

**Canopy Height Assessment in South Australian
Pinus Radiata Plantations using Sentinel-1: A
Comparative Analysis between INSAR and
Machine Learning Algorithms**

By

Gauravsingh Cheekhooree

*Thesis
Submitted to Flinders University
for the degree of*

Master of Geospatial Information Science

College of Science and Engineering

29th January 2024

TABLE OF CONTENTS

TABLE OF CONTENTS	I
GLOSSARY	IV
ABSTRACT	VII
DECLARATION	IX
ACKNOWLEDGEMENTS	X
LIST OF FIGURES	XI
LIST OF TABLES	XIII
LIST OF EQUATIONS	XIV
CHAPTER 1 INTRODUCTION	1
Remote sensing	1
Passive optical remote sensing.....	2
Active optical remote sensing	2
Issues with Passive and Active Optical Remote Sensing	2
Microwave Remote Sensing	3
Combination of Remote Sensing Techniques	3
Study area.....	4
Research aims	5
CHAPTER 2 INTRODUCTION TO SAR	6
SAR resolution	7
Azimuth resolution	7
Range resolution.....	7
SAR band Wavelength	7
SAR band penetration	8
SAR data characteristics	8
Object Dielectric Constant and Roughness.....	8
SAR polarisation	9
Interferometric SAR (INSAR).....	9
Types of INSAR systems	10
Along track INSAR.....	10
Repeat pass INSAR.....	11
INSAR principles	11
CHAPTER 3 LITERATURE REVIEW:	13
Operational SAR systems	13
Operational optical passive remote sensing systems.....	13
Operational active optical remote sensing systems	14
Literature review scope	15
Sentinel-1 INSAR for tree canopy height estimations	15
Sentinel-1 as a data source in MLM for tree canopy height estimations	17

X band SAR data for canopy height estimation.....	18
L band SAR data for canopy height estimation	19
Literature review outcomes and gaps	21
Summary of X and L band usage for tree height estimation	21
Summary of Sentinel-1 INSAR usage for tree height estimation	21
Gaps identified in literature for Sentinel-1 INSAR usage for tree height estimation	21
Summary of MLMs using Sentinel-1 for tree height estimation.....	22
Gaps identified in the MLMs using Sentinel-1 for tree height estimation.....	22
Research questions.....	22
CHAPTER 4 METHODS:	23
Processing of LIDAR ground truth data	23
Investigation of data.....	23
Selection of DSM and DTM Interpolation Technique.....	24
Overview of the interpolation concept	24
Overview of the interpolation techniques available.....	24
Interpolation techniques for DSM generation	25
Generation of LIDAR ground truth data.....	27
Setting up extraction areas based on completed thinning cycles	29
INSAR DSM Generation.....	30
Sentinel-1 description and data selection.....	30
INSAR processing workflow.....	32
Filtering methodologies.....	36
Pre-processing of INSAR DSM before adjustment.....	36
Adjustment of DSM.....	36
Generation of INSAR fusion CHM.....	38
RF MLM CHM generation.....	38
RF MLM input parameters	39
Training Dataset	39
Predictor Variables.....	39
Auxiliary data	40
Comparing INSAR fusion CHM with RF MLM CHM using LIDAR ground truth CHM.....	42
Resampling.....	42
Extraction of Canopy Height from Polygons.....	42
CHAPTER 5 RESULTS	44
Breakdown of the study area based on thinning cycles.....	44
Visualisation of the INSAR process data transformation	44
Best filtering method for INSAR DSM generation.....	49
INSAR DSM coherence analysis.....	50
INSAR DSM intensity analysis	51
PSM and GCP locations	52
PSM ADJUSTMENT	53

GCP ADJUSTMENT	54
RF MLM random forest variable importance	55
RF MLM accuracy assessment.....	56
INSAR fusion and RF MLM CHMs comparison with LIDAR CHM	57
CHAPTER 6 DISCUSSION	64
Overview of the INSAR DSM generation process.....	64
Understanding the impact of the different filtering processes on the INSAR accuracy.....	64
Effect of coherence of INSAR DSM accuracy	65
Effect of Pruning and Thinning on coherence.....	65
Effect of tree flexibility on coherence.....	66
Investigating the effect of SAR signal intensity on the INSAR accuracy.	67
Comparison between the GCP and PSM INSAR DSM adjustment methods.....	68
Limitations of INSAR based on climatic conditions and possible improvements.....	69
Effect of low resolution DEMs on the INSAR process and possible improvements	70
Limitations in the INSAR process due to SNAP	70
INSAR limitations over vegetated areas.....	70
RF MLM Overview.....	72
Selection of input variables	72
Comparison of Variable Importance.....	72
RF MLM Performance.....	74
Possible Improvements.....	74
Comparison between INSAR fusion CHM and RF MLM CGM with ground truth.....	75
Overall limitations of this study	77
Possible improvement by merging INSAR data and deep learning methodologies	77
CHAPTER 7 CONCLUSION.....	78
Further Research	78
CHAPTER 8 REFERENCES	80
CHAPTER 9 APPENDICES	93
Appendix A: Plantation years of <i>Pinus Radiata</i> present in the study area, adapted from ForestrySA (2016)	93
Appendix B: Code for the RF MLM adapted from Courage (2021) and used in the GEE platform	94

GLOSSARY

Abbreviation	Definition
AHD	Australian Height Datum
ALS	Airborne Laser Scanner
BM	Band Maths
CHM	Canopy Height Model
DEM	Digital Elevation Model
DSM	Digital Surface Model
DTM	Digital Terrain Model
EBK	Empirical Bayesian Kriging Method
ESRI	Environmental Systems Research Institute
ESA	European Space Agency
GCP	Ground Control Points
GDA	Geocentric Datum of Australia
GEDI	Global Ecosystem Dynamics Investigation
GFM	Goldstein Filtering Mask
GLCM	Gray-level Co-occurrence Matrix
GRD	Ground Range Detected
H	Reference Height
h	Topographic Height
IW	Interferometric Wide-Swath Mode
K_z	Vertical Wave Number
LIDAR	Light Detection and Ranging

ML	Multilooking
MLM	Machine Learning Model
NFR	Native Forest Reserve
NIR	Near Infrared
PSM	Permanent Scattering Marks
RF	Random Forest
RH	Height to the Ground
RMSE	Root Mean Square Error
RVoG	Random Volume over Ground Model
SAR	Synthetic Aperture Radar
SPC	Scattering Phase Centre
SNAP	Sentinel Application Platform
SWIR	Short Wave Infrared
SLC	Single Look Complex
SNAPHU	Statistical-Cost, Network-Flow Algorithm for Phase Unwrapping
SPC	Scattering Phase Centre
SRTM	Shuttle Radar Topography Mission
SWIR	Short Wave Infrared
T0	No Thinning Processes Applied
T1	1 Thinning Process Applied
T2	2 Thinning Processes Applied
T3	3 Thinning Processes Applied

$\Delta\phi$	Phase difference
γ	Correlation
Θ	Incidence Angle
HH	Horizontal send, Horizontal Receive
HV	Horizontal send, Vertical Receive
VH	Vertical Send, Horizontal Receive
VV	Vertical Send, Vertical Receive

ABSTRACT

In forestry applications, tree heights through their inclusion in specific allometric equations can be used to infer above-ground biomass and volume at both individual tree and plantation scales. However, the traditional field-based methods used, while accurate, are often cost-prohibitive due to extensive and time-consuming fieldwork. Remote sensing technologies, like airborne LIDAR, offer non-destructive alternatives by swiftly surveying large areas through the generation of Digital Surface Models (DSM) and Digital Ground Models (DGM), leading to the creation of a Canopy Height Model (CHM). Despite their efficacy, these surveys remain costly and often infrequent.

This study investigates the effectiveness of Synthetic Aperture Radar (SAR) to determine tree height using LIDAR data as the ground truth, focusing on Mount Panorama and Knot Hill Nature Forest Reserves in South Australia. These areas contain commercially planted *Pinus radiata* subjected to thinning cycles over their lifetimes (ranging from no thinning - T0, to one thinning - T1, two thinnings - T2, and three thinnings - T3). Initially, SAR VV data from the on-going C-band Sentinel-1 mission is used in interferometry (INSAR) to generate a corrected DSM from which a CHM is generated using the airborne LIDAR DGM. The second approach integrates Sentinel-1 SAR VV, VH, and their derived texture data with the multispectral bands from the Sentinel-2 mission and a DGM from the SRTM mission. These datasets feed into a random forest (RF) machine learning model (MLM), interpolating gaps in spaceborne LIDAR data collected by the GEDI instrument, resulting in another CHM.

Comparing the generated CHMs with the LIDAR ground truth CHM at the 90th height percentile revealed that only under T1 conditions, the INSAR method generated usable results with a 5.6% height underestimate and an RMSE of 4.2m. However, the RF MLM was appropriate under both T1 and T2 conditions, with an overestimated height of 9.3% and an RMSE of 2.3m, and a 4.4% height underestimate and an RMSE of 1.9m, respectively. Given the prevalence of T1 and T2 conditions in the study area, the RF MLM emerges as a potentially more suitable option due to its lower overall RMSE. Under T0 and T3 conditions, both the INSAR and RF MLM methods exhibit significant discrepancies from the ground truth, which can be attributed to the limited area in the study site under these conditions.

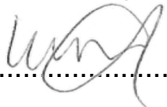
However, Sentinel-1 C-band data may not be optimal for INSAR over forests due to its upper canopy penetration and the current 12-day temporal resolution; these factors further limit its accuracy over forested areas which generally already exhibit low SAR coherence. Furthermore, the lack of GEDI input data over certain forested areas may have resulted in interpolation over larger data gaps than expected. Nevertheless, it was found that currently Sentinel-1 SAR data when used in combination with other optical and topographic data in MLMs results in more accurate canopy height predictions than when it is used directly through INSAR.

Further research using X-band SAR data that is better suited for INSAR, with its lower canopy penetration and higher temporal resolution, as well as the combination of SAR data with higher canopy penetration from the upcoming NISAR and BIOMASS missions in the RF MLM, could improve the accuracy of both methods for tree height estimations.

DECLARATION

I certify that this thesis:

1. does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university
2. and the research within will not be submitted for any other future degree or diploma without the permission of Flinders University; and
3. to the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

Signed.....

Date.....20/01/2024.....

ACKNOWLEDGEMENTS

I extend my deepest gratitude to my supervisor, David Bruce, whose unwavering support guided me through the challenging process of selecting a research focus. His introduction to the International Space University (ISU) proved pivotal, and the intensive 5-week course at ISU solidified my passion for remote sensing. I am thankful for his dedication in reviewing my work during unconventional hours and providing invaluable feedback. The opportunity for an exploration phase, where I could experiment with new software, was particularly enriching, and I appreciate the enthusiasm shown when some of these methods yielded success.

Special thanks go to Marcio Dasilva, who offered crucial assistance during the early stages of my thesis. His support was instrumental during moments when the complexity of the thesis seemed overwhelming. I am also grateful for his encouragement to respond affirmatively to the ISU's call for participation in the 3-month Space Studies Program (SSP), hosted in his home country—an experience that proved to be life-changing.

I would like to express my appreciation to the entire GIS team, from tutors to lecturers, for presenting a multitude of engaging topics throughout the master's program. The Belair field camp was a highlight, providing a hands-on demonstration of the practical application of our studies, even though it became evident that setting up surveying gear was not my strong suit!

Lastly, heartfelt thanks to my family and friends for their continuous support and encouragement. Without them, this thesis would never have come to fruition.

LIST OF FIGURES

Figure 1.1: Location and extent of the study area investigated.....	4
Figure 2.1: Typical configuration of a SAR sensor adapted from NISAR (2018)	6
Figure 2.2: Depiction of the different wavelength of the microwave signals used in SAR systems, and the corresponding naming of the section of the wavelengths into bands (Sartoretto, 2021).....	7
Figure 2.3: Canopy penetration of the different SAR bands (Wallington, 2006).	8
Figure 2.4: Types of scattering in SAR data, where the blue arrows correspond to rough surface scattering, the green arrows correspond to volume scattering, and the red arrows correspond to double bounce scattering (Meyer, 2019).	9
Figure 4.1: Flowchart showing the processing steps in processing the LIDAR ground truth data into DSM, DTM and CHM which were used in 1) The INSAR CHM generation 2) The accuracy assessment between the INSAR fusion CHM and the RF MLM CHM	27
Figure 4.2: Graphical overview adapted from ForestrySA (2018) of the timeline for the thinning process in ForestrySA <i>Pinus Radiata</i> plantations, highlighting the approximate years at which the thinning process takes place	29
Figure 4.3: Flowchart showing the INSAR processing steps carried out in ESA SNAP for the DSM generation and the subsequent DSM adjustment and INSAR fusion CHM generation in ESRI ArcGIS pro	32
Figure 4.4: Visualisation of the variation in strength of the Goldstein phase filtering where as shown in area A, the closer the fringes are to one another in the interferogram, the stronger the filtering, adapted from (Ghulam et al., 2010)	34
Figure 4.5: Flowchart showing the generation of a CHM based on an RF MLM carried out in Google earth engine (GEE).....	38
Figure 4.6: Flowchart showing the process used for the comparison of the INSAR fusion CHM with the RF MLM CHM and their overall accuracy assessment using the LIDAR CHM ground truth data.	42
Figure 5.1: Visual representations of the intermediary outputs generated in ESA SNAP during the following steps carried out for the INSAR process: A) Back geocoding, B) interferogram generation, C) Tops Deburst, D) Multilooking, E) Goldstein phase filtering, F) Subset, G) Phase unwrapping, H) Phase to elevation, I) Radiometric terrain correction.	48
Figure 5.2: Boxplot showing the mean pixel coherence over the combined thinning areas	50
Figure 5.3: Boxplot showing the mean intensity of the Sentinel-1 C band signal over the combined thinning areas	51
Figure 5.4: Location of the GCP and PSM used for the INSAR adjustment, with a buffer of 14m representing the INSAR DSM resolution applied to an example of a PSM and GCP.....	52
Figure 5.5: Distribution of the GEDI L2A datapoints used as the dependent variable over the study area in an RF MLM	55
Figure 5.6: Importance of predictor variables used in the RF MLM interpolating across GEDI data points for the creation of a CHM.....	55
Figure 5.7: Comparison between the height predicted by the RF MLM with the 70% of the GEDI training data	56
Figure 5.8: Comparison between the height predicted by the RF MLM with the 30% of the GEDI validation data.....	57
Figure 5.9: Comparison of the full height ranges (100 th percentiles) of the INSAR fusion CHM (A), RF MLM CHM (B), and the LIDAR ground truth CHM (C).	58

Figure 5.10: Comparison of the full height ranges (100th height percentiles) of INSAR fusion CHM (A), RF MLM CHM (B), and LIDAR ground truth CHM (C) that have been extracted in different thinning areas within the study area..... 59

Figure 5.11: Visual comparison of the INSAR fusion CHM (Chosen INSAR CHM), RF MLM CHM (GEDI ML CHM) and the LIDAR ground truth CHM, with all CHMs resampled to 14m over T0 areas 61

Figure 5.12: Visual comparison of the INSAR fusion CHM (Chosen INSAR CHM), RF MLM CHM (GEDI ML CHM) and the LIDAR ground truth CHM, with all CHMs resampled to 14m over T1 areas 61

Figure 5.13: Visual comparison of the INSAR fusion CHM (Chosen INSAR CHM), RF MLM CHM (GEDI ML CHM) and the LIDAR ground truth CHM, with all CHMs resampled to 14m over T2 areas 62

Figure 5.14: Visual comparison of the INSAR fusion CHM (Chosen INSAR CHM), RF MLM CHM (GEDI ML CHM) and the LIDAR ground truth CHM, with all CHMs resampled to 14m over T3 areas 62

LIST OF TABLES

Table 3.1 Representative sample of the SAR systems currently operational adapted from Jacobsen (2005); Paek et al. (2020); Zhang et al. (2020).....	13
Table 3.2: Representative samples of the currently operational passive optical systems adapted from (Ouhami et al., 2021)	14
Table 3.3 Current active spaceborne LIDAR sensors adapted from Coops et al. (2021)	14
Table 4.1: Characteristics of the LIDAR data used as ground truth	23
Table 4.2: Classification classes used by RPS Australia to classify the LIDAR points.....	23
Table 4.3: Selected processing parameters to be used for this study when a DSM and DTM have to be generated.....	26
Table 4.4: Breakdown of the study area based on the number of thinning processes carried out..	29
Table 4.5: Individual characteristics of the SAR images selected as input for the INSAR process	31
Table 4.6:Permutations of the filtering process carried out in the INSAR DSM generation process in ESA SNAP.....	36
Table 4.7 Predictors used and their sources to be used as input in the RF MLM, adapted from Luo et al. (2023).....	41
Table 5.1: Characteristics of the <i>Pinus radiata</i> plantation in the study area falling under the different thinning category.....	44
Table 5.2: Determination of the most accurate INSAR filtering methodology for each of the thinning areas.....	49
Table 5.3: Characteristics of the PSM adjusted INSAR DSM with its accuracy when compared to the LIDAR ground truth DSM	53
Table 5.4: Characteristics of the GCP adjusted INSAR DSM with its accuracy when compared to the LIDAR ground truth DSM	54
Table 5.5: Assessing the accuracy of the 90th percentile height derived from INSAR fusion CHM and RF MLM CHM against the 90th percentile height of ground truth LIDAR CHM across different thinning classes. All CHMs are resampled to a 14m resolution	60

LIST OF EQUATIONS

Equation 1.....	11
Equation 2.....	12
Equation 3.....	12

CHAPTER 1 INTRODUCTION

Forest inventories are an effective tool for the management and monitoring of forest resources, and their outputs usually provide information about the health and disturbances in the studied forest area. Information derived from forest inventories can be as input data in a range of models or allometric equations to estimate ecological and economic indicators such biomass estimations; total woody biomass for example can be used to estimate the total timber that might be obtained after a future harvest (De Petris, Sarvia, et al., 2022a; Fankhauser et al., 2018)

Canopy and tree height can be directly measured in the field (Frolking et al., 2009). The most accurate way to measure tree height is by measuring the length of a felled tree, but this approach is usually not employed due to its destructive nature (Jurjević et al., 2020). In the field, tree height can be measured directly using telescopic poles or by climbing the tree (Jurjević et al., 2020). Indirect methods also exist, such as using rangefinders or clinometers (Jurjević et al., 2020). According to Liang et al. (2022), inventories using conventional measurement techniques are the most reliable source of reference data.

However, it is known that both direct and indirect field methods of estimating tree height are costly and require extensive field work that is risky, time-consuming, and labour-intensive (Brown, 2002; Coops et al., 2021; Du et al., 2014; Jurjević et al., 2020; Liang et al., 2022). Additionally, these field surveys are often limited to a small spatial extent, a limited number of observations, and low frequency due to financial constraints faced by organizations without sufficient funding (Fankhauser et al., 2018; Liang et al., 2022). Direct methods of measuring tree height such as telescopic poles are limited to trees within a certain height, and indirect methods require a clear line of sight from the base of the tree to the treetop. This condition may not be possible in a natural environment due to terrain restrictions and complex overlapping tree crowns causing occlusion effects (Brown, 2002; Fankhauser et al., 2018; Liang et al., 2022; Wang et al., 2019).

Liang et al. (2022) further highlights the possibility that the data collected from field inventories may not be reliable. Moreover, tree height measurements from the field may have lower accuracy, depending on the type of the tree being measured. This can introduce bias, uncertainty, or inaccuracy in allometric relationships used to estimate biomass and carbon stock (Liang et al., 2022; Wang et al., 2019).

This is where remote sensing methods can come into play to measure tree canopy metrics, such as tree heights, and replace the need for traditional forest inventories.

Remote sensing

Remote sensing technologies have allowed for the monitoring of forests on a much larger scale than field inventories would allow, ranging from local scales when utilising airborne sensors to a global view when using satellite platforms (Torres de Almeida et al., 2022). Moreover, remote sensing has been found to be effective in the forestry context for extracting structural information, such as tree height (Mail et al., 2023; Wu, 2020). This technology can be categorised into passive and active remote sensing.

Passive optical remote sensing

Passive optical remote sensing, such as multispectral imaging, is a system that relies on measuring the reflectance of solar radiation. It has been observed to indirectly provide information about vegetation structure (Wu, 2020). Studies have shown that short-wave infrared-based indices, computed using 30m resolution Landsat data, are significant predictors of canopy height in Spanish pine forests. The additional bands and higher resolution offered by Sentinel-2 data are expected to contribute to more accurate predictions (Coops et al., 2021).

However, at resolutions of 10m (Sentinel-2) and higher, it has been observed that the spectral signature of pixels is no longer sufficient on its own to predict canopy height. Instead, texture-based indices that characterise pixel properties, such as roughness and shadowing, become stronger predictors (Lang et al., 2019).

Active optical remote sensing

The emergence of light detection and ranging (LIDAR) as an alternative to field measurements for obtaining information about forest structural attributes has been observed (Ismail et al., 2023). LIDAR has overshadowed traditional methods as the optical laser beam actively emitted by the LIDAR sensor can penetrate the tree canopy to provide insights about the tree top, tree base, and terrain that were previously inaccessible. It is capable of accurately capturing the 3D positions of objects with centimetre-scale precision (Coops et al., 2021; Wang et al., 2019). Additionally, the LIDAR data processing workflow can be automated, enabling analysis of larger survey areas, which has made this technology popular for forest surveys (Wang et al., 2019).

Issues with Passive and Active Optical Remote Sensing

Remote sensing technologies, when passive, are unable to penetrate the inner sections of forest canopies, such as the mid and understory. As a result, the availability of spectral signatures and texture information for these regions remains poor (Li et al., 2022; Tienaho et al., 2022). Moreover, the accuracy of tree attributes extracted from LIDAR data varies greatly depending on the platform used. Unpiloted aerial or low-altitude airborne LIDAR systems provide high-detail and high-resolution data about forest attributes but generally cover a small area, while plane-based systems offer a broader overview of forest structure. Spaceborne systems like the Global Ecosystem

Dynamics Investigation (GEDI) provide a global understanding of Earth's forests but compromise on the resolution of the resulting products, which are often at a much larger extent (Coops et al., 2021).

Finally, spaceborne optical remote sensing, regardless of being passive or active, is affected by atmospheric dynamics, most commonly the presence of clouds or other aerosols, and tends to overestimate areas with denser vegetation coverage (Chang et al., 2023; Roc Roc, 2019).

Microwave Remote Sensing

Microwave-based remote sensing techniques such as Synthetic Aperture Radar (SAR) have also been associated with the ability to measure vegetation structure (Wu, 2020). More specifically, SAR metrics such as the vegetation optical depth (VOD), which measures the attenuation of the incoming microwave signal by the vegetation, have been found to be more effective than optical-based indices for both the leafy and woody components of a forest. SAR, due to its longer wavelength of energy actively emitted by the sensor compared to the optical region, is also less affected by cloud presence and other atmospheric effects. Thus, its data can be acquired independent of daylight and cloud cover (Chang et al., 2023; Rüetschi et al., 2019).

Combination of Remote Sensing Techniques

Combining SAR with other remote sensing techniques such as LIDAR or optical methods, taking into account their individual advantages and limitations in a range of machine learning methods (MLM), could enhance the potential of remote sensing techniques to provide more accurate and spatially continuous mapping of forest canopy height (Chang et al., 2023; Froking et al., 2009; Y. Zhang et al., 2022). Furthermore these MLM are based on the principles that these combined datasets can be used to interpolate between data points area collected by spaceborne LIDAR based on assumptions of homogeneity or gradual change (Torres de Almeida et al., 2022)

Study area

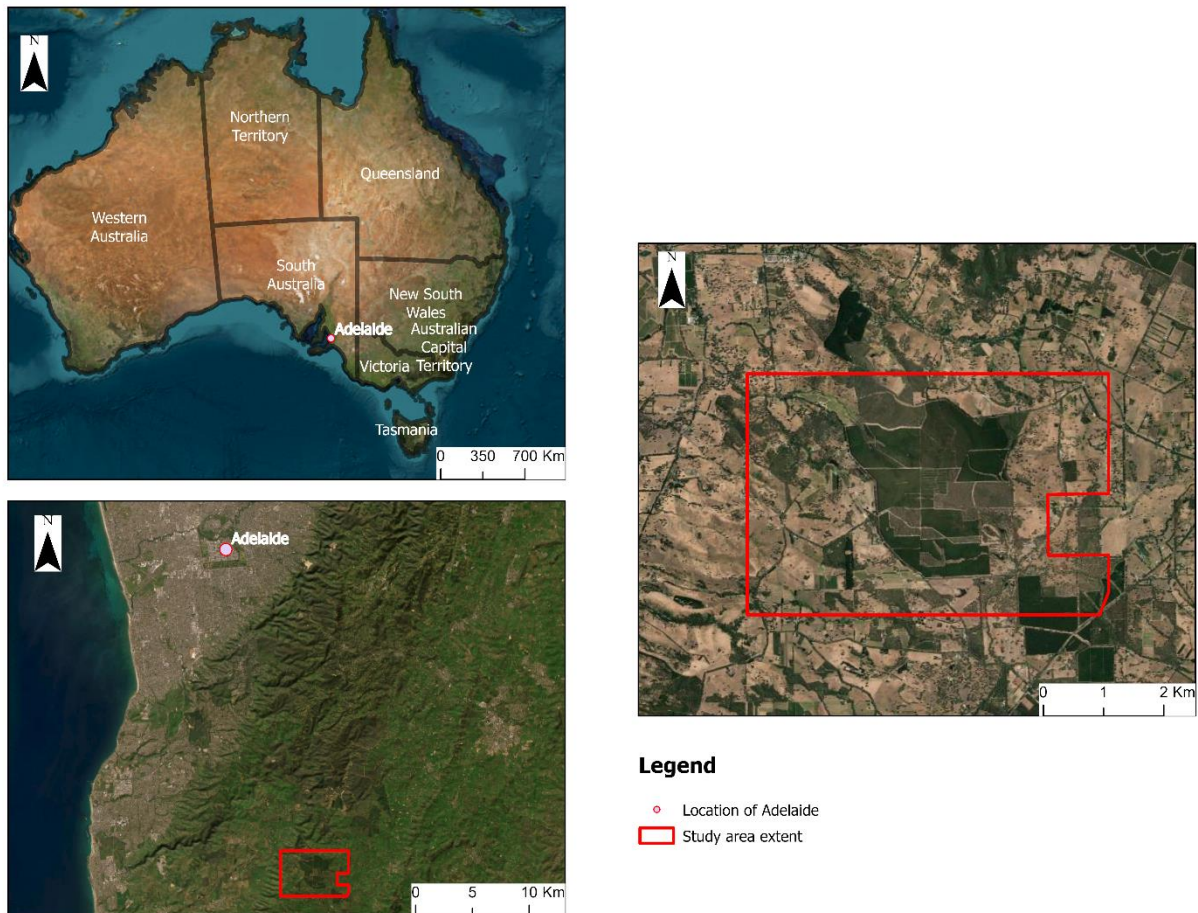


Figure 1.1: Location and extent of the study area investigated overlaid on ESRI Basemap.

The study area selected for this research, as shown in Figure 1.1, has an extent of 22.9 km². Within that area, 1.4 km² forms part of the native forest reserves (NFRs): Mount Panorama and Knott Hill, and 4.4 km² is part of an ongoing *Pinus radiata* plantation managed by ForestrySA. These reserves and plantations are part of the larger Kuitpo Forest Reserve, located in the southern section of the Mount Lofty Ranges, south of Adelaide, South Australia, Australia (ForestrySA, 2016)

A factor in the selection of the study site was the availability of freely accessible airborne LIDAR data in this section of the Kuitpo forest reserve. This data will be processed and used as ground truth in this study. Moreover, the site is relatively close to Flinders University, enabling timely access.

Since 1921, the land in these areas has been cleared to make room for pine plantations (*Pinus radiata*). Approximately only 15% of the native vegetation, consisting mainly of *Eucalyptus* species, has been preserved in dedicated sections of the reserve.

The study site experiences typical wet and cool winters, with average daily temperatures around 14°C. This transitions into dry and warm summers, with temperatures ranging from 30°C to 35°C, and an annual rainfall of 854mm (ForestrySA, 2016).

ForestrySA manages the pine plantations on a 40-year cycle. If the selected site is undergoing its first rotation of pine, the area is prepared to ensure proper water movement and prevent erosion. For second rotations, the residues from the previous cycle are minimised, and the site is cultivated to improve the underlying soil health (ForestrySA, 2018). Seedlings, approximately 25cm tall, are then planted at a density of around 1600 per hectare. Throughout their life span, the pine trees are thinned at a rate of 30% to 50% every 10 years and ultimately harvested when they reach 37 to 40 years of age. These trees are then used to provide sawlogs and industrial logs, serving as raw materials for various industries (ForestrySA, 2018, 2021). The study area was then broken down into different sections of *Pinus radiata* based on the number of thinning processes they have experienced so far, resulting on a further breakdown into areas such as T0, where no thinning was carried out, to T1, T2, and T3 where one, two and thinning cycles were carried out respectively.

Research aims

This study investigated the ability of SAR data from Sentinel-1 to be used on its own through INSAR and in combination with optical data from Sentinel-2 in a MLM. In both scenarios a canopy height model was generated and compared against LIDAR ground truth data. The objective was to determine which method of utilising SAR data was overall the more appropriate for large *Pinus radiata* plantations in South Australia which may contain trees that have different ages / maturities and thus have undergone a different number of thinning cycles.

CHAPTER 2 INTRODUCTION TO SAR

Synthetic Aperture Radar (SAR) is an active microwave remote sensing technique that can capture high-resolution 2D images of large study areas regardless of weather conditions, daylight intensity, and cloud coverage (Jagdhuber, 2012; Moreira et al., 2013). The synthetic nature of the SAR system arises from the coherent combination of received signals, which creates a virtual antenna that is much longer than a physical one (Moreira et al., 2013).

A space-borne SAR system consists of an active transmitter mounted on a satellite that periodically emits pulses to interact with the Earth's surface. The amplitude and phase of the reflected echoes are modified based on distance to the object as well as the physical or electrical properties of the objects present (Moreira et al., 2013; Santoro, 2003). The SAR system can be described as monostatic or bistatic, depending on whether the same antenna is used for both transmission and reception of the echoes or if different antennas are employed (Jagdhuber, 2012; Moreira et al., 2013). The SAR system is mounted in a side-looking configuration on the moving platform instead of the NADIR configuration to avoid signal saturation and positional ambiguities. In the nadir configuration, the SAR antenna would point straight down, resulting in a loss of image information due to saturation. Therefore, the side-looking configuration enables SAR to provide high-resolution images over a large area (Moreira et al., 2013; Santoro, 2003).

A typical SAR configuration is depicted in Figure 2.1

Figure 2.1 has been removed due to Copyright restrictions

Figure 2.1: Typical configuration of a SAR sensor adapted from NISAR (2018)

It is observed that the SAR system moves in an azimuth direction also described as the along track direction, with the signal pulses being transmitted in the direction of the slant range R , under an

incidence angle θ which, under the condition of flat terrain being observed, will be equal to θ_r , the RADAR look angle. Moreover, the radar swath is generated as the SAR platform moves in the azimuthal direction, and the SAR footprint, a function of the antenna design, is visible in the illuminated area.

SAR resolution

Azimuth resolution

In a SAR system, the beam width in the azimuth direction will define its capacity to separate between two objects A and B as shown in Figure 2.1 which are present on the same range, and is called the Azimuth resolution (Woodhouse, 2006).

Range resolution

The slant range resolution refers to the SAR system's ability to distinguish between two objects, in this case, A and C, that are found within the same position in the azimuth direction but along different sections in the range direction. On the other hand, the ground range resolution (GRR), often called the range resolution, is crucial for Earth observation. The slant range, based on the distance from the line of sight from the sensor to a particular point, is a geometric measurement that does not convey true resolution or ground position on its own. It is generally converted to ground range resolution through a projection onto a reference surface. It is observed that the ground range resolution is poorer in the near range and improves towards the far range.

Generally, the spatial resolution of a SAR system is defined in the Azimuth and Range direction (Az-R)

SAR band Wavelength

The backscatter received by the SAR sensor depends on three main parameters: surface roughness, dielectric constant, and viewing / object geometry (Jones & Vaughan, 2010). The wavelength of the microwave signal used by the SAR sensor is also known to affect the signal's penetration ability. Different microwave bands, consisting of a range of wavelengths, have been assigned specific names, as shown in Figure 2.2 (Sartoretto, 2021).

Figure 2.2 has been removed due to Copyright restrictions

Figure 2.2: Depiction of the different wavelength of the microwave signals used in SAR systems, and the corresponding naming of the section of the wavelengths into bands (Sartoretto, 2021)

SAR band penetration

A SAR system's ability to penetrate the canopy increases as the wavelength increases (Figure 2.3). Higher wavelengths, such as L and P bands, are suitable for assessing biomass dynamics, while lower wavelengths X and C bands are more appropriate for estimating canopy height (Chang et al., 2023; Coops et al., 2021). This distinction arises from the scattering phase centres (SPC), which are mathematical constructs used to describe the origin of the echo of an incoming SAR signal within a scattering area. X and C-band based SAR systems will exhibit SPC on close to the top of the canopy and within the upper tree crown, respectively (Chang et al., 2023; Coops et al., 2021). In contrast, L- and P-band systems will have significantly lower SPC within the canopy, as they penetrate it. Consequently, the backscatter recorded by these systems would primarily originate from either the tree trunks or the ground itself (Palm, 2022).

Figure 2.3 has been removed due to Copyright restrictions

Figure 2.3: Canopy penetration of the different SAR bands (Wallington, 2006).

SAR data characteristics

Object Dielectric Constant and Roughness

The penetration capability of the SAR system also depends on the reflection object dielectric constant, which determines the ratio of the transmitted microwave signal that will penetrate or be absorbed by the targeted medium and how much of the incoming signal will scatter at the surface (Meyer, 2019; Sartoretto, 2021). This constant depends on two characteristics: the density of the canopy and the object moisture content. The higher they are, the lower the penetration depth of the signal (Meyer, 2019). The roughness of the observed surface is a function of the incoming wavelength of the SAR signal and determines the amount of backscatter signal received by the SAR sensor. The longer the wavelength, the smoother the surface appears to the SAR sensor, resulting in specular reflection and limited backscatter return. In contrast, the shorter the wavelength, the rougher the surfaces appear, resulting in diffuse and stronger returns (Meyer, 2019).

SAR scattering mechanisms

The interaction and scattering of the incoming SAR data vary depending on the surface characteristics, as seen in Figure 2.4. Rough surface scattering occurs in areas composed of bare ground and low vegetation, double-bounce scattering occurs between the ground and tree trunks as well as in building corners, and finally, volume scattering occurs when the signal is reflected several times as it penetrates vegetation canopies (Meyer, 2019)

Figure 2.4 has been removed due to Copyright restrictions

Figure 2.4: Types of scattering in SAR data, where the blue arrows correspond to rough surface scattering, the green arrows correspond to volume scattering, and the red arrows correspond to double bounce scattering (Meyer, 2019).

SAR polarisation

The polarisation of the transmitted SAR signal measures its plane of oscillations, and different polarisations respond differently to the different types of scatterers present in a scene. The polarisations used in SAR systems range from single polarisations, such as horizontal (H) polarised wave transmitted and horizontal polarised wave received (HH), or vertical (V) polarised wave transmitted and vertical polarised wave received (VV), to dual polarisations where an H or V wave is transmitted and both H and V polarised waves can be received, HH and HV as well as VV and VH (Sartoretto, 2021). Finally, the final polarisation type is quad polarisations where both transmitted and received signals are orthogonal pairs (Sartoretto, 2021). The received or changed polarisation of the SAR signal can be used to infer the properties of objects observed in the scene (Sartoretto, 2021).

Interferometric SAR (INSAR)

INSAR is a powerful tool capable of measuring changes in various geophysical parameters, such as topography and surface alterations (Rosen et al., 2000; Yu et al., 2019). Regular SAR, which projects detected 3D objects onto resulting 2D SAR images based on their along-track position and slant range measured by only one antenna, cannot be used to determine topographic heights (Li &

Goldstein, 1990). However, INSAR adds another antenna to the system. Each antenna individually measures the phase of a specific pixel of interest. Due to their inherent differences in imaging geometry, the measured phase difference between both antennas, designed to be close together with a small baseline (Figure 2.5), will be influenced by the height of any object present in that pixel of interest (Li & Goldstein, 1990). This phase difference is related to the wavelength used in the SAR system, which makes INSAR measurements sensitive and precise for topographic information (Li & Goldstein, 1990).

Figure 2.5 has been removed due to Copyright restrictions

Figure 2.5 Showing a simplified depiction of INSAR in across-track systems, where two SAR scenes, SAR, and SAR', are used to determine the topographic height (h) at a scattering phase centre based on a reference height (H) and a look angle (θ)

Types of INSAR systems

Along track INSAR

One way to acquire INSAR pairs is through observations that are simultaneously acquired over the same target site, either through an along-track or cross-track mode, which relies on a system using a pair of antennae or, one antenna can be artificially split into two halves that can be combined for transmission but separated for reception (Bamler et al., 2006; S. Liu et al., 2021).

Repeat pass INSAR

Repeat pass or repeat track INSAR is another method that can be used with observations of the same site but from two different tracks acquired at different dates, through a process called repeat pass or repeat track INSAR (Bellec et al., 2005; S. Liu et al., 2021).

Additionally, SAR systems like TanDEM-X can operate in both acquisition modes since the system comprises two radar satellites flying in a configuration that can be modified depending on the need (Krieger et al., 2007).

INSAR principles

INSAR is mostly based on the concept of coherence which quantifies the similarity between the signals obtained from the INSAR antennae and thus is a measure of the correlation between two SAR images (Li & Bethel, 2008; Roc Roc, 2019; T. Wang et al., 2009). In areas with water bodies, low coherence is observed, while urban areas with buildings exhibit high coherence. It is worth noting that coherence can change over time, with agricultural fields experiencing a decrease in coherence from moderate to low due to plant growth. Coherence is a quantity with values between 0 and 1 where the lower the value the lower is the similarity between the SAR images (Roc Roc, 2019; T. Wang et al., 2009).

The correlation γ between two SAR scenes can be affected by several factors shown in equation 1 shown in Roc Roc (2019)

$$\gamma = \gamma_{geom} \cdot \gamma_{SNR} \cdot \gamma_{vol} \cdot \gamma_{temp}$$

Equation 1

Where γ_{geom} refers to the geometric correlation occurring due to the incidence angle, γ_{SNR} corresponds to the noise observed in the receive antennas of the SAR system, γ_{vol} is a measure of the volume scattering occurring and γ_{temp} is the temporal decorrelation occurring due to the difference in the SAR acquisition time (Roc Roc, 2019; Woodhouse, 2006)

The phase difference ($\Delta\phi$) for a SAR image is made up of three components as shown in equation 2, which was simplified from Roc Roc (2019). When equation 2 is expanded, the first component is the change in phase in the slant range direction from the SAR sensor and the scattering phase centre generated when the signal meets the ground (ΔR). The second component is the difference in elevation between the scattering phase centre and the reference height, which may be a reference ellipsoid, geoid, digital terrain model (DTM), or a flat earth. In the context of INSAR over

a tree canopy, this corresponds to the tree height (Δh). Finally, the third component is a 2π phase ambiguity, which is present as the phase value is initially bounded between $-\pi$ and π (Chen et al., 2000; Roc Roc, 2019; Soja, 2014)

$$\Delta\phi \approx K_z(\Delta R \cos\theta + \Delta h) + 2\pi n$$

Equation 2

Where k_z is the vertical wave number, which is a measure of the sensitivity of the phase of the SAR signal to variations in height. It represents the number of 2π cycles in the SAR signal that correspond to a vertical change in height of 1 metre, and n is a measure of the complete wavelengths in the phase difference (Asopa & Kumar, 2020; Soja, 2014).

Looking at equation 2, the first component on the RHS is first completely removed through a flattening process (Roc Roc, 2019). The third component ($2\pi n$) is then removed through a second process called the phase unwrapping which is a process that allows the number of 2π to now be unbounded from the initial $-\pi$ and π range as shown in figure 2.6

Figure 2.6 has been removed due to Copyright restrictions

Figure 2.6: Overview of the unwrapping process where the wrapped phase ϕ_w is unwrapped to phase ϕ , adapted from Chen et al. (2000)

Finally the only term left in equation 2 Δh corresponding to the tree height can be determined as shown in equation 3 adapted from Roc Roc (2019) where $\Delta\phi'$ is the phase difference after the unwrapping process

$$\Delta h = \frac{1}{k_z} \Delta\phi'$$

Equation 3

CHAPTER 3 LITERATURE REVIEW:

To thoroughly investigate the application of SAR data in estimating tree canopy height, a comprehensive assessment of presently operational SAR systems was conducted. This assessment involved identifying the specific SAR bands currently in use and operational. Subsequently, the literature review has been selectively narrowed to include only relevant and current studies aligned with the identified operational SAR bands. This targeted approach ensures a focused exploration of literature directly pertinent to the subject matter.

Operational SAR systems

A wide range of SAR systems are currently operational with some of their characteristics described in Table 3.1

Table 3.1 Representative sample of the SAR systems currently operational adapted from Jacobsen (2005); Paek et al. (2020); Zhang et al. (2020)

Satellite	Launch Year	Spatial resolution (m, Az-R)	Temporal Resolution (days)	Band Type	Pricing
TerraSAR-X	2007	1-1	11	X	Paid
TanDEM-X	2010	1-1	11	X	Paid
COSMO-SkyMed	2019	0.35-0.55	1-3	X	Paid
RADARSAT-2	2007	0.8-1.6	24	C	Free
ALOS-2 (Daichi-2)	2014	1-3	14	L	Paid
PAZ	2018	1-1	1	X	Paid
Sentinel-1A/B	2014,2016	4-2	6–12	C	Free

As shown in Table 3.1, it is seen that a wide range of SAR systems are currently operational. However, Only RADARSAT-2 and Sentinel-1 have their data currently freely available. Furthermore, as RADARSAT-2 data is only freely available as a yearly mosaic and not individual images, outputs from Sentinel-1 was chosen for this study as the data available is separated for each observation and not a mosaic, and thus can be used both for INSAR as well as MLMs

Operational optical passive remote sensing systems

Moreover, an investigation was carried out to identify the current passive and active optical remote sensing systems that are deployed to assess their potential to be included in MLM. (Table 3.2)

Table 3.2: Representative samples of the currently operational passive optical systems adapted from (Ouhami et al., 2021)

Satellite	Launch Year	Spatial Resolution (m)	Temporal Resolution (days)	Pricing
WorldView-2	2009	0.46 m	1.1	Paid
WorldView-4	2019	1.24 m (VNIR)	4.5	Paid
WorldView-3	2014	1.24 m (VNIR) 3.7 m (SWIR)	<1 <1	Paid
Pleiades-1A	2011	2.0 m (VNIR)	1	Paid
Jilin-1	2015	2.88 m	3.3	Paid
SPOT-6,7	2012,2014	6 m (VNIR)	1	Paid
TH-01	2010	10 m (VNIR)	5	Paid
Sentinel-2	2015	10 m, 20 m, 30 m	5	Free
Landsat 8/9	2021	30 m (VNIR + SWIR)	8	Free
ENMAP	2022	30 m	4	Paid
PRISMA	2019	30m	7	Paid

As shown in Table 3.2, it was observed that a wide variety of high spatial resolution optical spaceborne systems with high revisit times are currently deployed. However, it was noted that only Sentinel-2 and Landsat data were available for free. When comparing Sentinel-2 data to Landsat 8-9, it was found that Sentinel-2 had a higher temporal resolution, resulting in more images being available over the study period and increasing the likelihood of cloud-free observations. Additionally, four of Sentinel-2's spectral bands had a resolution of 10m, whereas Landsat consistently had a 30m resolution across all bands, except panchromatic. Therefore, if optical data were to be used in an MLM, the data source chosen was Sentinel-2.

Operational active optical remote sensing systems

Table 3.3 Current active spaceborne LIDAR sensors adapted from Coops et al. (2021)

LIDAR Sensor	Recording mode	Acquisition strategy	Wavelength	Approximate footprint size	Spatial extent
ATLAS (ICESat-2 satellite)	discrete returns	Profiling (1 footprint every 70 cm along laser ground track)	532 nm	~14 m	Global ($\pm 88^\circ$ N/S latitude)
GEDI (ISS)	Full waveform	Sampling (1 footprint every 60 m along ground track)	1064 nm	~25 m	Near-Global ($\pm 51.6^\circ$ N/S latitude)

It is observed from Table 3.3 that there are currently two active spaceborne LIDAR systems. Upon examining the approximate footprint size, it may be tempting to utilise ICESat-2 data. However, as demonstrated by A. Liu et al. (2021), it can be seen that ICESat-2 and GEDI employ different data sampling strategies. ICESat-2 employs a sampling strategy that prioritises higher boreal regions, resulting in reduced sampling density over temperate and tropical forests. Conversely, GEDI has a higher sampling rate in mid to low latitude regions, which aligns more closely with the location of the study site. Thus GEDI was utilised as the dependent variable in an MLM for this study due to its superior ability to retrieve canopy height compared to ICESat-2, as the latter tends to underestimate canopy heights under all forest types and conditions (A. Liu et al., 2021).

Literature review scope

The scope of this review has been narrowed down to primarily investigate the use of Sentinel-1 C-band data, initially through the INSAR process only, and then through a combination with other data sources, including Sentinel-2 in MLM. This review also aims to find the most appropriate MLM to be used in this context, as well as how the different data sources should be combined and processed before being used as input. Therefore, studies utilising the less accurate ICESAT-2 are not excluded, as they could still provide valuable insights into the optimal way of utilising input data such as Sentinel-1 and Sentinel-2 data.

Moreover, due to wide extent of X and L band SAR systems being deployed, an overview of their methodology in studies which aimed to estimate tree canopy height was also carried out to see potentially how this study may evolve in the future if these datasets were to be available.

Sentinel-1 INSAR for tree canopy height estimations

The use of sentinel-1 C-band data directly through INSAR was in the to be often dismissed category due to its intermediary canopy penetration ability which may not accurately describe the tree canopy height, and the low temporal resolution (Rüetschi et al., 2017). However, Kumar and Krishna (2019) conducted a study using a pair of VV Sentinel-1 datasets to generate a CHM. They achieved this by creating an interferogram and unwrapping the phase to obtain elevation information. The authors then applied radiometric corrections to generate a DSM. However, for the DTM from which the DSM needs to be subtracted to obtain a CHM, instead of using a DTM specific to the forest location, the authors searched for nearby areas that had been cleared of forest. They obtained the ground elevation from those areas and assumed that the elevation in those areas was the same as in the forest, and thus generated their CHM

Liu et al. (2022) used an alternative method to generate the CHM using INSAR. Instead of using an external digital elevation model (DEM) to obtain the ground elevations, they utilised the

assumption that the DSM generated from the VH polarised S1 INSAR pair corresponds to the scattering phase centre which is closer to the top of the tree canopy, and the DSM from the VV polarisation corresponds to the scattering phase centre which is closer to the ground, under the assumption that the VV scattering is due to the ground scattering, thus the phase equivalent of the DTM. Thus, the height of the vegetation was calculated by the canopy phase due to the vegetation (VH) minus the canopy phase due to the ground (VV)

Xie et al. (2023), on the other hand, used the INSAR process to generate a DSM using Sentinel-1. However, instead of using an external DTM or generating the phase information equivalent to the ground as in Liu et al. (2022), for the DTM generation, they used the INSAR PALSAR L band data, relying on the assumption that the L band has a greater penetration ability than the C band, and thus the DSM generated from the process will resemble a DTM more closely. Moreover, before creating a CHM using the DSM minus DTM formula, they assumed that there might be a systematic difference between the C band DSM and L band DTM. Thus, both DEMs were analysed over the same flat area in the study site, and adjusted by finding the average elevation over the flat polygon and removing it from both DEMs before generating the final CHM.

Kumar and Krishna (2019), Xie et al. (2023) and Liu et al. (2022) utilised the principle of phase unwrapping to generate their respective DEMs at some point. However, De Petris, Cuzzo, et al. (2022) maintained that phase unwrapping tends to fail in vegetated areas due to low coherence caused by vegetation. Instead, they employed an alternative approach that does not involve phase unwrapping. In De Petris, Cuzzo, et al. (2022), the INSAR process was halted at the interferogram generation stage and, as a result, only the phase component caused by topographic changes was eliminated.

To establish reference points, they selected the centroids of polygons within the study areas. These regions were both flat and highly coherent (coherence greater than 0.8). From these reference points, the phase information from the interferogram and the elevation from an external DTM were extracted. The absolute change in height was then calculated as the difference between the phase at any forest point and the phase at the reference point, multiplied by the sensitivity ratio of the phase to topographic changes. The canopy height was obtained by adding the absolute change in height to the height extracted from the DTM's nearest reference point.

Overall, it was observed that even within the traditional INSAR process using C-band data, where the Digital Surface Model (DSM) is generated after phase unwrapping, there are variations in how ground elevation is derived to generate a CHM. These variations include using an external DTM, utilising different INSAR data with greater capability to penetrate canopies, or employing different polarisations of the same data, all of which can better represent ground elevation. However, it was

shown that there exist a modified INSAR methodology using sentinel-1 data that circumvents the phase unwrapping process that often fails in vegetated areas, but is still able to generate a

Sentinel-1 as a data source in MLM for tree canopy height estimations

It has been observed that a wider range of studies have preferred utilising sentinel-1 data with other data sources into MLM to retrieve canopy heights directly through the INSAR process (Ghosh et al., 2020; Liu et al., 2019; Silveira et al., 2023; Sothe et al., 2022; Xi et al., 2022). This trend may be following the fact that recent research have outlined that the combination of SAR and optical data can accurately be used to derive forest canopy metrics such as tree canopy height. Moreover the use of MLM such as Random forest (RF), Classification and Regression Trees (CART), and Logistic regression (LR) has allowed the development of predictive models that can be built upon an existing reference database such as spaceborne LIDAR, and has been proven to be a viable alternative to field-collected data (Torres de Almeida et al., 2022).

Nandy et al. (2021) have used Sentinel-1 co and cross polarised data as well as a range of texture information (GLCM) derived from it as the independent variable and ICESAT-2 height canopy values as independent variable using an RF model to estimate the forest canopy height outside of the range covered by the ICESAT-2 data used. They found that when the output canopy heights were in line with the field data in the deciduous forest with an R^2 of 0.84.

Silveira et al. (2023) used a range of texture metrics derived from Sentinel-1 SAR data and vegetation indices from Sentinel-2 as inputs in a random forest algorithm to predict canopy height based on field data collected over a large scale. They found that their model predictions outperformed tree canopy height obtained by GEDI. Torres de Almeida et al. (2022) derived a range of Sentinel-1 features, such as the VV and VH polarizations and a series of derived SAR indices, as well as Sentinel-2 data in the form of raw reflectance bands and derived vegetation indices. These datasets were inputted separately, then combined in a series of machine learning algorithms such as LR, CART, and RF to estimate tree canopy height based on LIDAR data. They found that the RF algorithm with only Sentinel-2 outperformed the other combinations, even though the one based on both Sentinel-1 and Sentinel-2 performed well, the difference was not statistically significant.

However, Sothe et al. (2022) tested the use of spaceborne LIDAR data from ICESAT-2 and GEDI as an alternative to airborne LIDAR as used in Torres de Almeida et al. (2022) in the training and validation data in an RF model using both Sentinel-1 and Sentinel-2 data as predictors. The authors found that GEDI was closer to ALS measurements when compared to ICESAT2, but they both overestimated the canopy height. Furthermore, their most performant RF model in predicting forest canopy height was the one including both SAR and optical data. Overall, Sothe et al. (2022)

found that using spaceborne LIDAR data allowed the model to be used for more large-scale mapping applications.

Ghosh et al. (2020) used a range of Sentinel-1 and 2 data, as well as the SRTM DEM, to investigate the symbolic regression (SR) approach in the context of a mangrove forest and compare it to the RF algorithm. The SR approach can identify all the different regression models and come up with one that is interpretable and based on a selection of variables that are the most effective predictors, using ICESAT2 data for training and validation. However, it must be noted that using validation data as carried out by (Xi et al., 2022) and (Ghosh et al., 2020) that are of the same order of accuracy as the results data might thus not be adequate mechanism. The authors found the SR model to be more performant, reliable, and accurate as compared to the RF one when used to create a canopy height map. If the objective is to estimate the forest stand mean height.

Several of the studies that were investigated have shown that for Sentinel-1 data, cross-polarized data produces more accurate predictions for forest structure attributes (Liu et al., 2019; Silveira et al., 2023; Sothe et al., 2022). Sothe et al. (2022) suggested that co-polarized backscatter, such as HH, may be affected by double-bounce scattering whereas cross-polarized data indicated volume scattering from tree leaves and branches in the vegetation, which contributed more to the canopy height predictions. However, Ghosh et al. (2020) was the only study to reject the use of the VH polarization from Sentinel-1 due to the lower coherence caused by cross-polarized noise.

X band SAR data for canopy height estimation

It was seen that one of the most common ways of utilising X band SAR data for canopy height predictions was its inclusions in models such as the Random Volume Over Ground (RVoG) model, which has become a standard application in the forestry context (Gómez et al., 2021). The RVoG model allows for the relation of complex interferometric coherence through a function of only a few parameters such as the ground-to-volume backscatter ratio, vegetation height, extinction, and topography of the study area (Gómez et al., 2021; Guliaev et al., 2021).

Guliaev et al. (2021) and (Qi et al., 2019) have suggested a technique to estimate forest height in a tropical context with single-polarization TANDEM-X INSAR data using a constant value for the extinction coefficient that is relates the forest density and dielectric constant for the forest in the study site. They assume a ground-to-volume ratio of zero in a simplified RVoG model, resulting in a scenario where the tree height is related only to the interferometric coherence. DTM can be used to calculate the ground topographic phase, which can then be used to derive the forest height and extinction coefficient using an area of high coherence as reference, assuming a ground-to-volume amplitude ratio of 0, through the use of an RVoG model (Guliaev et al., 2021; Qi et al., 2019). Furthermore, to augment the coverage of the TANDEM-derived height estimate, the DTM can be

derived from GEDI data instead of Airborne Laser Scanning (ALS), as hypothesized by Qi et al. (2019).

Gómez et al. (2021) have modified the RVoG model to include two empirical parameters that can be empirically calibrated to match the data and provide more reliable forest height estimations within a condition that the estimated height is within a third of the height of ambiguity (HOA). Schlund et al. (2022) adapted the model and principle of parameterization mentioned by Gómez et al. (2021) by combining GEDI RH100 data following the model described in Qi et al. (2019) to retrieve one of the empirical parameters needed in their version of the RVoG model. This adaptation is aimed at tropical landscapes rather than the higher-latitude forest where the model has previously been applied (Gómez et al., 2021).

L band SAR data for canopy height estimation

Moreover similarly to X-band SAR data, L-band data use was favoured to be used in models instead of INSAR and MLM, potentially due to data not being suitable for INSAR and overall the amount of data needed in MLMs potentially due to their elevated cost (Meyer, 2019)

It is seen that there were attempts to modify the RVoG model has been proven as an effective model in predicting forest height or vertical vegetation structures through the use of Polisar data (Sa et al., 2023). However, it is seen in a repeat pass context, that an unmodified RVoG model may suffer from temporal decorrelation which is associated with change in the general environmental condition such as the moisture present, tree regrowth or even freezing, movement of scatterers; all of which might occur in between the next pass (Lei & Siqueira, 2014; Sa et al., 2023). Moreover, the presence of temporal decorrelation results will affect the RVoG model as it reduces the coherence in the interferometric data as well as causing phase shifts (Sa et al., 2023). When considering spaceborne systems like ALOS, it is seen that the changes in coherences due to the temporal correlation often surpasses that caused by the change in the objects to be investigated even when the vertical wave number is low as it is often below the inversion range threshold (Sa et al., 2023)

Lei and Siqueira (2014) have established specific values for the extinction coefficient in relation to L-band characteristics, including the presence of either dense or sparse forests, and the vertical wavenumber for a temporal baseline in an ALOS system. Additionally, coefficients related to temporal changes, such as dielectric fluctuation and motion standard deviation, were assigned constant values for various height values. These coefficients and constants were utilised in an RVoG model to estimate curve fitting parameters, which were then used to fit a modified version of the RVoG model to estimate forest height. To verify these parameters in areas where the forest height is known, LIDAR data was used, as well as the assumption of a low ground to volume ratio.

When compared to the reference LIDAR data, the output of the modified RVoG model had an RMSE of less than 4m (Lei & Siqueira, 2014).

Chen et al. (2021) developed a modified RVoG model to estimate tree canopy height using L band POLINSAR data from spaceborne systems like ALOS PALSAR. In their study, they accounted for the temporal decorrelation by adapting the original RVoG model. They assumed a constant ground vertical reflectivity (GVR) and derived an extinction coefficient from a lookup table based on L band data. The authors calculated and combined three decorrelation factors, including the signal-to-noise ratio (SNR) ratio, pre-processing decoherence due to the registration error, and baseline distance decoherence. To consider the time difference between the acquisition date, Chen et al. (2021) used the Normalized Difference Vegetation Index (NDVI) data from the same date as the SAR data acquisition to estimate a final time variation factor. They combined these four factors to divide the coefficient data from the original SAR data. Additionally, they included range slope information calculated from the SRTM DEM since it significantly affects canopy height estimation (Chen et al., 2021). Overall, the model that incorporated both the range slope data and the time decoherence factors outperformed the model that considered only the time decoherence factors (Chen et al. (2021)).

Managhebi et al. (2018) developed a new algorithm to address temporal decorrelation in estimating tree canopy height using L band polarimetric interferometric synthetic aperture radar (PolInSAR) data. They adapted the RVoG model and combined it with a volume over ground algorithm (RVoG + VTD) to improve the accuracy of the estimation. Two decorrelation coefficients, based on ground and volume coherences, were integrated into the original RVoG model and were in a range between zero and one, with the ground coherence coefficient being given a constant value of one as it was assumed that GVR was constant.

Instead of using a fixed extinction coefficient, Managhebi et al. (2018) estimated it based on a linear function that relied on the distance ratio index. This index is the ratio between the average distance of the scatterers with the canopy and the distance between the scatterers and the ground. They used L band data and estimated the parameters using a least squares method. Before the inversion step, the estimated extinction coefficient was applied to calculate the forest height. The output model had an RMSE of 2.47m, which was 5.47m lower than the traditional RVoG model that used L band data.

Sa et al. (2023) have made improvements to the original RVoG model, enabling it to be utilised for Repeat pass L-band spaceborne systems, such as ALOS PALSAR. In contrast to previous studies that assumed constant extinction coefficients and ground volume scattering ratios (GVRs) (Chen et al., 2021; Lei & Siqueira, 2014) or Just a constant GVR (Managhebi et al., 2018), Sa et al. (2023) included three terms in the model to correct for the effects of temporal decorrelation and to estimate the extinction coefficient that varies with height. The authors used phase centre histogram

data created through INSAR to develop this function. Furthermore, a customized GVR parameter was created that considered both the volume scattering caused by the ground layer and the volume later scattering. The use of a customized GVR prevented underestimation of the output forest height by accounting for the double bounce effect. Although LIDAR data was not available for the investigated site, the output structure was found to comply with the general scatterer distribution law (Sa et al., 2023).

Literature review outcomes and gaps

Summary of X and L band usage for tree height estimation

It was observed that the utilisation of X and L band data differed significantly from C-band Sentinel-1 data. In these bands, models such as the RVoG were preferred. Furthermore, it was noted that these input models were more intricate, with modifications and assumptions needed depending on the SAR band used. For instance, in the case of X-band, which has low canopy penetration but not zero canopy penetration, the ground to volume ratio, which indicates the depth of SAR signal penetration, was assumed to be 0. Additionally, it was observed that the models using L band data were greatly affected by decorrelation over vegetation.

Summary of Sentinel-1 INSAR usage for tree height estimation

It has been found that the use of Sentinel-1 C-band data has traditionally not been employed in the INSAR process for vegetated areas due to its low temporal resolution and limited canopy penetration capability. However, it has been observed that among the studies investigated, only the general methodology of the INSAR process was applied even in those low-coherence forest sites with variations in how the elevation of the ground under the forests were determined. Only recently did De Petris, Cuzzo, et al. (2022) suggested modifying the process to accommodate the study site in low-coherence forests.

Gaps identified in literature for Sentinel-1 INSAR usage for tree height estimation

It was observed that the INSAR studies assumed a study site where no previous remotely sensed forest inventories, such as LIDAR surveys, had been conducted. As a result, these studies derived their ground truth from either lower resolution global datasets or assumptions where the ground in unforested regions resembles forested areas. However, it has not been explored how the INSAR process could be advantageous in areas where pre-existing LIDAR-derived DTM data is available. Assuming no substantial changes in ground conditions in the forested areas, it is worth investigating how the traditional INSAR CHM generation workflow can be modified to make use of high-resolution LIDAR in both the correction of the INSAR DSM over flat areas carried out in Xie et al. (2023), and also the final CHM generation using the LIDAR DTM.

Summary of MLMs using Sentinel-1 for tree height estimation

it was observed that the most effective MLM for estimating tree canopy height in areas similar to the study site, using a combination of Sentinel-1 and other data sources, was the Random Forest (RF) model. In general, it was confirmed that MLM models performed best when GEDI was used as the dependent variable and when both Sentinel-1 and Sentinel-2 data were included.

Gaps identified in the MLMs using Sentinel-1 for tree height estimation

However, it was observed that even though there was agreement regarding the use of the RF model as the MLM and the combination of Sentinel-1 with other data sources available for this study, the way the data was processed before the MLM model varied significantly. Firstly, there was no consensus on the SAR polarisation to be used as input. Additionally, there was disagreement on how the input SAR and optical data should be incorporated; whether as the raw spectral bands or utilised to derive texture metrics and vegetation indices in the case of Sentinel-2 optical data.

Research questions

The initial research questions that were used in this study were:

- 1) Can the INSAR process be used using Sentinel-1 SAR data to generate a DSM from which Canopy Height Model (CHM) over a *Pinus radiata* plantation in Mount Panorama and Knott Hill can be generated by subtracting a DTM generated from the LIDAR data from it. This CHM would be a representation of a fusion approach of using both space based and terrestrial remote sensing ?
- 2) Can Sentinel-1 SAR data be used in combination with other data sources, such as Sentinel-2 optical data, in a Machine Learning Model (MLM) to generate a CHM based on entirely space based remote sensing data over a *Pinus radiata* plantation in Mount Panorama and Knott Hill?
- 3) Which one of the CHMs generated through the fusion INSAR approach and the entirely space based MLM is the most accurate when compared to a ground truth CHM generated using only terrestrial remote sensing over *the Pinus radiata* plantation in Mount Panorama and Knott Hill?

CHAPTER 4 METHODS:

Processing of LIDAR ground truth data

The ground truth data for this study was selected to be LIDAR data provided through the Elvis Elevation and Depth platform (<https://elevation.fsd.org.au/>) where data was available from a LIDAR survey conducted in 2022 over the area of interest, captured by RPS Australia East Pty Ltd on behalf of the Department for Environment and Water. The characteristics of the dataset are shown below in Table 4.1.

Table 4.1: Characteristics of the LIDAR data used as ground truth

Acquisition Date	21/01/2022 to 31/01/2022
Dataset number	2667
Average point spacing(pts/m ²)	15
Vertical Positional Accuracy(cm)	+/-4.2
Vertical datum	AHD (Ausgeoid2020)
Projection/ coordinate system	UTM MGA Zone 54

This dataset comprises full waveform LIDAR data collected in the LAS 1.4 format by a Riegl VQ-1560 II-S sensor over South Australia, specifically the Adelaide metropolitan area. The 23 1km by 1km tiles of data obtained were combined in CloudCompare to create one LAS Dataset.

Investigation of data

The LIDAR data over the area of interest had already been classified by RPS Australia East Pty Ltd, as shown in Table 4.2. The minimum elevation recorded in the ground classification code was 238.9m, and the maximum elevation of any point was classified as noise at 477.5m. The average point spacing in the dataset over the AOI was observed to be 0.21m.

Table 4.2: Classification classes used by RPS Australia to classify the LIDAR points

Classification Class	Point per class	Percent (%)
Unclassified	6515	0.0
Ground	20021381	11.0
Low Vegetation	39676384	21.9
Medium Vegetation	6047142	3.3
High Vegetation	52084805	28.7
Building	120800	0.1
Low Point(noise)	16890	0.0
Model_Key_Point	140281	0.0
Water	111764	0.0
Overlap Points	63322152	34.9
Total	181548114	100

Selection of DSM and DTM Interpolation Technique

Given that the LIDAR data from TPS Australia was utilised to produce the ground truth DSM, DTM, and ultimately a CHM, it is essential to consider potential sources of errors that may arise during the creation of these DEMs. One source of these errors can be attributed to incorrect interpolation selection, as emphasised in Gong et al. (2000). In this scenario, errors in the ground truth DEMs could lead to error propagation when applied to INSAR fusion CHM generation or the mischaracterisation of the INSAR and RF MLM for CHM generation.

Overview of the interpolation concept

Interpolation involves mathematical and statistical approximation and estimation, where sampled data values of a variable within a specific area are used to estimate values at unsampled locations (Burrough et al., 1998; Li et al., 2004). In the context of digital terrain modelling, interpolation is employed to estimate the height at a given point based on the known points in its vicinity (Li et al., 2004). It can be broadly classified into two main methods: deterministic and geostatistical. Deterministic methods create surfaces based on either the available sample points or the degree of similarity obtained from the existing smoothed surface (Liu, 2008; Salekin et al., 2018). However, these methods do not take into account any potential spatial processes in the input data (Ismail et al., 2016; Liu, 2008). On the other hand, geostatistical methods create surfaces based on statistical relationships among input points and the presence of spatial correlation (Ismail et al., 2016; Liu, 2008; Salekin et al., 2018).

Overview of the interpolation techniques available

In this study, ArcGIS Pro was used to interpolate the different DEMs from the classified LIDAR point cloud. where two deterministic interpolation methods, Binning and Triangulation (Delauney TIN) are available.

Triangulated Interpolation Network (TIN)

TIN interpolation constructs a mesh of triangles, with each triangle consisting of three data points, to connect the available data points. Within each triangle, the value at unsampled locations can be estimated using linear interpolation, with equal weight assigned to each vertex of the triangle (ESRI, 2023; Khan et al., 2021). Another interpolation method is natural neighbour, where values are interpolated based on the weights within the triangles, determined by the proximity to the point being interpolated and their influence within the triangle itself (Szypuła, 2016).

In the LAS to raster tool, a special TIN interpolation technique known as Delaunay triangulation is applied. This method creates triangles without data points within them and ensures no intersections occur between the triangles. The value of the point to be interpolated is determined by utilising the linear gradient of the triangle it resides in (Fan et al., 2014; Gonçalves, 2006).

Binning Interpolation

Binning is an interpolation method that assigns pixel values based on points falling within defined pixel boundaries. For LIDAR data, pixel values are determined by the LAS points present within the pixel (ESRI, 2023; Prior et al., 2022; Stack & Narine, 2022). Pixel values can be interpolated based on the average, minimum, or maximum of all data points within it, or using other methods such as Inverse Distance Weighting (IDW) and Nearest Neighbour. IDW assigns weights to data points that decrease as they move further from the calculated point, emphasising closer reference points (Gomroki et al., 2017; Li & Heap, 2008). In contrast, Nearest Neighbour assigns the value of the nearest reference point (Gomroki et al., 2017).

Interpolation techniques for DSM generation

After reviewing studies that use LIDAR data to generate DSMs over forested areas, the binning interpolation type was chosen (Bonneau et al., 2020; DiGiacomo et al., 2020; Stack & Narine, 2022). Bonneau et al. (2020) examined the average, maximum, IDW, and nearest neighbour binning methods. They found that the maximum binning method performed the best in estimating tree height, despite introducing more false positives. Therefore, the maximum binning method was selected, following the approach of Stack and Narine (2022). Additionally, there was a consensus on using the linear void fill method (Bonneau et al., 2020; Stack & Narine, 2022).

Interpolation techniques for DTM Generation

Binning interpolation was chosen for DTM generation (Bindzarova Gergelova et al., 2021; Bonneau et al., 2020). Moreover, the average binning type was commonly employed (Bindzarova Gergelova et al., 2021; Bonneau et al., 2020). However, Bindzarova Gergelova et al. (2021) suggested the use of IDW, while Razas et al. (2023) demonstrated that IDW might lead to biased results due to its sensitivity to clustered data and outliers (Bindzarova Gergelova et al., 2021; Muir et al., 2017). The adoption of a natural neighbour void filling method was also recommended, which aligns with the findings of Căţeanu and Ciubotaru (2021), indicating its suitability for forested regions with varying canopy densities.

The interpolation techniques selected for generating the DSM and DTM from the raw LIDAR

ground truth data, along with the parameters used during the process using the LAS dataset to raster in ArcGIS Pro, are outlined in Table 4.3.

Table 4.3: Selected processing parameters to be used for this study when a DSM and DTM must be generated

DEM type	Interpolation type	Cell assignment	Void fill
DSM	Binning	maximum	linear
DTM	Binning	Average	Nearest neighbour

Generation of LIDAR ground truth data

An overview of the process of creating the different DEMs from the raw LIDAR data supplied by RPS Australia is highlighted in Figure 4.1

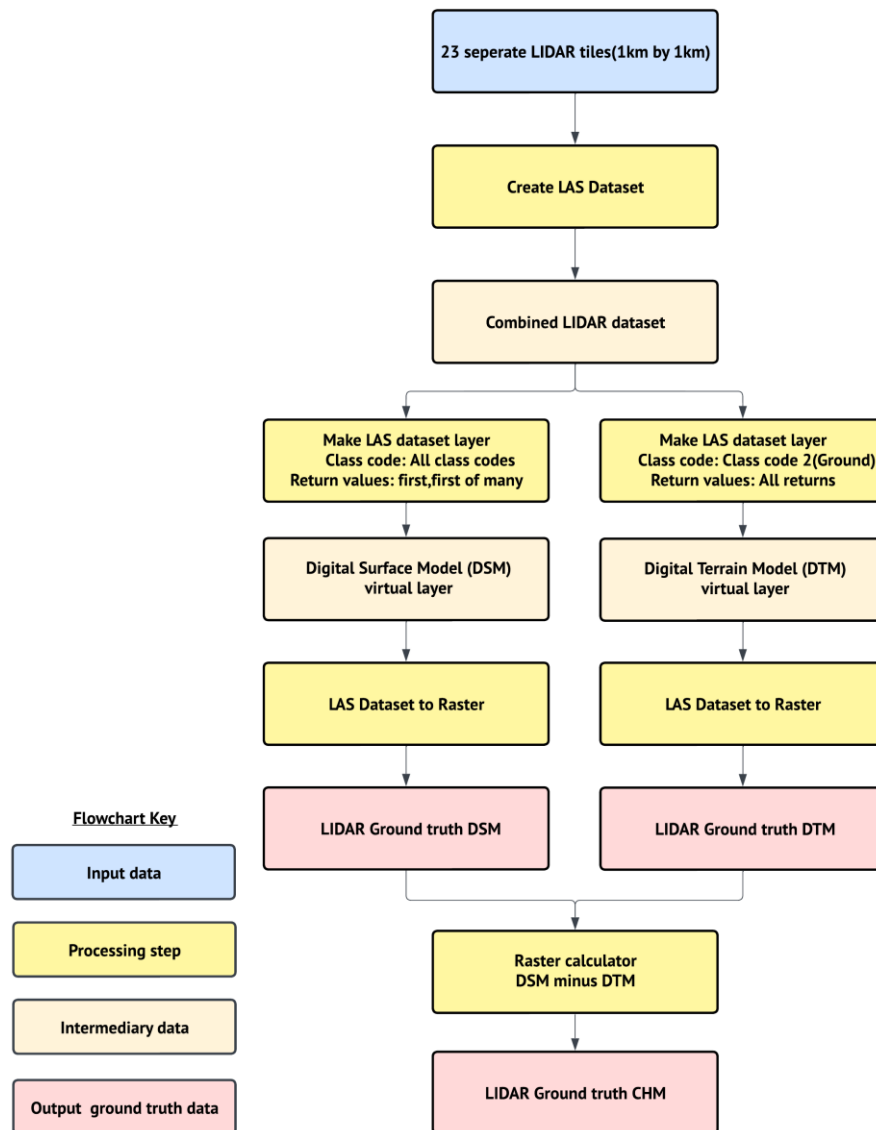


Figure 4.1: Flowchart showing the processing steps in processing the LIDAR ground truth data into DSM, DTM and CHM which were used in 1) The INSAR CHM generation 2) The accuracy assessment between the INSAR fusion CHM and the RF MLM CHM

Each of the 23 1km by 1km LIDAR LAS files were already classified by from RPS Australia, including specific classification for ground and noise points.

Typically, from LIDAR data, the points that are not classified as ground and correspond to the first returns when over a tree canopy can be used to generate a DSM that corresponds to the top of the vegetation (Vepakomma et al., 2008). Moreover, the remaining points that are classified as ground only, irrespective of their return number, can be used to generate a DTM. Finally, a CHM can be

generated by subtracting the DSM from the DTM (Vepakomma et al., 2008). To process the LIDAR data used as ground truth for the DSM, DTM, and CHM in this survey, the suite of tools in the ArcGIS Pro software from ESRI was used as shown in Figure 4.1. The coordinate system and projection used was GDA2020 MGA Zone 54 with its corresponding GDA 2020 AHD.

Merging of LIDAR tiles

To merge the 23 1km by 1km tiles that encompass the study area, the "Create LAS dataset" tool was used to generate a virtual layer that accurately combines the different LAS tiles into one.

DSM generation

Using the merged LIDAR virtual layer, the "Make LAS dataset to layer" tool was used. All class codes were selected except for the one corresponding to the classified noise, and only the return values corresponding to the first and first of many categories were selected. When executed, a virtual DSM is generated.

The "LAS dataset to Raster" tool was then used on the virtual layer to generate a DSM raster following the appropriate parameters outlined in Table 4.3 for a DSM over vegetated areas. Moreover, the DSM is generated with a cell size of 1m.

DTM generation

The merged LIDAR virtual layer created previously was then used in the "Make LAS dataset to layer" tool. Only class codes corresponding to ground points were selected, but this time, all the different returns are selected, creating a virtual DTM.

This virtual DTM was then used as input in the "LAS dataset to Raster" tool, using the parameters found to be more appropriate for a DTM over forested areas shown in Table 4.3, to generate a DTM raster with a resolution of 1m.

CHM generation

In the raster calculator, both the DSM and DTM rasters generated were used as input, and the CHM was created by subtracting the DSM raster from the DTM raster.

Setting up extraction areas based on completed thinning cycles

ForestrySA (2016) provides the year at which the different sections of the *Pinus radiata* plantation in the study area were planted, which include Knot Hill and Mount Panorama. The planting years range between 1993 and 2015. From these plantation years, the approximate age of the trees can be inferred, and thus the number of thinning cycles they have undergone. According to ForestrySA (2018), thinning cycles take place approximately every 10 years (Figure 4.2) where 30 to 50% of trees are felled until they are fully removed once they reach around 40 years.

Figure 4.2 has been removed due to Copyright restrictions

Figure 4.2: Graphical overview adapted from ForestrySA (2018) of the timeline for the thinning process in ForestrySA *Pinus Radiata* plantations, highlighting the approximate years at which the thinning process takes place

Therefore, the breakdown of the study area based on the plantation date (Appendix A) was categorised based on how many thinning processes the trees have undergone, as highlighted in Table 4.4. Moreover, these thinning categories were used throughout this study to compare the different accuracies and overall potential of INSAR and RF MLM in generating a usable CHM over *Pinus radiata* plantations.

Table 4.4: Breakdown of the study area based on the number of thinning processes carried out

Thinning Category	Thinning process	Plantation date assigned
T0	No thinning	2015
T1	First commercial thinning	2006,2007,2008
T2	Second commercial thinning	2000,2001,2004,2005
T3	Third commercial thinning	1993,1995

INSAR DSM Generation

Sentinel-1 description and data selection

The data utilised in this study were obtained from the Sentinel-1 mission, currently operated by the European Space Agency (ESA), and funded by the European Union. The mission comprises two identical satellites, Sentinel-1A and 1B, both orbiting at an altitude of 693 km in the same sun-synchronous and near-polar orbit, albeit phased (Fletcher, 2012; Torres et al., 2017). Under normal operation, the revisit time is expected to be 6 days. However, due to the unavailability of Sentinel-1B, the revisit period has been extended to 12 days (Carstairs et al., 2022).

The instrument onboard can operate in a dual-polarisation mode, acquiring both HH + HV polarisations or VV + VH when using the interferometric wide-swath mode (IW). This mode generates products with a 250 km swath width and range and azimuth resolutions of 3 m and 20 m, respectively, with incident angles ranging between 31° and 46° (Fletcher, 2012), making it suitable for this study. The chosen acquisition mode's resulting Level-1 product was the Slant-Range Single Look Complex (SLC), as it contains amplitude and phase information essential for INSAR operations which is absent in the Ground Range Detected (GRD) product (Fletcher, 2012).

In repeat-pass INSAR configurations like the Sentinel-1 constellation, the longer the time between image pair acquisition, the greater the potential degradation in the accuracy of obtaining forest height measurements. This is due, in large part, to the effects of wind and, to a lesser extent, the sensitivity of electromagnetic waves to tree components such as leaves (Seppi et al., 2022). Therefore, it was ensured that the selected image pair had the minimum temporal baseline possible—12 days.

Additionally, the Sentinel-1 image pairs were constrained to have perpendicular baselines between -200 m and 200 m to mitigate potential volume decorrelation, which may occur and is below the critical baseline of 5 km, beyond which there is a complete loss of correlation between images (Seppi et al., 2022). For this study, an additional constraint was added: the perpendicular baseline should be lower than 120 m to ensure accurate forest height estimation through the INSAR process (De Petris, Sarvia, et al., 2022b).

Furthermore, it was ensured that the selected INSAR image pairs were from days with no rainfall during or a few days prior to minimise disturbance to the phase signal caused by delays on the surface or within the atmosphere due to the presence of water vapour, which leads to a loss of coherence (Braun, 2021; Palm, 2022). It was also ensured that the INSAR image pairs were in a

timeframe similar to the ground truth LiDAR data acquisition date over the study area, as highlighted in Table 4.1 by RPS Australia.

The resulting INSAR image pair consists of an image from the 31st of January 22 and another from the 12th of February as a pair, with a perpendicular baseline of 118 m, a temporal baseline of 12 days, and a height of ambiguity of 150.3 m, each with individual characteristics as shown in Table 4.5

Table 4.5: Individual characteristics of the SAR images selected as input for the INSAR process

No	Acquisition date	Acquisition mode	Acquisition time(local)	Polarisation	Pass	Relative orbit
1	31 Jan 22	IW	06:27:38	VV +VH	Descending	162
2	12 Feb 22	IW	06:27:38	VV +VH	Descending	162

INSAR processing workflow

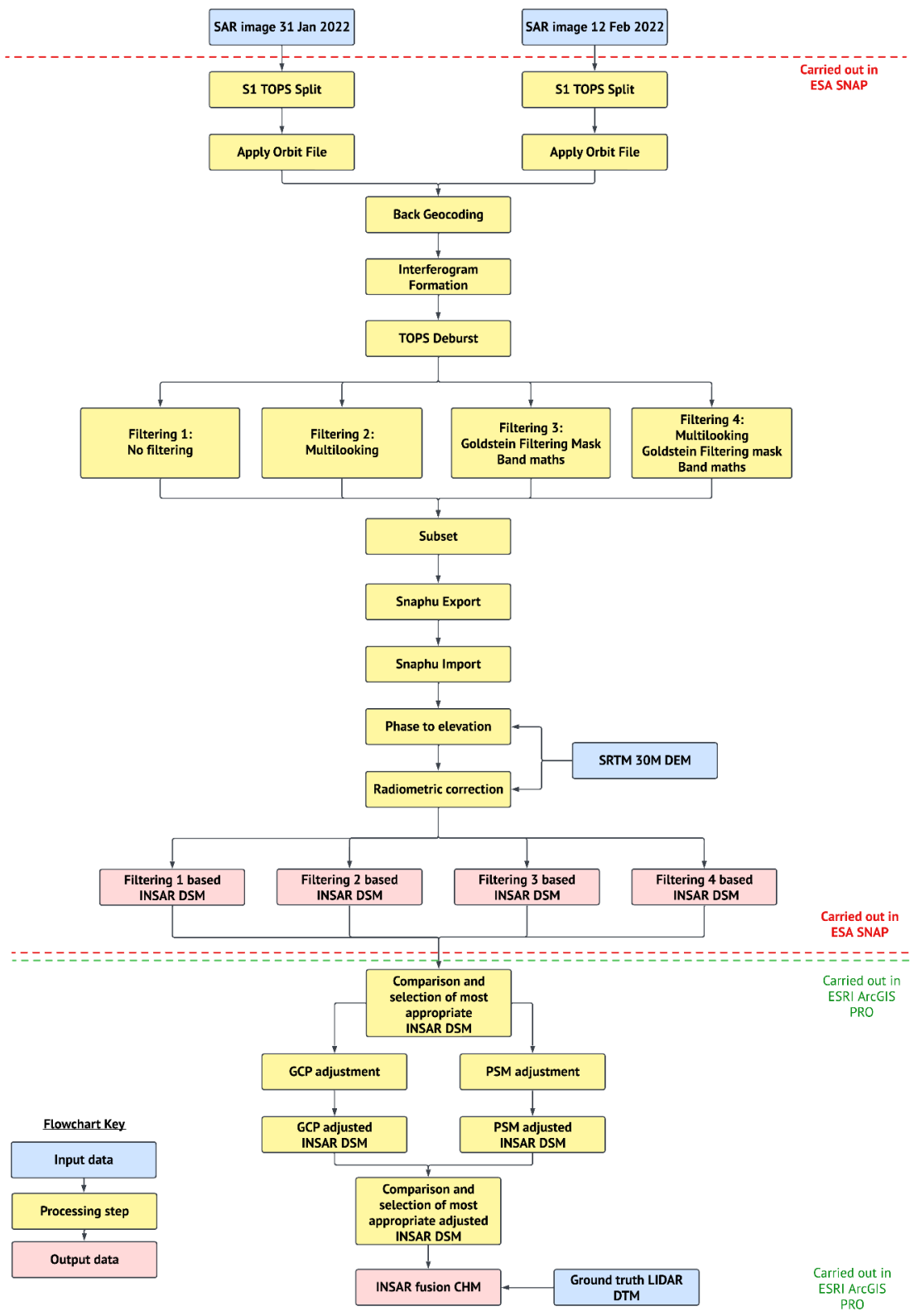


Figure 4.3: Flowchart showing the INSAR processing steps carried out in ESA SNAP for the DSM generation and the subsequent DSM adjustment and INSAR fusion CHM generation in ESRI ArcGIS Pro.

The initial processing involves individual treatment of the two Sentinel images. The Sentinel-1 image from 31st January was chosen as the master image, with the image from 12th February designated as the slave image (Ma et al., 2019; Zhou et al., 2009).

Utilising the Sentinel-1 toolbox, an open-source software developed for the Sentinel Application Platform (SNAP), the INSAR processing workflow was adapted from Braun (2021) to achieve accurate processing with the individual steps and chosen settings shown below

Individual SAR image preprocessing

The preliminary steps include selecting the available sub-swaths in the SLC data of both the master and slave images. The TOPSAR Split tool was used to isolate the VV polarisation, which is considered most suitable for canopy height retrieval (Braun, 2021; Palm, 2022).

Apply Orbit File

To ensure accurate positioning and velocity during acquisition, the precise orbit file, automatically updated and generated with a 3rd polynomial degree, was applied using the Apply Orbit File tool (Filipponi, 2019; Fomelis et al., 2018).

Combination of the SAR Scene

The S-1 Back Geocoding tool was employed to co-register sub-pixels from the master image to the slave image using the SRTM 1 arcsec DEM. A small window size of 5 range looks and one azimuth look was used to minimise resolution reduction during resampling (S1TBX, 2023).

Interferogram Generation

An interferogram was generated using the Interferogram Formation tool. The flat-earth phase was estimated and subtracted from the total phase information. Coherence between the two SAR scenes was estimated using specified parameters (Braun & Veci, 2021).

Topsar Deburst

Although the study area was covered within one burst, a debursting process was applied to maintain the integrity of the INSAR procedure. This process combines all bursts within the same sub-swath into a single image (Ab Rahman et al., 2023; Braun, 2021).

Multilooking

The multilooking (ML) procedure serves two purposes, the first of which involves transforming the interferogram from its initial spacing of 14.1m azimuth and 2.3m slant range (as depicted in Figure 2.1) to a coarser average ground range (GR) pixel size of 12.9m by 12.9m. In this study, a window size of 5 range looks and a single azimuth look were chosen for this process. By adopting this configuration, the multilooking operation contributes to improving coherence estimation and overall phase fidelity, albeit at the cost of spatial resolution. This is achieved by performing a spatial averaging, wherein the average value of the data within the moving pixel window is assigned to the resulting pixel with dimensions 12.9m by 12.9m (Ab Rahman et al., 2023; Bouaraba et al., 2018).

However as seen in Arslan et al. (2022), selecting all the input bands present in the interferogram, that describe phase, intensity and coherence way lead to errors in the multilooking process, and thus only the coherence information was selected to be multilooked

Goldstein filtering

The Goldstein filter was applied to the multilooked interferogram, enhancing its accuracy and minimizing residuals prior to the unwrapping phase. This enhancement results in improved interferogram fringe visibility through a reduction in phase noise (Baran et al., 2003; Mohammadi et al., 2020; Zylshal et al., 2021). The filter employs nonlinear algorithms, dynamically adjusting its strength based on the density of fringes in the selected window size. In areas with high fringe density and heterogeneity, the filter is more robust, while the opposite is true for low-density homogeneous areas. This adaptability is achieved through the application of Fast Fourier Transformations (FFT) (Y. Wu et al., 2023).

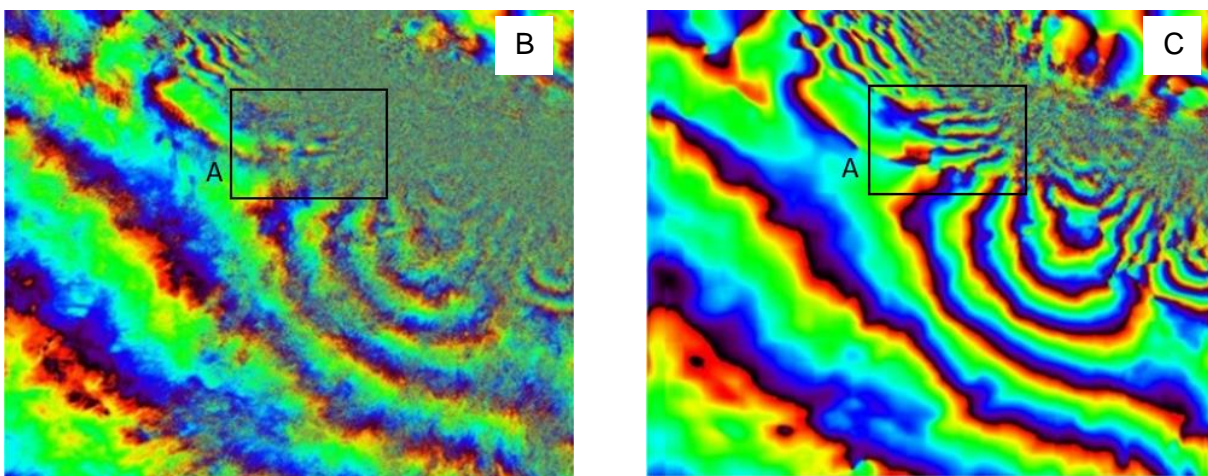


Figure 4.4: Visualisation of the variation in strength of the Goldstein phase filtering over an area A from where it is not applied (B) and where it is applied (C). the closer the fringes are to one another in the interferogram, the stronger the filtering, adapted from (Ghulam et al., 2010)

For this study, a default FFT size of 64 was utilised along with a window size of 3.

Furthermore, the Goldstein phase filtering mask (GFM) avoids applying the filter over areas with a coherence lower than 0.3 (threshold used in this study). This is done to prevent the filter from producing erroneous results in those areas (Y. Wu et al., 2023).

Band Maths

As the presence of low coherence areas in the input interferogram can lead to errors in the phase unwrapping process or even cause it to fail completely, a filter was generated for this study using the Band Maths (BM) functions in ESA SNAP. It was used to create IF statements based on coherence, where pixels with coherence lower than 0.3 – a threshold that is based on the average coherence for a dry forest as observed in Nurtyawan and Yulanda (2020) were completely removed from the interferogram and thus preventing error propagation (Bamler & Hartl, 1998; Bi & Wei, 2016; Li et al., 2020; Weike & Goulin, 2012).

Subset to the Study Area

The resulting SAR data was subset to the study area using the Subset from View tool, with a polygon projected in WGS84, leaving some data outside the polygon for trend analysis during the phase unwrapping process.

Unwrapping Process

Using the Snaphu export tool, the data was prepared for the unwrapping process, employing the Snaphu algorithm. The Minimum Cost Flow algorithm was chosen for its efficacy in similar studies (Palm, 2022; Persson & Fransson, 2017; Wang et al., 2022).

Phase to Elevation and Radiometric Correction

Unwrapped phase information was converted to elevation using a DEM as a reference. The output DSM underwent radiometric correction to address geometric distortions and was converted from WGS84, a global datum to a datum designed for the Australian region, GDA94 zone 54, and had a final pixel size of 14m (Filipponi, 2019; Mohammadi et al., 2020). The resulting DSM is ready for subsequent analysis in ArcGIS Pro.

Filtering methodologies

It was observed that ESA SNAP offers a range of filtering processes that can be applied to Sentinel-1 data. These processes aim to either enhance coherence and phase fidelity in the interferogram or eliminate areas of low coherence prior to the phase unwrapping process. The goal is to improve accuracy and success rate. Therefore, in this study, various combinations of these filters and masks were tested to identify the most accurate combination for generating an INSAR DSM over forested areas (Figure 4.3).

Due to the importance of the Goldstein phase filtering method highlighted in a range of different studies, regardless of the permutation, the main question was whether to use the Goldstein coherence mask or not. Furthermore, considering the similarities between the latter and the band math filter in targeting low coherence areas, it was decided to apply both simultaneously.

The different permutations for the filtering process are shown in Table 5.6. Moreover, a coherence threshold of 0.3 was used for BM and GFM.

Table 4.6: Permutations of the filtering process carried out in the INSAR DSM generation process in ESA SNAP

Filtering permutation	Filters used
Filtering 1	No Filters
Filtering 2	ML
Filtering 3	GFM, BM
Filtering 4	ML, GFM, BM

Pre-processing of INSAR DSM before adjustment

DSM Filtering, Study Area Extraction, and Reprojection

The INSAR DSM underwent filtering before any additional processing to eliminate outliers that might be present due to the incoherent sections in the INSAR pairs, which can be associated with noise (Sefercik et al., 2020). A focal statistics using a 3 by 3 median filter was deemed appropriate for outlier removal and was applied using the focal statistics tool in ArcGIS Pro (Bucher & Bayer, 2013; Sefercik et al., 2020). Furthermore, the DSM was extracted to the study area and reprojected to GDA2020 Zone 54.

Adjustment of DSM

The DSM was compared to the LiDAR DSM in areas where the latter showed as ground, with no visible tree canopy cover. It was observed that there was a discrepancy between the elevation values. Consequently, it was decided that the INSAR DSM should be adjusted to the LiDAR DSM to ensure that it accurately reflects the elevation of ground areas within the same elevation range

in both LiDAR and INSAR DSMs similar to the concept outlined in Yun et al. (2012). Two methods were investigated, involving the use of either existing Permanent Survey Marks (PSMs) or Ground Control Points (GCPs) which were manually generated.

PSM Adjustment

The locations of survey marks across South Australia were obtained from Location SA (2023), which is updated daily with survey marks of different types and purposes, currently comprising 190,000 entries across South Australia. Survey marks were extracted to the study area, and the points selected for the analysis were those classified as permanent survey marks, with both their horizontal and vertical values adjusted through a least square or any adjustment, respectively. The resulting 14 PSMs, measured in the MGA2020 and AHD vertical datum in MGA Zone 54 between October and November 2022, each with a vertical order of 5, were used in the analysis.

The elevation from the filtered INSAR DSM at those PSM locations was extracted with the "extract value to points" tool, and the difference in elevation was calculated by subtracting the PSM elevation from the INSAR DSM elevation. The resulting difference in elevation was used as input in the exploratory interpolation tool, which utilise the 20 interpolation techniques available in ArcGIS Pro to create and rank the interpolated layer. This interpolated layer aims to be the vertical INSAR correction layer based on the highest prediction accuracy. The interpolated correction layer that ranked first was then subtracted from the filtered DSM to create a corrected INSAR DSM.

GCP adjustment

A set of ground control points (GCPs) were manually established using planet imagery and the LiDAR CHM as a reference. The GCPs were in areas where within a 15m buffer no-forested areas were present to ensure that the GCP pixel in the INSAR DSM did not represent a forested pixel or a pixel with vertical objects such, as buildings, present. Additionally, efforts were made to distribute the GCPs evenly across the study area. A total of 56 GCPs were generated, and the elevations of the INSAR DSM and LiDAR DSM at these locations were extracted. The elevation difference was calculated by subtracting the LiDAR DSM elevation from the INSAR DSM elevation. Subsequently, the elevation difference for these 56 GCPs was interpolated using a similar method as the PSM adjustment to correct the INSAR DSM.

Generation of INSAR fusion CHM

Both the PSM and GCP-corrected INSAR DSMs were utilised to create a Canopy Height Model (CHM) by subtracting the LiDAR DTM from the corrected INSAR DSMs with the aid of the raster calculator tool. In both cases, a 3 by 3 median focal statistics filter was applied to the CHM to further eliminate any potential noise.

RF MLM CHM generation

In this method of estimating the CHM, Sentinel-1, Sentinel-2 data, as well as the SRTM DEM and the Land classification mask from ESA, were used and processed as predictor variables using a machine learning algorithm that GEDI products as training data. The data were acquired, where possible, over a similar time frame as that of the LIDAR ground truth survey, from the 1st of January 2022 to the 28th of February 2022, and were clipped to the study area. Figure 4.5 illustrates the methodology which was entirely carried out in the in the Google Earth Engine Platform (GEE).

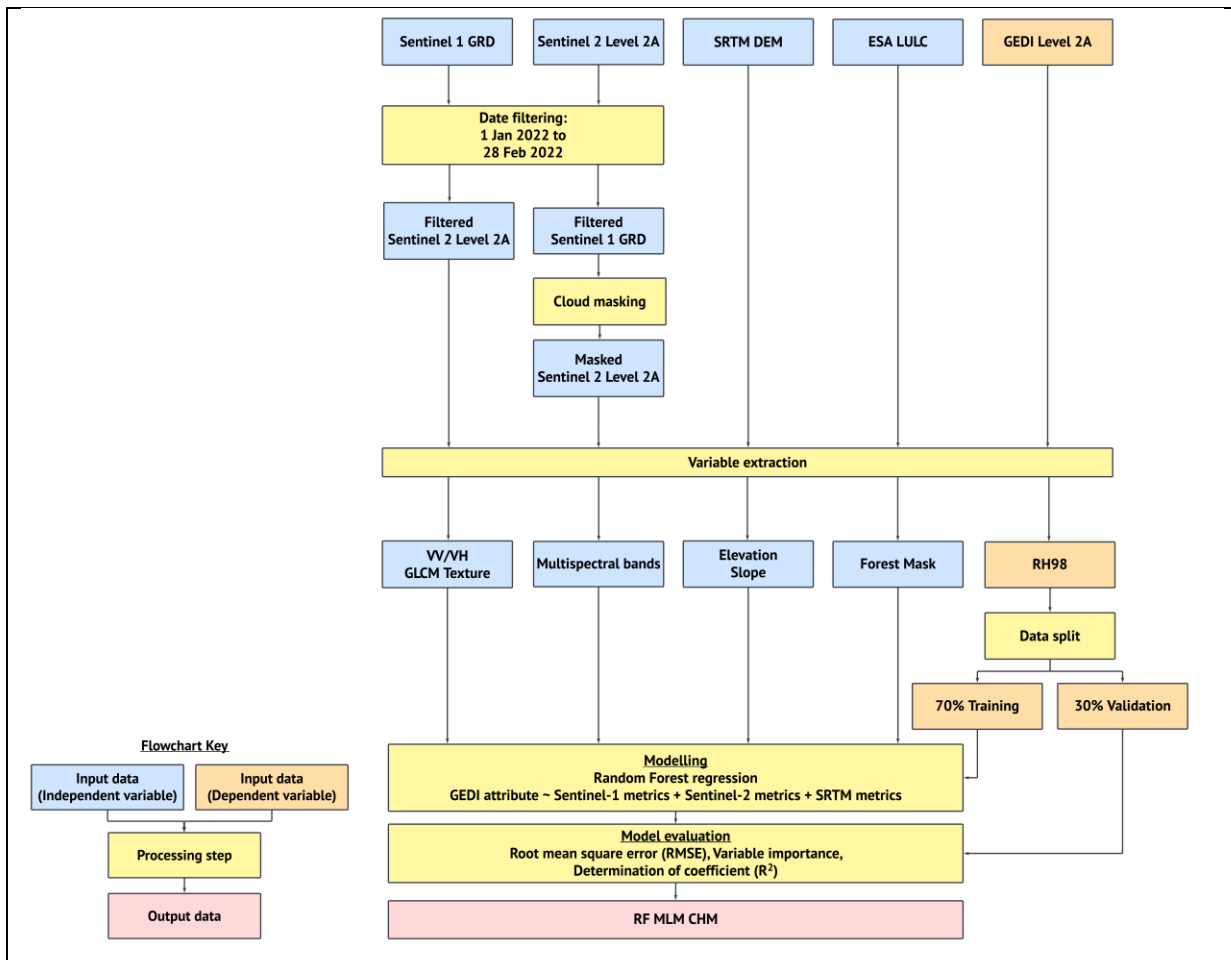


Figure 4.5: Flowchart showing the generation of a CHM based on an RF MLM carried out in Google Earth Engine (GEE)

RF MLM input parameters

The RF machine learning algorithm available in GEE was chosen for this analysis due to its proven accuracy in similar forest models. It establishes models by fitting functional relationships between forest attributes, specifically canopy height extracted from GEDI, and the predictor variables outlined in Table 4.7 (Tsao et al., 2023; Xi et al., 2022). Among the input parameters available for RF in GEE, the number of classification trees was set to 50, and the rest, as described in Teluguntla et al. (2018), were left with their default values based on the script from Courage (2021), which was adapted for this study. Ten thousand points were randomly selected across the study area, with 70% used for training data and the remaining 30% for validation purposes. The overall model accuracy was subsequently assessed using the R^2 and RMSE of observed and predicted tree canopy height values with the validation data.

Training Dataset

Data from the GEDI LIDAR mission on board the ISS was selected as the training dataset, specifically the Level 2A product. This product comprises percentiles representing the relative height to the ground (RH), describing the total energy returned to the sensor from the top of the tree canopy (Schwartz et al., 2022). In GEE, a raster version of Level 2A data was used, updated monthly, with an average spatial resolution of 25m. The RH98 data were extracted, as it is shown to be still representative of the top of a tree canopy while being less sensitive to noise and outliers than RH100 (Ghosh et al., 2022; Simard & Denbina, 2017).

Predictor Variables

Sentinel-1

The Sentinel-1 Ground Range Detected (GRD) products, already calibrated and ortho-corrected, were used to obtain the median values of backscatter of VV and VH data. Additionally, the median Gray-level Co-occurrence matrix (GLCM) information involving texture statistics such as contrast, variance, entropy, and dissimilarity was calculated from both VV and VH data, providing texture information about water content and overall tree structure (Numbisi et al., 2019). A total of 12 Sentinel-1 based texture features were generated and used as input.

Sentinel-2

Sentinel-2 LEVEL-2A from the Multispectral instrument was used as the optical data input. Using the QA60 band, the 10th and 11th bits assigned to clouds and cirrus were removed. Moreover, using the mask2cloud function, another cloud mask was generated, rejecting images with a cloudy pixel percentage higher than 10%. Finally, specific bands from the

L2A products were selected as input.

Auxiliary data

SRTM

The SRTM used for this analysis was SRTMGL2_003, a DEM with a resolution of 30m at the equator, used to generate elevation data. Moreover, it was used to derive slope information.

ESA Forest Mask

To ensure that only forested areas were selected for the CHM calculation, the ESA World Cover data V200 derived from Sentinel-1 and 2 data provided a 10m land cover classification dataset. The forest class category was used to mask the processing to only forested areas (Zanaga et al., 2022).

The predictors used in the RF MLM in the GEE platform for this study are showed in Table 4.7

Table 4.7 Predictors used and their sources to be used as input in the RF MLM, adapted from Luo et al. (2023)

Data source	Predictors	Description	References
Sentinel-1	VV	VV band	(Torres et al., 2012)
	VH	VH band	
	VV_ent	Entropy calculated by GLCM based on VV	(Haralick et al., 1973)
	VH_ent	Entropy calculated by GLCM based on VH	
	VV_contrast	Contrast calculated by GLCM based on VV	
	VH_contrast	Contrast calculated by GLCM based on VH	
	VV_var	Variance calculated by GLCM based on VV	
	VH_var	Variance calculated by GLCM based on VH	
	VV_corr	Correlation calculated by GLCM based on VV	
	VH_corr	Correlation calculated by GLCM based on VH	
	VV_diss	Dissimilarity calculated by GLCM based on VV	
VH_diss	Dissimilarity calculated by GLCM based on VH		
Sentinel-2	B2	Blue band (490 nm)	(Drusch et al., 2012)
	B3	Green band (560 nm)	
	B4	Red band (665 nm)	
	B5	Vegetation Red Edge band (705 nm)	
	B6	Vegetation Red Edge band (740 nm)	
	B7	Vegetation Red Edge band (783 nm)	
	B8	Near infrared (NIR) band (842 nm)	
	B11	Short Wave infrared (SWIR) band (1610nm)	
	B12	Short Wave infrared (SWIR) band (2190nm)	
SRTM	Elevation	30M DEM	(Farr et al., 2007)
	Slope	Slope derived from 30m DEM	
ESA world cover V200	Forest mask	10m Land classification	(Zanaga et al., 2022)

The predictors derived from Sentinel-1 and Sentinel-2 were extracted to their median value and resampled to 30m to account for the different range of resolutions present in the data.

Comparing INSAR fusion CHM with RF MLM CHM using LIDAR ground truth CHM

The overview of the methods for this process is shown in Figure 4.6

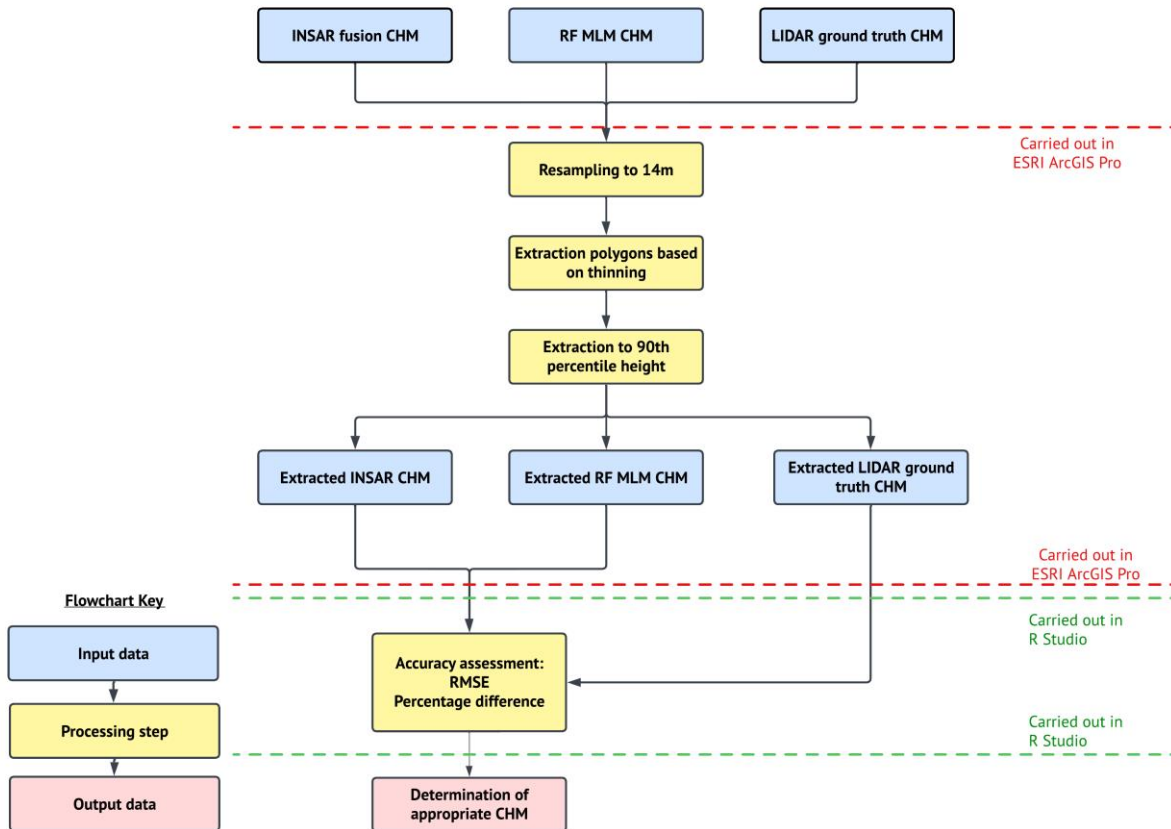


Figure 4.6: Flowchart showing the process used for the comparison of the INSAR fusion CHM with the RF MLM CHM and their overall accuracy assessment using the LIDAR CHM ground truth data.

Resampling

The input CHMs from the RF MLM and INSAR, as well as the LIDAR CHM, were all resampled to 14 m to match the lower resolution INSAR CHM using the resample tool in ArcGIS Pro. Moreover, it was ensured that all three CHMs were reprojected to GDA 2020 and its associated AHD if needed, to allow the comparison. The interpolation technique used for the resampling and reprojected was the bilinear interpolation technique.

Extraction of Canopy Height from Polygons

Canopy height information derived from INSAR and LIDAR ground truth was extracted from individual polygons, specifically the 90th percentile height. This extraction was performed using the Zonal Statistics as Table tool. The results were visualised in R Studio, where the average of

individual polygons was calculated to provide an overall value for the tree canopy height within the main polygon.

CHAPTER 5 RESULTS

Breakdown of the study area based on thinning cycles.

The breakdown of the study area, based on thinning cycles, resulted in 19 polygons that spanned 4.4 km² out of the original 22.9 km² of the study area. A simplified breakdown of the effective area used in this study is shown in Table 5.1, where the missing area can be allocated to bare ground, and the native vegetation is reserved in Mount Panorama and Knott Hill.

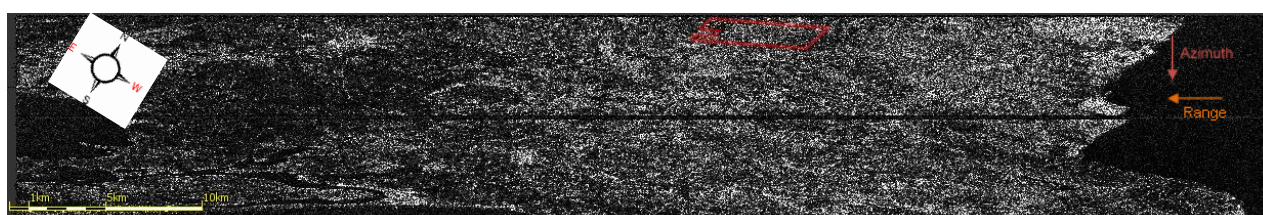
Table 5.1: Characteristics of the *Pinus radiata* plantation in the study area falling under the different thinning category

Thinning category	Number of distinct sections	Total area (km ²)
T0	2	0.28
T1	8	2.17
T2	7	1.75
T3	2	0.21

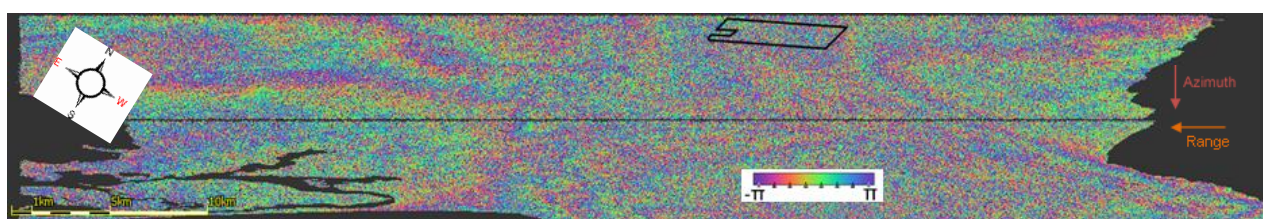
Visualisation of the INSAR process data transformation

It can be observed from the INSAR process outlined in Figure 4.3 that several steps are undertaken to transform, filter, and correct the input Sentinel-1 SAR data. The visualizations of the outputs of the primary processes were generated, as depicted in Figure 5.1, using the limited ESA SNAP software, which lacks customization options for various mapping features, including the scale bar and the north arrow. as these outputs cannot be exported to more suitable tools like ArcGIS.

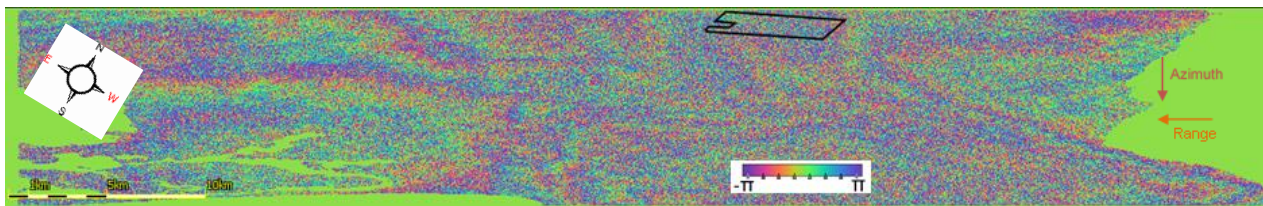
A



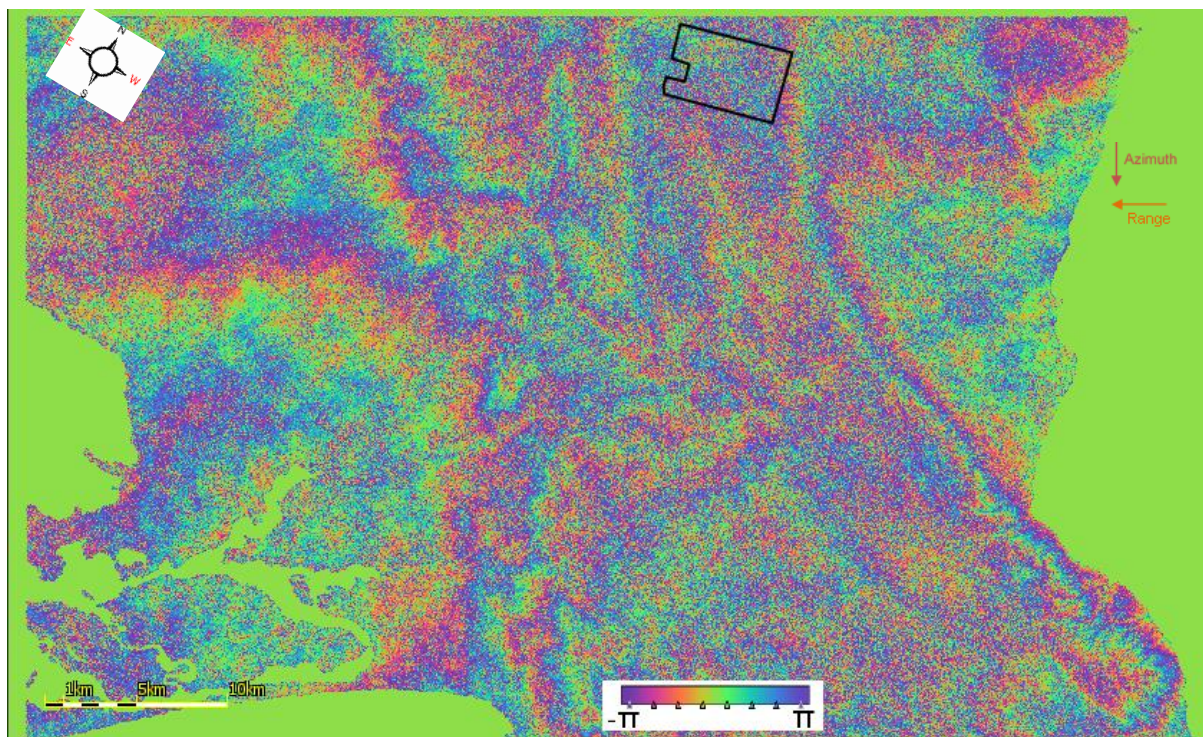
B

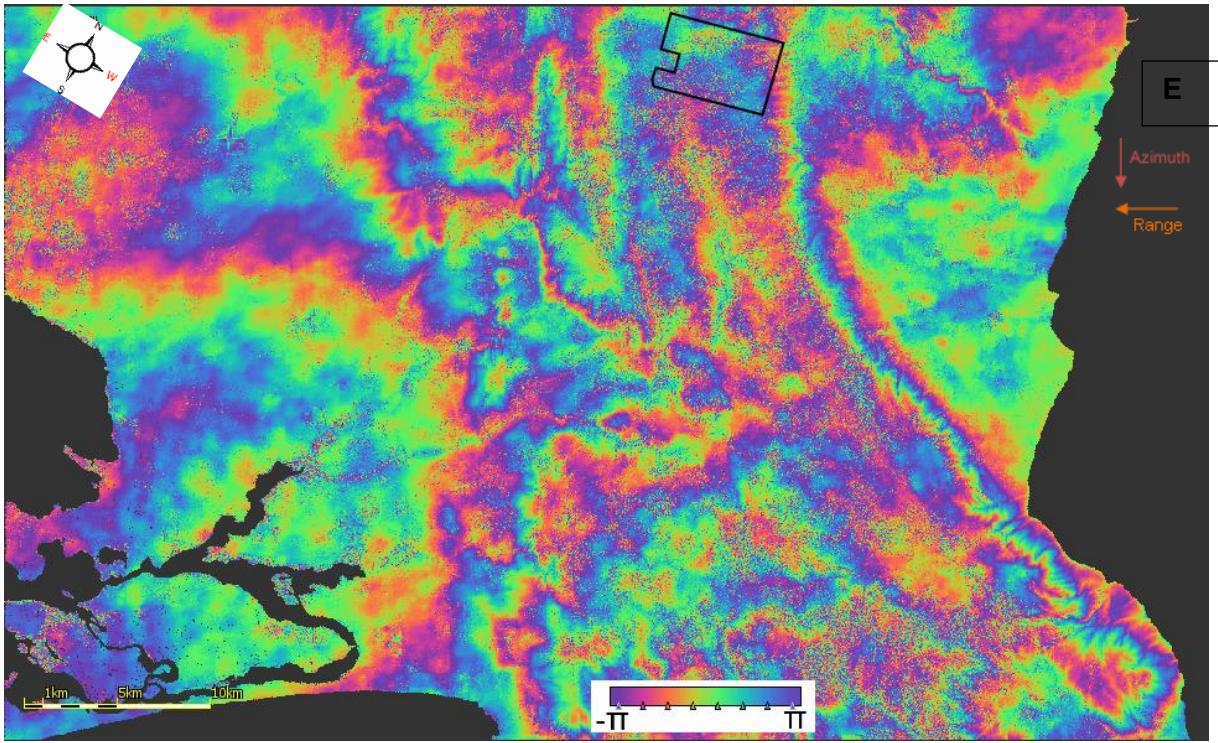


C

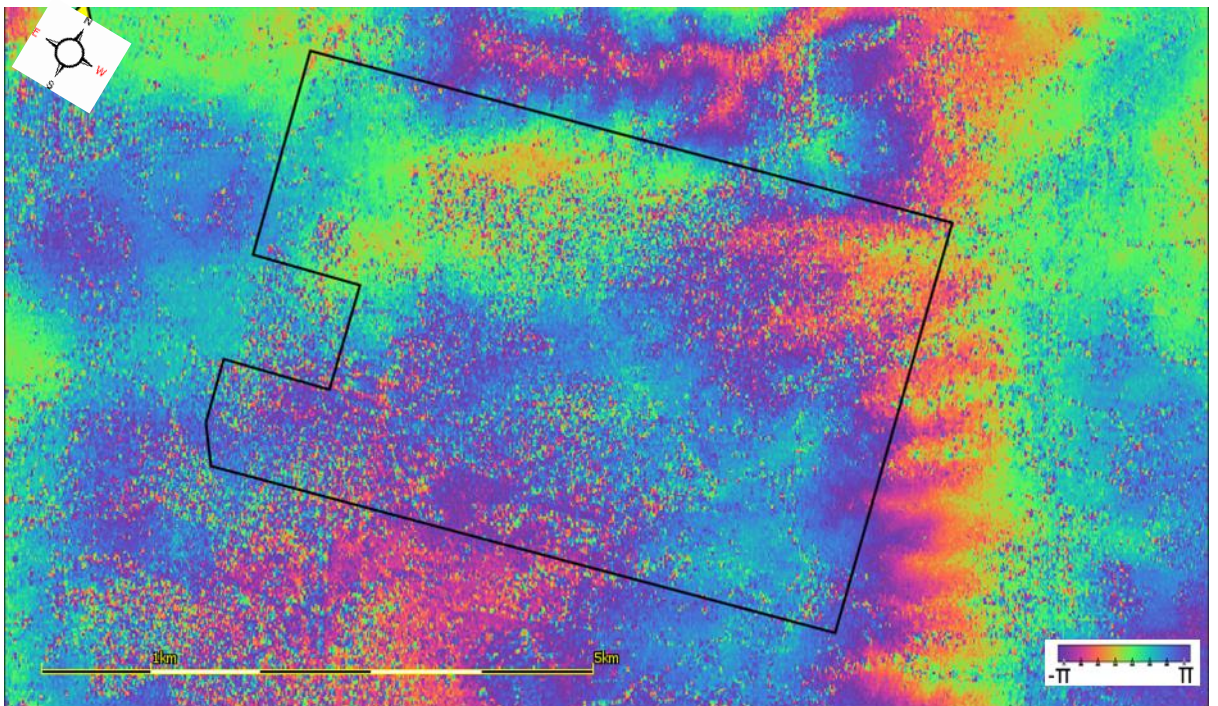


D

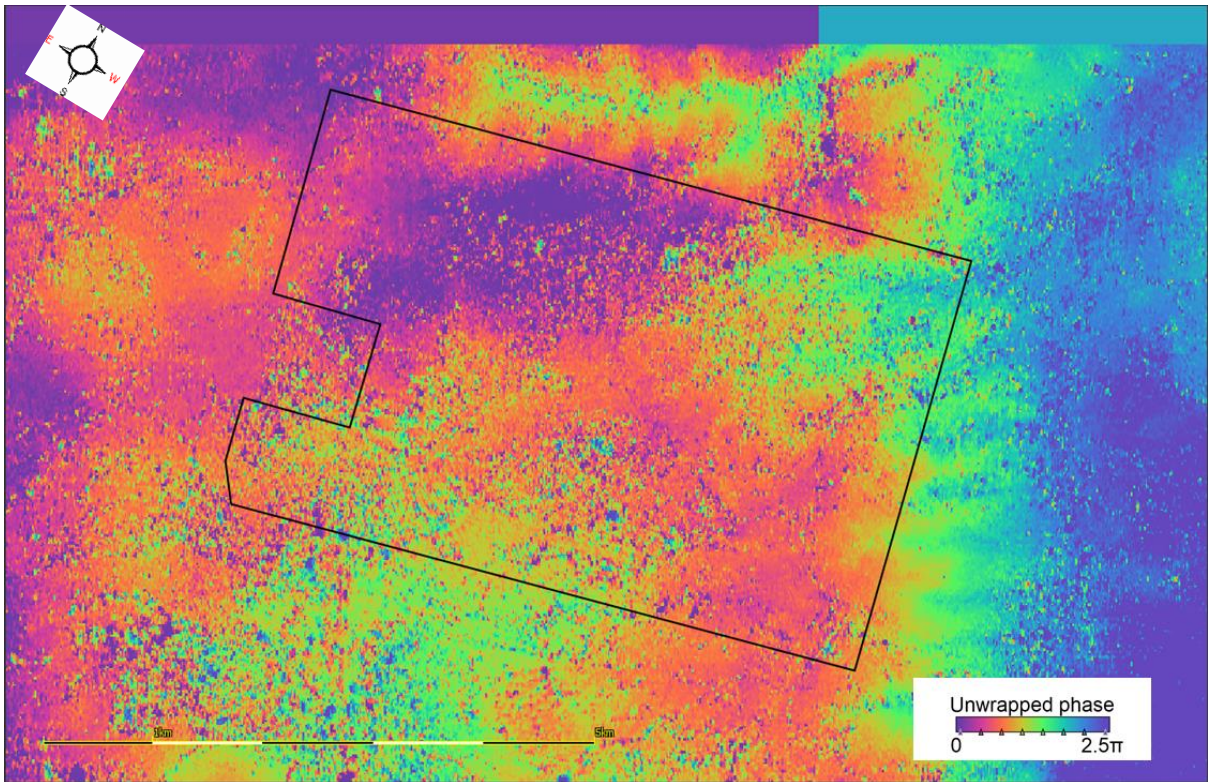




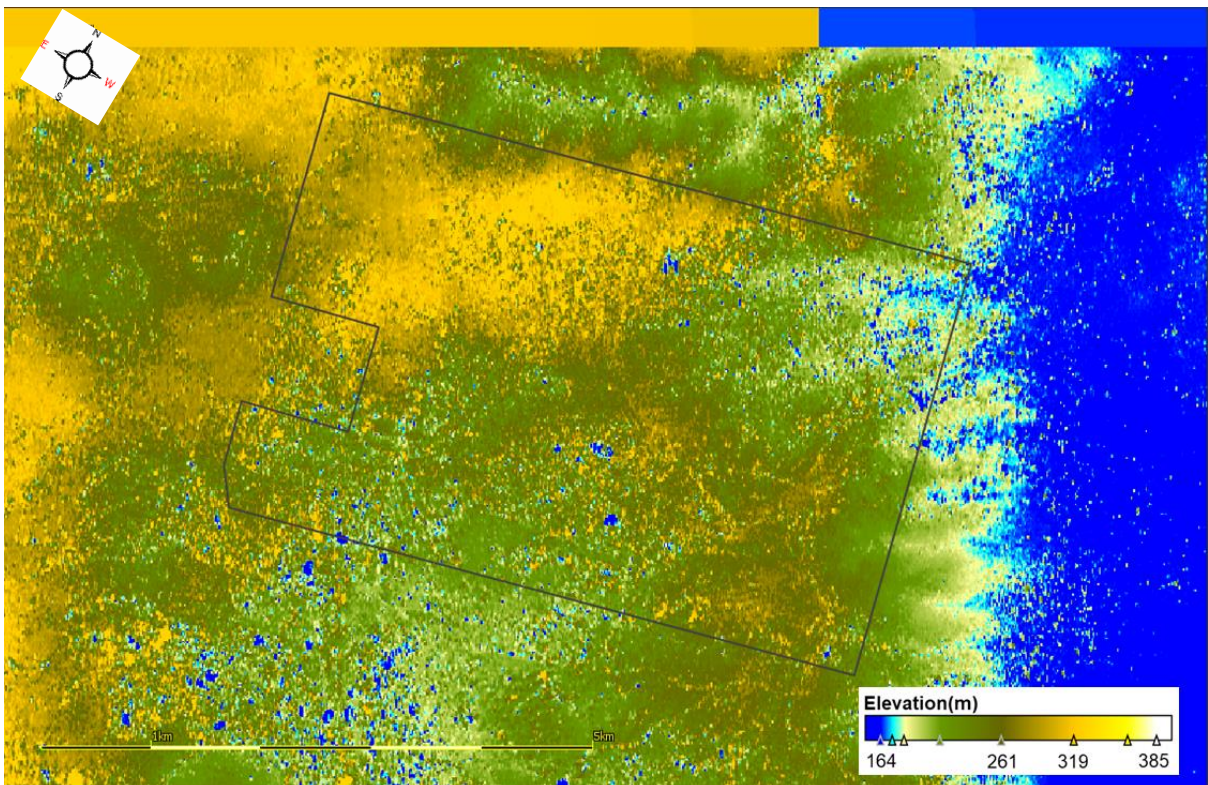
F



G



H



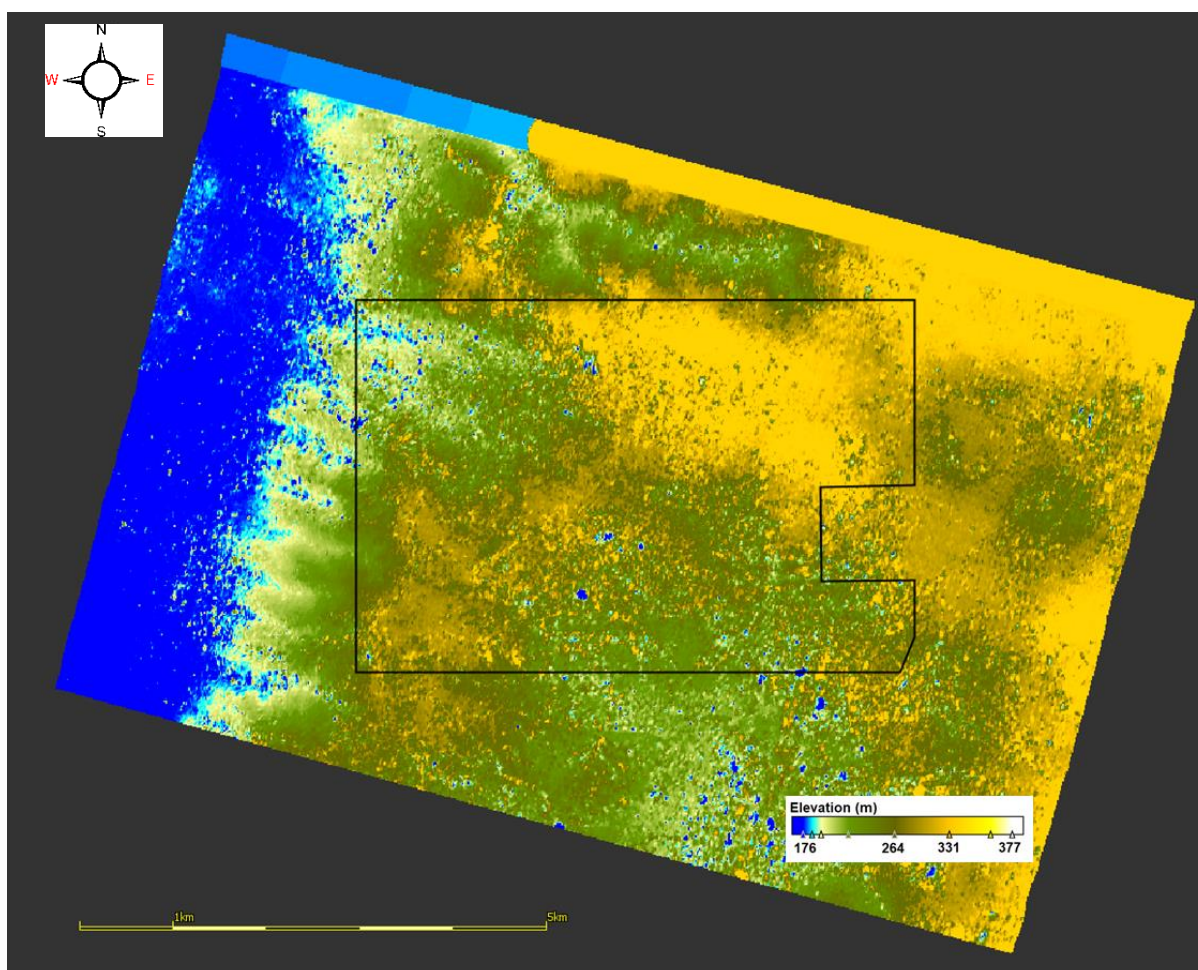


Figure 5.1: Visual representations of the intermediary outputs generated in ESA SNAP during the following steps carried out for the INSAR process: A) Back geocoding, B) interferogram generation, C) Tops Deburst, D) Multilooking, E) Goldstein phase filtering, F) Subset, G) Phase unwrapping, H) Phase to elevation, I) Radiometric terrain correction.

It has been observed that during the INSAR process, prior to multilooking and Goldstein filtering, the coregistered SAR scenes are compressed in the azimuth direction (Figure 5.1 (A),(B),(C) - SAR sensor flight path), but not in the range direction. Moreover, it was observed that the study site is in the near range of the SAR data and close to the beginning of a burst. Additionally, it was noted that the study area only covered one burst. The addition of the Goldstein phase filter significantly reduced the remaining noise after the multilooking step (Figure 5.1 (D),(E)). Furthermore, after the phase unwrapping step, the phase values in the area surrounding the study, after the subset process, jump from $-\pi$ to π , and then from 0 to 2.5π (Figure 5.1 (F),(G)). Moreover, after the phase-to-elevation process, the elevation range was initially between 154m and 385m, but changed to 178m and 377m after the radiometric terrain correction (Figure 5.1 (H),(I)). Lastly it was seen through all the steps until the radiometric correction that the East and West directions were flipped around.

Best filtering method for INSAR DSM generation

Table 5.2: Determination of the most accurate INSAR filtering methodology for each of the thinning areas

Thinning category	Filtering used in INSAR DSM generation	Mean elevation (m)	Standard deviation (m)	RMSE (m)	Percentage difference (%)
T0	Filtering 1	334.4	6.7	5.7	1.5
	Filtering 2	331.9	6.0	3.5	0.7
	Filtering 3	357.2	4.2	27.9	8.4
	Filtering 4	357.5	5.5	28.4	8.5
	Ground truth value	329.6	2.2	-	-
T1	Filtering 1	346.6	19.7	13.3	1.5
	Filtering 2	343.1	20.1	12.2	0.5
	Filtering 3	368.3	3.2	28.4	7.9
	Filtering 4	365.2	4.0	25.5	7.0
	Ground truth value	341.3	8.8	-	-
T2	Filtering 1	361.7	21.1	9.1	1.5
	Filtering 2	364.7	29.5	14.6	2.3
	Filtering 3	358.6	26.2	12.9	0.6
	Filtering 4	356.2	30.7	16.5	0.1
	Ground truth value	356.4	20.2	-	-
T3	Filtering 1	367.2	13.8	20.0	4.1
	Filtering 2	383.3	49.1	32.4	8.6
	Filtering 3	363.5	8.3	20.7	3.0
	Filtering 4	354.7	15.6	12.7	0.5
	Ground truth value	352.8	33.4	-	-

For both T0 and T1, it was observed that the most effective DSM generation method involved using only multilooking, without the Goldstein filtering coherence mask and band maths to eliminate areas of coherence below 0.3 from the interferogram before the unwrapping process. This method exhibited a percentage difference from the ground truth LIDAR DSM of 0.7% and 0.5% respectively, with an RMSE of 3.5m for T0 and 12.5m for T1.

Regarding the T2 and T3 areas, the most successful INSAR DSM generation method was the one that employed multilooking, Goldstein filtering mask, and band maths. This method resulted in a percentage difference from the LIDAR DSM of 0.1% for T2 and 0.5% for T3, with an associated RMSE of 16.5m for T2 and 12.7m for T3.

Moreover, it was observed that only under T0 conditions did the RMSE and percentage difference for the most effective INSAR DSM generation methods exhibit the lowest values overall, with a percentage difference of 0.7% and an RMSE of 3.5.

However, the INSAR DSM generation methodology using multilooking alone, without the Goldstein filtering mask and band maths, was found to be relatively effective across all management areas. Therefore, it was selected as the primary method for the required DSM adjustment and subsequent CHM generation.

INSAR DSM coherence analysis

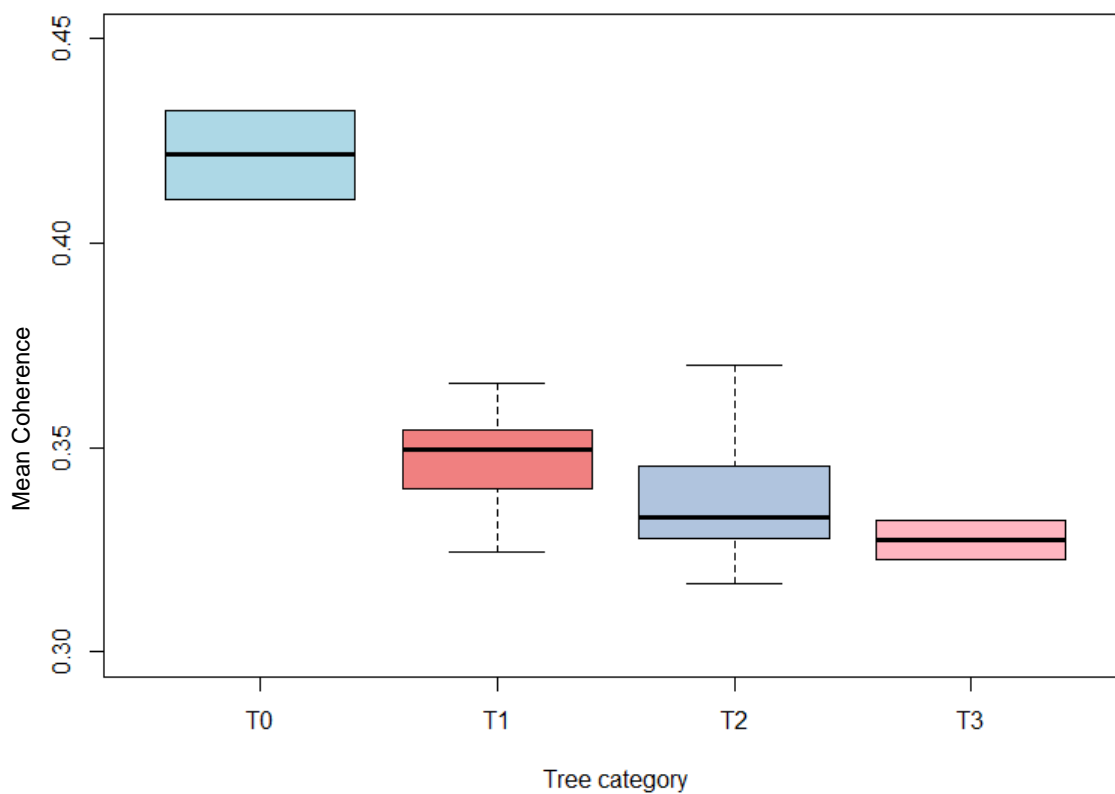


Figure 5.2: Boxplot showing the mean pixel coherence over the combined thinning areas

The coherence information from the two input Sentinel-1 images, obtained from the coregistered data on 31st January and 12th February, was extracted to the same investigation management areas using the chosen INSAR method. It was observed that the highest coherence value of 0.422, with a standard deviation of 0.015, was found in the T0 areas. Subsequently, the mean coherence values decreased from T1, T2, to T3, with values of 0.347, 0.338, and 0.327, respectively. The associated standard deviations for these values were 0.013, 0.017, and 0.007, respectively.

INSAR DSM intensity analysis

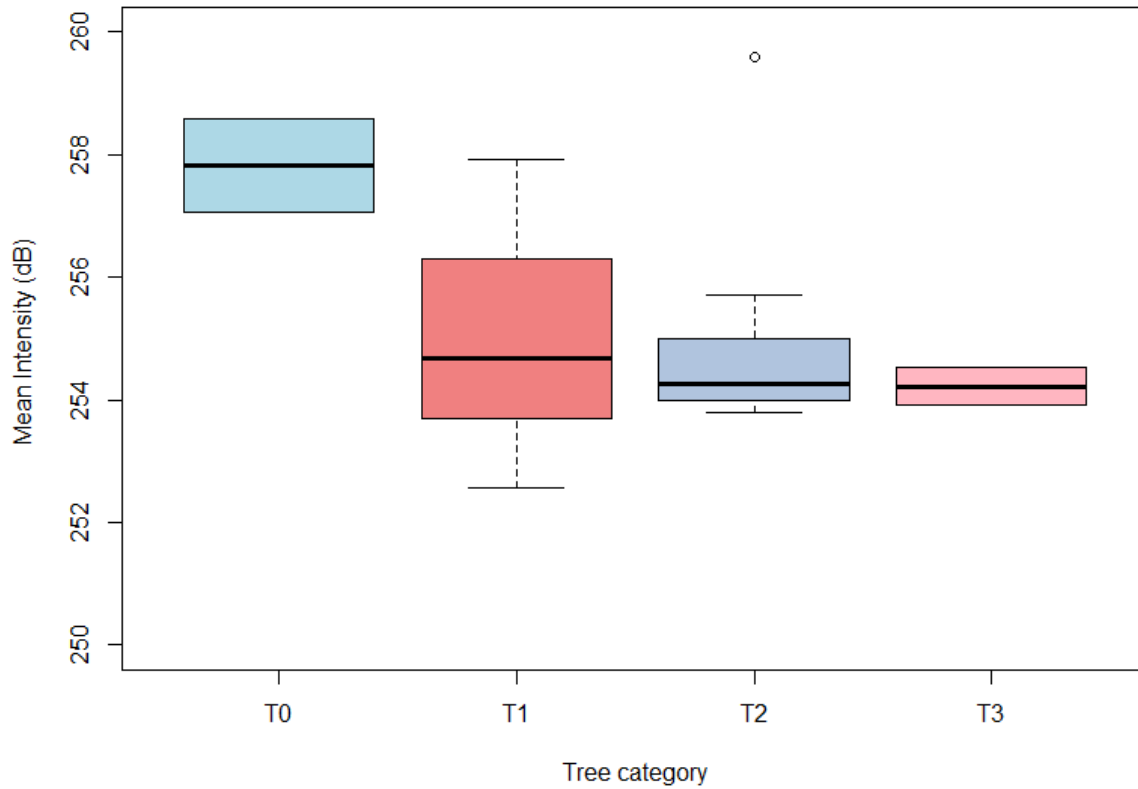


Figure 5.3: Boxplot showing the mean intensity of the Sentinel-1 C band signal over the combined thinning areas

The intensity (dB) of the returns from the coregistered images was extracted over the management areas. It was observed that the highest mean intensity was recorded over T0 areas, with a value of 257.8 dB and a standard deviation of 1.1 dB. The mean intensity of the returns over T1, T2, and T3 areas were approximately the same, measuring 255.0, 255.1, and 254.2, respectively. However, as shown in the Figure 5.3, the presence of an outlier in the T2 areas caused its corresponding mean value to be overestimated, as evident from its higher standard deviation of 2.08, compared to 1.86 for T1 and 0.43 for T3.

PSM and GCP locations

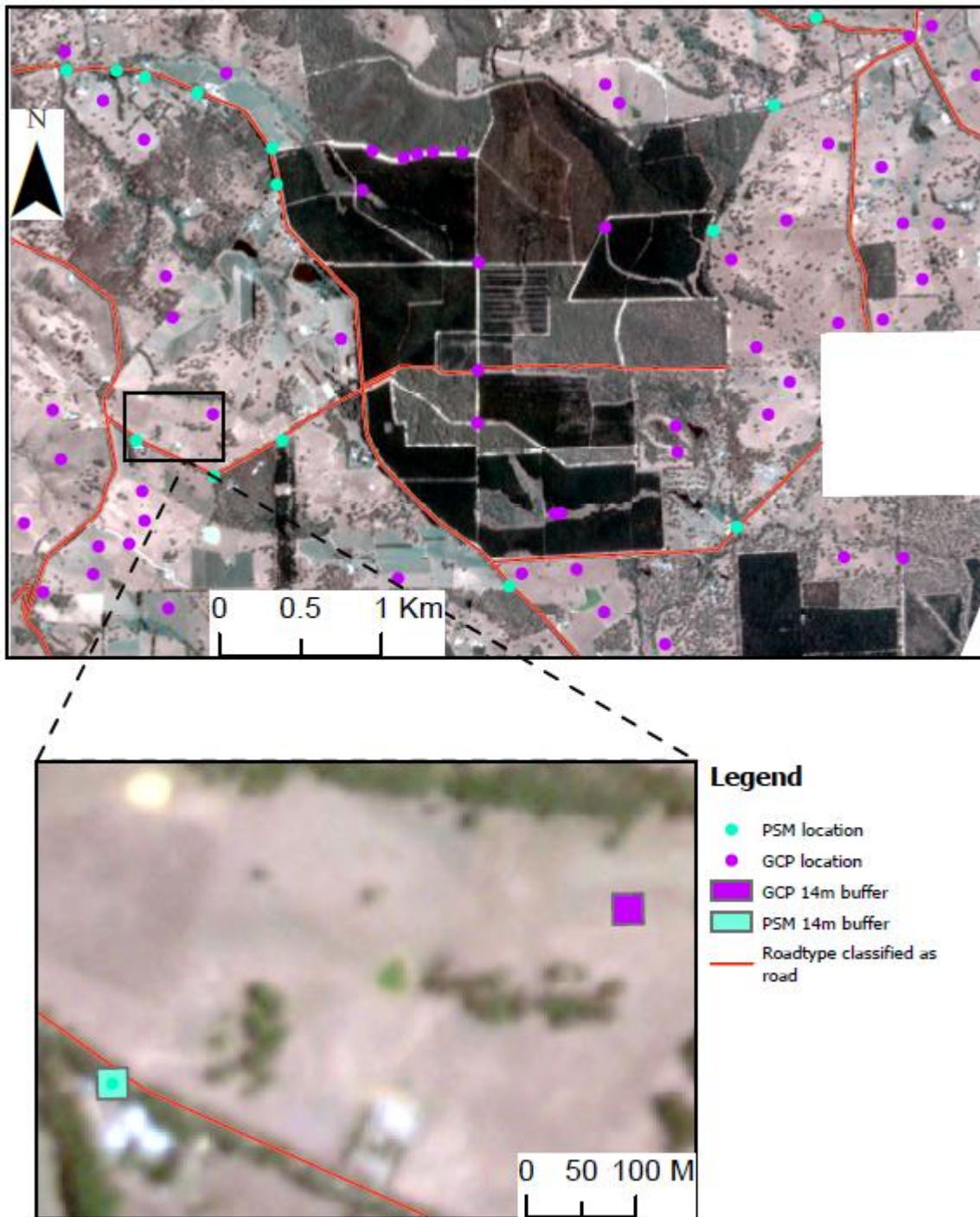


Figure 5.4: Location of the GCP and PSM used for the INSAR adjustment overlaid on ESRI Basemap, with a buffer of 14m representing the INSAR DSM resolution applied to an example of a PSM and GCP

From Figure 5.4, it is observed that the 52 GCPs are spread out across the study area, with most of them located in visually tree-free areas. However, it was found that the PSMs were concentrated mainly along roads, and none of the points were found within the thinned areas. Finally, when

examining the selected PSMs and GCPs with a 14m buffer that mimics the resolution of the INSAR DSM, it is evident that the PSMs along the roads within this buffer, there is the presence of vegetation, whereas within the GCP buffer only bare ground can be seen

PSM ADJUSTMENT

Table 5.3: Characteristics of the PSM adjusted INSAR DSM with its accuracy when compared to the LIDAR ground truth DSM

Thinning category	Filtering used in INSAR DSM generation	Mean elevation (m)	Standard deviation (m)	RMSE (m)	Percentage difference (%)
T0	Filtering 2	4.7	0.7	4.0	45.5
	Ground truth value	8.7	0.1	-	-
T1	Filtering 2	13.8	7.7	7.8	21.6
	Ground truth value	17.7	1.0	-	-
T2	Filtering 2	25.9	7.6	7.2	19.2
	Ground truth value	21.7	2.3	-	-
T3	Filtering 2	59.1	11.2	33.9	130.3
	Ground truth value	25.7	3.4	-	-

For each of the different thinning areas, the results of the INSAR DSM adjustment utilising PSM were found to have extremely high percentage difference values when compared with the lidar ground truth. Moreover, it was observed that the T3 regions had the highest combination of RMSE at 33.9m, as well as the highest percentage difference from the ground truth value at 130.3%. Similar values for RMSE and percentage difference were observed in T1 and T2 areas, with RMSEs of 7.8m and 7.2m, respectively, and associated percentage differences of 21.6% and 19.2%, respectively. However, in the T0 areas, it was observed to have the lowest RMSE at 4.0m, but with a large percentage difference from the ground truth at 45.5%.

GCP ADJUSTMENT

Table 5.4: Characteristics of the GCP adjusted INSAR DSM with its accuracy when compared to the LIDAR ground truth DSM

Thinning category	Filtering used in INSAR DSM generation	Mean elevation (m)	Standard deviation (m)	RMSE (m)	Percentage difference (%)
T0	Filtering 2	6.3	0.1	2.4	27.6
	Ground truth value	8.7	0.1	-	-
T1	Filtering 2	16.8	4.9	4.2	5.1
	Ground truth value	17.7	1.0	-	-
T2	Filtering 2	18.1	6.6	7.0	16.3
	Ground truth value	21.7	2.3	-	-
T3	Filtering 2	41.2	1.7	15.5	60.3
	Ground truth value	25.7	3.4	-	-

For each of the different thinning areas, the results of the INSAR DSM adjustment utilising PSM were found to have extremely high percentage difference values when compared with the LiDAR ground truth. Moreover, it was observed that the T3 regions had the highest combination of RMSE at 33.9 m, as well as the highest percentage difference from the ground truth value at 130.3%. Similar values for RMSE and percentage difference were observed in T1 and T2 areas, with RMSEs of 7.8 m and 7.2 m, respectively, and associated percentage differences of 21.6% and 19.2%, respectively. However, in the T0 areas, it was observed to have the lowest RMSE at 4.0 m, but with a large percentage difference from the ground truth at 45.5%.

Overall, from Table 5.3 and Table 5.4, it was seen that the GCP adjustment consistently had a lower RMSE and percentage difference value when compared to the PSM adjustment over the same area, and thus it was used to adjust the INSAR DSM before it being used for a CHM generation.

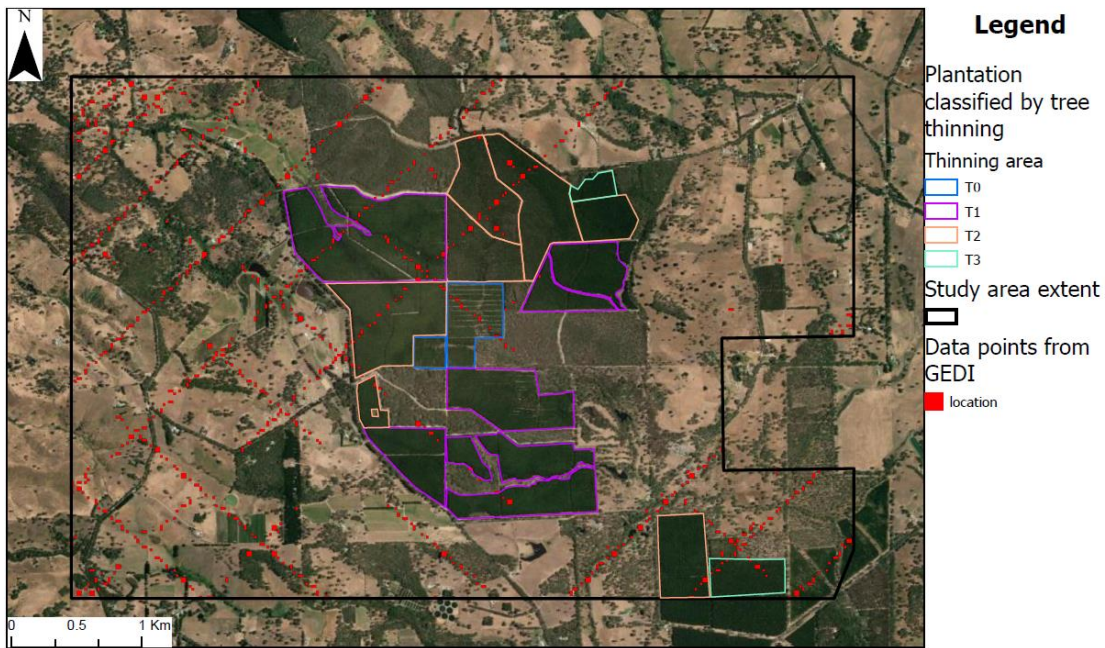


Figure 5.5: Distribution of the GEDI L2A datapoints overlaid on ESRI Basemap used as the dependent variable over the study area in an RF MLM

Most of the data points from GEDI are concentrated in the western section of the study area, but the majority are outside of the extraction polygons. Moreover, it was observed that the east and southeast directions suffer from an overall lack of data points (Figure 5.5).

RF MLM random forest variable importance

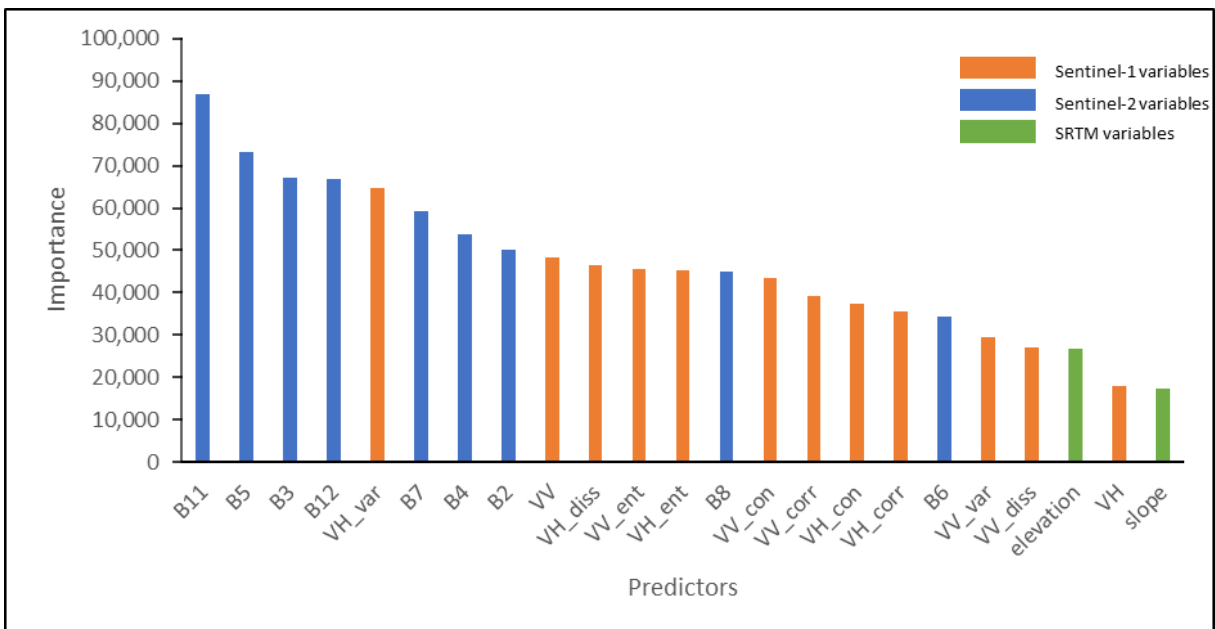


Figure 5.6: Importance of predictor variables used in the RF MLM interpolating across GEDI data points for the creation of a CHM

It was observed that the predictor with the highest overall importance in the RF algorithm was the B11 band from Sentinel-2, which corresponds to the shortwave infrared band with a wavelength of 1610nm. Most of the bands from Sentinel-2 tended to be ranked as more important. The predictor from Sentinel-1 that had the highest importance corresponded to the variance calculated based on the GLCM from the VH polarised data. Overall, the predictors derived from SRTM were among those with the lowest importance

RF MLM accuracy assessment

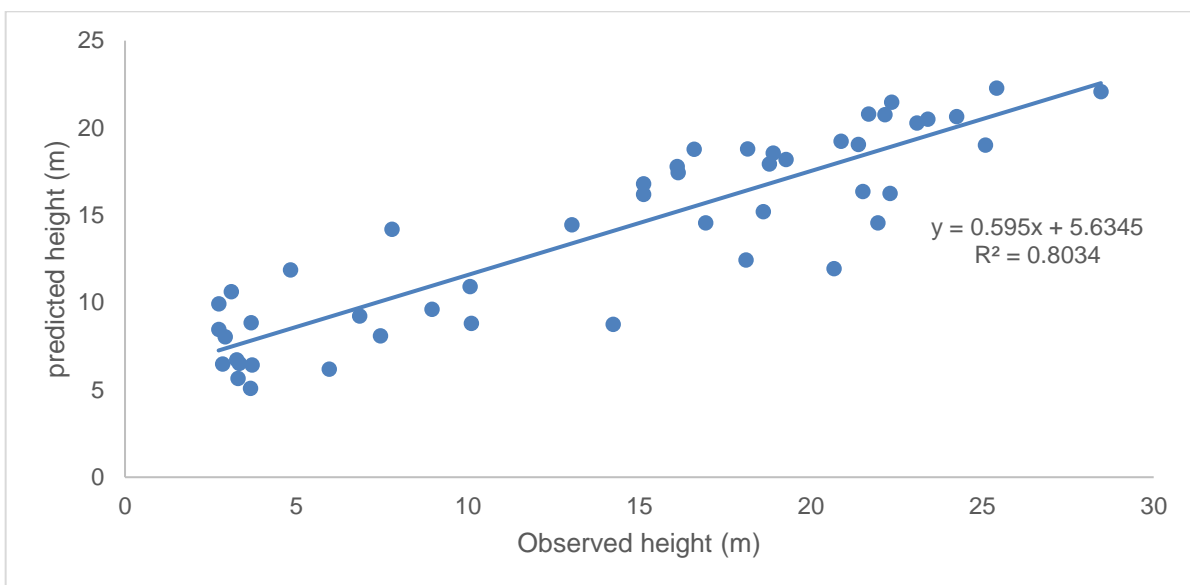


Figure 5.7: Comparison between the height predicted by the RF MLM (y-axis) with the GEDI rh98 height extracted at 70% of the 10000 points randomly distributed across the study area used as the training data(x-axis)

When comparing the GEDI CHM with the 70% of the 10,000 points assigned as training data over the entire study area, it was observed that it had an R^2 value of 0.803m and an associated RMSE of 3.99m. There was a constant overestimate of 0.595m in the predicted height.

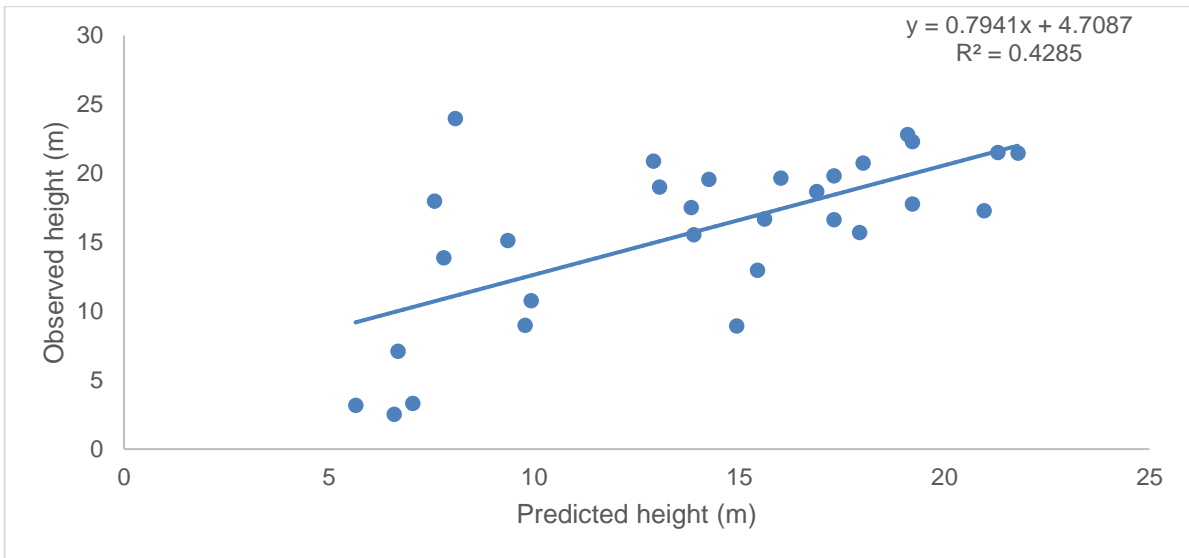
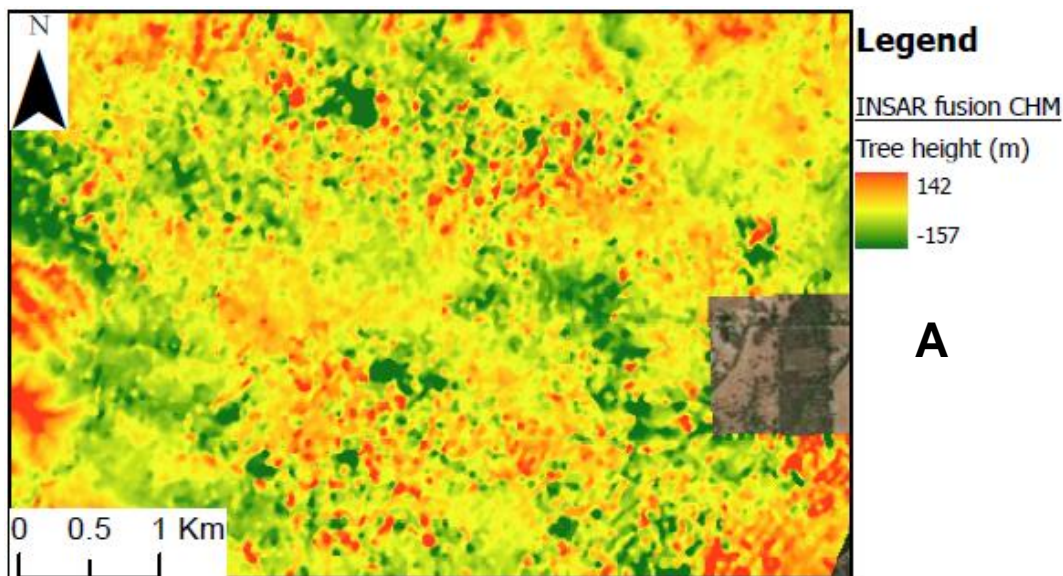


Figure 5.8: Comparison between the height predicted by the RF MLM(x-axis) compared with the rh98 height from GEDI extracted from the 30% of the 10000 points randomly distributed across the study area used as validation data (y-axis)

However, when comparing the observed height to the predicted height using the validation data, it was observed that the R^2 value dropped to 0.428m, while the overestimate in height increased to 0.794m compared to the training data. Moreover, the validation RMSE also exhibited a higher value of 4.94m overall.

INSAR fusion and RF MLM CHMs comparison with LIDAR CHM



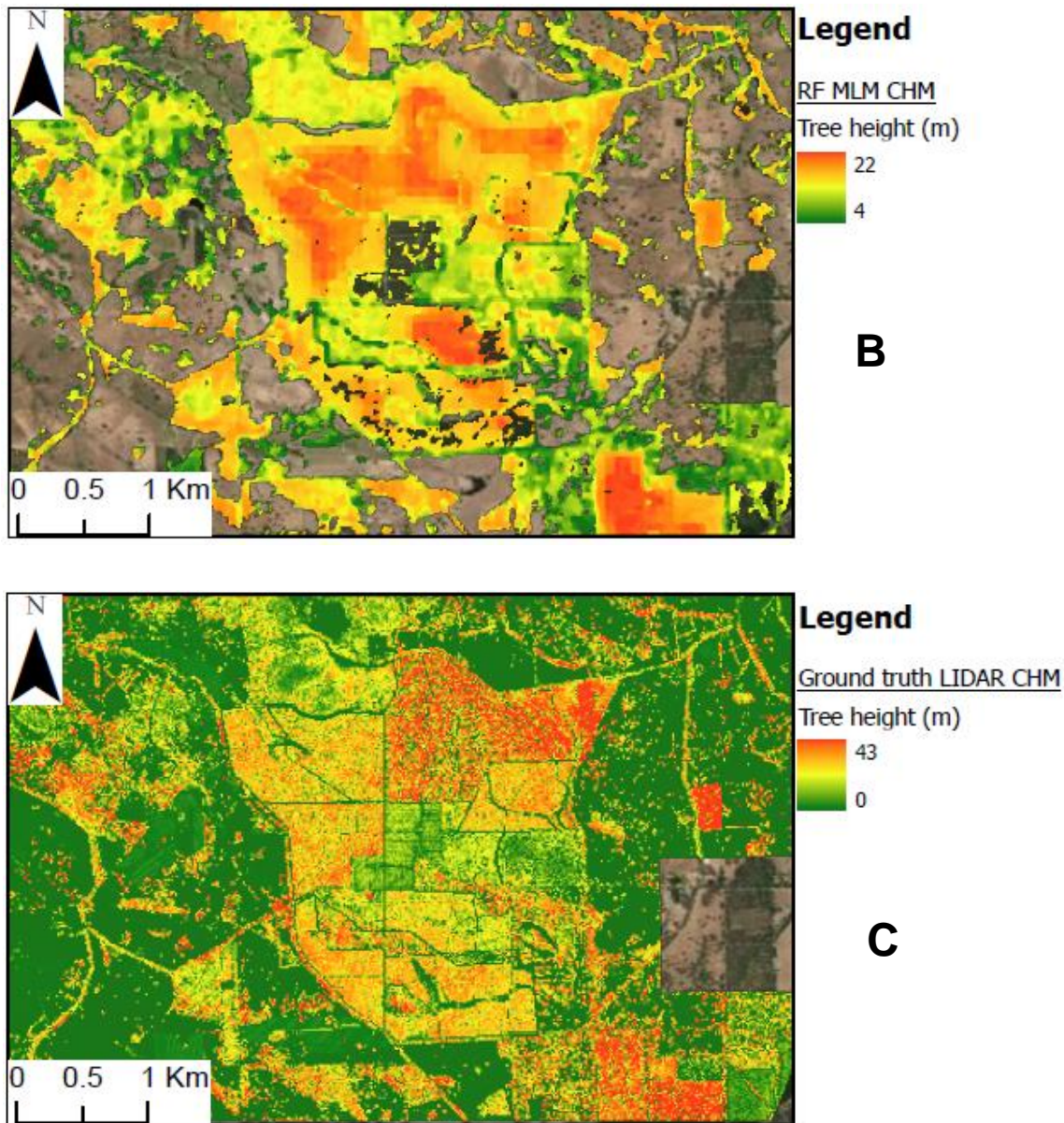


Figure 5.9: Comparison of the full height ranges (100th percentiles) of the INSAR fusion CHM (A), RF MLM CHM (B), and the LIDAR ground truth CHM (C) overlaid on ESRI Basemap.

It is evident that when comparing the INSAR CHM (Figure 5.9 (A)) with the ground truth and RF MLM CHM over the study area, no discernible pattern can be observed. Conversely, in the RF MLM CHM, a concentration of taller trees can be seen in the upper section of the study site, like the LIDAR CHM. However, it should be noted that the height values in the INSAR CHM vary greatly, ranging from -177m to 142m. As a result, no useful information can be extracted from it visually. On the contrary, the RF MLM CHM exhibits a more realistic height range across the entire study area, although it is approximately twice as low compared to the LIDAR CHM. Lastly, the RF MLM had certain areas within the study site that were masked, making it impossible to estimate the height. Visually, these areas mostly corresponded to the ground, but the roads shown in Figure 5.4 exhibited tree height estimates.

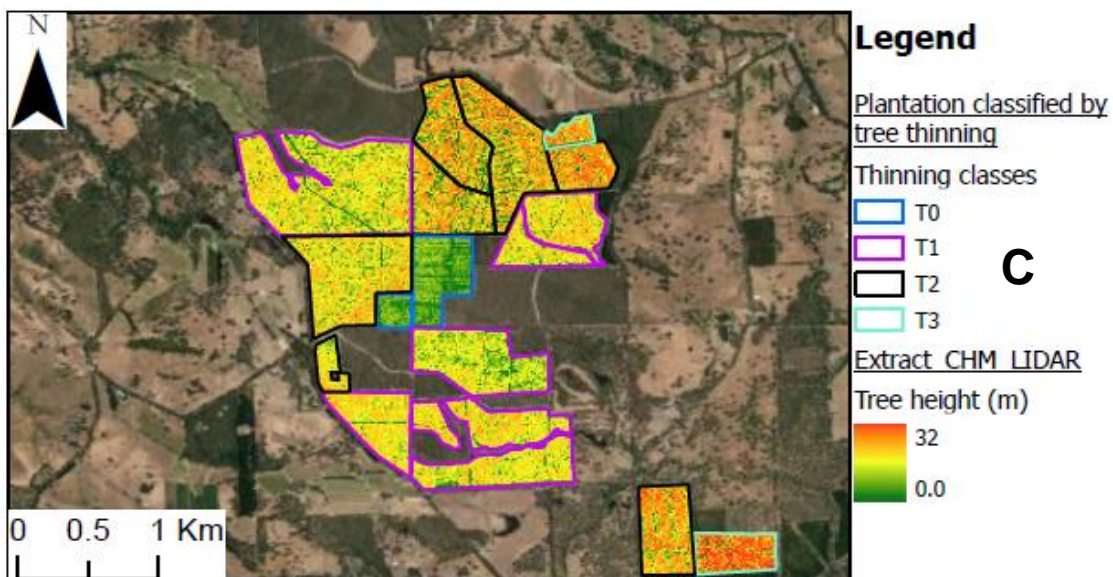
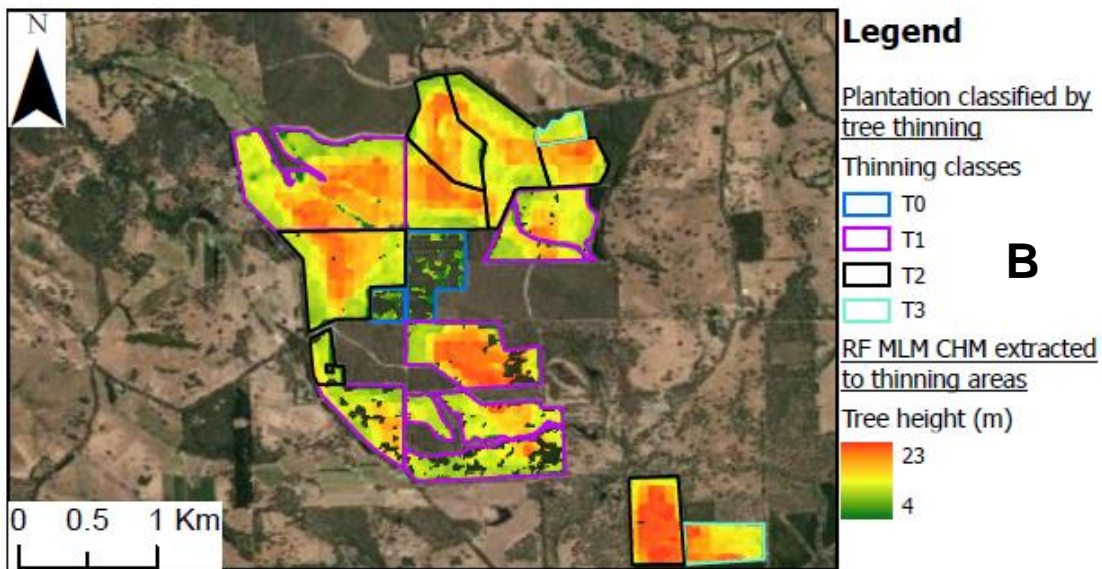
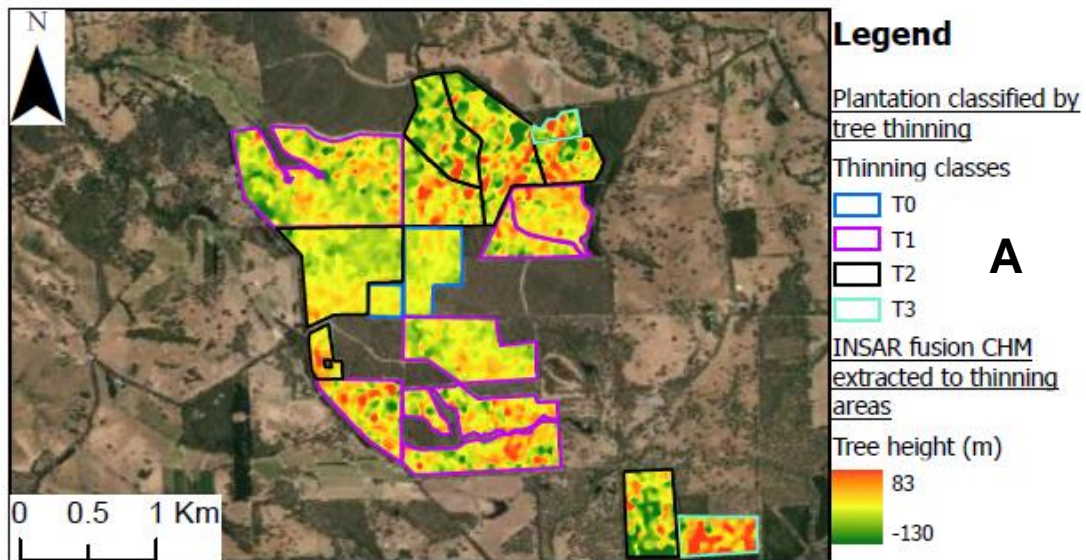


Figure 5.10: Comparison of the full height ranges (100th height percentiles) of INSAR fusion CHM (A), RF MLM CHM (B), and LIDAR ground truth CHM (C) that have been extracted in different thinning areas within the study area overlaid on ESRI Basemap.

A visual comparison between the RF MLM and the LIDAR ground truth CHM (Figure 5.10) suggests that the RF MLM has higher heights in the overall thinning areas, whereas the LIDAR CHM records the highest elevations in the southeast sections, specifically in the T3 areas. It is important to note that the peak height estimate of the RF MLM (23 m) is still lower than that of the LIDAR CHM (32 m). However, it is evident that the INSAR CHM provides widely inaccurate height estimates, ranging from a peak of 83 m to a lowest height of -130 m, which is not plausible. Furthermore, even when comparing the extracted thinning areas, no visual comparisons can be made with the LIDAR ground truth CHM.

Table 5.5: Assessing the accuracy of the 90th percentile height derived from INSAR fusion CHM and RF MLM CHM against the 90th percentile height of ground truth LIDAR CHM across different thinning classes. All CHMs are resampled to a 14m resolution

Thinning category	CHM source	Mean elevation (m)	Standard deviation (m)	RMSE (m)	Percentage difference (%)
T0	INSAR fusion	6.3	0.1	2.4	27.5
	RF MLM	16.3	0.16	7.7	88.9
	Ground truth LIDAR	8.7	0.09	-	-
T1	INSAR fusion	16.8	4.9	4.2	5.1
	RF MLM	19.2	1.1	2.3	8.8
	Ground truth LIDAR	17.7	1.0	-	-
T2	INSAR fusion	18.1	6.6	7.0	16.4
	RF MLM	20.6	1.1	1.9	5.1
	Ground truth LIDAR	21.7	2.3	-	0
T3	INSAR fusion	41.2	1.7	15.5	60.3
	RF MLM	21.0	0.7	5.0	18.1
	Ground truth LIDAR	25.7	3.4	-	-

Overall, the CHM obtained from the chosen INSAR methodology and the one derived from the RF MLM were compared to the LIDAR ground truth CHM, with all three resampled to 14m. It was observed that the highest percentage difference and associated RMSE from the INSAR methodology were observed in T3 conditions, with values of 60.3% and 15.4m, respectively. On the other hand, the lowest percentage difference was observed in T1 conditions, with a value of 5.1%. However, the lowest RMSE observed was not in T1, which had a value of 4.25m, but in T0 conditions, with 2.39m. It was observed, however, that under any other conditions, the percentage difference was at least 16.4% higher under T2 conditions, up to 60.3% in T3 conditions, with RMSE going up to 12.5m under those conditions.

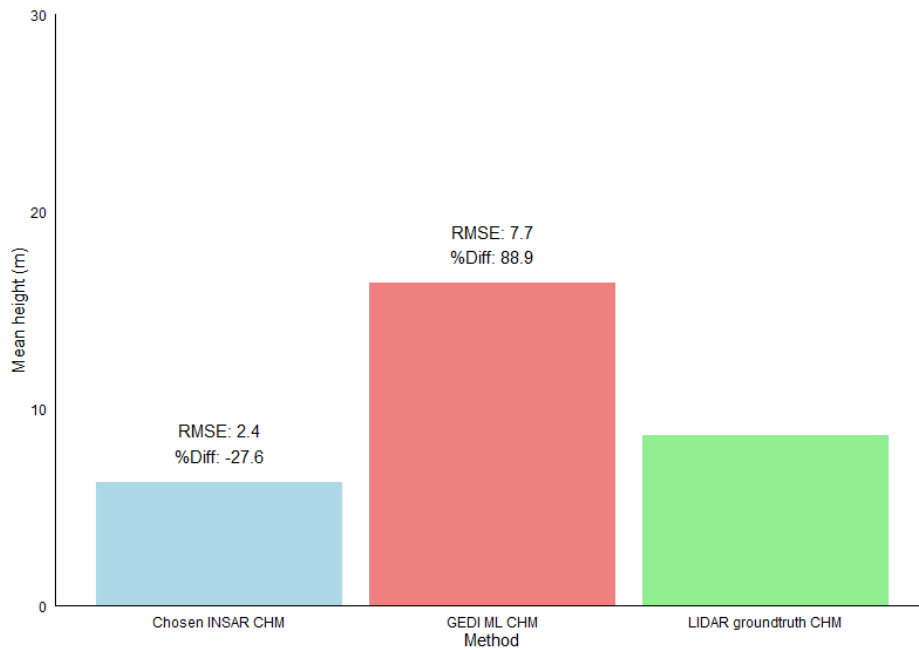


Figure 5.11: Visual comparison of the INSAR fusion CHM (Chosen INSAR CHM), RF MLM CHM (GEDI ML CHM) and the LIDAR ground truth CHM, with all CHMs resampled to 14m over T0 areas

Under T0 conditions, although neither INSAR nor RF MLM were capable of accurately estimating tree height due to the large percentage difference, it was observed that RF MLM was noticeably worse than INSAR.

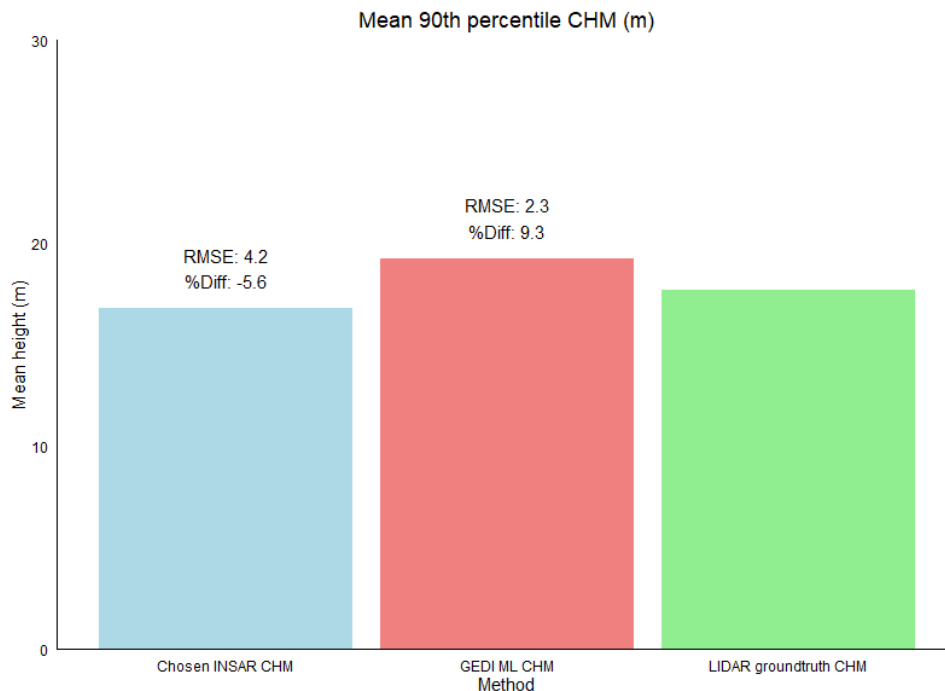


Figure 5.12: Visual comparison of the INSAR fusion CHM (Chosen INSAR CHM), RF MLM CHM (GEDI ML CHM) and the LIDAR ground truth CHM, with all CHMs resampled to 14m over T1 areas

Under T1 conditions, it was observed that INSAR underestimated the height by 5.6%, compared to RF MLM which overestimated the height by 9.3m. However, RF MLM had a lower RMSE of 2.3m.

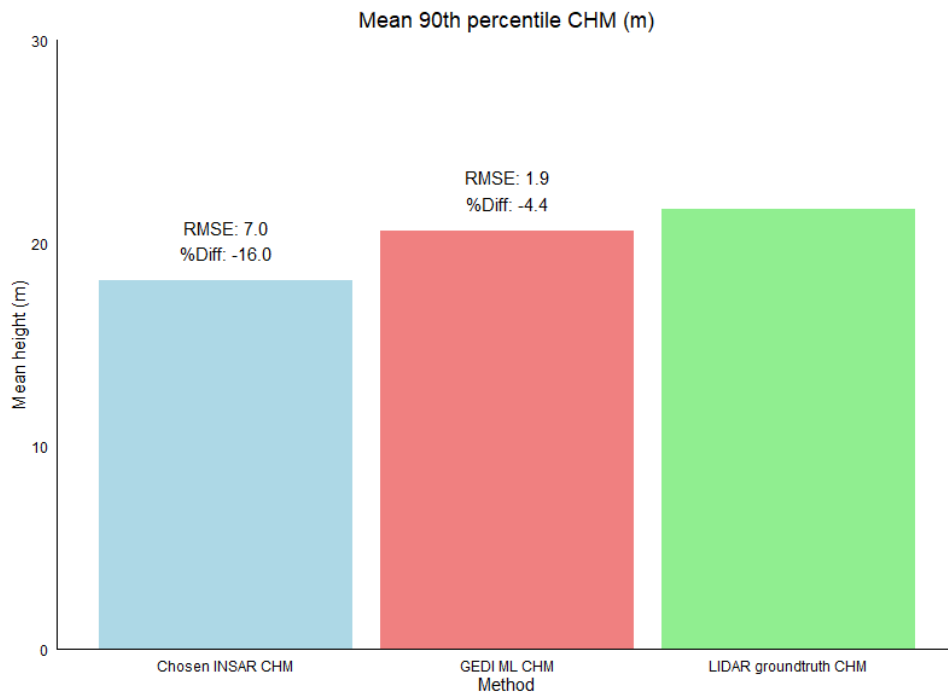


Figure 5.13: Visual comparison of the INSAR fusion CHM (Chosen INSAR CHM), RF MLM CHM (GEDI ML CHM) and the LIDAR ground truth CHM, with all CHMs resampled to 14m over T2 areas

Under T2 conditions, both INSAR and RF MLM underestimated the tree height, but RF MLM performed better due to the height underestimation and had an RMSE of only 4.4% and 1.9m, respectively.

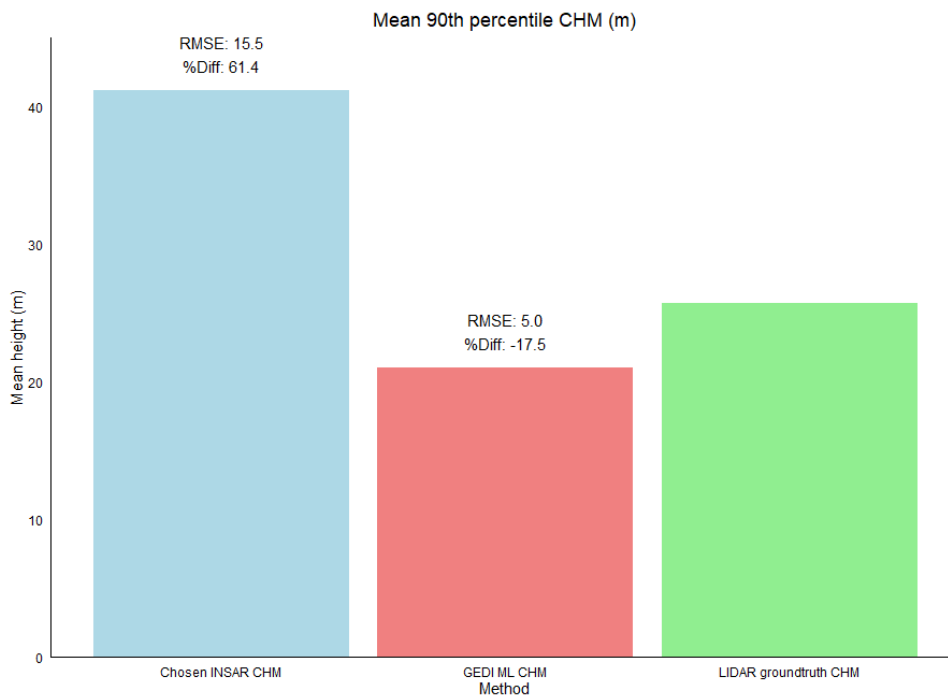


Figure 5.14: Visual comparison of the INSAR fusion CHM (Chosen INSAR CHM), RF MLM CHM (GEDI ML CHM) and the LIDAR ground truth CHM, with all CHMs resampled to 14m over T3 areas

Both INSAR and RF MLM were unable to accurately predict tree heights under T3 conditions. However, it was observed that INSAR was noticeably worse with the highest percentage difference and an RMSE of 61.4% and 15.5, respectively

CHAPTER 6 DISCUSSION

Overview of the INSAR DSM generation process

It was observed that two methods of generating the INSAR DSM proved to be effective over the management areas (Table 5.2). Filtering 2 was found to be the most suitable for T0 and T1 areas, while filtering 4 was deemed appropriate for T2 and T3 areas (Table 5.2). However, when comparing the overall accuracy of the INSAR DSM in relation to the ground truth LIDAR-based DSM, it became evident that the most appropriate INSAR DSM generation method varied depending on the investigated thinning area. This comparison was made against the LIDAR DSM resampled to 14m to match the resolution of the output INSAR DSM. However, when emphasising the significance of RMSE over the percentage difference, which is commonly used as a measure of vertical height accuracy and is easily interpretable in similar studies Braun (2021), it was determined that the overall most accurate INSAR generation method for most thinning areas was the one that employed filtering 2 in the process

Understanding the impact of the different filtering processes on the INSAR accuracy

It is well-known that the multilooking process helps reduce speckle noise in the interferogram by averaging SAR data within square pixels measuring 12.9m by 12.9m, but also decompresses the data in the azimuth direction as shown in Figure 5.1. This technique has also been observed to improve coherence estimation and, consequently, the overall phase fidelity of the interferogram, thereby minimising errors during unwrapping (Ab Rahman et al., 2023; Bouaraba et al., 2018). However, it was observed that the incorporation of both the Goldstein filtering coherence mask and the band maths mask did not effectively improve the accuracy of the generated INSAR DSM in most of the management areas, except for T3 (Table 5.2). This could be attributed to the fact that these two processes both aim to eliminate pixels with a coherence lower than 0.3 from the interferogram. The threshold value of coherence, suggested by Woodhouse (2017) and supported by Braun (2021) within the range of 0.1 to 0.4, was based on the assumption that areas with low coherence are affected by data noise and severe compromises in the phase information of these pixels, rendering them irrelevant. Consequently, it is necessary to remove these pixels before the phase unwrapping process to avoid errors that could result in an increase in the INSAR-derived height, especially when the coherence is below 0.2 (Braun, 2021; Woodhouse, 2017).

Nevertheless, it was observed that excluding pixels with low coherence prior to the phase unwrapping process in vegetated areas would lead to the formation of disconnected areas consisting only of pixels with higher coherence in the interferogram with no other pixels surrounding them. As a consequence, the MCF algorithm employed in this process would either

fail entirely or incur higher error rates (Braun, 2021). Moreover, this low coherence value thresholding is often valid over bare soil or urban areas, but not over mostly vegetated areas such as the ones in this study. This is the case as vegetated areas are themselves often associated with low to moderate coherence values owing to the instability of the top of the tree canopy structure between the two SAR image acquisitions (Singh et al., 2020). This lack of structural stability has been associated with an increase in volume scattering near the tree crown due to the movement of leaves and branches caused by wind effects (Askne et al., 1997). Additionally, the scattering phase centre of C band also occurs in the same wind-affected regions, resulting in much lower coherence over these vegetated areas (Singh et al., 2020).

The effect on coherence values has often been used to distinguish between agricultural and non-agricultural areas. A coherence value of 0.45 was used in Pandit et al. (2022), and it was found to be correlated with NDVI in the case of VV polarisations. Lower coherence values were associated with higher vegetation cover and were used to develop SAR-specific indices (Villarroya-Carpio et al., 2022). Cartus et al. (2021) found coherence values between 0.2 and 0.5 over forested areas containing a combination of pine and oaks, using similar VV C band data. However, a temporal baseline of 6 days was used, which is lower than the 12 days in this study. A longer temporal baseline would have resulted in even lower coherence values over the vegetated areas (Cartus et al., 2021; Monteith & Ulander, 2021).

Therefore, it can be concluded that masking pixels with a coherence of 0.3 over the study area would have removed data corresponding to the trees from the input interferogram. This explains why the most performant INSAR DSM methodology did not include any of the two coherence masking methods used, namely the Goldstein coherence mask and the band maths (Table 5.2).

Effect of coherence of INSAR DSM accuracy

It was observed that the highest coherence was observed over T0 areas and was drastically lower over other thinning areas, decreasing from T1, T2, and T3 (Figure 5.2). Moreover it was also seen that the higher the coherence in those areas the greater was the INSAR DSM generation accuracy as shown in Table 5.2 where the highest INSAR DSM accuracy was achieved using the INSAR DSM generation method that only involved multilooking and Goldstein phase filtering, without Goldstein phase filtering mask and band math, and was found in T0 areas. This accuracy decreased when moving from T1 to T2 and T3. The significantly higher initial coherence over T0 areas can be explained by several factors.

Effect of Pruning and Thinning on coherence

Two management practices are carried out throughout the life cycle of *Pinus radiata* plantations in our study site: a yearly pruning process between autumn and winter to mitigate fire risk and tree stress (ForestrySA, 2021; PIRSA, 2023), and the commercial thinning process carried out

approximately every 10 years. These processes, especially thinning, have been clearly identified in the study site, and the patches of trees associated with their level of thinning from T0, T1, T2, and T3, due to their differences in tree structure and biomass, will respond differently to the incoming C-band microwaves.

It is known that thinning of *Pinus radiata* contributes to an increase in the availability of space, airflow, temperature, and light penetration inside the existing canopy, which leads to an overall increase in tree diameter, as well as an increase in overall crown size (Fernández et al., 2017). Moreover, thinning also contributes to larger branches and an increase in tree taper, where the overall stem diameter will be substantially smaller than the higher the tree height (Burkhart et al., 2012; Fernández et al., 2017). Pruning under Chilean weather conditions for *Pinus radiata*, when carried out on young branches, was seen to decrease the overall tree growth rate and the photosynthetic ability of the tree, but the latter was compensated by additional growth of branch clusters (Pinkard & Beadle, 2000). When thinning and pruning were carried out together, it was observed that their individual effects described above occurred, leading to an overall higher tree diameter and increased foliage biomass when they were carried out on the 6th and 11th year of the *Pinus radiata* plantation (Fernández et al., 2017). However, when compared to the unmanaged sites that had not undergone either process by the 14th year, it was observed that the annual overall tree growth of the managed and unmanaged sites were similar, showing a level of canopy closure due to the crown repairing the foliage loss, but the managed trees retained their higher branch density (Fernández et al., 2017). If these effects are applied in the context of the Knott Hill and Mount Panorama study site, it can be assumed that between thinning and pruning cycles, the tree stem becomes wider, and more branch clusters develop. Then, the canopy recovers from the effects in preparation for the next thinning cycle, meaning that the more mature, taller trees that have undergone more thinning cycles will have more overall branches. Santoro et al. (2007) have shown that coherence is inversely proportional to the stem volume; it will decrease as the stem volume increases. Additionally, due to the limited penetration of C-band through the tree canopy, most of its scattering occurs at the crown and tree branches. Therefore, it can be assumed that more mature trees with a higher branch density will cause a higher scattering effect and potentially result in lower temporal coherence between the two SAR images over these patches. This assumption was supported by the findings of a box plot analysis (Figure 5.2) showing an overall decrease in coherence from 0.347 under T1, to 0.338 under T2, and finally, the lowest coherence observed at T3, at 0.327.

Effect of tree flexibility on coherence

Under open conditions, young *Pinus radiata* trees, within the first 10 years, produce wood that enables flexibility and avoids damage during the critical growth stage, particularly in windy conditions (Brennan et al., 2012). However, after this initial period, the increasing weight of the

trees results in corewood composition changes, leading to a stiffer structure that prevents self-buckling (Brennan et al., 2012). The reduction in flexibility over time increases the likelihood of branch movement with the wind in older trees, which may result in higher potential signal decorrelation between the pair of SAR images (Pinto et al., 2012). This phenomenon was observed by Castel et al. (2000), where even a 2-day difference in the acquisition of two C-band SAR images from the ERS-1 and ERS-2 missions, lower than the 12-day difference used in this study, resulted in lower coherence over mature pine trees compared to younger patches. The same phenomenon was observed in this study, with significantly higher coherence observed between the younger *Pinus radiata* trees in T0 areas that had not undergone any thinning process, compared to other patches that had undergone one to three thinning process (Figure 5.2)

Overall, the expectation that the higher the coherence of the area, the higher the confidence in the INSAR process, especially the phase unwrapping process, to be completed successfully, which may lead to a higher overall accuracy of the INSAR process, as highlighted in Li and Bethel (2008); Roc Roc (2019); T. Wang et al. (2009), was confirmed in this study.

Investigating the effect of SAR signal intensity on the INSAR accuracy.

The backscatter of SAR signal depends on the intrinsic characteristics of the data collection process, such as the wavelength band used, its associated polarisation, the incidence angle from the SAR sensor, and the characteristics of the surface being analysed, including the presence of vegetation, surface roughness, and dielectric constant (Alijani et al., 2021). In the C-band region, most of the backscatter over forested areas results from scattering mechanisms concentrated in the upper section of the canopy due to the smaller C-band wavelength compared to the size of the canopy (Hu et al., 2023). As described in Hu et al. (2023), volumetric scattering occurs on the canopy surface, double bound scattering mostly occurs on the upper canopy section and sometimes on larger and wider branches in the region, and finally, volumetric scattering also occurs in the upper canopy, mostly around small tree needles and secondary branches. It is known that the backscatter from VV polarised C-band data used in this study is correlated to forest structure, with its value decreasing as stem volume increases (Santoro et al., 2019). However, it is shown from Figure 5.3 that the mean intensity over the backscatter return over T1, T2, AND T3 hovers around the 255 dB mark, even considering their different management conditions and the increasing stem volume with more thinning processes and older tree patches. This lack of clear difference is explained by the overall weak sensitivity to the aforementioned relationship. In C VV-band data from Sentinel, the dynamic range of backscatter intensity observed over forested areas is only around 1 dB, as shown in Santoro et al. (2019). This explains the apparent lack of change in mean backscatter intensity over T1, T2, and T3. However, the mean intensity over T0 is higher than in the management areas, with a value of 257.8 dB, above the defined sensitivity threshold. This can be explained by the surface characteristics in T0 areas. It can be observed (Figure 5.5)

that in T0 areas, the overall area is not fully covered by *Pinus radiata*, and there is evidence of terrain manipulation, such as ridges in the soil present from ripping in past plantings. The increased backscatter observed might be due to a larger overall contribution of ground scattering caused by surface roughness, rather than scattering due to forest structure, due to the lower ratio of canopy coverage (Rüetschi et al., 2019). Moreover, it is observed that there is a greater sensitivity of VV polarised C-band data to surface roughness at incidence angles around 45°, which is the case in the far range of the Sentinel-1 data extracted over this study area, which reaches up to 45.3° in the far range (Alijani et al., 2021). Therefore, the higher sensitivity of VV polarised C-band data to rough surfaces, which cover a larger area in T0 regions, can explain the higher backscatter intensity observed, well above the expected sensitivity.

Comparison between the GCP and PSM INSAR DSM adjustment methods

It was observed that regardless of the source of the elevation data used for PSM and GCP adjustment, the interpolation done with the Empirical Bayesian kriging method (EBK) resulted in the lowest overall RMSE. This finding aligns with the principle behind EBK, which modifies the input data distribution to closely resemble a Gaussian distribution by eliminating local trends in the data. This allows for predictions of elevations to be made using semi-variogram models created with input data points near the prediction location (Krivoruchko, 2012). Furthermore, EBK is known to be more accurate than other kriging methods, even with smaller datasets, as observed in the case of GCP and PSM in this study (Krivoruchko, 2012).

Although both adjusted methods employed the same empirical Bayesian Kriging (EBK) interpolation method to generate the adjustment layer based on the reference height derived from the ground truth LIDAR for GCP and the measured elevation at the PSM, it was observed that the GCP points were more evenly distributed throughout the study area and were more abundant, with 52 GCPs compared to 14 PSMs (Figure 5.4). This may have contributed to the overall better performance of EBK interpolation (Krivoruchko & Gribov, 2019; Sunila & Kollo, 2007).

Moreover, the methodology used to create the GCPs may have also contributed to the higher effectiveness of the GCP DSM adjustment. When creating a GCP, visual examination of the 3m planet imagery ensured that there was no tall vegetation within a 14m buffer of the GCP. Therefore, the elevation extracted from the LIDAR DSM at that pixel would represent the ground elevation in the INSAR DSM, as the latter had a lower spatial resolution of 15m than the buffer size. On the other hand, it was observed that the 14 PSMs used in the PSM adjustment method were mostly located on roads passing through Knott Hill and Mount Panorama, where trees were present on both sides of the road within less than 14m (Figure 5.4). Consequently, the corresponding location of the PSM in the INSAR DSM pixel would represent a mixture of ground

and forest height information, resulting in an overestimated elevation for this pixel (D. Wang et al., 2009). As a result, the adjustment layer generated by subtracting the INSAR DSM height from the PSM height would be underestimated, leading to an overestimated overall Canopy Height Model (CHM) derived by subtracting the adjusted INSAR DSM from the ground truth 14m LIDAR DTM. This phenomenon is consistently reflected in the higher RMSE when comparing the CHM to the ground truth CHM, particularly in the investigated polygon using the GCP adjustment.

Limitations of INSAR based on climatic conditions and possible improvements

Rainfall

As rainfall would increase the overall soil moisture present in the ground, and that wet materials has a dielectric constant approximately 10 times higher than that of dry natural materials, the overall dielectric constant would be expected to decrease, reducing the SAR signal penetration ability (Cagnina et al., 2023; Periasamy & Ravi, 2020). Furthermore, the higher dielectric constant may cause an increase in overall backscatter, particularly in coniferous forests similar to the study site, due to the high rainwater storage potential of needle leaves (Cagnina et al., 2023). Overall, this higher backscatter may result in increased volume scattering effects, which can further cause decorrelation over vegetation, especially if one of the SAR images was taken in significantly drier conditions than the other (Beale et al., 2019; Gururaj et al., 2021; Periasamy & Ravi, 2020)

This was the reason why originally a better selection of INSAR pairs collected on the 19th and 31st of January was rejected due to rainfall prior to the SAR image capture, and when analysed resulted in large errors in the DSM generated.

High temperatures

High temperatures can also affect the overall water vapour content in the soil and the leaves, resulting in further degradation in the phase quality information derived from the INSAR data (Braun, 2021).

It was observed that at the time of collection of the first and second image used in the INSAR process, the temperature at the closest weather station in Kuipo was 27 degrees Celsius and 21 degrees Celsius, respectively, when the average minimum temperature for the two months was around 14 degrees Celsius (WillyWeather, 2023). Moreover, the time of the coldest temperatures differed on those days, between 3 to 4 am on the 31st of January image and between 6 am to 7 am on the 12th of February image. Finding a SAR acquisition time where both images are taken at the lowest temperatures will be difficult. Moreover, obtaining the times where the temperatures are the lowest over the forest is difficult due to the canopy, which absorbs and emits solar radiation, acting as a buffer against the daily temperature swings (Aussenac, 2000). Therefore, upon the

selection of the pair of SAR images, it must be ensured that they are night-time captures, and the temperatures recorded at the closest weather station or ideally a temperature probes that might be present in the forests are at their lowest, or at least, the temperature difference when the different SAR image acquisitions are taken is at their minimum. Moreover, as shown by Askne et al. (1997)

Windy conditions

Windy conditions might cause an increased loss in coherence between the two SAR images. Therefore, using SAR image acquisitions with the lowest wind speeds would be encouraged, especially considering the high temporal decorrelation that occurs as currently SAR images suitable for INSAR are acquired at least 12 days apart.

Effect of low resolution DEMs on the INSAR process and possible improvements

Throughout the INSAR process, the 30m SRTM DEM was used, which was derived from INSAR based on the C and X band antennas on board the space shuttle Endeavour in 2000 (Rabus et al., 2003). However, this generated DEM often exhibited poor accuracies over forested areas due to the SAR band data's limited ability to penetrate canopy cover. Consequently, the underlying subcanopy terrain information may not accurately reflect the actual conditions (Li et al., 2022; Rabus et al., 2003). Correcting the SRTM DEM over the forested areas in this study could greatly enhance the accuracy of the INSAR-generated DSM following the methodology outlined in Li et al. (2022)

Limitations in the INSAR process due to SNAP

It was observed that the phase-to-elevation tool used in SNAP during the INSAR process only used the SRTM elevation as a reference. Tampubolon et al. (2022) have shown that incorporating GCPs generated from GNSS measurements along with the automatically generated reference points from SRTM significantly improves the accuracy of the generated DSM. Additionally, the radiometric terrain correction tool does not support the use of the most up to date GDA 2020 and associated AHD2020 coordinate system, limiting it to GDA 1994. As a result, it may be necessary to explore and use alternative software for these critical processes instead of ESA SNAP.

INSAR limitations over vegetated areas

Due to the nature of the INSAR process, the generated DSM could not be directly evaluated against the ground truth LIDAR DSM using dispersed reference points throughout the areas. Instead, its accuracy was assessed using extraction polygons, specifically those based on the thinning process. Since the INSAR process was conducted over vegetation areas with inherently low coherence, elevation values had to be averaged over large areas to reduce the noise

associated with low coherence.

The presence of this noise led to the extraction of the 90th percentile height, which, upon investigation in this study, was found to be the most appropriate percentile for removing most error points while retaining important height metrics. Zandbergen (2011) highlights in their guidelines that elevation data may not always be normally distributed, and the use of percentiles is recommended in such cases. Furthermore, the authors emphasised that the use of RMSE statistics may not be accurate in this scenario. Therefore, based on these circumstances, the accuracy of the DSM could not be independently evaluated without utilising these extraction polygons, and an RMSE value could not be generated or easily compared with results from previous studies

RF MLM Overview

Selection of input variables

The initial objective was to compare the importance of these variables with other recent publications that used the same input data sources: Sentinel-1, Sentinel-2, and SRTM in a GEDI-based RF ML. However, it should be noted that there was no consensus on the types of variables to be used. For the Sentinel-1 variables, studies such as Guo et al. (2023) only used the VV and VH backscatter, while Kacic et al. (2023) and Tsutsumida et al. (2023) also incorporated percentile metrics based on the input VV and VH backscatter values, although with different percentile breakdowns. This was done to minimise the influence of temporal effects caused by atmospheric conditions and gaps in the time series when the Sentinel-1 data was collected. Chere et al. (2023) took a different approach and instead of the percentiles, included a combination of radar-based indices that utilise the input VV and VH backscatter as source data, with the latter also being included. Overall, none of the mentioned papers used texture variables such as contrast and entropy derived from the input VV and VH Sentinel-1 data used in this study.

Upon reviewing the same four publications which were selected as they represented the current methodology used to estimate canopy height using SAR data in MLMs, it was observed that none of them used only the raw reflectance bands of Sentinel-2 as input variables in the RF MLM, as was done in this study. Kacic et al. (2023) and Tsutsumida et al. (2023) agreed once again on including percentile metrics, though with different breakdowns based on the raw reflectance bands of Sentinel-2. Chere et al. (2023) chose to include a wide range of vegetation indices in addition to the raw Sentinel-2 spectral bands in their MLM. However, Guo et al. (2023) decided against adding the raw spectral data and instead included only derived vegetation indices and some biophysical variables. As for the DEM-based variables, Guo et al. (2023) and Chere et al. (2023) included aspect in their MLM, in addition to the derived slope and elevation used in this study. However, Kacic et al. (2023) and Tsutsumida et al. (2023) used only the DEM elevation.

Comparison of Variable Importance

It was observed that based on the variable importance (Figure 5.6), that a pattern can be observed, with variables based on Sentinel-2 ranked among the highest, followed by the Sentinel-1 based ones, and finally the SRTM based ones seen with the lowest overall importance. Overall, the trend of several Sentinel-2 based variables being ranked among the variables with the highest importance, as observed in this study, was also seen in the studies by Guo et al. (2023), Kacic et al. (2023) and Chere et al. (2023). The high importance of the B3 (visible green) and red-edge (B5) bands is not surprising, as they have been identified as having significant relevance to vegetation height compared to the other Sentinel-2 bands (Perez et al., 2022). Moreover, the inclusion of the

SWIR bands B11 and B12 among the ones with the highest importance aligns with results from previous models used to predict tree height, as highlighted in Perez et al. (2022).

In this study, the lowest importance of the topographic variables derived from the SRTM DEM does not follow the trend in past studies, where they were often among the strongest predictor variables for estimating forest height (Ghosh et al., 2022; Liu et al., 2019). Furthermore, RF MLM used by Chere et al. (2023), slope and elevation variables were ranked with the highest overall importance, and in the study by Kacic et al. (2023), elevation was the fifth most important variable. Moreover, a similar position in the importance of topographic variables as seen in this study was observed in Guo et al. (2023), where elevation and slope had an importance of only 0.5% and 0.18% respectively. It has been noted that this study and Guo et al. (2023) were conducted in relatively small study areas of 23 km² and 1400 km² respectively, compared to the ones in Chere et al. (2023) and Kacic et al. (2023), where the analysed zone covered approximately 6000 km² and 100000 km² respectively. Thus, it is possible that the variations in topographic variables were not significant enough over the smaller areas, leading to their weaker predictive value. Finally, overall, it was observed that in most cases, the Sentinel-1 variables were ranked lower than the ones based on Sentinel-2. This can be explained by the limited ability of the C band from Sentinel-1 to fully penetrate the tree canopy, resulting in lower signal strength and a higher chance of complex structures on the forest floor causing interference (Chere et al., 2023; Ghosh et al., 2022; Torres de Almeida et al., 2022). Additionally, when Sentinel-2 and Sentinel-1 data are used together, the latter often introduces noise in the MLM. This noise could be due to pre-processing of the initial Sentinel-1 data to generate backscatter values for both VV and VH polarizations or due to a mismatch in data acquisition days between the two sources. In the case of Sentinel-1 data collection being carried out under weather conditions such as rain, which greatly affects C band signal scattering and resulting correlation which was not filtered out before being used as input in the RF MLM therefore potentially inducing noise in the data (Braun, 2021; N. Zhang et al., 2022; Zhou et al., 2009).

However, in this study, VV backscatter was found to be more important in the RF model than VH (Figure 5.6) which was contrary to similar studies (Chere et al., 2023; Guo et al., 2023; Kacic et al., 2023). This discrepancy might be explained by the fact that the mentioned studies were conducted in environments containing both deciduous and coniferous trees, whereas the area analysed in this study only contained the coniferous *Pinus radiata*. VH polarisation is greatly influenced by foliage type, with broad leaves from deciduous trees resulting in higher backscattering and therefore higher predictor strength under those conditions compared to the needle leaves of coniferous trees, which have lower influence on VH backscatter and thus lower predictor strength as observed in this study (Ling et al., 2022). However, it was observed that VH variance was ranked as the most important Sentinel-1-based variable, and overall VH-derived variables performed better as

predictors compared to VV-based ones. This can be explained, as shown in Ling et al. (2022), by the fact that for coniferous forests, like the one in our study site, VH backscatter is negatively proportional to the length of needle-like leaves. This may explain why VH-derived texture variables may be stronger predictors than VV-based ones. Moreover, after the thinning process is carried out, debris is often left on the ground surface, increasing its interaction with the VH-polarised signal, which can now interact with the ground surface. As VH-based variables are more sensitive than VV-based texture variables to volume scattering on the surface, they will be better predictors in the RF ML.

RF MLM Performance

It was observed that the RF MLM used in this study had an R^2 value of 0.43 and an RMSE of 4.94m. This R^2 value was higher than that of Guo et al. (2023), who also used RF along with other MLM but reported an R^2 value of only 0.21. However, their RMSE was lower at 3.07m. Nonetheless, compared to similar studies utilising an RF MLM approach, it was found that this study performed relatively worse in terms of R^2 values. Kacic et al. (2023) and Tsutsumida et al. (2023) reported R^2 values of 0.64 and 0.63, respectively, although their RMSE values were higher at 6.6m and 5.8m. This suggests that the inclusion of percentiles of Sentinel-2 bands contributed to an increase in prediction accuracy, while the inclusion of percentiles of S1 data potentially introduced a large cumulative input noise due to the pre-processing required for Sentinel-1 data N. Zhang et al. (2022). Lastly, this study was significantly outperformed by Chere et al. (2023) whose RF MLM achieved an R^2 value of 0.86 and an associated RMSE of 3.63m.

Possible Improvements

These results indicate that instead of adding percentiles generated from the input data, the inclusion of a range of vegetation indices based on the raw Sentinel-2 bands, which were shown to have high importance in the MLM, would greatly improve the model's ability to predict tree canopy height. Moreover, the radar-based indices used in Chere et al. (2023) are derived from ratios or simple addition and averages of the VV and VH backscatter is a plausible way to increase the S1-based data in the model due to their essence having lower cumulative noise that would arise when using raw backscatter data and percentiles instead. One major difference between Chere et al. (2023) and this study was the incorporation of variable selection methods. In this case, the Pearson's correlation coefficient was evaluated, and variables with a coefficient higher than 0.95, indicating strong dependence, were removed as they were considered highly correlated and thus redundant. Additionally, variable selection methods can prevent the inclusion of data in the MLM that may not be reliable or have high associated input noise, thus potentially improving the overall predictive power and accuracy of the model (Chere et al., 2023; Geremew et al., 2023; Torres de Almeida et al., 2022). Moreover, as shown in Z. Wu et al. (2023), the addition of climate

information, such as temperature and precipitation, as input variables to the RF model demonstrated can result in an increase in the RF MLM predictive strength. These variables were used after a feature selection process as inputs in the MLM. The study also suggested incorporating PALSAR-2 data, based on L band SAR data, which has a higher tree canopy penetration ability and associated radar-based indices in addition to Sentinel-1, Sentinel-2, and SRTM-based variables. This addition was proposed to enhance the MLM capability to predict canopy height based on input GEDI data, despite the PALSAR dataset existing in a yearly mosaic format in the GEE platform, which was used in this study (Ghosh et al., 2022). One major limitation that may explain the lower performance of the GEDI-based RF model compared to other similar studies is the overall lack of GEDI data (over the forested part of the study area (Figure 5.4). This, coupled with the coarse resolution of the input Sentinel-1, Sentinel-2, and SRTM DEM, which was resampled to or originally at 30m to ensure input data homogeneity over a small study area of only 20 km², could contribute to the lower performance. One solution to overcome the scarcity of GEDI data was suggested by Z. Wu et al. (2023). They proposed using existing ICESAT-2 data, which was not initially favoured in the MLM due to its unfavourable sampling strategy and overall bigger ground distance between points when compared to GEDI as its different collection method may have allowed it to provide data points over the study area where GEDI did not. The researchers recommended using the ICESAT-2 data as predictor variables and the GEDI data as response variables in a genetic programming (GP) algorithm (Z. Wu et al., 2023). This produced a fused dataset of both ICESAT-2 and GEDI data, which was then used instead of solely GEDI data as the response variable in the MLM. The use of the fused GEDI and ICESAT-2 data resulted in an increased R² of 0.58 and a lowered RMSE of 3.63m for the RF model compared to using only GEDI data with an R² of 0.48 and an RMSE of 4.04m, both models having the same set of predictor variables. Additionally, alternative MLMs, such as the forest canopy height extrapolation (FCHE) model used in Z. Wu et al. (2023) and the multiple linear regression (MLR) model in Guo et al. (2023), outperformed the RF MLM under similar input conditions. These alternative models showed better performance than the RF MLM, which has traditionally been used in tree canopy height extraction studies (Chere et al., 2023; Kacic et al., 2023; Tsao et al., 2023; Tsutsumida et al., 2023; Xi et al., 2022)

Comparison between INSAR fusion CHM and RF MLM CHM with ground truth

The third objective of this study was to compare the performance of SAR data alone, using the INSAR process with Sentinel-1, and the incorporation of other data sources in a MLM to estimate tree canopy heights in a *Pinus radiata* plantation located in Mount Panorama and Knott Hill.

Overall, the INSAR-based CHM performed best in the T1 areas, exhibiting a low root mean square error (RMSE) of 4.24m and a percentage difference of 5.07% compared to the ground truth. The

RF MLM CHM, on the other hand, demonstrated superior performance in the T2 areas, with the lowest recorded RMSE of 1.92m and a percentage difference of 5.01%. Furthermore, the RF MLM method also performed well in the T1 areas, although the associated RMSE and percentage difference were higher compared to the T2 areas, measuring 2.34m and 8.77% respectively. However, it was deemed more reliable than the INSAR method due to its lower RMSE and thus the derived tree height may be more reliable. Both methods, however, proved to be inaccurate in the T0 and T3 areas. The higher performance of INSAR over T1 areas can be explained by the fact that there has been only one thinning process carried out, and thus the trees are still relatively free of the effect of visible ground between the trees. This exposed ground might be filled with pruned and thinned branches, which would increase the proportion of signal scattering and result in lower coherence. This is something expected in patches that have undergone more thinning processes. However, it would have been expected that T0 conditions might have the highest INSAR CHM accuracy due to its higher measured coherence and signal intensity. However, it was observed that the T0 areas had the smallest area in the study site, and thus the errors in the INSAR derived CHM over these sites could not be averaged out. The same could explain the poor performance in the T3 areas, which covered only around 1 km² more of the study area compared to T0 areas. The relatively small area of T0 and T3 areas, when compared to T1 and T2, can also explain the lack of accuracy over those areas. The relatively sparse number of inputs GEDI points over the study site might have been even more non-existent over those small patches, causing the MLM to be carried out on only a few input points. Moreover, it was observed that most of the T0 areas were classified as ground by the ESA forest mask, and data from the input predictor variables were masked out over these areas, further reducing the ML predicting power over these areas.

The higher performance of the RF MLM can be explained by the fact that in similar ML studies, it has been observed that Sentinel-1 SAR data is often not as significant in tree canopy estimation and often adds more noise to the ML than predictive power. This noise has also been observed in the INSAR process. As shown in N. Zhang et al. (2022), when combined with optical data as was done in this study with the RF MLM, the optical variables often have a higher ability to predict tree height. Thus, supplementing the SAR data with optical data generates an overall higher predictive model for tree canopy height.

Moreover, due to its relative ease of use once coded and its ability to be easily implemented over other study sites due to the global datasets used as input variables, the RF MLM model was found to make the best use of Sentinel-1 SAR data and overall would be the preferred methodology to estimate canopy heights over forest plantations.

Overall limitations of this study

This study primarily focuses on categorising different age groups of *Pinus Radiata* trees based on their planting years and the number of times they were thinned before being felled by Forestry SA. However, Figure 4.2 shows that the thinning process occurs over a wide range of ages. For instance, the second thinning may occur between 18 and 22 years. Unfortunately, the exact dates of thinning for each *Pinus radiata* patch investigated were not available. Therefore, the breakdown of the tree plantations is based on the approximate timeline provided by Forestry SA ForestrySA (2016). However, if the thinning process happened earlier or later than expected due to logistical reasons, there is a possibility of certain patches being allocated incorrectly, which may impact the confidence in the results obtained by the fusion INSAR CHM and the RF MLM.

Possible improvement by merging INSAR data and deep learning methodologies

The use of INSAR and MLM has been observed throughout this study as robust ways to estimate characteristics of forests, such as tree canopy height. However, there are novel methodologies that aim to combine these two principles and generate new, more accurate ways to predict tree canopy height. It is noted that MLM has been widely used for estimating these physical parameters, but there has recently been a new focus on the usage of deep learning methods to replace traditional regression models with great success. For example, Carcereri et al. (2023) demonstrated the ability to combine INSAR X band data and their derived geometry features with deep learning to successfully generate CHMs that rival the accuracy of RVoG models.

Given the nature of deep learning models and their inherent ability to iterate and have higher descriptive power over time, these methodologies can be expected to increase in accuracy as they become more refined and receive more data. This may result in an outcome where instead of being used separately as in Carcereri et al. (2023); Lang et al. (2019), different bands of SAR data, along with different optical data variables and spaceborne LIDAR, can be used as input in deep learning models to all together predict various biophysical parameters over forests

CHAPTER 7 CONCLUSION

The primary objective of this research was to investigate the use of Sentinel-1 data in estimating tree canopy heights, both independently through the INSAR process and in combination with other data sets in a MLM. The findings showed that the INSAR process yielded the most accurate results in *Pinus radiata* plantations that had undergone only one thinning process (T1). However, for this method to be effective, the areas needed to be large enough to average out the noise resulting from the INSAR process. Alternatively, it was observed that combining Sentinel-1 data with optical and topographic datasets to generate variables for an MLM was a better approach. This method proved suitable when compared with LIDAR ground truth, demonstrating its potential for application in a wider range of *Pinus radiata* plantations, including both T1 and T2 areas. Based on these findings, it can be concluded that the current freely available Sentinel-1 data can serve as an initial proxy for estimating tree canopy heights in plantations at regular intervals and for now would be better used in MLM instead as being used for INSAR. As the order of accuracy for the MLM was in the orders of metres, for now instead of replacing LIDAR surveys, foresters can initially use the MLM approach to initially investigate the *Pinus radiata* plantations and overtime identify areas that may suffer from abnormalities such as areas with a lower canopy height than expected based on the growth rates, and then target the more accurate LIDAR and field surveys only in those areas thus resulting in cost and time savings

Further Research

It is important to note that the Sentinel-1 data, due to its intermediate canopy penetration ability, is inherently not appropriate to be used over forested areas for tree height estimations. However, several missions are planned soon that aim to collect SAR data with enhanced canopy penetration ability, thanks to longer wavelengths.

One such mission is the NASA-ISRO Synthetic Aperture Radar (NISAR) mission, which is a collaboration between the US and Indian space agencies. The NISAR spacecraft planned to be launched in 2024, will be equipped with both L-band and S-band SAR instruments, providing global coverage with a 12-day repeat cycle (Fatoyinbo et al., 2021). The stability and capabilities of the spacecraft's orbit will allow for the use of longer wavelength L-band data in a repeat-pass INSAR process, similar to the one carried out in this study with Sentinel-1's C-band SAR scenes. This will enable more accurate estimation of tree canopy heights, particularly in the range of 10m to 30m, which is perfect for *Pinus radiata* plantations, as observed in the study site investigated in this research (Fatoyinbo et al., 2021).

However, the BIOMASS mission, planned to be launched in 2024, will be equipped with the P-band SAR instrument, which has higher penetration ability into the tree canopy compared to the wavelengths of the L, S, and C band data from NISAR and Sentinel-1 (Fatoyinbo et al., 2021). With a repeat coverage of only 3 days, the SAR data can also be used for multiple repeat pass interferometry over forests to produce forest tomography. However, it will have a much lower temporal decorrelation compared to the Sentinel-1 and NISAR data, with SAR scenes separated by 12 days instead of only 3 days in the case of BIOMASS (Fatoyinbo et al., 2021). Moreover, due to the range of different polarisations it can collect SAR data in, the BIOMASS mission will support the Polarimetric INSAR, which is expected to have higher tree height prediction abilities due to the usage of multiple SAR polarisations simultaneously instead of a single one as used in this study for INSAR CHM generation (Fatoyinbo et al., 2021).

Furthermore, as both datasets from NISAR and BIOMASS are expected to be freely available, their integration in online cloud platforms such as GEE will enable their fusion with other existing datasets, such as the one used in this study. The RF MLM model has already been found to be more accurate than the Sentinel-1 INSAR based method for predicting tree canopy heights. Therefore, these additional datasets will further enhance the accuracy of the RF MLM model (Fatoyinbo et al., 2021).

CHAPTER 8 REFERENCES

- Ab Rahman, A. A., Abd Majid, N., Ahli, N. A., Ab Latip, A. S., & Taib, A. M. (2023). The capability of SNAP software application to identify landslide using InSAR technique. *Physics and Chemistry of the Earth, Parts A/B/C*, 103427.
- Alijani, Z., Lindsay, J., Chabot, M., Rowlandson, T., & Berg, A. (2021). Sensitivity of C-Band SAR Polarimetric Variables to the Directionality of Surface Roughness Parameters. *Remote Sensing*, 13(11), 2210.
- Arslan, İ., Topakçı, M., & Demir, N. (2022). Monitoring Maize Growth and Calculating Plant Heights with Synthetic Aperture Radar (SAR) and Optical Satellite Images. *Agriculture*, 12(6), 800.
- Askne, J. I., Dammert, P. B., Ulander, L. M., & Smith, G. (1997). C-band repeat-pass interferometric SAR observations of the forest. *IEEE Transactions on Geoscience and Remote Sensing*, 35(1), 25-35.
- Asopa, U., & Kumar, S. (2020). UAVSAR tomography for vertical profile generation of tropical forest of Mondah National Park, Gabon. *Earth and Space Science*, 7(10), e2020EA001230.
- Aussenac, G. (2000). Interactions between forest stands and microclimate: ecophysiological aspects and consequences for silviculture. *Annals of forest science*, 57(3), 287-301.
- Bamler, R., & Hartl, P. (1998). Synthetic aperture radar interferometry. *Inverse problems*, 14(4), R1.
- Bamler, R., Kampes, B., Adam, N., & Suchandt, S. (2006). Assessment of slow deformations and rapid motions by radar interferometry. *Zeitschrift für Geo-Informationssysteme (GIS)*, 7, 22-27.
- Baran, I., Stewart, M. P., Kampes, B. M., Perski, Z., & Lilly, P. (2003). A modification to the Goldstein radar interferogram filter. *IEEE Transactions on Geoscience and Remote Sensing*, 41(9), 2114-2118.
- Beale, J., Snapir, B., Waive, T., Evans, J., & Corstanje, R. (2019). The significance of soil properties to the estimation of soil moisture from C-band synthetic aperture radar. *Hydrology and Earth System Sciences Discussions*, 1-32.
- Bellec, R., Legris, M., Khenchaf, A., Amate, M., & Hetet, A. (2005). Repeat-track SAS interferometry: feasibility study. Proceedings of OCEANS 2005 MTS/IEEE,
- Bi, H.-X., & Wei, Z.-Q. (2016). A new phase unwrapping method based on region recognition and region expansion. *International Journal of Remote Sensing*, 37(22), 5287-5303.
- Bindzarova Gergelova, M., Labant, S., Mizak, J., Sustek, P., & Leicher, L. (2021). Inventory of locations of old mining works using LiDAR data: a case study in Slovakia. *Sustainability*, 13(12), 6981.
- Bonneau, D. A., DiFrancesco, P.-M., & Hutchinson, D. J. (2020). A method for vegetation extraction in mountainous terrain for rockfall simulation. *Remote Sensing of Environment*, 251, 112098.

- Bouaraba, A., Belhadj-Aissa, A., & Closson, D. (2018). Drastic improvement of change detection results with multilook complex SAR images approach. *Progress In Electromagnetics Research C*, 82, 55-66.
- Braun, A. (2021). Retrieval of digital elevation models from Sentinel-1 radar data—open applications, techniques, and limitations. *Open Geosciences*, 13(1), 532-569.
- Braun, A., & Veci, L. (2021). TOPS Interferometry Tutorial. *ESA: Paris, France*.
- Brennan, M., McLean, J. P., Altaner, C. M., Ralph, J., & Harris, P. J. (2012). Cellulose microfibril angles and cell-wall polymers in different wood types of *Pinus radiata*. *Cellulose*, 19, 1385-1404.
- Brown, S. (2002). Measuring carbon in forests: current status and future challenges. *Environmental pollution*, 116(3), 363-372.
- Bucher, T., & Bayer, S. (2013). Application of Remote Sensing and GIS Methods for the Automatic Extraction of Single Trees Based on Digital Aerial Images and Elevation Models.
- Burkhart, H. E., Tomé, M., Burkhart, H. E., & Tomé, M. (2012). Tree form and stem taper. *Modeling forest trees and stands*, 9-41.
- Burrough, P. A., McDonnell, R. A., & Lloyd, C. D. (1998). *Principles of geographical information systems*. Oxford University Press, USA.
- Cagnina, C., Marino, A., Silva-Perez, C., Ruiz-Ramos, J., & Suarez, J. (2023). Assessment of the Impact of Surface Water Content for Temperate Forests in SAR Data at C-Band. *Remote Sensing*, 15(24), 5723.
- Carcereri, D., Rizzoli, P., Ienco, D., & Bruzzone, L. (2023). A deep learning framework for the estimation of forest height from bistatic TanDEM-X data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
- Carstairs, H., Mitchard, E. T., McNicol, I., Aquino, C., Chezeaux, E., Ebanega, M. O., Dikongo, A. M., & Disney, M. (2022). Sentinel-1 Shadows Used to Quantify Canopy Loss from Selective Logging in Gabon. *Remote Sensing*, 14(17), 4233.
- Cartus, O., Santoro, M., Wegmüller, U., Labrière, N., & Chave, J. (2021). Sentinel-1 coherence for mapping above-ground biomass in semiarid forest areas. *IEEE Geoscience and Remote Sensing Letters*, 19, 1-5.
- Castel, T., Martinez, J.-M., Beaudoin, A., Wegmüller, U., & Strozzi, T. (2000). ERS INSAR data for remote sensing hilly forested areas. *Remote Sensing of Environment*, 73(1), 73-86.
- Căţeanu, M., & Ciubotaru, A. (2021). The effect of lidar sampling density on DTM accuracy for areas with heavy forest cover. *Forests*, 12(3), 265.
- Chang, Z., Fan, L., Wigneron, J.-P., Wang, Y.-P., Ciais, P., Chave, J., Fensholt, R., Chen, J. M., Yuan, W., & Ju, W. (2023). Estimating aboveground carbon dynamic of China using optical and microwave remote sensing datasets from 2013 to 2019. *Journal of Remote Sensing*.

- Chen, W., Zheng, Q., Xiang, H., Chen, X., & Sakai, T. (2021). Forest canopy height estimation using polarimetric interferometric synthetic aperture radar (PolInSAR) technology based on full-polarized ALOS/PALSAR data. *Remote Sensing*, 13(2), 174.
- Chen, Y., Zhang, G., Ding, X., & Li, Z. (2000). Monitoring earth surface deformations with InSAR technology: principles and some critical issues. *Journal of Geospatial Engineering*, 2(1), 3-22.
- Chere, Z., Zewdie, W., & Biru, D. (2023). Machine learning for modeling forest canopy height and cover from multi-sensor data in Northwestern Ethiopia. *Environmental Monitoring and Assessment*, 195(12), 1452.
- Coops, N. C., Tompalski, P., Goodbody, T. R., Queinnec, M., Luther, J. E., Bolton, D. K., White, J. C., Wulder, M. A., van Lier, O. R., & Hermosilla, T. (2021). Modelling lidar-derived estimates of forest attributes over space and time: A review of approaches and future trends. *Remote Sensing of Environment*, 260, 112477.
- Courage, K. (2021). *Lesson 1.1 – Mapping Vegetation Structure using GEDI and Sentinel Data in Earth Engine*. Retrieved 6/05 from <https://aigeolabs.com/courses/course-6/module-1/1-1/>
- De Petris, S., Cuozzo, G., Notarnicola, C., & Borgogno-Mondino, E. (2022). Forest Height Estimation Using Sentinel-1 Interferometry. A Phase Unwrapping-Free Method Based on Least Squares Adjustment. Italian Conference on Geomatics and Geospatial Technologies,
- De Petris, S., Sarvia, F., & Borgogno-Mondino, E. (2022a). About Tree Height Measurement: Theoretical and Practical Issues for Uncertainty Quantification and Mapping. *Forests*, 13(7), 969.
- De Petris, S., Sarvia, F., & Borgogno-Mondino, E. (2022b). Uncertainties and perspectives on forest height estimates by Sentinel-1 interferometry. *Earth*, 3(1), 479-492.
- DiGiacomo, A. E., Bird, C. N., Pan, V. G., Dobroski, K., Atkins-Davis, C., Johnston, D. W., & Ridge, J. T. (2020). Modeling salt marsh vegetation height using unoccupied aircraft systems and structure from motion. *Remote Sensing*, 12(14), 2333.
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., & Martimort, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25-36.
- Du, L., Zhou, T., Zou, Z., Zhao, X., Huang, K., & Wu, H. (2014). Mapping forest biomass using remote sensing and national forest inventory in China. *Forests*, 5(6), 1267-1283.
- ESRI. (2023). *LAS To Raster function*. Retrieved 22/09 from https://pro.arcgis.com/en/pro-app/latest/help/analysis/raster-functions/las-to-raster-function.htm#ESRI_SECTION1_87A91E55B8554F6BA75E3DC79318CCE9
- Fan, L., Smethurst, J., Atkinson, P., & Powrie, W. (2014). Propagation of vertical and horizontal source data errors into a TIN with linear interpolation. *International journal of geographical information science*, 28(7), 1378-1400.

- Fankhauser, K. E., Strigul, N. S., & Gatzolis, D. (2018). Augmentation of traditional forest inventory and airborne laser scanning with unmanned aerial systems and photogrammetry for forest monitoring. *Remote Sensing*, *10*(10), 1562.
- Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., & Roth, L. (2007). The shuttle radar topography mission. *Reviews of geophysics*, *45*(2).
- Fatoyinbo, T., Armston, J., Simard, M., Saatchi, S., Denbina, M., Lavallo, M., Hofton, M., Tang, H., Marselis, S., & Pinto, N. (2021). The NASA AfriSAR campaign: Airborne SAR and lidar measurements of tropical forest structure and biomass in support of current and future space missions. *Remote Sensing of Environment*, *264*, 112533.
- Fernández, M. P., Basauri, J., Madariaga, C., Menéndez-Miguélez, M., Olea, R., & Zubizarreta-Gerendiain, A. (2017). Effects of thinning and pruning on stem and crown characteristics of radiata pine (*Pinus radiata* D. Don). *iForest-Biogeosciences and Forestry*, *10*(2), 383.
- Filipponi, F. (2019). Sentinel-1 GRD preprocessing workflow. International Electronic Conference on Remote Sensing,
- Fletcher, K. (2012). *SENTINEL 1: ESA's Radar Observatory Mission for GMES Operational Services*. European Space Agency.
- ForestrySA. (2016). Mount Panorama, Knott Hill and Christmas Hill Native Forest Reserves Management Plan. In: ForestrySA.
- ForestrySA. (2018). Plantation Forestry. In *How Does it Work?*
- ForestrySA. (2021). Mount Lofty Ranges Forest Management Plan. In.
- Foumelis, M., Blasco, J. M. D., Desnos, Y.-L., Engdahl, M., Fernández, D., Veci, L., Lu, J., & Wong, C. (2018). ESA SNAP-StaMPS integrated processing for Sentinel-1 persistent scatterer interferometry. IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium,
- Frolking, S., Palace, M. W., Clark, D., Chambers, J. Q., Shugart, H., & Hurtt, G. C. (2009). Forest disturbance and recovery: A general review in the context of spaceborne remote sensing of impacts on aboveground biomass and canopy structure. *Journal of Geophysical Research: Biogeosciences*, *114*(G2).
- Geremew, T., Gonsamo, A., Zewdie, W., & Pellikka, P. (2023). Extrapolation of canopy height and cover metrics of GEDI LiDAR in tropical montane forest ecosystem. *African Geographical Review*, 1-17.
- Ghosh, S. M., Behera, M. D., Kumar, S., Das, P., Prakash, A. J., Bhaskaran, P. K., Roy, P. S., Barik, S. K., Jeganathan, C., & Srivastava, P. K. (2022). Predicting the forest canopy height from LiDAR and multi-sensor data using machine learning over India. *Remote Sensing*, *14*(23), 5968.

- Ghosh, S. M., Behera, M. D., & Paramanik, S. (2020). Canopy height estimation using sentinel series images through machine learning models in a mangrove forest. *Remote Sensing*, *12*(9), 1519.
- Ghulam, A., Amer, R., & Ripperdan, R. (2010). A filtering approach to improve deformation accuracy using large baseline, low coherence DInSAR phase images. 2010 IEEE International Geoscience and Remote Sensing Symposium,
- Gómez, C., Lopez-Sanchez, J. M., Romero-Puig, N., Zhu, J., Fu, H., He, W., Xie, Y., & Xie, Q. (2021). Canopy height estimation in Mediterranean forests of Spain with TanDEM-X data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 2956-2970.
- Gomroki, M., Jafari, M., Sadeghian, S., & Azizi, Z. (2017). Application of intelligent interpolation methods for DTM generation of forest areas based on LiDAR data. *ISPRS International Journal of Geo-Information*, *6*(12), 227-241.
- Gonçalves, G. (2006). Analysis of interpolation errors in urban digital surface models created from LIDAR data. 7th international symposium on spatial accuracy assessment in natural resources and environmental sciences,
- Gong, J., Li, Z., Zhu, Q., Sui, H., & Zhou, Y. (2000). Effects of various factors on the accuracy of DEMs: an intensive experimental investigation. *Photogrammetric Engineering and Remote Sensing*, *66*(9), 1113-1117.
- Guliaev, R., Cazcarra-Bes, V., Pardini, M., & Papathanassiou, K. (2021). Forest height estimation by means of TanDEM-X InSAR and waveform lidar data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 3084-3094.
- Guo, Q., Du, S., Jiang, J., Guo, W., Zhao, H., Yan, X., Zhao, Y., & Xiao, W. (2023). Combining GEDI and sentinel data to estimate forest canopy mean height and aboveground biomass. *Ecological Informatics*, *78*, 102348.
- Gururaj, P., Umesh, P., & Shetty, A. (2021). Modeling of surface soil moisture using C-band SAR data over bare fields in the tropical semi-arid region of India. *Applied Geomatics*, *13*(4), 555-564.
- Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*(6), 610-621.
- Hu, Y., Nie, Y., Liu, Z., Wu, G., & Fan, W. (2023). Improving the Potential of Coniferous Forest Aboveground Biomass Estimation by Integrating C-and L-Band SAR Data with Feature Selection and Non-Parametric Model. *Remote Sensing*, *15*(17), 4194.
- Ismail, M. H., Shidiq, I. P. A., Ramli, M. F., Kamarudin, N., & Zaki, P. H. (2023). Rubber Trees and Biomass Estimation Using Remote Sensing Technology. In *Concepts and Applications of Remote Sensing in Forestry* (pp. 185-214). Springer.

- Ismail, Z., Abdul Khanan, M., Omar, F., Abdul Rahman, M., & Mohd Salleh, M. (2016). Evaluating error of lidar derived DEM interpolation for vegetation area. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 141-150.
- Jacobsen, K. (2005). High resolution satellite imaging systems-an overview. *Photogrammetrie Fernerkundung Geoinformation*, 2005(6), 487.
- Jagdhuber, T. (2012). *Soil parameter retrieval under vegetation cover using SAR polarimetry* RIMAX Publications, RIMAX, Deutsches GeoForschungsZentrum].
- Jones, H. G., & Vaughan, R. A. (2010). *Remote sensing of vegetation: principles, techniques, and applications*. Oxford university press.
- Jurjević, L., Liang, X., Gašparović, M., & Balenović, I. (2020). Is field-measured tree height as reliable as believed—Part II, A comparison study of tree height estimates from conventional field measurement and low-cost close-range remote sensing in a deciduous forest. *ISPRS journal of photogrammetry and remote sensing*, 169, 227-241.
- Kacic, P., Thonfeld, F., Gessner, U., & Kuenzer, C. (2023). Forest structure characterization in Germany: novel products and analysis based on GEDI, sentinel-1 and sentinel-2 data. *Remote Sensing*, 15(8), 1969.
- Khan, S., Fryirs, K. A., & Shumack, S. (2021). Semi-automating the calculation of catchment scale geomorphic controls on river diversity using publically available datasets. *Catena*, 203, 105354.
- Krieger, G., Moreira, A., Fiedler, H., Hajnsek, I., Werner, M., Younis, M., & Zink, M. (2007). TanDEM-X: A satellite formation for high-resolution SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 45(11), 3317-3341.
- Krivoruchko, K. (2012). Empirical bayesian kriging. *ArcUser Fall*, 6(10), 1145.
- Krivoruchko, K., & Gribov, A. (2019). Evaluation of empirical Bayesian kriging. *Spatial Statistics*, 32, 100368.
- Kumar, P., & Krishna, A. P. (2019). InSAR-based tree height estimation of hilly forest using multitemporal Radarsat-1 and Sentinel-1 SAR data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(12), 5147-5152.
- Lang, N., Schindler, K., & Wegner, J. D. (2019). Country-wide high-resolution vegetation height mapping with Sentinel-2. *Remote Sensing of Environment*, 233, 111347.
- Lei, Y., & Siqueira, P. (2014). Estimation of forest height using spaceborne repeat-pass L-band InSAR correlation magnitude over the US state of Maine. *Remote Sensing*, 6(11), 10252-10285.
- Li, F. K., & Goldstein, R. M. (1990). Studies of multibaseline spaceborne interferometric synthetic aperture radars. *IEEE Transactions on Geoscience and Remote Sensing*, 28(1), 88-97.
- Li, J., & Heap, A. D. (2008). A review of spatial interpolation methods for environmental scientists.
- Li, W., Niu, Z., Shang, R., Qin, Y., Wang, L., & Chen, H. (2020). High-resolution mapping of forest canopy height using machine learning by coupling ICESat-2 LiDAR with Sentinel-1,

- Sentinel-2 and Landsat-8 data. *International Journal of Applied Earth Observation and Geoinformation*, 92, 102163.
- Li, X., Ye, Z., Long, J., Zheng, H., & Lin, H. (2022). Inversion of Coniferous Forest Stock Volume Based on Backscatter and InSAR Coherence Factors of Sentinel-1 Hyper-Temporal Images and Spectral Variables of Landsat 8 OLI. *Remote Sensing*, 14(12), 2754.
- Li, Z., & Bethel, J. (2008). Image coregistration in SAR interferometry. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37, 433-438.
- Li, Z., Zhu, C., & Gold, C. (2004). *Digital terrain modeling: principles and methodology*. CRC press.
- Liang, X., Kukko, A., Balenovic, I., Saarinen, N., Junttila, S., Kankare, V., Holopainen, M., Mokroš, M., Surový, P., & Kaartinen, L. (2022). Close-Range Remote Sensing of Forests-The state of the art, challenges, and opportunities for systems and data acquisitions. *IEEE Geoscience and Remote Sensing Magazine (GRSM)*.
- Ling, Y., Teng, S., Liu, C., Dash, J., Morris, H., & Pastor-Guzman, J. (2022). Assessing the accuracy of forest phenological extraction from sentinel-1 C-band backscatter measurements in deciduous and coniferous forests. *Remote Sensing*, 14(3), 674.
- Liu, A., Cheng, X., & Chen, Z. (2021). Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals. *Remote Sensing of Environment*, 264, 112571.
- Liu, S., Tang, H., Feng, Y., Chen, Y., Lei, Z., Wang, J., & Tong, X. (2021). A Comparative Study of DEM Reconstruction Using the Single-Baseline and Multibaseline InSAR Techniques. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 8512-8521.
- Liu, X. (2008). Airborne LiDAR for DEM generation: some critical issues. *Progress in Physical Geography*, 32(1), 31-49.
- Liu, X., He, L., He, Z., & Wei, Y. (2022). Estimation of Broadleaf Tree Canopy Height of Wolong Nature Reserve Based on InSAR and Machine Learning Methods. *Forests*, 13(8), 1282.
- Liu, Y., Gong, W., Xing, Y., Hu, X., & Gong, J. (2019). Estimation of the forest stand mean height and aboveground biomass in Northeast China using SAR Sentinel-1B, multispectral Sentinel-2A, and DEM imagery. *ISPRS journal of photogrammetry and remote sensing*, 151, 277-289.
- Location SA. (2023). *Survey Marks Dataset 2239*.
https://location.sa.gov.au/LMS/Reports/ReportMetadata.aspx?p_no=2239&pu=y
- Luo, Y., Qi, S., Liao, K., Zhang, S., Hu, B., & Tian, Y. (2023). Mapping the Forest Height by Fusion of ICESat-2 and Multi-Source Remote Sensing Imagery and Topographic Information: A Case Study in Jiangxi Province, China. *Forests*, 14(3), 454.
- Ma, Z.-F., Jiang, M., Zhao, Y., Malhotra, R., & Yong, B. (2019). Minimum spanning tree co-registration approach for time-series Sentinel-1 TOPS data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(8), 3004-3013.

- Managhebi, T., Maghsoudi, Y., & Valadan Zoej, M. J. (2018). Four-stage inversion algorithm for forest height estimation using repeat pass polarimetric SAR interferometry data. *Remote Sensing*, 10(8), 1174.
- Meyer, F. (2019). Spaceborne Synthetic Aperture Radar: Principles, data access, and basic processing techniques. *Synthetic Aperture Radar (SAR) Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation*, 21-64.
- Mohammadi, A., Karimzadeh, S., Jalal, S. J., Kamran, K. V., Shahabi, H., Homayouni, S., & Al-Ansari, N. (2020). A multi-sensor comparative analysis on the suitability of generated DEM from Sentinel-1 SAR interferometry using statistical and hydrological models. *Sensors*, 20(24), 7214.
- Monteith, A. R., & Ulander, L. M. (2021). A tower-based radar study of temporal coherence of a boreal forest at P-, L-, and C-bands and linear cross polarization. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-15.
- Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., & Papathanassiou, K. P. (2013). A tutorial on synthetic aperture radar. *IEEE Geoscience and remote sensing magazine*, 1(1), 6-43.
- Muir, J., Goodwin, N., Armston, J., Phinn, S., & Scarth, P. (2017). An accuracy assessment of derived digital elevation models from terrestrial laser scanning in a sub-tropical forested environment. *Remote Sensing*, 9(8), 843.
- Nandy, S., Srinet, R., & Padalia, H. (2021). Mapping forest height and aboveground biomass by integrating ICESat-2, Sentinel-1 and Sentinel-2 data using Random Forest algorithm in northwest Himalayan foothills of India. *Geophysical Research Letters*, 48(14), e2021GL093799.
- NISAR. (2018). NASA-ISRO SAR (NISAR) Mission Science Users' Handbook. In. NASA Jet Propulsion Laboratory.
- Numbisi, F. N., Van Coillie, F. M., & De Wulf, R. (2019). Delineation of cocoa agroforests using multiseason Sentinel-1 SAR images: a low grey level range reduces uncertainties in GLCM texture-based mapping. *ISPRS international journal of geo-information*, 8(4), 179.
- Nurtyawan, R., & Yulanda, M. F. (2020). Lombok earthquakes using DInSAR techniques based on Sentinel 1A data (case study: Lombok, West Nusa Tenggara). IOP Conference Series: Earth and Environmental Science,
- Ouhami, M., Hafiane, A., Es-Saady, Y., El Hajji, M., & Canals, R. (2021). Computer vision, IoT and data fusion for crop disease detection using machine learning: A survey and ongoing research. *Remote Sensing*, 13(13), 2486.
- Paek, S. W., Balasubramanian, S., Kim, S., & de Weck, O. (2020). Small-satellite synthetic aperture radar for continuous global biospheric monitoring: A review. *Remote Sensing*, 12(16), 2546.

- Palm, R. (2022). Above-ground biomass estimation in boreal mountain birch forests using Sentinel-1 data. In.
- Pandit, A., Sawant, S., Mohite, J., & Pappula, S. (2022). Sentinel-1-derived coherence time-series for crop monitoring in Indian agriculture region. *Geocarto International*, 37(25), 9497-9517.
- Perez, G. G., Bourscheidt, V., Lopes, L. E., Takata, J. T., Ferreira, P. A., & Boscolo, D. (2022). Use of Sentinel 2 imagery to estimate vegetation height in fragments of Atlantic Forest. *Ecological Informatics*, 69, 101680.
- Periasamy, S., & Ravi, K. P. (2020). The Effect of Varying Moisture Content in the Retrieval of the Imaginary Part of Dielectric Constant from C-Band Frequency SAR. 2020 IEEE India Geoscience and Remote Sensing Symposium (InGARSS),
- Persson, H. J., & Fransson, J. E. (2017). Comparison between TanDEM-X-and ALS-based estimation of aboveground biomass and tree height in boreal forests. *Scandinavian Journal of Forest Research*, 32(4), 306-319.
- Pinkard, E., & Beadle, C. (2000). A physiological approach to pruning. *The International Forestry Review*, 295-305.
- Pinto, N., Simard, M., & Dubayah, R. (2012). Using InSAR coherence to map stand age in a boreal forest. *Remote Sensing*, 5(1), 42-56.
- PIRSA. (2023). Overview of farm based forestry in South Australia. In *FORESTRY FACT SHEET 20*.
- Prior, E. M., Czuba, J. A., Pingel, T. J., Thomas, V. A., Wynne, R. H., & Hession, W. C. (2022). INVESTIGATING FLOOD-VEGETATION INTERACTIONS THROUGH REMOTE SENSING AND MODELING. Virginia Space Grant Consortium Student Research Conference,
- Qi, W., Lee, S.-K., Hancock, S., Luthcke, S., Tang, H., Armston, J., & Dubayah, R. (2019). Improved forest height estimation by fusion of simulated GEDI Lidar data and TanDEM-X InSAR data. *Remote Sensing of Environment*, 221, 621-634.
- Rabus, B., Eineder, M., Roth, A., & Bamler, R. (2003). The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar. *ISPRS journal of photogrammetry and remote sensing*, 57(4), 241-262.
- Razas, M., Hassan, A., Khan, M., Emach, M., & Saki, S. (2023). A critical comparison of interpolation techniques for digital terrain modelling in mining. *Journal of the Southern African Institute of Mining and Metallurgy*, 123(2), 53-62.
- Roc Roc, D. (2019). Above-ground biomass estimation in boreal productive forests using Sentinel-1 data. In.
- Rosen, P. A., Hensley, S., Joughin, I. R., Li, F. K., Madsen, S. N., Rodriguez, E., & Goldstein, R. M. (2000). Synthetic aperture radar interferometry. *Proceedings of the IEEE*, 88(3), 333-382.

- Rüetschi, M., Schaepman, M. E., & Small, D. (2017). Using multitemporal sentinel-1 c-band backscatter to monitor phenology and classify deciduous and coniferous forests in northern switzerland. *Remote Sensing*, 10(1), 55.
- Rüetschi, M., Small, D., & Waser, L. T. (2019). Rapid detection of windthrows using Sentinel-1 C-band SAR data. *Remote Sensing*, 11(2), 115.
- S1TBX. (2023). *ESA Sentinel-1 Toolbox - Documentation*. In <http://step.esa.int>
- Sa, R., Nei, Y., & Fan, W. (2023). Combining Multi-Dimensional SAR Parameters to Improve RVoG Model for Coniferous Forest Height Inversion Using ALOS-2 Data. *Remote Sensing*, 15(5), 1272.
- Salekin, S., Burgess, J. H., Morgenroth, J., Mason, E. G., & Meason, D. F. (2018). A comparative study of three non-geostatistical methods for optimising digital elevation model interpolation. *ISPRS international journal of geo-information*, 7(8), 300.
- Santoro, M. (2003). *Estimation of biophysical parameters in boreal forests from ERS and JERS SAR interferometry* Department of Radio and Space Science, Chalmers University of Technology].
- Santoro, M., Cartus, O., Fransson, J. E., & Wegmüller, U. (2019). Complementarity of X-, C-, and L-band SAR backscatter observations to retrieve forest stem volume in boreal forest. *Remote Sensing*, 11(13), 1563.
- Santoro, M., Shvidenko, A., McCallum, I., Askne, J., & Schmillius, C. (2007). Properties of ERS-1/2 coherence in the Siberian boreal forest and implications for stem volume retrieval. *Remote Sensing of Environment*, 106(2), 154-172.
- Sartoretto, V. (2021). Precision farming and sustainability: optimisation of a radar microsatellites constellation for Italian agriculture.
- Schlund, M., Wenzel, A., Camarretta, N., Stiegler, C., & Erasmi, S. (2022). Vegetation canopy height estimation in dynamic tropical landscapes with TanDEM-X supported by GEDI data. *Methods in Ecology and Evolution*.
- Schwartz, M., Ciais, P., Ottlé, C., De Truchis, A., Vega, C., Fayad, I., Brandt, M., Fensholt, R., Baghdadi, N., & Morneau, F. (2022). High-resolution canopy height map in the Landes forest (France) based on GEDI, Sentinel-1, and Sentinel-2 data with a deep learning approach. *arXiv preprint arXiv:2212.10265*.
- Sefercik, U. G., Buyuksalih, G., & Atalay, C. (2020). DSM generation with bistatic TanDEM-X InSAR pairs and quality validation in inclined topographies and various land cover classes. *Arabian Journal of Geosciences*, 13, 1-15.
- Seppi, S. A., López-Martínez, C., & Joseau, M. J. (2022). Assessment of L-Band SAOCOM InSAR Coherence and Its Comparison with C-Band: A Case Study over Managed Forests in Argentina. *Remote Sensing*, 14(22), 5652.
- Silveira, E. M., Radeloff, V. C., Martinuzzi, S., Pastur, G. J. M., Bono, J., Politi, N., Lizarraga, L., Rivera, L. O., Ciuffoli, L., & Rosas, Y. M. (2023). Nationwide native forest structure maps

- for Argentina based on forest inventory data, SAR Sentinel-1 and vegetation metrics from Sentinel-2 imagery. *Remote Sensing of Environment*, 285, 113391.
- Simard, M., & Denbina, M. (2017). An assessment of temporal decorrelation compensation methods for forest canopy height estimation using airborne L-band same-day repeat-pass polarimetric SAR interferometry. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(1), 95-111.
- Singh, A., Kushwaha, S., & Kumar, S. (2020). Backscatter and coherence analysis using space borne C-band data for forest characterization. *Journal of Geomatics*, 14(1).
- Soja, M. J. (2014). *Modelling and retrieval of forest parameters from synthetic aperture radar data*. Chalmers Tekniska Hogskola (Sweden).
- Sothe, C., Gonsamo, A., Lourenço, R. B., Kurz, W. A., & Snider, J. (2022). Spatially Continuous Mapping of Forest Canopy Height in Canada by Combining GEDI and ICESat-2 with PALSAR and Sentinel. *Remote Sensing*, 14(20), 5158.
- Stack, V., & Narine, L. L. (2022). Sustainability at Auburn University: Assessing Rooftop Solar Energy Potential for Electricity Generation with Remote Sensing and GIS in a Southern US Campus. *Sustainability*, 14(2), 626.
- Sunila, R., & Kollo, K. (2007). A comparison of geostatistics and fuzzy application for digital elevation model. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(2).
- Szypuła, B. (2016). Geomorphometric comparison of DEMs built by different interpolation. *Landform Analysis*, 32, 45-58.
- Tampubolon, W., Reinhardt, W., & Amhar, F. (2022). Toward A New Standard for DEM Generation Based On Radar Remote Sensing Technology. Proceedings of the 2022 International Conference on Computer, Control, Informatics and Its Applications,
- Teluguntla, P., Thenkabail, P. S., Oliphant, A., Xiong, J., Gumma, M. K., Congalton, R. G., Yadav, K., & Huete, A. (2018). A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS journal of photogrammetry and remote sensing*, 144, 325-340.
- Tienaho, N., Yrttimaa, T., Kankare, V., Vastaranta, M., Luoma, V., Honkavaara, E., Koivumäki, N., Huuskonen, S., Hynynen, J., & Holopainen, M. (2022). Assessing Structural Complexity of Individual Scots Pine Trees by Comparing Terrestrial Laser Scanning and Photogrammetric Point Clouds. *Forests*, 13(8), 1305.
- Torres de Almeida, C., Gerente, J., Rodrigo dos Prazeres Campos, J., Caruso Gomes Junior, F., Providelo, L. A., Marchiori, G., & Chen, X. (2022). Canopy Height Mapping by Sentinel 1 and 2 Satellite Images, Airborne LiDAR Data, and Machine Learning. *Remote Sensing*, 14(16), 4112.

- Torres, R., Lokas, S., Di Cosimo, G., Geudtner, D., & Bibby, D. (2017). Sentinel 1 evolution: Sentinel-1C and-1D models. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS),
- Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Flouy, N., & Brown, M. (2012). GMES Sentinel-1 mission. *Remote Sensing of Environment*, 120, 9-24.
- Tsao, A., Nzewi, I., Jayeoba, A., Ayogu, U., & Lobell, D. B. (2023). Canopy Height Mapping for Plantations in Nigeria Using GEDI, Landsat, and Sentinel-2. *Remote Sensing*, 15(21), 5162.
- Tsutsumida, N., Kato, A., Osawa, T., & Doi, H. (2023). Mapping Forest Vertical Structure Attributes with GEDI, Sentinel-1, and Sentinel-2. IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium,
- Vepakomma, U., St-Onge, B., & Kneeshaw, D. (2008). Spatially explicit characterization of boreal forest gap dynamics using multi-temporal lidar data. *Remote Sensing of Environment*, 112(5), 2326-2340.
- Villarroya-Carpio, A., Lopez-Sanchez, J. M., & Engdahl, M. E. (2022). Sentinel-1 interferometric coherence as a vegetation index for agriculture. *Remote Sensing of Environment*, 280, 113208.
- Wallington, E. D. (2006). Forest height inventory from airborne Synthetic Aperture Radar. *Annexe Thesis Digitisation Project 2017 Block 16*.
- Wang, D., Wu, X., & Lin, D. (2009). Particle swarm mixel decomposition for remote sensing images. 2009 IEEE International Conference on Automation and Logistics,
- Wang, L., Zhou, Y., Shen, G., Xiong, J., & Shi, H. (2022). Forest Height Inversion Based on Time–Frequency RVoG Model Using Single-Baseline L-Band Sublook-InSAR Data. *Remote Sensing*, 15(1), 166.
- Wang, T., Liao, M., & Perissin, D. (2009). InSAR coherence-decomposition analysis. *IEEE Geoscience and Remote Sensing Letters*, 7(1), 156-160.
- Wang, Y., Lehtomäki, M., Liang, X., Pyörälä, J., Kukko, A., Jaakkola, A., Liu, J., Feng, Z., Chen, R., & Hyypä, J. (2019). Is field-measured tree height as reliable as believed—A comparison study of tree height estimates from field measurement, airborne laser scanning and terrestrial laser scanning in a boreal forest. *ISPRS journal of photogrammetry and remote sensing*, 147, 132-145.
- Weike, L., & Goulin, L. (2012). A new method of weighted choice in InSAR Least Squares unwrapping. *Geodesy and Geodynamics*, 3(1), 39-43.
- WillyWeather. (2023). *Kuitpo Forest Reserve*. Retrieved 29/12 from <https://www.willyweather.com.au/>
- Woodhouse, I. (2006). Introduction to microwave remote sensing. *Introduction to microwave remote sensing*.

- Woodhouse, I. H. (2017). *Introduction to microwave remote sensing*. CRC press.
- Wu, D. (2020). Measuring canopy structure for horticulture tree crops using remote sensing datasets.
- Wu, Y., Ren, H., & Madson, A. (2023). A Discussion on the Goldstein Filtering Parameters Within the Snap Software. IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium,
- Wu, Z., Yao, F., Zhang, J., Ma, E., Yao, L., & Dong, Z. (2023). Genetic Programming Guided Mapping of Forest Canopy Height by Combining LiDAR Satellites with Sentinel-1/2, Terrain, and Climate Data. *Remote Sensing*, *16*(1), 110.
- Xi, Z., Xu, H., Xing, Y., Gong, W., Chen, G., & Yang, S. (2022). Forest canopy height mapping by synergizing icesat-2, sentinel-1, sentinel-2 and topographic information based on machine learning methods. *Remote Sensing*, *14*(2), 364.
- Xie, Z., Liu, H., He, Y., Song, Y., Hu, F., & Guo, Q. (2023). Analysis of Forest Height Change in Helan Mountain Based on Multi-Band InSAR. 2023 3rd International Conference on Frontiers of Electronics, Information and Computation Technologies (ICFEICT),
- Yu, H., Lan, Y., Yuan, Z., Xu, J., & Lee, H. (2019). Phase unwrapping in InSAR: A review. *IEEE Geoscience and remote sensing magazine*, *7*(1), 40-58.
- Yun, Y., Zeng, Q., Jiao, J., Liang, C., Wang, Q., Zhou, X., & Yan, D. (2012). Calibration of airborne interferometric SAR data by external dem without artificial calibrators. 2012 IEEE International Geoscience and Remote Sensing Symposium,
- Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., & Fritz, S. (2022). ESA WorldCover 10 m 2021 v200.
- Zandbergen, P. A. (2011). Characterizing the error distribution of lidar elevation data for North Carolina. *International Journal of Remote Sensing*, *32*(2), 409-430.
- Zhang, C., Marzougui, A., & Sankaran, S. (2020). High-resolution satellite imagery applications in crop phenotyping: an overview. *Computers and Electronics in Agriculture*, *175*, 105584.
- Zhang, N., Chen, M., Yang, F., Yang, C., Yang, P., Gao, Y., Shang, Y., & Peng, D. (2022). Forest height mapping using feature selection and machine learning by integrating multi-source satellite data in Baoding City, North China. *Remote Sensing*, *14*(18), 4434.
- Zhang, Y., Liu, X., Liu, J., & Li, L. (2022). High-resolution Mapping of Forest Canopy Height by Integrating Sentinel and airborne LiDAR data. IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium,
- Zhou, X., Chang, N.-B., & Li, S. (2009). Applications of SAR interferometry in earth and environmental science research. *Sensors*, *9*(3), 1876-1912.
- Zylshal, Z., Saputra, M., Bayanuddin, A. A., Siwi, S. E., Setiyoko, A., & Arief, R. (2021). Space-borne Interferometric SAR for digital elevation model in the tropical region: An initial result of Sentinel-1 RADAR data over the Indonesian archipelago. 2021 7th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR),

CHAPTER 9 APPENDICES

Appendix A: Plantation years of *Pinus Radiata* present in the study area, adapted from ForestrySA (2016)

Appendix A has been removed due to Copyright restrictions

Appendix B: Code for the RF MLM adapted from Courage (2021) and used in the GEE platform

```
var boundary = ee.FeatureCollection("projects/ee-kunalkcg1997/assets/new_forest");
```

```
// Load Sentinel-1 for the post-rainy season.
```

```
var S1_PRS = ee.ImageCollection('COPERNICUS/S1_GRD')  
  
  .filterDate('2022-01-01', '2022-02-28')  
  
  .filter(ee.Filter.listContains('transmitterReceiverPolarisation', 'VV'))  
  
  .filter(ee.Filter.listContains('transmitterReceiverPolarisation', 'VH'))  
  
  .filter(ee.Filter.eq('instrumentMode', 'IW'))  
  
  .filter(ee.Filter.eq('orbitProperties_pass', 'DESCENDING'))  
  
  .filterBounds(boundary);
```

```
print(S1_PRS)
```

```
// Prepare Median
```

```
var S1_PRS_pc = S1_PRS.median();
```

```
// Convert to natural units (linear units, which can be averaged)
```

```
var S1_PRS_pc = ee.Image(10).pow(S1_PRS_pc.divide(10));
```

```
var S1_PRS_pc_Feats = S1_PRS_pc.select(['VH','VV']).clip(boundary);
```

```
// Reproject to WGS 84 UTM zone 54
```

```
var S1_PRS_pc_Feats = S1_PRS_pc_Feats.reproject({crs: 'EPSG:32754',scale: 30});
```

```
var S1_PRS_pc_Feats_VH = S1_PRS_pc_Feats
```

```
  .addBands((S1_PRS_pc_Feats.select('VH')));
```

```
var S1_PRS_pc_Feats_VV = S1_PRS_pc_Feats
```

```
.addBands((S1_PRS_pc_Feats.select('VV')));
```

```
// Display S1 inter-quartile range imagery
```

```
Map.addLayer(S1_PRS_pc_Feats.clip(boundary), {'bands': 'VV', min: 0,max:0.1}, 'Sentinel-1 IW  
VV');
```

```
Map.addLayer( S1_PRS_pc_Feats.clip(boundary), {'bands': 'VH', min: 0,max:0.1}, 'Sentinel-1 IW  
VH');
```

```
////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

```
//computing glcm metrcis
```

```
// Prepare Median
```

```
var S1_median = S1_PRS.median();
```

```
// Convert the image to a 32-bit integer type.
```

```
var S1_median_32 = S1_median.toInt();
```

```
// Add another band to S1_median_32.
```

```
var VH_median = S1_median_32.select('VH');
```

```
var VV_median = S1_median_32.select('VV');
```

```
var ratio = VH_median.divide(VV_median).rename('ratio');
```

```
S1_median_32 = S1_median_32.addBands(ratio);
```

```
// Define the GLCM texture parameters.
```

```
var textureParams = {
```

```
  size: 3,
```

```
  kernel: ee.Kernel.square(3, 'pixels'),
```

```
  average: true,
```

```
};
```

```

var S1_median_32 = S1_median.toInt8();

// Compute GLCM texture.

var glcm = S1_median_32.glcmTexture(textureParams).clip(boundary);

var glcm = glcm.reproject({crs: 'EPSG:32754',scale: 64,});

//var glcmBands = glcm.select(glcm.bandNames()).toBands();

var VV_contrast = glcm.select("VV_contrast");

var VH_contrast = glcm.select("VH_contrast");

var VV_ent = glcm.select("VV_ent");

var VH_ent = glcm.select("VH_ent");

var VV_var = glcm.select("VV_var");

var VH_var = glcm.select("VH_var");

var VV_corr = glcm.select("VV_corr");

var VH_corr = glcm.select("VH_corr");

var VV_diss = glcm.select("VV_diss");

var VH_diss = glcm.select("VH_diss");

////////////////////////////////////

// Load Sentinel-2 spectral reflectance data.

var s2 = ee.ImageCollection('COPERNICUS/S2_SR');

// Create a function to mask clouds using the Sentinel-2 QA band.

function maskS2clouds(image) {

  var qa = image.select('QA60');

  // Bits 10 and 11 are clouds and cirrus, respectively.

  var cloudBitMask = ee.Number(2).pow(10).int();

```

```

var cirrusBitMask = ee.Number(2).pow(11).int();

// Both flags should be set to zero, indicating clear conditions.

var mask = qa.bitwiseAnd(cloudBitMask).eq(0).and(

    qa.bitwiseAnd(cirrusBitMask).eq(0));

// Return the masked and scaled data.

return image.updateMask(mask).divide(10000);
}

// Filter clouds from Sentinel-2 for a given period.

var composite = s2.filterDate('2022-01-01', '2022-02-28')

    // Pre-filter to get less cloudy granules.

    .filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE', 10))

    .map(maskS2clouds)

    .filterBounds(boundary)

    .select('B2', 'B3', 'B4','B5','B6','B7','B8','B11', 'B12');

print(composite)

// Reproject to WGS 84 UTM zone 54s

var S2_composite = composite.median().reproject({crs: 'EPSG:32754', scale: 30});

// Check projection information

print('Projection, crs, and crs_transform:', S2_composite.projection());

// Display a composite S2 imagery

Map.addLayer(S2_composite.clip(boundary), {bands: ['B11', 'B8', 'B3'], min: 0, max: 0.3});

////////////////////////////////////

// Load SRTM

```



```

var SRTM = ee.Image("USGS/SRTMGL1_003");

// Clip Elevation

var elevation = SRTM.clip(boundary);

// Reproject 'elevation' to WGS 84 UTM zone 54s

var elevation = elevation.reproject({crs: 'EPSG:32754',scale: 30});

// Check projection information

print('Projection, crs, and crs_transform:', elevation.projection());

// Derive slope from the SRTM

var slope = ee.Terrain.slope(SRTM).clip(boundary);

// Reproject 'slope' to WGS 84 UTM zone 54

var slope = slope.reproject({crs: 'EPSG:32754',scale: 30});

// Check projection information

print('Projection, crs, and crs_transform:', slope.projection());

////////////////////////////////////

// Load ESA World Cover data

var dataset = ee.ImageCollection("ESA/WorldCover/v200").first();

// Clip the land cover to the boundary

var ESA_LC_2020 = dataset.clip(boundary);

// Extract forest areas from the land cover

var forest_mask = ESA_LC_2020.updateMask(

  ESA_LC_2020.eq(10) // Only keep pixels where class equals 2

);

// Display forests only

```

```

var visualization = {bands: ['Map'],};

Map.addLayer(forest_mask, visualization, "Trees");

////////////////////////////////////

// Merge the predictor variables

var mergedCollection = S2_composite

.addBands(S1_PRS_pc_Feats_VV

.addBands(S1_PRS_pc_Feats_VH

.addBands(VV_contrast

.addBands(VH_contrast

.addBands(VV_ent

.addBands(VH_ent

.addBands(VV_var

.addBands(VH_var

.addBands(VV_corr

.addBands(VH_corr

.addBands(VV_diss

.addBands(VH_diss

.addBands(elevation

.addBands(slope

.addBands(forest_mask))))))))))));

// Clip to the output image to Harare study area boundary.

var clippedmergedCollection = mergedCollection.clipToCollection(boundary);

print('clippedmergedCollection: ', clippedmergedCollection);

```

```

// Bands to include in the classification

var bands = ['B2', 'B3', 'B4', 'B5', 'B6', 'B7', 'B8', 'B11', 'B12', 'VV', 'VH', 'elevation', 'slope',
'Map', 'VV_contrast', 'VH_contrast', 'VV_ent', 'VH_ent', 'VV_var', 'VH_var', 'VV_corr', 'VH_corr',
'VV_diss', 'VH_diss'];

////////////////////////////////////

// Prepare training dataset

//The Global Ecosystem Dynamics Investigation (GEDI)

// GEDI's Level 2A Geolocated Elevation and Height Metrics Product (GEDI02_A) is primarily
composed of

// 100 Relative Height (RH) metrics, which collectively describe the waveform collected by GEDI.

// The original GEDI02_A product is a table of point with a spatial resolution (average footprint) of
25 meters.

// More information at https://lpdaac.usgs.gov/documents/986/GEDI02\_UserGuide\_V2.pdf

// Define a mask

var qualityMask = function(im) {

    return im.updateMask(im.select('quality_flag').eq(1))

        .updateMask(im.select('degrade_flag').eq(0));

};

// Import the "EDI02_A_002_MONTHLY: dataset

var dataset = ee.ImageCollection('LARSE/GEDI/GEDI02_A_002_MONTHLY')

    .map(qualityMask)

    .select('rh98').filterBounds(boundary);

// Create a pallete to visualize the dataset

var gediVis = {

```

```

min: 1,

max: 60,

palette: 'darkred,red,orange,green,darkgreen',

};

// Set the map center and visualize the dataset

Map.setCenter(138.6816,-35.2256, 10);

Map.addLayer(dataset, gediVis, 'rh98');

// Define projection and scale parameters

var projection = dataset.first().projection().aside(print);

var scale = projection.nominalScale().aside(print);

var mosaic = dataset.mosaic().setDefaultProjection({crs:projection, scale:scale});

// Sample 10,000 points from the dataset

var points = mosaic.sample({

  region: boundary,

  scale: scale,

  numPixels: 10000,

  projection: projection,

  geometries: true});

// Check the number of the training points

print(points.size());

print(points.limit(10));

// Display the training points

Map.addLayer(points);

```

```

// Add a random column (named random) and specify seed value for repeatability.

var datawithColumn = points.randomColumn('random', 27);

// Separate 70% for training, 30% for validation

var split = 0.7;

var trainingData = datawithColumn.filter(ee.Filter.lt('random', split));

// Print the training data

print('training data', trainingData);

var validationData = datawithColumn.filter(ee.Filter.gte('random', split));

// Print the testing (validation) data

print('validation data', validationData);

////////////////////////////////////

// Perform regression modeling using RF classifier

// Collect training data

var training = clippedmergedCollection.select(bands).sampleRegions({

  collection: trainingData,

  properties: ['rh98'],

  scale: 30 // Need to change the scale of training data to avoid the 'out of memory' problem

});

// Train a random forest classifier for regression

var classifier = ee.Classifier.smileRandomForest(50)

  .setOutputMode('REGRESSION')

  .train({

    features: training,

    classProperty: "rh98",

```

```

inputProperties: bands

});

//Run the classification and clip it to the boundary

var regression = clippedmergedCollection.select(bands).classify(classifier,
'predicted').clip(boundary);

// Load and define a continuous palette

var palettes = require('users/gena/packages:palettes');

// Choose and define a palette

var palette = palettes.colorbrewer.YIGn[5];

// Display the input imagery and the regression classification.

// get dictionaries of min & max predicted value

var regressionMin = (regression.reduceRegion({

  reducer: ee.Reducer.min(),

  scale: 30,

  crs: 'EPSG:32754',

  geometry: boundary,

  bestEffort: true,

  tileScale: 5

}));

var regressionMax = (regression.reduceRegion({

  reducer: ee.Reducer.max(),

  scale: 30,

  crs: 'EPSG:32754',

  geometry: boundary,

```

```

    bestEffort: true,

    tileScale: 5

  });

// Add to map

var viz = {palette: palette, min: regressionMin.getNumber('predicted').getInfo(), max:
regressionMax.getNumber('predicted').getInfo()};

Map.addLayer(regression, viz, 'Regression');

// Create the panel for the legend items.

var legend = ui.Panel({

  style: {

    position: 'bottom-left',

    padding: '8px 15px'

  }

});

// Create and add the legend title.

var legendTitle = ui.Label({

  value: 'TC Height',

  style: {

    fontWeight: 'bold',

    fontSize: '18px',

    margin: '0 0 4px 0',

    padding: '0'

  }

}

```

```

});

legend.add(legendTitle);

// create the legend image

var lon = ee.Image.pixelLonLat().select('latitude');

var gradient = lon.multiply((viz.max-viz.min)/100.0).add(viz.min);

var legendImage = gradient.visualize(viz);

// create text on top of legend

var panel = ui.Panel({

widgets: [

ui.Label(viz['max'])

],

});

legend.add(panel);

// create thumbnail from the image

var thumbnail = ui.Thumbnail({

image: legendImage,

params: {bbox:'0,0,10,100', dimensions:'10x200'},

style: {padding: '1px', position: 'bottom-center'}

});

// add the thumbnail to the legend

legend.add(thumbnail);

// create text on top of legend

var panel = ui.Panel({

```



```

widgets: [

  ui.Label(viz['min'])

],

});

legend.add(panel);

Map.add(legend);

// Zoom to the regression on the map

Map.centerObject(boundary, 11);

// Check model performance

// Get details of classifier

var classifier_details = classifier.explain();

// Explain the classifier with importance values

var variable_importance = ee.Feature(null, ee.Dictionary(classifier_details).get('importance'));

var chart =

  ui.Chart.feature.byProperty(variable_importance)

  .setChartType('ColumnChart')

  .setOptions({

    title: 'Random Forest Variable Importance',

    legend: {position: 'none'},

    hAxis: {title: 'Bands'},

    vAxis: {title: 'Importance'}

  });

// Plot a chart

```

```

print("Variable importance:", chart);

// Create model assessment statistics

// Get predicted regression points in same location as training data

var predictedTraining = regression.sampleRegions({collection:trainingData, geometries: true});

// Separate the observed (agb_GEDI) and predicted (regression) properties

var sampleTraining = predictedTraining.select(['rh98', 'predicted']);

// Create chart, print it

var chartTraining = ui.Chart.feature.byFeature(sampleTraining, 'rh98', 'predicted')

.setChartType('ScatterChart').setOptions({

title: 'Predicted vs Observed - Training data ',

hAxis: {'title': 'observed'},

vAxis: {'title': 'predicted'},

pointSize: 3,

trendlines: { 0: {showR2: true, visibleInLegend: true} ,

1: {showR2: true, visibleInLegend: true}}});

print(chartTraining);

// Compute Root Mean Squared Error (RMSE)

// Get array of observation and prediction values

var observationTraining = ee.Array(sampleTraining.aggregate_array('rh98'));

var predictionTraining = ee.Array(sampleTraining.aggregate_array('predicted'));

// Compute residuals

var residualsTraining = observationTraining.subtract(predictionTraining);

```

```

// Compute RMSE with equation and print the result

var rmseTraining = residualsTraining.pow(2).reduce('mean', [0]).sqrt();

print('Training RMSE', rmseTraining);

////////////////////////////////////

//Perform validation

// Get predicted regression points in same location as validation data

var predictedValidation = regression.sampleRegions({collection:validationData, geometries: true});

// Separate the observed (rh98) and predicted (regression) properties

var sampleValidation = predictedValidation.select(['rh98', 'predicted']);

// Create chart and print it

var chartValidation = ui.Chart.feature.byFeature(sampleValidation, 'predicted', 'rh98')

.setChartType('ScatterChart').setOptions({

title: 'Predicted vs Observed - Validation data',

hAxis: {'title': 'predicted'},

vAxis: {'title': 'observed'},

pointSize: 3,

trendlines: { 0: {showR2: true, visibleInLegend: true} ,

1: {showR2: true, visibleInLegend: true}}});

print(chartValidation);

// Compute RMSE

// Get array of observation and prediction values

```

```

var observationValidation = ee.Array(sampleValidation.aggregate_array('rh98'));

var predictionValidation = ee.Array(sampleValidation.aggregate_array('predicted'));

// Compute residuals

var residualsValidation = observationValidation.subtract(predictionValidation);

// Compute RMSE with equation and print it

var rmseValidation = residualsValidation.pow(2).reduce('mean', [0]).sqrt();

print('Validation RMSE', rmseValidation);

////////////////////////////////////

// Export the image, specifying scale and region.

Export.image.toDrive({

  image: regression,

  description: 'GEDI_new_forest_14m_optimised_2022',

  scale: 14,

  crs: 'EPSG:32754', // EPSG:32735 (WGS 84 UTM Zone 54S)

  maxPixels: 6756353855,

  region: boundary

```