictions are distributed to fine resolution pixels using TPS, and the final high resolution image at target date is produced by combining these residuals, weighted by both spectral and spatial information. This method excels at preserving spatial details while adapting to land cover changes over time, making it particularly useful in heterogeneous landscapes. An improved version of FSDAF introduces a constrained least-squares process to combine the increments from unmixing and coarse image interpolation, enhancing the model's ability to handle complex landscapes (M. Liu et al., 2019).

A range of other hybrid models integrate different STF approaches to achieve better results. For instance, the STARFM and unmixing-based model (USTARFM) combines STARFM with unmixing techniques to refine predictions by handling mixed pixels more effectively (Xie et al., 2016). The Spatial-Temporal Remotely Sensed Images and Land Cover Maps Fusion Model (STIMFM) integrates the spectral linear mixing model and Bayesian framework, allowing it to fuse multi-temporal coarse images and fine-resolution land cover maps, producing a series of fine-resolution land cover maps (X. Li et al., 2017).In addition, the enhanced linear STF method characterizes the residuals caused by systematic biases in the linear model, refining the slope and intercept using spectral unmixing theory, followed by a weight-based strategy to improve the final prediction (Ping et al., 2018). Hybrid methods have also been applied in various domains, such as temperature monitoring. The BLEnd Spatiotemporal Temperature

(BLEST) model adapts the FSDAF framework for land surface temperature (LST) fusion by blending data from sources with different spatial resolutions (Quan et al., 2018).

In summary, STF enables the generation of high-resolution time series by leveraging the complementary strengths of multiple sensors. A fundamental consideration in STF is the selection of spectrally compatible bands. Fusion must only be performed between bands from different sensors that lie within the same or closely aligned spectral ranges; attempting to combine non-corresponding bands from different parts of the electromagnetic spectrum leads to physically invalid results. As such, STF is limited to the spectral bands that are common to both the HSLT and LSHT sensors.

Spatio-temporal-spectral Fusion

Most fusion methods in remote sensing have been developed independently and target a single reconstruction task, focusing on either spatial, temporal, or spectral fusion. There is, however, growing recognition of the need for a unified framework that can simultaneously leverage complementary information across all dimensions to improve data completeness. (Ng et al., 2017) took an early step in this direction by proposing an adaptive weighted low-rank tensor model (AWTC) for reconstructing remote sensing images with missing data. This method capitalizes on spatial, spectral, and temporal information to create a more holistic model for recovery. Similarly, (X. Li et al., 2016) presented a spatial-spectral-temporal approach using group sparse representation, which extended single-patch sparse representation methods to encompass multiple patches, effectively capturing correlations across both local and nonlocal spatial regions. Despite progress, integrated frameworks that address spatial, temporal, and spectral data from multiple sensors remain relatively rare, with most fusion techniques limited to just two of these dimensions or only one or two sensors. Authors in (Shen, 2012) were among the first to propose an integrated spatio-temporal-spectral fusion framework for multisensor remote sensing, although this approach was validated only on simulated data. (Huang et al., 2013) further extended this work, exploring relationships between spatio-spectral and spatio-temporal fusion approaches; while this method was tested on real data, it was still restricted to only two sensors. But these approaches did not fully incorporate all three fusion dimensions, spatial, temporal, and spectral simultaneously.

2.3.2 Multimodal fusion

Multimodal fusion in remote sensing refers to the integration of data acquired from different types of sensors such as optical, SAR, LiDAR, and thermal imaging etc, to exploit their complementary strengths. Each sensor modality captures different physical properties of the Earth's surface: optical sensors provide rich spectral information under cloud-free conditions; SAR offers structural details independent of weather and lighting; LiDAR captures precise elevation and 3D shape information; and thermal sensors measure surface temperature variations. By combining these sources, multimodal fusion enhances the quality, robustness, and reliability of remote sensing applications.

Fusion can occur at different processing levels, typically categorized as pixel-level, feature-level, and decision-level fusion. Pixel-level fusion directly combines raw or preprocessed

data from different modalities, often through mathematical operations or alignment-based reconstruction. Feature-level fusion involves extracting and integrating representative features (e.g., texture, gradients, shape) from each modality, making it particularly effective in classification or segmentation tasks. Decision-level fusion merges the outputs of separate classifiers or models to reach a consensus, typically used in scenarios where sensor-specific models are first independently applied (Shen et al., 2019).

These fusion strategies have been widely applied across various remote sensing domains, including land cover classification, change detection, biomass estimation, environmental monitoring, and disaster response. The ability to combine multi-sensor information allows researchers and practitioners to overcome the limitations of individual data sources such as cloud cover in optical imagery, or the lack of spectral detail in SAR, thus enabling more comprehensive and reliable Earth observation.

A wide variety of sensor combinations have been explored in remote sensing, each offering unique advantages for specific applications. LiDAR–optical fusion has been extensively used in land use and land cover (LULC) classification, urban mapping, and forest structure analysis (J. Zhang & Lin, 2017). Optical imagery provides detailed spectral signatures useful for distinguishing surface materials, while LiDAR offers precise elevation and 3D structural information, enhancing the discrimination of vegetation layers, buildings, and bare soil. This synergy has proven particularly effective in complex environments such as dense forests and urban areas (Y. Chen et al., 2017; Ghamisi et al., 2016). LiDAR–hyperspectral fusion further expands classification capabilities, particularly in vegetation species identification, ecosystem monitoring, and infrastructure assessment. The fine spectral resolution of hyperspectral imagery captures subtle biochemical and biophysical variations, while LiDAR contributes detailed canopy and terrain structure. This integration supports applications in forest biomass estimation, biodiversity mapping, and urban planning (H. Li et al., 2018; X. Wang et al., 2025). In environmental monitoring and precision agriculture, the combination of thermal and optical data has enabled improved assessments of evapotranspiration, crop water stress, and soil moisture (Comba et al., 2019). Thermal sensors capture land surface temperature variability, while optical imagery contributes spectral indices that are sensitive to vegetation health. This fusion supports smart irrigation systems, yield forecasting, and sustainable agricultural practices. Change detection has also benefited from multimodal integration, particularly by combining SAR, LiDAR, and optical data to identify both spectral and structural changes. These approaches have been widely used for disaster response, land surface monitoring, and urban expansion mapping (X. Li et al., 2021; Shangguan et al., 2024). Incorporating multiple data types improves temporal consistency and reduces false detections due to sensor-specific artifacts or illumination variability. Recent developments in deep learning architectures, such as graph-based networks and decision-level fusion modules, have further improved the robustness of change detection models (Zhao et al., 2020). Multimodal fusion also plays a central role in terrain modeling and biomass estimation. For instance, combining LiDAR and SAR improves digital elevation models (DEMs), especially in regions where canopy cover or complex topography limits single-sensor accuracy (Kahraman & Bacher, 2021). In biomass modeling, structural data from LiDAR or SAR can be complemented by spectral reflectance information from optical or hyperspectral sensors, enabling more accurate estimations across forest and agricultural systems (Alonzo et al., 2014).

Among the various multimodal combinations, the fusion of optical and SAR data has received significant attention due to the complementary nature of the two modalities. Optical sensors offer detailed spectral and radiometric information but are limited by cloud cover and illumination conditions. In contrast, SAR sensors provide structural and backscatter data that are unaffected by weather or daylight, making them highly reliable in regions with frequent atmospheric interference. This complementary relationship enables optical–SAR fusion to support a wide range of remote sensing tasks that demand consistent, high-resolution observations.

Applications of this fusion strategy span land use and land cover classification, flood mapping, change detection, and disaster response. For instance, optical–SAR fusion has been shown to improve flood detection accuracy during the 2009–2010 Data Fusion Contest, outperforming single-modality approaches in dynamic environments (Debes et al., 2014). More recent developments have integrated optical–SAR data into deep learning architectures for infrastructure monitoring, urban growth analysis, and environmental assessment (Shangguan et al., 2024; J. Wang et al., 2022).

A particularly important and challenging application of optical–SAR fusion is cloud removal, the reconstruction of cloud-contaminated optical images using structural and temporal cues from SAR. This task is critical for enabling the generation of cloud-free, temporally continuous time series in areas where persistent cloud cover hinders the usability of optical imagery. The following section explores the state of the art in SAR-optical cloud removal methods and presents the contributions made in this thesis to advance this area.

One of the most critical and actively explored applications of multimodal fusion is the cloud removal in optical satellite imagery. Clouds obstruct approximately 55% of Earth's land surface at any given time (Bar-Or et al., 2011), severely limiting the availability of cloud-free observations required for time series analysis, environmental monitoring, land use classification, and precision agriculture. By leveraging the complementary nature of SAR and optical sensors, these methods aim to reconstruct cloud-free optical imagery using the structural and temporal consistency of SAR data. Unlike temporal gap-filling or traditional inpainting approaches, SAR–optical fusion directly integrates physically meaningful observations, enabling more robust reconstruction in persistently cloudy or dynamically changing environments.

Early approaches attempted direct translation from SAR to optical using CNN-based models (Bermudez et al., 2018; Fuentes Reyes et al., 2019). However, due to the fundamental differences in the spectral signatures of SAR and optical modalities, these methods often failed to capture the fine spectral details necessary for high-fidelity reconstruction, particularly in areas with diverse land cover. To address these limitations, more advanced multimodal deep learning architectures have been developed, combining SAR and optical imagery to exploit their complementary information. One notable example is FusionNet, proposed by (J. Hu et al., 2017), which introduced a two-branch CNN architecture to extract distinct features from hyperspectral and SAR inputs before merging them. This design allowed the model to learn modality-specific representations while leveraging their joint spatial structure. Later, (X. Liu et al., 2020) improved the generalizability of this framework by introducing sparse constraints on batch normalization layers, which helped reduce feature redundancy and improved model robustness across varying landscapes.

Another significant advancement is DSen2-CR, introduced by (Meraner et al., 2020), which employs a deep residual network trained on real S1 and S2 imagery. The model integrates SAR as auxiliary input to guide the reconstruction of cloud-covered areas in S2 images. A key innovation in DSen2-CR is its cloud-adaptive loss function, designed to prioritize the restoration of cloud-occluded regions while minimizing distortion in cloud-free areas. This loss function dynamically adjusts the learning objective based on cloud presence, enabling the model to handle both thin and thick clouds effectively. A related line of work, authors in (Cresson et al., 2022) proposed a model which uses a U-Net architecture (Ronneberger, 2015) in place of residual networks. U-Net's encoder–decoder structure with skip connections enables efficient multi-scale feature extraction while preserving spatial resolution. This architectural shift reduces computational cost and memory usage compared to fully convolutional designs. Furthermore, an extended version of the model incorporates DEM data, allowing the network to better distinguish between terrain features, cloud shadows, and true cloud cover particularly beneficial in topographically complex areas. The inclusion of DEM as an additional modality enhances the model's capacity to resolve ambiguities related to elevation-induced illumination differences.

Beyond CNNs, recent efforts have explored generative approaches, particularly GANs for cloud removal tasks. These models are designed to learn high-dimensional mappings from cloudy to cloud-free images, often producing visually coherent and perceptually realistic results. One example is the method proposed by (Grohnfeldt et al., 2018), which applies a conditional GAN (cGAN) to generate cloud-free S2 imagery from SAR–optical input pairs. The model is trained with unpaired data, enhancing its adaptability and reducing dependency on perfectly aligned image sets.

Building on this direction, (J. Gao et al., 2020) proposed a two-stage GAN pipeline. In the first stage, a simulated optical image is generated from the SAR input. In the second stage, this simulated image is concatenated with the original cloudy optical and SAR images to reconstruct the final cloud-free result. This approach improves spectral consistency and contextual integration by incorporating both structural and radiometric priors in the reconstruction process.

Further improvements were introduced in AMGAN (M. Xu et al., 2022), a GAN-based model enhanced with attention mechanisms. AMGAN incorporates an attentive recurrent network to identify cloud-covered regions and extract cloud-specific features, followed by an attentive residual network that selectively removes clouds using these attention maps. The final reconstruction is performed by a dedicated decoder network. This modular architecture allows the model to focus on cloud-affected areas while preserving details in unaffected regions, resulting in high-quality outputs even under dense cloud cover. Attention mechanisms also help mitigate artifacts commonly seen in generative models, especially along cloud boundaries.

As SAR–optical fusion techniques continue to evolve, these deep learning approaches highlight the field's growing sophistication in handling diverse cloud conditions. The integration of auxiliary data such as DEM, the use of attention for feature selection, and the shift toward generative paradigms all contribute to improving the accuracy, realism, and adaptability of cloud removal models in complex and dynamic landscapes.

2.4 Gaps in Research

As outlined in the literature review, achieving dense time series with high spatial, spectral, and temporal resolution is essential for accurately monitoring dynamic environmental processes. High-quality time series datasets provide the level of detail needed to track subtle transitions and rapid changes over time, offering invaluable insights for applications in ecology, agriculture, urban planning, and climate science. However, producing such dense datasets poses significant challenges, as individual satellite sensors typically excel in only one resolution dimension—spatial, spectral, or temporal—but rarely all three. STF methods have emerged as key solutions to address this limitation by producing enhanced time series imagery that retains both fine spatial detail and frequent temporal sampling. In doing so, STF extends the capacity of remote sensing to monitor surface dynamics more comprehensively than is possible with individual sensors alone.

Despite its utility, a key limitation of STF techniques lies in their assumption that input images from different sensors possess identical spectral characteristics (F. Gao et al., 2006; Q. Wang & Atkinson, 2018). Each sensor has unique spectral resolution, often leading to spectral mismatches when their data are fused. This assumption of identical spectral profiles can introduce distortions, especially in heterogeneous landscapes, where accurate spectral information is crucial to distinguish different types of land cover. While some STF methods attempt to address these spectral discrepancies, they generally do so by adjusting a single spectral band, which does not fully resolve the spectral inconsistencies across the entire dataset (Cao et al., 2020; J. Li et al., 2021; S. Liu et al., 2022). As a result, the quality of fused imagery can suffer, reducing its reliability for applications that depend on high spectral fidelity.

An additional limitation of STF methods is their reliance on cloud-free images to maintain temporal continuity and data clarity. In practice, cloud-free imagery is not consistently available, especially in regions with frequent cloud cover, such as mountainous or tropical areas. Persistent clouds disrupt the quality and consistency of the time series, complicating efforts to track changes accurately.

This dependency on cloud-free images highlights the need for effective cloud removal techniques to complement STF processes by providing clear imagery even in cloud-prone regions. When cloud removal techniques are applied in conjunction with STF, they ensure that dense, high resolution time series data remain continuous and reliable. However, existing cloud removal techniques still face limitations, particularly in complex terrains like mountainous areas where cloud cover is persistent and interacts dynamically with topography. Traditional cloud removal methods, including multitemporal and multispectral approaches, often struggle to distinguish between clouds, snow, and shadows in these regions, leading to errors in image reconstruction (Lin et al., 2012; M. Xu et al., 2019). The lack of integrated topographic information further limits the ability of these models to effectively handle cloud interference in high-elevation areas where elevation plays a crucial role in image quality (Immerzeel et al., 2020; R. Wu et al., 2023).

Addressing these challenges, the research in Chapters 3 and 4 introduces methods to tackle the specific limitations in STF and cloud removal techniques. Chapter 3 focuses on enhancing

spectral consistency across multiple bands in STF, bridging the gap left by existing fusion methods that prioritize spatial and temporal integration over spectral alignment. By systematically adjusting the spectral characteristics of S3 to match those of S2 data. This approach enables more accurate, high-quality time series that are suitable for applications requiring consistent spectral information. Chapter 4, in turn, advances cloud removal by incorporating topographical data, specifically designed for mountainous regions where existing models struggle with thick, persistent clouds. By leveraging SAR data and integrating DEM information to help distinguishes between clouds, snow, and shadows, even in conditions of extensive cloud cover. This topography-sensitive approach allows for continuous monitoring in challenging terrains, supporting high-quality time series data for environmental and hydrological applications.

Chapter 3

Spectral Adjustment for Spatiotemporal Fusion of Sentinel-2 and Sentinel-3

As seen in Chapter 2, most STF techniques operate under the assumption that multispectral input images have identical spectral characteristics, disregarding spectral discrepancies, and focused on resolving spatial and temporal differences. Spectral differences between multispectral optical sensors stem from variations in the number, width and spectral range of their spectral bands. Consequently, sensors with narrow spectral bands may have multiple overlaps with the broader bands of other sensors Despite these discrepancies, current STF methods that combine images from multiple sensors to improve spatial and temporal resolution typically overlook these spectral relationships. This oversight can lead to spectral distortions in fused images, especially in heterogeneous landscapes, where accurate spectral information is essential for effective land cover monitoring.

Some STF approaches do account for spectral differences, but inputs are often required to have the same number of bands, with fusion typically performed on a band-by-band basis. This approach presents challenges when dealing with sensors that have differing spectral configurations. Rather than utilizing all the available spectral information, these methods often select only a single band or a limited subset of bands for the fusion process (Xue et al., 2017). While this simplification reduces computational complexity, it limits the use of valuable spectral information potentially reducing the quality and accuracy of the fused products, as even minor overlaps in spectral bands can introduce inconsistencies that degrade image quality.

In the absence of standardized guidelines for handling spectral mismatches, STF methods are less effective in complex environments where precise spectral data are vital for detecting subtle changes in land cover.

To address spectral differences between sensors, various band adjustment techniques have been proposed. One such method is WiSpeR, a wavelet-based spatiospectral fusion technique developed by (Otazu et al., 2005) for pansharpening. WiSpeR considers SRF differences when

determining how much information from the panchromatic image should be injected into multispectral bands, thereby improving the spectral fidelity of fused images. Similarly,(Aiazzi et al., 2007) quantify SRF discrepancies by calculating linear regression coefficients between panchromatic and multispectral images, thereby enhancing the spectral quality of pansharpened images.

In the STF domain, recent studies have applied RMs models to account for variations in land cover and spectral bands. These methods calculate weights for methods such as STARFM, improving the selection of neighboring pixels that are spectrally similar and improving the fusion accuracy in diverse landscapes (Cao et al., 2020). Building on the STARFM method, this approach replaces the use of predefined weights with weights dynamically calculated based on spectral differences between sensors. Likewise, (J. Li et al., 2021) proposed a linear regression-based approach that derived fitting coefficients between the MODIS and Landsat bands, generating higher-resolution images with reduced spectral errors in fused images. Overall, statistical regression has become a popular approach for compensating for spectral differences, allowing for more precise cross-calibration across multispectral sensors.

Building on these methods, our research introduces a novel preprocessing step called Multispectral Temporal Band Adjustment (MSTBA) to directly address spectral inconsistencies between sensors. The MSTBA method leverages all available spectral information from overlapping bands to create adjusted bands that better align the spectral characteristics of the input images. By incorporating this adjustment process before applying STF methods, our approach enhances the spectral fidelity of multiespectral fused images, reducing distortions and improving their quality.

3.1 Theoretical Basis of MSTBA in STF

3.1.1 Understanding Sensor Characteristics

The sensor's SRF represents the probability that a given sensor will detect a photon at a given frequency (ν) reflecting its physical and design characteristics. In this work, two SRFs should be considered: the SRF corresponding to $HSLT$, designed as $R_h(\nu)$ and the SRF of the ith band of the $LSHT$, designed as $R_{l_i}(\nu)$, with $i = 1, 2, ...n$. n being the number of bands of the $LSHT$ sensor. The probability that a photon will be detected by the $HSLT$ sensor can be defined as the probability of the event h.

$$P(h) = \int R_h(\nu)\, d\nu \tag{3.1}$$

Similarly, the probability of the detection of a photon by the $LSHT$ sensor for a band i can be defined as the probability of the event l_i (equation 3.2).

$$P(l_i) = \int R_{l_i}(\nu)\, d\nu \tag{3.2}$$

The probability of events (h) and (l_i) can be geometrically understood as the area below their SRFs.

Taking into account n_h and n_{l_i} are the number of photons detected by $HSLT$ and $LSHT_i$, respectively, the total photons detected simultaneously by the $HSLT$ and $LSHT$ sensors n_{h,l_i} could be defined as:

$$n_p = \sum_i n_{h,l_i} \tag{3.3}$$

with

$$n_{h,l_i} = P(h \,|\, l_i) \cdot n_{l_i} \tag{3.4}$$

But equation 3.3 is only correct if all the area below $R_h(\nu)$ is contained within the $R_{l_i}(\nu)$. In terms of photons, this statement holds true when all the photons detected by the $HSLT$ are also detected by one of the $LSHT$ sensors, which is not the general case.

If the number of $HSLT$ photons that are simultaneously below both the $R_h(\nu)$ and $R_{l_i}(\nu)$ functions is known, we can express the previous equations in terms of these shared photons as follows:

$$P(l_i|h_m) = \frac{P(l_i \bigcap h_m)}{P(h_m)} \tag{3.5}$$

$$P(h_m|l_i) = \frac{P(h_m \bigcap l_i)}{P(l_i)} \tag{3.6}$$

h_m being the event corresponding to a $HSLT$ photon simultaneously below the $R_h(\nu)$ and $R_{l_i}(\nu)$.

Given the number n_h of photons detected by the $HSLT$ sensor, we can estimate the number n'_{l_i} of photons that the $LSHT$ sensor is expected to detect (Otazu et al., 2005). In the context of image fusion, equation 3.7 provides the number of $HSLT$ photons that contain spatial details. Specifically, it indicates the proportion of high-resolution spatial information captured by the $HSLT$ sensor that can be integrated with the coarse-resolution information from the $LSHT$ sensor:

$$n'_{l,i} = \alpha_p \frac{P(l_i \,|\, h_m)}{P(h_m \,|\, l_i)} n_h \tag{3.7}$$

with

$$\alpha_p = \frac{\int \min(R_h, \max(R_{l_1}, R_{l_2}, \ldots, R_{l_n}))d\nu}{\int R_h(\nu)d\nu} \tag{3.8}$$

The relationship between the digital number (DN) value and the top-of-atmosphere (TOA) reflectance is defined as follows (Chander et al., 2009):

$$DN = \frac{\rho \cdot E_s \cdot \cos\theta}{\text{Gain} \cdot d^2 \cdot \pi} \tag{3.9}$$

where ρ represents the TOA reflectance, E_s is the solar irradiance at the TOA, θ is the solar zenith angle, Gain is the band-specific rescaling gain factor, and d is the solar-Earth distance.

The spectral irradiance at the sensor aperture is averaged over the solar irradiance across the SRF range. Therefore DN value can be approximated by integrating the solar radiance weighted by the product of the SRF and TOA reflectance (Cao et al., 2020; Choi et al., 2010):

$$L_t = \frac{\cos\theta_{l_t}}{\pi \cdot d^2 \cdot \text{Gain}_{l_t}} \int_{\lambda_{min}}^{\lambda_{max}} R_l(\lambda)\rho_{l_t}(\lambda)E_t(\lambda)\, d\lambda \tag{3.10}$$

where L_t is the DN value of the $LSHT$ sensor at time t, $R_l(\lambda)$ is the spectral response of the $LSHT$ band at wavelength λ, $\rho_{l_t}(\lambda)$ is the TOA surface reflectance at wavelength λ at time t, and $E_t(\lambda)$ is the TOA solar irradiance at wavelength λ at time t.

Similarly, the DN value for the $HSLT$ sensor at time t is given by:

$$H_t = \frac{\cos\theta_{h_t}}{\pi \cdot d^2 \cdot \text{Gain}_{h_t}} \int_{\lambda_{min}}^{\lambda_{max}} R_h(\lambda)\rho_{h_t}(\lambda)E_t(\lambda)\, d\lambda \tag{3.11}$$

Considering the response function of the $LSHT$ sensor within the range $[\lambda_{min}, \lambda_{max}]$, assuming no overlap between the bands of the same sensor, R_l can be treated as a linear function between λ_{min} and λ_{max}, and can be expressed as:

$$R_l = \sum_{i=1}^{m} R_l^i \tag{3.12}$$

where R_l^i represents the sensor response corresponding to band i, with $i = 1, 2, \ldots, m$, and m being the total number of $LSHT$ bands within the interval $[\lambda_{min}, \lambda_{max}]$.

Applying equation (3.12), equation (3.10) can be reformulated as:

$$L_t = \frac{\cos\theta_{l_t}}{\pi \cdot d^2 \cdot \text{Gain}_{l_t}} \sum_{i=1}^{m} \int_{\lambda_{min}}^{\lambda_{max}} R_l^i(\lambda)\rho_{l_t}(\lambda)E_t(\lambda)\, d\lambda \tag{3.13}$$

Equations (3.11) and (3.13) can be expressed in matrix form to simplify calculation:

$$H_t = \frac{\cos\theta_{h_t}}{\pi \cdot d^2 \cdot \text{Gain}_{h_t}} R_h \rho_{h_t} E_t \tag{3.14}$$

$$L_t = \frac{\cos\theta_{l_t}}{\pi \cdot d^2 \cdot \text{Gain}_{l_t}} \sum_{i=1}^{m} R_l^i \rho_{l_t} E_t \tag{3.15}$$

Note that

$$L_t^i = \frac{\cos\theta_{l_t}}{\pi \cdot d^2 \cdot \text{Gain}_{l_t}} R_l^i \rho_{l_t} E_t \tag{3.16}$$

Here, L_t^i represents the DN value of the $LSHT$ sensor at time t for the band i that overlaps with the corresponding $HSLT$ sensor band.

The physical characteristics of the sensors, including their SRF and photon detection capabilities, establish a foundational understanding of their spectral behavior (C. Song et al., 2001).

This understanding is critical for maintaining consistency in temporal analysis, as it ensures that variations in DN values across different time points stem from actual environmental changes rather than sensor-related spectral discrepancies (Villaescusa-Nadal et al., 2019). Building on this, we examine how these physical properties support the accurate tracking of temporal changes.

3.1.2 Temporal Consistency of Spectral Properties

Building on the established understanding of sensor characteristics, this section focuses on their role in ensuring temporal consistency. By leveraging the overlap between spectral responses and the derived relationships, we can reliably compare data across time points, such as t_1 and t_2. This approach ensures that observed variations in reflectance or radiance are attributed to actual environmental changes, rather than inconsistencies in sensor behavior. Similarly, we can determine the DN values for both sensors for the narrow and wide bands at two different dates (t_1 and t_2):

$$H_{t_1} = \frac{\cos\theta_{h_{t_1}}}{\pi \cdot d^2 \cdot \text{Gain}_{h_{t_1}}} R_h \rho_{h_{t_1}} E_{t_1} \tag{3.17}$$

$$H_{t_2} = \frac{\cos\theta_{h_{t_2}}}{\pi \cdot d^2 \cdot \text{Gain}_{h_{t_2}}} R_h \rho_{h_{t_2}} E_{t_2} \tag{3.18}$$

$$L^i_{t_1} = \frac{\cos\theta_{l_{t_1}}}{\pi \cdot d^2 \cdot \text{Gain}_{l_{t_1}}} R^i_l \rho_{l_{t_1}} E_{t_1} \tag{3.19}$$

$$L^i_{t_2} = \frac{\cos\theta_{l_{t_2}}}{\pi \cdot d^2 \cdot \text{Gain}_{l_{t_2}}} R^i_l \rho_{l_{t_2}} E_{t_2} \tag{3.20}$$

Maintaining temporal consistency of the spectral characteristic across sensor data allows STF methods to accurately capture real-world changes over time rather than sensor-related inconsistencies. In most STF methods, a band from each sensor is fused at the base time t_1, meaning that $L^i_{t_1}$ is used in the fusion process. To calculate the temporal changes from t_1 to t_2 for the coarse sensor bands and the fine sensor band, we have:

$$H_{t_2} - H_{t_1} = \frac{\cos\theta_{h_{t_1}}}{\pi \cdot d^2 \cdot \text{Gain}_{h_{t_1}}} R_h \left(\rho_{h_{t_2}} E_{t_2} - \rho_{h_{t_1}} E_{t_1}\right) \tag{3.21}$$

$$L^i_{t_2} - L^i_{t_1} = \frac{\cos\theta_{l_{t_2}}}{\pi \cdot d^2 \cdot \text{Gain}_{l_{t_2}}} R^i_l \left(\rho_{l_{t_2}} E_{t_2} - \rho_{l_{t_1}} E_{t_1}\right) \tag{3.22}$$

Equations 3.21 and 3.22 ensure temporal consistency of the spectral characteristics by explicitly accounting for the physical and sensor-specific factors that influence radiance measurements over time. By addressing variations in environmental conditions and sensor behavior, they ensure that observed changes reflect true surface dynamics rather than inconsistencies introduced by the sensors. This approach aligns with studies such as (Chander et al., 2009) which

emphasize the importance of maintaining spectral and temporal consistency in multi-sensor analyses.

This temporal consistency forms the foundation for accurately capturing and analyzing spatiotemporal dynamics, a crucial step in reliable STF methods. However, while temporal consistency addresses changes over time, a comprehensive approach must also account for spatial differences in reflectance between coarse and fine resolution sensors. These spatial discrepancies, influenced by sensor characteristics and calibration, can introduce systematic errors that affect data integration. Building on this, in the next section, we delve into the relationship between spatial reflectance differences and their integration with temporal consistency.

3.1.3 Integrating Temporal Consistency with Spatial Reflectance Differences

To extend the temporal consistency of the sensors characteristics established in previous sections to a fully integrated spatiotemporal framework, it is necessary to address spatial differences in reflectance between coarse and fine resolution sensors. For a given pixel (x, y), we assume that the difference in reflectance between those sensors is influenced by system differences inherent to the different sensors. The relationship between the reflectances associated with coarse resolution pixels and fine resolution pixels on a specific date can be expressed as a linear function (J. Wang & Huang, 2017), as:

$$\rho_h(x, y, t_1) = a_1 * \rho_l(x, y, t_1) + b_1 \tag{3.23}$$

Here, ρ_h and ρ_l represent the reflectances of the fine and coarse resolution sensors for the pixel (x, y) on the date t_1, respectively; while a and b are the coefficients of the linear function. These parameters compensate for systematic differences in magnitude between the two resolutions, which can arise due to differences in the sensor's sensitivity or calibration. The linear regression equation is particularly valuable as it directly addresses spatial and spectral inconsistencies, allowing for a smoother integration of data across resolutions. Assuming that the system differences and land cover remain unchanged between dates t_1 and t_2, the function parameters will be consistent, such that $a_1 = a_2 = a$. Similarly to equation (3.23), and under the assumption that b_1 is small, we can obtain:

$$\rho_h(x, y, t_2) - \rho_h(x, y, t_1) = a * (\rho_l(x, y, t_2) - \rho_l(x, y, t_1)) \tag{3.24}$$

and from (3.23) we get:

$$\frac{\rho_h(x, y, t_2)}{\rho_l(x, y, t_2)} = \frac{\rho_h(x, y, t_1)}{\rho_l(x, y, t_1)} = a \tag{3.25}$$

Equation (3.25) suggests that there is no change in land cover over time in the relationship between ρ_h and ρ_l at the location (x, y). Consequently, it is assumed that the ratio of TOA reflectance remains constant across different wavelengths over time, given that the SRF does

not vary over time (Otazu et al., 2005). Similarly, we can derive:

$$\frac{\rho_{h_{t_2}}}{\rho_{l_{t_2}}} = \frac{\rho_{h_{t_1}}}{\rho_{l_{t_1}}} = \frac{\rho_h}{\rho_l} \cdot I \tag{3.26}$$

With I being the identity matrix.

Multiplying the equation 3.21 by $\frac{\pi \cdot d^2 \cdot \text{Gain}_{h_{t_1}}}{\cos\theta_{h_{t_1}}} \cdot R_h^{-1}$, we get:

$$R_h^{-1} \cdot \frac{\pi \cdot d^2 \cdot \text{Gain}_{h_{t_1}}}{\cos\theta_{h_{t_1}}} \cdot (H_{t_2} - H_{t_1}) = \rho_{h_{t_2}} E_{t_2} - \rho_{h_{t_1}} E_{t_1}. \tag{3.27}$$

from equation 3.26 we get

$$R_h^{-1} \cdot \frac{\pi d^2 \cdot \text{Gain}_{h_{t_1}}}{\cos\theta_{h_{t_1}}} \cdot (H_{t_2} - H_{t_1}) = ((\frac{\rho_{h_{t_1}} \cdot \rho_{l_{t_2}}}{\rho_{l_{t_1}}})E_{t_2} - \rho_{h_{t_1}} \cdot \frac{\rho_{l_{t_1}}}{\rho_{l_{t_1}}} E_{t_1}). \tag{3.28}$$

And then, we can obtain

$$\rho_{l_{t_1}} . R_h^{-1} \cdot \frac{\pi \cdot d^2 \cdot \text{Gain}_{h_{t_1}}}{\cos\theta_{h_{t_1}}} \cdot (H_{t_2} - H_{t_1}) = \rho_{h_1}(\rho_{l_{t_2}} E_{t_2} - \rho_{l_{t_1}} E_{t_1}) \tag{3.29}$$

Same way, multiplying both sides of the equation 3.22 $\frac{\pi \cdot d^2 \cdot \text{Gain}_{l_{t_1}}}{\cos\theta_{l_{t_1}}} \cdot (R_l^i)^{-1}$ gives:

$$(R_l^i)^{-1} \cdot \frac{\pi d^2 \cdot \text{Gain}_{l_{t_1}}}{\cos\theta_{l_{t_1}}} \cdot (L_{t_2}^i - L_{t_1}^i) = \rho_{l_{t_2}} E_{t_2} - \rho_{l_{t_1}} E_{t_1}. \tag{3.30}$$

combining equations 3.29 and 3.30

$$\rho_{l_{t_1}} \cdot R_h^{-1} \cdot \frac{\pi d^2 \cdot \text{Gain}_{h_{t_1}}}{\cos\theta_{h_{t_1}}} \cdot (H_{t_2} - H_{t_1}) = \rho_{h_1} \cdot (R_l^i)^{-1} \cdot \frac{\pi d^2 \cdot \text{Gain}_{l_{t_1}}}{\cos\theta_{l_{t_1}}} \cdot (L_{t_2}^i - L_{t_1}^i). \tag{3.31}$$

Using the equations presented earlier, given a coarse-fine pair of images at time t_1 (L_{t_1} and H_{t_1}) and a coarse image at time t_2, the fine image at time t_2, denoted as H_{t_2}, can be modeled as:

$$H_{t_2} = H_{t_1} + \alpha^i (L_{t_2}^i - L_{t_1}^i) \tag{3.32}$$

This equation combines both temporal and spatial information, representing a generalized framework that aligns with principles commonly employed in data fusion methods. Specifically, it enables the derivation of high-resolution estimates at t_2 by effectively combining coarse-resolution temporal changes with fine-resolution spatial details. Each term plays a specific role:

- **Base Fine-Resolution Image** H_{t_1}: This term provides a starting reference, capturing fine spatial details at the base time t_1. It serves as a baseline for the fine-resolution image, ensuring that H_{t_2} retains the detailed spatial structure found at t_1.

- **Difference** $(L^i_{t_2} - L^i_{t_1})$: This component represents the temporal change in reflectance captured by the coarse-resolution sensor from t_1 to t_2. By including this difference, the equation incorporates information about changes that occurred over time, allowing H_{t_2} to reflect temporal dynamics in the scene.
- **Coefficient** α^i: The coefficient α^i adjusts the contribution of the temporal change $(L^i_{t_2} - L^i_{t_1})$ to match the fine-resolution context. It accounts for any systematic differences between coarse and fine resolutions, ensuring that the temporal changes detected at the coarse scale are accurately transferred to the fine scale.

In this chapter, we specifically focus on the scenario where multiple narrow bands overlap with a single wide band. In the case of multiple spectral narrow bands overlapped with one wide band, we get:

$$\begin{aligned} H_{t_2} &= H_{t_1} + \alpha^1 \left(L^1_{t_2} - L^1_{t_1}\right) \\ H_{t_2} &= H_{t_1} + \alpha^2 \left(L^2_{t_2} - L^2_{t_1}\right) \\ &\vdots \\ H_{t_2} &= H_{t_1} + \alpha^m \left(L^m_{t_2} - L^m_{t_1}\right) \end{aligned} \tag{3.33}$$

we can sum these equations for all $i = 1, 2, \ldots, m$

$$m \cdot H_{t_2} = m \cdot H_{t_1} + \left(\alpha^1 (L^1_{t_2} - L^1_{t_1}) + \cdots + \alpha^m (L^m_{t_2} - L^m_{t_1})\right) \tag{3.34}$$

where m is the number of spectral bands. Taking the average across the m bands, we obtain:

$$H_{t_2} = H_{t_1} + \frac{1}{m} \sum_{i=1}^{m} \alpha^i (L^i_{t_2} - L^i_{t_1}). \tag{3.35}$$

The coefficients α^i can be expressed as:

$$\alpha^i = R_h \left(\frac{\rho_h}{\rho_l} \cdot I\right) \left(R^i_l\right)^{-1} \frac{\text{Gain}_{l_{t_1}} \cdot \cos \theta_{h_{t_1}}}{\text{Gain}_{h_{t_1}} \cdot \cos \theta_{l_{t_1}}} \tag{3.36}$$

The calculation of the coefficients α^i is straightforward, as most SRFs are available in the literature. Additionally, the TOA reflectance for the high and low resolution sensors at time t_1 can be calculated from the DN values using equation 3.9.

The use of equations similar to equation 3.32 can be found in previous studies, albeit with different interpretations. For instance, in (Zhu et al., 2010), the coefficients α^i represent the temporal variation from t_1 to t_2, whereas in (J. Wang & Huang, 2017), they are used to account for system error. In (Weng et al., 2014), these coefficients describe the spatial variation between $HSLT$ and $LSHT$ sensors. In (H. Song & Huang, 2012), they are interpreted as representing the high-pass modulation between two sparsely sampled low-resolution images. Furthermore, in (Shen et al., 2013), the coefficients correspond to the spectral differences between sensors. However here, the coefficients α^i are specifically defined as the contribution of the SRFs to the spectral band adjustment between $HSLT$ and $LSHT$ sensors.

The relationship between TOA radiance and Bottom-Of-Atmosphere (BOA) radiance can be simplified using the following expression (Schott, 2007; S. Zhou & Cheng, 2023):

$$\mathrm{TOA}_\lambda = \mathrm{BOA}_\lambda \cdot \tau_\lambda + I_\lambda^\uparrow, \tag{3.37}$$

where TOA_λ is the TOA spectral radiance for a given wavelength λ, BOA_λ is the corresponding BOA spectral radiance, $I_\lambda^\uparrow$ represents the atmospheric upwelling radiance, and τ_λ is the atmospheric transmittance. For $HSLT$ and $LSHT$ sensors, this relationship becomes:

$$L_{\mathrm{BOA}} \cdot \tau = L_{\mathrm{TOA}} - I^\uparrow, \tag{3.38}$$

$$H_{\mathrm{BOA}} \cdot \tau = H_{\mathrm{TOA}} - I^\uparrow, \tag{3.39}$$

where L_{BOA} and H_{BOA} are the BOA radiances for $HSLT$ and $LSHT$ sensors, respectively, and L_{TOA} and H_{TOA} are the corresponding TOA radiances. Dividing these two equations, we obtain:

$$\frac{H_{\mathrm{BOA}}}{L_{\mathrm{BOA}}} = \frac{H_{\mathrm{TOA}} - I^\uparrow}{L_{\mathrm{TOA}} - I^\uparrow}. \tag{3.40}$$

Under favorable atmospheric conditions, with minimal aerosols or haze, the upwelling radiance $I^\uparrow$ becomes small compared to the TOA radiances. In such cases, $I^\uparrow \ll H_{\mathrm{TOA}}$ and $I^\uparrow \ll L_{\mathrm{TOA}}$, and the relationship simplifies to (Gilabert et al., 1994):

$$\frac{H_{\mathrm{BOA}}}{L_{\mathrm{BOA}}} = \frac{H_{\mathrm{TOA}}}{L_{\mathrm{TOA}}}. \tag{3.41}$$

This result is consistent with equation 3.25, indicating that the linearity between the coarse and the fine sensors remains valid at the BOA level as well.

Based on this theoretical foundation, we propose a linear regression model to spectrally align the fine-resolution bands with the coarse-resolution bands. This linear regression model for band adjustment integrates the physical properties of the SRFs, temporal consistency, and spatial differences established in previous sections. The relationship at time t_1 can be expressed as:

$$F_{t_1} = \sum_{i=1}^{N} a_i C'_{i,t_1}, \tag{3.42}$$

where F_{t_1} represents the fine-resolution band at time t_1, C'_{i,t_1} is the spatially downscaled coarse-resolution band at time t_1; and a_i is the regression coefficient. The number N represents the overlapping coarse-resolution bands.

The same relationship holds at time t_2:

$$F_{t_2} = \sum_{i=1}^{N} a_i C'_{i,t_2}, \tag{3.43}$$

The change between times t_1 and t_2 can be described as:

$$F_{t_2} - F_{t_1} = \sum_{i=1}^{N} a_i \left(C'_{i,t2} - C'_{i,t1} \right). \tag{3.44}$$

Here, the regression coefficients a_i are directly proportional to the SRF coefficients α_i, making the two approaches mathematically equivalent. Therefore, the coefficients a_i represent the contribution of each coarse-resolution band to the adjustment, harmonizing the $HSLT$ and $LSHT$ coarse and fine data both temporally and spectrally.

In summary, the theoretical foundation presented here establishes a comprehensive framework for integrating spectral information into STF methods, ensuring spectral, temporal, and spatial consistency between $HSLT$ and $LSHT$ sensors. The SRF overlap is used primarily to define the overlapping bands between sensors, setting the foundation for aligning spectral information. By examining the physical properties of each sensor, we understand how SRFs influence photon detection across bands, providing a basis for consistent alignment between sensors. Temporal consistency is further ensured by modeling reflectance changes over time, so observed differences reflect true environmental dynamics rather than sensor discrepancies. Using a linear regression model for band adjustment, we align $HSLT$ and $LSHT$ bands, harmonizing both temporal changes and spectral characteristics across resolutions. The model's regression coefficients, directly linked to SRF properties, minimizing spectral discrepancies.

3.2 Methodology, Materials, and Implementation

In this section, we detail the methodology and materials used to develop and test our proposed MSTBA strategy for STF of S2 and S3 satellite imagery. The study area selection is described, featuring two distinct environments that present unique challenges for STF methods, allowing for a robust evaluation of the proposed strategy. We outline the data preparation workflow, including atmospheric correction, co-registration, and resampling, to ensure that both datasets are accurately aligned and comparable. Next, the MSTBA process is presented, highlighting how spectral discrepancies are minimized through linear regression modeling of overlapping S3 bands to align with S2 spectral characteristics. Finally, we describe the integration of adjusted bands within STF methods, explaining how this approach enhances fusion accuracy by maintaining spectral consistency over time.

Based on the theoretical foundation established in the previous section, we propose the MSTBA strategy designed to address spectral discrepancies in STF by selectively adjusting overlapping bands between S2 and S3. This method improves spectral similarity by focusing on S2 bands that overlap with multiple S3 bands, allowing for enhanced spectral consistency in the fused images across three STF methods. By leveraging the additional spectral information from overlapping bands, our approach ensures a closer match in spectral properties between S2 and S3, ultimately leading to more accurate and consistent fusion results.

This study focuses exclusively on the S2 bands that overlap with more than one S3 band, as outlined in Figure 3.1 and detailed in Table 3.1. These bands were selected due to the additional spectral information offered by multiple overlaps, which provides greater precision during the spectral adjustment process. By using several overlapping S3 bands, the MSTBA

strategy facilitates the creation of adjusted S3 bands whose spectral characteristics closely align with those of the corresponding S2 bands. The key S2 bands considered in this work include the Blue, Red, and NIR bands, which overlap with corresponding S3 bands.

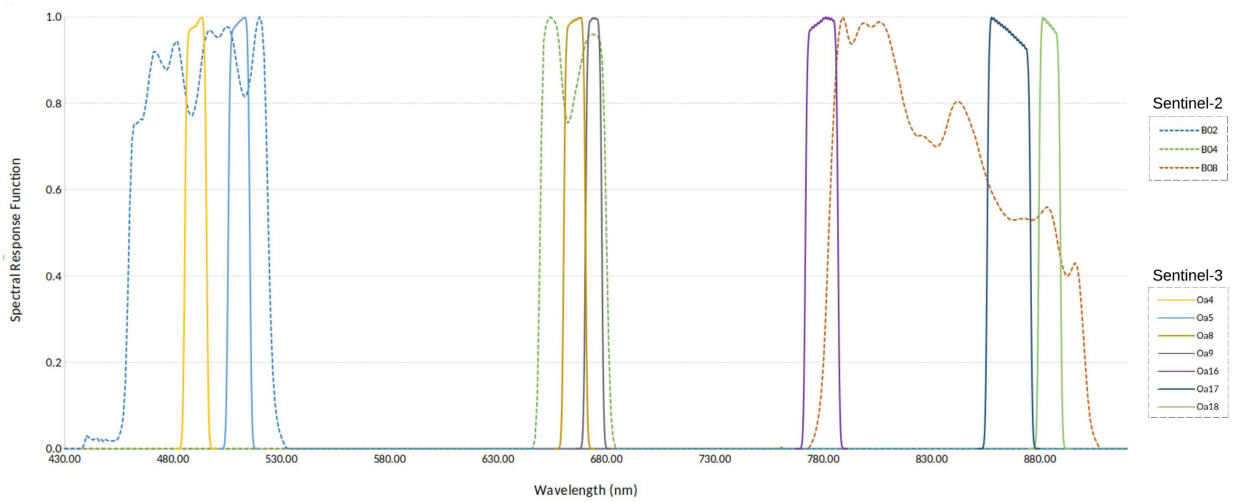


Figure 3.1: SRFs corresponding to S2 (MSI) and S3 (OLCI). S2 SRF is presented by a dashed line and S3 SRF is presented by a continuous line.

Table 3.1: Summary of the overlapping Sentinel-2 and Sentinel-3 OLCI bands presented in figure 3.1

Specification	**Band**	**Sentinel-2**		**Sentinel-3**	
		MSI band	**Wavelength range (nm)**	**OLCI band**	**Wavelength range (nm)**
Band width	Blue	B02	458-523	Oa4 Oa5	485-495 505-515
	Red	B04	650-680	Oa8 Oa9	660-670 670-677
	NIR	B08	780-885	Oa16 Oa17 Oa18	771-786 855-875 880-890
Spatial resolution		10 m		300 m	
Temporal resolution		5 days		1.5 days	

3.2.1 Study areas

The first site is Waterbank, located in northwestern Australia, which experiences a hot semi-arid climate with distinct wet and dry seasons. The wet season, from December to June,

brings around 580 mm of rainfall, with record peaks like 910 mm in January, while the dry season (July to November) sees less than 30 mm of rain. Waterbank is prone to frequent fires, especially during the dry season when vegetation dries out, creating fuel for wildfires. This site provides an ideal test for STF methods in a rapidly changing landscape when fire occurs. A total of 40 cloud-free S2 and S3 OLCI image pairs were collected between January 2019 and November 2020, covering an area of 324 km^2.

The second site is Maspalomas, a natural reserve in the Canary Islands, Spain. Known for its subtropical desert climate, Maspalomas experiences hot, dry summers and mild winters. Average temperatures range from 20 to 29°C, with most of the rainfall occurring in the winter, averaging 35 mm. The site is characterized by diverse landscapes, including coastal dunes, the Maspalomas lagoon, and terrestrial areas, alongside frequent sandstorms. The complexity of the ecosystem, combined with human activity, makes it a challenging environment for STF methods. A total of 40 cloud-free image pairs were collected between January 2019 and December 2020, covering 81 km^2.

Figure 3.2 shows the location of the study areas. Details of the dates for each image pair collected for both study sites can be found in Annex 1.

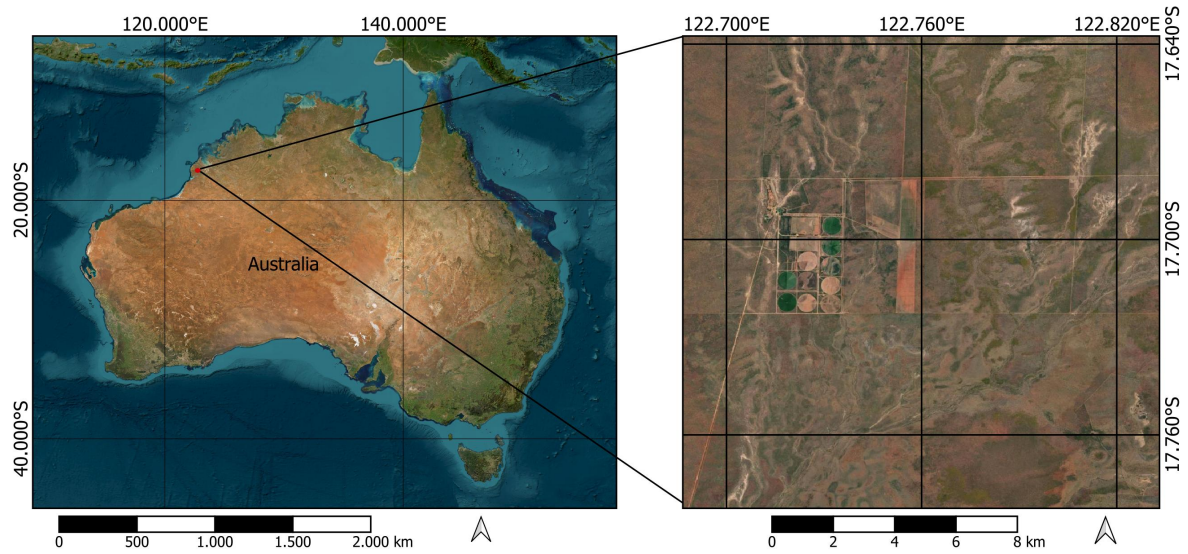
3.2.2 Data preparation

The first step in the preprocessing workflow was atmospheric correction. Although various atmospheric correction algorithms are available, only a few are compatible with both S2 Level-1C and S3 OLCI Level-1B data. In this study, atmospheric correction was conducted using the Image Correction for Atmospheric Effects (iCOR) algorithm (De Keukelaere et al., 2018), which is specifically designed for these datasets. iCOR is capable of processing satellite images collected over different types of land and water surfaces, and it follows a series of steps to correct atmospheric effects.

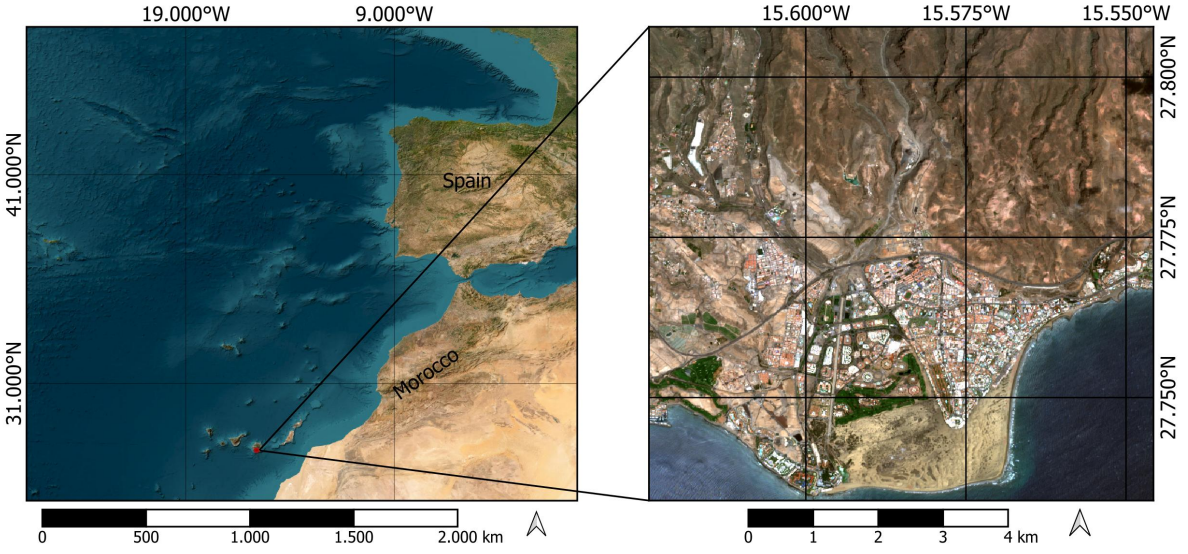
Initially, iCOR applies a band threshold to distinguish between water and land pixels. The aerosol optical thickness (AOT) is then estimated from the land pixels and extrapolated over water areas. Following this, adjacency correction is carried out using the SIMilarity Environmental Correction (SIMEC) approach. iCOR further incorporates solar and view angles, including the Sun zenith angle (SZA), view zenith angle (VZA), and relative azimuth angle (RAA), along with a Digital Elevation Model (DEM), to perform the final atmospheric correction. This correction references MODTRAN5 Look-Up-Tables (LUTs) to ensure accurate results (De Keukelaere et al., 2018). The iCOR algorithm was accessed through the SNAP software plugin[1]. After atmospheric correction, the S3 data were reprojected to the UTM/WGS84 projection to match the projection of the S2 data.

Geometric alignment (co-registration) is crucial for effective time series processing of remotely sensed data, especially when integrating data from different sensors (Rufin et al., 2020). The primary problem in time series analysis without co-registration is that geometric inconsistencies—like misalignments exceeding a pixel's width—can lead to significant errors. These errors hinder accurate tracking of changes and comparisons across datasets, especially between differ-

[1] https://step.esa.int/main/download/snap-download/ (last access March 2025)



(a)



(b)

Figure 3.2: Geographic location of (a) Waterbank, North Australia. (b) Maspalomas, Gran Canarias, Spain.

ent sensors (Townshend et al., 1992). For this reason, it is essential to not only co-register each pair of images but also to ensure consistent geometric alignment across the entire time series. This approach minimizes cumulative spatial discrepancies that could otherwise be amplified over multiple observations, ensuring that the changes observed over time are due to actual landscape dynamics rather than geometric misalignment. The co-registration process was completed in two stages. First, a single pair of S2 and S3 images was manually co-registered using ENVI 5.6.1 software. Following this initial co-registration, resampling was conducted in ENVI to ensure that the spatial resolution and grid structure of the images matched

those of S2. This resampled reference pair was then used to automatically co-register the entire time series of both S2 and S3 images using AROSICS software[2] (Scheffler et al., 2017), which employs cross-correlation in Fourier space to achieve precise alignment. By applying co-registration across both time series, we maintain geometric consistency throughout the dataset. The data preparation workflow is illustrated in Figure 3.3.

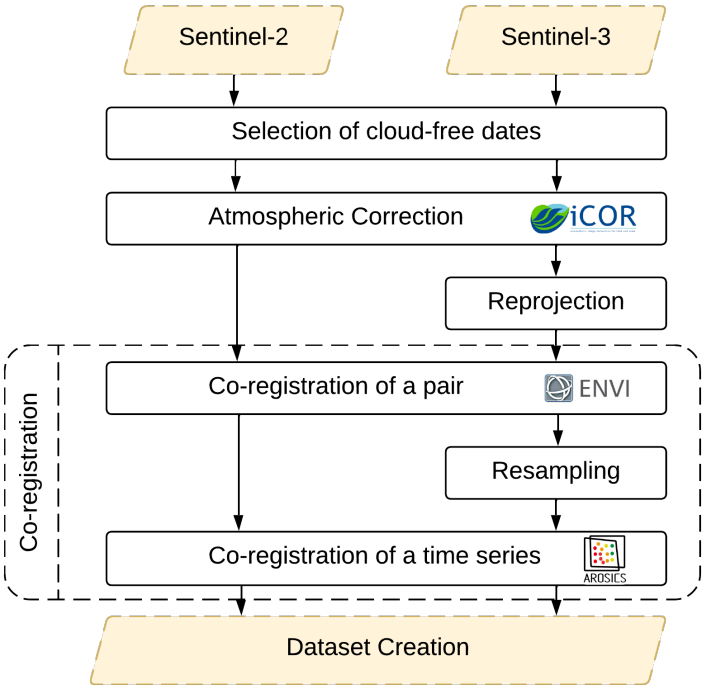


Figure 3.3: The data preparation workflow

3.2.3 MSTBA method

The MSTBA begins by calculating a linear regression model (LRM) at the base date t_1 using the equation 3.42. The coefficients obtained a_i (where $i = 1, 2, \ldots, n$, being n the number of overlapping S3 bands) present the information contribution of each narrow S3 band to match the spectral characteristics of S2. These coefficients quantify each S3 band's contribution to forming a spectral profile that closely matches S2. Since incorporating all S3 bands individually into the STF method is not feasible, we use the regression-derived coefficients to create an adjusted S3 band. This band aggregates the contributions of relevant narrow bands into a single, spectrally aligned representation, ensuring harmony with S2.

By applying the same set of coefficients to the S3 bands on both dates (t_1 and t_2), the adjustment process maintains spectral consistency over time. The adjusted S3 bands replace the original bands for subsequent STF process. Figure 3.4 presents the flowchart of the MSTBA method.

[2] https://helmholtz.software/software/arosics

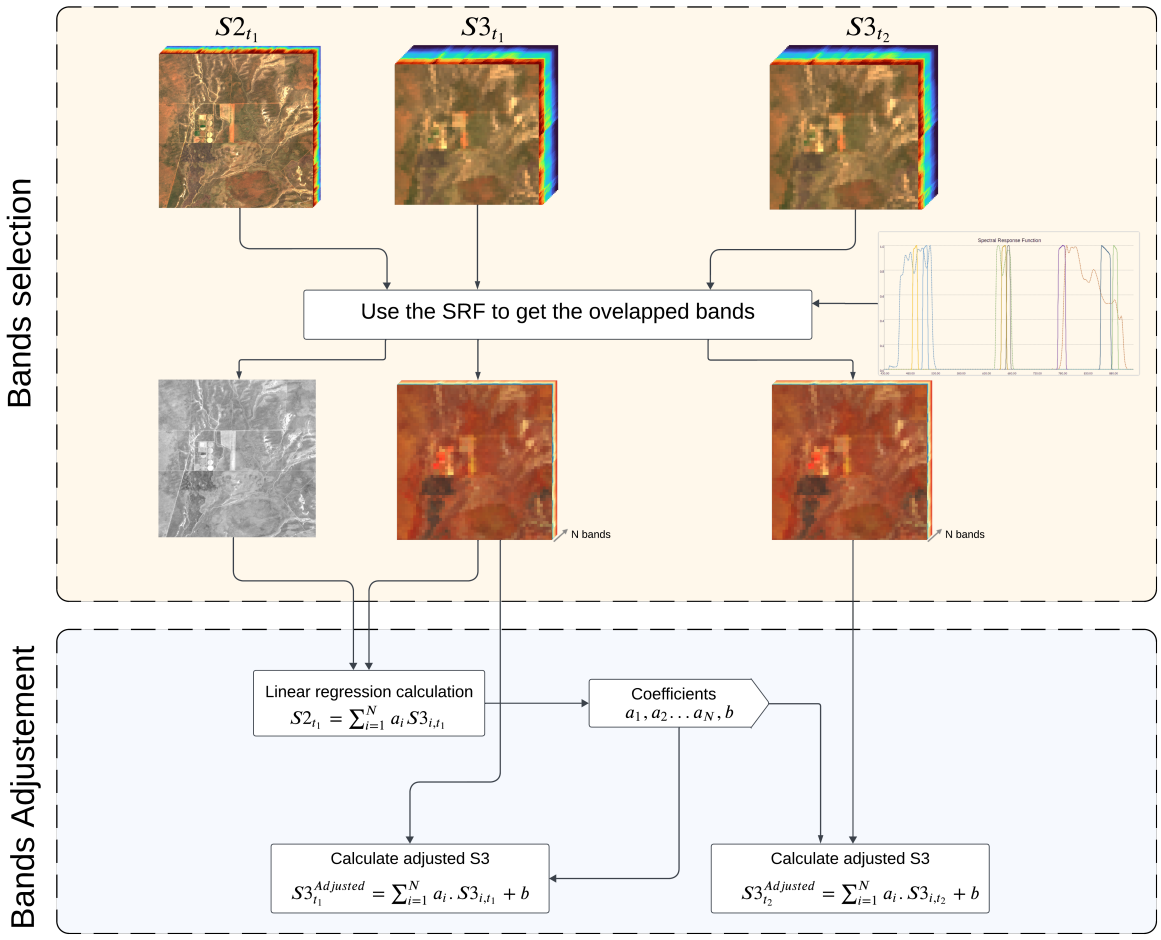


Figure 3.4: Flowchart of the proposed band adjustment method including the selection of the narrow and wide bands

Once the S3 bands have been adjusted to match the S2 bands at both t_1 and t_2, they are used as input for the STF methods, alongside the S2 bands at t_1. In the fusion process, the adjusted S3 bands replace the original S3 bands at both t_1 and t_2, ensuring a seamless integration and enhancing the accuracy of the fusion.

3.2.4 Experimental setup

In this study, we conducted experiments to assess the effectiveness of the proposed MSTBA strategy for improving STF accuracy between S2 and S3 data across the study sites: Waterbank (Australia), and Maspalomas (Spain). The experimental setup consisted of the following key elements: selection of bands, STF methods, and evaluation metrics, each described below. The pairs were carefully chosen to ensure consistency and temporal coverage, allowing for a robust analysis under varying environmental conditions. The dates for each site are detailed in Annex 6.3. For each site, the focus was placed on three key spectral bands: blue, red, and near-infrared (NIR). These bands were selected due to the unique overlap between each S2

band and multiple S3 bands, allowing the MSTBA strategy to leverage additional spectral information and achieve a more precise alignment. Specifically, the bands used are detailed in Table 3.1 and they were used in the fusion as follows:

- **Blue Band**: S2 band B02, aligned with S3 bands Oa4 and Oa5.
- **Red Band**: S2 band B04, aligned with S3 bands Oa8 and Oa9.
- **NIR Band**: S2 band B08, aligned with S3 bands Oa16, Oa17, and Oa18.

The experiment was conducted in two parts. In the first part, fusion was performed using the original S3 bands individually. For each S2 band, we fused it with each overlapping S3 band separately—meaning, for example, that the S2 blue band (B02) was fused with S3 bands Oa4 and Oa5 individually, while the S2 red band (B04) was fused with S3 bands Oa8 and Oa9 individually, and so on. This approach allowed us to assess the performance of STF methods using unadjusted spectral data from S3, analyzing how each original S3 band performed in the fusion. In the second part of the experiment, the MSTBA strategy was applied to combine information from overlapping S3 bands, producing a single adjusted S3 band for each spectral range (blue, red, and NIR) that more closely matched the spectral characteristics of the corresponding S2 band. This adjusted band was then fused with the corresponding S2 band to evaluate the impact of the spectral adjustment on fusion accuracy.

To assess the impact of the adjusted bands in the fusion quality, we applied three STF methods: STARFM, FSDAF, and Fit-FC. Each method was selected for its ability to handle specific challenges associated with STF. For each STF method, the default parameters outlined in the original research were employed to ensure a standardized evaluation process. More details about those STF methods are in the previous chapter, section 2.3, and they were applied using their respective default settings to ensure consistency. The STF methods produced predicted images at a target time t_2 for each dataset, allowing us to evaluate the accuracy and fidelity of the fusion results under both conditions using either original or adjusted bands.

3.2.5 Evaluation metrics

In image fusion quality assessment, there is no single metric capable of comprehensively evaluating all aspects of image quality. Each metric captures specific characteristics, such as spatial structure, spectral fidelity, or noise levels, but none provide a complete picture alone. Therefore, using multiple metrics is essential to thoroughly assess the fused image's quality (Zhu et al., 2022). This combination of metrics enables a more robust and balanced evaluation, ensuring that both spatial and spectral information is preserved while minimizing artifacts and distortions introduced during fusion. Employing multidimensional approach helps validate that the fusion method effectively meets the diverse requirements of the intended application.

RMSE

The Root Mean Square Error (RMSE) measures the average squared differences between corresponding pixel values of the fused image and a reference image, thus quantifying the pixel-level deviations. RMSE serves as a fundamental metric for assessing the overall error introduced during the fusion process, making it especially useful for numerical accuracy

evaluation. Lower RMSE values indicate higher similarity to the reference image, meaning less information loss or alteration. However, RMSE lacks sensitivity to structural or perceptual elements of the image, limiting its usefulness for applications that require high spatial fidelity in fused results.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(I_{\text{fused}}(i) - I_{\text{ref}}(i)\right)^2} \tag{3.45}$$

where N represents the total number of pixels, and $I_{\text{fused}}(i)$ and $I_{\text{ref}}(i)$ are pixel values at position i in the fused and reference images, respectively.

SSIM

Structural Similarity Index (SSIM) evaluates structural similarity between the fused and reference images by considering luminance, contrast, and structure components (Z. Wang et al., 2004). It ranges from -1 to 1, with values closer to 1 indicating higher similarity.

$$\text{SSIM}(I_{\text{fused}}, I_{\text{ref}}) = \frac{(2\mu_{\text{fused}}\mu_{\text{ref}} + C_1)(2\sigma_{\text{fused, ref}} + C_2)}{(\mu_{\text{fused}}^2 + \mu_{\text{ref}}^2 + C_1)(\sigma_{\text{fused}}^2 + \sigma_{\text{ref}}^2 + C_2)} \tag{3.46}$$

where: μ_{fused} and μ_{ref} are the mean values of I_{fused} and I_{ref}, respectively, σ_{fused} and σ_{ref} are the standard deviations, $\sigma_{\text{fused, ref}}$ is the covariance between I_{fused} and I_{ref}, C_1 and C_2 are small constants to avoid division by zero.

SAM

Spectral Angle Mapper (SAM) is widely used in remote sensing applications, measuring the spectral similarity between two images by calculating the angle between their pixel vectors (Kruse et al., 1993). Smaller SAM values indicate closer spectral similarity.

$$\text{SAM} = \frac{1}{N}\sum_{i=1}^{N}\arccos\left(\frac{\langle \mathbf{I}_{\text{fused}}(i), \mathbf{I}_{\text{ref}}(i)\rangle}{\|\mathbf{I}_{\text{fused}}(i)\|\|\mathbf{I}_{\text{ref}}(i)\|}\right) \tag{3.47}$$

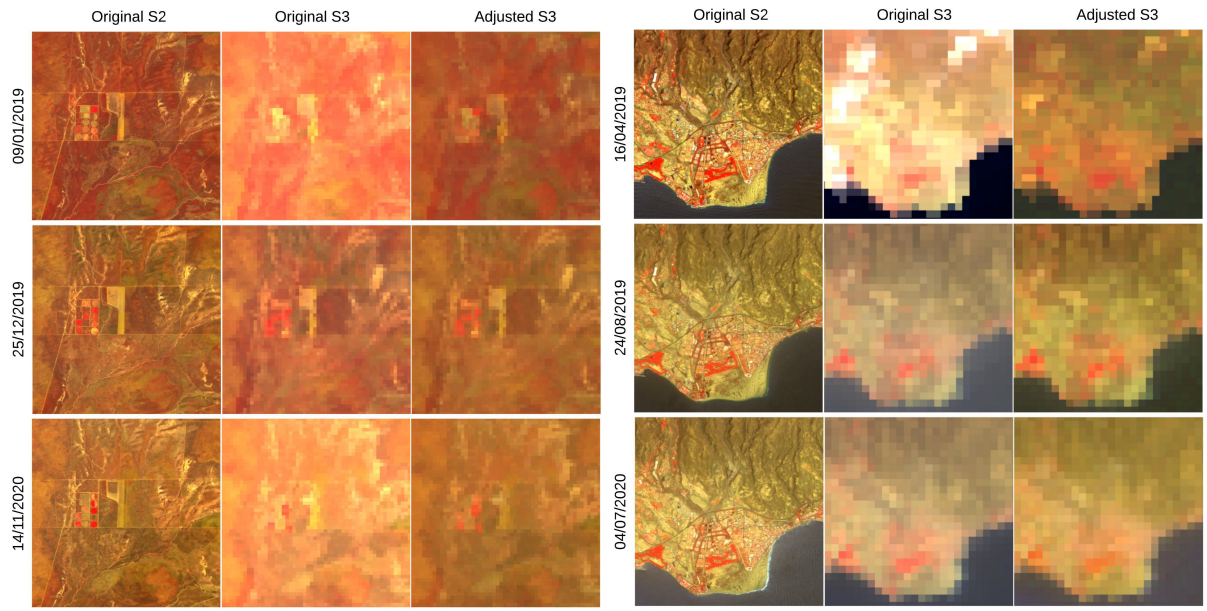
where $\mathbf{I}_{\text{fused}}(i)$ and $\mathbf{I}_{\text{ref}}(i)$ represent the spectral vectors at the i-th pixel in the fused and reference images.

3.3 Results

For all visualizations in this section, an NIR-Red-Blue composite was used for both S2 and S3 imagery to facilitate consistent visual comparisons. Specifically, for original S3, the composite was created using bands Oa17 (NIR), Oa8 (Red), and Oa4 (Blue), aligning with the corresponding spectral bands in S2.

Figure 3.5 showcases the NIR-Red-Blue composite images for the two study areas, contrasting the original and spectrally adjusted S3 images with S2 images across three distinct dates. In the original S3 composites, significant discrepancies in color representation and contrast are

observed when compared to the S2 reference. These differences indicate a notable spectral mismatch between the original S3 and S2 data, highlighting the need for spectral adjustments to enable accurate comparison and integration of S3 and S2 data. After applying the spectral adjustment procedure, the adjusted S3 composites show a marked improvement in their spectral alignment with the S2 references for both studied sites. For Waterbank, areas that previously appeared to fade in the original S3 images are now more distinctly defined, reflecting a closer match to the spectral characteristics of the S2 data. Similarly, in Maspalomas, the adjusted S3 images exhibit improved contrast, particularly in coastal zones and vegetated areas. The adjusted S3 composites show better alignment with the spectral characteristics of the S2 data, as evidenced by the more consistent color tones and enhanced contrast across different land cover types. The overall improvement in spectral fidelity demonstrates the effectiveness of the adjustment in reducing discrepancies between the sensors. This is further supported by the quantitative and qualitative results, which highlight the reduced spectral mismatch and improved alignment with the S2 reference. The consistent results across both sites further validate the robustness of the adjustment technique for different geographic settings, ensuring that the adjusted bands are more suitable for tasks that require high spectral accuracy.



(a) Waterbank (Australia) (b) Maspalomas (Spain)

Figure 3.5: NIR-Red-Blue composites of S2, S3 (Oa17, Oa8, Oa4) and adjusted S3 images

To visualize the impact and for the sake of simplicity, we calculated the differences Δ between the metrics values mentioned in the experimental setup (Section 3.2.4) obtained in these two cases. These differences have been visualized as heatmaps.

3.3.1 Scenario 1: Waterbank Site

The analysis begins with a quantitative evaluation of the fusion results for the Waterbank site, as summarized in Table 3.2. The table presents the average values of the quality metrics, including the SAM, SSIM, and RMSE, for the entire dataset. The metrics were computed for fused images generated using the sets of inputs mentioned earlier. The results indicate that using the adjusted bands (A-Blue, A-Red and A-NIR) consistently improves the quality of the fused images across the three spectral bands. The average SAM values are lower for the adjusted bands, suggesting a closer match to the spectral characteristics of the ground-truth S2 images. Similarly, the RMSE values are reduced, while the SSIM values are higher when the adjusted bands are used, reflecting improvements in both spectral fidelity and spatial detail.

Table 3.2: Average of the quality metrics for the whole dataset for the Waterbank site

	STARFM			FSDAF			Fit-FC		
	RMSE ↓	SSIM ↑	SAM ↓	RMSE ↓	SSIM ↑	SAM ↓	RMSE ↓	SSIM ↑	SAM ↓
Oa4	0.010	0.964	0.126	0.011	0.939	0.125	0.010	0.967	0.116
Oa5	0.011	0.960	0.133	0.012	0.934	0.131	0.011	0.965	0.122
A-Blue	**0.009**	**0.973**	**0.111**	**0.009**	**0.972**	**0.110**	**0.009**	**0.974**	**0.105**
Oa8	0.018	0.942	0.124	0.021	0.920	0.120	0.018	0.953	0.112
Oa9	0.018	0.941	0.126	0.021	0.918	0.123	0.018	0.952	0.113
A-Red	**0.016**	**0.950**	**0.109**	**0.017**	**0.951**	**0.106**	**0.016**	**0.956**	**0.101**
Oa16	0.039	0.926	0.083	0.045	0.903	0.080	0.040	0.911	0.088
Oa17	0.045	0.921	0.089	0.049	0.896	0.085	0.045	0.906	0.101
Oa18	0.046	0.920	0.090	0.050	0.894	0.086	0.046	0.904	0.105
A-NIR	**0.034**	**0.934**	**0.071**	**0.035**	**0.935**	**0.070**	**0.034**	**0.943**	**0.066**

A-Blue = adjusted blue, A-Red = adjusted red, A-NIR = adjusted NIR.

The heatmaps included in Figure 3.6 offers a comprehensive visualization of the differences in performance metrics between the two fusion cases: images generated using the original S3 bands and those using adjusted bands.

For each case, performance metrics (SAM, SSIM and PSNR) were first computed by comparing the fused image with the reference S2 image. The values from the two cases were then subtracted from each other to highlight the differences in performance. Each heatmap represents the differences of RMSE, SSIM, and SAM metrics. The red color in the heatmaps signifies that the fusion metrics using the original bands as input are better than those of fusion using the adjusted bands, while the green color indicates improved performance with the adjusted bands.

The heatmaps summarize these differences across the 40 image pairs analyzed, providing a side-by-side comparison of the metrics for each STF method. Specifically, SAM values for the adjusted bands show a notable decrease, with reductions reaching up to 0.2, indicating a closer alignment with the spectral characteristics of the reference images S2. Additionally, both SSIM and RMSE values display significant improvements when using adjusted bands.

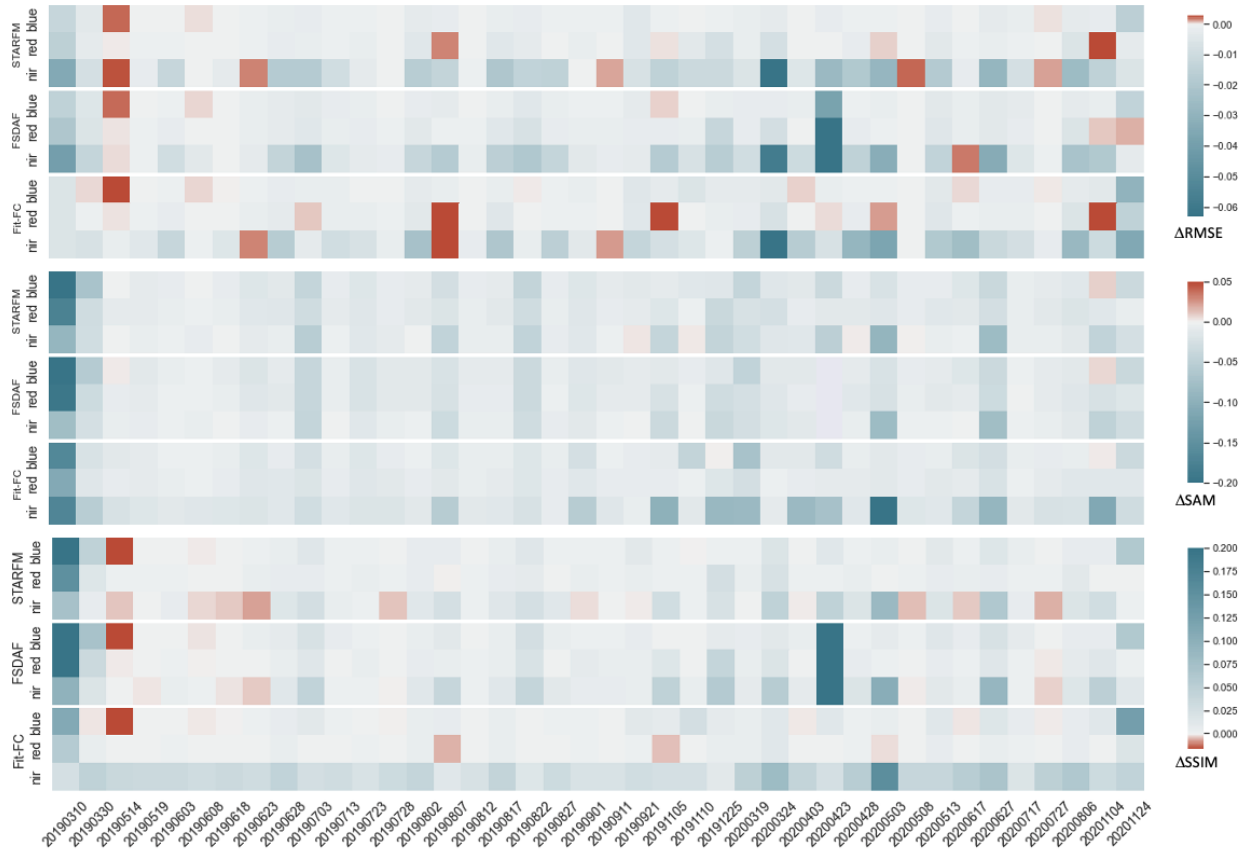


Figure 3.6: Heatmap comparison of quality metrics for fused images using adjusted versus original input bands at the Waterbank site.

Specifically, negative values in RMSE indicate better quality in the fusion results with adjusted bands, as they reflect reduced error compared to the original bands. Similarly, higher SSIM values for the adjusted bands imply greater structural similarity to the S2 reference images, further confirming the enhanced performance of the adjusted bands in the fusion process.

One exception to this trend is observed for the blue band on May 14th, 2019, where fusion results generated with original bands yielded better metrics than those with adjusted bands. This anomaly may be attributed to the presence of smoke from a fire recorded on a previous date, which could have been carried over by the adjustment process, even though the smoke was not present on the prediction date itself. This smoke may have introduced spectral inconsistencies, affecting the ability of the adjusted bands to reflect the true conditions. Saturation effects and inconsistencies, as illustrated in Figure 3.5 for January 9th, 2019 (first row), can lead to inaccurate spectral information and hinder fusion quality. Despite this specific case, the general trend in Figure 3.6 supports the use of adjusted bands, as they more reliably produce results closer to the reference S2, particularly in SAM, in most image pairs and spectral bands. Consistently, the heatmaps reveal that fusion results produced with adjusted bands tend to outperform those generated with original bands, suggesting an overall enhancement in the fusion quality.

The results for May 3, 2020, at the Waterbank site, which experiences rapid changes due to forest fires and vegetation dynamics, are illustrated in Figure 3.7. The comparison involves images generated by the different STF algorithms using both the original and adjusted S3 spectral bands. In the first row (a) and (b) represents the original S3 bands, while figures (c), (d), and (e) show the predicted images produced by STARFM, FSDAF, and Fit-FC methods, respectively, using the original S3 bands. The second row includes the adjusted S3 bands in figures (f) and (g), figure (h), (i), and (j) displaying the predictions using the adjusted bands with the same methods. The results indicate that images generated with the adjusted S3 bands show better spectral similarity to the reference S2 image, particularly in areas depicting burned vegetation. This is evident in the intensity of the green regions, which are more accurately represented in the predictions using adjusted bands. The closer spectral alignment suggests that the band adjustment enhances the ability of the STF algorithms to capture better spectral characteristics of the landscape.

To further quantify these improvements, Figure 3.8 presents the difference maps between the reference image and the fused images across the three bands. For visualization purposes, different scales were used in each difference map to better illustrate the improvements achieved when using the adjusted bands. These difference maps reveal areas where spectral discrepancies are reduced, highlighting the impact of band adjustment. The fused images using original S3 bands show larger discrepancies across all three bands, indicating deviations from the reference image, particularly in regions with complex terrain and vegetation. The adjusted bands, however, consistently reduce these spectral discrepancies, with a noticeable improvements in the blue and red bands, where the predictions align more closely with the reference image. In the map, the reductions in errors are especially evident in the details, primarily along the borders of the burned areas and in the fine details of the structures. The difference maps in Figure 3.8(c) reveal that the discrepancies between the fused and reference images are markedly reduced when using the adjusted NIR band. As NIR is sensitive to vegetation health and density. The original S3 predictions exhibit higher errors, especially in the burned areas located at the bottom, depicted in dark green within the false-color composite. In contrast, the adjusted NIR band captured better this variations in plant health, providing a more accurate representation of vegetation conditions. Additionally, figure 3.8 shows that the Fit-FC algorithm benefits from the adjusted bands by exhibiting a noticeable reduction in the blurring effect that was present when using the original bands.

The blurring observed in the difference maps highlights the substantial improvement achieved with the adjusted bands compared to the original. Regions where the original predictions were less accurate such as the bottom right area in the map show higher difference values, indicating that the adjusted bands led to sharper, more accurate fused images. In the original Fit-FC results, a slight blur is evident, particularly in regions with sharp transitions in vegetation density or terrain features, which diminishes the clarity of the predicted images. The adjustment reduces this blurring effect, leading to a more defined boundaries in the vegetation structures. This improvement enhances the visual quality and ensures a more precise spectral representation.

To capture the finer details of the fusion results, we included zoomed-in views of selected regions within the Waterbank site presented as white boxes in Figure 3.7. These close-ups

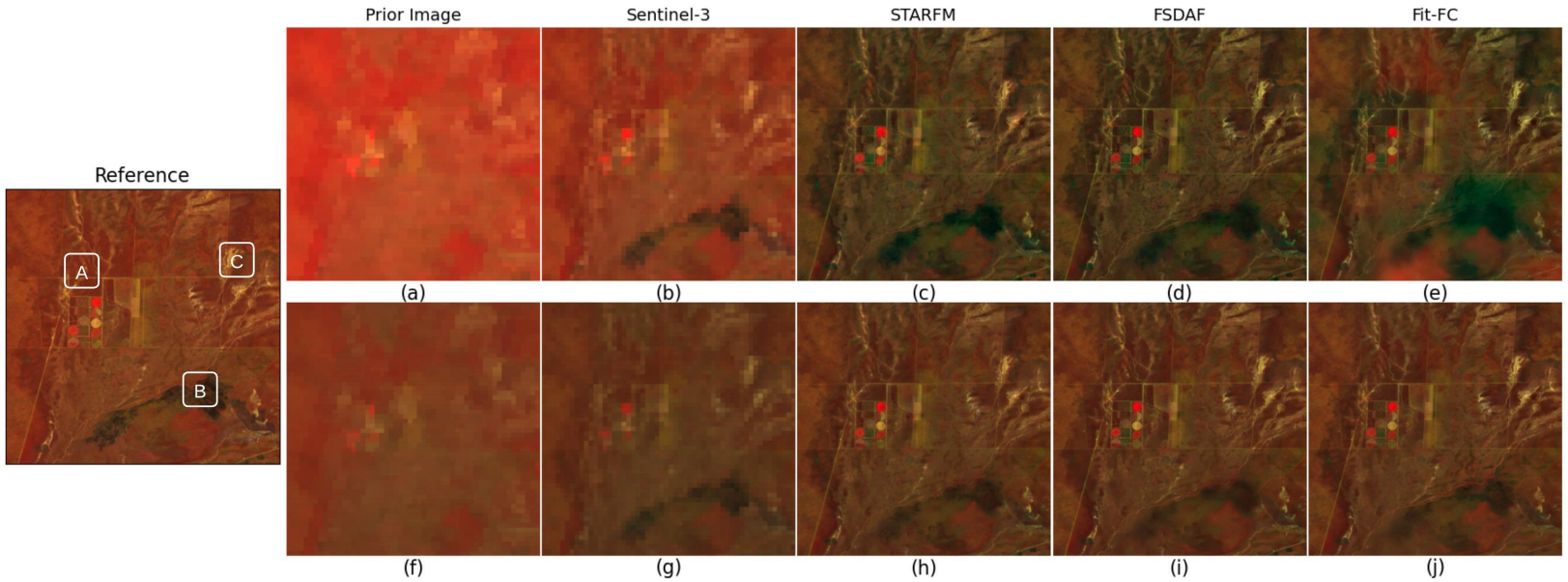
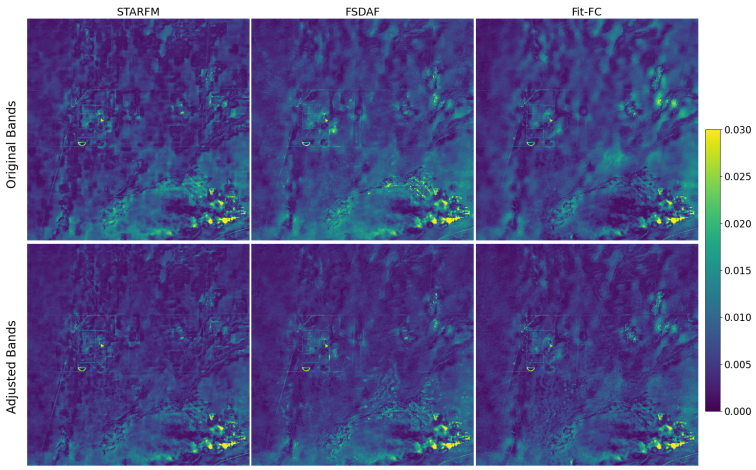
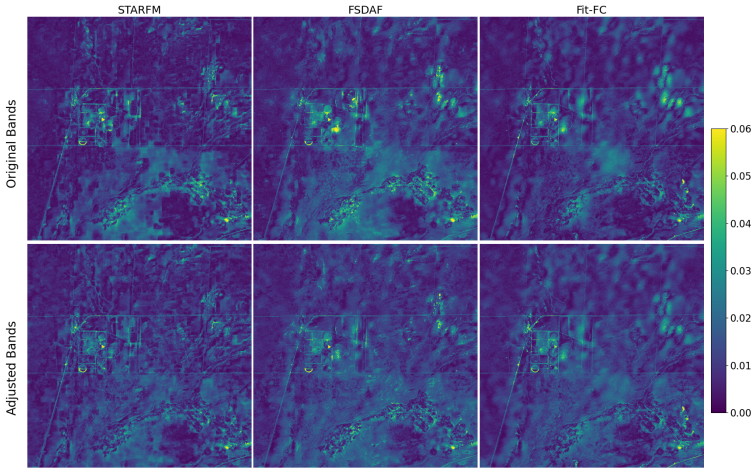


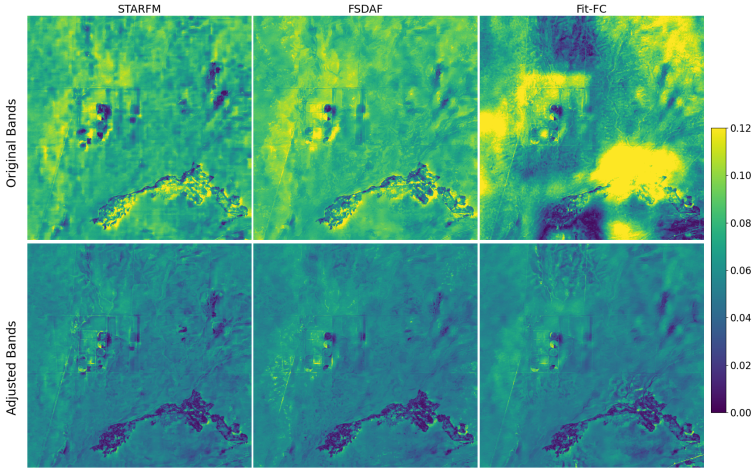
Figure 3.7: Color composition of ground truth (S2), original S3, and STF algorithm predictions for the May 3, 2020. The first row shows S3 original bands, and the second row shows adjusted bands. In each row: (a)/(f) and (b)/(g) represent the prior and prediction dates, while (c)/(h), (d)/(i), and (e)/(j) are predictions by STARFM, FSDAF, and Fit-FC, respectively.



(a) Band blue



(b) Band red



(c) Band NIR

Figure 3.8: Difference between the reference and fused images using original and adjusted S3 bands, with STARFM, FSDAF, and Fit-FC for the WaterBank site.

allow for pixel-level examination, revealing subtle differences and details that may not be immediately apparent in the full-scale figures. By zooming-in, we can better observe how each STF method handles complex textures, small-scale structures, and transitions in land cover. This level of detail is crucial for assessing the effectiveness of the adjusted versus original bands, as it highlights improvements in sharpness, accuracy, and spectral consistency that might otherwise be lost in broader views. The zoomed-in areas provide a clearer insight into how well the fusion captures the true characteristics of the landscape at a granular level.

In Figure 3.9 the fusion results for the zoomed-in regions are presented, illustrating the performance of STF methods. Two prediction dates are chosen –November 5, 2019, and May 3, 2020– based on the significant environmental changes that occurred during this period, including a fire event. The structure of the figure try to facilitate a clear comparison between the original and adjusted band results. The first row (Reference) shows the ground-truth S2 images. For each STF method, the fusion results using the original S3 bands are presented in the left columns (2nd, 4th, and 6th), while the fusion results using the adjusted S3 bands are displayed in the right columns (3rd, 5th, and 7th).

A visual inspection of the figure reveals that the fusion results using the adjusted bands are consistently closer to the reference images, particularly in zones A and B, where notable improvements are observed. This improvement is especially evident in challenging regions such as zone B, where the burned areas caused by the fire appear more naturally in the predicted images, with less spectral distortion when the adjusted bands are used. In the original band results, these areas often show unwanted greenish hues and poor spectral representation.

The STF methods exhibit varying levels of success in predicting both spatial and spectral information. STARFM maintains reasonable spatial details, but its spectral accuracy suffers, especially in regions affected by rapid environmental changes. FSDAF shows better performance in terms of spatial clarity, as evidenced by the sharper contours of roads and other landscape features. However, it still struggles with spectral fidelity, particularly in the burned areas, where the original bands introduce significant color distortion. Fit-FC presents the poorest performance in spectral prediction, especially with the original bands. The predicted images using Fit-FC lack important spectral details, appearing flat and devoid of texture, particularly in zones where rapid changes have occurred, such as zone B. Nevertheless, Fit-FC shows notable improvement with the adjusted bands, particularly in zone B with burned areas, where spectral information aligns more closely with the reference image.

The fire that occurred between the two observed dates poses a challenge for all three STF methods, especially in accurately predicting the burned regions. None of the methods fully capture the extent of the fire damage, but the adjusted bands contribute to a slight enhancement in spectral prediction, reducing the extent of the spectral artifacts in these areas. While spatial details are reasonably well-preserved in FSDAF and, to a lesser extent, in STARFM, spectral accuracy remains an issue, particularly with Fit-FC, where the original bands fail to capture the complexity of the changes. In conclusion, the figure demonstrates that the adjusted bands contribute to significant improvements in both spatial and spectral prediction across all three STF methods. The proposed adjustments help mitigate spectral artifacts and improve color fidelity, especially in regions experiencing rapid environmental changes.

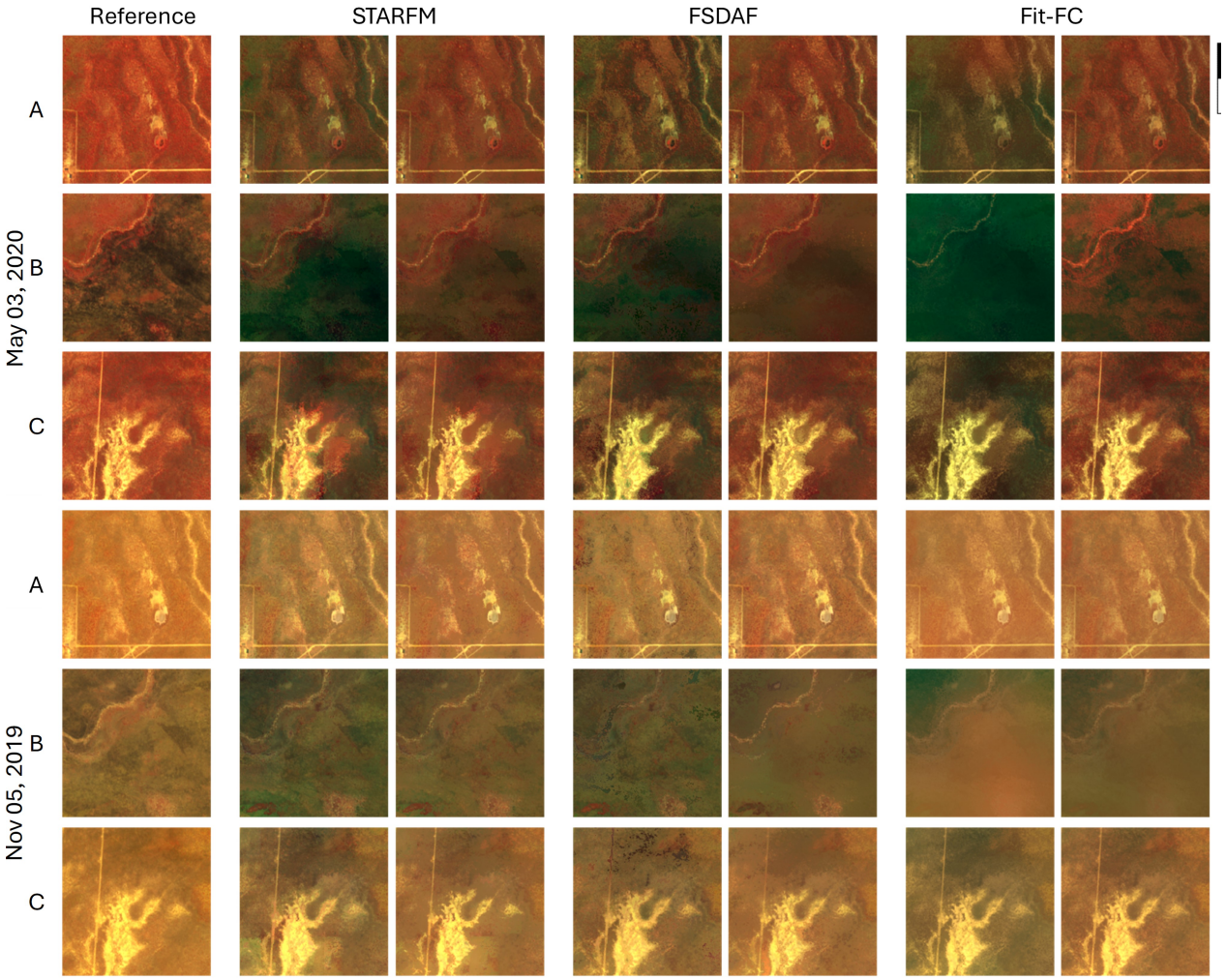


Figure 3.9: Illustrative comparison in Waterbank site for two prediction dates for the three zoom-in area mentioned in Figure 3.7 (200 x 200 S2 pixels).

3.3.2 Scenario 2: Maspalomas Site

The average results of the quantitative evaluation of the Maspalomas site are presented in Table 3.3 similarly to Table 3.2. The evaluation clearly shows that the fusion results obtained with the adjusted bands outperform those produced using the original bands across all STF methods for the entire dataset. Specifically, the adjusted bands yield the best mean results for all three metrics. The RMSE values are consistently lower for the adjusted bands, indicating greater accuracy in the fusion process. Similarly, SSIM scores are higher when using the adjusted bands, demonstrating that these bands better preserve the spatial and structural features of the images. Finally, the SAM values are also improved with the adjusted bands, reflecting a more accurate reproduction of the spectral information. In general, the quantitative evaluation confirms that the proposed adjusted bands significantly enhance the performance of STF methods.

Similarly to Figure 3.6, Figure 3.10 provides a detailed comparison of the performance metrics

Table 3.3: Average of the quality metrics for the whole dataset for the Maspalomas site.

	STARFM			FSDAF			Fit-FC		
	RMSE ↓	SSIM ↑	SAM ↓	RMSE ↓	SSIM ↑	SAM ↓	RMSE ↓	SSIM ↑	SAM ↓
Oa4	0.024	0.934	0.537	0.026	0.926	0.604	0.023	0.927	0.610
Oa5	0.025	0.932	0.551	0.026	0.925	0.621	0.023	0.923	0.633
A-blue	**0.021**	**0.945**	**0.468**	**0.022**	**0.943**	**0.475**	**0.021**	**0.940**	**0.485**
Oa8	0.028	0.922	0.694	0.030	0.915	0.767	0.030	0.902	0.840
Oa9	0.028	0.922	0.695	0.031	0.915	0.768	0.030	0.902	0.843
A-red	**0.022**	**0.939**	**0.592**	**0.023**	**0.938**	**0.605**	**0.023**	**0.920**	**0.678**
Oa16	0.031	0.922	0.723	0.033	0.916	0.786	0.035	0.899	0.910
Oa17	0.031	0.921	0.728	0.034	0.914	0.794	0.036	0.897	0.921
Oa18	0.031	0.921	0.730	0.034	0.914	0.795	0.036	0.897	0.922
A-NIR	**0.027**	**0.931**	**0.661**	**0.027**	**0.932**	**0.675**	**0.029**	**0.909**	**0.790**

A-Blue = adjusted blue, A-Red = adjusted red, A-NIR = adjusted NIR.

between two sets of fused images: those generated using the original S3 bands as input and those using the adjusted bands. This comparison is based on the 40 image pairs evaluated at the Maspalomas site. In the majority of cases, the heatmaps show that the use of adjusted bands leads to superior results. Specifically, the adjusted bands tend to produce lower RMSE and SAM values and higher SSIM values. On average, the adjusted bands reduced SAM by approximately 0.1 and increased SSIM by 0.1. Two specific dates –June 30, 2019, and July 19, 2020– show particularly significant improvements in performance when using the adjusted bands. For these dates, the SSIM increased by 0.2, indicating much closer structural similarity to the reference images. Additionally, there were decreases of 0.04 in RMSE and 0.2 in SAM, reflecting improved accuracy in both the spatial and spectral domains. However, it is worth noting that on February 20, 2020, the STF methods produced better results using the original bands than the adjusted bands, as indicated by the positive difference values in the heatmap. This exception suggests that while the adjusted bands generally improve fusion quality, there may be specific instances or environmental conditions where the original bands yield better performance. This exception could be attributed to the presence of a sandstorm, which are frequent in this study site. Similar to the smoke effect noted in the previous case, sand particles in the atmosphere may have been carried over by the adjustment process, introducing spectral inconsistencies that were not present on the actual prediction date. However, the overall trend observed in Figure 3.10 strongly supports the use of adjusted bands as they typically lead to more accurate and reliable predictions in STF methods.

In Figure 3.11, the predicted images obtained from the different STF methods on June 30, 2019, are displayed, using both the original S3 bands (top row) and the adjusted S3 bands (bottom row). The first row shows that all methods were able to enhance the spatial resolution despite the significant scale difference between the coarse S3 and fine S2 images. The improvement is particularly evident in urban areas, where more detailed structures, such as roads and buildings, are visible. However, when examining the coastal area, the performance of the STF methods is less accurate, as the algorithms struggle to capture the complex spectral signatures of the water and land interface. In the second row, the adjusted bands yield images with better spatial and spectral accuracy, particularly for the STARFM

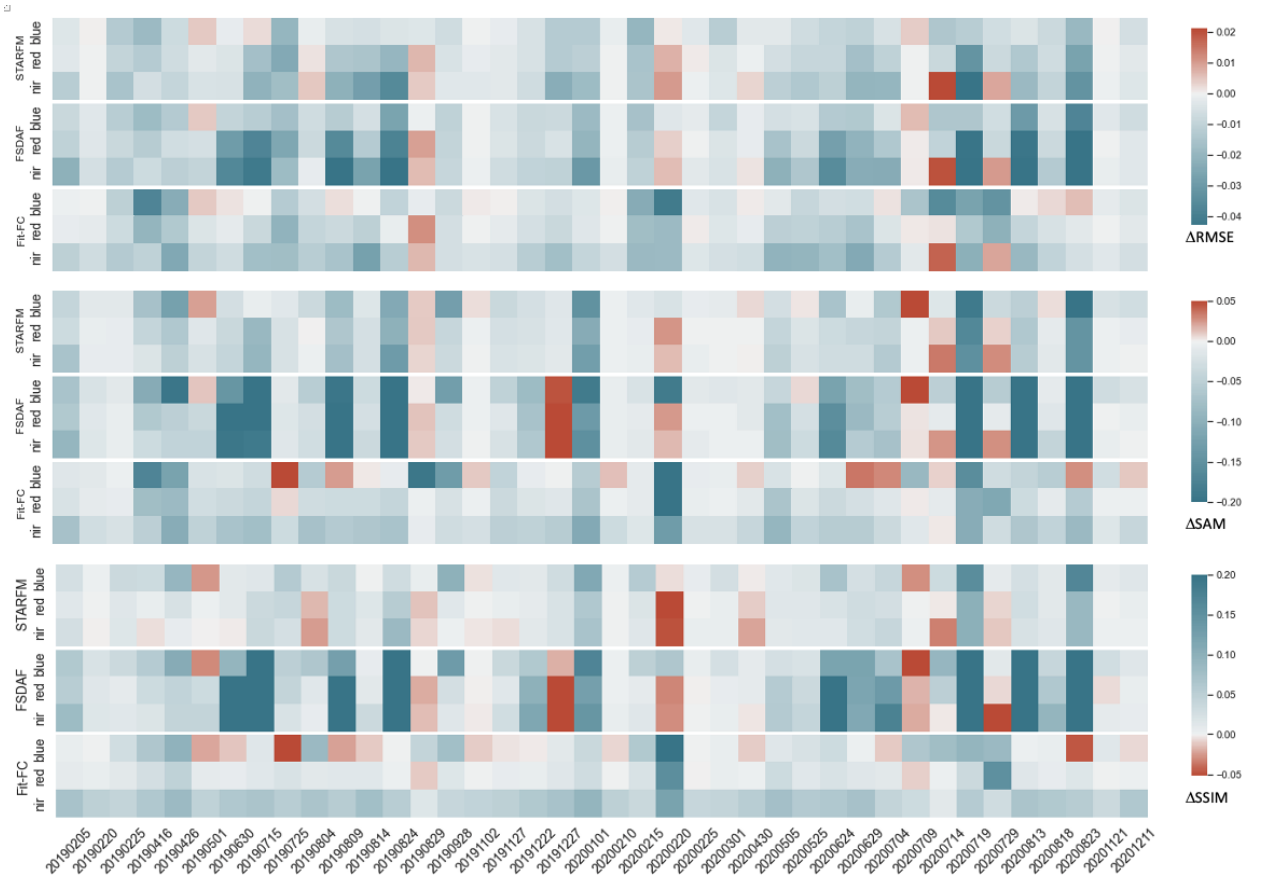


Figure 3.10: Differences between the metrics obtained when evaluating the quality of fused images using as input the adjusted bands and the original bands for the Maspalomas site.

and FSDAF methods. The coastal region, while still challenging, shows slightly improved delineation of land-water boundaries. Urban areas also exhibit better definition, with clearer textures and more pronounced boundaries between different land cover types.

To further quantify the improvements introduced by the adjusted bands, Figure 3.12 shows the differences between the reference image and the fused images. These difference maps reveal the areas where spectral discrepancies are reduced, clearly highlighting the impact of using adjusted bands as input in STF methods. In the first row of each figure, where the original S3 bands were used as input, significant discrepancies can be observed between the fused images and the reference. This is particularly notable in coastal areas and parks, which feature complex land-water boundaries and varied vegetation cover. The regions along the coastline, especially, demonstrate higher deviations, as indicated by the brighter areas in the difference maps. These errors suggest that the original bands struggle to accurately capture the spectral complexity of the landscape, particularly in areas where the contrast between land and water is significant. When examining the second row, which shows the results using the adjusted bands, the improvements are immediately apparent. Across all spectral bands the adjusted bands yield fused images that exhibit smaller discrepancies from the reference, as indicated by the darker (lower difference) regions in the maps. This is

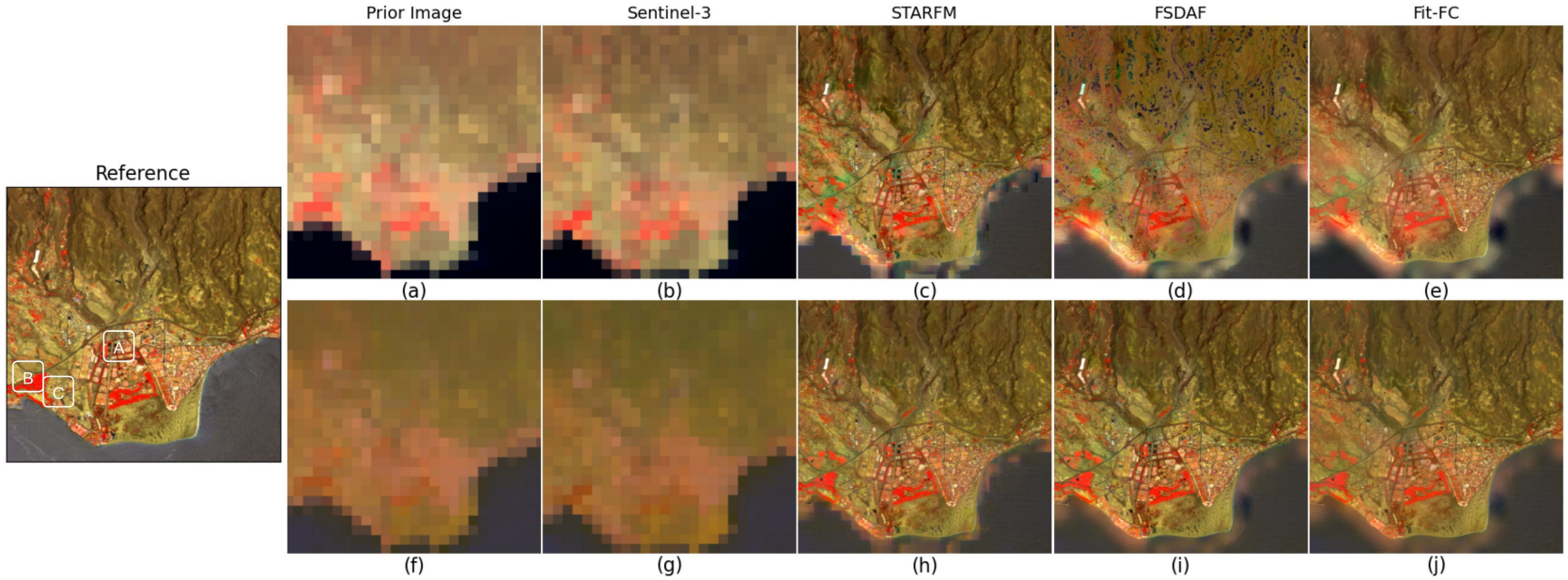


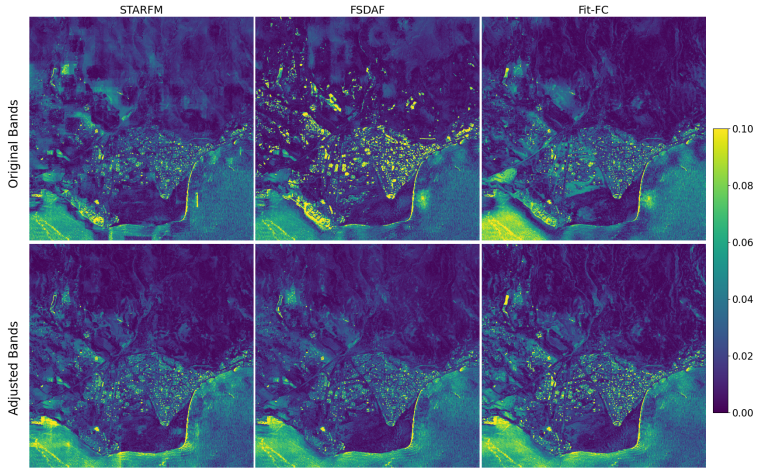
Figure 3.11: Color composition of ground truth (S2), original S3, and STF algorithm predictions for the June 30, 2019. The first row shows S3 original bands, and the second row shows adjusted bands. In each row: (a)/(f) and (b)/(g) represent the prior and prediction dates, while (c)/(h), (d)/(i), and (e)/(j) are predictions by STARFM, FSDAF, and Fit-FC, respectively.

particularly true in coastal areas and parks, suggesting that adjusted bands provide a better representation of the spectral properties of these complex environments. Among the STF methods, FSDAF shows the highest level of improvement when using the adjusted bands. The difference maps for FSDAF show significantly lower discrepancies in both the coastal and vegetated areas compared to the other methods. The reduction in error is particularly evident in the red and NIR bands (Figures 3.12 and 3.12(c)), where FSDAF shows the most substantial alignment with the reference image, as seen by the relatively uniform dark areas in the adjusted band maps. STARFM also benefits from the adjusted bands, though the improvements are less pronounced compared to FSDAF. In the coastal regions, STARFM still shows some level of discrepancy, particularly in the blue band, where the spectral response of water and surrounding land appears more difficult to predict accurately. Adjusted bands reduce these discrepancies.

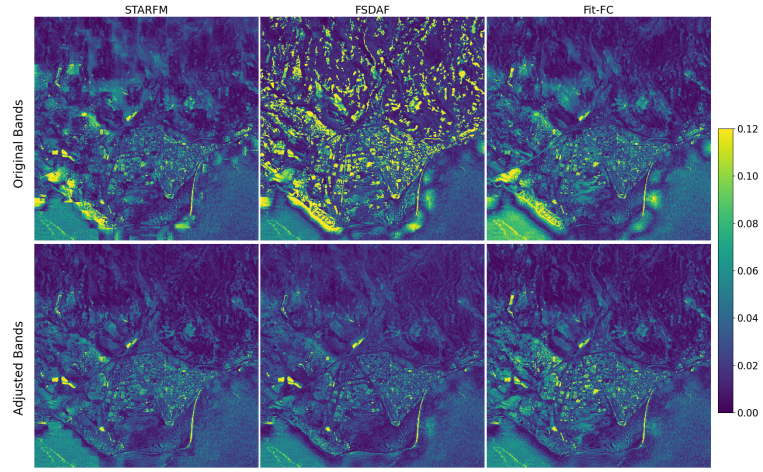
Similarly to Figure 3.9, Figure 3.13 provides a comparative analysis of fusion results for two prediction dates (June 30, 2019, and July 04, 2020) at the Maspalomas site with three zoomed-in areas (A, B, and C) of the site. In the fusion results using the original S3 bands, in general there is a noticeable loss of spatial detail across all three STF methods. The boundaries of buildings, roads, and other fine structures are poorly defined, particularly in zone A, where the blurred edges make it difficult to differentiate between urban and vegetated areas. Additionally, spectral artefacts are visible in several areas. For example, in zones A and C, STARFM produces unwanted greenish tones, deviating from the reference image. FSDAF exhibits incorrect pixel values, and Fit-FC shows noise, particularly in the red regions. This reduces the overall quality of the fusion results, especially in regions with complex land cover. The fusion results using the adjusted S3 bands show a marked improvement in both spatial and spectral quality. The adjusted bands help recover fine textures and restore color accuracy, making the fused images closely resemble the reference S2 images. In zone C, the adjusted bands allow for clearer delineation of buildings and roads, with sharper boundaries that improve the visibility of urban features. The spectral artefacts present in the results for the original bands are corrected. The greenish tones in STARFM disappear, the pixel inaccuracies in FSDAF are minimized, and the noise in Fit-FC is significantly reduced. This results in more reliable predictions across all areas, particularly in the heterogeneous landscapes present in zones A, B, and C. Among the three methods, FSDAF shows the highest level of improvement when using the adjusted bands. The spatial detail in zone C is particularly well-preserved, with more accurate representations of both built and natural environments. While STARFM and Fit-FC also benefit from the adjusted bands, the improvements in FSDAF are the most pronounced. In conclusion, the use of adjusted S3 bands significantly enhances the performance of all three STF methods, improving both the spatial clarity and spectral fidelity of the fused images. This is particularly evident in areas with complex land cover, where the adjusted bands lead to better alignment with reference S2 images.

3.4 Discussion

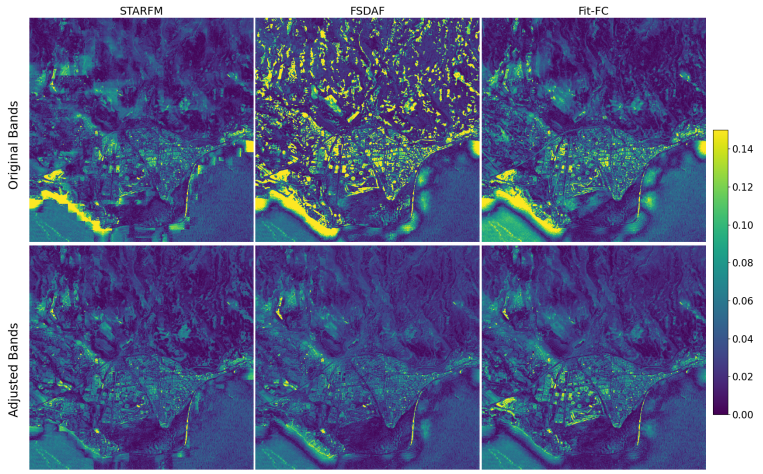
This chapter addresses a critical gap in the field of STF methods by emphasizing the role of SRF in the adjustments and the importance of aligning spectral bands to improve fusion accuracy. Notably, previous research has given limited attention to the challenge of selecting



(a) Band blue



(b) Band red



(c) Band NIR

Figure 3.12: Difference maps between the reference and fused images using original and adjusted S3 bands, with STARFM, FSDAF, and Fit-FC for Maspalomas site

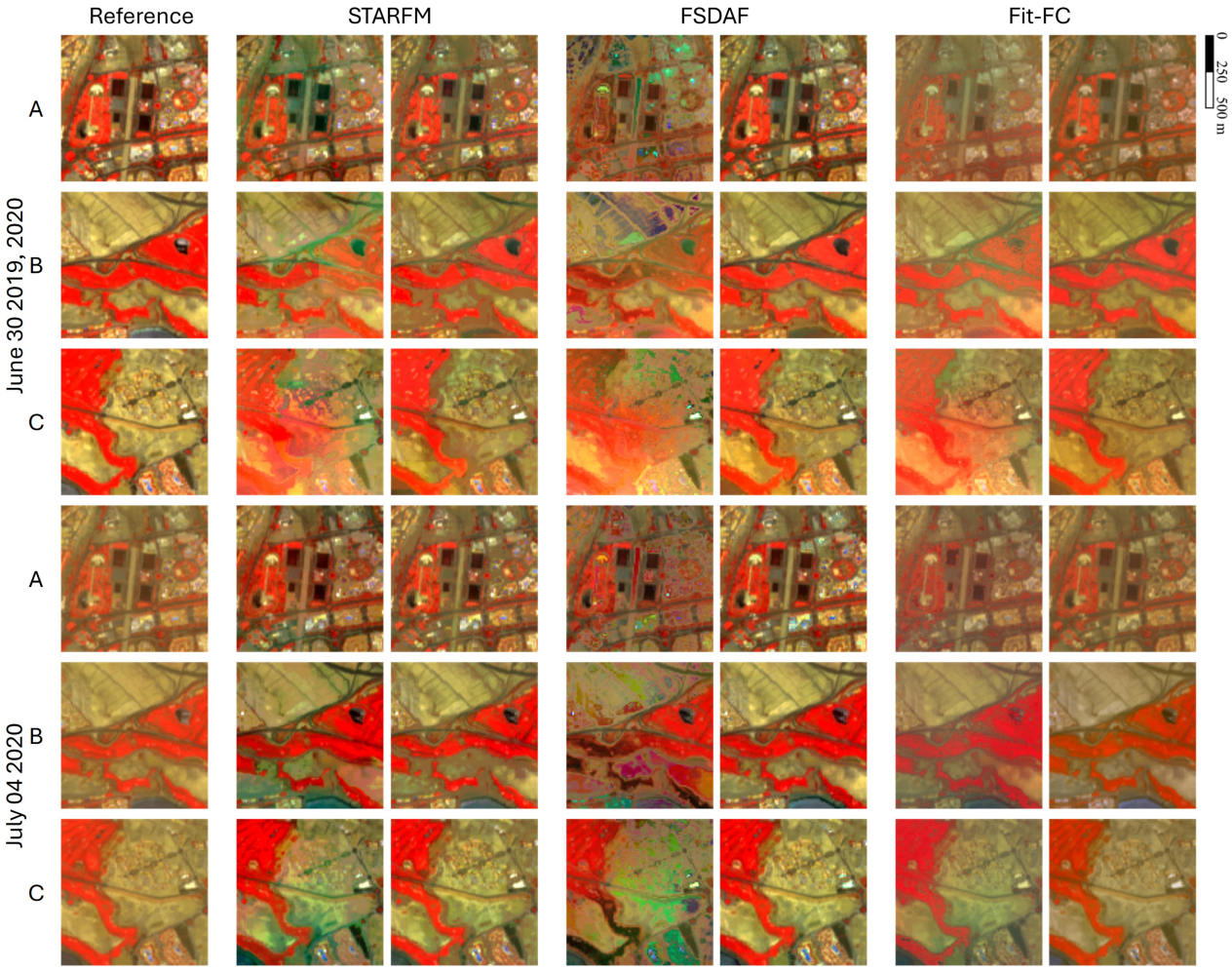


Figure 3.13: Illustrative comparison in Waterbank site for two prediction dates for the three zoom-in area mentioned in figure 3.11(100 x 100 S2 pixels).

the most suitable narrow spectral band when multiple narrow bands overlap with a single, wide spectral band. This common practice of using only one narrow band often neglects valuable spectral and spatial information contained in other overlapping bands, which can increase spectral discrepancies between sensors and, consequently, reduce fusion accuracy. While some previous studies have addressed similar challenges in the context of pansharpening, this study is innovative in applying a comprehensive approach that leverages multiple overlapping narrow bands within the wider spectral range to enhance STF performance. Drawing from this foundational understanding, the study proposed a new preprocessing strategy that aims to harness spectral information from multiple narrow bands rather than limiting the process to one band. By adjusting these bands before using them as an input to the STF methods, this approach seeks to maintain the integrity of both spectral and spatial data without modifying the underlying STF techniques.

To evaluate the effectiveness of the proposed strategy, a series of experiments were conducted under various conditions, including testing with multiple narrow bands of different bandwidths

and employing different STF methods across different sites with distinct challenges. The findings consistently demonstrated that the use of adjusted bands enhances the quality of the fused images, significantly improving the retention of spatial details and reducing spectral distortions. This improvement is particularly crucial in heterogeneous landscapes where maintaining both spatial and spectral integrity can be particularly difficult.

One of the major contributions of this study is the demonstration that the proposed methodology can generate adjusted bands with spectral properties much closer to the reference than the original bands, particularly in the NIR-Red-Blue composition. These adjusted bands enhance the spectral fidelity by incorporating information from several narrow bands and aligning them with the wider spectral bands, reducing the discrepancies caused by the different reflectance values detected by each sensor. This reduction in spectral discrepancies was achieved through the application of regression coefficients that help establish a relationship between the bands of different sensors, such as S2 and S3. These coefficients facilitate the transfer of the spectral information of S3 bands to match the spectral properties of S2.

While some studies in the field have favored simulated data over real data for STF processes, the use of real S2 and S3 data in this study stands out. Simulations, as in the work of (Q. Wang & Atkinson, 2018), were used to evaluate the Fit-FC method in reducing uncertainties caused by substantial disparities in spatial scale and spectral characteristics between sensors.

Another significant finding from this research is the confirmation that the SRF is essential in STF, as it guides the selection of overlapping bands to maximize fusion quality. The SRF's importance lies in its ability to leverage the spectral data from multiple overlapping narrow bands within a wider spectral range, which helps reduce both spectral and spatial discrepancies in the fusion process. By including information from all relevant narrow bands, the STF process benefits from a more comprehensive spectral representation, leading to higher quality fused images. Conversely, relying on only one narrow band when multiple are available reduces fusion accuracy, highlighting the need for a more inclusive approach to utilizing overlapping spectral bands.

The study addresses the challenge of rapid land cover changes, which complicates the accuracy of most STF methods, especially in landscapes where objects smaller than the coarse-resolution pixels introduce prediction errors. Incorporating additional spectral information from multiple narrow bands significantly improved the spectral fidelity of fusion results, effectively reducing artifacts and enhancing fusion accuracy. These findings confirm the critical role of SRF sensitivity in capturing land cover details, especially in heterogeneous environments where traditional STF methods often struggle to resolve small-scale structures. Coarse pixels frequently encompass mixed objects, leading to errors when their reflectance changes are assumed to directly represent finer-resolution features. Adjusted bands mitigate these effects by capturing small spatial structures, improving the representation of features like buildings and roads.

Numerical comparisons further support the effectiveness of adjusted bands, consistently yielding fused images that align more closely with the reference data (S2 at the prediction date) across different STF methods and spectral ranges. Adjusted bands led to substantial improvements in key metrics, with SAM increasing by 16% to 37%, depending on the method

and spectral range. These results indicate that adjusted bands offer a more reliable solution for addressing spectral and spatial inconsistencies in dynamic or complex environments.

Finally, while the current study presents promising results, there is room for further refinement, particularly in the development of algorithms that can better account for temporal variance and spectral differences. The findings underscore the need for more research on the application of the MSTBA approach to other sensors, as the issue of spectral band overlap is common across many remote sensing platforms. Future studies should focus on examining the impact of SRF differences on fusion results and developing more sophisticated techniques for integrating spectral information to achieve higher accuracy in STF methods.

Chapter 4

Multisource Topographic-Enhanced Cloud Removal for Remote Sensing in Mountainous Landscapes

In mountainous environments, remote sensing is indispensable for ecological and hydrological monitoring (S. Yang et al., 2022), disaster management (Zhong et al., 2020), and sustainable resource planning (Tarolli & Straffelini, 2020). The rugged landscapes and dynamic weather conditions typical of these regions require continuous high quality time series data to accurately capture environmental processes such as snow accumulation, vegetation health, and water resource availability (W. Deng et al., 2022; Z. Zhang, 2021). However, cloud cover presents a significant obstacle to optical remote sensing. Frequent cloud interference can lead to substantial data gaps, reducing the temporal resolution and continuity required for reliable monitoring (Jing et al., 2022; Xiong et al., 2022).

Efforts to address cloud interference in remote sensing imagery have resulted in the development of various cloud removal techniques, which aim to reconstruct missing data by integrating information from cloud-free imagery. Traditional cloud removal methods, which span spatial, spectral, temporal, and hybrid approaches, have achieved varying levels of success, especially in the removal of thin clouds and haze (Benabdelkader & Melgani, 2008; Lin et al., 2013; M. Xu et al., 2015). However, these approaches often struggle with persistent and thick cloud cover and complex mountain terrains where the interplay between clouds, snow, and shadows complicates image reconstruction (R. Wu et al., 2023). Moreover, such methods lack the flexibility to adapt to the unique topographic features of mountainous areas, limiting their effectiveness in these environments.

Recent advances in DL have introduced powerful, non-linear approaches for cloud removal, leveraging the pattern recognition capabilities of CNNs and GANs to improve cloud removal precision (D. Ma et al., 2023; Sarukkai et al., 2020; Q. Zhang et al., 2018). By learning intricate patterns and relationships within the data, DL models can outperform traditional methods in distinguishing cloud cover from other landscape features. However, even the most sophisticated DL models face limitations in mountainous regions, where complex topography

and heterogeneous cloud formations present additional challenges to cloud differentiation and reconstruction (Immerzeel et al., 2020).

Recognizing these challenges, this chapter proposes Cloud Removal with Topographic information UNet (CRT-UNet), a novel cloud removal model for mountainous landscapes. By integrating S1 data, S2 imagery, and Digital Elevation Model (DEM) data, CRT-UNet utilizes topographical information to enhance cloud differentiation and improve the model's ability to accurately reconstruct obscured imagery. While some existing models incorporate radar data as auxiliary inputs, mostconventional cloud removal models, which typically rely on optical imagery alone. CRT-UNet extends this by leveraging DEM-based insights into elevation, slope, aspect, and hillshade. This incorporation of topographic data enables the model to better distinguish between clouds, snow, and shadow's features that are often misclassified in complex terrain.

The key contributions of this chapter include the integration of topographic data to enhance cloud removal performance. The proposed CRT-UNet model demonstrates improved accuracy in mountainous regions by effectively addressing the limitations of both traditional deep learning-based methods, which often fail due to the absence of topographic context.

Additionally, CRT-UNet demonstrates robust performance in thick cloud conditions, reconstructing spatial details that reflect underlying terrain features and providing more consistent, high-quality data for environmental monitoring.

4.1 Material & Method

4.1.1 Data Description

The data used in this chapter include S1 imagery, S2 multispectral optical data, DEM and its derived topographic features. While S1 and S2 were previously described in Section 2.1, their specific relevance to cloud removal in mountainous regions is highlighted here. S1, a radar satellite from ESA's Copernicus program, provides consistent, all-weather, day-and-night data, making it valuable for cloud-prone regions. Its dual-polarization capabilities capture surface features like vegetation and moisture, allowing it to complement optical imagery in cloud-obscured areas. S2 offers high-resolution optical data across 13 spectral bands, enabling detailed land-surface analysis. Its bands are critical for detecting vegetation, snow, and cloud properties, while its 10 to 60 m spatial resolution and 5-day revisit cycle ensure rich temporal and spectral data. However, its reliance on clear skies limits its effectiveness in mountainous areas with persistent cloud cover.

Digital Elevation Model

The Shuttle Radar Topographic Mission (SRTM), launched in 2000 by NASA in collaboration with the National Geospatial-Intelligence Agency (NGA), was designed to generate high-resolution global elevation data. Covering approximately 80% of the Earth's land surface, the SRTM produces a DEM with a spatial resolution of 1 arc-second (approximately 30 meters). These data have become a cornerstone for scientific research, offering detailed topographic

information that is widely used in various disciplines (Sun et al., 2003). SRTM DEM data are generated using X-band and C-band Interferometric Synthetic Aperture Radar (InSAR) sensors, operating at a wavelength of 5.6 cm and a frequency of 5.3 GHz. These radar measurements enable precise elevation modeling even in regions where optical measurements are challenging, such as areas with heavy vegetation or persistent cloud cover. As a result, SRTM DEM has become an invaluable resource for applications including geological studies, environmental monitoring, hydrology, urban planning, and disaster management (Mashimbye & Loggenberg, 2023). For mountain regions, the SRTM DEM provides critical insights into terrain elevation and morphology, forming the basis for deriving additional topographic information like slope, aspect, and hillshade. These derivatives are particularly relevant in cloud removal applications, where terrain features play a key role in distinguishing clouds from snow, shadows, and other landscape elements.

Topographic Information

Topographic information derived from DEM data is essential for understanding the physical characteristics of terrain. These features support a range of analyses, from environmental modeling to infrastructure planning, and are crucial for enhancing the model's ability to understand and to process mountainous regions.

Slope Slope measures the steepness or incline of the terrain, expressed in degrees or as a percentage. It is a critical parameter for assessing terrain stability, potential erosion, and suitability for land use. In the context of mountainous regions, slope gradients often vary widely, influencing land accessibility, vegetation dynamics, and the spatial distribution of snow and water (Peng et al., 2008). Slope helps differentiate terrain features, particularly in areas where steep gradients create complex shadow patterns that could be misclassified as clouds or snow.

Here, slope is calculated from the DEM using the algorithm implemented in the GDAL tool gdaldem[1], which is based on the method introduced in (Horn, 1981).

The slope at each pixel is derived using a 3×3 moving window, where the elevation values of the surrounding pixels are weighted to approximate the first-order partial derivatives in the x (east–west) and y (north–south) directions. The formulas used are:

$$\begin{aligned} \frac{\partial z}{\partial x} &= \frac{(z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)}{8 \cdot \text{cell size}} \\ \frac{\partial z}{\partial y} &= \frac{(z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3)}{8 \cdot \text{cell size}} \end{aligned} \tag{4.1}$$

where z_1 to z_9 represent elevation values in a 3×3 neighborhood, ordered as follows:

[1] https://gdal.org/en/stable/programs/gdaldem.html

$$\begin{bmatrix} z_1 & z_2 & z_3 \\ z_4 & z_5 & z_6 \\ z_7 & z_8 & z_9 \end{bmatrix} \tag{4.2}$$

The slope angle in radians is then calculated using:

$$\text{slope} = \arctan\left(\sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2}\right) \tag{4.3}$$

Aspect Aspect refers to the directional orientation of a slope, measured in degrees from 0° (north) to 360° (north again, completing the circle), with 90° representing east, 180° south, and 270° west. This parameter is vital for understanding how solar radiation interacts with terrain, as it influences temperature, moisture levels, vegetation patterns, and snow dynamics (Marsh et al., 2012). In mountainous regions image analysis, aspect plays a significant role in shaping microclimates and ecosystems, and it allows to differentiate snow, shadows, and cloud cover by accounting for illumination angles.

The aspect angle (in radians) can be calculated using the same way the slope was calculated, the same moving window and the partial derivatives in the x (east–west) and y (north–south) directions as in equation 4.1.

Then the aspect can be expressed as :

$$\text{aspect} = \arctan 2\left(\frac{\partial z}{\partial y}, -\frac{\partial z}{\partial x}\right) \tag{4.4}$$

To express the aspect in degrees clockwise from north, the result is converted as follows:

$$\text{aspect (degrees)} = \left(180/\pi \cdot \arctan 2\left(\frac{\partial z}{\partial y}, -\frac{\partial z}{\partial x}\right)\right) \tag{4.5}$$

If the computed angle is negative, 360° is added to obtain a compass bearing in the range [0°, 360°]:

$$\text{aspect (final)} = \begin{cases} 360 + \text{aspect}, & \text{if aspect} < 0 \\ \text{aspect}, & \text{otherwise} \end{cases} \tag{4.6}$$

Hillshade Hillshade is a visualization technique that simulates the appearance of terrain under specific lighting conditions by incorporating shadows and highlights based on slope and aspect. This shaded relief representation improves the visual interpretation of landforms, revealing ridges, valleys, and other surface features that might otherwise be obscured (Van Den Eeckhaut et al., 2005). Hillshade provides context information about surface illumination and shadowing effects, especially in mountainous areas.

This method combines slope, aspect, and the position of the sun—defined by its azimuth (sun direction from north, in degrees) and zenith (sun height above the horizon, in degrees). The hillshade value is computed using the following equation:

$$\begin{aligned}\text{Hillshade} = 255 \cdot \Big[& \cos(\theta_{\text{zenith}}) \cdot \cos(\theta_{\text{slope}}) \\ & + \sin(\theta_{\text{zenith}}) \cdot \sin(\theta_{\text{slope}}) \cdot \cos(\phi_{\text{azimuth}} - \phi_{\text{aspect}}) \Big]\end{aligned} \tag{4.7}$$

With θ_{zenith} is solar zenith angle (in radians), computed as ($90° - \text{altitude}$). θ_{slope} is slope of the terrain (derived from the DEM), in radians, ϕ_{azimuth} is solar azimuth angle (in radians), ϕ_{aspect} is aspect angle, i.e., the direction the slope is facing, in radians.

Both angles can be retrieved from the metadata of S2 images and they are changing based on the acquisition date and time. The result is scaled to a grayscale range from 0 (completely shaded) to 255 (fully illuminated).

4.1.2 Dataset

Study Area

The study area is located in northern Chilean Patagonia, located in southern Chile, and spans seven S2 tiles. It covers a geographical range extending from 41.5°S to 45.5°S in latitude and from 71°W to 74°W in longitude (Figure 4.1). This region is characterized by its unique geographical and climatic conditions and includes several inhabited towns and critical infrastructure. Among these is the "Carretera Austral" (Route 7), a vital north-south highway that traverses the primary valley and serves as a critical connection between Chilean Patagonia and central Chile (Morales et al., 2021). The climate of northern Chilean Patagonia is heavily influenced by persistent westerly winds, commonly known as the westerlies. These winds transport moist air masses from the Pacific Ocean, driving storms against the western slopes of the southern Andes. As a result, the region experiences substantial precipitation, with local rates exceeding 4,000 mm per year in some areas. Mean annual temperatures vary significantly across the study area due to its diverse topography. Temperatures range from approximately 14°C in the northeast to 10°C in the southwest. In the high-altitude regions of the Andes, temperatures drop further, with minimum annual averages around 6°C. These temperature variations are closely tied to elevation and proximity to the Southern Ocean (Sauter, 2020).

The Andes Mountains, in conjunction with the Southern Ocean, create atmospheric conditions conducive to frequent cloud formation. These conditions lead to a high prevalence of both precipitating and non-precipitating clouds, significantly influencing the region's microclimate and ecological processes. The Figure 4.2 illustrates the monthly average percentage of cloud cover across the study area, derived from S2 data spanning 2018 to 2023. The percentage of cloud cover was calculated using the S2 cloud mask, implemented within Google Earth Engine (GEE). The analysis highlights the temporal variability of cloud cover. Lower cloud cover values are observed during the summer months (December to February), ranging between 30% and 40%, while the winter months (June to August) record significantly higher values,

often exceeding 70%. Climatic conditions make frequent the presence of both precipitating and non-precipitating clouds along the years in this region.

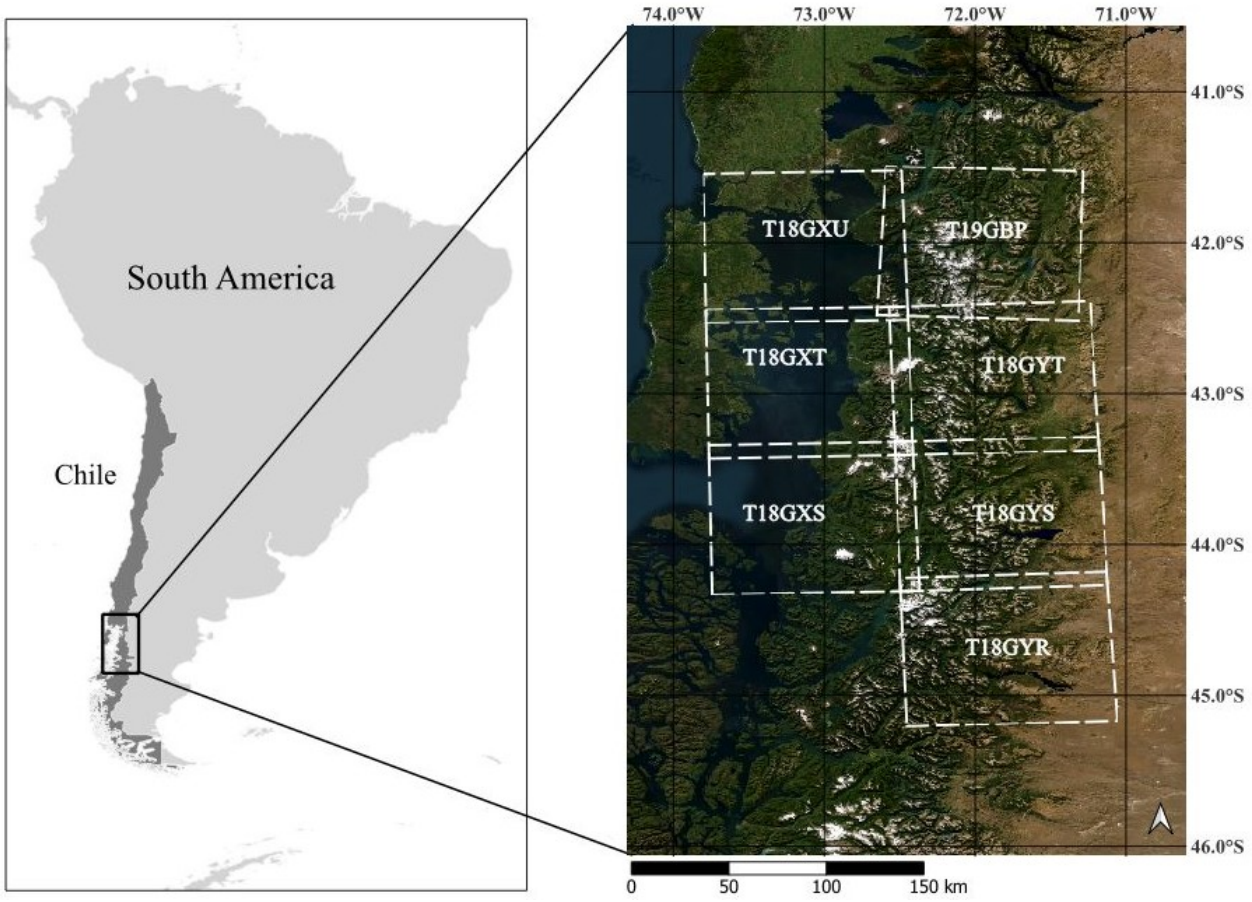


Figure 4.1: The study area covers seven S2 tiles in northern Chilean Patagonia.

Preprocessing workflow

The preprocessing workflow, summarized in Figure 4.3, involved the acquisition, preparation, and processing of S2 and S1 satellite imagery, along with the generation of terrain data. This comprehensive process ensured high-quality, spatially coherent datasets for subsequent analyses. S2 Level-1C data were downloaded from the PEPS platform[2] (Plateforme d'Exploitation des Produits Sentinel) using the EODAG library[3]. Atmospheric correction was performed using the MAJA (MACCS-ATCOR Joint Algorithm) tool (Colin et al., 2023). MAJA played a pivotal role in reducing atmospheric interferences, including aerosols, water vapor and ozone, resulting in precise surface reflectance values. In addition to correcting atmospheric effects, MAJA addressed adjacency and topographic effects, which are essential for accurate reflectance measurements in complex terrain. Furthermore, it generated comprehensive cloud masks that identified thin and thick clouds as well as their shadows, ensuring that only high-quality, cloud-free images were utilized for analysis.

[2]https://peps.cnes.fr/ (last access December 2024)
[3]https://eodag.readthedocs.io/en/stable/

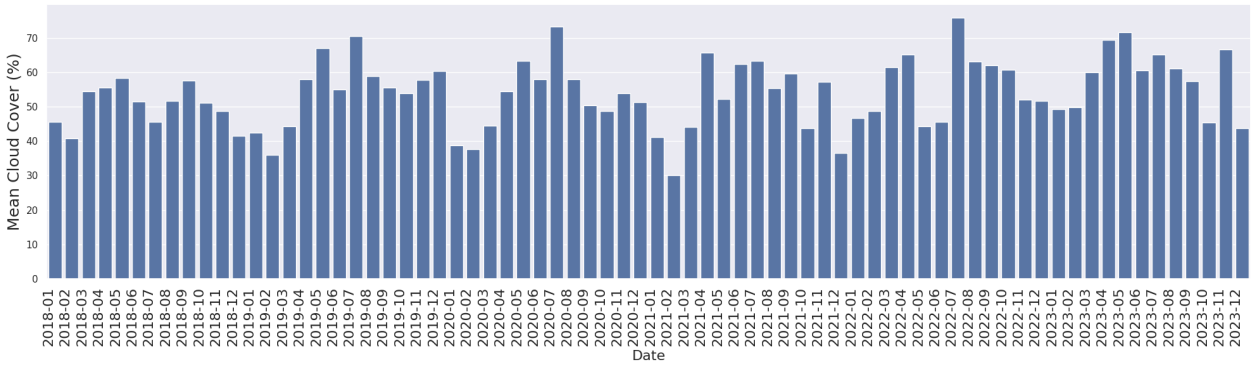


Figure 4.2: Monthly average cloud cover percentage in northern Chilean Patagonia.

S1 imagery was acquired and processed using the S1Tiling tool[4], which automated both data acquisition and preprocessing through ORFEO ToolBox[5]. This tool performed essential tasks such as orthorectification and radiometric calibration, yielding sigma nought (σ^0) values for both VV (Vertical-Vertical) and VH (Vertical-Horizontal) SAR channels. To ensure spatial alignment with the S2 imagery, the S1Tiling tool also reprojected and resampled the S1 images to match the coordinate reference system and the pixel grid of the S2 data. This alignment facilitated seamless integration of datasets, achieving a uniform spatial resolution of 10 m.

To incorporate terrain characteristics, the DEM was downloaded and resampled to a spatial resolution of 20 m. From this DEM, topographic attributes such as aspect, slope, and hillshade were calculated at the same resolution. These attributes enriched the dataset with critical information about terrain variability, providing valuable inputs for analyzing the complex landscape of northern Chilean Patagonia.

Dataset Generation

The dataset generation process involved the integration of S2 and S1 satellite imagery, as well as the DEM data, to create a robust dataset. Careful preprocessing steps and selection criteria were applied to ensure data consistency and high-quality results.

Sentinel-2 Data Preparation S2A and S2B even they are twin satellites but they are operating on slightly offset orbits that result in a temporal difference in image acquisition. Although both sensors are nearly identical, this temporal offset introduces variations in solar illumination between acquisitions. Non-synchronized orbits lead to variations in sun positions, altering shadow lengths, orientations, and overall appearance, especially in hilly or sloped regions. This issue becomes particularly critical in mountainous regions, where complex topography interacts strongly with lighting conditions. Even small shifts in sun angle between acquisitions can cause noticeable differences in the way slopes and valleys are illuminated, making image comparison and data fusion between S2A and S2B more challenging. Such inconsistencies are especially relevant in the context of this study area, characterized by steep slopes, varying elevations, and frequent terrain-induced shadows (F. Chen et al., 2018). To

[4]https://gitlab.orfeo-toolbox.org/s1-tiling/s1tiling (Last access March 2025)
[5]https://www.orfeo-toolbox.org/

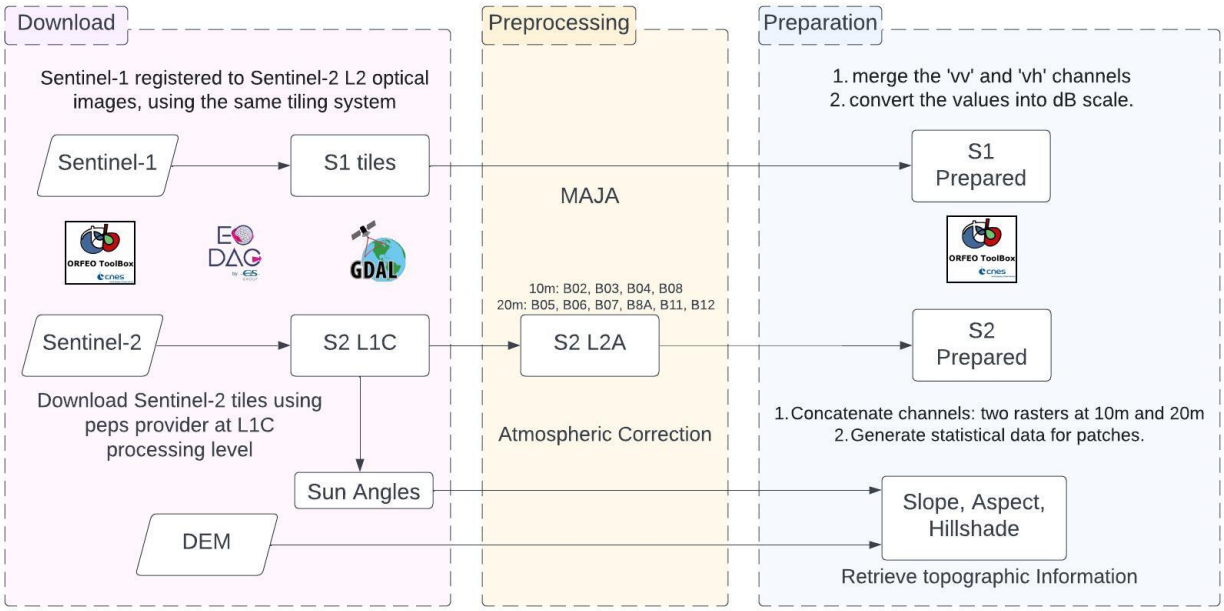


Figure 4.3: Workflow for preprocessing and preparing S1 and S2 data. This workflow involves atmospheric correction of S2 images, resampling from 20 m to 10 m resolution, and merging of bands post-correction. It also includes topographic information extraction from the DEM, such as slope, aspect, and hillshade. For S1 data, the workflow includes tiling to align with the S2 grid system and dB normalization of the VV and VH polarization bands.

address these challenge and minimize potential errors, only S2A data were used with the ascending S1 data.

All S2 bands at 20 m resolution were resampled to 10 m using the nearest-neighbor interpolation technique. This resampling maintained compatibility with other bands while preserving essential spatial information and not introducing new values. Additionally, two supplementary rasters were generated:

- **Cloud Mask**: Produced by the MAJA atmospheric correction algorithm, indicating both thin and thick clouds, as well as shadows.
- **Pixel Validity**: Highlighting cloud-free and cloudy pixels, aiding in the selection of high-quality images for analysis.

Sentinel-1 Data Preparation The preprocessing of the S1 data followed a workflow similar to that of S2. Only ascending passes of S1 imagery were. The two polarization channels, VV and VH, were merged into a single raster, and the raw intensity values were converted to the decibel (dB) scale to enhance interpretability. An additional raster was created to calculate the count of valid and non-valid pixels within the dataset. This allowed for quality assurance and ensured that only valid pixels were used.

Topographic Data Preparation Topographic data was derived from a high resolution DEM. To generate hillshade maps, we consider the lighting and shadow conditions of the study area based on the Sun position (Zenith and Azimuth angles) at the time of S2 image acquisition. These angles were directly obtained from the S2 L1C metadata. This approach avoided reliance on default values, enhancing the realism of the shadow modeling.

Dataset Construction A total of seven S2 tiles from the Chilean Patagonia region, captured between January 2018 and December 2022, were processed, resulting in 34,000 patches. Figure 4.4 illustrates examples of patch triplets, including S2 cloudy, S2 cloud-free, and S1 images, along with DEM and topographic data.

The dataset construction process involved combining S2 cloudy, S2 cloud-free (target), and S1 rasters into tuples. The following criteria were applied:

- Temporal Gaps: A maximum 72-hour gap was allowed between the S1 and S2 image acquisitions, and a maximum 7-day gap was set between the cloud-free and cloudy S2 images (Cresson et al., 2022). This assumption ensured minimal changes in land cover, allowing images to be treated as though they were captured on the same day.
- Patch Selection: Only land pixels were included, excluding water bodies such as seas and lakes. Each selected patch consisted of 100% land pixels with a size of 256x256 pixels.
- Dataset Partitioning: The patches were randomly divided into training (80%, 27315 images), validation (15%, 5167), and testing (5%, 1789) sets, ensuring no overlap between these subsets.

The inclusion of real cloud images in the training set was a critical consideration, particularly given the mountainous terrain of the study area. This strategy allowed the model to effectively learn cloud and shadow identification, significantly improving its performance in challenging regions.

Cloud Cover and Cloud Type Analysis To better understand the variability and challenges presented by the dataset, an analysis of cloud characteristics was conducted across the training, validation, and testing splits. This included both the distribution of cloud types and the range of cloud cover percentages.

Figure 4.5 presents a stacked bar chart showing the distribution of cloud cover percentages across the training, validation, and testing subsets. The x-axis represents cloud cover intervals grouped in 10% bins, ranging from 0–10% up to 90–100%, while the y-axis indicates the percentage of samples within each bin, relative to the total dataset. Each bar is segmented to show the contribution of each subset within that cloud cover range. As shown in the Figure, dataset is dominated by high cloud cover samples, with over 60% of all patches falling within the 90–100% interval. This trend is consistent across all three subsets, with the training set alone accounting for nearly half of the highly clouded scenes. In contrast, lower cloud cover ranges—from 0–10% to 80–90%—are more evenly distributed, each contributing roughly between 3% and 7% of the dataset. This consistency across the lower intervals indicates that all splits include samples with varying degrees of cloudiness, although to a much lesser

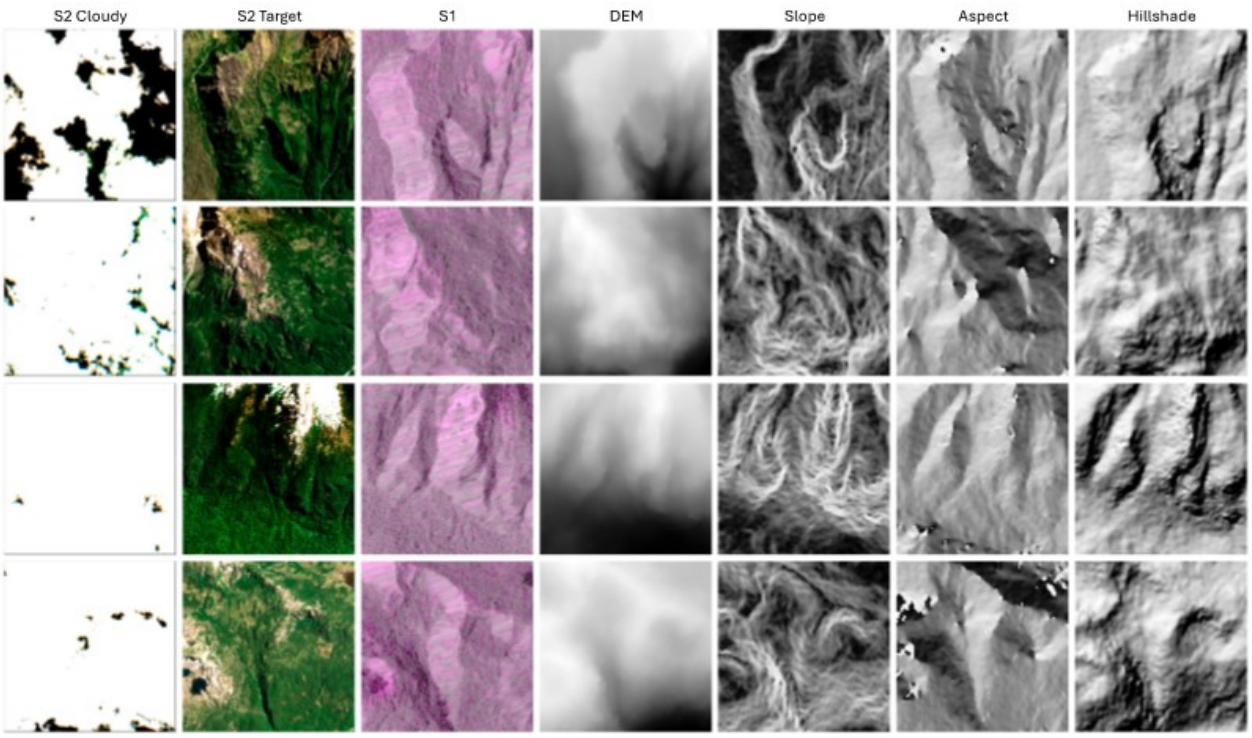


Figure 4.4: Examples of 256x256 pixels patch input from the dataset. The columns exhibit various images: RGB S2 cloudy images and S2 target cloud-free optical images, and a composite of the two polarization channels of S1 (where R = VV, G = VH, B = VV). Additionally, the patch includes grayscale representations of DEM, Slope, Aspect, and Hillshade, with white color indicating higher values.

extent than the highly clouded category. This distribution reflects the real-world conditions of the Chilean Patagonia region, which is characterized by persistent and dense cloud cover, especially in high-altitude areas. Including a high number of heavily clouded samples in all dataset splits is critical for training models capable of handling such challenging environments. At the same time, the presence of low and moderate cloud cover patches enhances the model's ability to generalize and adapt to a broader range of atmospheric conditions.

Figure 4.6 presents a detailed analysis of cloud types across the training, validation, and testing sets, segmented by cloud cover intervals. The figure is structured as a series of stacked bar charts—one for each dataset split—where the x-axis represents cloud cover ranges (in 10% intervals from 0–10% to 90–100%), and the y-axis indicates the average percentage composition of three cloud-related categories: thin clouds, thick clouds, and shadows. Each bar shows how the total cloud-related content in patches of a given cloud cover range is distributed among these three categories. The percentages of each cloud type and shadow were calculated using the cloud mask generated by the MAJA algorithm. This comprehensive cloud mask enabled precise identification and categorization of cloud types, ensuring an accurate analysis of their distribution within the test set.

The top panel of Figure 4.6 illustrates the distribution of cloud types in the training set. In the training set, thick clouds dominate the majority of cloud cover intervals, especially in

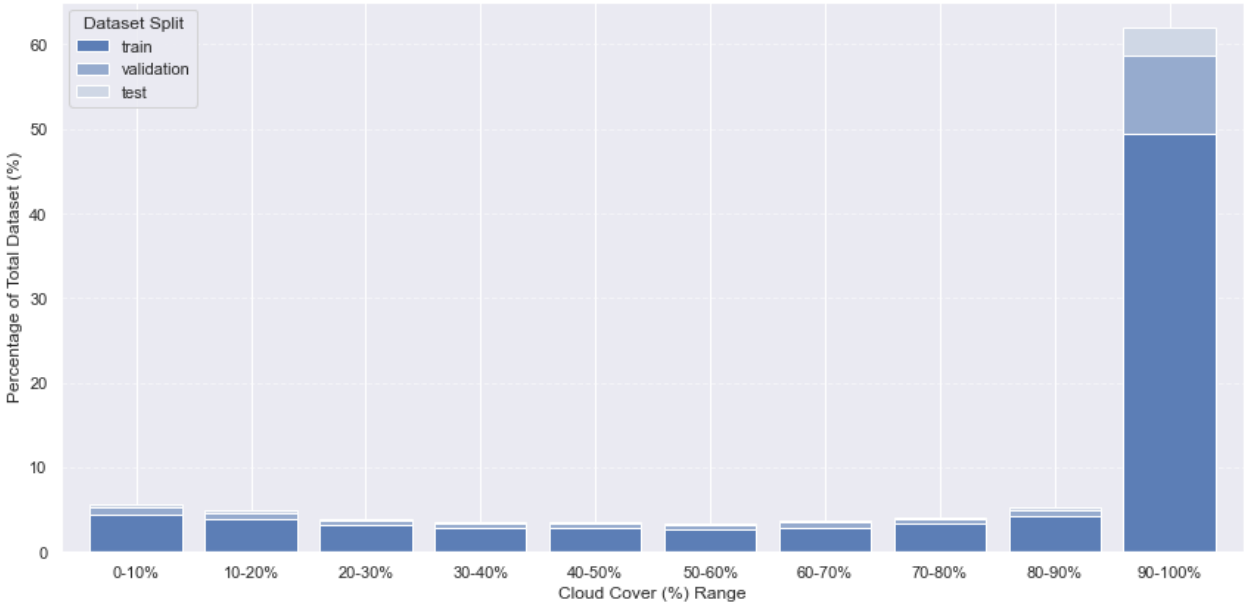


Figure 4.5: Distribution of cloud coverage percentages in the dataset including the three sets (training, testing and validation). The dataset spans a wide range of cloud cover levels, from 0–10% to 90–100%.

the upper range (e.g., 80–100%), where they account for over 60% of the total. Shadows consistently make up a significant portion across all intervals, generally ranging between 30% and 40%, with slight increases around the mid-cloud ranges (40–70%). Thin clouds are present but less prominent, usually contributing less than 15–20%. The middle panel of Figure 4.6 shows the same breakdown for the validation set. The cloud type distribution follows a similar pattern to the training set but with some notable variations in proportions. As with training, thick clouds remain the most dominant class across all intervals, though their share slightly decreases in the 60–90% ranges compared to the training set. Shadows continue to play a substantial role, especially in the mid-to-high cloud intervals, where they sometimes approach parity with thick clouds. Thin clouds maintain a relatively small but consistent presence, again representing the smallest portion of the composition. The bottom panel of Figure 4.6 presents the cloud type breakdown for the test set. It maintains the same structure and intervals as the other two panels. In this set, thick clouds again lead, particularly in the higher cloud cover intervals, where they contribute more than half of the total. Interestingly, shadows appear slightly more prominent in the test set than in the validation set, especially in the 40–70% range. Thin clouds show a modest increase in some lower cloud cover intervals (10–30%), but their overall presence remains limited. This consistent yet slightly varied distribution in the test set ensures that model performance is evaluated not only under heavy cloud conditions but also under more mixed and subtle atmospheric scenarios. This distribution suggests that the dataset is not only cloud-heavy but also dominated by optically dense cloud structures, which are typically more challenging to reconstruct in cloud removal tasks. The consistent presence of shadows—particularly in a mountainous study area—adds further complexity. However, the inclusion of a wide range of cloud and shadow types across all splits supports the development and evaluation of robust

models capable of handling diverse atmospheric conditions in real-world scenarios.

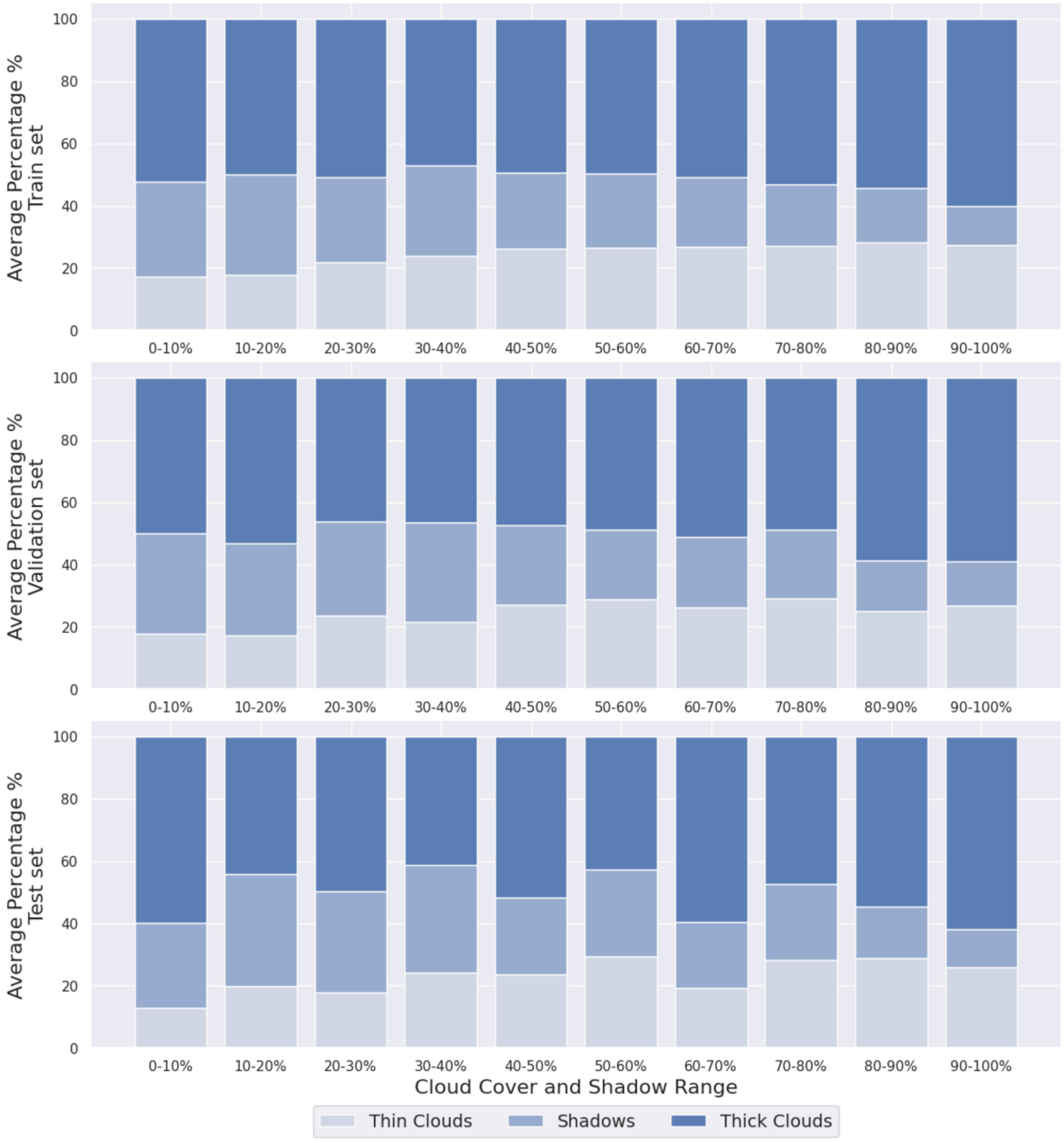


Figure 4.6: Average percentage of cloud types (thick clouds, thin clouds) and shadows across ten cloud cover intervals for the training, validation, and test sets (top to bottom).

4.1.3 Enhanced U-Net Architecture for Mountainous Terrain Cloud Removal (CRT-UNet)

The U-Net architecture, introduced by (Ronneberger, 2015), has established itself as a versatile solution for image segmentation and reconstruction tasks. It features a symmetric encoder-decoder structure that balances global contextual understanding with fine spatial detail preservation. The encoder, or contraction path, extracts hierarchical features through 3x3 convolutional layers followed by ReLU activations and 2x2 max-pooling operations, progressively reducing the spatial dimensions of the feature maps while increasing the feature depth. The decoder, or expansion path, reconstructs the input image resolution by upsampling with transposed convolutions, followed by 3x3 convolutions and ReLU activations. A key feature of U-Net is its use of skip connections, which directly transfer high-resolution features from the encoder to the corresponding layers in the decoder. This mechanism preserves critical spatial details and enhances the model's ability to reconstruct occluded regions, making U-Net a robust foundation for tasks requiring both local precision and global coherence, such as cloud removal (Reddy & Sasikala, 2023).

The ability of U-Net to integrate multi-scale contextual information has been demonstrated in remote sensing applications, where reconstruction quality and structural coherence are critical (Z. Chen et al., 2021; J. Gao et al., 2020). This makes it an ideal baseline for extending reconstruction capabilities in challenging contexts like mountainous terrains, where the interplay of clouds, shadows, and topography demands precise localization and robust feature extraction.

Building on the strengths of U-Net, and inspired by the work of (Cresson et al., 2022) we present CRT-UNet, in Figure 4.7, a U-Net based model to address the challenges of cloud removal in mountainous terrain. These challenges include differentiating cloud cover from terrain-induced shadows, preserving fine terrain details, and maintaining coherence in reconstructed imagery. CRT-UNet introduces several key modifications, most notably the integration of multi-source inputs, including S2 and S1 imagery alongside DEM-derived features such as slope, aspect, and hillshade. This allows CRT-UNet to leverage spectral, radar, and elevation data, improving its ability to handle complex terrains and atmospheric conditions. CRT-UNet employs a dual-input strategy to process two distinct groups of data. In the first block of the encoder, the model processes high-resolution inputs, including four bands of S2 at 10 m resolution, six bands of S2 at 20 m resolution resampled to match the 10 m resolution, which makes a total of 10 S2 bands, and the VV and VH polarization bands from S1. These inputs are aligned spatially and normalized to ensure compatibility, allowing the model to extract critical spectral and structural features needed to identify surface characteristics obscured by cloud cover.

In the second block of the encoder, DEM-derived features—slope, aspect, and hillshade—are introduced. The DEM-derived inputs, at 20 m, are concatenated with the feature maps from the first block, ensuring that the model incorporates terrain morphology and elevation-dependent characteristics into its processing.

The encoder continues to downsample the combined feature maps through convolutional and pooling layers, progressively reducing spatial resolution while increasing feature abstraction.

These multiscale features are passed to the decoder, which uses transposed convolutions to upsample the feature maps back to their original resolution. Additional convolutional layers refine the reconstruction, and skip connections between corresponding encoder and decoder layers ensure that high-resolution spatial details, including terrain features, are preserved. This approach enables CRT-UNet to maintain both global structural coherence and local detail, even in areas with complex terrain. Figure 4.7 provides a detailed illustration of the CRT-UNet architecture, highlighting its dual-input strategy.

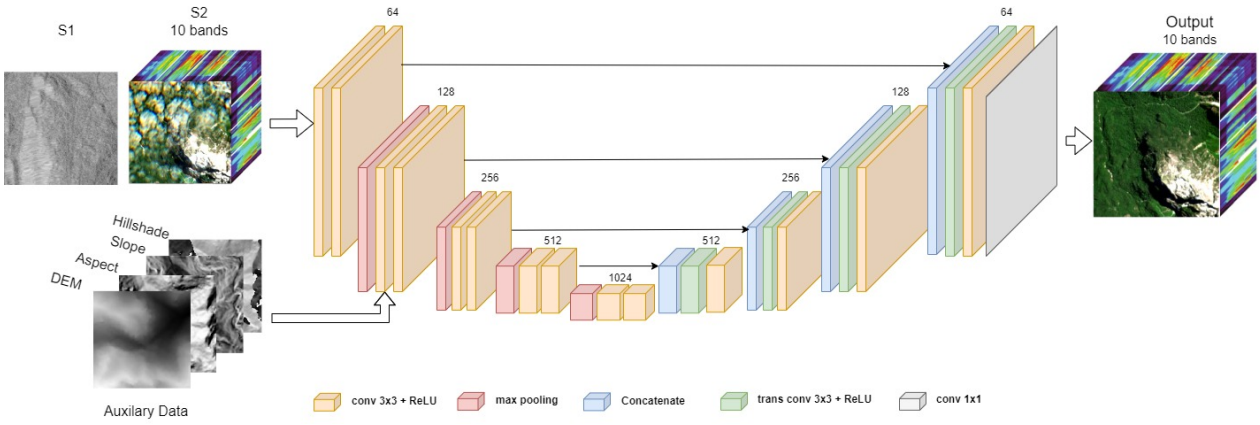


Figure 4.7: Structure diagram of CRT-UNet model

4.2 Experimental Setup

4.2.1 Data Augmentation

To enhance the diversity and robustness of the training dataset, a set of data augmentation techniques was applied. These augmentations techniques were designed to simulate different spatial configurations and orientations, ensuring that the model is generalized well to unseen conditions. The augmentation pipeline included vertical and horizontal flipping, each with a probability of 50%, as well as random 90° rotations, also applied with a probability of 50%. These transformations introduced variability in the spatial orientation of the input data, which is particularly beneficial for satellite imagery that lacks a fixed orientation. The augmentation was implemented using the albumentations.ai[6] library, which efficiently handles multichannel input data. The pipeline was applied consistently across all input channels. By ensuring that all input modalities underwent the same transformations, spatial alignment between the channels was preserved, preventing artifacts or misalignment during training. For validation and testing datasets, no augmentations were applied to ensure consistency in evaluating the model's performance.

Input image values are clippled to remove anomalous pixels and normalizing them to the range [0, 1]. The normalization step is useful for stabilizing the training process and ensuring consistent scaling across input modalities.

[6] https://albumentations.ai (last access March 2025)

4.2.2 Model Training and Optimization

The training of the CRT-UNet was conducted with a systematic approach to ensure effective learning and robust generalization. The model was trained for a maximum of 150 epochs, with the training ending earlier if the learning curve demonstrated stability, indicating convergence. A batch size of 16 was used, striking a balance between computational efficiency and gradient stability.

The optimization process employed the Adam optimizer (Kingma & Ba, 2014), known for its adaptive learning rate capabilities, which efficiently handled the gradient updates. The initial learning rate was set to 1.10^{-4}, and $L2$ regularization with a weight decay of 1.10^{-4} was applied to mitigate overfitting by penalizing large weight magnitudes. The $L1$ loss function was used as the primary loss function to guide the training process. This function minimizes the mean absolute error between the predicted and reference cloud-free images, making it particularly effective for image reconstruction tasks where preserving fine details is crucial. It is defined as:

$$\mathcal{L}_{\text{L1}} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{4.8}$$

where y_i and $\hat{y}_i$ denote the ground truth and predicted pixel values, respectively, and N is the total number of pixels.

The choice of $L1$ loss was motivated by its robustness to outliers and its ability to produce smooth and visually coherent outputs.

The CRT-UNet model was implemented using the PyTorch framework (version 2.1), leveraging its flexibility and efficient handling of GPU resources. All experiments were conducted on a high-performance system equipped with an AMD Ryzen Threadripper 3960X 24-Core CPU, 64 GB of RAM, and an NVIDIA GeForce RTX 3090 GPU. This computational setup provided the necessary resources to process large-scale satellite data and train the model effectively.

4.2.3 Evaluation metrics

In addition to RMSE, SAM, and SSIM were previously defined in Section 3.2.5, where their relevance to cloud removal and image quality assessment was discussed in detail. Peak Signal to Noise Ratio (PSNR) was also used in this chapter to evaluate pixel-level reconstruction accuracy, complementing the other metrics by quantifying the overall similarity between the predicted and reference images in terms of signal strength. PSNR measures the ratio between the maximum possible pixel value and the power of the noise affecting the image (Korhonen & You, 2012). Higher PSNR values indicate a better quality image with less distortion relative to the reference.

$$\text{PSNR} = 10\log_{10}\left(\frac{MAX_{I_{\text{ref}}}^2}{\text{MSE}}\right) \tag{4.9}$$

where $\text{MAX}_{I_{\text{ref}}}$ is the maximum possible pixel value in I_{ref}, and $\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(I_{\text{fused}}(i) - I_{\text{ref}}(i)\right)^2$ is the Mean Squared Error.

4.3 Results

We assessed the performance of our model by comparing it with state-of-the-art models, the baseline UNet-based model described in (Cresson et al., 2022), namely a ResNet model called DSen2-CR introduced in (Meraner et al., 2020), and a GAN-based model called AMGAN presented in (M. Xu et al., 2022). To evaluate the model performance, we used a test set that consists of 1,789 S2 images that cover a wide range of cloud coverage levels, from 1% to 100%. It was used to thoroughly evaluate the performance of CRT-UNet. The results in this section directly reflect the model's effectiveness in handling the diverse cloud conditions captured within this dataset, showcasing its robustness across varying levels of cloud coverage.

Table 4.1 shows the average values of the evaluation metrics calculated across the entire test set. The quantitative results highlight the performance of the CRT-UNet model compared to state-of-the-art cloud removal methods, including baseline U-Net model, DSen2-CR and AMGAN.

CRT-UNet consistently outperformed these methods in all evaluated metrics, demonstrating its effectiveness in addressing the challenges of cloud removal, particularly in regions with diverse terrain and varying cloud coverage levels. Among the evaluated metrics, RMSE and SAM provide a direct measure of reconstruction accuracy, with lower values indicating better performance. CRT-UNet achieved the lowest RMSE (0.069) and SAM (0.267), outperforming the baseline U-Net (0.080 RMSE, 0.279 SAM) and significantly surpassing AMGAN, which struggled with dense cloud cover (0.176 RMSE, 0.322 SAM). These results indicate CRT-UNet's superior capability to reconstruct pixel-level details and preserve spectral consistency. The structural fidelity of the reconstructed images, captured by the SSIM metric, was highest for CRT-UNet (0.823), representing a significant improvement over the U-Net (0.741) and AMGAN (0.466). This improvement can be attributed to CRT-UNet's ability to integrate global and local information through its use of DEM features and skip connections, enabling it to accurately reconstruct terrain features blocked by clouds. Similarly, CRT-UNet achieved the highest PSNR value (27.423 dB), reflecting the model's superior ability to reconstruct high-quality images with minimal noise. In comparison, DSen2-CR and AMGAN, despite their advanced architectures, were limited in their performance (26.507 dB and 19.106 dB, respectively), likely due to their inability to effectively handle the interaction between cloud cover and terrain-induced shadows. These results demonstrate the robustness of CRT-UNet in handling diverse cloud coverage levels and terrain complexities, confirming its effectiveness as a state-of-the-art method for cloud removal.

Table 4.1: Quantitative evaluation of the proposed method CRT-UNet against state-of-the-art approaches in terms of RMSE, PSNR, SSIM, and SAM metrics.

	UNet	DSen2-CR	AMGAN	CRT-UNet
RMSE ↓	0.080	0.086	0.176	**0.069**
SAM ↓	0.279	0.299	0.322	**0.267**
SSIM ↑	0.741	0.726	0.466	**0.823**
PSNR ↑	27.094	26.507	19.106	**27.423**

* Best values are in bold.

4.3.1 Ablation Study

To evaluate the impact of individual input features on the CRT-UNet performance, we conducted an ablation study by systematically removing key topographical inputs: DEM, aspect, slope, and hillshade. Both quantitative and qualitative analyses reveal the essential role of these features in achieving optimal performance. Table 4.2 presents the average values of key metrics—RMSE, SAM, SSIM, and PSNR—calculated across the entire test set, illustrating the performance degradation observed when specific features are excluded. Removing the DEM resulted in the most significant decline, with RMSE increasing from 0.069 to 0.122, SSIM dropping from 0.823 to 0.743, and PSNR decreasing from 27.423 to 24.795. These results underscore the critical importance of elevation information in accurately modeling terrain features and correcting elevation-dependent artifacts, particularly in snow-covered regions where DEM plays a vital role. Aspect, while less impactful than DEM, still showed a noticeable influence on the model's output. Excluding aspect increased RMSE to 0.099, with SSIM and PSNR dropping to 0.762 and 24.737, respectively. These results suggest that aspect is essential for preserving structural integrity in areas where directional lighting significantly affects surface appearance. Slope and hillshade contributed more moderately to the model's overall performance. Excluding slope resulted in an RMSE of 0.087, an SSIM of 0.774, and a PSNR of 25.511. This indicates that slope is particularly useful for modeling steep elevation changes but has a less pronounced effect on flatter terrains. Similarly, the exclusion of hillshade led to a performance decline, with RMSE increasing to 0.099, SSIM decreasing to 0.757, and PSNR falling to 24.752. Hillshade provides critical depth cues, especially in regions with heavy shadow. The proposed model, which incorporates all topographical inputs, achieved the best performance across all metrics, with an RMSE of 0.069, SSIM of 0.823, and PSNR of 27.423. These results demonstrate that each input plays a complementary role in improving the model's accuracy and reconstruction quality.

Table 4.2: Ablation Study Results for the proposed cloud removal model

Method	RMSE ↓	SAM ↓	SSIM ↑	PSNR ↑
Zeroing out DEM	0.122	0.377	0.743	24.795
Zeroing out Aspect	0.099	0.381	0.762	24.737
Zeroing out Slope	0.087	0.353	0.774	25.511
Zeroing out Hillshade	0.099	0.355	0.757	24.752
Proposed Model	**0.069**	**0.267**	**0.823**	**27.423**

* Best values are in bold.

The qualitative results, shown in Figure 4.8, further demonstrate the impact of excluding each input feature on the reconstructed images. The removal of DEM resulted in significantly darker and less detailed images, particularly in regions with high elevation or snow cover. This highlights DEM's role in capturing critical elevation information and correcting terrain-based shading. Excluding aspect produced blurred images with noticeable artifacts in areas where sunlight distribution is strongly influenced by terrain orientation, emphasizing the importance of the slope in preserving structural details under varying directional lighting conditions. Similarly, removing slope led to washed-out images with poorly defined features in steep regions, indicating its role in modeling terrain steepness accurately. Finally, omitting

hillshade caused the images to lose depth and appear flat and less realistic reconstructions, as hillshade contributes to the perception of realistic landforms by providing shadow information. By contrast, the proposed model, incorporating all topographical inputs, produced visually coherent outputs that closely resemble the ground truth.

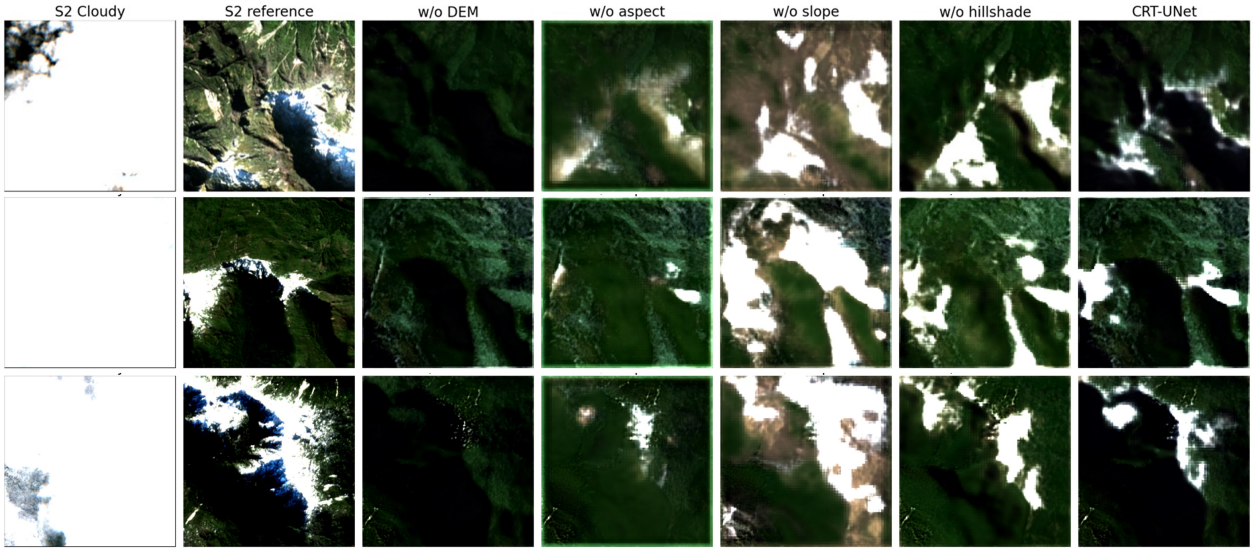


Figure 4.8: Qualitative ablation study across three scenes. For each scene, the images are displayed from left to right as follows: the original cloudy image, the cloud-free reference image, followed by images with the DEM, aspect, slope, and hillshade individually zeroed out (labeled as "w/o DEM," "w/o aspect," "w/o slope," and "w/o hillshade," respectively), and finally, the output from the proposed model. Each image is sized at 256×256 pixels.

These results presented in Table 4.2 and Figure 4.8 conclusively demonstrate the necessity of incorporating all topographical features into the CRT-UNet model to achieve optimal performance. While DEM has the most significant impact due to its ability to capture elevation-dependent variations, aspect, slope, and hillshade each contribute uniquely to structural preservation, error minimization, and realistic shading. The absence of a single feature leads to measurable declines in both quantitative metrics and visual quality, confirming the complementary roles of these inputs. By leveraging these features, the proposed CRT-UNet achieves superior accuracy and reconstruction quality, highlighting the importance of integrating topographic data for effective cloud removal in complex terrains.

4.3.2 Influence of Cloud Coverage Levels

Figure 4.9 presents a set of four line plots showing the performance of four cloud removal methods including DSen2-CR, U-Net, AMGAN, and CRT-UNet across different levels of cloud coverage. Each subplot corresponds to a different evaluation metric. The x-axis in all plots represents the cloud coverage percentage, ranging from 0% to 100%. The y-axis present the metric value.

As shown in Figure 4.9, CRT-UNet consistently outperformed other models (DSen2-CR,

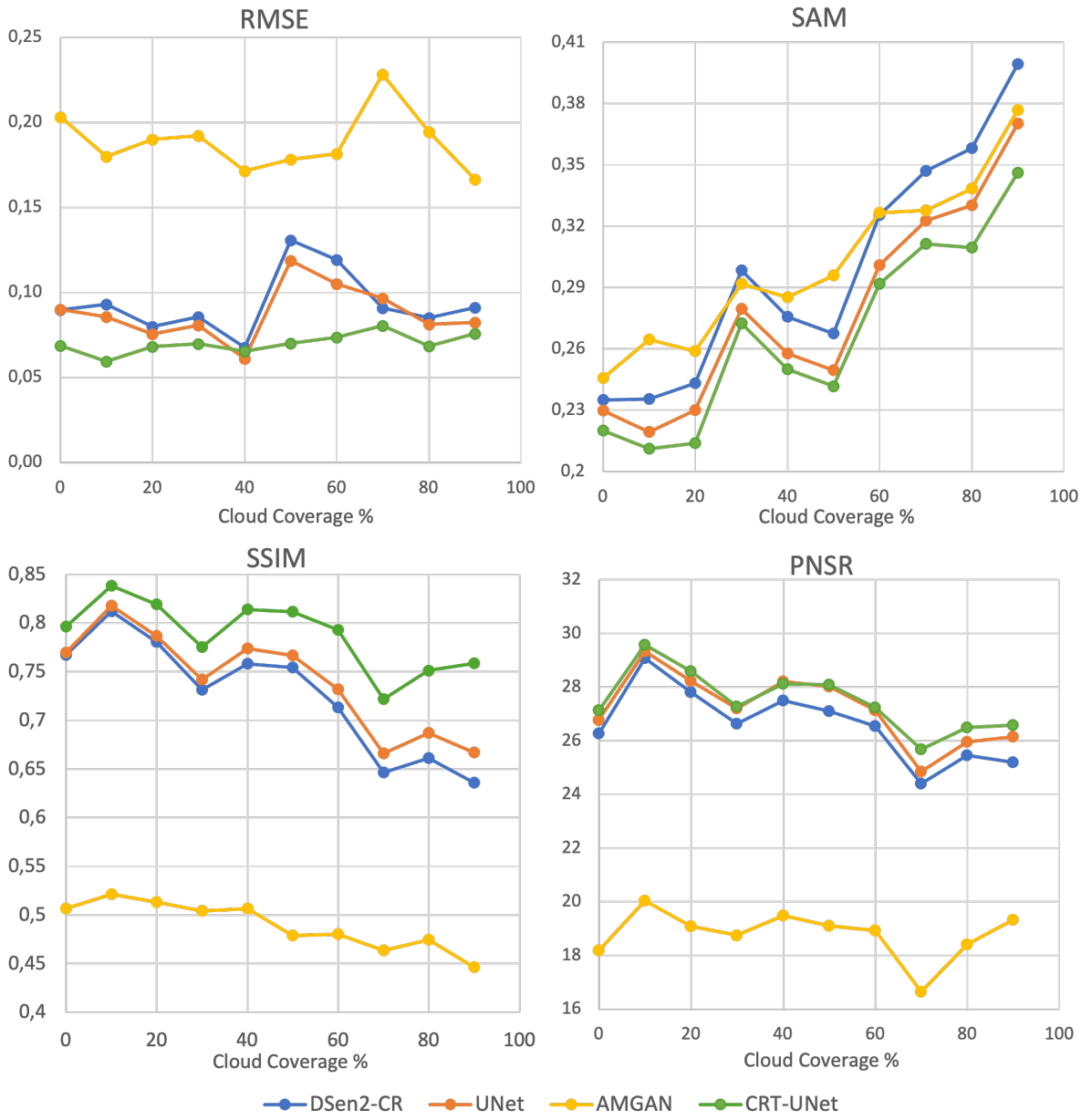


Figure 4.9: Quantitative comparisons of proposed CRT-UNet to state-of-the-art methods on different cloud cover levels.

AMGAN, and the baseline U-Net) across all cloud coverage intervals. For lower cloud cover levels (e.g., 0–20%), CRT-UNet demonstrated exceptional performance, achieving the lowest RMSE and SAM values alongside the highest SSIM and PSNR scores. This indicates that the model can effectively remove clouds and restore underlying terrain features when cloud density is minimal. At medium cloud cover levels (e.g., 40–60%), the performance of all models begins to decline, as clouds obscure more terrain information. Nevertheless, CRT-UNet maintains its superiority by leveraging its integration of topographical inputs and contextual information. Interestingly, an anomaly was observed in the 60–70% cloud coverage range, where all models,

including CRT-UNet, exhibited a noticeable dip in performance. This is reflected in increased RMSE and SAM values, alongside lower SSIM and PSNR scores. As shown in Figure 4.6, this interval has one of the highest proportions of thick clouds (59.55%) compared to other cloud coverage ranges. Thick clouds pose significant challenges for cloud removal due to their high optical density, which completely hides both spatial and spectral information from the underlying landscape. This forces the model to rely heavily on learned patterns and contextual cues from surrounding regions, increasing the likelihood of reconstruction errors.

The predominance of thick clouds in this interval likely intensifies the difficulty of reconstructing accurate cloud-free images. These clouds often occur alongside shadows, further compounding the complexity by obscuring terrain details and reducing the availability of reliable reference information for reconstruction. Additionally, the interplay between thick and thin clouds within this interval introduces heterogeneity in the input data, requiring the model to distinguish between regions of varying opacity. While CRT-UNet's use of DEM and topographical data aids in terrain reconstruction, these challenging conditions reveal limitations in its ability to fully restore spectral and structural fidelity in such scenarios.

At higher cloud coverage levels (e.g., 90–100%), CRT-UNet continues to outperform other models but exhibits a natural decline in performance due to the increased density and opacity of the clouds. Thick clouds dominate in this interval, accounting for 61.78% of cloud types, while thin clouds and shadows decrease significantly. The results highlight the robustness of CRT-UNet in varying levels of cloud coverage, with consistent improvements over other models in both quantitative metrics and visual quality. Overall, CRT-UNet's ability to effectively handle both low and high cloud densities demonstrates its potential for real-world applications in remote sensing, particularly in regions with frequent and diverse cloud cover.

Figure 4.10 illustrates the qualitative performance of CRT-UNet and baseline models (UNet, DSen2-CR, and AMGAN) at different cloud coverage levels, showcasing reconstructed outputs for various scenes. Each row in the figure represents a specific cloud coverage level, ranging from low (36%) to high (100%), and each column displays the cloudy input image, outputs from the baseline models, the CRT-UNet output, and the ground truth cloud-free reference image. At lower levels of cloud cover, such as 36% (Figure 4.10(a)), all models produce visually clear output, but CRT-UNet excels in preserving terrain details and avoiding spectral artifacts. Other models like AMGAN and DSen2-CR introduce some residual noise or subtle distortions, particularly in shadowed regions. CRT-UNet effectively reconstructs cloud-free terrain with high fidelity, closely resembling the reference image. In scenes with moderate cloud coverage, such as 71% (Figure 4.10(c)), CRT-UNet demonstrates significant advantages over the baseline models. While UNet and DSen2-CR manage to remove most clouds, they leave residual artifacts and fail to restore fine spatial details, particularly in complex regions like hills and water bodies. AMGAN struggles further, introducing blurring and spectral inconsistencies across the scene. In contrast, CRT-UNet reconstructs the terrain with minimal artifacts, accurately restoring the structure of vegetation, slopes, and shadows. In cases of high cloud coverage, such as 89% and 100% (Figure 4.10(d), 4.10(e) respectively), CRT-UNet continues to outperform the baseline models, although the task becomes increasingly challenging. DSen2-CR and AMGAN exhibit significant spectral distortions and fail to recover large sections of the terrain covered by dense clouds. UNet performs slightly better but

retains noticeable thin clouds and pixelation effects. CRT-UNet successfully mitigates these issues, producing outputs with greater spectral consistency and clearer spatial details, such as hills and vegetation structures. While some residual errors remain in extremely dense cloud regions, CRT-UNet's outputs closely approximate the ground truth, demonstrating its ability to reconstruct realistic and high-quality cloud-free images even under challenging conditions. These qualitative results underscore the robustness and adaptability of CRT-UNet in handling various cloud coverage scenarios. Using global contextual information and integrating topographical features, CRT-UNet achieves better reconstruction quality compared to baseline models, making it particularly effective for real-world remote sensing applications.

To further assess the robustness of CRT-UNet, its performance was analyzed across individual S2 bands at both 10 m and 20 m resolutions. This band-wise evaluation offers deeper insights into the model's ability to handle varying spectral ranges, as each band is sensitive to specific atmospheric and surface features. By examining the metrics of RMSE, SAM, SSIM, and PSNR across cloud coverage intervals, we can better understand how CRT-UNet adapts to spectral differences and cloud removal challenges.

For the 10 m resolution bands (B02, B03, B04, and B08), Figure 4.11 shows CRT-UNet's consistent superiority over baseline models. In most of the cases, CRT-UNet achieves the best results. As cloud cover increases, RMSE and SAM values rise gradually for all models, indicating the increasing difficulty of removing dense clouds. However, CRT-UNet continues to outperform the baselines, even under high cloud cover conditions. In SSIM and PSNR, CRT-UNet also consistently achieves the highest values, preserving spatial and spectral fidelity effectively across all cloud coverage intervals. Bands B03 and B04 (Green and Red), while performing well overall, exhibit slightly lower SSIM and PSNR values compared to Band B08 (NIR). This can be attributed to the greater atmospheric scattering that affects shorter wavelengths, making cloud removal more challenging. In contrast, AMGAN struggles significantly across all metrics, with severe spectral distortions and reduced reconstruction quality, particularly in intervals of higher cloud coverage.

For the 20 m resolution bands, CRT-UNet demonstrates similar trends, as shown in Figure 4.12. At this resolution, Bands B05 and B06 and Band B12 show particularly better results in term of metrics values, highlighting CRT-UNet's ability to leverage topographical features and contextual information to reconstruct vegetation and terrain accurately. At lower cloud cover levels, the model performs with minimal RMSE and high SSIM and PSNR values. As cloud density increases, CRT-UNet continues to maintain superior performance over baselines, demonstrating its ability to handle challenging scenarios effectively. The SWIR bands (B11 and B12), critical for detecting water bodies and snow cover, particularly benefit from CRT-UNet's capacity to reconstruct features obscured by dense clouds. These bands are often difficult to reconstruct due to the high absorption properties of clouds in these wavelengths, yet CRT-UNet's integration of contextual and spatial information allows it to outperform baseline models. By contrast, AMGAN and DSen2-CR exhibit pronounced spectral distortions, as reflected in higher SAM values, which indicate greater spectral discrepancies. Additionally, these methods fail to recover fine terrain details, especially under high cloud coverage conditions.

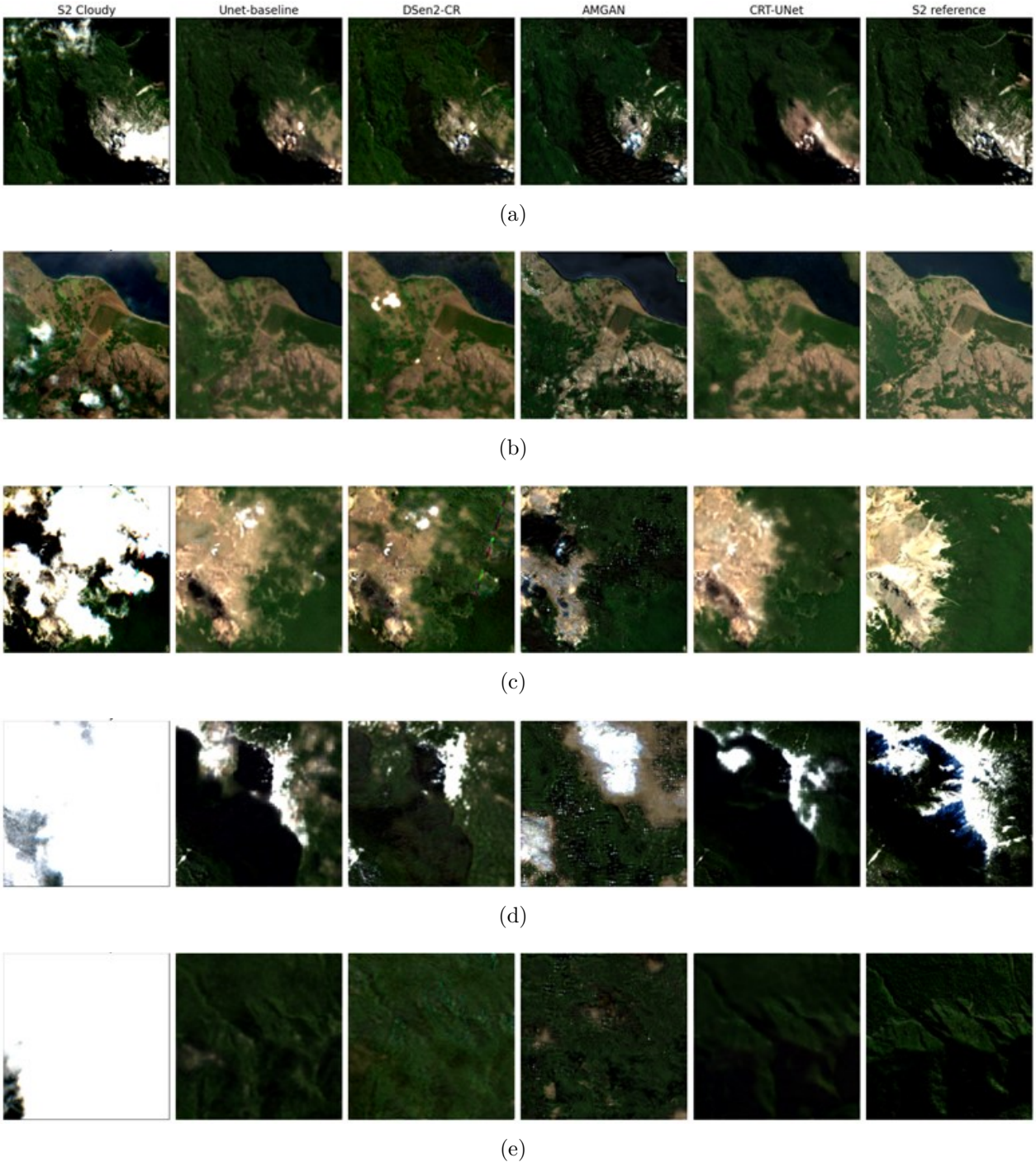


Figure 4.10: Qualitative results of cloud removal models for different scenes with different cloud cover levels. Rows depict cloud cover percentages from top to bottom 36%, 54%, 71%, 89%, and 100%, respectively. Within each scene, the horizontal sequence showcases the cloudy image, UNet, DSen2-CR, AMGAN, CRT-UNet output, and the corresponding reference cloud-free image. All images have a size of 256 x 256 pixels.

The analysis reveals that CRT-UNet performs particularly well in high-wavelength bands,

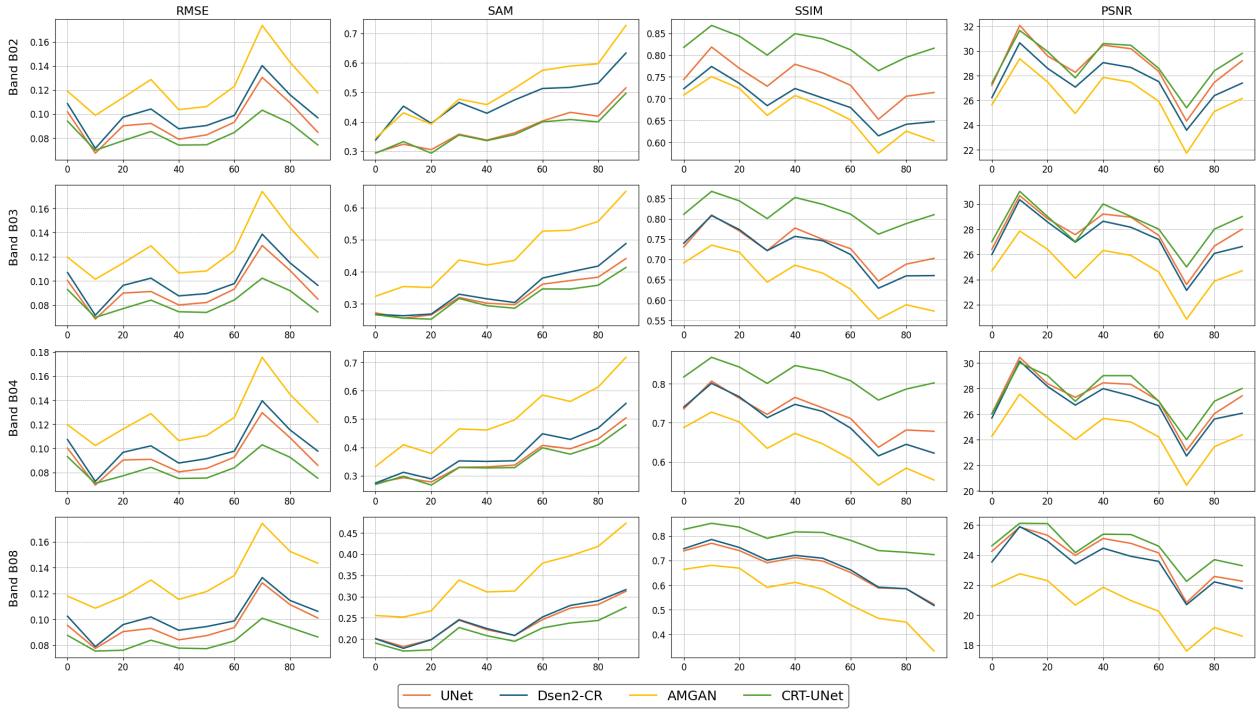


Figure 4.11: Quantitative performance of CRT-UNet and baseline models (UNet, DSen2-CR, AMGAN) across different cloud coverage levels for Sentinel-2 bands at 10 m resolution.

such as NIR and SWIR, which are less affected by atmospheric scattering and provide critical information for vegetation and terrain analysis. Its ability to reconstruct cloud-free images with high structural and spectral fidelity is evident in both the quantitative metrics and its consistent advantage over the baseline models. However, slight performance declines are observed in shorter wavelengths, such as Bands 2 and 3 (Blue and Green), likely due to the increased sensitivity of these bands to atmospheric effects, including scattering and absorption. Overall, the band-wise analysis demonstrates CRT-UNet's ability to handle diverse spectral characteristics, making it highly suitable for remote sensing tasks that rely on accurate spectral and spatial information.

4.3.3 Performance in fully cloudy images

Fully cloud-covered images represent one of the most challenging scenarios for cloud removal models, as the absence of any visible surface information makes accurate reconstruction particularly difficult. In such cases, models must rely entirely on auxiliary data sources such as S1 or topographic information to infer the underlying land surface. These conditions test a model's ability to generalize beyond partial cloud occlusions and to synthesize spatial and spectral patterns without direct visual cues. The complexity increases further in heterogeneous landscapes, where the variation in land cover types, seasonal changes, and atmospheric conditions must be accounted for during reconstruction. This section focuses on evaluating the performance of the models on fully cloud-covered scenes, both quantitatively and qualitatively, to assess their capacity for reconstructing plausible and coherent outputs in the absence of

Figure 4.12: Quantitative performance of CRT-UNet and baseline models (UNet, DSen2-CR, AMGAN) across different cloud coverage levels for Sentinel-2 bands at 20 m resolution.

optical surface data.

This section focuses exclusively on evaluating model performance in the most challenging scenarios images with cloud cover between 90% and 100%. These fully cloud-obscured scenes provide little to no optical information, making reconstruction heavily reliant on auxiliary inputs and the model's generalization capability. To assess performance under these extreme conditions, we computed the average values of four key evaluation metrics for all test samples falling within this cloud cover interval. Table 4.3 summarizes the quantitative results, enabling direct comparison between CRT-UNet and baseline methods on severely clouded inputs.

For the fully cloudy subset, CRT-UNet achieves the best performance across all four metrics. It records the lowest RMSE (0.075) and SAM (0.346), indicating more accurate pixel-level and spectral reconstruction, respectively. Similarly, it yields the highest SSIM (0.758), reflecting superior structural similarity with the reference images, and the highest PSNR (26.593), suggesting better signal fidelity. These results highlight the model's ability to effectively

Table 4.3: Quantitative evaluation of the proposed method CRT-UNet against state-of-the-art approaches in terms of RMSE, PSNR, SSIM, and SAM metrics for the fully cloudy images.

	UNet	DSen2-CR	AMGAN	CRT-UNet
RMSE ↓	0.082	0.091	0.166	**0.075**
SAM ↓	0.370	0.399	0.376	**0.346**
SSIM ↑	0.667	0.635	0.446	**0.758**
PSNR ↑	26.162	25.210	19.328	**26.593**

* Best values are in bold.

reconstruct plausible and coherent outputs in the absence of visible surface information.

To complement the quantitative analysis, Figure 4.13 presents visual examples of fully cloud-covered scenes from the test set, each with a cloud coverage of 90–100%. These examples illustrate the model's ability to reconstruct surface features in the complete absence of visible optical data. In some cases, the underlying terrain includes snow-covered areas, which further complicate the reconstruction due to their spectral similarity with clouds for some bands.

The first example in Figure 4.13(a) showcases a snow-covered peak obscured by thick clouds and shadows. CRT-UNet successfully reconstructs the snow-covered terrain with minimal artifacts, accurately capturing the slopes and shadows, while AMGAN introduces severe spectral distortions and fails to remove all cloud features. DSen2-CR, while removing most clouds, leaves residual artifacts and does not restore fine spatial details in snow-covered areas. The baseline UNet produces better results than AMGAN and DSen2-CR but retains a thin cloud and loses the structural integrity of the slopes. The second example in Figure 4.13(b), featuring dense vegetation and cloud-shadow interactions, CRT-UNet outperforms the baseline models by reconstructing the hillside with high fidelity and preserving vegetation structures. The AMGAN model fails to reconstruct the hillside altogether, while DSen2-CR introduces blurring and spectral inconsistencies. UNet, though performing slightly better, suffers from pixelation effects and fails to fully restore the finer details of the vegetation. Figure 4.13(c), with dense clouds obscure the majority of the terrain, CRT-UNet demonstrates its ability to handle extreme conditions. The model restores critical terrain features, such as slopes and vegetation cover, with greater spectral and structural consistency than the baseline models. While some residual artifacts remain in regions of dense cloud cover, CRT-UNet's outputs closely approximate the ground truth and maintain the overall coherence of the reconstructed scene.

These qualitative results underscore CRT-UNet's adaptability and effectiveness in handling the most challenging cloud cover conditions. In fully cloud-obscured scenes—where the optical signal is entirely blocked, CRT-UNet consistently outperforms baseline methods by reconstructing coherent and spectrally plausible outputs. The model is able to preserve key surface patterns, mitigate spectral distortion, and enhance visual consistency, even in the complete absence of visible input. This capability is particularly valuable for remote sensing applications in persistently cloud-covered regions, where conventional optical methods fail to provide usable observations.

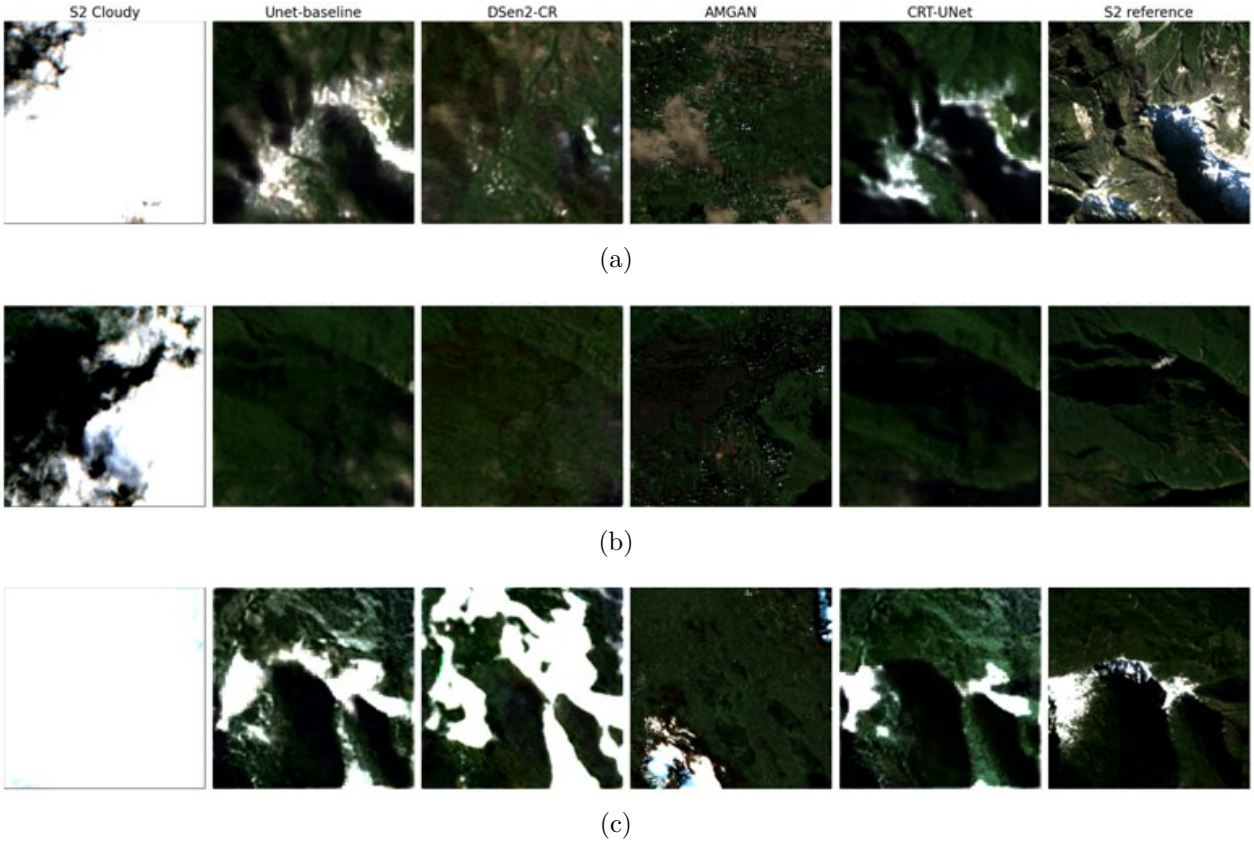


Figure 4.13: Exemplary results of three different mountainous scenes. Rows: three different samples with mountain. Columns : S2 Cloudy, UNet, DSen2-CR, AMGAN, CRT-UNet outputs and S2 reference

CRT-UNet's enhanced performance in fully cloudy conditions can be attributed to its integration of auxiliary inputs, particularly S1 and topographic data. The SAR backscatter provides structural and moisture-related information independent of cloud presence, while DEM-derived features offer spatial context such as elevation gradients and shadow-prone areas. DEM provides critical elevation information for understanding terrain variations, enabling the model to reconstruct snow-covered peaks and steep slopes with high precision. Slope inputs capture the gradient information necessary for modeling sharp elevation changes, while hillshade enhances the model's understanding of shadow patterns caused by terrain features. These inputs work in synergy to allow CRT-UNet to differentiate between clouds, shadows, and terrain elements, resulting in reconstructions that closely match the ground truth. In contrast, baseline models lack access to such contextual information, limiting their capacity to reconstruct scenes with dense or complete cloud coverage.

4.3.4 Large-Scale Scene Reconstruction

Large-scale scene reconstruction is a critical benchmark for evaluating the practical applicability of cloud removal models in real-world scenarios. Unlike small patches, which are often used for testing, large scenes encompass diverse terrain types, complex cloud structures, and

varying atmospheric conditions, making them more representative of operational challenges. These scenes test a model's ability to maintain spatial coherence, preserve spectral fidelity, and ensure consistent reconstruction quality over extended areas. For applications such as agriculture, disaster management, and environmental monitoring, the ability to reconstruct large-scale images accurately is essential. Cloud removal in these scenarios must integrate terrain features, minimize artifacts, and recover spectral information across multiple bands to provide reliable inputs for downstream analysis. This section evaluates CRT-UNet's scalability and robustness in handling large-scale S2 scenes, highlighting its effectiveness in producing high-quality cloud-free outputs over extensive regions.

Figure 4.14 presents examples of CRT-UNet's performance on large-scale S2 scenes, covering approximately 20 $km \times 20$ km. The results are shown in both RGB (top-row) and false-color compositions using bands B12, BA8, and B06 (bottom row) to highlight the model's ability to reconstruct cloud-free images while preserving essential spectral and spatial details. Each column includes the input cloudy image, CRT-UNet's predicted output, and the ground truth cloud-free reference. The first column represents the input cloudy image, where thick cloud cover obscures much of the landscape, rendering critical surface features such as vegetation, water bodies, and terrain details difficult to discern. Dense cloud patches, thin clouds, and shadows are prominent in these images, presenting significant challenges for cloud removal. The second column displays CRT-UNet's predicted output after cloud removal. In the RGB composition, the model successfully eliminates the cloud cover, revealing underlying features with remarkable clarity. Key elements such as forested areas, valleys, and water bodies such as lakes are reconstructed with a high level of spatial detail and minimal artifacts. The restored snow cover is particularly evident in high-altitude regions, where the model captures the distinct brightness and structural consistency of snow, closely matching the patterns in the ground truth image shown in the third column.

In the false-color composition (B12, BA8, B06), the results further emphasize CRT-UNet's effectiveness in recovering spectral information critical for remote sensing applications. Vegetation and water bodies, which are essential indicators for environmental and agricultural monitoring, are reconstructed with high spectral fidelity. For example, areas previously obscured by dense clouds now reveal well-defined vegetation indices and water features that are consistent with the ground truth. Snow-covered regions are accurately reconstructed, reflecting their distinct spectral properties in SWIR and NIR bands. These bands are particularly sensitive to snow and water, allowing the model to differentiate between snow, vegetation, and shadows with high precision. The model preserves the spectral integrity of snow-covered slopes, ensuring accurate representation even in areas initially obscured by dense clouds. The predicted outputs demonstrate CRT-UNet's ability to accurately generate all 10 m bands of S2 imagery, effectively managing landscape variability and dense cloud coverage. CRT-UNet consistently produces seamless, high-quality cloud-free images across both RGB and false-color compositions, preserving fine details and spatial coherence. These results underscore its suitability for large-scale remote sensing applications requiring accurate and reliable reconstructions.

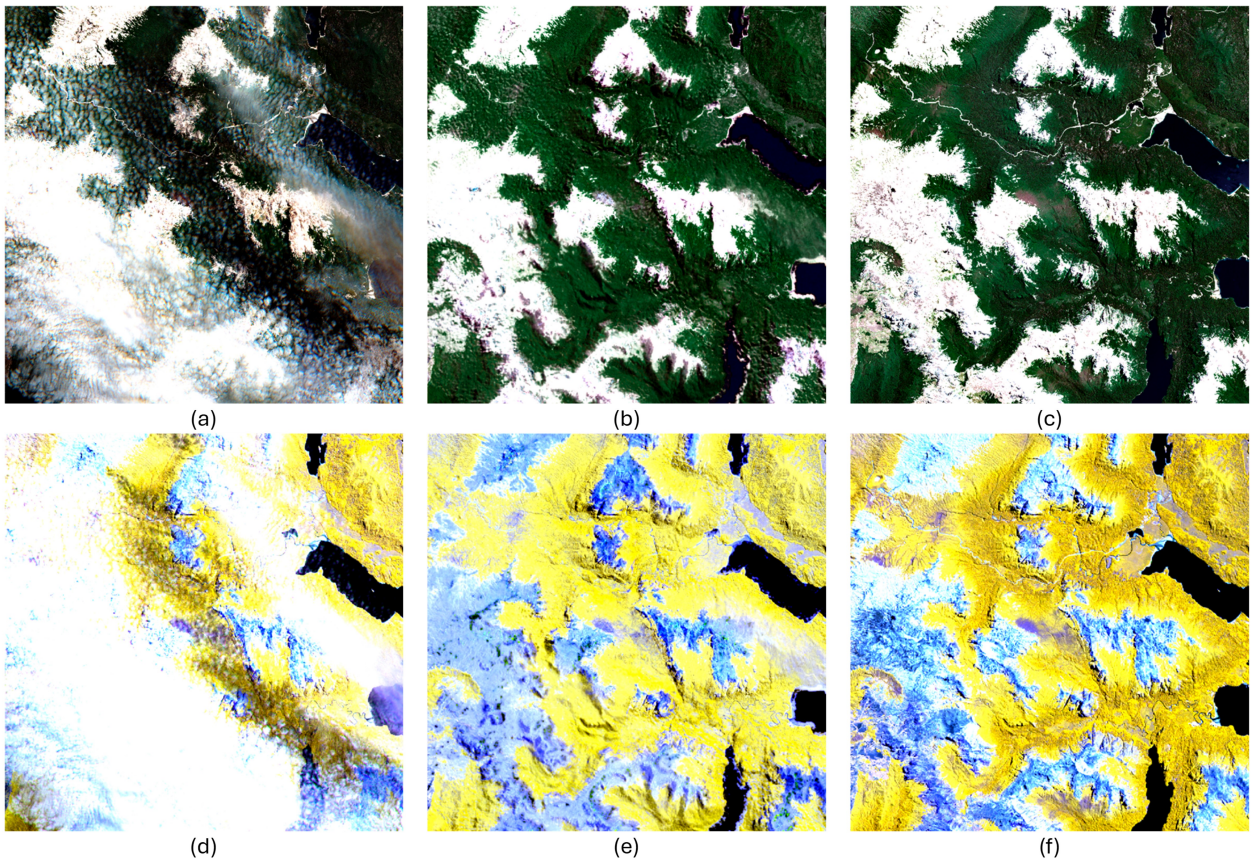


Figure 4.14: Example results using CRT-UNet on large-scale S2 images in RGB (a, b and c) and false-color (B12, BA8, B06) compositions (d, e and f). Columns represent the cloudy input (left), predicted cloud-free output (middle), and target image (right). Size of the image is 20 km x 20 km.

4.4 Discussion

Cloud removal in optical remote sensing data remains a significant challenge, particularly in mountainous regions characterized by complex topographic and atmospheric conditions. In this context, CRT-UNet was introduced as a novel model to address these challenges by leveraging multi-source data, including S2 and S1 imagery, alongside DEM-derived topographic features such as slope, aspect, and hillshade. By integrating these complementary inputs, CRT-UNet demonstrated its ability to reconstruct cloud-covered areas effectively while preserving critical terrain features, even in the most challenging scenarios.

The model's performance was validated using a comprehensive test dataset with varying cloud coverage levels, consistently outperforming state-of-the-art methods such as DSen2-CR and AMGAN across multiple quantitative metrics. The quantitative evaluation results provided robust evidence of CRT-UNet's superiority. Metrics such as RMSE, SSIM, and SAM consistently demonstrated the model's effectiveness in reducing errors, preserving structural details, and maintaining spectral consistency. Visual assessments emphasizes these findings, highlighting CRT-UNet's ability to handle challenging terrains and restore features accurately, even under

dense cloud coverage. The comparison with baseline models underscored CRT-UNet's unique advantages, particularly its terrain-aware design, which enabled superior performance in regions where traditional models struggled. These improvements were particularly evident in complex terrains, such as steep slopes, shaded valleys, and snow-covered regions, where the inclusion of elevation-dependent information proved critical for accurate reconstructions.

One of the key strengths of CRT-UNet lies in its integration of topographical inputs. The inclusion of DEM data enabled the model to differentiate between clouds, shadows, and terrain features, facilitating the restoration of features such as snow-covered slopes and shaded valleys. The use of slope, hillshade, and aspect further enhanced the model's ability to recover depth and spatial consistency, ensuring realistic reconstructions even in areas with complex morphology. This capability is particularly valuable for environmental applications such as snow cover monitoring, where CRT-UNet's ability to accurately estimate snow-covered areas supports downstream tasks like streamflow prediction and seasonal snowpack assessment.

Beyond patch-based evaluations, CRT-UNet also demonstrated robustness and scalability in large-scale scene reconstructions. The model maintained spatial coherence and spectral fidelity across extensive areas, effectively removing dense clouds while recovering features such as vegetation, water bodies, and urban structures. Its ability to preserve spectral information in SWIR and NIR bands, critical for vegetation indices and hydrological analyses, further highlights its operational applicability. Unlike baseline models, CRT-UNet consistently avoided spectral distortions, residual cloud artifacts, and incomplete reconstructions, making it suitable for a wide range of remote sensing applications, including agricultural monitoring, disaster response, and environmental conservation.

While CRT-UNet delivered notable advancements, it also exhibited minor limitations. In some cases, pixelation artifacts were observed, particularly in regions with intricate textures or where resolution mismatches between input bands affected the spatial quality of the outputs. Additionally, extremely dense cloud conditions posed challenges, as limited spectral and spatial information hindered complete reconstruction. Future refinements could focus on integrating advanced techniques such as attention mechanisms, domain-adaptive weighting, or higher-resolution datasets to address these issues and further enhance the model's performance.

In conclusion, CRT-UNet represents a significant advancement in cloud removal methodologies for remote sensing. By integrating multi-source data and leveraging a robust architectural design, it addresses the spectral and spatial complexities inherent in cloud-covered imagery, particularly in mountainous terrains. The model's demonstrated accuracy in snow cover estimation and its ability to scale to large scenes make it a valuable tool for applications such as agricultural monitoring, hydrology, and climate impact analysis. Future developments should aim to address the identified limitations, expanding CRT-UNet's applicability to diverse terrains and ensuring its continued relevance in both operational and research settings.

Chapter 5

Multisource and Multimodal Data Fusion for High Quality Time Series Generation

High-quality time series are essential for capturing dynamic environmental processes and monitoring vegetation and land cover changes over time. Generating such time series is particularly challenging due to the trade-off between spatial and temporal resolution in satellite imagery and the persistent presence of clouds in many regions, such as mountainous areas as detailed in Chapter 2.

This chapter focuses on the necessity of generating continuous, high-quality time series by integrating S1, S2, and S3 data. The proposed framework leverages cloud-free S2 images generated using S1 and topographic data, aligns S3 data through spectral adjustment, and applies STF techniques to produce temporally dense and spatially detailed time series.

Building on methodologies developed in previous chapters, this chapter demonstrates the practical application of cloud removal and STF to generate continuous time series for vegetation and land cover monitoring. By addressing the challenges posed by cloud cover and spectral inconsistencies, the framework ensures that high-resolution data is available even in cloud-prone regions. These methodologies include:

- Multi-Sensor Harmonization: Chapter 3 addressed the challenges of aligning datasets from different sensors with varying resolutions and characteristics. Techniques for preprocessing, spatial alignment, and spectral adjustment were developed to ensure compatibility between S2, and S3. These steps are critical for enabling seamless integration and accurate spatiotemporal fusion of multi-sensor data.
- Cloud-Removal Techniques: In Chapter 4, a cloud-removal model was developed to mitigate the impact of persistent cloud cover on optical data. By integrating S1's radar data with S2, this approach generated high-quality, cloud-free S2 images. The model's effectiveness in creating gap-free imagery lays the foundation for constructing a temporally consistent dataset.

This chapter leverages these methodologies to tackle a specific challenge: generating a high-quality resolution time series that retains the spectral and the spatial detail necessary for analyzing vegetation and land cover changes. By applying the methods in a practical context, this chapter demonstrates their broader utility. The ability to track changes in vegetation and land cover over time is essential for addressing numerous environmental and societal challenges. High-quality time series enable the detection of short-term and transient changes, such as vegetation responses to extreme weather events or human-induced disturbances. They support seasonal and phenological analyses, helping to monitor vegetation cycles and understand the impacts of climate variability.

Specifically, this chapter seeks to:

- Utilize the cloud-removal techniques established in earlier chapters to generate gap-free S2 images.
- Preprocess S3 data to align temporally, spatially, and spectrally with S2, ensuring compatibility for fusion.
- Apply the STARFM method to combine S2 and S3 data, generating a temporally dense and spatially detailed dataset suitable for analyzing vegetation and land cover changes.

This chapter contributes to advancing the field of remote sensing and environmental monitoring in several key ways. It shows the integration of cloud-removal and STF techniques in a real-world application to generate high-quality time series, providing a template for other regions and environmental variables. It provides a framework for continuous time series generation in cloud-prone areas, ensuring accurate vegetation and land cover monitoring. The ability to produce such time series addresses critical gaps in current monitoring capabilities and enhances remote sensing applications across various domains, including agriculture, climate change mitigation, and environmental management. Furthermore, it highlights the potential of multi-sensor fusion to enhance remote sensing applications across various domains. By overcoming the limitations of individual sensors, this chapter provides a pathway for generating high-quality datasets that meet the demands of dynamic environmental monitoring.

5.1 Material & Method

In this chapter, we utilize S1, S2, and S3 data (previously described in Section 2.1), along with topographic information, to address the challenges of generating high-quality time series for vegetation and land cover monitoring. This imagery is integrated to ensure both spatial detail and temporal continuity.

This study directly applies the previously established methodologies without repeating their descriptions, focusing instead on how they contribute to generating high-quality time series.

5.1.1 Study site

In addition to the North Patagonia region presented in Section 4.1.2, we expanded the study area to include several mountainous regions across Europe between 2018 and 2022, which

complement the previous study, introducing a diverse array of environmental conditions, as illustrated in Figure 5.1.

These additional sites were selected to enhance the analysis of multisource data fusion in diverse topographic and climatic conditions, further testing the robustness of the proposed methodology. It is important to note that the study focuses only on the countries covered by the selected S2 tiles. The expanded study areas cover:

- **The Alps**: Spanning France, Italy, Germany, and Switzerland, the Alps are one of Europe's most extensive mountain systems, reaching elevations of over 4,600 meters at Mont Blanc. The region experiences significant annual precipitation, ranging from 1000 to 3000 mm, depending on altitude and exposure. Cloud cover is frequent, especially during the winter and transitional seasons, which complicates satellite-based observations. Snow cover is prominent during winter months, with lower elevations experiencing intermittent snow and higher altitudes maintaining persistent snow and glaciers. Vegetation in the Alps varies from dense forests at mid-elevations, including beech, fir, and spruce, to alpine meadows and bare rock zones at higher elevations.
- **The Picos de Europa**: Located in northern Spain, the Picos de Europa are part of the Cantabrian Mountains, reaching elevations of up to 2650 meters. This region is heavily influenced by the Atlantic Ocean, leading to high annual precipitation levels averaging 800 mm to 3000 mm, with frequent rainfall throughout the year. Cloud cover is common, particularly in autumn and winter, and low-hanging clouds are a typical feature of its valleys. Snow cover is usually restricted to higher altitudes during winter. The area is characterized by lush vegetation, including deciduous forests of oak and beech, alongside grasslands and karst landscapes.
- **The Pyrenees Mountains**: Stretching between Spain and France, the Pyrenees form a natural border between the two countries, with peaks reaching up to 3400 meters. The region experiences a mix of climatic influences, with wetter conditions on the Atlantic-facing slopes and drier, Mediterranean-influenced climates on the southern slopes. Precipitation averages between 800 and 2000 mm annually, depending on elevation and orientation. Snowfall is significant at higher elevations during the winter, with snow-covered peaks lasting into spring. Vegetation varies widely, from Mediterranean woodlands of holm oak and pine at lower elevations to montane forests and alpine meadows at higher altitudes.
- **The Monti Simbruini**: Located in central Italy, it is a part of the Apennine mountain range and form the core of the Parco Naturale Regionale Monti Simbruini. It covers rugged peaks, deep valleys, and vast forests, with peaks reaches 2156 meters. Due to their elevation and geographical position, the region experience a temperate mountain climate with distinct seasonal variations. Snowfall is common in winter, especially above 1500 meters, covering the landscape from December to March. Annual precipitation is abundant, ranging from 1200 to 1800 mm , with the highest amounts occurring in autumn and early spring. The karstic nature of the terrain plays a crucial role in the region's hydrology, feeding numerous underground water reserves and contributing to the formation of important springs and rivers such as the Aniene River. The vegetation

in Monti Simbruini varies with altitude, from beech forests to subalpine meadows and sparse vegetation, featuring juniper and alpine grasslands rich in endemic plant species.

The variability of these regions in cloud cover, precipitation patterns, snow dynamics, and vegetation types across these mountainous areas offers an opportunity to analyze environmental interactions in regions with complex topography and climates. Together, these sites provide a broad spectrum of ecological and climatic conditions for study.

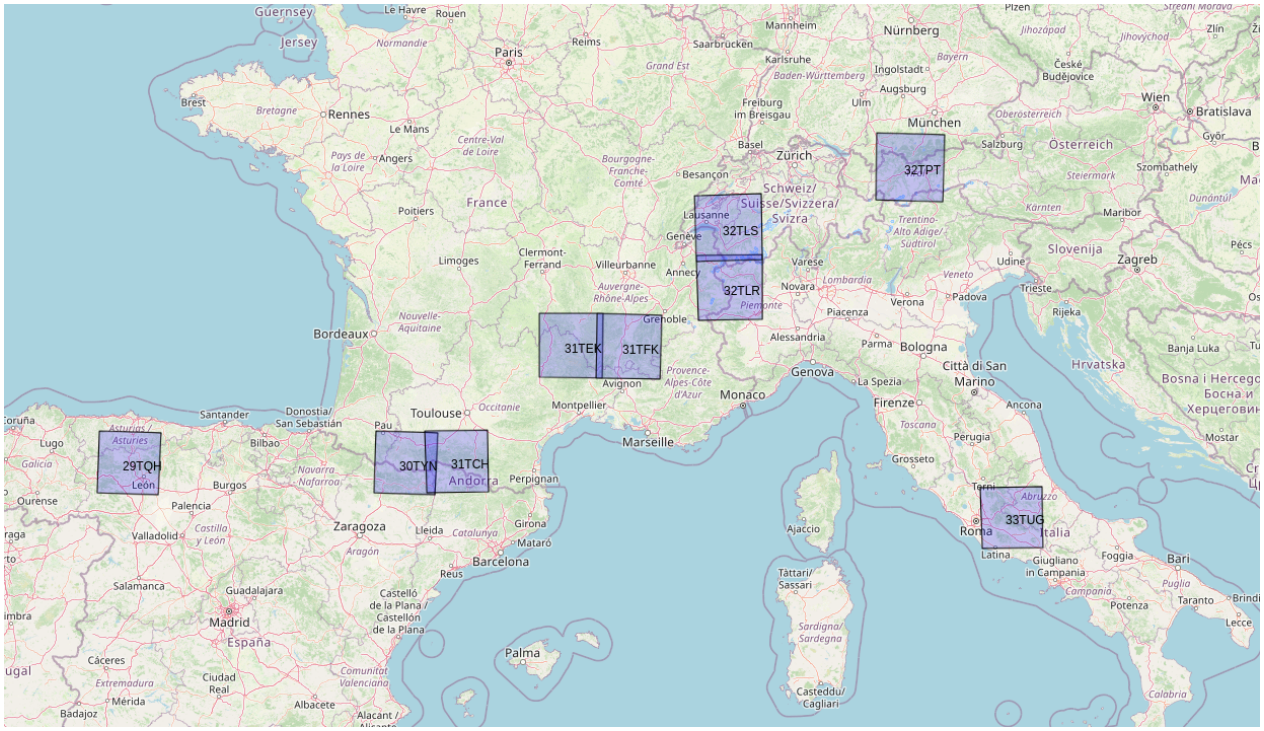


Figure 5.1: Expanded study areas in Europe, showing the S2 tiles covering the Alps (France, Italy, and Switzerland), the Picos de Europa (northern Spain), the Pyrenees Mountains (Spain and France), the Monti Simbruini (Italy).

5.1.2 Methodology

The methodology for this study is divided into three main parts, each addressing specific steps required to generate high-quality time series. These steps build on the techniques and workflows developed earlier in this thesis and adapt them for the expanded study area, as summarized in the workflow schema shown in Figure 5.2.

The first part involves preprocessing the S1 and S2 imagery for the additional mountainous locations, as described in Section 4.1.2. The same preprocessing pipeline, including radiometric calibration, atmospheric correction, cloud masking, and terrain alignment, was applied to ensure consistency with the North Patagonia datasets. The processed S1 and S2 dataset with the topographic data were used to retrain the neural network architecture introduced in Section 4.1.3. This ensures the models are capable of handling the unique climatic and topographic conditions of the new regions while maintaining consistency with the established

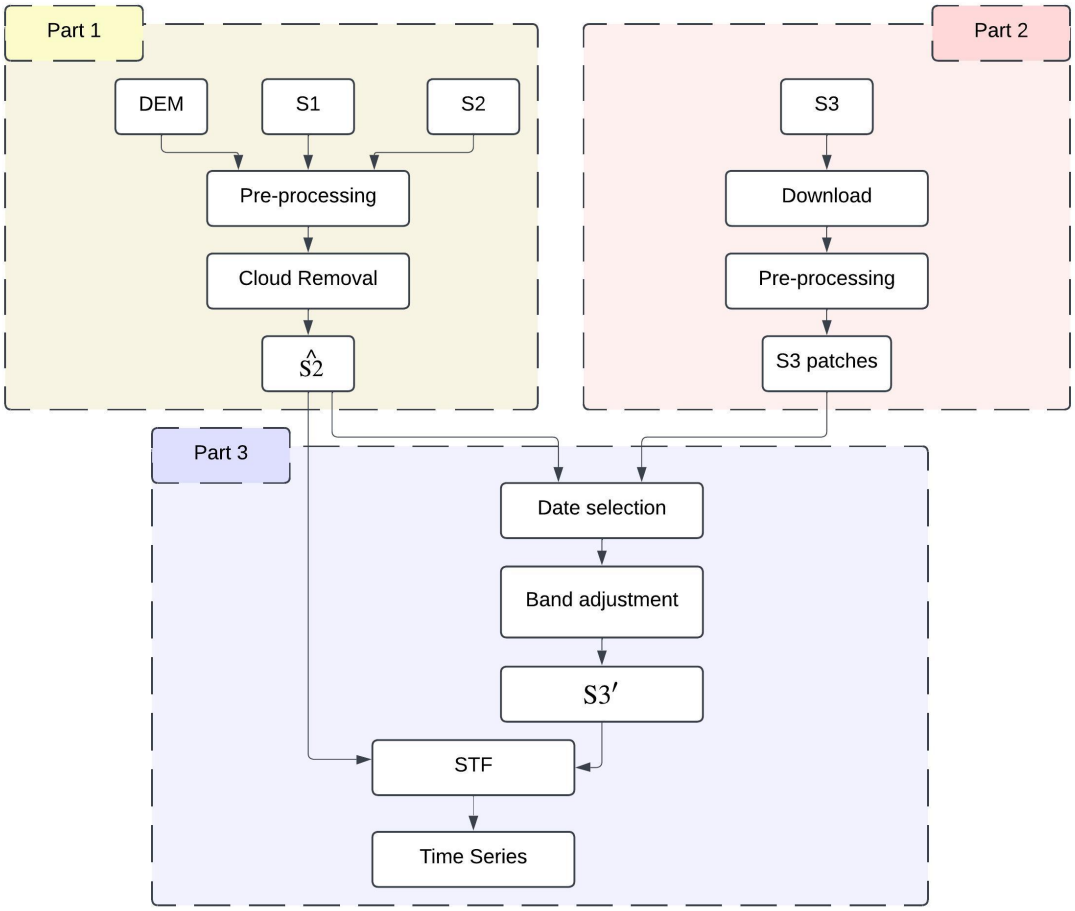


Figure 5.2: Workflow schema of the methodology, divided into three main parts: (1) preprocessing S1 and S2 data and the cloud removal model, (2) preprocessing S3 data to spatially harmonize it with S2, and (3) selecting overlapping bands and performing band adjustments for spatiotemporal fusion.

methodology. This step results in $\hat{S2}$, the cloud-free S2 images predicted by the cloud removal model, which are integral for the subsequent steps.

The second part focuses on preprocessing S3 data to spatially harmonize it with S2. S3's coarser spatial resolution was resampled to match S2 imagery, and atmospheric corrections were applied to ensure radiometric consistency. This pipeline ensures that S3 data can effectively complement S2 observations by filling temporal gaps without introducing spatial inconsistencies. The output of this step is S3′, the adjusted S3 dataset, which is compatible with S2 data and prepared for further integration.

The third part involves selecting the overlapping spectral bands between S2 and S3 and identifying dates where observations from both sensors are available. A band adjustment process was conducted to minimize spectral discrepancies between the two datasets. This adjustment ensures that the overlapping bands are harmonized and suitable for use as inputs in the spatiotemporal fusion process. The harmonized bands and selected dates form the

foundation for subsequent fusion steps, enabling the generation of a consistent and high-quality time series.

To contextualize the results, it is important to first illustrate the problem addressed by the proposed framework and the workflow employed to solve it. Figure 5.3 visually demonstrates the challenge of generating high-quality, cloud-free S2 images at specific time steps (e.g. t_2) using multisource data. At t_1, we have a cloud-contaminated S2 image, and S3 image. At t_2 only a S3 image is available. The goal of the framework is to first generate a cloud-free S2 image at t_1 ($\hat{\text{S2}}_{t_1}$) using the cloud removal model. This cloud-free image is then used as a critical input, along with $\hat{\text{S3}}$ data, to predict a high-resolution S2 image at t_2 ($\hat{\text{S2}}_{t_2}$) using STF techniques. This figure highlights the framework's ability to address key challenges in remote sensing, including cloud contamination in optical data and the temporal gaps in high-resolution imagery.

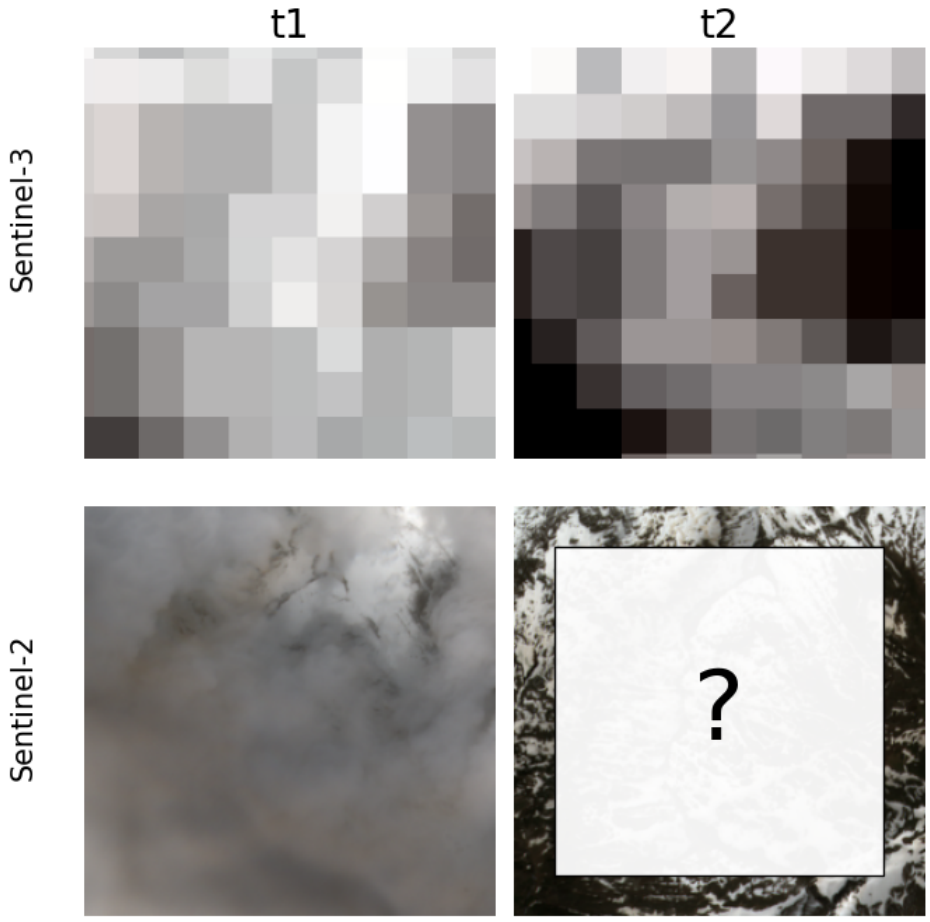


Figure 5.3: The challenge of the STF method with the existence of clouds at time t_1.

Cloud-Free Sentinel-2

To prepare the data for training, S1 and S2 images were preprocessed using the pipeline described in Section 4.1.2. S1 SAR imagery was spatially aligned with S2 images by resampling

to match their resolution, ensuring compatibility between the datasets. Cloud masking was performed on S2 images using the cloud mask generated with MAJA and auxiliary data to identify and mask cloud-covered areas. Additionally, all images were normalized to maintain consistent pixel value ranges, facilitating model training and reducing the influence of outliers. To enhance the model's robustness and generalization, data augmentation techniques such as random rotations, flips, and cropping were applied. These augmentations increased variability in spatial patterns and environmental conditions seen by the model during training.

The dataset was split spatially, meaning that patches were assigned to either the training, validation, or testing sets based on their geographic location. Specifically, each tile was divided into non-overlapping subsets, with approximately 80% of the patches allocated to training, 15% to validation, and 5% to testing. This spatial split ensures that the model is evaluated on locations it has not seen during training, enabling a more realistic and robust assessment of its generalization capability.

The training set includes patches covering the period from 2018 to 2023, allowing the model to learn from a wide temporal span, including multiple seasonal cycles and atmospheric conditions. In contrast, the validation and testing sets include patches from 2019 to 2023, selected exclusively from locations not used in training. This guarantees that evaluation is performed on spatially distinct regions, preventing data leakage and overfitting. To further support generalization without introducing entirely unfamiliar spatial contexts, historical imagery (from 2018–2019) corresponding to the validation and testing regions was also included in the training set. This means that while the model does not see the target patches used for evaluation, it is exposed to earlier observations from the same geographic locations. This strategy allows the model to learn about the underlying landscape characteristics of those regions—such as vegetation structure, land cover type, and topography—before being tasked with reconstructing cloud-free images in different time periods. This spatial-temporal splitting strategy balances the need for spatial independence in evaluation with the benefits of temporal familiarity.

Table 5.1 provides a summary of the years represented in each dataset split, highlighting how training, validation, and testing patches are distributed over time, it includes also the number of images in each set.

Table 5.1: Yearly distribution of image patches across training, validation, and testing sets.

Year	Training Set	Validation Set	Testing Set
2018	✓	–	–
2019	✓	✓	✓
2020	✓	✓	✓
2021	✓	✓	✓
2022	✓	✓	✓
2023	✓	✓	✓
Total images	141062	26550	9037

The training dataset, comprising preprocessed and augmented S1 and S2 images from North Patagonia and the European locations, was then used to train the cloud removal model

described in Section 4.1.3. After training, we apply the model to the testing set. This step produced a cloud-free time series for each tile at the spatial and temporal resolution of S2. These predicted images were later used for subsequent analyses. Figure 5.4 present the detailed workflow of this part.

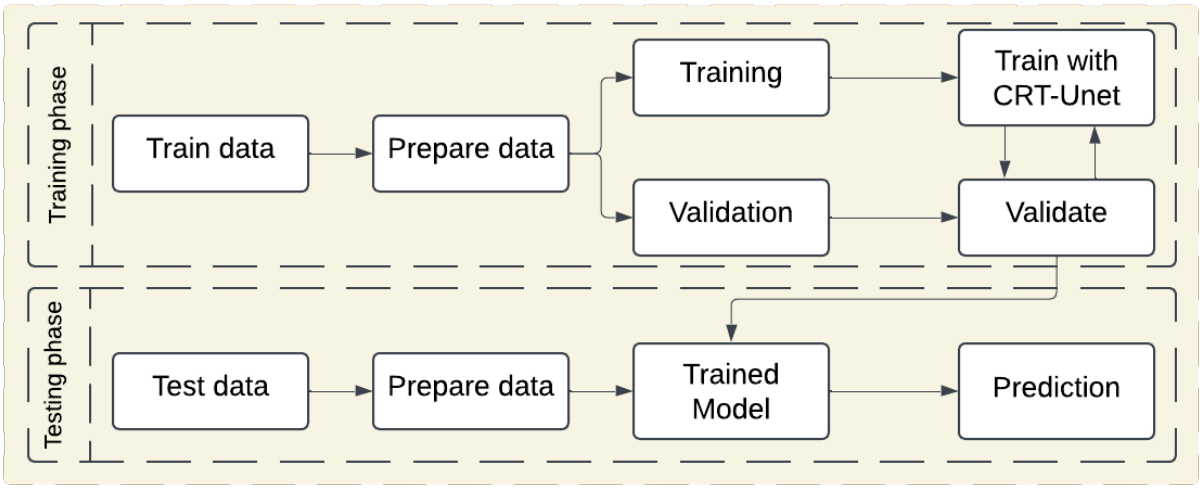


Figure 5.4: Workflow of the first part of the methodology, illustrating the training and the testing phase.

Sentinel-3 Preprocessing pipeline

The second part of the methodology focuses on preprocessing S3 OLCI Level 1 images to align them spatially with the S2 images predicted in the first part of the methodology. This preprocessing step ensures compatibility between the datasets, forming the foundation for subsequent analysis and integration.

To begin, the available S3 OLCI Level 1 images were retrieved, corresponding to the same geographic locations as the S2 cloud-free images ($\hat{S2}$) generated in the first part. Each S3 image matching these criteria was downloaded for processing.

The preprocessing workflow for S3 images starts with atmospheric correction, performed using the Acolite algorithm (Vanhellemont & Ruddick, 2021). This step provides surface reflectance values comparable to those of S2 imagery. After atmospheric correction, the preprocessed S3 images are subjected to multiple additional processing steps using the Graph Processor Tool (GPT) in the ESA SNAP software. The first step is the reprojection of the S3 OLCI data to the same coordinate reference system as the S2 imagery, to ensure spatial alignment between the two datasets. Then, a band selection is applied where only the reflectance bands relevant for analysis were retained. This filtering reduced data volume and focused the processing on the most meaningful spectral information. To conserve computational memory, the images are then clipped with a buffer to roughly match the spatial extent of the corresponding S2 image tiles. The S3 images are then resampled to a 10 m resolution using the nearest neighbor method. The use of nearest neighbor resampling preserves the original reflectance values

without introducing new interpolated values, maintaining data integrity. Finally, the images underwent a second, precise clipping step to exactly match the spatial extent of the S2 image patches generated in the first part. Figure 5.5 presents the workflow of the S3 preprocessing pipeline.

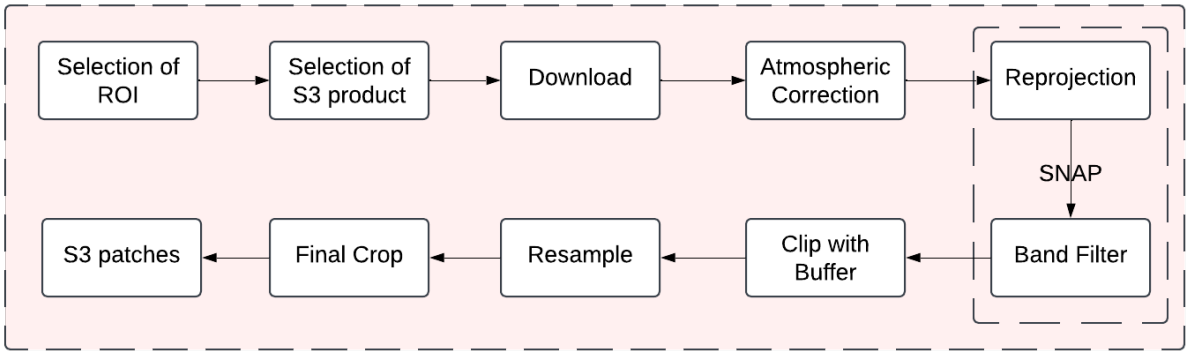


Figure 5.5: Preprocessing pipeline of Sentinel-3

Band selection and Adjustment

The third part of the methodology focuses on aligning S2 and S3 datasets temporally and spectrally, ensuring compatibility for their integration in the STF process. This step bridges the differences between the sensors and prepares the datasets for generating high-temporal-resolution outputs. The process began with selecting matching dates from S2 and S3 data to create temporal triples. For each time step t_1, corresponding S2 ($S2_{t1}$) and S3 ($S3_{t1}$) images were paired, along with the next available S3 observation at t_2 ($S3_{t2}$). These temporal triples ($S2_{t1}, S3_{t1}, S3_{t2}$) formed the basis for subsequent spectral alignment and fusion steps.

Once the temporal triples were established, all overlapping bands between S2 and S3 were identified. These overlapping bands were critical for ensuring spectral consistency between the datasets. Linear coefficients were then calculated for each pair of overlapping bands, following the methodology described in detail in Section 3.2.3. These coefficients were derived to minimize spectral discrepancies between S2 and S3 reflectance values, effectively harmonizing their spectral responses. The calculated coefficients were applied to the S3 reflectance values, producing spectrally adjusted S3 data ($S3'$). This spectral adjustment ensured that the adjusted S3 data was comparable to S2 in terms of both magnitude and spectral characteristics. The $S3'$ dataset was prepared as input for the STF process.

The STF method used in this study was the STARFM, which was described in detail in Chapter 2. STARFM leverages the temporal continuity of S3 data ($S3'$) and the high spatial resolution of S2 ($S2_{t1}$) to predict fine-resolution images for intermediate time steps. By combining the strengths of both sensors, STARFM generates a temporally dense and spatially detailed time series, suitable for monitoring vegetation and land cover dynamics.

5.2 Results

The results of this study present the outcomes of the methodology applied to generate high-temporal-resolution time series through the integration of S1, S2, and S3 imagery. The focus is on evaluating the effectiveness of the cloud removal process, the spectral adjustment of S3 images, and the STF approach in producing seamless and accurate outputs for vegetation and land cover monitoring. To demonstrate the results of the framework, we used a time series from a single patch located in Sentinel-2 [tile 31TCH], as shown in Figure 5.1. This patch was selected due to its diverse landscape, which includes variations in vegetation cover, topography, and seasonal changes, making it an ideal test case to evaluate the performance of the framework under different environmental conditions.

After applying the proposed methodology, we generated images with enhanced spatial and temporal resolution, allowing for more complete time series coverage. However, while the integration of multimodal data improves the continuity of observations, the results exhibit variable performance across the evaluation period.

Figure 5.6 presents the temporal evolution of SSIM values, computed between the reconstructed images and the corresponding reference Sentinel-2 acquisitions. These values reflect the structural similarity of the fusion outputs and serve as an indicator of the model's effectiveness at different time points.

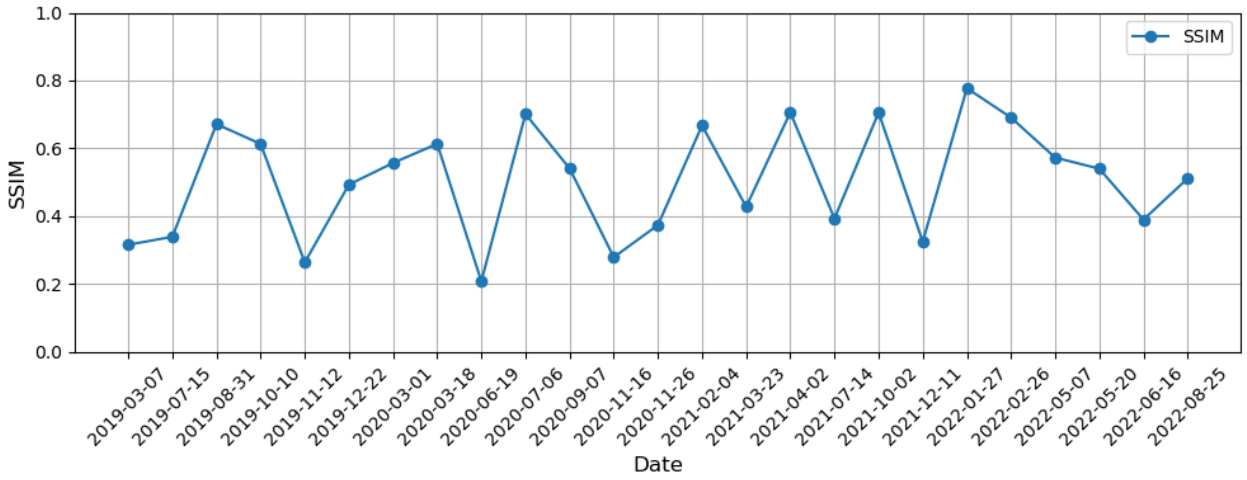


Figure 5.6: Time serie of SSIM values between the generated images and its corresponding Sentinel-2 reference image

The SSIM values show noticeable variability, ranging from approximately 0.2 to 0.8, suggesting that while the model is capable of producing satisfactory reconstructions in many instances, it also faces difficulties under certain conditions. This fluctuation can be attributed to multiple factors, including variations in land cover dynamics, scene complexity, and environmental conditions. Additionally, lower SSIM values tend to occur during periods with fewer available cloud-free input images, which limits the temporal context available for reliable fusion. Sparse data can hinder the model's ability to reconstruct fine spatial details and spectral patterns accurately.

Higher SSIM values, on the other hand, are observed in periods with less snow cover and more vegetation cover, showing the framework's ability to produce consistent and accurate outputs under typical conditions.

Similarly, Figure 5.7 presents the RMSE values calculated between the reconstructed and reference S2 images (when cloud-free observations were available). While the RMSE values remain within a moderate range throughout the testing period, they exhibit noticeable variability, reflecting differences in scene conditions, land cover dynamics, and input data availability.

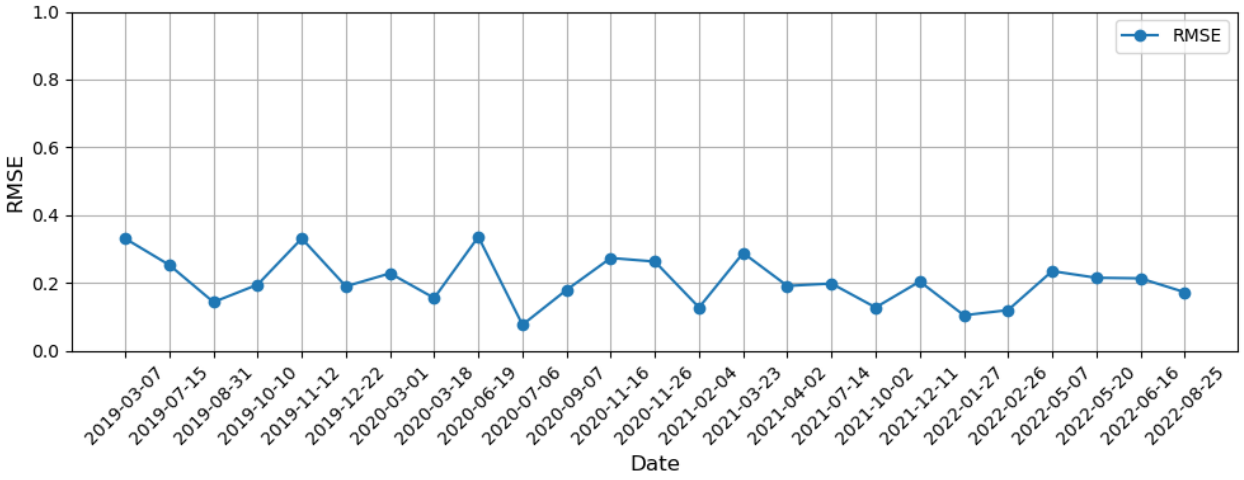


Figure 5.7: Time serie of RMSE values between the generated images and its corresponding Sentinel-2 reference image

Higher RMSE values observed at certain time points suggest that the fusion process is more challenging under specific conditions particularly when fewer high-quality or temporally close input images are available to guide reconstruction. These discrepancies may also stem from surface changes or complex reflectance patterns that are difficult to model with limited data support. Conversely, lower RMSE values are recorded during periods with more temporally rich input data and stable surface conditions, where the model can better align spectral and spatial features with the reference.

Overall, while the RMSE trend suggests that the framework is capable of generating plausible outputs across a range of scenarios, these observations underscore the need for further refinement to improve performance in cases where supporting data is sparse or environmental conditions are more variable.

Figure 5.8 presents an RGB composite of results for a series of prediction dates using the proposed fusion framework. From top to bottom, the rows display: (i) S3 imagery at the base date $t1$, (ii) S2 imagery at the same base date (typically affected by clouds), (iii) S3 imagery at the prediction date $t2$, (iv) the fusion output generated for $t2$, and (v) the corresponding cloud-free S2 reference image used for evaluation. The columns represent different dates, capturing a variety of landscape and seasonal conditions, including scenes with cloud cover, vegetation, and snow.

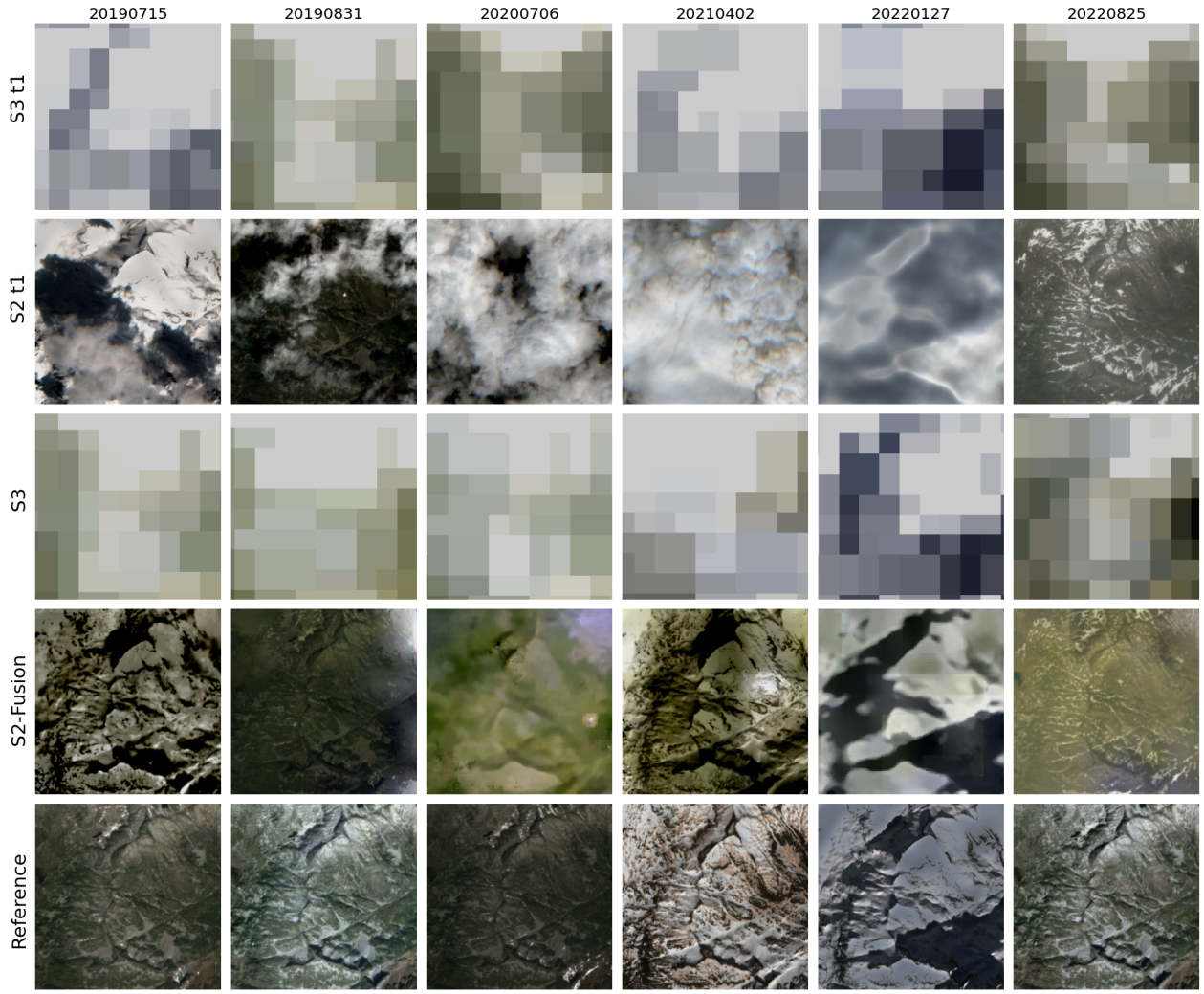


Figure 5.8: Sentinel-3, spatio-temporal fusion results and Sentinel-2 RGB compositions on different dates.

The fusion results show a spatial enhancement over the coarse-resolution S3 input, with improved delineation of terrain features and finer structural detail. This illustrates the ability of the framework to incorporate spatial characteristics from S2 while leveraging the temporal availability of S3 data. The fusion outputs demonstrate noticeable improvements in spatial detail compared to the original low-resolution S3 inputs. Fine textures, landscape features, and terrain structures are better defined, indicating that the method successfully transfers spatial information from the high-resolution sensor while benefiting from the dense temporal coverage of the coarse-resolution imagery.

In several examples, the fused images visually approximate the reference S2 scenes, preserving major land cover features and seasonal patterns. However, the results also exhibit certain limitations. A recurring artifact across some predictions is a greenish tint, particularly visible in vegetated or mixed land cover areas. This suggests a spectral imbalance in the fusion process, possibly caused by the limitation of STARFM fusion method. These discrepancies

highlight the need for further refinement of the spectral harmonization strategy and the spatiotemporal fusion to better preserve realistic color distributions and surface reflectance properties.

Additionally, in snow-covered scenes such as the one from 2022-01-27, the fused outputs tend to over-smooth the fine structures and lack the detailed texture observed in the reference image. This limitation may stem from the challenges posed by high reflectance, terrain-induced variability, and limited spatial context in the input data. Overall, while the visual results highlight the potential of the proposed framework to generate enhanced spatiotemporal products, they also underscore areas where performance could be improved, particularly in challenging conditions such as snow, shadowed terrain, or rapid surface changes.

To demonstrate the utility of the generated time series, the Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974), was calculated for the S3 original data, the S2 reference data (when available) and the fusion results ($\hat{S2}$). The NDVI, a key vegetation index, provides a quantitative measure of vegetation health and density, making it an ideal metric for assessing the quality of the fusion outputs, it can be calculated as:

$$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}} \tag{5.1}$$

where: NIR is reflectance in the Near-Infrared band, RED is reflectance in the Red band. This index ranges from -1 to 1, with higher values indicating more vigorous vegetation. By comparing the NDVI time series derived from the fusion results to those from the S2 reference and S3 original data, we can evaluate the framework's ability to capture temporal vegetation dynamics while maintaining spatial accuracy.

Figure 5.9 presents the NDVI time series calculated from the S2 reference data, the S3 original data, and the fusion results ($\hat{S2}$). The NDVI values are shown as discrete points. This comparison highlights the temporal dynamics of vegetation and evaluates the effectiveness of the fusion framework in replicating the high-resolution NDVI patterns of S2.

The S3 NDVI values, represented by the orange star symbol, show a smoother temporal trend but lack fine detail due to the sensor's coarser spatial resolution as presented in Figure 5.8. The fusion results, shown as the green circle, closely follow the reference S2 NDVI values while maintaining the temporal continuity of S3. The peaks and troughs in the fusion NDVI align well with the S2 reference, demonstrating the framework's ability to accurately capture vegetation dynamics, such as seasonal growth and senescence cycles.

Notably, discrepancies are observed during certain periods, particularly when NDVI values are lower. These differences could be attributed to environmental factors like snow cover or rapid vegetation changes that challenge the fusion model, it could also be attributed to the quality of the cloud free S2 image that is used as an input to the STF. Despite this, the fusion results generally outperform the S3 NDVI by providing greater spatial detail and keeping alignment with the high-resolution S2 reference.

To summarize, the results demonstrate the utility effectiveness of the proposed framework in addressing the challenges associated with cloud contamination, spectral differences, and temporal gaps in remote sensing data. The cloud-free S2 images ($\hat{S2}$) generated by the

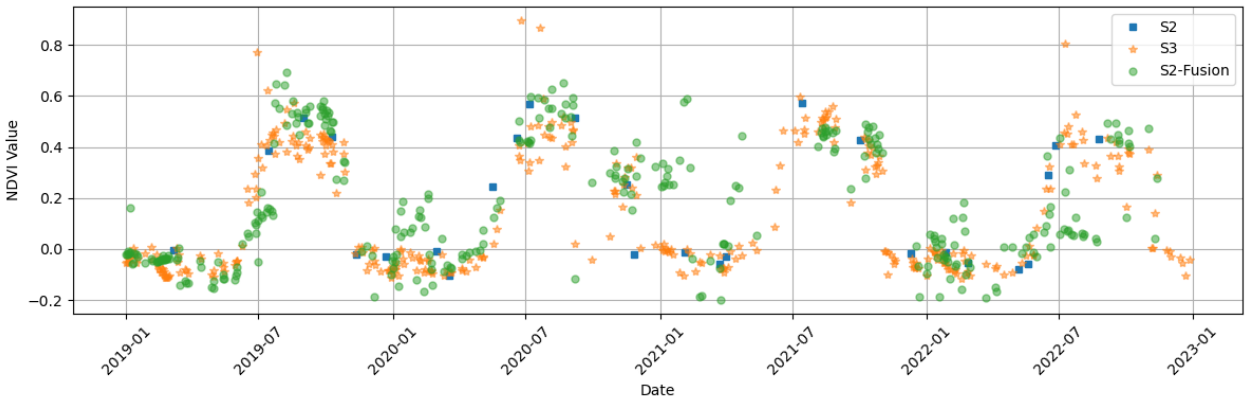


Figure 5.9: NDVI time series based on Sentinel-3 fusion and Sentinel-2 data.

framework preserve critical spatial and spectral information, while the spectral adjustment of S3 data ensures compatibility for STF. The fusion process produced spatially enhanced outputs that captured the temporal variability of the landscape and, in many cases, approximated the structure and reflectance of reference S2 observations. This was particularly evident in periods with stable surface conditions and good data availability. However, the framework exhibited sensitivity in more complex scenarios, such as snow-covered areas or dates with rapid vegetation changes. In these cases, minor artifacts and spatial or spectral deviations were observed, reflecting limitations in input data quality, spectral mismatches, or fusion model generalization.

5.3 Discussion

The proposed framework demonstrates a novel approach to multisource and multimodal data fusion, integrating S1, S2, and S3 data to generate high spatial and temporal resolution imagery. By combining the strengths of radar, optical (both coarse and fine), the framework addresses key challenges in remote sensing, particularly in cloud-contaminated regions. The framework begins with the generation of cloud-free S2 images ($\hat{S2}$) using S1 and DEM data, ensuring the availability of high-resolution optical images even under persistent cloud cover. These cloud-free images are then spectrally adjusted to align with S3 data, reducing distortions caused by differences in sensor characteristics, before performing STF to generate temporally dense and spatially accurate outputs.

One of the primary strengths of the framework lies in its ability to address the temporal gaps in S2 observations, which are often exacerbated in cloud-prone regions. Previous studies have shown that the time difference between the base time (t_1) and the prediction time (t_2) significantly affects the quality of spatiotemporal fusion results (J. Zhou et al., 2021). By ensuring the availability of cloud-free S2 images at t_1, the framework minimizes these temporal gaps and improves the accuracy of the fusion outputs. However, challenges remain in regions where cloud cover is persistent, especially during the winter months, when the frequency of cloud-free observations is significantly reduced. In such cases, the next available cloud-free image may represent a completely altered landscape due to snow cover or rapid vegetation

changes, making the fusion results less reliable.

The results confirm the framework's ability to generate stable and temporally consistent time series, with better spatial detail than S3 while maintaining temporal alignment and behavior consistent with S3 observations. This performance makes the framework suitable for applications such as vegetation monitoring and phenological studies, where both spatial and temporal resolution are critical. However, limitations are observed in areas with significant snow cover, which introduces high reflectance values, particularly in the visible bands. These spectral distortions can propagate through the fusion process, leading to artifacts or inaccuracies in the final output. This highlights the need for specialized handling of snow-covered regions, such as adaptive spectral adjustment techniques or the incorporation of additional snow-specific indices.

Although the framework performs robustly under diverse environmental conditions, there is room for improvement. For example, the inclusion of more sophisticated cloud-detection and removal algorithms could enhance the quality of cloud-free S2 images. Additionally, integrating ancillary datasets such as weather information or snow cover maps could help mitigate the challenges posed by snow and seasonal variability. Enhancing the spectral adjustment process to account for dynamic environmental conditions, such as changing illumination or atmospheric effects, could further improve the spectral consistency between S2 and S3. Finally, extending the framework to include more advanced STF techniques, such as those incorporating machine learning, could enhance the ability to model complex temporal patterns and transitions.

To sum up, the proposed framework offers a practical and effective solution for generating high-temporal and high-spatial-resolution datasets through the fusion of S1, S2, and S3 data. Despite some limitations, the framework provides a solid foundation for advancing remote sensing applications and improving environmental monitoring capabilities. Future work could focus on addressing these limitations and exploring opportunities to expand the adaptability and accuracy of the framework in various conditions and applications.

Chapter 6

Conclusions and future work

6.1 Conclusions

This thesis addressed the critical challenge of generating high-resolution, temporally dense, and spectrally consistent time series in remote sensing by developing a multimodal, multisource fusion framework. Existing methodologies often treat spatial, temporal, and spectral dimensions separately and fail to resolve the compounded issues arising from cloud contamination, sensor-specific spectral discrepancies, and rapidly changing land cover. In response, this thesis proposed a cohesive solution that integrates S1 (SAR), S2 (high-resolution optical), and S3 (coarse-resolution optical) data, supported by topographic information and advanced deep learning models. The proposed framework tackles the three central challenges outlined in the thesis objectives and each chapter of the thesis directly supports the overarching goal.

The integration of S1 (SAR), S2 (fine-spatial resolution optical), and S3 (coarse-spatial resolution optical) data enables improved STF while overcoming limitations posed by persistent cloud cover, spectral misalignment, and temporal inconsistencies. The contributions of this thesis provided a robust solution to these challenges, advancing the state-of-the-art in remote sensing by ensuring continuous, cloud-free, and spectrally accurate observations. The framework developed in this thesis presents a step forward in data fusion methodologies, making significant improvements in the quality and applicability of remote sensing time series data.

A primary objective of this work was the development of a unified framework that integrates multiple sensors, topographic data and fusion techniques for time series reconstruction. Traditional STF methods focus on merging fine and coarse images, such as S2 and S3, to improve spatial resolution but depend on cloud-free optical data, limiting their effectiveness in regions with persistent cloud cover. By enabling STF beyond cloud-free scenarios, the framework enhanced the overall robustness of the framework.

The first contribution of this thesis lies in addressing spectral misalignment issues between sensors (Chapter 3). Differences in SRF between S2 and S3 often result in spectral inconsistencies in the fusion products. This thesis proposed a spectral adjustment strategy that aligns the reflectance values of S3 with S2, significantly reducing spectral distortions and ensuring

consistency in fused time series data. Spectral inconsistencies are a major source of error in remote sensing applications that rely on multisensor fusion, particularly for vegetation indices, land cover classification, and environmental monitoring. The spectral adjustment strategy introduced in this work enhances the compatibility of these datasets, allowing for improved accuracy in change detection and other analytical applications.

The second contribution of this work was the development of a cloud removal model (CRT-UNet) to enhance the availability of high-quality S2 images (Chapter 4). This method effectively reconstructs cloud-contaminated optical data by leveraging S1 SAR and topographic information, addressing a major limitation in conventional STF approaches that depend on cloud-free observations. Cloud cover is a persistent issue in many regions worldwide, particularly in mountainous and humid tropical areas, making reliable cloud removal essential for consistent time series generation. By applying DL and multisource data integration, CRT-UNet significantly improves the quality of recovered optical imagery, ensuring that cloud-free observations can be more reliably incorporated into STF workflows.

A key advancement was the integration of S1 SAR data to enhance temporal alignment in STF. Traditional fusion methods assume minimal change between observation dates, an assumption that fails in dynamic landscapes such as agricultural areas, wetlands, and urban environments. By incorporating SAR data, which remains unaffected by atmospheric conditions, this research provides additional structural information, stabilizing the generated time series and ensuring meaningful reconstructions even when optical observations are unavailable for extended periods. The ability to integrate radar and optical data effectively is a major step forward to ensure the continuity and reliability of remote sensing observations under varying environmental conditions.

Beyond methodological advancements, this thesis introduced an adaptive multimodal fusion strategy that optimally combines cloud removal, spectral alignment, and STF techniques. Rather than relying on a singular fusion method, the framework dynamically selects and integrates the most appropriate technique at each stage of reconstruction. This makes it highly adaptable for various remote sensing applications, including environmental monitoring, vegetation phenology studies. Adaptability is critical in remote sensing workflows, as different environments and data availability scenarios require flexible methodologies to maximize the utility of observations. By incorporating multiple models and adjusting the fusion strategy accordingly, the framework ensures that the highest quality outputs are generated for a diverse range of applications.

The findings of this thesis confirmed that multisource and multimodal data fusion is essential for producing stable, high-quality time series, particularly in cloud-prone regions. The proposed framework enhances spatial and spectral fidelity while preserving the temporal behavior of observed landscapes. The ability to maintain high accuracy in both spatial and temporal domains is crucial for detecting subtle environmental changes, making the proposed framework a strong tool for scientific and operational applications. This research demonstrates that a well-integrated fusion framework can bridge the temporal gaps in S2 observations and minimize spectral distortions, leading to more reliable and consistent datasets for long-term environmental studies.

Each chapter of this thesis contributes to the overarching goal of improving time series images. Chapter 3 introduces a novel approach for spectral band selection in fusion processes, demonstrating the benefits of incorporating multiple spectral bands to enhance the quality of fused images. This approach ensures that the full spectral richness of available datasets is leveraged, reducing information loss and improving the accuracy of derived products. Chapter 4 presents CRT-UNet, a model designed to address cloud contamination issues, ensuring the availability of high-quality S2 data for fusion. The implementation of DL within this workflow provides a scalable and automated approach to cloud removal, allowing for more widespread adoption of cloud-free data generation methodologies. Chapter 5 integrates multiple fusion techniques to construct a multimodal, multisource fusion framework, reducing spectral distortions, improving spatial resolution, and ensuring temporal consistency. This chapter synthesizes the contributions of the previous methodologies into a unified approach.

This thesis establishes a strong foundation for the next generation of remote sensing time series reconstruction methods, emphasizing the critical role of Sentinel mission satellites in multisource and multimodal fusion. By integrating data from Sentinel-1, Sentinel-2, and Sentinel-3, this research highlights the advantages of leveraging the complementary capabilities of these satellites to generate high-quality, time series. Despite its successes, this research acknowledges several challenges that require further attention. The limitations open new lines of research that would improve the results obtained so far.

Furthermore, this work contributed a publicly available benchmark dataset for spatiotemporal fusion between Sentinel-2 and Sentinel-3, addressing the gap in existing fusion research that largely focuses on MODIS and Landsat data. The dataset comprises 12 environmentally diverse sites across different regions of the world, promoting generalizability and robust model evaluation. It includes 10 S2 bands (4 at 10 m and 6 at 20 m resolution) and 16 S3 bands at 300 m resolution, covering a wide spectral range from the visible to SWIR. This configuration supports a variety of applications, particularly the testing and benchmarking of data fusion algorithms using new-generation European sensors.

This thesis emphasizes the pivotal role of Sentinel satellites in advancing next-generation remote sensing time series reconstruction. By integrating Sentinel-1, Sentinel-2, and Sentinel-3 data, the proposed framework demonstrates the power of multisource and multimodal fusion to bridge temporal gaps, reduce spectral inconsistencies, and ensure data availability under challenging conditions. These capabilities are essential for supporting applications in environmental monitoring, land cover change analysis, vegetation phenology, and agricultural management.

6.2 Future Work and Directions

Building on the contributions of this research, several promising directions can be pursued to further advance time series generation through multimodal, multisource fusion techniques. One key area for future exploration is the integration of emerging DL architectures, such as transformer-based models and graph neural networks, which could enhance the adaptability and efficiency of cloud removal and STF processes. These models have demonstrated superior capabilities in handling complex spatial-temporal relationships and could further improve the

consistency of generated time series.

Another potential advancement is the incorporation of hyperspectral satellite data, in combination with existing multispectral observations, to provide richer spectral information and improve the accuracy of environmental monitoring applications. Future studies could also investigate the fusion of Sentinel data with upcoming missions such as the Landsat Next or commercial high-resolution satellite constellations to enhance spatial and temporal coverage. Expanding the operational scalability of the proposed framework is another essential aspect of future research. Implementing the fusion framework in cloud computing environments, such as Google Earth Engine would enable near real-time processing, facilitating large-scale monitoring applications across diverse geographic regions. This would allow for a broader adoption of the framework in operational remote sensing applications, making high-quality time series data more accessible to researchers and policymakers.

Future research could also explore adaptive fusion frameworks that incorporate dynamic environmental variables, such as weather conditions, soil moisture levels, and vegetation indices, to refine cloud removal and spectral adjustment processes. The integration of such contextual data could significantly improve the reliability of STF outputs in dynamic landscapes, making the framework more robust and generalizable across various environmental conditions. Furthermore, extending the applicability of the framework beyond optical and SAR data by integrating LiDAR and UAV-based observations could provide finer spatial details for high-resolution mapping in critical applications such as disaster response, precision agriculture, and urban planning. This multimodal integration would create a more comprehensive and flexible remote sensing system capable of capturing subtle land cover changes with high accuracy.

Lastly, future research should focus on developing automated and user-friendly implementations of the framework to facilitate its adoption in operational remote sensing workflows. User-friendly implementations could also improve usability among end-users.

6.3 Author's Contribution

To conclude this chapter, the most relevant publications, derived from this research and developed in previous chapters, are referenced here.

- **Boumahdi, M.**, Gonzalo-Martin, C., A., Lillo-Saavedra, Somos-Valenzuela, M., García-Pedrero, A., 2025. Multisource Topographic-Enhanced Cloud Removal for Remote Sensing in Mountainous Landscapes. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. https://doi.org/10.1109/JSTARS.2025.3572379 (Chapter 4).
- **Boumahdi, M.**, García-Pedrero, A., Lillo-Saavedra, M. Gonzalo-Martin, C. A new benchmark for spatiotemporal fusion of Sentinel-2 and Sentinel-3 OLCI images. Earth Sci Inform 18, 349 (2025). https://doi.org/10.1007/s12145-025-01855-4 (Chapter 3)
- **Boumahdi, M.**, García-Pedrero, A., Lillo-Saavedra, M., & Gonzalo-Martin, C. (2025). Evaluation of Spatiotemporal Fusion Methods Using Sentinel-2 And Sentinel-3: A New

Benchmark Dataset And Comparison [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.14860220 (Chapter 3)

- **Boumahdi, M.**, García-Pedrero, A., Gonzalo-Martín, C., Lillo-Saavedra, M. 2024. Eliminación de nubes en imágenes Sentinel-2 de entornos montañosos mediante Sentinel-1: Un caso de estudio de la Patagonia chilena. En Teledetección y Cambio Global: Retos y Oportunidades para un Crecimiento Azul, Actas del XX Congreso de la Asociación Española de Teledetección, pp. 749-752. 2024, Cádiz. (Chapter 4)
- **Boumahdi, M.**, García-Pedrero, A., Lillo-Saavedra, M. and Gonzalo-Martin, C., 2023. Adjustment of Sentinel-3 Spectral Bands With Sentinel-2 to Enhance the Quality of Spatio-Temporally Fused Images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, pp.584-600. (Chapter 3)

In addition, contributions indirectly related to the work performed during this dissertation are listed below.

- Gonzalo-Martín, C., González-Delgado, J., García-Pedrero, A., **Boumahdi, M.**, Lillo-Saavedra, M. A Comparative Analysis of a New Long-term Burned Area Product and High-resolution Burned Area Datasets. European Journal of Remote Sensing (In revision).
- Romero-Bermeo, D., Chuizaca-Espinoza, I. A., Maia, A. B., Ramírez San Miguel, V., Bulatov, D., Helmholz, P., **Boumahdi, M.**, Molano Cárdenas, S. M., Rodrigues Santos, D. C., Lima Raiol, L., and Velastegui-Montoya, A. 2025. Mining in The Southern Ecuadorian Amazon: Hotspots Analysis and Regulatory Efficiency. In IGARSS 2025 IEEE International Geoscience and Remote Sensing Symposium (Accepted).
- González-Delgado, J., Gonzalo-Martín, C., García-Pedrero, A., **Boumahdi, M.**, and Lillo Saavedra, M. 2024. Validation Of A New Long-term Burned Area Product Compared With High-Resolution Burned Area Data Sets. In Proceedings of 13th EARSeL Workshop on Forest Fires : Remote Sensing of Forest Fires: Lessons learned and future challenges under a changing climate.
- **Boumahdi, M.**, Gonzalo-Martin, C., Garcia-Pedrero, A. and Lillo-Saavedra, M., 2024, July. A Random Forest Approach for Generating Daily Surface Reflectance Time Series Integrating MODIS and Sentinel Data. In IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium (pp. 9037-9040). IEEE.
- González-Delgado, J., Gonzalo-Martín, C., García-Pedrero, A., **Boumahdi, M.**, and Lillo-Saavedra, M. 2024. Generación de una serie temporal de área quemada de larga duración. En Teledetección y Cambio Global: Retos y Oportunidades para un Crecimiento Azul, Actas del XX Congreso de la Asociación Española de Teledetección, pp. 257-260. Cádiz.
- García-Pedrero, A., Gonzalo-Martín, C., González-Delgado, J., **Boumahdi, M.**, and Rodríguez-Esparragón, D. 2024. Análisis temporal del NDVI y su relación con la precipitación y la temperatura en el Parque Nacional de Garajonay. En Teledetección y Cambio Global: Retos y Oportunidades para un Crecimiento Azul, Actas del XX

Congreso de la Asociación Española de Teledetección, pp. 619-622. 2024, Cádiz.

- Gonzalo-Martín, C., González-Delgado, J., García-Pedrero, A., **Boumahdi, M.**, Lillo-Saavedra, M., 2023. Integration of Multiple Global Burned Area Products To Improve Fire Monitoring and Management. Conference on Big Data from Space (BiDS'23), pp. 305-308.
- Gonzalo, C., **Boumahdi, M.**, García-Pedrero, A., and Lillo-Saavedra, M., 2022. Estimación de series diarias de reflectancia superficial en el NIR mediante Random Forests. XIX Congreso de la Asociación Española de Teledetección.

References

Aiazzi, B., Baronti, S., & Selva, M. (2007). Improving component substitution pansharpening through multivariate regression of ms + pan data. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(10), 3230–3239.

Alonzo, M., Bookhagen, B., & Roberts, D. A. (2014). Urban tree species mapping using hyperspectral and lidar data fusion. *Remote sensing of environment*, *148*, 70–83.

Bar-Or, R., Altaratz, O., & Koren, I. (2011). Global analysis of cloud field coverage and radiative properties, using morphological methods and modis observations. *Atmospheric Chemistry and Physics*, *11*(1), 191–200.

Belgiu, M., & Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote sensing of environment*, *204*, 509–523.

Benabdelkader, S., & Melgani, F. (2008). Contextual spatiospectral postreconstruction of cloud-contaminated images. *IEEE Geoscience and remote sensing letters*, *5*(2), 204–208.

Bermudez, J. D., Happ, P. N., Oliveira, D. A. B., & Feitosa, R. Q. (2018). Sar to optical image synthesis for cloud removal with generative adversarial networks. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4*, 5–11.

Boulze, H., Korosov, A., & Brajard, J. (2020). Classification of sea ice types in sentinel-1 sar data using convolutional neural networks. *Remote Sensing*, *12*(13), 2165.

Buntilov, V., & Bretschneider, T. (2008). Investigation on image fusion of remotely sensed images with substantially different spectral properties. *Image and Signal Processing for Remote Sensing XIV*, *7109*, 11–22.

Burt, P. J., & Adelson, E. H. (1987). The laplacian pyramid as a compact image code. In *Readings in computer vision* (pp. 671–679). Elsevier.

Caballero, I., Ruiz, J., & Navarro, G. (2019). Sentinel-2 satellites provide near-real time evaluation of catastrophic floods in the west mediterranean. *Water*, *11*(12), 2499.

Campbell, J. B., & Wynne, R. H. (2011). *Introduction to remote sensing*. Guilford press.

Cao, Z., Chen, S., Gao, F., & Li, X. (2020). Improving phenological monitoring of winter wheat by considering sensor spectral response in spatiotemporal image fusion. *Physics and Chemistry of the Earth, Parts A/B/C*, *116*, 102859.

Carper, W., Lillesand, T., Kiefer, R., et al. (1990). The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data. *Photogrammetric Engineering and remote sensing*, *56*(4), 459–467.

Chander, G., Markham, B. L., & Helder, D. L. (2009). Summary of current radiometric calibration coefficients for landsat mss, tm, etm+, and eo-1 ali sensors. *Remote sensing of environment*, *113*(5), 893–903.

Chen, F., Ming, C., Li, J., Wang, C., & Claverie, M. (2018). A comparison of sentinel-2a and sentinel-2b with preliminary results. *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spain*, 22–28.

Chen, Y., Li, C., Ghamisi, P., Jia, X., & Gu, Y. (2017). Deep fusion of remote sensing data for accurate classification. *IEEE Geoscience and Remote Sensing Letters*, *14*(8), 1253–1257.

Chen, Z., Wang, C., Li, J., Xie, N., Han, Y., & Du, J. (2021). Reconstruction bias u-net for road extraction from optical remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 2284–2294.

Choi, J., Yu, K., & Kim, Y. (2010). A new adaptive component-substitution-based satellite image fusion by using partial replacement. *IEEE transactions on geoscience and remote sensing*, *49*(1), 295–309.

Colin, J., Hagolle, O., Landier, L., Coustance, S., Kettig, P., Meygret, A., Osman, J., & Vermote, E. (2023). Assessment of the performance of the atmospheric correction algorithm maja for sentinel-2 surface reflectance estimates. *Remote Sensing*, *15*(10), 2665.

Comba, L., Biglia, A., Aimonino, D. R., Barge, P., Tortia, C., & Gay, P. (2019). 2d and 3d data fusion for crop monitoring in precision agriculture. *2019 IEEE international workshop on metrology for agriculture and forestry (MetroAgriFor)*, 62–67.

Cresson, R., Narçon, N., Gaetano, R., Dupuis, A., Tanguy, Y., May, S., & Commandré, B. (2022). Comparison of convolutional neural networks for cloudy optical images reconstruction from single or multitemporal joint sar and optical images. *arXiv preprint arXiv:2204.00424*.

De Keukelaere, L., Sterckx, S., Adriaensen, S., Knaeps, E., Reusen, I., Giardino, C., Bresciani, M., Hunter, P., Neil, C., Van der Zande, D., et al. (2018). Atmospheric correction of landsat-8/oli and sentinel-2/msi data using icor algorithm: Validation for coastal and inland waters. *European Journal of Remote Sensing*, *51*(1), 525–542.

Debes, C., Merentitis, A., Heremans, R., Hahn, J., Frangiadakis, N., van Kasteren, T., Liao, W., Bellens, R., Pižurica, A., Gautama, S., et al. (2014). Hyperspectral and lidar data fusion: Outcome of the 2013 grss data fusion contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(6), 2405–2418.

Deng, L.-J., Feng, M., & Tai, X.-C. (2019). The fusion of panchromatic and multispectral remote sensing images via tensor-based sparse modeling and hyper-laplacian prior. *Information Fusion*, *52*, 76–89.

Deng, W., Wang, K., Liu, X., Zhang, T., Liu, H., & Liu, J. (2022). Research on remote sensing detection method for distributed subsurface targets inside mountain bodies. *2022 International Conference on Artificial Intelligence and Computer Information Technology (AICIT)*, 1–6.

Dian, R., Li, S., Sun, B., & Guo, A. (2021). Recent advances and new guidelines on hyperspectral and multispectral image fusion. *Information Fusion*, *69*, 40–51.

Donlon, C., Berruti, B., Mecklenberg, S., Nieke, J., Rebhan, H., Klein, U., Buongiorno, A., Mavrocordatos, C., Frerick, J., Seitz, B., et al. (2012). The sentinel-3 mission: Overview

and status. *2012 IEEE International Geoscience and Remote Sensing Symposium*, 1711–1714.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., et al. (2012). Sentinel-2: Esa's optical high-resolution mission for gmes operational services. *Remote sensing of Environment*, *120*, 25–36.

Emelyanova, I. V., McVicar, T. R., Van Niel, T. G., Li, L. T., & Van Dijk, A. I. (2013). Assessing the accuracy of blending landsat–modis surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection. *Remote Sensing of Environment*, *133*, 193–209.

Fuentes Reyes, M., Auer, S., Merkle, N., Henry, C., & Schmitt, M. (2019). Sar-to-optical image translation based on conditional generative adversarial networks—optimization, opportunities and limits. *Remote Sensing*, *11*(17), 2067.

Gao, F., Anderson, M. C., Zhang, X., Yang, Z., Alfieri, J. G., Kustas, W. P., Mueller, R., Johnson, D. M., & Prueger, J. H. (2017). Toward mapping crop progress at field scales through fusion of landsat and modis imagery. *Remote Sensing of Environment*, *188*, 9–25.

Gao, F., Masek, J., Schwaller, M., & 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), 2207–2218.

Gao, J., Yuan, Q., Li, J., Zhang, H., & Su, X. (2020). Cloud removal with fusion of high resolution optical and sar images using generative adversarial networks. *Remote Sensing*, *12*(1), 191.

Geudtner, D., Torres, R., Snoeij, P., Davidson, M., & Rommen, B. (2014). Sentinel-1 system capabilities and applications. *2014 IEEE geoscience and remote sensing symposium*, 1457–1460.

Gevaert, C. M., & García-Haro, F. J. (2015). A comparison of starfm and an unmixing-based algorithm for landsat and modis data fusion. *Remote sensing of Environment*, *156*, 34–44.

Ghamisi, P., Höfle, B., & Zhu, X. X. (2016). Hyperspectral and lidar data fusion using extinction profiles and deep convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(6), 3011–3024.

Ghamisi, P., Rasti, B., Yokoya, N., Wang, Q., Hofle, B., Bruzzone, L., Bovolo, F., Chi, M., Anders, K., Gloaguen, R., et al. (2019). Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, *7*(1), 6–39.

Ghassemian, H. (2016). A review of remote sensing image fusion methods. *Information Fusion*, *32*, 75–89.

Ghosh, P., Mandal, D., Bhattacharya, A., Nanda, M. K., & Bera, S. (2018). Assessing crop monitoring potential of sentinel-2 in a spatio-temporal scale. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, 227–231.

Gilabert, M., Conese, C., & Maselli, F. (1994). An atmospheric correction method for the automatic retrieval of surface reflectances from tm images. *International Journal of Remote Sensing*, *15*(10), 2065–2086.

Grohnfeldt, C., Schmitt, M., & Zhu, X. (2018). A conditional generative adversarial network to fuse sar and multispectral optical data for cloud removal from sentinel-2 images.

IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, 1726–1729.

Hakim, W. L., Achmad, A. R., Eom, J., & Lee, C.-W. (2020). Land subsidence measurement of jakarta coastal area using time series interferometry with sentinel-1 sar data. *Journal of Coastal Research*, *102*(SI), 75–81.

Han, X.-H., Zheng, Y., & Chen, Y.-W. (2019). Multi-level and multi-scale spatial and spectral fusion cnn for hyperspectral image super-resolution. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 4330–4339.

Hilker, T., Wulder, M. A., Coops, N. C., Linke, J., McDermid, G., Masek, J. G., Gao, F., & White, J. C. (2009). A new data fusion model for high spatial-and temporal-resolution mapping of forest disturbance based on landsat and modis. *Remote Sensing of Environment*, *113*(8), 1613–1627.

Horn, B. K. (1981). Hill shading and the reflectance map. *Proceedings of the IEEE*, *69*(1), 14–47.

Hu, J., Mou, L., Schmitt, A., & Zhu, X. X. (2017). Fusionet: A two-stream convolutional neural network for urban scene classification using polar and hyperspectral data in: 2017 joint urban remote sensing event (jurse). *IEEE. https://doi. org/10.1109/JURSE*.

Hu, X., Ren, H., Tansey, K., Zheng, Y., Ghent, D., Liu, X., & Yan, L. (2019). Agricultural drought monitoring using european space agency sentinel 3a land surface temperature and normalized difference vegetation index imageries. *Agricultural and Forest Meteorology*, *279*, 107707.

Huang, B., Zhang, H., Song, H., Wang, J., & Song, C. (2013). Unified fusion of remote-sensing imagery: Generating simultaneously high-resolution synthetic spatial–temporal–spectral earth observations. *Remote sensing letters*, *4*(6), 561–569.

Hutengs, C., & Vohland, M. (2016). Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sensing of Environment*, *178*, 127–141.

Immerzeel, W. W., Lutz, A. F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., Hyde, S., Brumby, S., Davies, B. J., Elmore, A. C., et al. (2020). Importance and vulnerability of the world's water towers. *Nature*, *577*(7790), 364–369.

Jiang, J., Johansen, K., Tu, Y.-H., & McCabe, M. F. (2022). Multi-sensor and multi-platform consistency and interoperability between uav, planet cubesat, sentinel-2, and landsat reflectance data. *GIScience & Remote Sensing*, *59*(1), 936–958.

Jing, R., Duan, F., Lu, F., Zhang, M., & Zhao, W. (2022). Cloud removal for optical remote sensing imagery using the spa-cyclegan network. *Journal of Applied Remote Sensing*, *16*(3), 034520–034520.

Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., Jepsen, M. R., Kuemmerle, T., Meyfroidt, P., Mitchard, E. T., et al. (2016). A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sensing*, *8*(1), 70.

Kahraman, S., & Bacher, R. (2021). A comprehensive review of hyperspectral data fusion with lidar and sar data. *Annual Reviews in Control*, *51*, 236–253.

Kawakami, R., Matsushita, Y., Wright, J., Ben-Ezra, M., Tai, Y.-W., & Ikeuchi, K. (2011). High-resolution hyperspectral imaging via matrix factorization. *CVPR 2011*, 2329–2336.

Khanal, S., Kc, K., Fulton, J. P., Shearer, S., & Ozkan, E. (2020). Remote sensing in agriculture—accomplishments, limitations, and opportunities. *Remote Sensing*, *12*(22), 3783.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Klemas, V. (2015). Remote sensing of floods and flood-prone areas: An overview. *Journal of Coastal Research*, *31*(4), 1005–1013.

Korhonen, J., & You, J. (2012). Peak signal-to-noise ratio revisited: Is simple beautiful? *2012 Fourth international workshop on quality of multimedia experience*, 37–38.

Kruse, F. A., Lefkoff, A., Boardman, y. J., Heidebrecht, K., Shapiro, A., Barloon, P., & Goetz, A. (1993). The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment*, *44*(2-3), 145–163.

Kwarteng, P., & Chavez, A. (1989). Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis. *Photogramm. Eng. Remote Sens*, *55*(1), 339–348.

Laben, C. A., & Brower, B. V. (2000, January). Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening [US Patent 6,011,875].

Lapucci, C., Antonini, A., Böhm, E., Organelli, E., Massi, L., Ortolani, A., Brandini, C., & Maselli, F. (2023). Use of sentinel-3 olci images and machine learning to assess the ecological quality of italian coastal waters. *Sensors*, *23*(22), 9258.

Li, H., Ghamisi, P., Soergel, U., & Zhu, X. X. (2018). Hyperspectral and lidar fusion using deep three-stream convolutional neural networks. *Remote Sensing*, *10*(10), 1649.

Li, J., Li, Y., Cai, R., He, L., Chen, J., & Plaza, A. (2021). Enhanced spatiotemporal fusion via modis-like images. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–17.

Li, W., Cao, D., Peng, Y., & Yang, C. (2021). Msnet: A multi-stream fusion network for remote sensing spatiotemporal fusion based on transformer and convolution. *Remote Sensing*, *13*(18), 3724.

Li, X., Ling, F., Foody, G. M., Ge, Y., Zhang, Y., & Du, Y. (2017). Generating a series of fine spatial and temporal resolution land cover maps by fusing coarse spatial resolution remotely sensed images and fine spatial resolution land cover maps. *Remote Sensing of Environment*, *196*, 293–311.

Li, X., Du, Z., Huang, Y., & Tan, Z. (2021). A deep translation (gan) based change detection network for optical and sar remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *179*, 14–34.

Li, X., Shen, H., Li, H., & Zhang, L. (2016). Patch matching-based multitemporal group sparse representation for the missing information reconstruction of remote-sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(8), 3629–3641.

Lillo-Saavedra, M., & Gonzalo, C. (2006). Spectral or spatial quality for fused satellite imagery? a trade-off solution using the wavelet à trous algorithm. *International Journal of Remote Sensing*, *27*(7), 1453–1464.

Lin, C.-H., Lai, K.-H., Chen, Z.-B., & Chen, J.-Y. (2013). Patch-based information reconstruction of cloud-contaminated multitemporal images. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(1), 163–174.

Lin, C.-H., Tsai, P.-H., Lai, K.-H., & Chen, J.-Y. (2012). Cloud removal from multitemporal satellite images using information cloning. *IEEE transactions on geoscience and remote sensing*, *51*(1), 232–241.

Liu, M., Yang, W., Zhu, X., Chen, J., Chen, X., Yang, L., & Helmer, E. H. (2019). An improved flexible spatiotemporal data fusion (ifsdaf) method for producing high spatiotemporal resolution normalized difference vegetation index time series. *Remote sensing of environment*, *227*, 74–89.

Liu, S., Zhou, J., Qiu, Y., Chen, J., Zhu, X., & Chen, H. (2022). The first model: Spatiotemporal fusion incorrporting spectral autocorrelation. *Remote Sensing of Environment*, *279*, 113111.

Liu, X., Liu, Q., & Wang, Y. (2020). Remote sensing image fusion based on two-stream fusion network. *Information Fusion*, *55*, 1–15.

Lu, C.-H., Ni, C.-F., Chang, C.-P., Yen, J.-Y., & Chuang, R. Y. (2018). Coherence difference analysis of sentinel-1 sar interferogram to identify earthquake-induced disasters in urban areas. *Remote Sensing*, *10*(8), 1318.

Lu, Z., Dzurisin, D., Jung, H.-S., Zhang, J., & Zhang, Y. (2010). Radar image and data fusion for natural hazards characterisation. *International Journal of Image and Data Fusion*, *1*(3), 217–242.

Ma, D., Wu, R., Xiao, D., & Sui, B. (2023). Cloud removal from satellite images using a deep learning model with the cloud-matting method. *Remote Sensing*, *15*(4), 904.

Ma, J., Zhang, W., Marinoni, A., Gao, L., & Zhang, B. (2018). An improved spatial and temporal reflectance unmixing model to synthesize time series of landsat-like images. *Remote Sensing*, *10*(9), 1388.

Ma, J., Yu, W., Chen, C., Liang, P., Guo, X., & Jiang, J. (2020). Pan-gan: An unsupervised pan-sharpening method for remote sensing image fusion. *Information Fusion*, *62*, 110–120.

Malinowski, R., Lewiński, S., Rybicki, M., Gromny, E., Jenerowicz, M., Krupiński, M., Nowakowski, A., Wojtkowski, C., Krupiński, M., Krätzschmar, E., et al. (2020). Automated production of a land cover/use map of europe based on sentinel-2 imagery. *Remote Sensing*, *12*(21), 3523.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, *11*(7), 674–693.

Mao, R., Li, H., Ren, G., & Yin, Z. (2022). Cloud removal based on sar-optical remote sensing data fusion via a two-flow network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 7677–7686.

Marsh, C. B., Pomeroy, J. W., & Spiteri, R. J. (2012). Implications of mountain shading on calculating energy for snowmelt using unstructured triangular meshes. *Hydrological Processes*, *26*(12), 1767–1778.

Mashimbye, Z. E., & Loggenberg, K. (2023). A scoping review of landform classification using geospatial methods. *Geomatics*, *3*(1), 93–114.

Meng, X., Shen, H., Yuan, Q., Li, H., Zhang, L., & Sun, W. (2018). Pansharpening for cloud-contaminated very high-resolution remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(5), 2840–2854.

Meraner, A., Ebel, P., Zhu, X. X., & 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*, 333–346.

Morales, B., Lizama, E., Somos-Valenzuela, M. A., Lillo-Saavedra, M., Chen, N., & Fustos, I. (2021). A comparative machine learning approach to identify landslide triggering factors in northern chilean patagonia. *Landslides*, *18*(8), 2767–2784.

Munawar, H. S., Hammad, A. W., & Waller, S. T. (2022). Remote sensing methods for flood prediction: A review. *Sensors*, *22*(3), 960.

Nascimento, J. M., & Dias, J. M. (2005). Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE transactions on Geoscience and Remote Sensing*, *43*(4), 898–910.

Nationen, V. (2015). *Transforming our world: The 2030 agenda for sustainable development: A/res/70/1*. United Nations, Division for Sustainable Development.

Ng, M. K.-P., Yuan, Q., Yan, L., & Sun, J. (2017). An adaptive weighted tensor completion method for the recovery of remote sensing images with missing data. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(6), 3367–3381.

Nkonya, E. M., et al. (2019). Climate change and land: An ipcc special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.

Otazu, X., Gonzalez-Audicana, M., Fors, O., & Nunez, J. (2005). Introduction of sensor spectral response into image fusion methods. application to wavelet-based methods. *IEEE Transactions on Geoscience and Remote Sensing*, *43*(10), 2376–2385. https://doi.org/10.1109/TGRS.2005.856106

Paolini, L., Grings, F., Sobrino, J. A., Jiménez Muñoz, J. C., & Karszenbaum, H. (2006). Radiometric correction effects in landsat multi-date/multi-sensor change detection studies. *International Journal of Remote Sensing*, *27*(4), 685–704.

Peng, W., Zhou, J., Yang, C.-j., & He, Z. (2008). Analysis on slope uncertainty based on different resolution level dem: A case study. *Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Advanced Spatial Data Models and Analyses*, *7146*, 295–306.

Persson, M., Lindberg, E., & Reese, H. (2018). Tree species classification with multi-temporal sentinel-2 data. *Remote Sensing*, *10*(11), 1794.

Pettorelli, N., Laurance, W. F., O'Brien, T. G., Wegmann, M., Nagendra, H., & Turner, W. (2014). Satellite remote sensing for applied ecologists: Opportunities and challenges. *Journal of Applied Ecology*, *51*(4), 839–848.

Pettorelli, N., Schulte to Bühne, H., Tulloch, A., Dubois, G., Macinnis-Ng, C., Queirós, A. M., Keith, D. A., Wegmann, M., Schrodt, F., Stellmes, M., et al. (2018). Satellite remote sensing of ecosystem functions: Opportunities, challenges and way forward. *Remote Sensing in Ecology and Conservation*, *4*(2), 71–93.

Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V. R., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 data for land cover/use mapping: A review. *Remote Sensing*, *12*(14), 2291.

Ping, B., Meng, Y., & Su, F. (2018). An enhanced linear spatio-temporal fusion method for blending landsat and modis data to synthesize landsat-like imagery. *Remote Sensing*, *10*(6), 881.

Potin, P., Bargellini, P., Laur, H., Rosich, B., & Schmuck, S. (2012). Sentinel-1 mission operations concept. *2012 IEEE international geoscience and remote sensing symposium*, 1745–1748.

Potin, P., Rosich, B., Grimont, P., Miranda, N., Shurmer, I., O'Connell, A., Torres, R., & Krassenburg, M. (2016). Sentinel-1 mission status. *Proceedings of EUSAR 2016: 11th European conference on synthetic aperture radar*, 1–6.

Potin, P., Rosich, B., Miranda, N., Grimont, P., Shurmer, I., O'Connell, A., Krassenburg, M., & Gratadour, J.-B. (2019). Copernicus sentinel-1 constellation mission operations status. *IGARSS 2019-2019 IEEE international geoscience and remote sensing symposium*, 5385–5388.

Prudente, V. H. R., Martins, V. S., Vieira, D. C., e Silva, N. R. d. F., Adami, M., & Sanches, I. D. (2020). Limitations of cloud cover for optical remote sensing of agricultural areas across south america. *Remote Sensing Applications: Society and Environment*, *20*, 100414.

Quan, J., Zhan, W., Ma, T., Du, Y., Guo, Z., & Qin, B. (2018). An integrated model for generating hourly landsat-like land surface temperatures over heterogeneous landscapes. *Remote Sensing of Environment*, *206*, 403–423.

Reddy, B. S. J. N., & Sasikala, D. (2023). Cloud cover removal from remote sensing data using gans based on attention mechanism. *2023 Seventh International Conference on Image Information Processing (ICIIP)*, 651–657.

Reiche, J., Lucas, R., Mitchell, A. L., Verbesselt, J., Hoekman, D. H., Haarpaintner, J., Kellndorfer, J. M., Rosenqvist, A., Lehmann, E. A., Woodcock, C. E., et al. (2016). Combining satellite data for better tropical forest monitoring. *Nature Climate Change*, *6*(2), 120–122.

Reyes-Muñoz, P., Pipia, L., Salinero-Delgado, M., Belda, S., Berger, K., Estévez, J., Morata, M., Rivera-Caicedo, J. P., & Verrelst, J. (2022). Quantifying fundamental vegetation traits over europe using the sentinel-3 olci catalogue in google earth engine. *Remote sensing*, *14*(6), 1347.

Ronneberger, O. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., et al. (1974). Monitoring vegetation systems in the great plains with erts. *NASA Spec. Publ*, *351*(1), 309.

Roy, D. P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., & Lindquist, E. (2008). Multi-temporal modis–landsat data fusion for relative radiometric normalization, gap filling, and prediction of landsat data. *Remote Sensing of Environment*, *112*(6), 3112–3130.

Rufin, P., Frantz, D., Yan, L., & Hostert, P. (2020). Operational coregistration of the sentinel-2a/b image archive using multitemporal landsat spectral averages. *IEEE Geoscience and Remote Sensing Letters*, *18*(4), 712–716.

Salcedo, A. P., & Cogliati, M. G. (2014). Snow cover area estimation using radar and optical satellite information. *Atmospheric and Climate Sciences*, *2014*.

Sarukkai, V., Jain, A., Uzkent, B., & Ermon, S. (2020). Cloud removal from satellite images using spatiotemporal generator networks. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1796–1805.

Sauter, T. (2020). Revisiting extreme precipitation amounts over southern south america and implications for the patagonian icefields. *Hydrology and Earth System Sciences*, *24*(4), 2003–2016.

Scheffler, D., Hollstein, A., Diedrich, H., Segl, K., & Hostert, P. (2017). Arosics: An automated and robust open-source image co-registration software for multi-sensor satellite data. *Remote sensing*, *9*(7), 676.

Schmitt, M., & Zhu, X. X. (2016). Data fusion and remote sensing: An ever-growing relationship. *IEEE Geoscience and Remote Sensing Magazine*, *4*(4), 6–23.

Schott, J. R. (2007). *Remote sensing: The image chain approach*. Oxford University Press.

Senf, C., Leitão, P. J., Pflugmacher, D., van der Linden, S., & Hostert, P. (2015). Mapping land cover in complex mediterranean landscapes using landsat: Improved classification accuracies from integrating multi-seasonal and synthetic imagery. *Remote Sensing of Environment*, *156*, 527–536.

Shang, J., Liu, J., Poncos, V., Geng, X., Qian, B., Chen, Q., Dong, T., Macdonald, D., Martin, T., Kovacs, J., et al. (2020). Detection of crop seeding and harvest through analysis of time-series sentinel-1 interferometric sar data. *Remote Sensing*, *12*(10), 1551.

Shangguan, Y., Li, J., Chen, Z., Ren, L., & Hua, Z. (2024). Multi-scale attention fusion graph network for remote sensing building change detection. *IEEE Transactions on Geoscience and Remote Sensing*.

Shen, H. (2012). Integrated fusion method for multiple temporal-spatial-spectral images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *39*, 407–410.

Shen, H., Wu, J., Cheng, Q., Aihemaiti, M., Zhang, C., & Li, Z. (2019). A spatiotemporal fusion based cloud removal method for remote sensing images with land cover changes. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(3), 862–874.

Shen, H., Wu, P., Liu, Y., Ai, T., Wang, Y., & Liu, X. (2013). A spatial and temporal reflectance fusion model considering sensor observation differences. *International journal of remote sensing*, *34*(12), 4367–4383.

Simoes, M., Bioucas-Dias, J., Almeida, L. B., & Chanussot, J. (2014). A convex formulation for hyperspectral image superresolution via subspace-based regularization. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(6), 3373–3388.

Sola, I., García-Martín, A., Sandonís-Pozo, L., Álvarez-Mozos, J., Pérez-Cabello, F., González-Audícana, M., & Llovería, R. M. (2018). Assessment of atmospheric correction methods for sentinel-2 images in mediterranean landscapes. *International journal of applied earth observation and geoinformation*, *73*, 63–76.

Solberg, R., Huseby, R. B., Koren, H., & Malnes, E. (2008). Time-series fusion of optical and sar data for snow cover area mapping. *Proceedings of the EARSeL Land Ice and Snow Special Interest Group Workshop, Berne, Switzerland*, *2224*.

Solórzano, J. V., Mas, J. F., Gallardo-Cruz, J. A., Gao, Y., & de Oca, A. F.-M. (2023). Deforestation detection using a spatio-temporal deep learning approach with synthetic

aperture radar and multispectral images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *199*, 87–101.

Son, N.-T., Chen, C.-F., Chang, L.-Y., Chen, C.-R., Sobue, S.-I., Minh, V.-Q., Chiang, S.-H., Nguyen, L.-D., & Lin, Y.-W. (2016). A logistic-based method for rice monitoring from multitemporal modis-landsat fusion data. *European Journal of Remote Sensing*, *49*(1), 39–56.

Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., & Macomber, S. A. (2001). Classification and change detection using landsat tm data: When and how to correct atmospheric effects? *Remote sensing of Environment*, *75*(2), 230–244.

Song, H., & Huang, B. (2012). Spatiotemporal satellite image fusion through one-pair image learning. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(4), 1883–1896.

Song, H., Huang, B., Zhang, K., & Zhang, H. (2014). Spatio-spectral fusion of satellite images based on dictionary-pair learning. *Information Fusion*, *18*, 148–160.

Song, H., Liu, Q., Wang, G., Hang, R., & Huang, B. (2018). Spatiotemporal satellite image fusion using deep convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(3), 821–829.

Spoto, F., Sy, O., Laberinti, P., Martimort, P., Fernandez, V., Colin, O., Hoersch, B., & Meygret, A. (2012). Overview of sentinel-2. *2012 IEEE international geoscience and remote sensing symposium*, 1707–1710.

Starck, J.-L., Fadili, J., & Murtagh, F. (2007). The undecimated wavelet decomposition and its reconstruction. *IEEE transactions on image processing*, *16*(2), 297–309.

Sun, G., Ranson, K. J., Kharuk, V. I., & Kovacs, K. (2003). Validation of surface height from shuttle radar topography mission using shuttle laser altimeter. *Remote Sensing of Environment*, *88*(4), 401–411.

Tarolli, P., & Straffelini, E. (2020). Agriculture in hilly and mountainous landscapes: Threats, monitoring and sustainable management. *Geography and sustainability*, *1*(1), 70–76.

Toming, K., Kutser, T., Uiboupin, R., Arikas, A., Vahter, K., & Paavel, B. (2017). Mapping water quality parameters with sentinel-3 ocean and land colour instrument imagery in the baltic sea. *Remote Sensing*, *9*(10), 1070.

Torres, D. L., Turnes, J. N., Soto Vega, P. J., Feitosa, R. Q., Silva, D. E., Marcato Junior, J., & Almeida, C. (2021). Deforestation detection with fully convolutional networks in the amazon forest from landsat-8 and sentinel-2 images. *Remote Sensing*, *13*(24), 5084.

Townshend, J. R., Justice, C. O., Gurney, C., & McManus, J. (1992). The impact of misregistration on change detection. *IEEE Transactions on Geoscience and remote sensing*, *30*(5), 1054–1060.

Turner, W., Rondinini, C., Pettorelli, N., Mora, B., Leidner, A. K., Szantoi, Z., Buchanan, G., Dech, S., Dwyer, J., Herold, M., et al. (2015). Free and open-access satellite data are key to biodiversity conservation. *Biological Conservation*, *182*, 173–176.

Van Den Eeckhaut, M., Poesen, J., Verstraeten, G., Vanacker, V., Moeyersons, J., Nyssen, J., & Van Beek, L. (2005). The effectiveness of hillshade maps and expert knowledge in mapping old deep-seated landslides. *Geomorphology*, *67*(3-4), 351–363.

Vanhellemont, Q., & Ruddick, K. (2021). Atmospheric correction of sentinel-3/olci data for mapping of suspended particulate matter and chlorophyll-a concentration in belgian turbid coastal waters. *Remote Sensing of Environment*, *256*, 112284.

Villaescusa-Nadal, J. L., Franch, B., Roger, J.-C., Vermote, E. F., Skakun, S., & Justice, C. (2019). Spectral adjustment model's analysis and application to remote sensing data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(3), 961–972.

Wang, H., Skau, E., Krim, H., & Cervone, G. (2018). Fusing heterogeneous data: A case for remote sensing and social media. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(12), 6956–6968.

Wang, J., & Huang, B. (2017). A rigorously-weighted spatiotemporal fusion model with uncertainty analysis. *Remote Sensing*, *9*(10), 990.

Wang, J., Li, W., Gao, Y., Zhang, M., Tao, R., & Du, Q. (2022). Hyperspectral and sar image classification via multiscale interactive fusion network. *IEEE Transactions on Neural Networks and Learning Systems*, *34*(12), 10823–10837.

Wang, Q., & Atkinson, P. M. (2018). Spatio-temporal fusion for daily sentinel-2 images. *Remote Sensing of Environment*, *204*, 31–42.

Wang, X., Song, L., Feng, Y., & Zhu, J. (2025). S3f2net: Spatial-spectral-structural feature fusion network for hyperspectral image and lidar data classification. *IEEE Transactions on Circuits and Systems for Video Technology*.

Wang, Y., Zhang, W., Chen, W., Chen, C., & Liang, Z. (2024). Mffnet: Multimodal feature fusion network for synthetic aperture radar and optical image land cover classification. *Remote Sensing*, *16*(13), 2459.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE transactions on image processing*, *13*(4), 600–612.

Weng, Q., Fu, P., & Gao, F. (2014). Generating daily land surface temperature at landsat resolution by fusing landsat and modis data. *Remote sensing of environment*, *145*, 55–67.

Wu, M., Niu, Z., Wang, C., Wu, C., & Wang, L. (2012). Use of modis and landsat time series data to generate high-resolution temporal synthetic landsat data using a spatial and temporal reflectance fusion model. *Journal of Applied Remote Sensing*, *6*(1), 063507–063507.

Wu, R., Liu, G., Lv, J., Fu, Y., Bao, X., Shama, A., Cai, J., Sui, B., Wang, X., & Zhang, R. (2023). An innovative approach for effective removal of thin clouds in optical images using convolutional matting model. *Remote Sensing*, *15*(8), 2119.

Xie, D., Zhang, J., Zhu, X., Pan, Y., Liu, H., Yuan, Z., & Yun, Y. (2016). An improved starfm with help of an unmixing-based method to generate high spatial and temporal resolution remote sensing data in complex heterogeneous regions. *Sensors*, *16*(2), 207.

Xiong, C., Yang, J., Pan, J., Lei, Y., & Shi, J. (2022). Mountain snow depth retrieval from optical and passive microwave remote sensing using machine learning. *IEEE Geoscience and Remote Sensing Letters*, *19*, 1–5.

Xu, M., Deng, F., Jia, S., Jia, X., & Plaza, A. J. (2022). Attention mechanism-based generative adversarial networks for cloud removal in landsat images. *Remote sensing of environment*, *271*, 112902.

Xu, M., Jia, X., Pickering, M., & Jia, S. (2019). Thin cloud removal from optical remote sensing images using the noise-adjusted principal components transform. *ISPRS Journal of Photogrammetry and Remote Sensing*, *149*, 215–225.

Xu, M., Pickering, M., Plaza, A. J., & Jia, X. (2015). Thin cloud removal based on signal transmission principles and spectral mixture analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(3), 1659–1669.

Xu, Y., Huang, B., Xu, Y., Cao, K., Guo, C., & Meng, D. (2015). Spatial and temporal image fusion via regularized spatial unmixing. *IEEE Geoscience and Remote Sensing Letters*, *12*(6), 1362–1366.

Xue, J., Leung, Y., & Fung, T. (2017). A bayesian data fusion approach to spatio-temporal fusion of remotely sensed images. *Remote Sensing*, *9*(12), 1310.

Yang, J., Zhou, J., Göttsche, F.-M., Long, Z., Ma, J., & Luo, R. (2020). Investigation and validation of algorithms for estimating land surface temperature from sentinel-3 slstr data. *International Journal of Applied Earth Observation and Geoinformation*, *91*, 102136.

Yang, J., Zhao, Y.-Q., & Chan, J. C.-W. (2018). Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network. *Remote Sensing*, *10*(5), 800.

Yang, J., Fu, X., Hu, Y., Huang, Y., Ding, X., & Paisley, J. (2017). Pannet: A deep network architecture for pan-sharpening. *Proceedings of the IEEE international conference on computer vision*, 5449–5457.

Yang, S., Zhou, B., Lou, H., Wu, Z., Wang, S., Zhang, Y., Pan, Z., & Li, C. (2022). Remote sensing hydrological indication: Responses of hydrological processes to vegetation cover change in mid-latitude mountainous regions. *Science of the Total Environment*, *851*, 158170.

Zhang, F., Zhu, X., & Liu, D. (2014). Blending modis and landsat images for urban flood mapping. *International journal of remote sensing*, *35*(9), 3237–3253.

Zhang, H., Song, Y., Han, C., & Zhang, L. (2020). Remote sensing image spatiotemporal fusion using a generative adversarial network. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(5), 4273–4286.

Zhang, J. (2010). Multi-source remote sensing data fusion: Status and trends. *International Journal of Image and Data Fusion*, *1*(1), 5–24.

Zhang, J., & Lin, X. (2017). Advances in fusion of optical imagery and lidar point cloud applied to photogrammetry and remote sensing. *International Journal of Image and Data Fusion*, *8*(1), 1–31.

Zhang, Q., Yuan, Q., Zeng, C., Li, X., & Wei, Y. (2018). Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(8), 4274–4288. https://doi.org/10.1109/TGRS.2018.2810208

Zhang, X., Qiu, Z., Peng, C., & Ye, P. (2022). Removing cloud cover interference from sentinel-2 imagery in google earth engine by fusing sentinel-1 sar data with a cnn model. *International Journal of Remote Sensing*, *43*(1), 132–147.

Zhang, Y., & Jiang, J. (2014). Why optical images are easier to understand than radar images?–from the electromagnetic scattering and signal point of view. *PIERS Proceedings*.

Zhang, Z. (2021). Retracted article: Mountain rainfall forecast and regional environmental economic development model based on remote sensing images. *Arabian Journal of Geosciences*, *14*(12), 1176.

Zhao, X., Tao, R., Li, W., Li, H.-C., Du, Q., Liao, W., & Philips, W. (2020). Joint classification of hyperspectral and lidar data using hierarchical random walk and deep

cnn architecture. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(10), 7355–7370.

Zhong, C., Liu, Y., Gao, P., Chen, W., Li, H., Hou, Y., Nuremanguli, T., & Ma, H. (2020). Landslide mapping with remote sensing: Challenges and opportunities. *International Journal of Remote Sensing*, *41*(4), 1555–1581.

Zhou, J., Chen, J., Chen, X., Zhu, X., Qiu, Y., Song, H., Rao, Y., Zhang, C., Cao, X., & Cui, X. (2021). Sensitivity of six typical spatiotemporal fusion methods to different influential factors: A comparative study for a normalized difference vegetation index time series reconstruction. *Remote Sensing of Environment*, *252*, 112130.

Zhou, S., & Cheng, J. (2023). A new bottom-of-atmosphere (boa) radiance-based hybrid method for estimating clear-sky surface longwave upwelling radiation from modis data. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–25.

Zhu, X., Cai, F., Tian, J., & Williams, T. K.-A. (2018). Spatiotemporal fusion of multisource remote sensing data: Literature survey, taxonomy, principles, applications, and future directions. *Remote Sensing*, *10*(4), 527.

Zhu, X., Chen, J., Gao, F., Chen, X., & Masek, J. G. (2010). An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sensing of Environment*, *114*(11), 2610–2623.

Zhu, X., Helmer, E. H., Gao, F., Liu, D., Chen, J., & Lefsky, M. A. (2016). A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote Sensing of Environment*, *172*, 165–177.

Zhu, X., Zhan, W., Zhou, J., Chen, X., Liang, Z., Xu, S., & Chen, J. (2022). A novel framework to assess all-round performances of spatiotemporal fusion models. *Remote Sensing of Environment*, *274*, 113002.

Zhukov, B., Oertel, D., Lanzl, F., & Reinhackel, G. (1999). Unmixing-based multisensor multiresolution image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(3), 1212–1226.

Appendix A

Table 1: Dates of the pairs for both sites

Pair	Waterbank	Maspalomas
1	20190109	20190106
2	20190310	20190205
3	20190330	20190220
4	20190514	20190225
5	20190519	20190416
6	20190603	20190426
7	20190608	20190501
8	20190618	20190630
9	20190623	20190715
10	20190628	20190725
11	20190703	20190804
12	20190713	20190809
13	20190723	20190814
14	20190728	20190824
15	20190802	20190829
16	20190807	20190928
17	20190812	20191102
18	20190817	20191127
19	20190822	20191222
20	20190827	20191227
21	20190901	20200101
22	20190911	20200210
23	20190921	20200215
24	20191105	20200220
25	20191110	20200225
26	20191225	20200301
27	20200319	20200430
28	20200324	20200505
29	20200403	20200525
30	20200423	20200624
31	20200428	20200629
32	20200503	20200704
33	20200508	20200709
34	20200513	20200714
35	20200617	20200719
36	20200627	20200729
37	20200717	20200813
38	20200727	20200818
39	20200806	20200823
40	20201104	20201121
41	20201124	20201211