InSAR Time series Analysis and Machine Learning for Ground Subsidence Monitoring and Susceptibility Mapping in Midvaal, South Africa

By: Thobani Maluleka
MLLTHO015

Department of Architecture, Planning and Geomatics: Division of Geomatics
Faculty of Engineering and the Built Environment

Dissertation submitted in fulfilment of the requirements for the Degree of
Master of Science in Engineering (Geomatics) at the
University of Cape Town

Supervisor:
Dr Siphiwe Matthews Mphuthi

January 2025

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

Copyright Declaration

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive licence granted to UCT by the author.

Authorship Declaration

I know that plagiarism is wrong and that I cannot use another author's work and pretend that it is my own. This thesis is my own work.

Each significant contribution to, and quotation in this thesis from the work(s) of other people has been attributed and has been cited and referenced.

The work in this thesis has not previously been published in whole or in part for the award of any degree.

Signed by candidate Date: 13/01/ 2025

Ethics Declaration

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA · UNIVERSITEIT VAN KAAPSTAD

2024/10/11

EBE/00757/2024

RE: Research Ethics Committee Project Approval Letter

Dear Thobani Maluleka,

Your application for ethics review of your project titled

InSAR Time-Series Analysis and Machine Learning for Ground Subsidence Monitoring and Susceptibility Mapping in Midvaal, South Africa

has been reviewed and evaluated by the

Faculty of Engineering & Built Environment Research Ethics Committee (REC).

Based on the information supplied your application has been successful and is approved.

You may proceed with your research project.

Please note that should:

(i) any serious or adverse effects to participants occur and/or,

(ii) aspect(s) of your current project change and/or

(iii) any unforeseen events that might affect continued ethical acceptability of the project occur then you should immediately report this to the approving REC. You may be required to submit an amendment to this application, in order to determine whether the changed aspects increase the ethical risks of your project.

Please note the following additional conditions associated with this approval:

(i) This application should have been processed via PSQ which would have been much quickest. Nevertheless, the student should include the research methodology in the research proposal.

Regards,

APG: School of Architecture, Planning & Geomatics

Research Ethics Committee

Abstract

The Midvaal region in Gauteng faces a significant challenge of ground subsidence, which poses significant risks, including sinkhole formation, disruptions to infrastructure, and threats to human settlements. Despite the severity of this issue, existing methodologies for monitoring and mapping ground subsidence susceptibility are limited by inefficiencies, labour-intensive processes, and reliance on sparse data sources, leading to delayed and inaccurate risk assessments.

This study addresses these challenges by integrating Interferometric Synthetic Aperture Radar (InSAR) time series analysis with Machine Learning (ML) techniques to monitor ground subsidence and develop ground subsidence susceptibility maps for Midvaal, South Africa. The key research objectives included evaluating the effects of various InSAR pre-processing and processing software on ground subsidence monitoring accuracy, examining the impact of Digital Elevation Model (DEM) vertical accuracy on ground subsidence monitoring accuracy, identifying the spatial and temporal (spatio-temporal) patterns of ground subsidence, and employing machine learning to create predictive models for ground subsidence susceptibility mapping. Furthermore, the study provides actionable recommendations for rural-urban planning and infrastructure management based on its findings.

This study assessed various InSAR processing methods, including SNAP, ISCE, HyP3, StaMPS, and MintPy. ISCE PS-InSAR was the most accurate for tectonic deformation monitoring, with minimal errors and stable velocity estimates. In contrast, HyP3 SBAS-InSAR performed well in detecting localised deformations but had higher error margins. Additionally, A semi-automated preprocessing tool, StoSAP, was developed using SNAP and Python-3 to streamline workflows for PS and SBAS methods.

DEM vertical accuracy effect was another critical factor examined. Among the AW3D30, COP30, SRTMGL1, and ASTER DEM evaluated, AW3D30 displayed the highest vertical and velocity modelling accuracy, particularly in less complex terrains, whereas ASTER performed poorly in areas with dense vegetation, urban infrastructure, and mining activities. These findings underscore the importance of high-quality DEMs in enhancing ground subsidence monitoring and susceptibility mapping.

Spatio-temporal analyses revealed that ground subsidence and uplift patterns were strongly influenced by land use, hydrological factors, vegetation cover, soil moisture, precipitation, and ground water storage anomalies. Subsequently, machine learning algorithms including Random Forest, XGBoost, LightGBM, and CNN were employed to create ground subsidence susceptibility maps and classify the region into five risk zones: Very Low, Low, Moderate, High, and Very High. Random Forest (RF) achieved the highest predictive accuracy with a Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R^2 score) of 3.621 mm/year, 5.039 mm/year, and 84.98% respectively,

identifying factors such as NDVI, lithology, and hydrological parameters as primary contributors to ground subsidence through SHapley Additive exPlanations and Frequency Ratio analysis. Furthermore, Random Forest classified the region into five risk zones, namely, Very Low, Low, Moderate, High, and Very High with values 0.1%, 87.9%, 9.4%, 1.7%, and 0.9% respectively indicating that the largest area in the study region is characterised by “Low” ground subsidence susceptibility. This suggests that the majority of the ground in the study region does not pose a significant threat to the environment. However, the combined occurrence of “High” and “Very High” susceptibility covers at least 2.6% of the area, equivalent to approximately 107.28 square kilometres about the size of Midrand, Gauteng, South Africa (152.87 square kilometres). This highlights a potentially alarming impact, given the significant proportion of high-risk areas.

This study advances the understanding and management of ground subsidence by integrating InSAR time series analysis, machine learning, and environmental modelling into a robust, scalable framework for real-time monitoring and risk mitigation. The proposed methodology not only enhances the detection and prediction of ground subsidence in the Midvaal region but also holds potential for broader geohazard applications. Future research should explore the integration of multi-sensor data, real-time environmental monitoring, and further model refinement to improve performance in complex terrain.

Acknowledgements

Primarily, I want to express my gratitude to Dr Siphiwe Mphuthi, whose counsel was essential in developing the research objectives and strategies as well as providing me with the funding and materials I needed to conclude my dissertation. Throughout the writing of this dissertation, I received insightful advice, constructive criticism, and support, for which I am very appreciative.

I would also like to express my gratitude to the European Space Agency (ESA), National Aeronautics and Space Administration (NASA), NASA Jet Propulsion Laboratory (JPL), Alaska Satellite Facility (ASF) Vertex, OpenScienceLab, Chief Directorate: National Geospatial Information (CD: NGI), Gauteng administration, Statistics South Africa, Department of Mineral Resources and Energy (DMRE), and Others for providing the datasets used in this research.

Furthermore, my sincere appreciation to the University of Cape Town staff and lecturers, and the Faculty of Engineering (Geomatics) for providing me with the undergraduate academic knowledge and skills that serves as a foundation for this research.

Finally, I want to pay tribute to my family, Miss N.I Maluleka, Mr B.S Ndlovu, Miss S.N Ndlovu; my amazing partner, Miss B.P Mahlangu and my parents, Mrs P.A Maluleka and Mr B.W Maluleka, for their unwavering support, patience, guidance and help throughout my academic career. When things went tough, your words of encouragement, support and prayers carried me through; I sincerely appreciate your affection and concern.

Dedication

This dissertation is dedicated to my daughter Melokuhle Ofentse Ariana Maluleka, whose love and admiration made this dissertation paper possible. Thank you for bearing with my erratic attention span amid my hectic schedules and for putting up with my time spent away from home while I finished this dissertation.

Thus, I hope that this dissertation will serve as inspiration for you to pursue a successful academic career and make contributions to a more sustainable future.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence
ALOS	Advanced Land Observing Satellite
APS	Atmospheric Phase Screen
ASF DAAC	ASF Distributed Active Archive Centre
ASF	Alaska Satellite Facility
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AUC	Area Under the Curve
AW3D30	Advanced Land Observing Satellite World 3D, 30 m spatial resolution
CD: NGI	Chief Directorate: National Geo-spatial Information
cGNSS	Continuous Global Navigation Satellite Systems
CNN	Convolutional Neural Networks
Combined PSDS	Time series phase change of both PS and DS targets
DEFO	Deformation
DEM	Digital Elevation Models
DFFE	Department of Fisheries, Forestry, and Energy
D-InSAR	Differential InSAR
DS	Distributed Scatterer
ECMWF	European Centre for Medium-Range Weather Forecasts
EM	Electromagnetic
ENU	East-North-Up
GLDAS	Global Land Data Assimilation System
GNSS	Global Navigation Satellite Systems
GS	Ground Subsidence
HRAO	Hartebeesthoek Radio Astronomy
InSAR	Interferometric Synthetic Aperture Radar
ISCE	InSAR Scientific Computing Environment
IW	Interferometric Wide
LULC	Land Use and Land Cover
LOS	Line of Sight
MCF	Minimum Cost Flow
MintPy	Miami InSAR Time series software in Python
ML	Machine Learning
MP	Measurement Point
NASA	National Aeronautics and Space Administration

NESD	Network Enhanced Spectral Diversity
NDVI	Normalised Difference Vegetation Index
PS	Persistent Scatterer
PSI	Persistent Scatterer Interferometry
RF	Random Forest
RMSE	Root Mean Square Error
SBAS	Small Baseline
SDG	Sustainable Development Goal
SLC	Single Look Complex
SNAP	Sentinel Application Platform
SNR	Signal-to-Noise Ratio
SRTMGL1	Shuttle Radar Topography Mission Global Level 1
StaMPS	Stanford Method for Persistent Scatters
STDV	Standard Deviation
TOPO	Topography
TOPS	Terrain Observation with Progressive Scan
TWI	Topographical Wetness Index
COP30	Copernicus DEM, thirty metres spatial resolution
WGS84	World Geodetic System 1984

1 Introduction

1.1 Background

Imagine waking up one morning to find a massive sinkhole has opened up in your neighbourhood, swallowing homes and roads alike. This nightmare scenario has become all too common in Gauteng, South Africa, where the occurrence of over 3,000 sinkholes has raised concerns about ***ground subsidence***, a phenomenon exacerbated by rapid urbanisation and increasing demands for underground water resources (Richardson, 2013). Ground subsidence is the gradual or sudden sinking of the earth's surface (Ashraf and Cawood, 2015; Nur et al., 2024; Ranjgar et al., 2021). It poses significant threats to infrastructure, water resources, the environment, and human settlements in the region. This phenomenon stems from a combination of *natural and anthropogenic factors.* Natural factors include topographic and geological processes, such as sediment compaction and the dissolution of underlying rock due to lithological conditions. Conversely, anthropogenic factors include excessive groundwater extraction for agriculture and domestic consumption, as well as mineral extraction in mining areas within the Gauteng region (Constantinou and Van Rooy, 2018; Montjane et al., 2020; Theron et al., 2017).

In Gauteng, ground subsidence contributes to the formation of sinkholes through the gradual dissolution of the underlying limestone rock (dolomite soil), creating voids that eventually collapse and form sinkholes (Constantinou and Van Rooy, 2018; Theron et al., 2017). Moreover, the extraction of groundwater for domestic consumption and commercial farming lowers the water table, leading to sediment consolidation due to reduced pore water pressure, and subsequent ground subsidence (Constantinou and Van Rooy, 2018; Montjane et al., 2020). Furthermore, ground subsidence disrupts natural drainage patterns and increases the risk of flooding, as the altered topography can lead to water accumulation in urban areas (F. Li et al., 2023; Pacheco-Martínez et al., 2015). These consequences have led to significant economic damage of around 1 billion-rand, destruction of infrastructure and loss of life in Gauteng (Richardson, 2013; Theron et al., 2017). Understanding and monitoring ground subsidence are crucial for mitigation strategies and sustainable development. These efforts help address topographic, hydrological, geological, and land-use challenges in the region.

Ground subsidence monitoring in Midvaal aligns with global sustainability efforts. Specifically, it supports SDG 11 ("Sustainable Cities and Communities") and Target 15.3 of SDG 15 ("Life on Land"). SDG 11 aims to create sustainable urban environments by preventing infrastructure damage and safeguarding human settlements. Additionally, this effort aligns with SDG 15's goal of combatting flooding, restoring degraded land, and achieving land degradation neutrality (The Global Goals, 2023). By reducing land degradation, habitat disruption and water resource contamination, addressing ground subsidence not only supports South Africa's commitment to the 2030 Agenda but also fosters resilience and environmental

sustainability while promoting rural-urban growth through ground subsidence monitoring (National Planning Commission, 2011).

Ground subsidence monitoring in the region has historically relied on geodetic and remote sensing techniques such as continuous Global Navigation Satellite System (GNSS), precise levelling, differential Interferometric Synthetic Aperture Radar (InSAR), and to a lesser extent microgravity, and extensometers (Ashraf and Cawood, 2015; Montjane et al., 2020; Theron et al., 2017). These techniques play a crucial role in accurately quantifying ground subsidence, aiding in early detection, and effective mitigation (Argyrakis et al., 2020; Cigna et al., 2021; Khan et al., 2014). Their essential role lies in offering ground-level data and complementing remote sensing techniques like InSAR, ensuring a comprehensive understanding of ground subsidence patterns and facilitating informed decisions for sustainable development and infrastructure protection.

However, each of these techniques has their own capabilities and limitations. Continuous GNSS (cGNSS) and precise levelling provide high accuracy in ground deformation measurement but suffer from sparse point distribution over large areas, which limits spatial resolution due to the time-consuming and expensive process of data retrieval and maintenance (Chaussard et al., 2013). In contrast, Differential InSAR (D-InSAR) provides measurements with high spatial resolution but is prone to signal decorrelation, which compromises measurement accuracy (Theron et al., 2017).

These limitations significantly impact ground subsidence monitoring and related geophysical research in the region. Sparse networks hinder the collection of precise continuous ground subsidence data, making it difficult to promptly detect and address ground subsidence issues. While signal decorrelation impedes broader geophysical and geodynamics research by reducing the accuracy of ground deformation measurements, essential for comprehending topographic and hydrological processes, geological processes and potential South African geological hazards (geohazards). Addressing these network limitations is vital to improve both ground subsidence monitoring and geophysical research in the region.

In the study region Midvaal, there is a scarcity of dependable continuous GNSS network stations, with only two cGNSS/TrigNET stations (HEID, VERG) managed by Chief Directorate: National Geo-spatial Information (CD: NGI). This limited size and sparse distribution of the reliable geodetic network in Midvaal, Gauteng restrict its applicability for vast region deformation/ground subsidence monitoring.

However, this limitation due to sparse points and limited size of geodetic measurements can be overcome by Radar satellites, Radar satellites employ the InSAR technology to effectively monitor ground subsidence across large regions (Kirui et al., 2022; Yuan et al., 2023). This active sensor allows for continuous day and night monitoring, regardless of weather conditions, and offers wide geographical coverage with high temporal and spatial resolution (Yalvac, 2020;

Zhang et al., 2022). Differential InSAR, a variant of InSAR, has been employed in the past for studying ground subsidence, which measures ground deformation by comparing differences between interferograms. While D-InSAR has been used in the past for studying ground subsidence, it is susceptible to signal decorrelations, such as spatial decorrelation and atmospheric disturbance, which can inaccurately compensate and limit the precision of ground subsidence measurements or velocities obtained from complete interferograms (Ho Tong Minh et al., 2020; Theron et al., 2017). However, the precision of D-InSAR data can be improved to a millimetre level by employing InSAR time series analysis techniques (Hakim et al., 2023b).

InSAR time series analysis approaches are devoted to dealing with signal decorrelations (e.g., atmospheric disturbances, spatial decorrelations and temporal decorrelations) in order to determine deformation and atmospheric signal (Ho Tong Minh et al., 2020); Different approaches for InSAR time series analysis deformation monitoring can be categorised into three groups: Persistent Scatterer (PS) InSAR for selecting highly coherent permanent or dominant scatterer, Small Baseline (SBAS) InSAR for selecting highly coherent distributed scatterers (DS), and the combination of PS , and SBAS or DS known as Combined PSDS (Ho Tong Minh et al., 2020). While InSAR time series analysis enhances ground deformation measurements accuracy, it reduces measurement point (MP) density to discrete points, and the velocities are relative to the chosen reference point, not absolute (trend offset present).

To address the problem of non-continuous measurements and trend offsets in ground subsidence measurements obtained through InSAR time series analysis methods, Machine Learning (ML) algorithms can be used. These algorithms enable the interpolation of points or surfaces based on processed InSAR ground subsidence/deformation measurements and their adjustment to higher order GNSS data through co-registration and filtering. Various machine learning algorithms, including Random Forest (RF), Convolutional Neural Network (CNN), eXtreme Gradient Boosting (XGBoost), Light Gradient-Boosting Machine (LightGBM), etc, have been employed for MP interpolation and optimization. Moreover, these machine learning algorithms have also been applied to assess factors contributing to ground subsidence, and ground subsidence susceptibility mapping worldwide (Achour et al., 2021; Bagheri et al., 2024; Lee et al., 2023; Sekkeravani et al., 2022).

Susceptibility mapping for ground subsidence employs various techniques that integrate spatial data such as InSAR time series analysis measurements, topographic information, geological maps, and hydro-geological parameters (Zhao et al., 2024). Traditional susceptibility mapping methods like Analytic Hierarchy Process (AHP), Fuzzy AHP, Multi-Criteria Decision-Making Analysis (MCDMA), Weight of Evidence (WoE), Fuzzy Logic, and Neuro-Fuzzy approaches, each possess distinct advantages and limitations. For example, AHP may introduce subjectivity due to expert judgment, while MCDMA excels at integrating criteria but can become unwieldy with increasing complexity (Ashraf and Cawood, 2015; Choi et al., 2010; Lee et al., 2012; Yalcin et al., 2011). In contrast, data-driven methods, such as machine learning algorithms like

neural networks and ensemble methods are increasingly preferred for their capacity to handle extensive datasets, explore the factors contributing to ground subsidence, and reveal subtle patterns overlooked by traditional methods. This approach produce susceptibility maps that categorise areas into different risk levels for ground subsidence, and other hazards.

Therefore, the integration of InSAR time series analysis and machine learning techniques represents a groundbreaking advancement in the field of accurate ground subsidence monitoring and susceptibility mapping (Fadhillah et al., 2022; Gharechaee et al., 2023). This integration promises high spatial and temporal resolution, and accurate ground deformation measurements. As noted by Bagheri-Gavkosh et al. (2021), remotely sensed methods are now used in 38.07% of ground subsidence monitoring studies, demonstrating a significant trend towards the adoption of InSAR and similar technologies due to the time-consuming, labour-intensive, and costly nature of traditional ground-based methods.

By harnessing the power of SAR satellite imagery, InSAR time series analysis, and ML advanced algorithms, this approach promises to revolutionise our understanding of ground subsidence patterns and facilitate more effective mitigation strategies. However, despite its potential benefits, this integration has been unexplored in Midvaal, Gauteng depicted in Figure 1.1 due to reasons outlined in section-1.2.

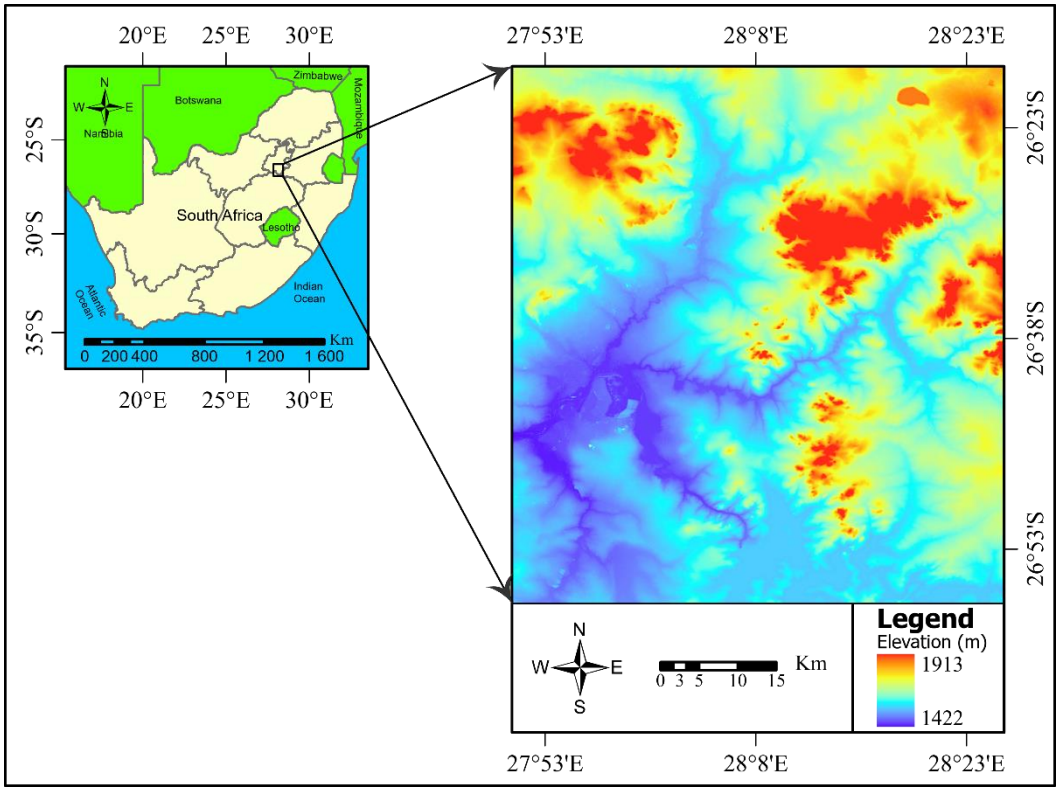


Figure 1.1: Geographical location and topographic elevation map of Midvaal, South Africa.

1.2 Problem Statement

The Midvaal region in Gauteng faces a significant challenge of ground subsidence, which poses substantial risks, including the formation of hazardous sinkholes, disruptions to critical infrastructure, and threats to human settlements. Compounding this issue is the lack of a comprehensive and efficient methodology for ground subsidence monitoring and susceptibility mapping. Existing methodologies, often reliant on limited data sources or labour-intensive processes, hinder the accurate and timely assessment of ground subsidence patterns and risks.

The integration of InSAR time series analysis and machine learning techniques presents an untapped opportunity for ground subsidence monitoring and susceptibility mapping in Midvaal. These advanced technologies can offer large coverage, highly accurate and timely data, yet their potential remains underutilised due to a historical reliance on conventional methods, limited awareness of these innovative techniques, and a lack of comprehensive initiatives to adopt them. Additionally, the absence of trained personnel and infrastructure tailored to this approach further impedes its implementation.

Moreover, challenges remain with InSAR time series analysis itself. The availability of multiple tools and techniques for InSAR data processing leads to inconsistent results, and there has been no, or little assessment or comparison of the accuracy and uncertainties associated with these tools to determine the best option for ground subsidence monitoring. Furthermore, the selection of DEMs and the propagation of DEMs vertical errors into ground subsidence susceptibility maps have not been adequately evaluated, despite their critical importance in ensuring accurate results.

Addressing these barriers is crucial to unlocking the full potential of InSAR time series analysis and machine learning, ensuring the safety, sustainability, and resilience of Midvaal's rural-urban development. By leveraging these technologies and addressing the identified gaps, Midvaal can enhance its ability to monitor ground subsidence accurately, respond to risks promptly, and plan for future urban development more effectively.

1.3 Aim and Objectives

The aim of this research is to develop an InSAR time series analysis framework, integrated with machine learning, for monitoring ground subsidence and generating susceptibility maps in Midvaal, South Africa. This approach seeks to improve the detection and understanding of ground subsidence patterns while assessing vulnerable areas for better risk management.

Therefore, the primary objectives of this dissertation paper are to:

- Investigate available InSAR time series analysis techniques and open-source processing tools.

- Develop tools and software to complement open-source InSAR pre-processing and processing software.
- Assess the effects of various InSAR pre-processing and processing technique on ground subsidence monitoring accuracy, clarifying their impact on the research outcomes.
- Evaluate the impact of Digital Elevation Model (DEM) vertical accuracy on ground subsidence monitoring accuracy, providing context for the importance of this factor.
- Investigate the rates and spatio-temporal patterns of ground subsidence in Midvaal, Gauteng, with clear criteria for assessing the success of this investigation.
- Integrate machine learning algorithms with InSAR time series analysis data to create predictive models for ground subsidence susceptibility mapping in Midvaal, highlighting the potential advantages of this integration.
- Provide practical recommendations for rural-urban planning and infrastructure management based on the obtained results of ground subsidence monitoring and susceptibility mapping, specifying the stakeholders and decision-makers who can benefit from these recommendations.

1.4 Research Questions

The research questions are:

- What are the available InSAR time series analysis techniques for monitoring ground subsidence, and what tools facilitate continuous big-data monitoring using these techniques?
- How does the choice of InSAR processing tools affect the accuracy of ground subsidence monitoring?
- What is the influence of employing specific InSAR time series analysis technique on accuracy, considering the suitability of different scatterers for ground deformation monitoring and the implications of technique selection?
- What are the effects of DEM's vertical accuracy on the accuracy of computed or predicted ground deformation measurements?
- What are the capabilities and limitations of Sentinel-1 and InSAR time series techniques in accurately detecting ground deformation?
- What are the rates and spatio-temporal patterns of ground subsidence in Midvaal, South Africa?
- What are the contributing factors to ground subsidence in Midvaal, South Africa?

- How can ground subsidence susceptibility maps be created to identify areas prone to ground subsidence in Midvaal, South Africa?

1.5 Research Scope and Limitations

This dissertation focuses on the application of InSAR time series analysis and machine learning for ground subsidence monitoring and susceptibility mapping in Midvaal, South Africa. The choice of Midvaal is due to limitations in processing resources and time, as well as its normal distribution of geological features and land cover, which facilitates fair and consistent analysis. The study will use Sentinel-1 C-band satellite data, offering a reliable and consistent data source for long-term monitoring. The analysis will cover the period from January 2019 to December 2021 and incorporate available ground-measurement data for validation and comparison.

Additionally, it is important to note that no study will be conducted using sinkhole data, and no correlation will be drawn between ground subsidence and sinkhole formation. This limitation arises due to issues with data sharing and transfer, as access to sinkhole data was denied by the Council for Geoscience.

1.6 Research Outline

This dissertation begins with an introduction that establishes the background, outlines the problem statement, and sets forth the research aim and objectives. This section provides the necessary framework for understanding the context and significance of the study. The subsequent chapter, the literature review, offers an in-depth foundation of theory and examination of existing studies on ground subsidence, InSAR time series analysis, digital elevation models, and machine learning applications in ground subsidence monitoring and susceptibility mapping. It serves to lay a strong theoretical foundation by reviewing current knowledge and methods relevant to the research.

The subsequent chapter, the methodology and materials chapter describes the study area, datasets, and computational tools employed. It provides a comprehensive discussion on the processing workflows for InSAR data, including the use of software such as ISCE, SNAP, HyP3, MintPy, and StaMPS, while also covering the effects of InSAR pre-processing and processing. The chapter further explores the role of DEMs and examines the spatial and temporal patterns of ground subsidence in Midvaal. This is followed by an analysis of ground subsidence susceptibility mapping, highlighting the factors that contribute to ground subsidence and identifying spatial patterns of susceptibility within the study area.

The subsequent chapter, the results and discussion chapter presents the key findings, which include the observed ground subsidence patterns derived from InSAR time series analysis and the development and performance of machine learning models for ground subsidence

susceptibility mapping. A comparative assessment of the different methods and their accuracy is also included.

The final chapter concludes by summarising the primary findings and their contributions to the field of ground subsidence monitoring and susceptibility mapping. It also offers recommendations for future research, addressing the limitations encountered and identifying potential areas for advancement.

2 Literature Review

This chapter presents a thorough review of literature (foundation of theory, empirical research, and identification of the research gaps) relevant to ground subsidence monitoring and susceptibility mapping, with a specific emphasis on InSAR time series analysis and Machine Learning. It begins by contextualising ground subsidence and the role of InSAR as a pivotal tool in monitoring this phenomenon. The discussion advances to explore various InSAR time series analysis techniques, essential pre-processing and processing tools, and validation methods using cGNSS data. Finally, the review critically assesses InSAR technology's advantages, limitations, and challenges.

The chapter then shifts to ground subsidence susceptibility mapping, highlighting the significance of factor prioritisation and multicollinearity analysis. It also examines the integration of machine learning, with a focus on model explainability through SHapley Additive exPlanations and statistical methods. Moreover, it reviews the impact of DEMs and ground subsidence related factors, addressing both current challenges and potential advancements. Additionally, case studies, are presented to identify research gaps, and illustrate practical applications of the InSAR time series analysis and machine learning methods. Finally, the chapter concludes with a summary of key insights and research gaps derived from the literature.

2.1 Ground Subsidence Overview

Ground subsidence is a complex geophysical phenomenon caused by both natural processes and human activities, with its impacts varying across spatial and temporal scales. Understanding its underlying causes, impacts, and monitoring techniques is critical for effective mitigation and management strategies.

2.1.1 Causes of Ground Subsidence

Apart from the previously discussed human and natural factors such as excessive groundwater extraction, agricultural expansion, and rapid urbanisation, several other key contributors play a significant role in ground subsidence. These include *human factors* such as land use and land cover changes due to urbanisation and agriculture, underground oil and gas extraction, mineral extraction from mining, and distance from roads (Bagheri-Gavkosh et al., 2021). Mining activities induce ground subsidence by creating voids during underground mineral extraction as illustrated by Figure 2.1 (Behera and Singh Rawat, 2023). Additionally, oil and natural gas extraction can contribute to ground subsidence by depleting underground reservoirs (Cian et al., 2019). Moreover, the distance from roads can induce ground subsidence through activities like cut and fill operations during road construction, altering soil structures and contributing to ground deformation (Edmonds, 2018). Furthermore, the construction of heavy infrastructure,

such as buildings and highways, can also induce local ground subsidence due to soil compression (M. Sun et al., 2023).

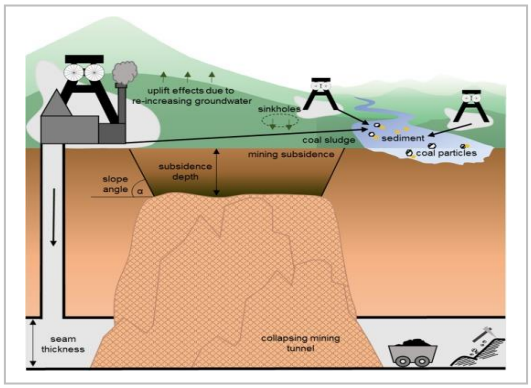


Figure 2.1: Ground subsidence induced by mineral extraction (Source: Maaß and Schüttrumpf, 2018).

In contrast, *natural factors* include lithology, distance from rivers, precipitation, temperature, topographic factors, and tectonic plate movements. Specific lithological conditions can increase groundwater infiltration and the dissolution of subsurface rocks, creating voids that contribute to ground subsidence and sinkhole formation (Bianchini et al., 2019; Na et al., 2021). Areas far from rivers may experience limited natural groundwater replenishment, while areas close to rivers may be subjected to erosion (Han et al., 2017; Saether and Caritat, 2020).

Moreover, precipitation can influence ground subsidence by affecting groundwater levels (H. Wang et al., 2024), while temperature can indirectly affect ground subsidence through soil expansion and contraction, leading to ground movement. This is supported by Bagheri-Gavkosh et al. (2021), who note that dry temperate regions often experience higher ground subsidence velocities, while warm temperate regions typically have lower rates. Additionally, semi-arid climates with frequent droughts are more susceptible to ground subsidence due to groundwater depletion (Aslan and Cakir, 2019).

Furthermore, topographic factors, including ground elevation, slope, aspect, topographical wetness index, and curvature, can affect ground subsidence by influencing groundwater dynamics, soil stability, and stress variations (Najafi et al., 2020; Ranjgar et al., 2021). Lastly, Tectonic plate movements can also induce ground subsidence through processes like crustal compression, tension, and lateral fault movements, causing subsurface faulting or collapse. However, these account for a relatively small proportion of ground subsidence cases over decades due to their slow and infrequent occurrence (Bagheri-Gavkosh et al., 2021).

2.1.2 Impacts of Ground Subsidence

Regardless of the cause, ground subsidence poses significant risks and can profoundly impact both the environment and human societies. These impacts of ground subsidence are significant and wide-ranging, categorised into geological, hydrogeological, environmental, and socio-economic impacts (Bagheri-Gavkosh et al., 2021; Zhao et al., 2023). *Geological impacts*

include the development of ground fissures leading to ground failure, structural failure, and infrastructure destruction, such as the destruction of geodetic control points and buildings. Sinking ground can cause the cracking and shifting of buildings, roads, and pipelines, resulting in costly repairs and potential safety hazards as illustrated in Figure 2.2. Ground subsidence can also disrupt underground utilities, such as water and sewer lines, leading to leaks, contamination, or complete system failure (Bagheri-Gavkosh et al., 2021; Salehi Moteahd et al., 2019).

Figure 2.2: Ground subsidence in Gauteng resulting in destruction of road infrastructure and sanitation system (Image by Delwyn Verasamy).

In contrast, the *environmental impacts* include increased flooding, particularly in coastal areas, as ground subsidence exacerbates the risk of coastal flooding and saltwater intrusion, increasing the vulnerability of coastal communities (Brempong et al., 2023). Ground subsidence can also alter the hydrological dynamics of an area, affecting natural water systems, including rivers, lakes, and wetlands (Alogayell et al., 2021; Modoni et al., 2013). Moreover, *hydrogeological impacts* encompass changes in well casings, soil properties, seepage and stress fields, and pore water characteristics. These effects can damage wells, reduce aquifer storage, and alter the hydro-geological landscape (P. Li et al., 2023).

Lastly, the *socio-economic impacts* include damage to historical sites, failures in drainage systems, impediments to sustainable economic development, harm to surface and subsurface infrastructure, and the destruction of pipelines. Structural deterioration of buildings, roads, and bridges, difficulties in foundation engineering, and the occurrence of cracks and structural failures in civil constructions are also significant concerns. Additionally, the loss of life can occur when structural failures happen in populated regions (Josset et al., 2024; Kok and Costa, 2021; Stouthamer et al., 2020).

Given the potential consequences of ground subsidence, monitoring, susceptibility mapping and mitigating this phenomenon are of utmost importance. Regular monitoring and susceptibility mapping allows for early detection of ground subsidence and helps assess its rate and spatial extent. It is therefore apparent that effective ground subsidence susceptibility assessment and monitoring are crucial for addressing and mitigating this geohazard and other related geohazards.

2.1.3 Monitoring Techniques of Ground Subsidence

Ground subsidence monitoring techniques can be categorised into ground-based and remote sensing (airborne) techniques. Ground based techniques provide accurate measurements of ground subsidence and deformation. However, they have several limitations, including small monitoring coverage, high costs, time consumption, and potential inaccuracies if the geodetic reference point is unstable. Examples of such ground-based techniques include GNSS, precise levelling, Very Long Baseline Interferometry (VLBI), Satellite Laser Ranging (SLR), and Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS). Other methods include tripod-based Light Detection and Ranging (LiDAR), ground-based InSAR, extensometers, and microgravity (Cai et al., 2023; Han et al., 2023).

Conversely, remote sensing techniques provide extensive monitoring coverage and efficiency but require significant resources and expertise for data extraction. Examples of such remote sensing techniques include satellite InSAR, airborne LiDAR, photogrammetry, and satellite gravimetry. Among these, InSAR delivers a broader spatial coverage compared to LiDAR, satellite gravimetry and photogrammetry. Additionally, InSAR's high temporal resolution, characterised by its shorter revisit period, makes it a preferred method for monitoring ground subsidence, offering more frequent and detailed observations than the other techniques (Auriol and Tourain, 2010; Cohen, 1973; Dehne et al., 2009; Di Stefano et al., 2022; Fabris et al., 2014; Hofmann-Wellenhof et al., 2008; Ignjatović Stupar et al., 2020; Linder, 2003; Monterroso et al., 2020; Ogundare, 2015; Pan and Tian, 2016; Sheikh, 2020; Sośnica et al., 2015; Touboul et al., 2016; Wahyudi et al., 2021; Willis et al., 2006; Zaki et al., 2024; Zhang and Zhu, 2023).

Thus, integrating these techniques, particularly GNSS and satellite InSAR, provides a robust framework for comprehensive ground subsidence monitoring. By combining GNSS's high temporal resolution and accuracy with InSAR's extensive spatial coverage and frequent revisit capabilities, this approach capitalises on the strengths of both techniques. Moreover, this synergy facilitates a detailed analysis of both rapid and gradual ground subsidence phenomena, enhancing the overall understanding of ground subsidence dynamics.

2.2 InSAR and InSAR time series Analysis

Interferometric Synthetic Aperture Radar and InSAR time series analysis techniques has revolutionised the monitoring and measurement of ground deformation, providing unprecedented insights into both rapid and gradual ground subsidence through spaceborne SAR images. By harnessing the combined strengths of InSAR and InSAR time series analysis, which offer high temporal resolution, extensive spatial coverage, and high precision measurements, this section will explore the history and fundamentals of InSAR, various InSAR time series analysis techniques, an overview of InSAR pre-processing and processing tools, and applications of InSAR time series analysis for ground subsidence monitoring.

The development of SAR images for earth observation originates from the launch of Seasat in 1978. Initially designed to measure and monitor variations in ocean characteristics, Seasat marked the beginning of a significant transformation in the field of space SAR. Since then, SAR development has expanded globally, involving ten government agencies worldwide and substantial contributions from private entities, as depicted in Figure 2.3, which provides an overview of space SAR missions (Ferretti, 2014; Flores et al., 2019; Woodhouse, 2006).

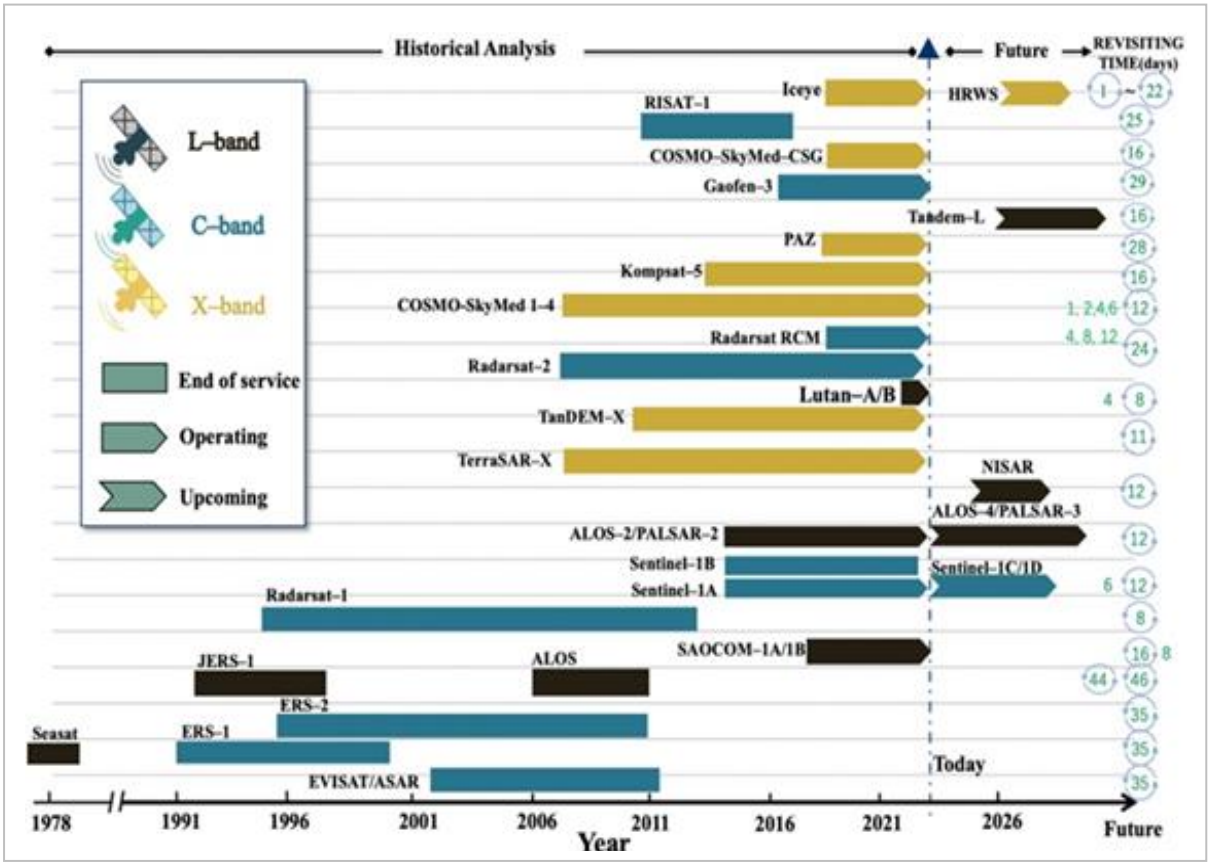


Figure 2.3: Synthetic Aperture Radar satellite missions (Source: S. Wu et al., 2023).

Based on the mission's success, SAR has proven to be an exceptionally versatile and powerful imaging radar technique, capable of characterising a wide range of natural disasters and Earth processes, including ground deformation, biomass estimation, vegetation monitoring, hydrology studies, and ice or glacial changes. Ground deformation assessment, for example, is achieved by analysing interferometric phase differences between radar images acquired at different times, as illustrated in Figure 2.4. This interferometric phase is highly sensitive to changes in the Earth's surface, such as ground subsidence and uplift (Orellana et al., 2022; Wang et al., 2023; S. Wu et al., 2023).

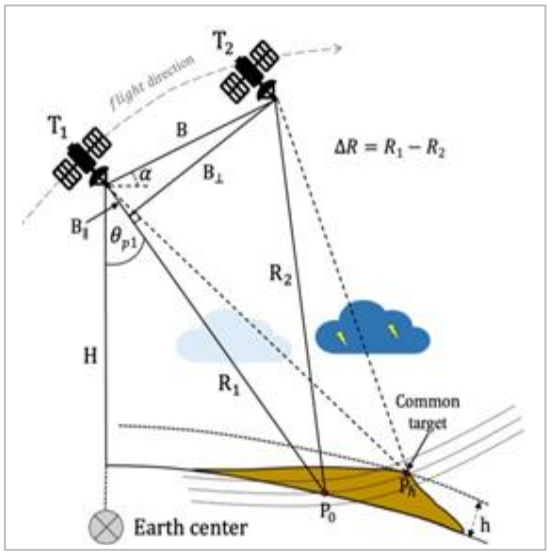


Figure 2.4: Range measurements by a InSAR sensor over time (Source: S. Wu et al., 2023).

A key factor in SAR's effectiveness is its unique side-looking geometry, where a satellite-mounted antenna moves in the azimuth (flight) direction perpendicular to its orbit, creating a 'footprint' on the ground, with the observation direction along the 'line of sight' (LOS) known as the 'slant-range direction,' as depicted in Figure 2.5. This side-looking geometry significantly contributes to SAR's ability to mitigate signal ambiguity, capture and understand a broad spectrum of Earth's phenomena (Ferretti, 2014; Wang et al., 2023; Woodhouse, 2006).

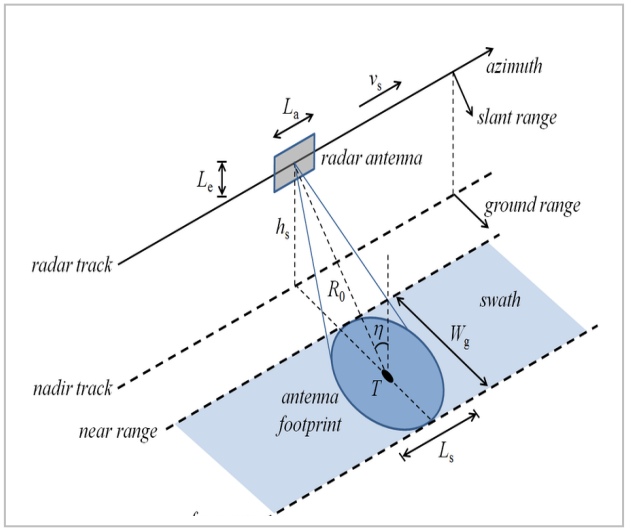


Figure 2.5: Synthetic Aperture Radar geometry (Source: Martone, n.d.).

Operating within the microwave spectrum, typically ranging from 0.3 to 300 GHz (1 mm to 1 m) as shown in Appendix A, InSAR can penetrate clouds and function both day and night. Moreover, InSAR satellites use various microwave bands tailored to specific applications, as EM waves are most sensitive to objects whose size is roughly equivalent to or larger than the EM wavelength, rendering smaller objects nearly invisible (Emery and Camps, 2017; Woodhouse, 2006).

Table 1: An overview of the various SAR microwave bands (source: Woodhouse, 2006)

Standard Radar Nomenclature	
Band	**Frequency Range**
HF	3 - 30 MHz
VHF	30 - 300 MHz
P (UHF)	0.3 - 1 GHz
L	1 - 2 GHz
S	2 - 4 GHz
C	4 - 8 GHz
X	8 - 12 GHz
Ku	12 - 18 GHz
Ka	18 - 27 GHz
V	27 - 40 GHz
W	75 - 110 GHz
Mm	110 - 300 GHz

Additionally, different radar frequencies offer varying penetration depths into surfaces. For example, shorter wavelengths (X and C bands) are well-suited for sensing canopy and leaf structures, while longer wavelengths (S and L bands) penetrate deeper, making them ideal for imaging tree trunks and limbs, as illustrated in Figure 2.6. Moreover, ground surface properties such as roughness and moisture content also significantly influence radar signal return, known as backscatter, which directly affects radar image characteristics (Emery and Camps, 2017; Flores et al., 2019).

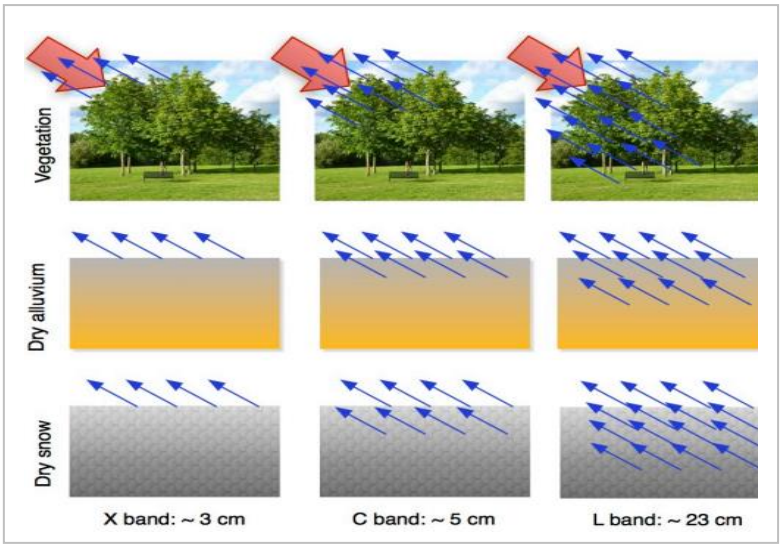


Figure 2.6: SAR penetration depths versus radar wavelengths (Image by European Space Agency).

Furthermore, radar signals can experience decorrelation due to various factors. Spatial decorrelation occurs when the radar resolution is insufficient to resolve small-scale surface features, leading to a loss of coherence in the radar return signal. While, Temporal decorrelation, on the other hand, arises from changes in the Earth's surface between radar

acquisitions, such as vegetation growth or ground movement, which affects the phase coherence between images. Lastly, atmospheric disturbances, such as moisture and turbulence, can induce phase variations, complicating the accurate measurement of ground deformation and other surface characteristics. These challenges can be effectively addressed by employing InSAR time series analysis techniques.

2.2.1 InSAR Time series Analysis Techniques

InSAR time series analysis techniques are devoted to dealing with atmospheric disturbances, spatial decorrelations and temporal decorrelations to determine deformation and atmospheric signals from the interferometric phase (Ferretti, 2014; Ho Tong Minh et al., 2020). The interferometric phase is composed of phase contributions related to ground deformation (ϕ_{defo}), residual topography (ϕ_{topo}), atmospheric phase screens (ϕ_{aps}), and noise (ϕ_{noise}). Assuming the two images have been coregistered and the phase contributions due to orbital errors and terrain topography have been compensated for on a reference grid, the wrapped interferometric phase can be expressed as shown in equation (2.1) below:

$$\Delta\phi = \phi_{topo} + \phi_{defo} + \phi_{aps} + \phi_{noise} + 2k\pi \quad (2.1)$$

In many geodetic and geophysical applications, the phase contribution of interest is the phase due ground deformation. Differential InSAR directly addresses ground deformation by comparing SAR images acquired at different times. This technique subtracts the phase of one image from another to measure surface deformation with high accuracy, typically ranging from metres to centimetres. However, D-InSAR does not directly resolve atmospheric phase screens and noise, and can be affected by temporal decorrelation, spatial decorrelation, and phase unwrapping errors, which can impact measurement accuracy (Ding and Huang, 2011; Fan et al., 2021; Tampubolon et al., 2023; Theron et al., 2017). Therefore, for continuous and reliable deformation monitoring, it is necessary to compensate for the other phase contributions and artefacts that are not of interest from the measured residual phase using InSAR time series analysis.

InSAR time series analysis techniques can be categorised into two main groups as discussed below:

- **Persistent Scatterer InSAR**

  Persistent Scatterer InSAR is a technique that focuses on identifying and analysing stable, persistent scatterers in a series of InSAR images. These scatterers maintain consistent reflection over time, enabling precise deformation measurements, often with millimetre accuracy per year. PS-InSAR is particularly useful in urban environments where man-made structures serve as strong reflectors, providing high-density measurements. The technique excels in areas with long-term, slow-moving ground

deformation, and is less sensitive to temporal and spatial decorrelation compared to D-InSAR.

The initial selection of PS points employs the amplitude dispersion technique, which measures phase stability in high Signal-to-Noise Ratio (SNR) regions (Hooper et al., 2012). A high SNR allows for the precise estimation of parameters in equation (2.1); And the amplitude dispersion (D_a) is then mathematically expressed by equation (2.2) below.

$$D_a = \frac{\sigma_a}{\mu_a} \tag{2.2}$$

where σ_a is the amplitude standard deviation and μ_a is the amplitude mean of a pixel over time. Preliminary identification of PS candidates uses amplitude dispersion with a user-defined threshold, commonly $D_a = 0.4$, to isolate potential PS pixels. And the final PS points selection is chosen to maximise the complex ensemble coherence ($\hat{\gamma}_{pq}$) in time, given by equation (2.3) below:

$$\hat{\gamma}_{pq} = \frac{1}{N-1}\sum_{n=1}^{N-1} exp(i\Delta w_{pq}^{n}) \tag{2.3}$$

where Δw_{pq}^{n} is the residual phase representing differences between the model-based and actual atmospheric phase screens, and N is the number of InSAR acquisitions. PS candidates with coherence below a user-defined threshold (e.g., 0.7) are excluded.

Additionally, while PS-InSAR accurately measures high densities of PS points, it requires a large number of images and sophisticated processing algorithms, making it computationally demanding (Bianchini et al., 2019; Cian et al., 2019; Cigna et al., 2021; Hakim et al., 2023b). Moreover, PS-InSAR is less effective in intensely vegetated areas primarily due to spatial and temporal decorrelation; however, this limitation can be mitigated by employing SBAS-InSAR as discussed below.

- **Small Baseline Subset InSAR**

  Small Baseline Subset InSAR technique reduces the spatial and temporal decorrelation by using a series of InSAR images with small temporal and spatial baselines. This method divides the data into subsets where the differential interferograms are formed, reducing noise and improving the reliability of ground deformation measurements. SBAS-InSAR can effectively monitor both slow and fast deformation rates and is less sensitive to decorrelation and atmospheric artifacts compared to traditional D-InSAR and PS-InSAR. It is particularly useful in areas where persistent scatterers are sparse, such as rural or intensive vegetated regions (Ajayi et al., 2023; Cigna et al., 2021; Kirui et al., 2022; Yunjun et al., 2019).

  Similar to PS-InSAR, SBAS points are initially selected based on amplitude dispersion difference as mathematically expressed by equation (2.4) below (Kirui et al., 2022).

$$D_{\Delta a} = \frac{\sigma_{\Delta a}}{\mu_{\Delta a}} \tag{2.4}$$

And the final selection is based on temporal coherence, which is a function of measured phase (ϕ), the adjusted phase ($\hat{\phi}$) , and design matrix (A) as shown by equation (2.5) below.

$$\gamma_{temp} = \frac{1}{M} exp(j(\phi - A\hat{\phi})) \tag{2.5}$$

Moreover, combined PS and SBAS techniques, known as Combined PSDS, integrate these approaches for comprehensive ground deformation monitoring. The implementation of these techniques is facilitated through InSAR time series analysis processing tools, which require InSAR pre-processing tools to prepare the data, as discussed below.

2.2.2 InSAR Pre-processing and Tools

Several InSAR datasets pre-processing software options are available to enable extraction of valuable insights from SAR data, with commercial offerings such as Gamma Remote Sensing (GAMMA) and SARscape (formerly known as ENVI SARscape). On the other hand, open-source alternatives include SentiNel Application Platform (SNAP), InSAR Scientific Computing Environment (ISCE), Generic Mapping Tools for Synthetic Aperture Radar (GMTSAR) and Repeat Orbit Interferometry Package (ROI_PAC), Alaska Satellite Facility (ASF) HyP3 (pronounced “hype”), as well as Python GMTSAR (PyGMTSAR).

Based on the dissertation’s goals emphasising promotion of open-source software, availability of computational resources, and a review of widely used InSAR pre-processors, ISCE, SNAP, and HyP3 were selected for discussion below.

- **InSAR Scientific Computing Environment**

  InSAR Scientific Computing Environment is a robust software framework developed by NASA’s Jet Propulsion Laboratory (JPL) for processing InSAR data. It is renowned for its capability to extract precise ground deformation information from InSAR data, providing accurate deformation measurements, displacement maps, and coherence estimation. As an open-source platform, ISCE fosters collaboration among researchers, benefiting from an active user and developer community that ensures regular updates and support. However, ISCE has a steep learning curve, requires significant computational resources, and involves complex parameter configurations, which may limit its accessibility to novice users and those without access to high-performance computing facilities. The version discussed in this dissertation is ISCE version 2.6, commonly referred to as ISCE2 (Fattahi et al., 2017; Rosen et al., 2012).

- **SentiNel Application Platform**

  The SentiNel Application Platform is a robust software developed by the European Space Agency (ESA) and European Commission (EC) for processing and analysing data from Earth observation satellites, particularly the Sentinel missions. SNAP's architecture is grounded in the Graph Processing Framework (GPF), facilitating the creation of complex processing chains through a graphical interface, and enhancing flexibility and scalability for tailored workflows. SNAP supports various data formats including Sentinel-1, Sentinel-2, Landsat, and more, and it also caters to a wide range of applications in geospatial research. Additionally, its user-friendly interface simplifies the creation and execution of processing chains, aiding users in navigating its extensive capabilities. As an open-source platform, SNAP benefits from ongoing development and global user support. Nonetheless, some tasks in SNAP may demand substantial computational resources, and while its interface is accessible, mastering all features may require time (Foumelis et al., 2018).

- **Alaska Satellite Facility HyP3**

  The Alaska Satellite Facility HyP3 (Vertex) is a user-friendly cloud-based geospatial platform tailored for efficiently downloading and processing Sentinel-1 datasets. ASF Vertex streamlines the generation of InSAR products such as interferograms, displacement maps, and look vectors, eliminating the complexities of processing and data retrieval. Leveraging ASF's HyP3 processing platform with GAMMA Software, the service delivers processed products in UTM-projected GeoTIFF format, accompanied by detailed guidance in the ASF Sentinel-1 InSAR Product Guide. To initiate ASF HyP3 On-Demand processing, users navigate the ASF Data Search - Vertex portal to select scenes, utilising various search functionalities like Geographic Search, SBAS Search, and Baseline Search. Upon scene selection, users submit processing requests through the On-Demand queue, with up to 10,000 jobs allowed per month. Subsequently, monitoring processing status is facilitated through the On-Demand Products search, with completed products typically available for download within an hour and accessible for two weeks post-processing.

  In essence, ASF's Vertex platform streamlines Sentinel-1 dataset processing for InSAR products, fostering accessibility and usability across diverse geospatial applications through its intuitive interface and efficient processing capabilities. Nevertheless, it is equally important to note that some countries will be prohibited from using this service if they are blacklisted by the United States of America government, underscoring the effects of politics in geospatial analysis (hyp3-docs.asf.alaska.edu, n.d.; Wegnüller et al., 2016; Werner et al., 2000).

- **ISCE, SNAP, and ASF HyP3 Processing Workflow**

  The processing of Sentinel-1 SLC images employs the stack Sentinel processor within ISCE, StoSAP processor within SNAP, and GAMMA software within ASF HyP3, allowing for the simultaneous processing of single and multi-reference interferogram stacks as outlined by the steps below.

  1. Applying Orbits: Refined precise orbit state vectors from the ESA GNSS Hub, available 20 days post-acquisition, are used for accurate coregistration and interferogram generation.
  2. TOPS Coregistration with NESD: This involves resampling the secondary acquisition into the reference acquisition grid (frame) using Network Enhanced Spectral Diversity (NESD) and geometric coregistration, employing orbit vectors and a DEM.
  3. Interferogram Formation: Complex interferograms are computed by removing the flat-earth phase using precise orbits or a polynomial estimation.
  4. Topographic Phase Removal: A DEM is used to estimate and remove topographic phase contributions, interpolated using Delaunay Triangulation.
  5. Multilooking: This reduces speckle noise via space-domain averaging or frequency-domain techniques, improving interpretability.
  6. Goldstein Phase Filtering: This step minimises residues in phase unwrapping using a nonlinear adaptive filtering algorithm.
  7. Phase Unwrapping: SNAPHU resolves ambiguities in interferometric phase measurements, obtaining spatially continuous phase values.
  8. Exporting: The coregistered stack and interferograms are converted to a time series software-compatible format.

  For further detailed information on ISCE, SNAP, and ASF HyP3 processing workflows, the reader is referred to Fattahi et al. (2017), Rosen et al. (2012), Foumelis et al. (2018), Wegnüller et al. (2016), and Werner et al. (2000).

Following InSAR pre-processing, the exported compatible coregistered stack and interferograms are then ingested into an InSAR time series processing tool in Section 2.2.3, where time series processing is conducted as discussed below.

2.2.3 InSAR Time series Analysis Processing Tools

Several InSAR time series analysis processing software options are available, with commercial offerings such as GAMMA and SARscape. On the other hand, open-source alternatives include Generic InSAR Analysis Toolbox (GIAnT), Miami INsar Time series software in Python

(MintPy), Stanford Method for Persistent Scatterers (StaMPS), SqueeSAR, Python tool for InSAR Rate and Time series Estimation (PyRate), Multidimensional Analysis and Processing of SAR (MAAP), as well as SARProz.

Based on the dissertation's goals emphasising promotion of open-source software, availability of computational resources, and a review of widely used processors, MintPy, and StaMPS were selected for discussion below.

- **Miami InSAR Time series software in Python**

  Miami InSAR Time series software in Python is a powerful open-source tool designed for InSAR time series data processing and analysis, leveraging Python's flexibility and accessibility. Its core strength lies in identifying distributed scatterers, leveraging spatial and temporal coherence to derive highly accurate ground displacement measurements. Moreover, its compatibility with various InSAR data pre-processing tools like GMTSAR, SNAP, ISCE, HyP3, and ARIA enhances its versatility and interoperability. Furthermore, MintPy features a user-friendly command-line interface and interactive plotting, ensuring accessibility even for researchers with limited programming experience. However, mastering MintPy's full potential may require overcoming a learning curve, particularly for newcomers to InSAR time series analysis. Additionally, processing large datasets can strain computational resources, and reliance on a command-line interface might present challenges for users accustomed to graphical interfaces (Ajayi et al., 2023; Yunjun et al., 2019).

  MintPy SBAS-InSAR key processing workflow steps includes:

  1. Loading Stack: The unwrapped interferograms stack is loaded into the system.
  2. Network Modification: Incoherent interferograms are filtered out based on spatial and temporal coherence.
  3. Reference Point Selection: A stable reference point with high spatial and temporal coherence is chosen.
  4. Unwrapping Error Correction: Techniques such as "Bridging of Reliable Region" and "Interferogram Triplets Phase Closure" are employed to correct unwrapping errors.
  5. Network Inversion: Raw phase time series are processed, minimising errors via L2 norm least squares.
  6. Tropospheric Delay Correction: Atmospheric phase screens corrections are made using Global Atmospheric Models (GAM) or a DEM.
  7. Phase Deramping: Linear and quadratic phase ramps are estimated and removed.

8. Topographic Residual Correction: Residual topographic phase from DEM errors is evaluated and corrected.
9. Noise Evaluation: The residual phase is assessed to determine noise levels in the time series.
10. Re-execution and Reference Selection: Noisy SAR acquisitions are excluded, and the optimal reference SAR acquisition date is selected.
11. Average Velocity Estimation: The average velocity is computed from the time series displacements.

- **Stanford Method for Persistent Scatterers**

The Stanford Method for Persistent Scatterers is a pivotal open-source software package in InSAR time series analysis, crucial for detecting and monitoring ground deformations using satellite SAR data from sources like Sentinel-1, ALOS PALSAR, and TerraSAR-X. Its core strength lies in identifying persistent scatterers, leveraging temporal coherence to derive highly accurate ground displacement measurements (Hakim et al., 2023b; Kirui et al., 2022). This capability makes StaMPS invaluable for geohazard assessment, urban planning, and environmental monitoring. However, StaMPS imposes significant computational demands, particularly for large datasets, potentially limiting its use on personal computers. Additionally, its performance in monitoring rapid ground deformation events may be constrained due to its reliance on persistent scatterers, and its dependency on MATLAB, a paid software, could limit accessibility due to financial constraints (Foumelis et al., 2018; Hooper et al., 2012; Osmanoğlu et al., 2016).

StaMPS PS-InSAR and SBAS-InSAR key processing workflow includes:

1. Loading Stack: The wrapped interferogram stack is loaded into StaMPS or MATLAB.
2. Phase Noise Estimation: Phase noise is iteratively estimated for each pixel candidate.
3. PS Pixel Selection: PS candidates are initially selected based on noise characteristics.
4. Non-PS Pixel Weeding: Non-PS candidate pixels are removed based on amplitude dispersion.
5. Wrapped Phase Correction: The wrapped phase is corrected for “Spatially Uncorrelated Look Angle” (SULA) errors.
6. Phase Unwrapping: SNAPHU three-dimensional unwrapping resolves phase ambiguities for PS candidates.

7. SCLA Error Estimation: Errors from DEM inaccuracies, and master atmosphere and orbit errors are estimated and corrected.
8. Atmospheric Filtering: The atmospheric phase screens are filtered using the TRAIN toolbox or Global Atmospheric Models.

For further detailed information on MintPy and StaMPS processing workflows, the reader is referred to Yunjun et al. (2019), Foumelis et al. (2018), Hooper et al. (2012), and Osmanoğlu et al. (2016).

Following InSAR time series analysis processing, the displacements and velocity measurements are exported for post-processing analysis. This includes generating ground subsidence susceptibility maps, modelling other geohazards such as floods, generating Digital Elevation Models, and most importantly, validating the obtained results using ground-based measurements such as those from precise levelling or GNSS, as discussed in Section 2.2.4 below.

2.2.4 Validation of InSAR using GNSS

Validation analysis is crucial in InSAR time series analysis as it assesses the reliability and accuracy of derived measurements. Uncertainties can arise from atmospheric artifacts, noise, decorrelation, and unwrapping errors, which highlight the importance of validation. To validate InSAR results, the East-North-Up (ENU) geographic velocities obtained from continuous GNSS measurements are converted into LOS velocities. These LOS velocities are then compared with the InSAR-derived velocities to ensure their accuracy and reliability (S. Wu et al., 2023).

The conversion of cGNSS measurements to LOS velocities is essential because InSAR detects deformation along the LOS direction. This direction is determined by two angles: the heading angle (β), also known as the look angle, and the local incident angle (θ), which define the LOS orientation concerning the ground surface's normal line as illustrated in Figure 2.7. The conversion formula for ENU to LOS displacements is given by equation (2.6) below.

$$D_{LOS} = -D_E \sin\theta \cos\beta + D_N \sin\theta \sin\beta + D_U \cos\theta \quad (2.6)$$

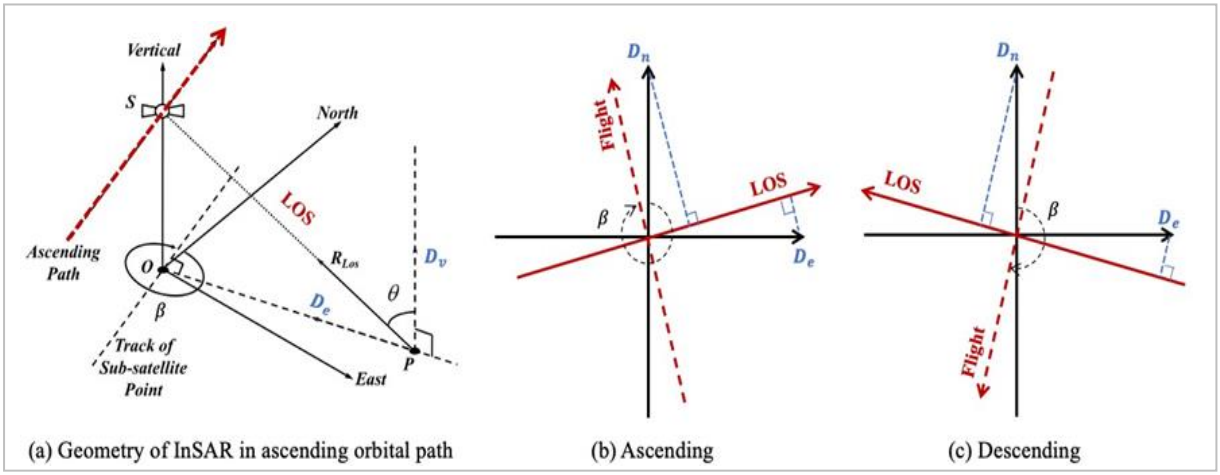


Figure 2.7: Spaceborne SAR observational geometry (Source: S. Wu et al., 2023).

Subsequently, the LOS velocity is estimated using the converted LOS displacements. By comparing these estimated cGNSS LOS velocities with InSAR-derived velocities, the absolute discrepancy in ground deformation can be computed as in equation (2.7) below:

$$V_{error} = V_{GNSS} - V_{InSAR} \tag{2.7}$$

Additionally, the relative discrepancy can be computed by considering the baseline differences between InSAR and cGNSS velocities as shown in equation (2.8) below:

$$\Delta V_{error} = (V_m - V_n)_{cGNSS} - (V_m - V_n)_{InSAR} \tag{2.8}$$

where m and n are subscripts representing the two points that form the baseline. These validation techniques enable the assessment of InSAR velocities accuracy using a single SAR acquisition orbit direction. Alternatively, another approach involves converting InSAR LOS velocities back to ENU velocities to distinguish East-West, North-South, and Up-Down components of ground deformation. This alternative approach can be found in Appendix L.

Having established the techniques for InSAR and InSAR time series analysis processing, discussed the uncertainties present in the measurements, and outlined the validation of InSAR results, it is important to recognise the capabilities and limitations of each technique. Understanding these aspects ensures a comprehensive evaluation of the InSAR results.

2.2.5 Advantages, Limitations and Challenges of InSAR

The capabilities provided by InSAR and InSAR time series analysis are unparalleled, offering detailed spatial and temporal data with exceptional precision. Equally significant are the limitations and challenges associated with these techniques and technology. Therefore, this section aims to present a comprehensive insight into InSAR's capabilities and limitations, which must be considered when using this technique for a wide range of scientific and practical applications.

- **Advantages**

  InSAR offers several advantages for ground deformation monitoring. Firstly, it provides wide-area coverage, allowing for the assessment of ground deformation over large regions. Additionally, InSAR can detect both vertical and horizontal (LOS) displacements, making it suitable for various ground deformation phenomena. It also offers high measurement accuracy, often in the millimetre to centimetre range, which is critical for capturing subtle ground movements. Moreover, InSAR provides high-resolution images, enabling the detection of small-scale features; and multiple InSAR acquisitions over time enable monitoring of dynamic processes like ground subsidence. Additionally, InSAR is unaffected by weather conditions and can operate day and night, ensuring continuous monitoring capabilities.

- **Limitations and Challenges**

  While InSAR is a powerful technique for ground deformation monitoring, it does have limitations and challenges. One significant limitation is the inability to directly measure the absolute vertical and horizontal displacement. Moreover, InSAR provides relative measurements, meaning it can detect changes in ground position but requires ground-based reference points or other techniques to establish the absolute ground deformation.

  Another challenge is the sensitivity of InSAR to factors such as vegetation cover, atmospheric conditions, and topography. Temporal decorrelation of SAR signals due to vegetation growth, land cover changes, or surface roughness variations can hinder the accurate detection of small deformation signals, resulting in data gaps or errors in displacement measurements. Also, Atmospheric disturbances, such as changes in moisture content, pressure and temperature, can also introduce noise into the interferograms. Furthermore, complex topography or steep slopes can cause phase unwrapping errors, leading to inaccuracies in the derived ground subsidence values. Thus, developing advanced noise filtering techniques, and robust atmospheric correction techniques is crucial to mitigate these effects and improve the reliability of InSAR time series analysis.

  Additionally, the large amount of data generated by InSAR satellites poses computational challenges for processing and analysis. Dealing with this volume of data requires advanced computing infrastructure and algorithms that can efficiently handle the processing and interpretation, this can be a barrier for researchers and organisations with limited resources.

  Lastly, integrating and analysing multi-temporal InSAR datasets with other geospatial data sources, such as cGNSS measurements or environmental information, poses another challenge. Combining different data types and sources requires effective data

fusion techniques and the development of comprehensive models to extract meaningful information and insights.

Finally, having discussed the fundamentals of InSAR and InSAR time series analysis, this section delved into several techniques and tools used, along with their inherent limitations, challenges, and advantages. The discussion emphasised the importance of understanding and addressing uncertainties in the measurements and the critical role of validation in ensuring accurate and reliable results. By recognising the strengths and weaknesses of each technique, a comprehensive evaluation of InSAR data and its applications in ground deformation studies was achieved.

Subsequently, we now transition to a new aspect of its application: ground subsidence susceptibility mapping. This next section provides an overview of how InSAR time series analysis data can be used to map areas prone to ground subsidence, an essential step for risk assessment and mitigation in various geological and rural-urban environments. By understanding the methodologies and criteria for ground subsidence susceptibility mapping, we can further appreciate the practical applications of InSAR time series analysis in real-world scenarios.

2.3 Ground Subsidence Susceptibility Mapping Overview

Susceptibility mapping is a crucial tool for assessing the likelihood of ground subsidence in a particular area. It involves identifying and mapping areas prone to ground subsidence based on several topographic, land use and land cover, hydro-geological, and geotechnical factors. By mapping the susceptibility of an area to ground subsidence, authorities and researchers can effectively prioritise mitigation measures and land use planning, aiming to reduce the potential consequences of ground subsidence occurrences.

The integration of InSAR time series analysis data with other geospatial and geotechnical information is fundamental to accurately identifying and assessing areas at risk of ground subsidence. InSAR time series analysis provides precise measurements of ground deformation over time, which, when combined with geological, topo-hydrological, and land use data, enhances the reliability of susceptibility maps. This integrated approach allows for a more comprehensive understanding of ground subsidence phenomena and supports the development of effective mitigation and management strategies.

2.3.1 Susceptibility Mapping Techniques

Several techniques are employed in susceptibility mapping for ground subsidence, floods, and landslides. These techniques involve the integration of spatial data such as land cover and land use, topographic information, and hydro-geological parameters. Traditional susceptibility mapping methods include the Analytic Hierarchy Process, Fuzzy AHP, Multi-Criteria Decision-Making Analysis, Weight of Evidence, and Fuzzy Logic, including Neuro-Fuzzy

approaches (Ashraf and Cawood, 2015; Choi et al., 2010; Lee et al., 2012; Yalcin et al., 2011). Each of these methods has its advantages and disadvantages as discussed below:

- **Analytic Hierarchy Process**
  The AHP technique provides a structured framework for decision-making and is straightforward to understand and implement. However, it is subjective due to its reliance on expert judgement and may struggle with complex, high-dimensional data.
- **Fuzzy AHP**
  Fuzzy AHP incorporates fuzzy logic, in an attempt to address AHP's subjectivity and better handles uncertainty and imprecision. Nonetheless, it can still be subjective, relies on expert judgement, and becomes computationally intensive with a large number of criteria factors.
- **Multi-Criteria Decision-Making Analysis**
  MCDMA is effective in integrating multiple criteria and providing comprehensive analysis. However, MCDMA can become unwieldy and difficult to calibrate as the number of criteria increases, necessitating significant computational resources and expertise.
- **Weight of Evidence**
  WoE is useful for integrating different types of spatial data and relatively straightforward to implement. However, this method may oversimplify complex relationships and can be biased if not properly calibrated.
- **Fuzzy Logic and Neuro-Fuzzy**
  Fuzzy Logic and Neuro-Fuzzy offers flexibility in handling uncertainties and linguistic terms, and capable of modelling complex systems with imprecise information. However, this approach is challenging to interpret and parameterise, requiring significant computational power and expertise.

Overall, these traditional susceptibility mapping techniques offer structured and comprehensive decision-making frameworks but tend to be subjective, computationally intensive, and difficult to manage with increasing complexity.

In contrast, data-driven approaches such as machine learning algorithms, are commonly used to analyse data and generate susceptibility maps that categorises areas into different levels of susceptibility to ground subsidence (Hakim et al., 2023b; Mohammady et al., 2019; Saha et al., 2024). Machine learning techniques, such as neural networks and ensemble methods, are increasingly adopted due to their ability to handle large, complex datasets and uncover patterns that traditional methods might miss (Hakim et al., 2023b; Sekkeravani et al., 2022). Moreover, statistical models including frequency ratio (FR) analysis, certainty factor analysis are used to

analyse and explain the inducing factors of ground subsidence, complimenting machine learning algorithms.

Furthermore, the integration of InSAR time series analysis with Machine Learning for susceptibility mapping allows for a more accurate and detailed assessment of ground subsidence prone areas. InSAR time series analysis provides spatially extensive precise measurements of ground deformation, which can be used to validate and refine susceptibility maps, leading to improved understanding and prediction of ground subsidence. While Machine learning algorithms enhance this process by effectively processing and analysing the extensive data provided by InSAR time series analysis, resulting in more reliable and precise susceptibility assessments. This integration provide valuable information for decision-makers to prioritise areas for monitoring, preventive measures, and land-use planning.

To ensure the effectiveness of machine learning algorithms in this integration, it is essential to undertake two crucial preliminary processes: factor prioritisation and multicollinearity analysis. These steps help in selecting the most significant factors influencing ground subsidence and in mitigating potential issues of redundancy among variables, thereby enhancing the accuracy and reliability of the ground subsidence susceptibility mapping.

2.3.2 Factor Prioritisation

Factor prioritisation using Information Gain (IG) analysis aims to identify the most effective factors for modelling output features, such as ground subsidence velocities or susceptibility maps. Information Gain analysis is a method used in feature selection to assess the relevance of each predictor variable in relation to the target variable. It measures the reduction in entropy or uncertainty about the target variable achieved by splitting the data on a particular predictor variable. Higher information gain values suggest greater predictive power of the variable (Hakim et al., 2023b; Mohammady et al., 2019).

Once the effective factors have been determined through factor prioritisation, the next step is to perform multicollinearity analysis.

2.3.3 Multicollinearity Analysis

The multicollinearity analysis using Variance Inflation Factor (VIF), Tolerance, and Pearson correlation aims to identify and remove redundant factors that may be highly correlated with one another (inter-dependent). Factors exhibiting high multicollinearity may introduce redundancy in the model, leading to inflated coefficients and inaccurate predictions. Therefore, by identifying and eliminating these redundant factors, the analysis ensures that the model remains robust and that the predictive power of each factor is not diminished by multicollinearity (Hakim et al., 2023b; Sekkeravani et al., 2022).

- **Pearson Correlation Analysis**

  The Pearson correlation metric is used to measure the strength and direction of the linear relationship between two continuous variables (Li et al., 2024). It assesses the degree of association between variables, aiding in understanding how changes in one variable are associated with changes in another. The Pearson correlation coefficient (r) is calculated as the covariance of the two variables divided by the product of their standard deviations (Jing et al., 2024). For variables x and y, it is given by equation (2.9) below:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \ \sum(y_i - \bar{y})^2}} \tag{2.9}$$

  Where $\bar{x}$ and $\bar{y}$ are the means of variables x and y, respectively. A correlation coefficient close to +1 indicates a strong positive linear relationship, while a value close to -1 indicates a strong negative linear relationship. And a coefficient near 0 suggests no linear relationship between the variables.

- **VIF and Tolerance Analysis**

  Variance Inflation Factor and Tolerance are statistical measures used to assess multicollinearity among predictor variables in regression analysis. Multicollinearity occurs when independent variables in a regression model are highly correlated with each other, potentially leading to unreliable coefficient estimates (Sekkeravani et al., 2022).

  VIF measures how much the variance of an estimated regression coefficient is increased due to multicollinearity. It is calculated for each predictor variable x_i as shown in equation (2.10) below:

$$VIF(x_i) = \frac{1}{1 - r_i^2} \tag{2.10}$$

  Where r_i^2 is the coefficient of determination from regressing r on all other predictor variables. Tolerance, on the other hand, quantifies the proportion of variance in a predictor variable that is not explained by the other predictor variables. It is calculated as in equation (2.11) below:

$$Tolerance(x_i) = 1 - r_i^2 \tag{2.11}$$

  Low tolerance values (below 0.1) indicate high multicollinearity, whereas high VIF values (typically above 5) suggest problematic multicollinearity.

After confirming the effectiveness and independence of the selected factors through factor prioritisation and multicollinearity analysis, a solid foundation is established for applying machine learning techniques in ground subsidence susceptibility mapping.

2.3.4 Machine Learning for Susceptibility Mapping

The incorporation of machine learning algorithms has emerged as a viable way to extract and predict useful insights from complex data in the field of ground subsidence monitoring and susceptibility mapping. Machine learning is a subset of artificial intelligence (AI) that focuses on developing algorithms and statistical models, enabling computers to perform tasks without explicit instructions (Ayati et al., 2022). By learning from data and making predictions or decisions based on this information, ML has revolutionised various fields, including geospatial analysis. In geospatial problems, ML is applied to enhance the understanding and management of complex spatial phenomena.

For instance, in ground subsidence monitoring, ML algorithms analyse InSAR time series analysis data to detect and predict ground deformation. This provides critical insights for infrastructure management and natural hazard assessment. Similarly, in susceptibility mapping, ML techniques process diverse geospatial datasets to identify areas prone to natural hazards like ground subsidence, landslides, floods, or earthquakes, aiding in risk assessment and mitigation strategies.

These applications underscore the transformative potential of ML in addressing pressing environmental and societal challenges through advanced geospatial analysis (Lary et al., 2016; Tien Bui et al., 2016). Thus, this section delves into four prominent machine learning algorithms and their applications in ground subsidence monitoring and susceptibility mapping, and similar geospatial contexts.

- **Convolutional Neural Networks**
  Convolutional Neural Networks are a type of deep learning model particularly effective for processing grid-like data. CNNs use convolutional layers that apply filters to input data to detect patterns, which makes them highly suitable for image recognition tasks (LeCun et al., 2015). The architecture of CNNs typically includes convolutional layers, pooling layers, and fully connected layers, enabling them to learn hierarchical feature representations.

  Key aspects of CNNs include:

  - **Convolutional Layers:** These layers apply convolution operations to the input data, detecting local patterns. The operation is defined as shown in equation (2.12) below:

$$(f * g)(t) = \sum_{a=-\infty}^{\infty} f(a) \cdot g(t-a) \tag{2.12}$$

  - **Pooling Layers:** These layers reduce the dimensionality of the data, preserving important features while minimising computational complexity. The max pooling operation, for example, can be expressed as shown in equation (2.13) below:

$$y_{i,j} = max\left(x_{i\times m+k, j\times n+l}\right) \quad (2.13)$$

where m and n are the pooling dimensions.

- **Fully Connected Layers:** These layers integrate features extracted by convolutional and pooling layers for final classification, defined as shown in equation (2.14) below:

$$y = W \cdot x + b \quad (2.14)$$

where W is the weight matrix, x is the input, and b is the bias.

In geospatial analysis, CNNs have been widely applied to tasks such as land cover classification, object detection, and change detection (Liu et al., 2019; Memon et al., 2021; Zhiqiang and Jun, 2017). Their application to InSAR time series analysis data for ground subsidence monitoring has shown promising results, allowing for the detection and mapping of ground subsidence patterns with high accuracy (Hakim et al., 2023b). Additionally, CNNs have been used for susceptibility mapping in geohazards, enhancing the ability to predict areas prone to ground subsidence (Bagheri et al., 2024; Hakim et al., 2023b; Z. Wu et al., 2023).

- **eXtreme Gradient Boosting**

Extreme Gradient Boosting is an efficient and scalable implementation of gradient boosting algorithms (Yang et al., 2022). It enhances the performance of gradient boosting by incorporating system optimization and algorithmic enhancements, including tree pruning, regularisation, and handling of missing values (Chen and Guestrin, 2016). XGBoost ability to handle large-scale data and prevent overfitting makes it a powerful tool for predictive modelling.

Key aspects of XGBoost include:

- **Boosting:** Sequentially builds models, each correcting errors of the previous ones. Here, the objective function is defined as shown in equation (2.15) below:

$$\mathcal{L}(\theta) = \sum_i l\left(y_i, \hat{y_i}^{(t)}\right) + \sum_k \Omega(f_k)] \quad (2.15)$$

where l is a differentiable convex loss function, and Ω is a regularisation term.

- **Regularisation:** Incorporates L1 and L2 regularisation to prevent overfitting as shown in equation (2.16) below:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \quad (2.16)$$

where T is the number of leaves, and w_j are the leaf weights.

- **Scalability:** Efficiently handles large datasets and is highly parallelisable.

In geospatial analysis, XGBoost has been applied to various tasks such as land cover classification, crop yield prediction, and urban heat island analysis (Bushenkova et al., 2024; Georganos et al., 2018; Ravi and Baranidharan, 2020). Its robustness and high predictive accuracy make it suitable for ground subsidence monitoring and susceptibility mapping, where it can effectively model complex relationships in InSAR time series analysis data (Achour et al., 2021; Costache et al., 2022; Eghrari et al., 2023; Naghibi et al., 2022; Shi et al., 2020; Toomanian et al., 2022).

- **Light Gradient Boosting Machine**

Light Gradient Boosting Machine is a gradient boosting framework developed by Microsoft that focuses on efficiency and scalability. It uses a histogram-based algorithm and leaf-wise tree growth strategy to reduce memory usage and increase training speed (Ke et al., 2017). LightGBM is well-suited for large datasets and high-dimensional data, offering excellent accuracy and performance.

Key aspects of LightGBM include:

- **Histogram-based Learning:** Buckets continuous feature values into discrete bins, reducing computational cost. Here, the binning process can be described as shown in equation (2.17) below:

$$\mathrm{Bin}(x) = \left\lfloor \frac{x - \min(x)}{\Delta} \right\rfloor \tag{2.17}$$

  where Δ is the bin width.

- **Leaf-wise Tree Growth:** Grows tree's leaf-wise rather than level-wise, leading to deeper trees with better accuracy.

- **Parallel and GPU Learning:** Supports parallel and GPU-based training for enhanced performance.

Applications of LightGBM in geospatial analysis include hyperspectral image classification, and flood prediction (Xu et al., 2023; Yoon et al., 2023). Its efficient handling of large datasets makes it ideal for ground subsidence monitoring and susceptibility mapping, enabling the processing of extensive InSAR time series analysis data with high accuracy (Chai et al., 2024; Lee et al., 2023).

- **Random Forest**

Random Forest is a powerful and widely used supervised learning algorithm that excels in both classification and regression tasks. It works by constructing multiple decision trees during training and outputting the mode of the classes (classification) or mean prediction (regression) of the individual trees (Le et al., 2024). Introduced by Breiman

in 2001, the algorithm builds each tree using a different bootstrap sample of the data and selects splits from a random subset of the features, which helps in reducing overfitting and improving generalisation (Breiman, 2001).

Key aspects of Random Forest includes:

- **Bagging:** Random subsets of data and features are used to build each tree, promoting diversity. The prediction of a Random Forest equation (2.18) below:

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} h_i(x) \tag{2.18}$$

  where h_i is the prediction of the $i - th$ tree.

- **Robustness:** Aggregating multiple trees reduces variance and improves generalisability.

- **Feature Importance:** Provides insights into the significance of each feature in the prediction process.

Random Forest has been proven to be highly effective in geospatial applications, where it has been used for tasks such as land cover classification, object detection, and change detection (Ai et al., 2024; Avci et al., 2023; Gislason et al., 2006). Its ability to handle large datasets and high-dimensional spaces makes it particularly suitable for InSAR time series analysis data. Applications in ground subsidence monitoring include mapping ground subsidence areas and predicting ground subsidence based on ground subsidence related data, leveraging its robustness and high accuracy (Achour et al., 2021; Eghrari et al., 2023; Hosseinzadeh et al., 2024; Mohammady et al., 2019; Qiao et al., 2024; Zhao et al., 2024).

Having discussed the foundational principles of machine learning algorithms and their applications in ground subsidence monitoring and susceptibility mapping, the next crucial step is to optimise their performance. Effective application of these algorithms often requires fine-tuning their hyperparameters, which significantly impact the model's accuracy and generalisation capability. One widely used method for achieving this is Hyperparameter Optimization with GridSearchCV.

- **Hyperparameter Optimization with GridSearchCV**

  GridSearchCV is a method for hyperparameter tuning that exhaustively searches through a predefined set of hyperparameters to find the optimal combination for model performance. For each algorithm, hyperparameters such as learning rate, number of estimators, epochs, batch size, filters, and tree depth are systematically varied and evaluated using cross-validation. Subsequently, to assess the effectiveness of these

hyperparameters and the overall ML model performance, appropriate model validation metrics must be employed (Petro and Palvo, 2019).

- **Model Validation Metrics**

  Model validation is crucial to ensure the reliability and accuracy of predictive models, providing confidence in their ability to generalise to new data and make informed decisions. To evaluate the performance of the regression models for ground subsidence susceptibility mapping, the following metrics are employed:

  - **Root Mean Squared Error (RMSE):** Measures the average magnitude of the error between predicted ($\hat{y}_i$) and actual values (y_i) of ground subsidence as shown in equation (2.19) below.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.19}$$

  - **Mean Absolute Error (MAE):** Measures the average magnitude of the errors in a set of ground subsidence predictions, without considering their direction as shown in equation (2.20) below.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2.20}$$

  - **R^2 score (coefficient of determination):** Measures the proportion of the variance in the dependent variable that is predictable from the independent variables as shown in equation (2.21) below.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \hat{y})^2} \tag{2.21}$$

    where $\hat{y}$ is the mean of the observed ground subsidence data.

  After regression, the predicted values are categorised into classes using thresholds optimised for the specific problem domain. To evaluate the performance of the classification models on the reclassified data, the following metrics are employed:

  - **Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC):** ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings, while AUC quantifies the overall ability of the model to distinguish between classes as shown in equation (2.22) below.

$$\text{TPR} = \frac{TP}{TP + FN} \ , \ \text{FPR} = \frac{FP}{FP + TN} \tag{2.22}$$

Where TP, TN, FP, and FN are true positive, true negative, false positive, and false negative rate respectively (Ayati et al., 2022).

Subsequently, to gain comprehensive insights into the predictions made by our validated hyper-parameterised model and to evaluate the relative importance of each feature in these predictions, we adopted the use of SHapley Additive exPlanations.

- **Model Explainability with SHAP Values**

  Shapley Additive exPlanations (SHAP) is a powerful method for attributing a model's prediction to individual features, providing a clear understanding of how each feature influences the model's decisions. This technique enhances model interpretability, supports informed decision-making in model refinement, and aids in feature selection (Qiao et al., 2024).

  SHAP values make model outputs more comprehensible by associating each data point with a value that reflects the importance of each feature. These values rank features based on their significance.

  For a given prediction, the SHAP value of each feature indicates its impact on the model's decision compared to a baseline value. Positive SHAP values imply that the feature increased the output relative to the baseline (high probability), whereas negative SHAP values suggest a decrease (low probability).

  Key aspects of SHAP values include:

  - **Additivity:** Ensures that the sum of the SHAP values across all features equals the difference between the model's prediction and the average prediction as shown in equation (2.23) below.

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i \tag{2.23}$$

    where ϕ_0 is the base value and ϕ_i are the SHAP values for each feature.

  - **Consistency:** Guarantees that an increase in a feature's contribution does not decrease its corresponding SHAP value.

  - **Global and Local Interpretability:** Provides both an overview of feature importance across the dataset and detailed explanations for individual predictions.

To further understand the factor classes contributing to ground subsidence and their impact across different regions, the integration of statistical analysis techniques is essential.

2.3.5 Ground subsidence explainability with statistical models

Statistical techniques such as Frequency Ratio analysis play a crucial role in complementing machine learning approaches. These techniques provide deeper insights into the relationships between ground subsidence occurrences and influencing factor classes. By enhancing the interpretability of machine learning results, these statistical models enrich the overall understanding of ground subsidence dynamics, as discussed below.

- **Frequency Ratio Analysis**

  Frequency ratio technique assess the spatial frequency of ground subsidence incidents within various classes associated with each ground subsidence related factor. A frequency ratio value exceeding 1 indicates a stronger correlation between a particular class within each factor and the incidents of ground subsidence. Higher frequency ratio values, along with a greater number of classes showing higher FR values, indicate stronger associations between the related factor and the occurrence of ground subsidence (Hakim et al., 2023b; Ranjgar et al., 2021; Sekkeravani et al., 2022). The frequency ratio is computed as shown by equation (2.24) below:

$$FR = \frac{S_i}{\sum_{i=1}^{n} S_i} * \frac{A_{study}}{A_{class}} \tag{2.24}$$

  Where S_i represents the total ground subsidence in each class, A_{study} represents the total area of the study region, A_{class} represents the area of each class, and n represents the total number of classes.

Having explored statistical methods for explaining ground subsidence, and machine learning for susceptibility mapping. It is important to note that the CNN, LightGBM, Random Forest, and XGBoost ML supervised approaches rely on selected features to make accurate predictions. Including, the explainability of ground subsidence inducing factors using SHAP and statistical methods like Frequency Ratio analysis. Therefore, the subsequent section extends this foundation by discussing factors associated with ground subsidence, which are crucial for training ML models in susceptibility mapping.

2.3.6 Ground Subsidence Related Factors

This sub-section explores factors contributing to ground subsidence, crucial for understanding its impact and training machine learning models for accurate ground subsidence susceptibility mapping. Factors considered include elevation, slope, aspect, curvature, topographic wetness index, distance to roads, distance to rivers, Modified Normalised Difference Water Index, New Built-Up Index, Normalised Built-Up Area Index, Normalised Difference Built-up Index, Normalised Difference Moisture Index, Normalised Difference Water Index, Normalised Difference Vegetation Index, and Soil-Adjusted Vegetation Index, as discussed below.

- **Elevation**

  Elevation is crucial in ground subsidence monitoring and susceptibility mapping as it directly influences gravitational potential energy and surface water flow. Typically, higher elevations experience less ground subsidence compared to lower elevations due to decreased pressure from overlying materials. Elevation data helps identify areas prone to ground subsidence, with lower elevations being more susceptible.

- **Slope**

  Slope represents the steepness of terrain and is a key factor in ground subsidence monitoring and susceptibility mapping. Steeper slopes are generally less prone to ground subsidence because gravitational forces are less concentrated, reducing the potential for ground movement. In contrast, gentle slopes may indicate areas of potential ground subsidence. The slope is mathematically expressed as shown in equation (2.25) below:

$$Slope = tan^{-1}\left(\sqrt{\left(\frac{\Delta z}{\Delta x}\right)^2 + \left(\frac{\Delta z}{\Delta y}\right)^2}\right) \tag{2.25}$$

- **Aspect**

  Aspect, the compass direction a slope faces, affects ground subsidence patterns due to varying amounts of solar radiation impacting soil moisture and vegetation growth. Aspect information helps identify areas prone to rapid drying or excessive moisture accumulation, contributing to ground subsidence. The mathematical expression for aspect is as shown in equation (2.26) below:

$$Aspect = tan^{-1}\left(-\frac{\Delta z}{\Delta y} \times \frac{\Delta x}{\Delta z}\right) \tag{2.26}$$

- **Curvature**

  Curvature represents the rate of change in slope along a surface. Positive curvature indicates convex surfaces, which may have tensional forces increasing ground subsidence susceptibility, while negative curvature indicates concave surfaces, potentially stabilising the ground. Curvature analysis identifies areas where ground movement is likely based on surface deformation patterns. Curvature is calculated as shown in equation (2.27) below:

$$Curv = -\frac{\left(\frac{\Delta z}{\Delta x^2} \cdot \frac{\Delta x}{\sqrt{(\Delta x)^2 + (\Delta y)^2}}\right)^2 + \left(\frac{\Delta z}{\Delta y^2} \cdot \frac{\Delta y}{\sqrt{(\Delta x)^2 + (\Delta y)^2}}\right)^2}{1 + \left(\frac{\Delta z}{\sqrt{(\Delta x)^2 + (\Delta y)^2}}\right)^2} \tag{2.27}$$

- **Topographic Wetness Index**

  TWI represents a landscape's capacity to accumulate and retain water, it is calculated based on slope and specific catchment area. Higher TWI values indicate wetter areas more susceptible to ground subsidence due to increased groundwater saturation and decreased soil strength. TWI is expressed as shown in equation (2.28) below:

$$TWI = ln\left(\frac{A}{tan(Slope)}\right) \tag{2.28}$$

  where A is the upslope contributing area.

- **Distance to Roads**

  Distance to roads influence ground subsidence due to construction materials, drainage patterns, and associated human activities. Areas near roads may experience ground subsidence due to compaction, alteration of natural drainage pathways, and changes in groundwater flow. Proximity to roads is considered a factor in increased ground subsidence risk.

- **Distance to Rivers**

  Distance to rivers affect ground subsidence through hydrology, sediment transport, and erosion. Riverbanks are often susceptible to ground subsidence due to these hydraulic processes. Proximity to rivers helps identify areas prone to ground subsidence, with river proximity and characteristics quantified for analysis.

- **Modified Normalised Difference Water Index**

  Modified Normalised Difference Water Index (MNDWI) highlights water bodies and moisture content in vegetation, with positive values indicating water presence. Areas with high positive MNDWI values may have high moisture content, contributing to ground subsidence through soil softening. MNDWI is calculated as shown in equation (2.29) below:

$$MNDWI = \frac{Green - SWIR}{Green + SWIR} \tag{2.29}$$

- **New Built-Up Index**

  New Built-Up Index (NBI) identifies newly built-up areas from satellite imagery. High NBI values indicate recent urban expansion, potentially increasing ground subsidence risk due to changes in land use and drainage patterns. NBI is expressed as shown in equation (2.30) below:

$$NBI = \frac{RED - SWIR2}{NIR} \tag{2.30}$$

- **Normalised Built-Up Area Index**

  Normalised Built-Up Area Index (NBAI) quantifies the extent of built-up areas relative to their surroundings, with high values indicating dense urban areas. Built-up areas may be more susceptible to ground subsidence due to land use changes and increased anthropogenic activities. NBAI is calculated as shown in equation (2.31) below:

$$NBAI = \frac{SWIR2 - SWIR1/GREEN}{SWIR2 + SWIR1/GREEN} \tag{2.31}$$

- **Normalised Difference Built-up Index**

  Normalised Difference Built-up Index (NDBI) highlights built-up areas, with high values indicating urban areas susceptible to ground subsidence due to land use changes, in contrast less or no ground subsidence due to persistent scatterers over buildings. NDBI is expressed as shown in equation (2.32) below:

$$NDBI = \frac{SWIR2 - NIR}{SWIR2 + NIR} \tag{2.32}$$

- **Normalised Difference Moisture Index**

  Normalised Difference Moisture Index (NDMI) assesses vegetation moisture content, with positive values indicating higher moisture levels. Areas with higher NDMI values may have more stable soils, reducing ground subsidence susceptibility. NDMI is calculated as shown in equation (2.33) below:

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR} \tag{2.33}$$

- **Normalised Difference Water Index**

  Normalised Difference Water Index (NDWI) identifies water bodies and vegetation moisture content. High negative values indicate water presence, contributing to ground subsidence through soil softening. NDWI is expressed as shown in equation (2.34) below:

$$NDWI = \frac{Green - NIR}{Green + NIR} \tag{2.34}$$

- **Normalised Difference Vegetation Index**

  Normalised Difference Vegetation Index assesses vegetation health and density, with higher values indicating denser vegetation that stabilises soils and reduces ground subsidence risk. NDVI is calculated as shown in equation (2.35) below:

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{2.35}$$

- **Soil Adjusted Vegetation Index**

  Soil Adjusted Vegetation Index (SAVI) minimises soil background influence on vegetation indices, particularly in bright soil areas. SAVI values range from -1 to 1, with higher values indicating denser vegetation. SAVI is calculated as shown in equation (2.36) below:

$$SAVI = \left(\frac{NIR - Red}{NIR + Red + L} \right) \times (1 + L) \tag{2.36}$$

  where L is a constant parameter ranging from 0 to 1 (i.e., L=0.5), adjusting for soil background effects.

Finally, it is worth noting that the majority of these factors, including elevation, slope, aspect, curvature, and TWI, are derived from Digital Elevation Models. Therefore, the subsequent section, explores the critical aspects of DEMs, their associated errors, and the research gap concerning how these errors propagate into susceptibility mapping. Understanding these intricacies is essential for refining the accuracy and reliability of predictive models used to assess ground subsidence and susceptibility mapping.

2.3.7 Digital Elevation Models for Susceptibility Mapping

Digital Elevation Models plays a pivotal role in InSAR and Machine Learning computation for ground subsidence monitoring and susceptibility mapping. A DEM is a three-dimensional representation of Earth's surface topography, essential for correcting and interpreting InSAR data (Ferretti, 2014; Maune, 2007). DEMs are generated using methods such as photogrammetry, LiDAR, ground surveys, and radar altimetry (Gómez et al., 2012; Guth et al., 2021). Their integration in InSAR time series analysis and machine learning processing ensures high-quality ground deformation measurements and supports effective risk assessment and mitigation strategies.

In InSAR and machine learning processing, DEMs are used to derive topo-hydrological factors related to ground subsidence and to remove topographic phase components from InSAR data, isolating surface deformation signals. Accurate DEMs enable precise measurement of ground deformation, enhancing the reliability of ground subsidence monitoring. Moreover, in ground subsidence susceptibility mapping, DEMs help identify and analyse potential hazard areas by providing critical ground subsidence related data (Gharechaee et al., 2023; Mohammady et al., 2019).

However, DEMs have inherent vertical errors, and the propagation of these errors and their impact on ground subsidence modelling, particularly when using machine learning techniques, remains unclear. Addressing this gap is critical to improving the accuracy and reliability of ground subsidence models. Therefore, below is a brief review of the DEMs adopted in this research to address this research gap.

- **Advanced Land Observing Satellite World 3D**

  The Advanced Land Observing Satellite World 3D (AW3D), 30 metres spatial resolution, is a high-resolution global DEM developed from stereo data collected by the "Japan Aerospace Exploration Agency" (JAXA). AW3D30's creation involved the use of 3 million optical images captured by the PRISM panchromatic stereo mapping sensor on "Advanced Land Observing Satellite" (ALOS) satellite between 2006 to 2011 to generate precise elevation information (Li et al., 2021). AW3D30 geographic coordinate system is referenced to EGM96 vertical datum and WGS84 horizontal datum; and it is reported to have an absolute vertical error of no more than 5 metres (Tadono et al., 2014; Takaku et al., 2016, 2014).

  The AW3D30 DEM has been used for both ground subsidence monitoring and susceptibility mapping in studies conducted by Orhan et al. (2020), H. Sun et al. (2023), and S. Wang et al. (2024).

- **Advanced Spaceborne Thermal Emission and Reflection Radiometer**

  The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global DEM is another significant dataset in the realm of DEM. This product is derived from 1.7 million stereo images collected by the ASTER sensor aboard NASA's Terra satellite between the years 2000 and 2013. It provides a global elevation dataset from 83° North to 83° South at 30 metres resolution, offering detailed topographic information. ASTER's geographic coordinate system is referenced to EGM96 vertical datum and WGS84 horizontal datum; and it is reported to have a vertical error of no more than 20 metres at a 95% confidence level (Reuter et al., 2009).

  The ASTER DEM has been applied for ground subsidence monitoring and susceptibility mapping in several studies. Specifically, Ranjgar et al. (2021) used the ASTER DEM for ground subsidence monitoring, while both Naghibi et al. (2022) and Ranjgar et al. (2021) employed it for ground subsidence susceptibility mapping.

- **Shuttle Radar Topography Mission Global Level 1**

  The Shuttle Radar Topography Mission Global Level 1 (SRTMGL1) DEM is a collaborative effort between NASA, the "German and Italian Space Agencies" (ASI), and the "National Geospatial-Intelligence Agency" (NGA). It was created using C-band and X-band radar imaging technology during an 11-day mission in February 2000. Providing 1 arc-second resolution (30 m spatial resolution), the SRTMGL1 dataset delivers highly precise elevation information on a near-global scale from 60° N to 56° S (about 80% of the world topography), referenced to Mean Sea Level (MSL) based on the EGM96 vertical datum and WGS84 for the horizontal datum. It is reported to have a vertical error of no more than 16 metres at a 90% confidence level (Yang et al., 2011).

The SRTMGL1 DEM has been applied for ground subsidence monitoring and susceptibility mapping in several studies. Specifically, Bagheri et al. (2024), Long Chai et al. (2024), Costache et al. (2022), Hakim et al. (2023b), Naghibi et al. (2022), and Z. Wu et al. (2023) used the SRTMGL1 DEM for ground subsidence monitoring. Additionally, Long Chai et al. (2024), Costache et al. (2022), Hakim et al. (2023b), Qiao et al. (2024), Z. Wu et al. (2023), and Yang et al. (2022) used it for ground subsidence susceptibility mapping.

- **Copernicus DEM**

  The Copernicus DEM, 30 metres resolution (GLO-30 or COP30), is derived from the esteemed WorldDEM dataset, which originates from SAR data acquired by the TanDEM-X satellite mission. TanDEM-X mission is a joint venture between the German Aerospace Center (DLR) and Airbus Defence and Space which conducted its data acquisition operations between December 2010 and January 2015, to generate the WorldDEM. The WorldDEM dataset, serves as the foundation for Copernicus GLO-30 DEM, offering global coverage with a spatial resolution of 30 metres. COP30 geographic coordinate system is referenced to EGM2008 vertical datum and WGS84 horizontal datum; and it is reported to have the best absolute vertical accuracy of no more than 10 metres (Airbus, 2020).

  The COP30 DEM has been applied for ground subsidence monitoring, as demonstrated by Hakim et al. (2023a). However, there have been no or very few studies on susceptibility mapping using the COP30 DEM.

Moreover, in South Africa, a study conducted by Malindi and Odera (2022) evaluated the vertical accuracy of three prominent DEMs: AW3D30, SRTMGL1, and ASTER. The findings revealed vertical accuracies of 5.09 metres, 7.03 metres, and 9.20 metres, respectively. These results highlight the variability in accuracy among different DEMs and underscore the need for careful selection based on the specific requirements of a given application.

Further validation was performed in Cape Town, South Africa, by Okolie et al. (2024). This study assessed the accuracy of COP30, AW3D30, and ASTER, reporting vertical accuracies of 2.34 metres, 3.72 metres, and 7.18 metres, respectively. The validation analysis of these DEMs reinforces the notion that each model offers distinct levels of reliability and accuracy.

These studies collectively emphasise the importance of DEM validation using ground levelling points. Accurate ground-truth data is essential for assessing and improving DEM accuracy, ensuring that these models can be confidently used in geospatial analysis and applications. The variability in accuracy across different DEMs also suggests the need for tailored approaches to DEM selection and validation, considering the specific geographic and research requirements.

- **Ground Levelling Elevation**

  The South African geodetic reference network comprises approximately 30,000 ground levelling (trigonometric) beacons referenced to Hartebeesthoek94 (WGS84) horizontal datum, and the Land Levelling Datum (LLD) vertical datum. The Hartebeesthoek94 initial point is defined by Hartebeesthoek Radio Astronomy (HRAO) telescope, and LLD initial point is defined by 'BM1' benchmark in Cape Town harbour.

  These beacons provide accurate terrain representation but are sparsely distributed, which presents challenges for continuous topographic modelling. Nonetheless, they play a critical role in DEM adjustment and fusion techniques and, more importantly, serve as pivotal reference points for validating DEMs.

- **DEM Validation**

  The precision of the DEM plays a crucial role in ground subsidence modelling and analysis. Any systematic errors, blunders, and random errors (inaccuracies) within the DEM will proportionally affect the computation and will become evident in the subsequent analysis. These inaccuracies stem from the techniques used during data collection and the generation process of the DEM model. Hence, it is vital to evaluate the accuracy of DEM's and use the most precise one for geospatial analysis.

  In order to evaluate the accuracy of the DEM datasets, it is necessary to identify and evaluate the height differences associated with the DEM. This involves computing the height differences (H_{error}) between the ground levelling height (H_{TRIG}) and the DEM implied height (H_{DEM}) at the corresponding horizontal position (latitude and longitude) as shown by equation (2.37) below.

$$H_{error} = H_{TRIG} - H_{DEM} \tag{2.37}$$

  Once the height differences are obtained, statistical analysis can be conducted using metrics such as mean, range, root mean square error, and standard deviation to determine the reliability and accuracy of each DEM. Moreover, the height differences can be correlated with other product errors, such as velocity errors, to understand the propagation of DEM vertical errors into geospatial modelled products, such as ground subsidence rates and susceptibility maps.

Overall, DEMs provide essential elevation data crucial for ground subsidence monitoring and susceptibility mapping. However, results from DEM validations indicate that these models have inherent vertical errors, underscoring the importance of validating DEMs to select the best-performing models. Additionally, understanding how DEM vertical errors propagate into ground subsidence modelling and susceptibility mapping is essential. Despite their limitations, DEMs offer unique advantages in terms of resolution, coverage, and accuracy, significantly

contributing to diverse scientific investigations and geographical analyses across different regions.

2.3.8 Challenges, Potential Advancements and Future Directions

There are several challenges associated with ground subsidence susceptibility mapping, one of the key challenges is the availability and quality of input data. Obtaining high-resolution and up-to-date data, including terrain characteristics, geological information, and historical hazard records, is crucial for accurate ground subsidence susceptibility mapping. Limited data availability or poor data quality can lead to uncertainties and inaccuracies in the ground subsidence susceptibility mapping results. Another challenge is the complexity and variability of natural hazard processes. Different factors, such as geology, topography, rainfall patterns, and human activities, can influence the occurrence and magnitude of hazards. Capturing this complexity and accounting for the spatial and temporal variations in hazard processes require advanced modelling techniques and robust statistical methods.

Additionally, the interpretation and integration of multiple data sources and models can be challenging. Ground subsidence susceptibility mapping often involves the combination of various data layers, such as environmental data, remote sensing data, terrain attributes, and statistical models. Developing effective methodologies for data integration and model fusion is essential to generate reliable ground subsidence susceptibility maps. Furthermore, communicating the results of ground subsidence susceptibility mapping to stakeholders and decision-makers poses another challenge. Translating complex scientific information into actionable insights and ensuring effective communication to policymakers and communities is crucial for implementing appropriate risk reduction strategies and land-use planning.

Despite the challenges, there are promising advancements and future directions in both InSAR time series analysis and ground subsidence susceptibility mapping. In the case of InSAR, advancements in computing power and cloud-based processing platforms such as “OpenScience Lab” and the High-Performance Clusters (HPC) have the potential to address the computational challenges associated with large datasets. Machine learning, particularly ensemble and deep learning techniques can also contribute to the development of more efficient algorithms for InSAR data analysis and interpretation.

In susceptibility mapping, the integration of high-resolution remote sensing data, such as InSAR time series analysis, can enhance the accuracy of terrain characterisation and hazard identification. Furthermore, the integration of real-time monitoring data, such as precipitation data or seismic measurements, can improve the temporal aspect of ground subsidence susceptibility mapping and enable timely warnings and hazard assessment. Incorporating advanced modelling techniques, such as machine learning algorithms, can help capture the complex interactions between different factors and improve the predictive capabilities of susceptibility mapping models.

2.4 Review on Ground Subsidence Monitoring and Susceptibility Mapping, and InSAR Tools

This section provides a critical review of existing case studies on ground subsidence monitoring, susceptibility mapping, and the combined approach of these techniques in South Africa and globally. The review aims to deepen understanding of methodologies, identify research gaps, and highlight the integration of InSAR time series analysis and machine learning for effective ground subsidence monitoring and susceptibility mapping.

2.4.1 Case Studies of Ground Subsidence Monitoring

Abdikan et al. (2011) investigated ground subsidence in the Kozlu mine, Turkey, using ENVISAT ASAR images, ROI_PAC, and StaMPS for InSAR time series analysis in both PS and SBAS modes. Their findings of ground subsidence rates ranging from 20 to 140 mm between 2003 and 2010, highlight the importance of employing multiple InSAR techniques for accurate monitoring (Abdikan et al., 2011). However, the study does not thoroughly address potential errors or limitations associated with these methods, revealing a critical gap in understanding the effectiveness of different techniques and tools.

Blöthe (2018) investigated landslide detection using SNAP, StaMPS, and PS and SBAS InSAR across various locations. The study highlighted significant differences between PS and SBAS results, emphasising the importance of careful technique selection (Blöthe, 2018). However, it does not explore the practical implications of these differences, pointing to an area for further investigation.

Ajayi et al. (2023) applied various InSAR processing tools, including HyP3, MintPy, SNAP, and StaMPS, to monitor ground deformation in Warri, Nigeria. The study reported varying ground deformation rates depending on the tool used, which underscores the variability in results (Ajayi et al., 2023). However, it lacks a comparative analysis of the accuracy and reliability of these tools, indicating a need for further research into how tool selection impacts the accuracy of ground subsidence monitoring.

Hrysiewicz et al. (2023) analysed ground deformation in Campi Flegrei Caldera and Long Valley Caldera using Sentinel-1, ISCE, MintPy, StaMPS, and PS and SBAS InSAR. The results showed significant ground deformation rates in Campi Flegrei Caldera but less pronounced in Long Valley Caldera (Hrysiewicz et al., 2023). This comprehensive approach highlights the importance of multi-method validation but could benefit from a discussion on the implications of these findings for ground subsidence monitoring.

Dong et al. (2023) monitored mine dump ground subsidence using Sentinel-1, ISCE, StaMPS, SBAS-InSAR, and GNSS RTK positioning, showing maximum ground deformation of 20 mm weakly associated with rainfall (Dong et al., 2023). The study highlights the need for further

research into additional contributing factors, given its limitations in correlating ground deformation with rainfall.

Argyrakis et al. (2020) used cGNSS data to study ground deformation in Thessaly, Greece, observing up to 550 mm of cumulative displacement correlated with groundwater decline. While the study demonstrates the effectiveness of cGNSS for capturing long-term trends, the reliance on a single data source limits its comprehensiveness (Argyrakis et al., 2020). Integrating additional data sources could strengthen the study's findings and provide a more complete picture.

Aslan and Cakir (2019) examined ground subsidence factors in Northwest Turkey using data from TerraSAR-X, Sentinel-1A, GMTSAR, StaMPS, and cGNSS. They reported creep rates and secular deformation, highlighting the impact of groundwater exploitation on ground subsidence (Aslan and Cakir, 2019). To corroborate and expand upon these findings, additional data sources should be integrated, addressing the study's current limitations.

Dauteuil et al. (2023) focused on rainfall-induced ground subsidence in Southern Africa using GPS and GRACE data, revealing a 2 to 3-month lag (Dauteuil et al., 2023). While this provides useful insights, the study's scope may limit its broader applicability.

Dawod et al. (2023) analysed ground subsidence using cGNSS stations, noting ground deformation rates between 2.6 and -7.2 mm/year (Dawod et al., 2023). The study's emphasis on the importance of considering ground deformation in infrastructure planning is crucial, though further discussion on implications for different infrastructure types could provide more actionable recommendations.

Chaussard et al. (2013) studied ground subsidence in western Indonesia using ALOS PALSAR, ROI_PAC, and SBAS InSAR, reporting rates up to 220 mm/year due to groundwater extraction (Chaussard et al., 2013). While focusing on high-risk areas is valuable, a broader examination of contributing factors would enhance the understanding of ground subsidence in these regions.

Cian et al. (2019) analysed ground subsidence rates in Nigeria and Gambia using Sentinel-1, TerraSAR-X, COSMO-SkyMed, Envisat-ASAR, SNAP, StaMPS, and PS-InSAR (Cian et al., 2019). Their integration of various data sources highlights regional variations in ground subsidence but would benefit from a more detailed analysis of discrepancies to better understand their causes (Cian et al., 2019).

Cigna et al. (2021) assessed the accuracy of Sentinel-1, SNAP, StaMPS, and PS and SBAS InSAR in Mexico, revealing vertical ground deformation rates up to -150 mm/year (Cigna et al., 2021). The study's validation with cGNSS and levelling data is robust; however, discussing the practical implications of these results could enhance their applicability to real-world scenarios.

Ahmad et al. (2022) employed Sentinel-1 imagery, SNAP, StaMPS, and PS-InSAR to analyse ground subsidence in Pakistan, uncovering ground deformation rates from 15 mm/year to -114 mm/year. They identified groundwater extraction, mining, and construction overloading as significant contributors (Ahmad et al., 2022). Although their study offers valuable insights into human-induced ground subsidence, a more detailed analysis of the interactions between these factors would provide a richer understanding of their combined effects.

Papoutsis et al. (2020) investigated ground deformation in Greece with Sentinel-1, ISCE, StaMPS, and PS InSAR, reporting ground deformation rates from 10 mm/year to -10 mm/year (Papoutsis et al., 2020). The study adds to the understanding of regional ground deformation patterns but would benefit from a more detailed exploration of causative factors.

Ayani et al. (2021) studied crustal deformation in Surabaya, Indonesia, using Sentinel-1, ISCE, GIAnT, and SBAS InSAR, revealing LOS velocities ranging from -14.8 to 10.8 mm/year (Ayani et al., 2021). While the study contributes to understanding ground subsidence, further contextual analysis would be beneficial.

Kirui et al. (2022) investigated ground subsidence in Nairobi, Kenya, using ISCE, StaMPS, and SBAS InSAR techniques. Their findings of up to 62 mm/year ground subsidence related to groundwater level decline are significant (Kirui et al., 2022). However, a deeper exploration of urbanisation and other factor impacts could provide more insights into the factors influencing ground subsidence.

Bagheri-Gavkosh et al. (2021) conducted a review of 290 case studies across 41 countries, finding that InSAR was used in 38% of monitoring methods for monitoring ground subsidence. Their review provides insight into global trends, particularly human-induced ground subsidence and groundwater-related issues (Bagheri-Gavkosh et al., 2021). However, it lacks depth in individual case analyses, suggesting a need for more detailed studies in specific regions.

These case studies collectively highlight the critical importance of using a range of InSAR techniques and processing tools to accurately monitor ground subsidence; moreover, they highlight the prevalence of SNAP, ISCE, HyP3, MintPy, and StaMPS for InSAR pre-processing and processing. Furthermore, the observed variability in InSAR results and the gaps identified in understanding tool effectiveness emphasize the need for further research into comparative analyses and methodological improvements. Addressing these gaps will support the selection/development of more accurate and reliable ground subsidence monitoring methods, aligning with the objective of advancing our understanding of ground subsidence.

2.4.2 Review on Automation in SNAP, ISCE, HyP3, MintPy, and StaMPS

This section focuses on the automation capabilities and limitations of the prevalent InSAR processing tools, namely SNAP, ISCE, HyP3, MintPy, and StaMPS. While ISCE, HyP3,

MintPy, and StaMPS have been designed with built-in automation capabilities, such as ISCE's use of StackSentinel, HyP3's integration with GAMMA, and MintPy and StaMPS reliance on command-line scripting. On the other hand, SNAP presents a distinct challenge in this regard; Although it does not inherently support full automation, SNAP offers some automation potential through its Graph Processing Framework, which can be extended to semi-automate workflows.

Despite the availability of processors like *SNAP2StaMPS* and *SBAS_SNAP2StaMPS* for SAR SNAP automated data processing, these tools have limitations that hinder seamless automation. They support only one processing mode at a time (either PS or SBAS), rely on outdated Python versions, and face installation difficulties due to multiple dependencies. Furthermore, infrequent software updates, lack of support for processing data across multiple swaths, and the need for manual sorting of datasets adds additional layers of complexity. Additional drawbacks include limited support for digital elevation models, the absence of integrated error monitoring, and increased storage requirements due to data duplication. Together, these limitations reduce the efficiency and compatibility of these tools, especially with newer software versions, making them less effective for comprehensive ground deformation monitoring.

To overcome some of these challenges, the development of the SNAP to StaMPS Automated Processing (StoSAP) becomes increasingly important. StoSAP aims to streamline automation by integrating SNAP's GPF with StaMPS workflows, thereby addressing the issue of automation support within SNAP. This development can significantly enhance efficiency in SAR data processing and provide a more user-friendly environment for ground subsidence monitoring and susceptibility mapping.

2.4.3 Case Studies of Susceptibility Mapping for Ground Subsidence

Alogayell et al. (2021) employed nine factors namely, lithology, land cover/land use, elevation, slope, aspect, rainfall, distance to faults, earthquake epicentre, and distance from streams, along with an integrated AHP and MCDMA approach for ground subsidence susceptibility mapping in Saudi Arabia. Their study achieved a high accuracy of 97.1% when validated against 52 ground subsidence data locations. The findings revealed that lithology, rainfall, distance from faults, and earthquake epicentres were the most significant factors contributing to ground subsidence (Alogayell et al., 2021). This highlights the importance of incorporating diverse factors for accurate susceptibility mapping and provides a benchmark for assessing similar methodologies in other regions.

Gharechaee et al. (2023) applied Random Forest, KNN, and CART machine learning models to landslide susceptibility mapping using twenty factors and InSAR data. Their results indicated that high and very high landslide susceptibility classes were predominantly located on or near irrigated agricultural land. Contrary to expectations, groundwater withdrawals were

not the primary cause of ground subsidence in the study region. Instead, the distance from dams, and proximity to anticlines, faults, and mines were identified as the most crucial factors. The models' accuracy, assessed using the coefficient of determination, was 0.77 for Random Forest, 0.69 for KNN, and 0.71 for CART (Gharechaee et al., 2023). This study underscores the need for region-specific analyses and highlights the varying impacts of different factors on susceptibility.

Fadhillah et al. (2020) used ten factors, StaMPS PS InSAR data, and ensemble machine learning models for ground subsidence susceptibility mapping in Seoul. They achieved ground deformation rates between 6 and 12 mm/year, with a dominance of Very Low susceptibility classes across 64.44% of the study area. Their use of the frequency ratio to analyse the spatial correlation between ground subsidence and each factor revealed that land use and groundwater extraction were the primary influences on ground subsidence risk. The models demonstrated a strong performance with an AUC of up to 0.883 (Fadhillah et al., 2020). This study illustrates the effectiveness of ensemble models in capturing complex ground subsidence susceptibility patterns and provides a foundation for improving ground subsidence susceptibility assessments in other urban areas.

Bagheri et al. (2024) applied Convolutional Neural Networks (CNNs) with clustering and regression modelling strategies to predict the Ground Subsidence Vulnerability Index (GSVI) in the Tabriz aquifer, northwest Iran. This region faces severe water table decline due to groundwater over-abstraction. The study incorporated 7 ALPRIFT data layers and ground subsidence maps from 15 SAR (ALOS-2 PALSAR-2) images. The CNN models, including Regression-CNN and Clustering-CNN, significantly improved accuracy compared to the traditional ALPRIFT model, achieving AUC values of 0.87, 0.83, and 0.56, respectively (Bagheri et al., 2024). This highlights the potential of CNNs in reducing subjectivity and enhancing the accuracy of vulnerability assessments.

Hakim et al. (2023b) used Improved Combined Scatterers Interferometry with Optimised Point Scatterers (ICOPS) to enhance measurement point density for ground subsidence analysis in Semarang, Indonesia. They compared ICOPS optimization using Support Vector Regression (SVR) and CNN, with ICOPS-CNN showing superior performance and higher measurement point density. The standalone CNN algorithm achieved an accuracy of 0.833 in ground subsidence susceptibility mapping, outperforming SVR (0.820). The hybrid CNN with Grey Wolf Optimiser (CNN-GWO) achieved the highest accuracy at 0.907, followed by SVR-GWO at 0.843. Factors such as lithology (IGR = 0.97) and elevation (IGR = 0.92) were identified as crucial contributors (Hakim et al., 2023b). This study demonstrates the benefits of advanced optimization techniques, hybrid models, and effective factor selection in improving ground subsidence susceptibility assessments.

Eghrari et al. (2023) assessed ground subsidence susceptibility in the Kashan Plain, Iran, using Random Forest and XGBoost machine learning algorithms. They incorporated twelve factors including, elevation, groundwater decline, distance to faults, slope, aspect, and lithology, using 291 ground subsidence points for training (70%) and testing (30%) the models. XGBoost outperformed Random Forest with an R^2 of 0.9032, RMSE of 0.3764 cm, and MAE of 0.1217 cm, compared to Random Forest's R^2 of 0.8355, RMSE of 0.4906 cm, and MAE of 0.3050 cm. Elevation, groundwater decline, and distance to faults were identified as the most influential factors (Eghrari et al., 2023). This highlights the effectiveness of XGBoost in improving prediction accuracy and provides valuable insights into key factors influencing ground subsidence.

Chai et al. (2024) assessed ground subsidence risk in the Shanghai Metro network using PS-InSAR for ground subsidence monitoring and ten causal factors, including distance to rivers, distance to roads, TWI, and slope. The LightGBM machine learning algorithm was employed for ground subsidence susceptibility analysis, achieving the highest performance with an AUC of 0.902 and an accuracy of 0.817. This outperformed GBDT (AUC = 0.895, accuracy = 0.811) and Random Forest (AUC = 0.855, accuracy = 0.772). The analysis revealed higher ground subsidence risk in the southwest of the study area, with risk levels varying across the metro network (Chai et al., 2024). This study demonstrates the effectiveness of LightGBM over other algorithms in providing accurate detailed risk assessments.

Lee et al. (2023) evaluated the risk of ground subsidence in urban areas using attribute information of underground pipelines and historical ground subsidence data. The study divided the target area into grids of 100m×100m, 300m×300m, and 500m×500m, creating datasets based on pipeline diameter, years of use, and density. The LightGBM achieved the highest performance with an F1-score of 0.75 and an AUC of 0.83, followed by XGBoost (F1-score: 0.66, AUC: 0.85) and Random Forest (F1-score: 0.67, AUC: 0.84). LightGBM proved to be the most effective for predicting ground subsidence risk (Lee et al., 2023). This study highlights the importance of integrating infrastructure data into ground subsidence susceptibility models to improve predictive accuracy, moreover the effectiveness of LightGBM over other algorithms in providing accurate detailed risk assessments.

These case studies collectively underscore the crucial importance of incorporating diverse factors and employing advanced machine learning techniques in ground subsidence susceptibility mapping. The widespread use of CNN, LightGBM, Random Forest, and XGBoost techniques highlights their significance in this field, with hydrological factors (such as groundwater extraction) and land use/land cover factors (such as agriculture) being particularly prominent. The variations in accuracy and effectiveness observed across different ML models and techniques underscore the ongoing need for research to refine methodologies and enhance predictive capabilities. Additionally, these case studies reveal a gap in the explainability of ML models, which is essential for understanding the causative factors of

ground subsidence as inferred by the models. To address this, SHAP has been adopted in this study to improve model explainability.

2.4.4 Case Studies of Ground Subsidence and Susceptibility Mapping in South Africa

Engelbrecht and Inggs (2013) investigated mining-induced ground subsidence in the Witbank Coalfields, Gauteng, using D-InSAR and ALOS PALSAR data. Their results revealed a strong correlation between D-InSAR ground deformation measurements and the locations of active underground mining activities, underscoring the potential value of incorporating D-InSAR in ground subsidence monitoring. However, the study also identified limitations in D-InSAR effectiveness, notably signal decorrelation caused by vegetation dynamics, which impacted the accuracy of ground subsidence measurements. Additionally, phase noise effects observed in interferograms further hindered phase unwrapping, thereby reducing the precision of the ground subsidence data (Engelbrecht and Inggs, 2013).

Theron (2017) investigated the correlation between ground subsidence and sinkhole formation, revealing that ground subsidence frequently precedes sinkhole collapse. By employing D-InSAR, Theron identified three previously unknown ground deformation basins, which were subsequently confirmed through field observations. The study also documented a water supply pipeline burst occurring 6 months after confirming the ground deformation basins, underscoring the practical impacts of ground subsidence. The maximum sinkhole displacement recorded using D-InSAR was 70 mm over 55 days, however decorrelation issues in InSAR were noted. This case underscores the critical importance of timely detection and monitoring to mitigate ground subsidence-related risks, also addressing signal decorrelation in InSAR using time-series analysis (Theron et al., 2017).

Constantinou and Van Rooy (2018) examined karst-related sinkholes and ground subsidence in Gauteng, South Africa, focusing on the severe impact of these phenomena in the region underlain by Chuniespoort Group dolomite bedrock. Despite the widespread presence of dolomite across several South African provinces, Gauteng has experienced a disproportionately high number of sinkhole and ground subsidence events over the past 50 years. The study highlighted the fragmented and incomplete nature of historical data, previously gathered by various entities without a central reporting requirement. Their findings underscore the pressing need for a monitoring method that can capture both historical and current data for deeper insights into ground subsidence and sinkhole patterns in Gauteng (Constantinou and Van Rooy, 2018). While the study highlights the importance of a centralised database to improve future assessments and decision-making, the reliance on available but often incomplete data limits the understanding of the full scope and dimensions of these hazards.

Isiaka et al. (2019) conducted a high-resolution seismic reflection survey in Benoni, near Johannesburg, to investigate ground subsidence and sinkholes at an abandoned coal mine site,

where a lack of historical mining records and the presence of weathered dolomitic rocks complicate ground subsidence assessment. The study revealed that ground subsidence is primarily due to the collapse of the colliery roof into voids formed by coal extraction at depths of 7 to 13 metres and the presence of dissolution cavities in the dolomite formation 50 metres below the surface (Isiaka et al., 2019). While the findings highlight significant ground subsidence challenges in Gauteng, the study's reliance on time-consuming, expensive, and low-coverage seismic techniques limits broader monitoring efforts, underscoring the need for more efficient approaches to managing ground subsidence risks.

Montjane et al. (2020) investigated the severe karstic sinkhole development in Centurion, a rapidly growing area underlain by fractured dolomite formations of the Malmani Dolomite Group. The study utilised fieldwork mapping, petrology, and remote-sensing techniques to analyse the influence of geological fractures on sinkhole formation. Findings revealed that sinkhole development is primarily controlled by the underlying fracture systems, which direct fluid circulation and enhance karstic processes like dissolution. Additionally, the study highlighted that sinkhole occurrence is exacerbated by human activities. However, the study's focus on a single geological formation (dolomite) and the reliance on time-consuming, low-coverage, and high-cost observation methods limit its scope (Montjane et al., 2020). Incorporating additional factors such as topography, hydrology, and land use/land cover, as well as employing high coverage monitoring methods such as InSAR, could provide deeper insights into the spatial and temporal dynamics of ground subsidence and sinkhole formation.

These case studies collectively highlight the critical need for improved monitoring, susceptibility mapping, and comprehensive analysis of ground subsidence. They underscore the challenges associated with current methodologies, including limitations due to signal decorrelation, high costs, and low coverage of observation techniques. Additionally, the limited integration of these findings into practical mitigation strategies highlights a gap in the application of these technologies. Addressing these gaps, alongside the adoption of innovative techniques such as InSAR time series analysis and machine learning, is essential for enhancing the accuracy and timeliness of ground subsidence susceptibility mapping. Moreover, developing integrated approaches that incorporate diverse factors such as topography, hydrology, and land use, as well as employing high coverage monitoring methods, will improve the understanding and management of ground subsidence and sinkhole risks in the Midvaal region and beyond.

2.5 Chapter Conclusion

This chapter has reviewed key literature on ground subsidence monitoring and susceptibility mapping, with a focus on InSAR time-series analysis and machine learning techniques. It highlighted the prevalent use of open-source software such as SNAP, ISCE, HyP3, StaMPS,

and MintPy, prevalent use of techniques such as PS-InSAR and SBAS-InSAR and noted a critical gap in comparative studies of their results.

The review also identified gaps related to the accuracy of DEMs and their effects on ground subsidence susceptibility mapping. The importance of water and land use/land cover (vegetation) factors in ground subsidence was also noted, supporting the need for spatio-temporal analysis incorporating NDVI, MNDWI, groundwater storage anomaly, and precipitation. The methodology will include a detailed examination of these factors to enhance understanding of ground subsidence patterns.

Moreover, the review highlighted the need for improved ground subsidence monitoring and susceptibility mapping by leveraging InSAR time-series analysis and machine learning algorithms. It also emphasized the importance of understanding correlations between ground subsidence and related factors using advanced analytical tools such as SHAP and frequency ratio analysis.

In summary, the literature review establishes the need for further investigation into InSAR software and technique performance, DEM accuracy, and the application of InSAR time series analysis and machine learning techniques for ground subsidence monitoring and susceptibility mapping. The subsequent methodology chapter will address these needs and provide a robust framework for advancing ground subsidence research.

3 Methodology and Materials

In this chapter, the methodologies, software, tools, platforms, and datasets employed in this investigation to achieve the research objectives and answer the research questions are outlined. Additionally, the diagram (Figure 3.1) presented below provides a visual representation of the methods employed, the computations and statistical analysis to be conducted, and the quantitative results that will be discussed in the conclusion.

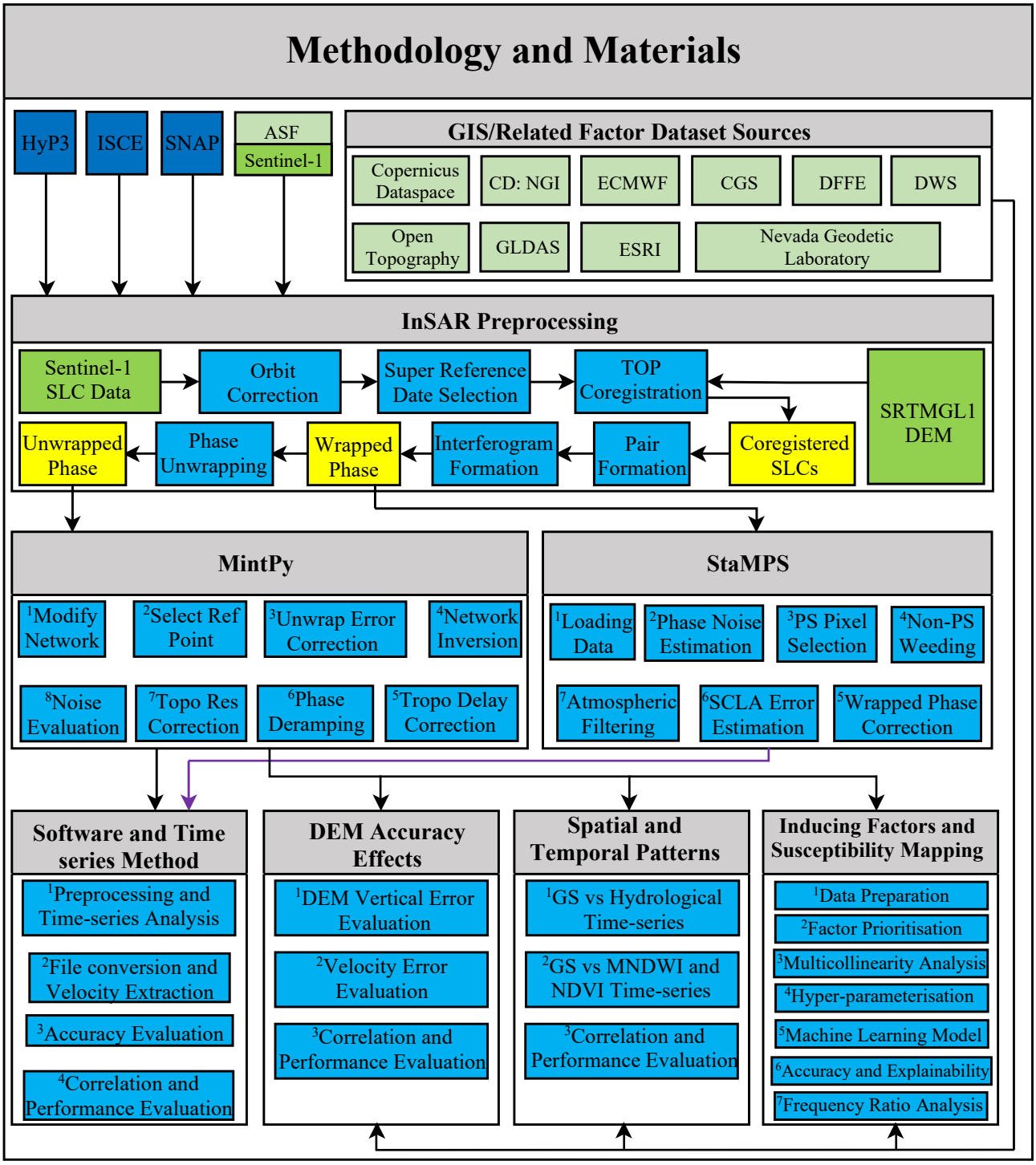


Figure 3.1: Generalised Methodological Framework for Research Design.

This chapter delineates the methodology employed to develop a comprehensive InSAR time series analysis approach for monitoring and predicting ground subsidence in Midvaal, Gauteng. It commences with an examination of existing InSAR time series analysis methods and open source processing tools, including SNAP, ISCE, ASF HyP3, MintPy, and StaMPS. These tools are assessed based on their transparency, documentation quality, and efficacy in ground deformation monitoring. A significant challenge identified is the scarcity of automated tools for SNAP PS and SBAS InSAR stack processing, which impacts both efficiency and reliability.

To address this issue, the chapter introduces StoSAP, a semi-automated stack processor designed to improve the pre-processing of both Persistent Scatterer and SBAS InSAR data within the SNAP framework. StoSAP aims to overcome automation limitations and enhance functional capabilities. Subsequently, the chapter evaluates the impact of software selection and methodological choices on ground subsidence monitoring accuracy. This is achieved through a comparative analysis of PS and SBAS InSAR processing across various software combinations, with validation against continuous GNSS stations.

The chapter further investigates the effects of Digital Elevation Models on the accuracy of ground subsidence prediction and monitoring. It explores the relationship between DEM accuracy and ground deformation measurements, employing cGNSS and ground levelling data to mitigate error impacts. Additionally, methodologies for analysing InSAR data to discern spatio-temporal trends in ground subsidence are outlined, with a focus on ground subsidence rates and patterns in Midvaal.

Furthermore, the integration of machine learning algorithms with InSAR time series analysis data is explored to develop predictive models of ground subsidence susceptibility. This approach aims to deepen the understanding of ground subsidence dynamics.

3.1 Study Area

The study is conducted in Midvaal, South Africa, which located in the north-eastern part of South Africa, and spans approximately between latitudes 26° 18’ 38’’ S and 26° 56’ 53’’ S, as well as longitudes 27° 50’ 40’’ E to 28° 25’ 52’’ E. This region has distinct climate pattern characterised by a rainy season that typically prevails from November through March. During this period, the highest average monthly precipitation of approximately 150 mm in January and the lowest values of about 10 mm in June and July are recorded. Midvaal study area covers an estimated ground area of 4,126.34 square kilometres, other than the Midvaal local municipality, which is about 1,722 square kilometres. It features a densely populated rural-urban landscape, with a population size of 112,254 residents as of 2022, making it one of the populous regions in South Africa. Midvaal is also known for its economic significance and urban development. Figure 3.2 provides an overview of the geographical extent of the region, and it geological features.

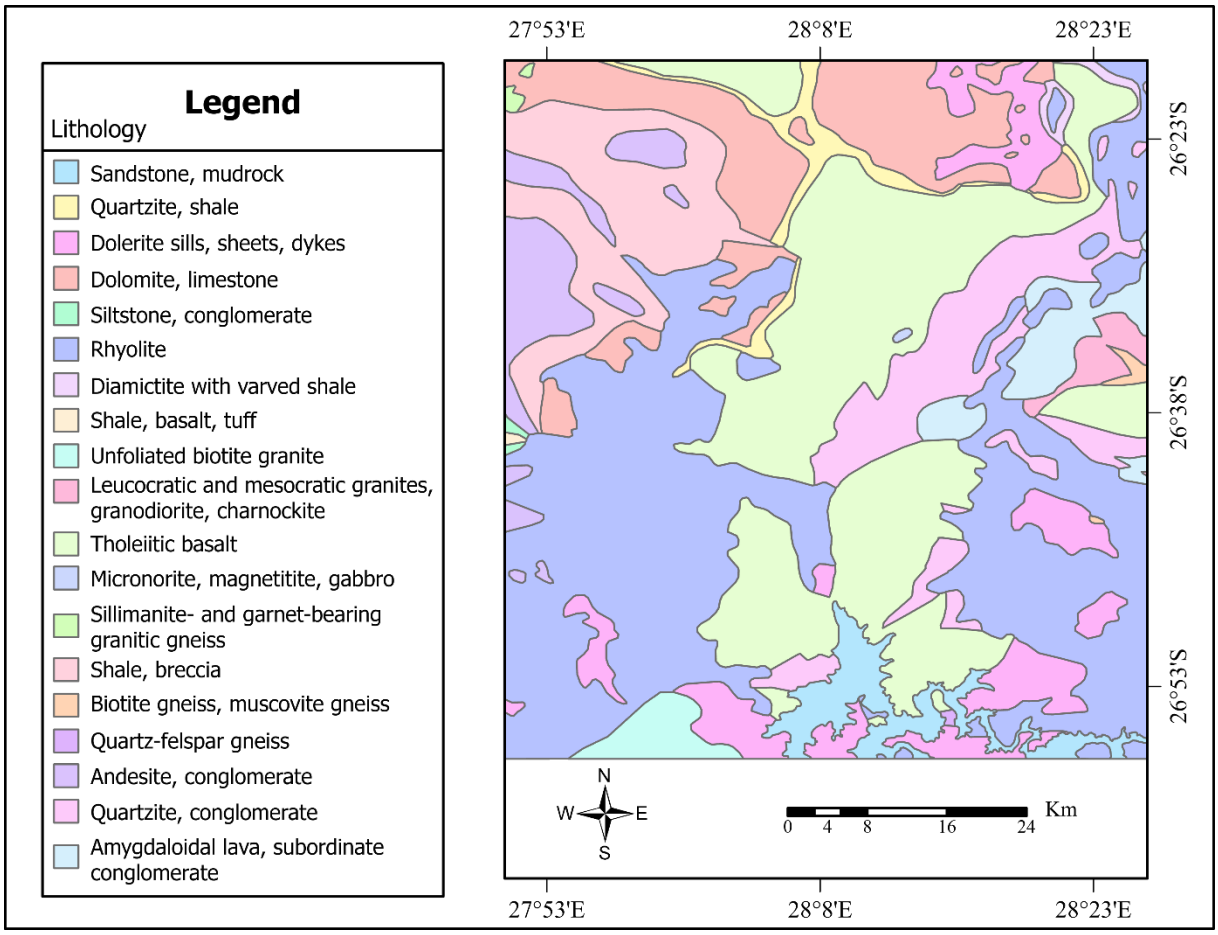


Figure 3.2: Midvaal, Gauteng geological formation.

Midvaal, Gauteng geological characteristics are diverse and play a crucial role in shaping its landscape. The region's geological history has led to the presence of various rock formations and mineral deposits. Notable geological features in Midvaal include the Witwatersrand Basin, which is renowned for its gold deposits, as well as a variety of sedimentary rocks, volcanic formations, and granite intrusions.

The Witwatersrand Basin, in particular, has had a profound impact on the region history and economy. It is one of the world's richest gold-bearing regions and has played a central role in South Africa's mining industry for over a century. The province's geological diversity has also contributed to the presence of valuable minerals like platinum, chromium, and manganese.

The geological distribution and formations in Midvaal have significant implications for groundwater availability and quality. The presence of certain rock types, such as dolomite, can create unique hydrogeological conditions, including the formation of extensive underground aquifers. However, the vulnerability of these aquifers to pollution and ground subsidence poses challenges for sustainable groundwater management.

Moreover, rapid urbanisation, agricultural activities and industrialisation have placed considerable stress on Midvaal's water resources. High population density, rapid urban growth, and industrial activities have increased water demand, leading to concerns about excessive

underground water extraction, potential groundwater contamination, ground subsidence and sinkhole formation.

3.2 Computing Platforms, Software's and Tools

This study leverages diverse array of computing platforms, software, and tools used in the pre-processing and analysis of SAR data for PS and SBAS InSAR, as well as InSAR time series analysis applications. The pre-processing of PS and SBAS data relies on the robust functionalities of SNAP and ISCE software packages, while InSAR time series analysis is facilitated by StaMPS running on the UCT High Performance Computing (HPC) Facility platform, specifically leveraging the ada partition with 4TB free storage, 384GB of RAM and 40 CPUs per user. Conversely, SBAS InSAR processing is also conducted using HyP3 and MintyPy via OpenScienceLab (OpenSARLab) on the cloud, offering 500GB of free storage, RAM configurations ranging from 10GB to 16GB, and a CPU limit of 4 CPUs per user.

Notably, the processing automation in ISCE uses the stackSentinel.py processor, while SNAP employs StoSAP.py processor and HyP3 leverages Gamma Remote Sensing software. A key advantage of UCT HPC over OpenSARLab lies in its backup power infrastructure, ensuring uninterrupted processing even during power outages, and providing analysts with greater flexibility in software installation and customisation, moreover the large storage space, RAM and number of CPUs makes it a great choice. However, UCT HPC has limitations in that it is only available to researchers affiliated with University of Cape Town, while OpenSARLab is completely open-source and available to all public users.

3.3 Data Sources

3.3.1 Sentinel-1 SLC

The Sentinel-1 mission consists of a pair of satellites, Sentinel-1A and Sentinel-1B. The launch of Sentinel-1A in April 2014 marked a milestone in Earth observation technology. Equipped with a C-band SAR, it provided valuable data for applications such as ground deformation monitoring, disaster management, and environmental surveillance. However, initial observations within specific regions were marred by data gaps, limiting its utility for complex multitemporal analysis required for research.

Fortunately, the launch of Sentinel-1B in April 2016 brought a significant enhancement to the Sentinel-1 mission. It addressed the data consistency issues observed with its predecessor and substantially improved the reliability and quality of observations. Consequently, our study benefited from a more coherent dataset, featuring reduced revisit times starting in 2017; this transition from a 24-day to a 12-day revisit cycle significantly boosted the temporal resolution of our analysis. Also, this leap in data quality had profound implications for various scientific disciplines and practical application.

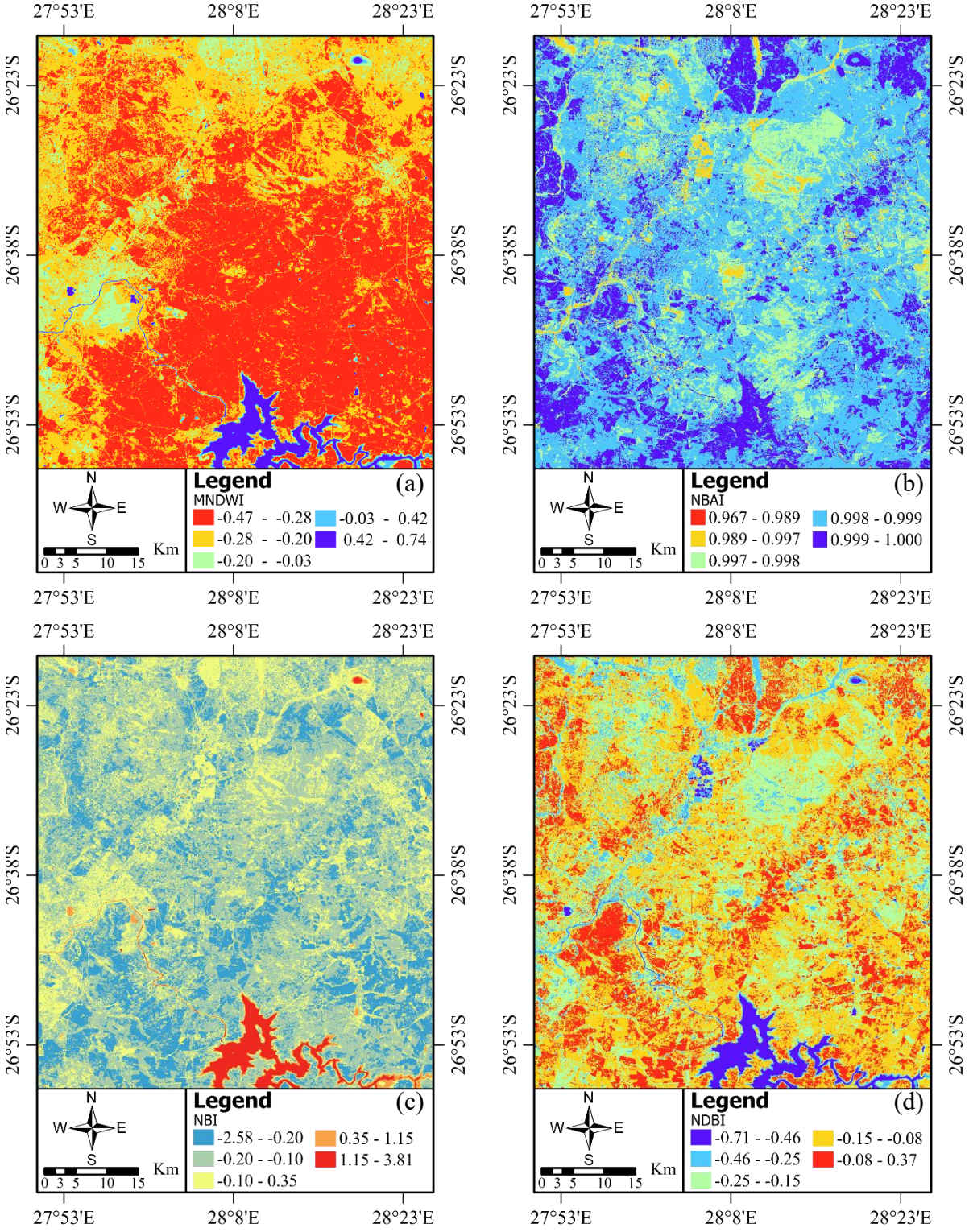
Thus, this study uses Sentinel-1 C-band imagery, the choice of Sentinel-1 C-band imagery for monitoring ground deformation in this research is well-justified based on several key factors. Firstly, the C-band frequency (5.5cm wavelength) is known for its ability to penetrate vegetation cover effectively, making it ideal for detecting changes in the Earth's surface beneath forests or agricultural areas. Additionally, the Interferometric Wide (IW) mode of Sentinel-1 offers three sub-swaths, each independently processable, providing wide coverage (250 kilometres swath width) and enabling comprehensive monitoring of ground deformation. The spatial resolution of 5 metres in the range and 20 metres in the azimuth direction ensures detailed observation of surface features, crucial for accurately capturing and analysing ground deformation processes.

Moreover, the extensive temporal coverage spanning from January 2019 to December 2021, with a total of 84 and 90 selected images in descending and ascending orbit direction respectively, allows for the detection of both short-term fluctuations and long-term trends in the study area. By consistently including frame 677 from the descending orbit path 50, and frame 1089 from the ascending orbit path 116, potential biases or variations in data acquisition are minimised, enhancing the reliability and consistency of the analysis. Finally, the direct sourcing of Sentinel-1 data from the ASF Distributed Active Archive Centre (ASF DAAC) ensures the data's quality, reliability, and accessibility for researchers (Hakim et al., 2023b; Kirui et al., 2022). These characteristics collectively make Sentinel-1 C-band imagery a highly suitable choice for achieving the research objectives of monitoring ground deformation in the region.

3.3.2 Sentinel-2 L2A

The Sentinel-2 satellite, launched on 23 June 2015, as part of the European Union's Copernicus program, is a high-resolution multispectral imaging system capturing a series of images of the Earth's surface in multiple spectral bands. Since its launch, it has been operational, providing valuable data for various environmental and ground monitoring applications. This research study uses the abundant data resources offered by Sentinel-2 Level 2A imagery to explore various land cover factors related to ground subsidence.

Moreover, the top-of-canopy Sentinel-2 Level 2A data, in conjunction with ArcPy (ArcGIS Pro) software, is used to derive a suite of crucial GIS or ground subsidence related factors, including datasets such as the mean MNDWI, NBAI, NBI, NDBI, NDMI, NDVI, NDWI, and SAVI for the period from 01 January 2019 to 01 January 2022, which play pivotal roles in this study, as shown in Figure 3.3. The combination of Sentinel-2 L2A imagery with the ArcPy Raster Calculator and ArcPy Reclassify tool, using the Natural Jenks algorithm, enables us to generate valuable geospatial datasets that form the foundation of our study, enhancing our understanding of land use, land cover dynamics effect on ground subsidence, facilitating comprehensive analyses and informed decision-making.


27°53'E
28°8'E
28°23'E
26°23'S
26°38'S
26°53'S
N
W
E
S
Km
0 3 5 10 15
Legend
(a)
MNDWI
-0.47 - -0.28
-0.28 - -0.20
-0.20 - -0.03
-0.03 - 0.42
0.42 - 0.74
(b)
NBAI
0.967 - 0.989
0.989 - 0.997
0.997 - 0.998
0.998 - 0.999
0.999 - 1.000
(c)
NBI
-2.58 - -0.20
-0.20 - -0.10
-0.10 - 0.35
0.35 - 1.15
1.15 - 3.81
(d)
NDBI
-0.71 - -0.46
-0.46 - -0.25
-0.25 - -0.15
-0.15 - -0.08
-0.08 - 0.37

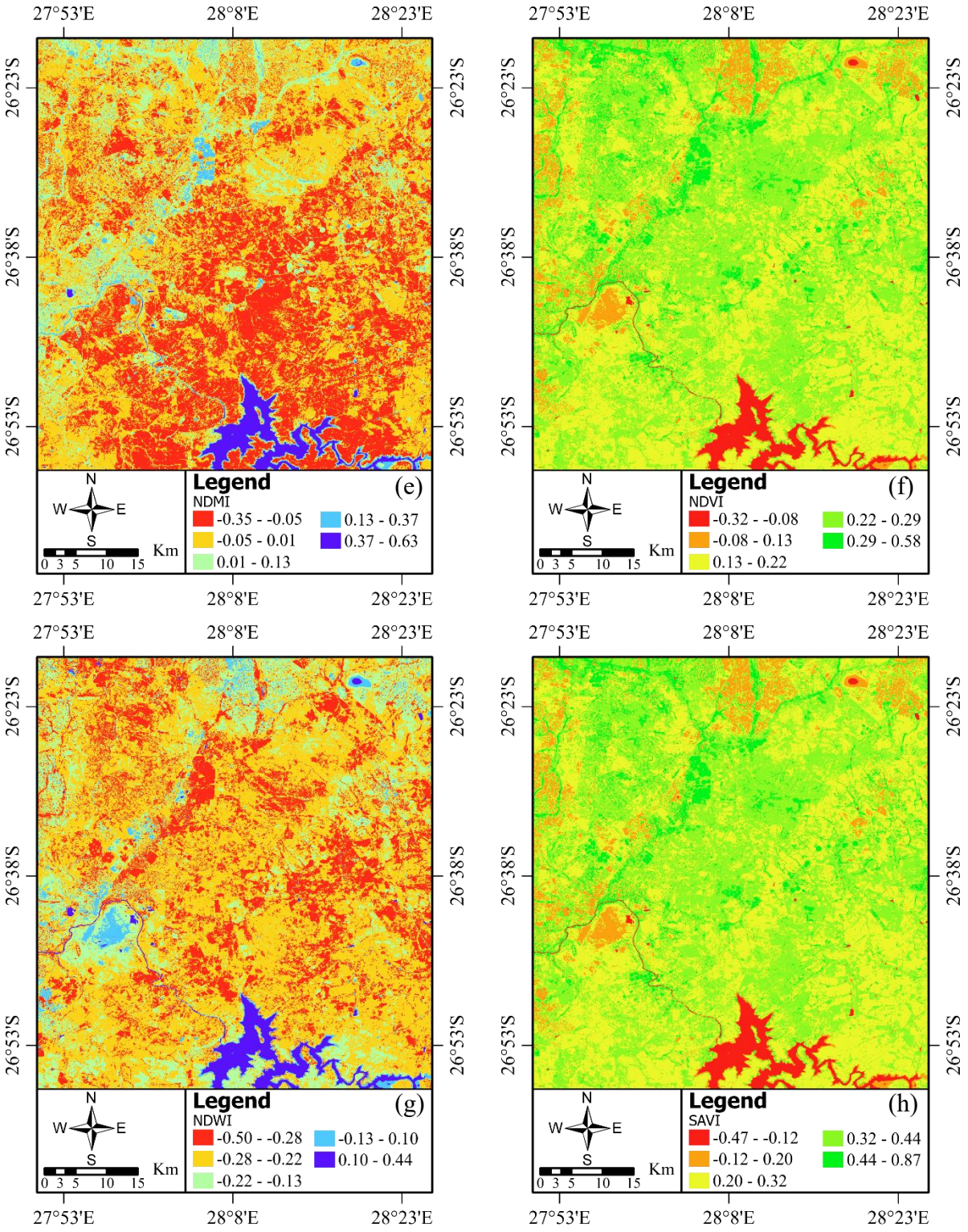


Figure 3.3: Ground subsidence related factors: (a) MNDWI, (b) NBAI, (c) NBI, (d) NDBI, (e) NDMI, (f) NDVI, (g) NDWI, (h) SAVI.

3.3.3 Continuous GNSS

The use of data from the cGNSS stations HEID and VERG, carefully referenced to the WGS84 ellipsoid, as shown in Figure 3.4, provides a solid basis for validation. These stations provide raw data originating from the CD: NGI and processed by the Nevada Geodetic Laboratory

(Blewitt et al., 2018). This results in ENU ground displacements that are critical for deriving higher order LOS velocities and displacements. The validation process is an essential part of ensuring the accuracy and reliability of the ground deformation analysis results. Through a comparative analysis with LOS velocities derived from Sentinel-1 C-band imagery, the consistency between the datasets is assessed, demonstrating the robustness of the results. In addition, the validation process identifies discrepancies between ground-based measurements and satellite-based observations, ensuring accuracy by investigating possible errors in data processing or environmental influences.

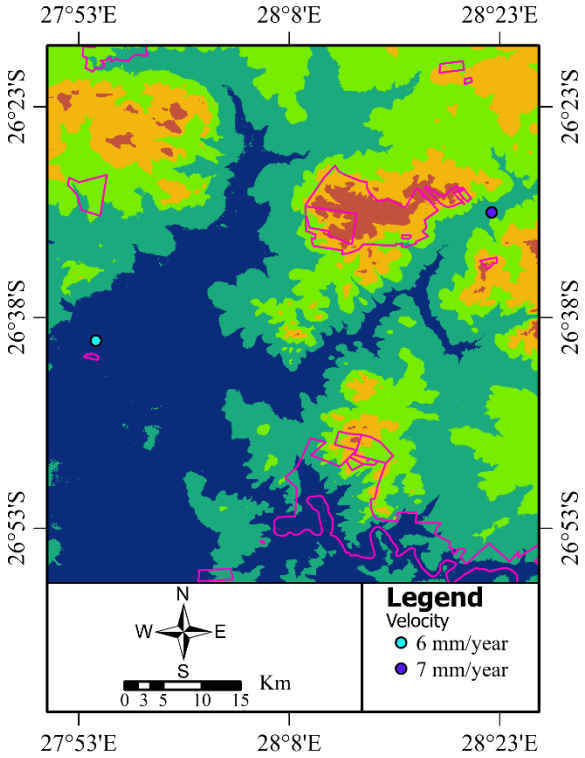


Figure 3.4: The cGNSS points distribution, station VERG in light blue and HEID in dark purple. The pink polygons represent the South African Protected Areas.

The validation results not only serve as a quality assurance mechanism, but also strengthen the scientific integrity of the study. Proof of agreement between the velocities derived from cGNSS/TrigNET and those from the Sentinel-1 images strengthens confidence in the accuracy and reliability of the analysis. This validation framework, underpinned by rigorous ground-based measurements, ensures the robustness of the results and facilitates informed decision-making in the monitoring and management of ground deformation.

3.3.4 Digital Elevation Models

To investigate various ground subsidence-related topographic and hydrological factors, this study uses DEM datasets as outlined in section 2.3.6, namely AW3D30, ASTER, SRTMGL1 and COP30 to support our analysis, as shown in Figure 3.5. These four DEMs provide a spatial resolution of 1 arc-second (30m). The datasets AW3D30, ASTER, COP30 and SRTMGL1 dataset are sourced from the Open Topography website, a trusted source for accessing high-

quality elevation data. The four DEMs are retrieved from the above source in GeoTIFFs format, AW3D30, ASTER, SRTMGL1 are referenced to WGS84 and EGM96 for horizontal and vertical datum respectively. While COP30 is referenced to WGS84 and EGM2008 for horizontal and vertical datum respectively.

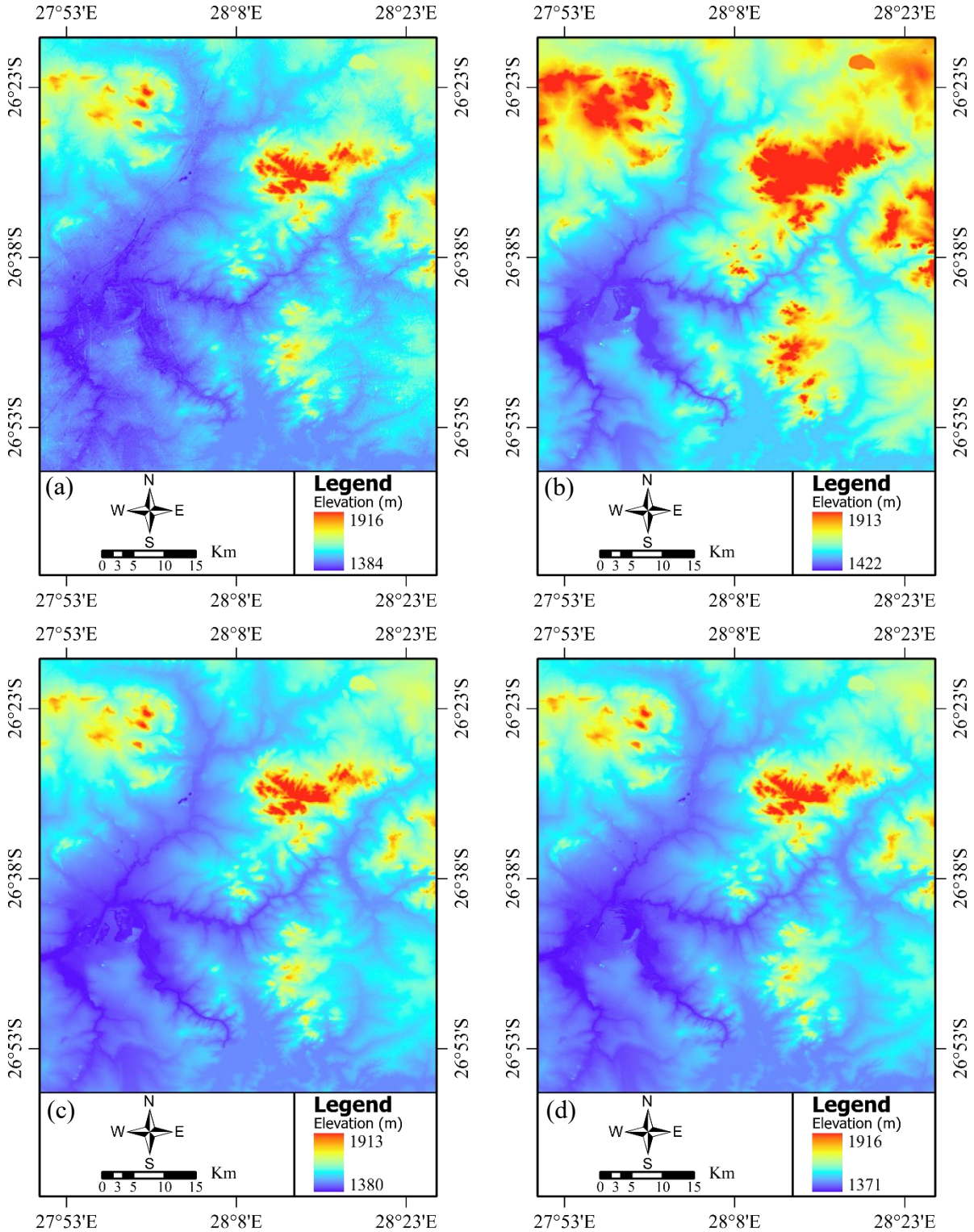
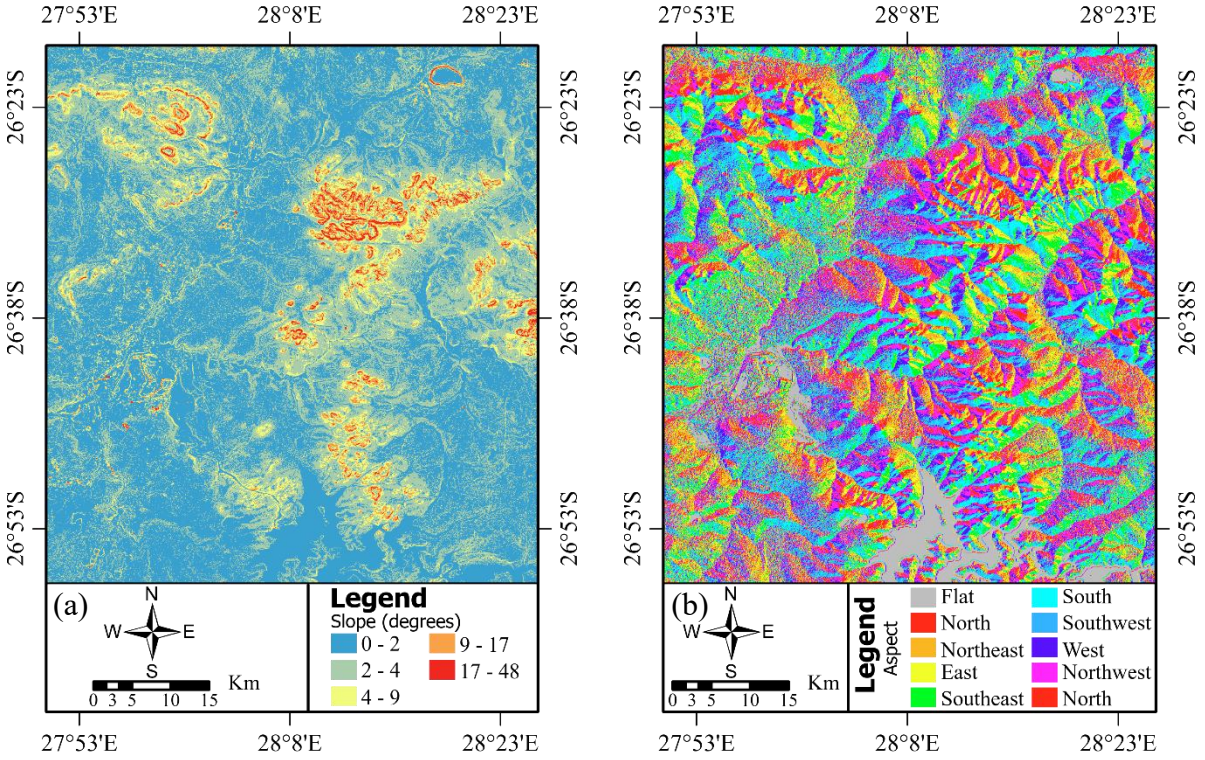


Figure 3.5: Digital Elevation Models in Midvaal, South Africa: (a) ASTER, (b) AW3D30, (c) COP30, and (d) SRTMGL1.

These DEM datasets plays a crucial role in our study, offering detailed elevation information that contributes to the accuracy and precision of our analyses. Their global coverage, high accuracy and fine spatial resolution render them an ideal choice for our research needs.

Moreover, the AW3D30 DEM data, in conjunction with ArcPy software, is used to derive a suite of crucial GIS or ground subsidence related factors, including datasets such as the slope, aspect, curvature (total curvature), and the topographic wetness index, which play pivotal roles in this study, as shown in Figure 3.6. The combination of AW3D30 DEM with the ArcPy Surface Parameters, Raster Calculator and ArcPy Reclassify tool, using the Natural Jenks algorithm, enables us to generate valuable geospatial datasets that form the foundation of our study, enhancing our understanding of topo-hydrological dynamics effect on ground subsidence, facilitating comprehensive analyses and informed decision-making.



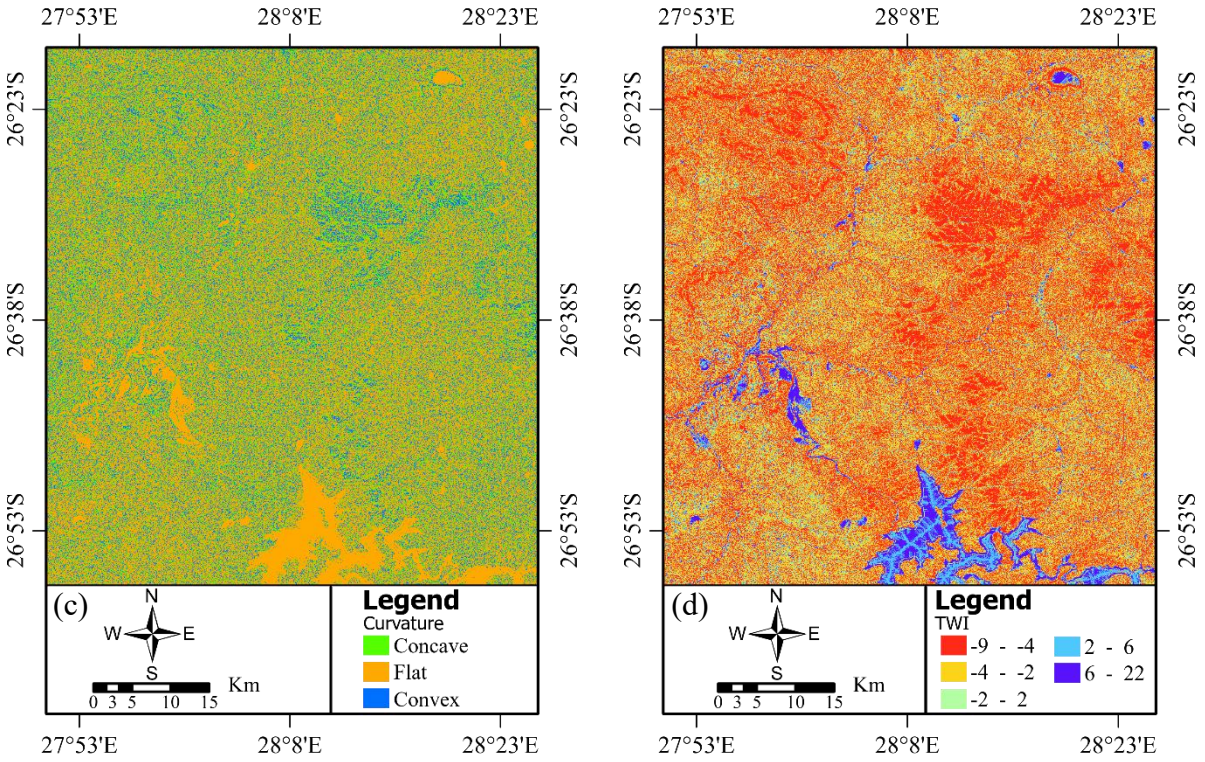


Figure 3.6: Ground subsidence related factors: (a) slope, (b) aspect, (c) curvature, (d) topographic wetness index.

3.3.5 Environmental, Geological, and Infrastructure Data

1. **Road Networks**

The road networks feature layer dataset used in our study is sourced from ESRI, providing detailed information on road networks, including main arteries and boundaries, as shown in Figure 3.7. This dataset is crucial for understanding transportation infrastructure, traffic patterns and distribution effects on induced ground subsidence. However, potential limitations, such as data inaccuracies and temporal variations in road network digitisation, must be acknowledged. Despite these considerations, the ESRI roads dataset significantly contributes to ground subsidence modelling and analysis, and transportation planning in the study area.

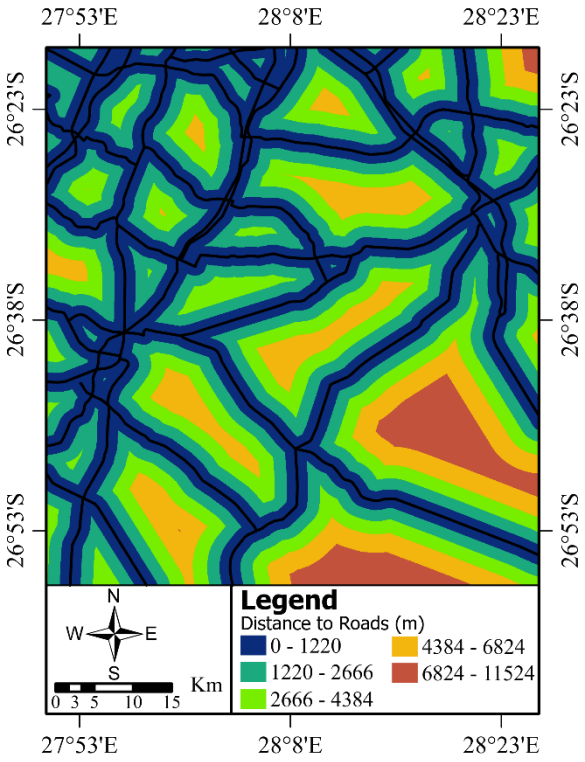


Figure 3.7: Distance to roads as ground subsidence related factor.

2. River Networks

The river networks feature layer dataset used in our study is sourced from ESRI, providing detailed information on main river channels, and watershed boundaries, as shown in Figure 3.8. This dataset is crucial for understanding surface water dynamics, and erosion risk assessment which contributes to ground deformation. However, potential limitations, such as data inaccuracies and temporal variations in river morphology, must be acknowledged. Despite these considerations, the ESRI river dataset significantly contributes to ground subsidence modelling and analysis in the study area.

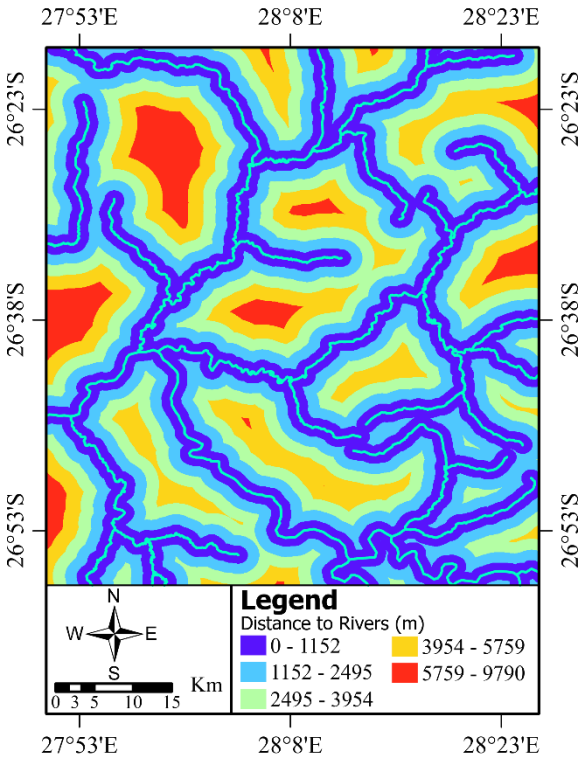


Figure 3.8: Distance to rivers as ground subsidence related factor.

3. Ground Water Storage

The Ground Water Storage (GWS) anomalies are sourced from the Global Land Data Assimilation System (GLDAS), which comprises three products: GLDAS 2.0 which is forced with only meteorological data, GLDAS 2.1 forced with observational data, and Catchment Land Surface Model (CLSM) or Variable Infiltration Capacity (VIC) model or Noah model, and GLDAS 2.2, employing data assimilation from the Gravity Recovery and Climate Experiment (GRACE) since February 2003.

The GLDAS-2.2 GRACE-DA product used in this study is simulated using Catchment-F2.5 in the Land Information System (LIS) , as shown in Figure 3.9. The simulation incorporates conditions from the GLDAS-2.0 Daily Catchment model, forced with meteorological analysis fields from the ECMWF Integrated Forecasting System. The terrestrial water anomaly observations from the GRACE satellite were assimilated (Li et al., 2019). Consequently, this study uses GWS derived from the CLSM model, a physics-based model focusing on natural effects rather than human-induced ones. The integration of CLSM and GRACE yields the GLDAS Catchment Land Surface Model L4 daily 0.25 x 0.25-degree GRACE-DA1 V2.2 (GLDAS_CLSM025_DA1_D 2.2) dataset, accessible via Earth Data (GES DISC). These GWS anomalies can be expressed mathematically as follows:

$$\Delta GWS = \Delta TWS - (\Delta CWE + \Delta SM + \Delta QS + \Delta SWE) \quad (3.1)$$

Where ΔGWS is the ground water storage anomaly in mm, ΔTWS is the terrestrial water storage anomaly in mm, ΔCWE is the canopy water storage, ΔSM is the soil moisture anomaly, ΔQS is the quick-flow surface runoff, ΔSWE is the snow water equivalent. From equation (3.1), it is evident that meteorological data only accounts for natural effects and neglects anthropogenic factors. This limitation represents a disadvantage in this method of determining Ground Water Storage.

While the CLSM model has limitations in neglecting surface water and human water use budget components, it provides valuable insights into ground water storage dynamics. This dataset, generated through sophisticated modelling techniques, offers high spatial resolution of 0.25 x 0.25 degrees. Leveraging data from the GRACE mission enables daily monitoring of ground water changes, facilitating detailed analysis and interpretation of hydrological processes at regional and global scales. Accessible through the Global Earth System Data Centre (GES DISC), this dataset serves as a critical resource for hydrological modelling, water resource management, and ground deformation studies.

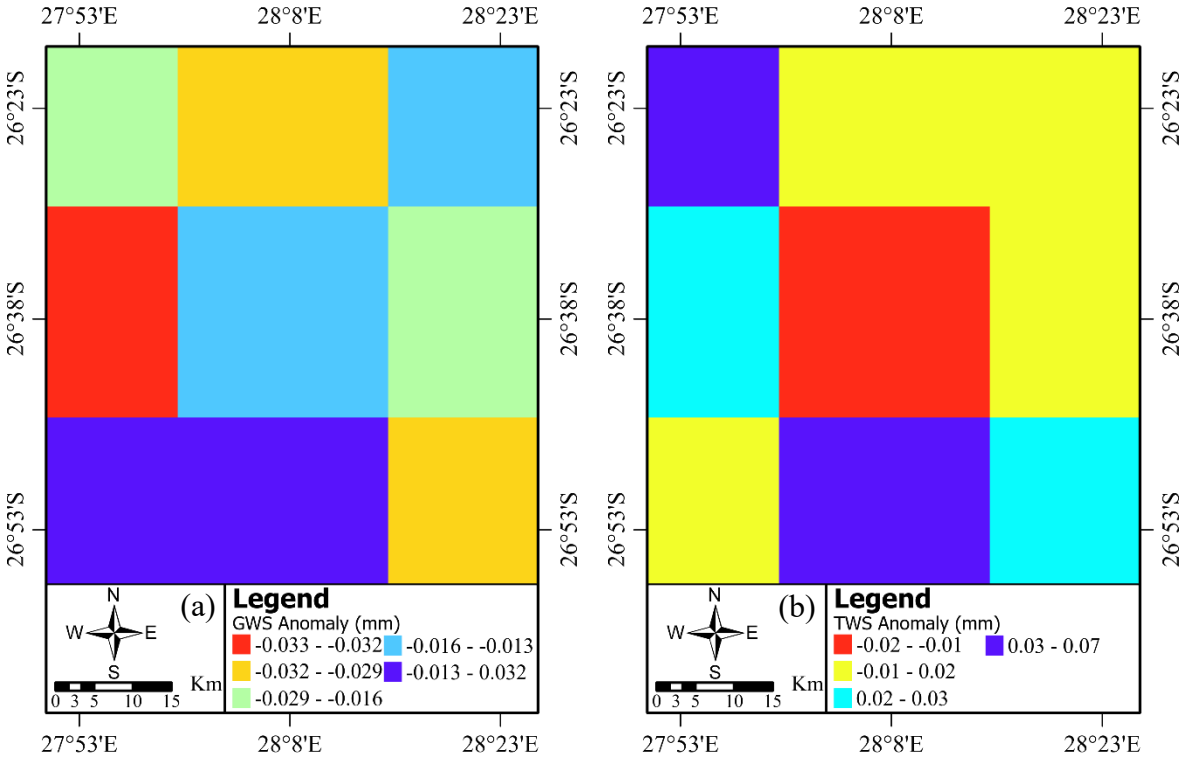


Figure 3.9: Ground subsidence related factors: (a) ground water storage anomalies, (b) terrestrial water storage anomalies.

4. Land Use and Land Cover

The South African National Land Cover 2018 (SANCL2018) is sourced from the Department of Forest, Fisheries and the Environment (DFFE) in South Africa, as shown in Figure 3.10. The SANLC2018 dataset is derived from multi-seasonal 20 metres resolution Sentinel-2 satellite imagery, it offers a comprehensive land-cover mapping

of South Africa and extends 100 metres into neighbouring countries and 10 kilometres into coastal waters. Generated through automated mapping models, it encompasses various land-cover components such as water bodies, mining areas, and forest plantations. Available in GeoTIFF raster format, it employs both Geographic Coordinates and Albers Equal Area projection. Representing a new era in land-cover datasets, SANLC2018 ensures standardised naming conventions and ongoing updates through the Computer Automated Land-Cover (CALC) system. With an overall map accuracy of 90.14%, it serves as a valuable resource for ground subsidence modelling and analysis in this study.

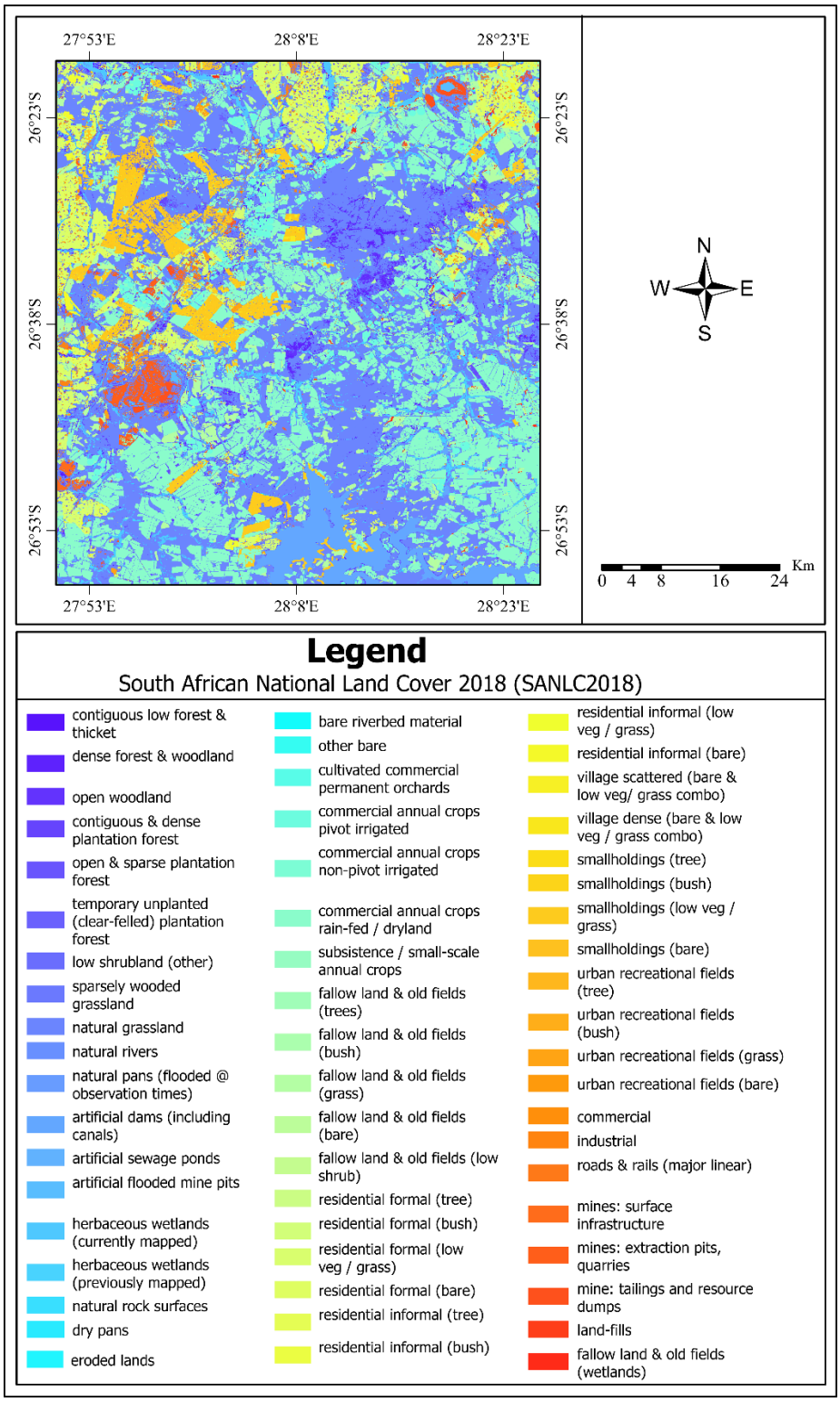


Figure 3.10: Land cover and land use as ground subsidence related factor.

5. Geological Formation

The geological (lithology) data for the study area were obtained from the Council of Geoscience (CGS) in South Africa. The CGS plays a pivotal role in mapping and understanding the geological composition of the country, including the Midvaal region in Gauteng. In our study, we accessed and retrieved lithology information from the CGS

to enhance our understanding of the subsurface characteristics within the study area. This dataset provides valuable insights into the geological formations, soil types, and rock compositions present in the Midvaal region. By analysing this data, we better comprehend the hydrogeological properties of the area, including groundwater flow patterns, aquifer characteristics, and geological hazards such as ground subsidence posing potential risks.

6. Climate: Precipitation and Temperature

The daily aggregated precipitation and temperature data, spanning from 1 January 2019 to 31 January 2021 is sourced from the European Centre for Medium-Range Weather Forecasts (ECMWF), ERA5, as shown in Figure 3.11. The ECMWF ERA5 data offers a comprehensive overview of weather patterns over this period, it provides detailed information on temperature and precipitation, crucial for various climatological effects studies on ground subsidence modelling and monitoring. This dataset offers valuable insights into long-term climate trends, extreme weather events, and regional variations in temperature and precipitation patterns. With its high spatial resolution of 0.25° x 0.25°, and daily temporal resolution, ERA5 data is a valuable resource for this research study evaluating weather and climate dynamics effects on ground subsidence.

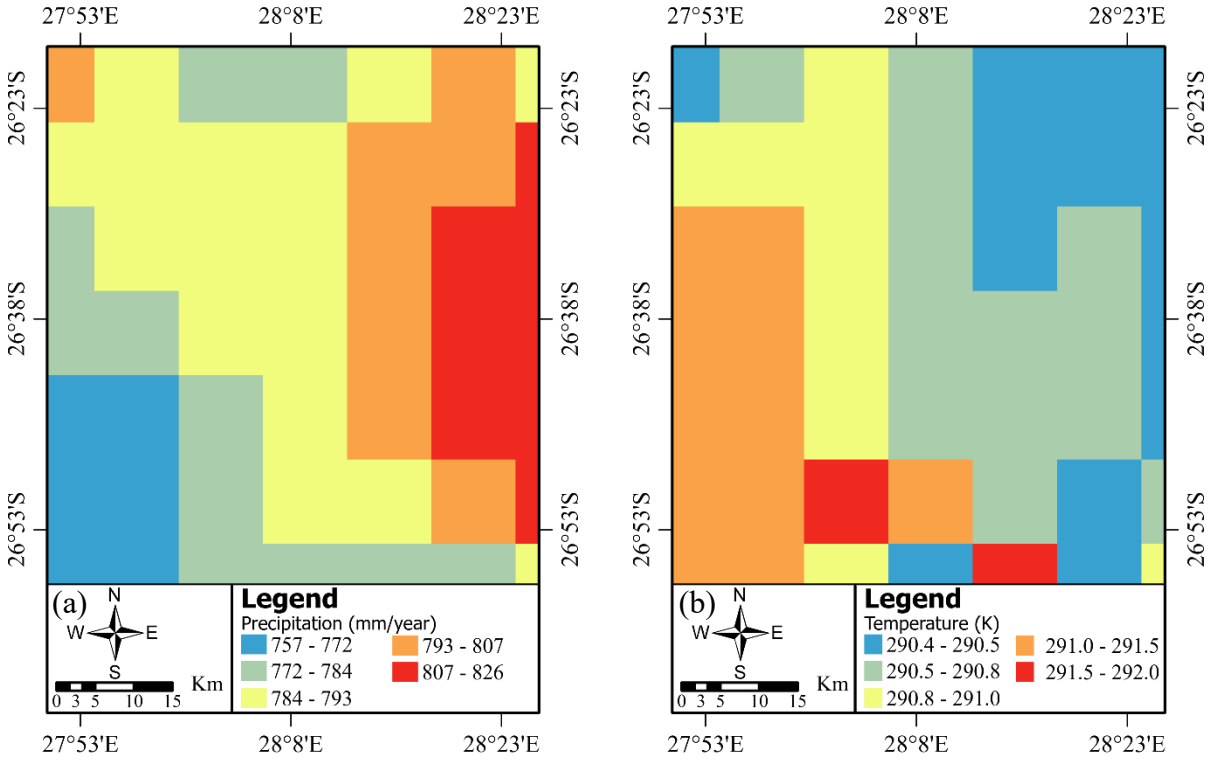


Figure 3.11: Ground subsidence related factors: (a) Annual Precipitation, (b) Temperature.

3.4 Development of SNAP to StaMPS semi-Automated Processor

The development of the *SNAP to StaMPS Semi-Automated Processor* addresses key limitations in existing InSAR processors, particularly SNAP. StoSAP automates the InSAR pre-processing

of Sentinel-1 SLC imagery for PS and SBAS InSAR techniques. By leveraging Python scripting and SNAP's Graph Processing Framework, StoSAP streamlines InSAR workflows, enhances processing efficiency, and improves functionality.

At the core of StoSAP's development is the integration of SNAP's GPF with Python scripting to enable full automation. This was achieved by defining all SNAP processing steps and graph XML codes within Python using object-oriented programming (OOP). The following are StoSAP's key features:

- **Class Definition for SNAP Graphs:** SNAP graph XML codes are defined within a Python class, making workflows accessible, modifiable, and extendable, allowing seamless updates when processing parameters change.
- **Iterative Processing:** StoSAP uses an iterative approach to handle multiple datasets, repeating SNAP processing steps for each acquisition to ensure efficient processing.
- **Error Handling and Log Management:** Automated error detection at each step enables early intervention, which is crucial for large-scale workflows.
- **Integration of PS and SBAS InSAR Modes:** StoSAP combines Persistent Scatterer and Small Baseline Subset processing into one tool, offering versatility in InSAR analysis.
- **Updated Dependencies and Python-3 Migration:** StoSAP is built on updated dependencies and migrated to Python-3, ensuring better compatibility and security.
- **Simplified Installation:** Dependencies are minimised to streamline the installation and setup process, reducing complexity for users.
- **Comprehensive Documentation:** Up-to-date documentation is provided to guide users through installation, configuration, and use of the processors.
- **Regular Update Cycle:** StoSAP follows a regular release cycle for updates and bug fixes to maintain its relevance and improve functionality.
- **Compatibility with Latest SNAP Versions:** The processor is designed to work seamlessly with the latest versions of SNAP, leveraging all new features and improvements.
- **Enhanced Multi-Swath Processing and Error Monitoring:** StoSAP includes support for processing data from multiple swaths and integrates robust error monitoring throughout the workflow.
- **Optimised Data Management:** Efficient data management strategies are implemented to reduce duplication and optimise storage usage.

- **Support for External DEMs and State Vector Orbit Types:** StoSAP enhances terrain correction and orbit accuracy by integrating external Digital Elevation Models and supporting various state vector orbit types.
- **User-Friendly Orbit Download Scripts:** Custom scripts are provided for direct downloading of orbit data from satellite repositories like ESA, simplifying data acquisition for users.

3.5 Effects of InSAR Pre-processing and Processing Software, and Time series Analysis Method

In this section, the methodology employed to assess the effects of InSAR pre-processing and processing software, as well as InSAR time series analysis methods is outlined. The approach commenced with a systematic evaluation of InSAR time series analysis techniques and open-source processing tools, focusing on identifying software capable of pre-processing Sentinel-1 C-band data in descending interferometric wide mode.

3.5.1 InSAR Pre-Processing and Time series Analysis

The pre-processing of InSAR data was undertaken using the ISCE, SNAP, and ASF HyP3 software suites. The PS InSAR workflow utilised ISCE-2.6 and SNAP-9.0, while the SBAS InSAR analysis employed ISCE-2.6. The pre-processing parameters included the use of 15 July 2020 as the super-reference acquisition, sub-swaths IW1 and IW2, bursts 1 to 10, and the "VV" polarisation mode.

Key processing steps began with the application of precise orbit vectors to ensure accuracy in the datasets. This was followed by TOPS Coregistration with NESD using the SRTMGL1 DEM and applied orbit vectors for precise alignment of sub-swaths and bursts. Interferogram formation was conducted with a temporal baseline of 24 days, after which topographic phase removal was applied using the SRTMGL1 DEM. The datasets were then multilooked, with range and azimuth looks set to 1 and 1, respectively. Goldstein Phase Filtering was applied with a default strength of 0.7 to reduce noise, and phase unwrapping was performed using the minimum cost flow algorithm. Finally, the coregistered stack and interferograms were exported in a format compatible with StaMPS for further analysis.

The StaMPS software was utilised for InSAR time series analysis, implementing both the PS and SBAS InSAR methodologies. The processing involved setting amplitude dispersion thresholds of 0.4 for PS and 0.6 for SBAS InSAR. The pre-processed stack of interferograms was loaded into StaMPS, and phase noise was estimated to identify coherent signals. Persistent Scatterer pixel candidates were then selected, and non-coherent pixels were removed. Subsequently, the wrapped phase corrections were made to estimate and remove SULA errors, followed by phase unwrapping with SNAPHU. Ground deformation corrections were further refined by estimating SCLA errors based on SRTMGL1 DEM data. Then, atmospheric filtering

was performed using a linear-based correction within the TRAIN toolbox. Finally, velocity estimates were derived from the corrected unwrapped phase measurements and exported to a CSV file.

Additional InSAR pre-processing and analysis were carried out using HyP3 and MintPy. Sentinel-1 C-band SLC descending mode data were processed in HyP3 using all sub-swaths (IW1, IW2, and IW3), bursts 1 to 10, and the "VV" polarisation mode. Pre-processing steps included the application of precise orbit vectors and radiometric calibration to ensure accurate reflectivity measurements. Then, TOPS Coregistration with NESD was undertaken using the SRTMGL1 DEM and orbit vectors, and interferogram formation employed a temporal baseline of 24 days. Subsequently, topographic phase removal was applied using the SRTMGL1 DEM, and Multilooking was configured with range and azimuth looks set to 10 and 2, respectively. Then, Goldstein Phase Filtering was applied with a filter strength of 0.7, followed by phase unwrapping using the minimum cost flow algorithm. Finally, the pre-processed data were then exported in a format compatible with MintPy.

In MintPy, the SBAS InSAR analysis was conducted. Spatial and temporal coherence thresholds were set to 0.7, and the processed stack was loaded into MintPy. Network modification and reference point selection were performed to optimise the analysis. Subsequently, unwrapping error corrections were applied, followed by network inversion using L2 norm least squares estimation. Then, tropospheric delay corrections were conducted using ECMWF ERA5 GAM data, and phase Deramping was used to remove long-wavelength trends. After that, topographic residual corrections were applied to address DEM inaccuracies, and noise evaluation and iterative re-execution ensured the reliability of results. The final step involved the estimation of average velocity from the corrected phase measurements, producing ground deformation data in H5 file format for further interpretation.

3.5.2 Point Velocity Extraction Methods

Ground deformation velocity extraction from the CSV file generated by StaMPS was carried out using the *stamps_nearby* python module implemented in StoSAP as shown in Appendix N. This involved interpolating the HEID and VERG cGNSS station velocity from the nearest 25 PS/DS points within a 2 km radius of the station. For h5 file outputs from MintPy, the H5 file was converted into tiff raster format using *h5_to_raster* module within StoSAP as shown in Appendix Q, subsequently the actual values at the cGNSS locations were extracted using the 'Extract Multi Values to Points' tool in ArcPy.

3.5.3 Conversion of StaMPS and MintPy Output Data to Raster Format

The StaMPS output data in CSV format and the MintPy output data in H5 format were converted to TIFF raster format using the *stamps_to_raster* and *h5_to_raster* modules,

respectively, as implemented within the StoSAP processor (see Appendix P and Appendix Q). These conversions facilitated visualisation and data manipulation.

The *stamps_to_raster* module converted geographic coordinates to corresponding pixel indices. When multiple points shared the same pixel indices, their velocities were averaged, and the resulting average value was used to generate the raster; while missing points were assigned null values. A 3 arc-second resolution was chosen to grid PS/DS points and generate velocity and standard error rasters.

The *h5_to_raster* module retrieved grid data from the H5 file and extracted the coordinate system from a reference grid raster (EPSG:4326). It then wrote the output data in TIFF format, preserving the reference grid's coordinate system and H5 file spatial resolution.

3.5.4 Accuracy Assessment

The evaluation of ground subsidence monitoring accuracy employed cGNSS stations HEID and VERG as validation datasets. The methodology involved comparing cGNSS derived velocities with results obtained from various software and processing techniques, emphasising Persistent Scatterer and Small Baseline Subset methods. Accuracy was assessed using multiple metrics, including the Root Mean Square Error, which was calculated to quantify the precision of continuous variables as shown in equation (3.2) below:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i^j - x_i^k\right)^2} \tag{3.2}$$

where x_i^j represents the cGNSS implied velocity and x_i^k represents the InSAR implied velocity.

The Mean Absolute Error was computed to measure the mean of absolute differences as shown in equation (3.3) below:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}\left| x_i^j - x_i^k \right| \tag{3.3}$$

Range Error was assessed to determine the interval length between maximum and minimum values as shown in equation (3.4) below.

$$R = max(x_i^j - x_i^k) \; - \; min(x_i^j - x_i^k) \tag{3.4}$$

3.5.5 Correlation and Performance Evaluation

Additionally, to perform our correlation analysis, we resampled our HyP3 - MintPy SBAS-InSAR velocities into a 3 arc-second grid corresponding to the StaMPS reference grid rasters. Once the StaMPS and MintPy rasters are coregistered, the Pearson correlation was used to analyse the relationship among common pixels in the velocity maps generated by different

techniques. Scatter plots visualised variations and relationships, while frequency distribution plots illustrated velocity class patterns. This detailed approach ensured a robust evaluation of the performance and accuracy of each processing tool, providing insights into their effectiveness for ground subsidence monitoring and susceptibility analysis.

3.6 Effects of Digital Elevation Models

In this section, the methodology employed to assess the impact of Digital Elevation Model accuracy on machine learning derived ground subsidence velocities is outlined. The approach begins with an evaluation of DEM accuracy, followed by an assessment of velocity accuracy using levelling data as detailed below.

Before evaluating DEM accuracy, data voids were addressed. Using ArcPy in ArcGIS Pro, we performed conditional map algebra expressions to classify data into "No Data" and "Valued Data" categories. Voids were then filled using ArcPy's void fill function to interpolate and complete missing pixel values.

3.6.1 Digital Elevation Model Accuracy Assessment

The evaluation of the impact of Digital Elevation Models on ground subsidence susceptibility mapping involved undertaking a thorough assessment. Initially, DEM elevations were compared with ground levelling data from 247 levelling points using equation (2.37) within the study area to determine vertical accuracy. This comparison involved calculating mean errors, standard deviations, and root mean square errors for each DEM, including ASTER, AW3D30, COP30, and SRTMGL1.

Following this, the DEMs were coregistered to align their grid centroids, and elevation errors were recalculated to assess the effect of coregistration on accuracy. This step aimed to enhance the alignment of DEMs, improve the precision of elevation data, and reduce processing complexities and time which may results from incompatible grids.

The elevation errors were further visualised through spatial interpolation using the Inverse Distance Weighting (IDW) algorithm in ArcPy, highlighting spatial variations in elevation errors across the study area. Additionally, frequency and bar graph analyses were conducted to evaluate the distribution of elevation errors within specific ranges. The relationship between DEM elevation errors and DEM-implied elevations was also examined to understand how elevation errors vary with different DEMs.

3.6.2 Predicting Levelling and DEM-Implied Velocity

The velocity prediction was performed using the XGBoost regression algorithm, with geographic latitude, longitude, and ground levelling or DEM height as input features, yielding the predicted velocities $V_{levelling}$ and V_{DEM}. It was assumed that inaccuracies in the predicted velocities arose from error estimation within the prediction model and propagated consistently

across all predictions. Consequently, when the difference between $V_{levelling}$ and V_{DEM} are computed, these errors will effectively cancel out. Thus, the remaining differences are attributed to inherent vertical errors in the DEM. Additionally, the choice of prediction algorithm is assumed to be insignificant as the differences are attributed to DEM vertical errors.

3.6.3 Velocity Accuracy Assessment Using Levelling Data

The investigation into how DEM errors affect ground subsidence velocity prediction involved a statistical analysis of discrepancies between velocities derived from DEMs (V_{DEM}) and those obtained from ground levelling ($V_{levelling}$) data was performed as shown by equation (3.5) below.

$$V_{error} = V_{levelling} - V_{DEM} \tag{3.5}$$

This analysis involved calculating RMSE, mean absolute errors ($\bar{x}$), and range error (R) for velocities predicted using each DEM as shown in equation (3.6), (3.7), and (3.8) respectively below.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(V_{error})^2} \tag{3.6}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}|V_{error}| \tag{3.7}$$

$$R = max(V_{error}) - min(V_{error}) \tag{3.8}$$

The spatial distribution of velocity errors was visualised using IDW interpolation in ArcPy to identify areas with significant discrepancies. The analysis was extended to compute frequency distribution graphs to assess the prevalence of velocity errors within different ranges. Additionally, relationships between velocity errors and other variables, including velocity itself, elevation, and elevation errors, were examined. This aimed to determine how these factors influence velocity error prediction and to identify the need for more advanced models.

3.7 Spatial and Temporal Patterns of Ground Subsidence

The assessment of the spatial and temporal patterns of ground subsidence in Midvaal involved the selection of two continuous GNSS stations and ten randomly chosen points strategically distributed across all five classes of ground subsidence susceptibility. This approach facilitated an in-depth understanding of patterns within each class, and the continuous ground deformation data further strengthens the robustness of the spatio-temporal pattern analysis. The analysis encompassed spatial evaluation of ground subsidence, correlation with hydrological factors, vegetation and water index correlations, as well as statistical analysis as detailed below.

3.7.1 Spatial Analysis of Ground Subsidence

The spatial analysis of ground subsidence in Midvaal was conducted using HyP3 SBAS-InSAR velocity measurements from Sentinel-1 C-band data. To compare ground deformation patterns, both ascending and descending orbit measurements were used. The study involved calculating the maximum rates of ground subsidence and uplift and mapping their spatial distribution. For relative accuracy, the spatial patterns of ground deformation from ascending and descending orbits were compared using statistical measures such as the coefficient of determination and absolute difference statistics. This comparison aimed to assess the consistency and reliability of ground deformation patterns captured by the two orbits.

3.7.2 Statistical Analysis

Statistical analyses were employed to quantify the relationships and discrepancies observed in the data. The mean, standard deviation, and percentiles of the absolute differences between ascending and descending orbit measurements were calculated. This statistical analysis provided insights into the consistency and reliability of the measurements and highlighted any potential outliers or significant differences.

3.7.3 Temporal Dynamics and Sample Point Analysis

The exploration of temporal dynamics involved selecting ten random sample points representing a range of susceptibility levels for in-depth analysis. These points were chosen to provide a comprehensive understanding of ground subsidence patterns over time. The analysis focused on temporal variations in ground subsidence, using velocity data from the ascending orbit due to minimal differences between ascending and descending velocities. The temporal data were examined for trends and patterns to understand how ground deformation evolves over time.

3.7.4 Correlation with Hydrological Factors

The relationship between ground subsidence and hydrological factors such as groundwater storage anomaly and precipitation was assessed. Groundwater storage anomaly data from the GLDAS CLMS model were compared with ground deformation data. The analysis aimed to determine whether changes in groundwater storage correlate with ground subsidence or uplift. Similarly, daily precipitation data were analysed to evaluate its influence on ground deformation. The goal was to identify any significant correlations between hydrological variables and ground subsidence.

3.7.5 Vegetation and Water Index Correlations

The analysis explored the influence of vegetation and water content on ground deformation by assessing correlations between ground subsidence and vegetation indices (NDVI) and water

indices (MNDWI). The NDVI and MNDWI values were compared with ground subsidence data to assess how changes in vegetation cover and water content impact ground deformation. The analysis focused on both short-term and long-term variations in NDVI and MNDWI to evaluate their relationship with ground subsidence patterns.

3.8 Ground subsidence Susceptibility Mapping, Inducing Factors, and Spatial Patterns in Midvaal, Gauteng

In this section, the methodology for evaluating ground subsidence inducing factors, susceptibility mapping, and spatio-temporal patterns in Midvaal, Gauteng is outlined. The approach began with the preparation and georeferencing of data, followed by rigorous factor prioritisation and multicollinearity analysis. Machine learning models were then employed to generate susceptibility maps, with performance evaluated through accuracy metrics and model explainability techniques. The analysis concluded with Frequency Ratio analysis to explore the spatial relationships between ground subsidence and various influencing factors, providing insights into their correlations and impacts.

3.8.1 Data Preparation and Georeferencing

Data preparation began with the careful selection and creation of a reference grid, which underpins the subsequent analysis. This grid is positioned with its initial coordinates at latitude 26° 18’ 38" S and longitude 27° 50’ 40" E, covering the entire study area with a high spatial resolution of 1 arc-seconds. It consists of dimensions of 2112 by 2295 rows and columns respectively, with a spatial resolution of 0.00027777777 degrees in both latitude and longitude. This setup ensured comprehensive coverage while maintaining computational efficiency.

Following grid establishment, raster images representing factors related to ground subsidence were generated and co-registered to the reference grid. Subsequently, ground subsidence data from HyP3 SBAS-InSAR ascending orbit velocity measurements was georeferenced (adjusted using station HEID). The un-georeferenced dataset initially identified 1,190,550 pixels experiencing ground subsidence. After georeferencing with station HEID, this number was refined to 303,949 pixels, highlighting the critical role of precise georeferencing in accurately identifying subsiding and uplifting areas.

The georeferenced HyP3 SBAS-InSAR ascending orbit velocity was divided into 70% training data and 30% test data, in line with methodologies outlined by Hakim et al. (2023a) and Mohammady (2019). The co-registered subsidence-related factors and training data formed the basis for subsequent analytical techniques.

3.8.2 Factor Prioritisation and Multicollinearity Analysis

Factor prioritisation was performed using the Information Gain analysis, which assessed the significance of each ground subsidence related factor. Subsequently, the multicollinearity

analysis employed Pearson Correlation, Variance Inflation Factor, and Tolerance. Factors with a VIF value greater than 5 and a Tolerance value below 0.1 were considered redundant and were removed to prevent bias. Pearson Correlation was used to further assess factor interdependence, identifying high correlations (0.7 threshold) that indicated potential redundancy. Following the removal of correlated factors, the prioritisation and multicollinearity analyses were recalculated with the remaining twelve factors.

3.8.3 Machine Learning Model Implementation

Machine learning models including Convolutional Neural Networks, LightGBM, Random Forest, and XGBoost were used to generate ground subsidence susceptibility maps. Data preparation involved integrating the georeferenced HyP3 SBAS-InSAR measurements with twelve effective factors. Each model was evaluated using mean absolute error, root mean square error, and coefficient of determination. The default models were first evaluated, followed by hyperparameter optimization using GridSearchCV.

For the CNN, various hyperparameters were examined, including epochs, batch sizes, kernel sizes, and filter counts. The analysis determined that the optimal settings were 20 epochs, a batch size of 64, a kernel size of 128, and 5 filters. These model parameters and values are well justified as they provided a balance between training efficiency and model accuracy, preventing overfitting while ensuring sufficient learning capacity.

In contrast, for LightGBM, the focus was on the number of estimators, learning rates, and maximum depth. The best configuration for LightGBM included 1000 estimators, a maximum depth of 9, and a learning rate of 0.2. This combination of parameters and values was selected because it offered strong predictive performance without excessive training time, with the learning rate of 0.2 ensuring faster convergence during optimization.

Regarding Random Forest, the key hyperparameters explored were the number of estimators, minimum samples split, and maximum depth. The most effective configuration for Random Forest comprised 1000 estimators, a maximum depth of “None”, and minimum samples split of 4. Setting maximum depth to “None” allowed each tree to grow fully, capturing complex patterns in the data, while a minimum samples split of 4 prevented overfitting by requiring a slightly larger sample to create splits.

Finally, for XGBoost, the optimization involved varying the number of estimators, learning rates, and maximum depths, with the optimal settings being 1000 estimators, a maximum depth of 7, and a learning rate of 0.1. This configuration was chosen because it achieved a good trade-off between model complexity and generalisation, with a relatively shallow maximum depth reducing the risk of overfitting, and a conservative learning rate of 0.1 allowing for gradual model updates.

3.8.4 Machine Learning Model Accuracy Evaluation

The model performance was assessed using a test dataset comprising 30% of the data. The evaluation was based on several accuracy metrics, including R^2 score (coefficient of determination), Root Mean Square Error, and Mean Absolute Error. Additionally, the accuracy of each model was further validated through Receiver Operating Characteristic and Area Under the Curve analyses. These methods provided a comprehensive assessment of the model's effectiveness in predicting ground subsidence susceptibility.

3.8.5 Machine Learning Model Explainability

The interpretability of the machine learning models was enhanced through the use of various explainability techniques. Feature importance analysis was conducted across all four models using mean SHAP values to identify the key factors influencing ground subsidence predictions.

SHapley Additive exPlanations values were used to provide insights into the contributions of individual features to the model's outputs. This analysis involved generating SHAP bar plots to illustrate the overall importance of each feature, and SHAP beeswarm plots to visualise the distribution and impact of feature values on model predictions. These visualisations facilitated a deeper understanding of how different features, such as geographic position, lithology, elevation, soil moisture, land use, etc, influenced the model's predictions and their relative importance.

3.8.6 Frequency Ratio Analysis

Frequency Ratio analysis was employed to investigate the spatial relationship (correlation) between ground subsidence and various influencing factors. A Frequency Ratio value greater than 1 indicates a positive correlation, suggesting that the factor is associated with higher instances of ground subsidence. Conversely, a value less than 1 implies a negative or no significant correlation.

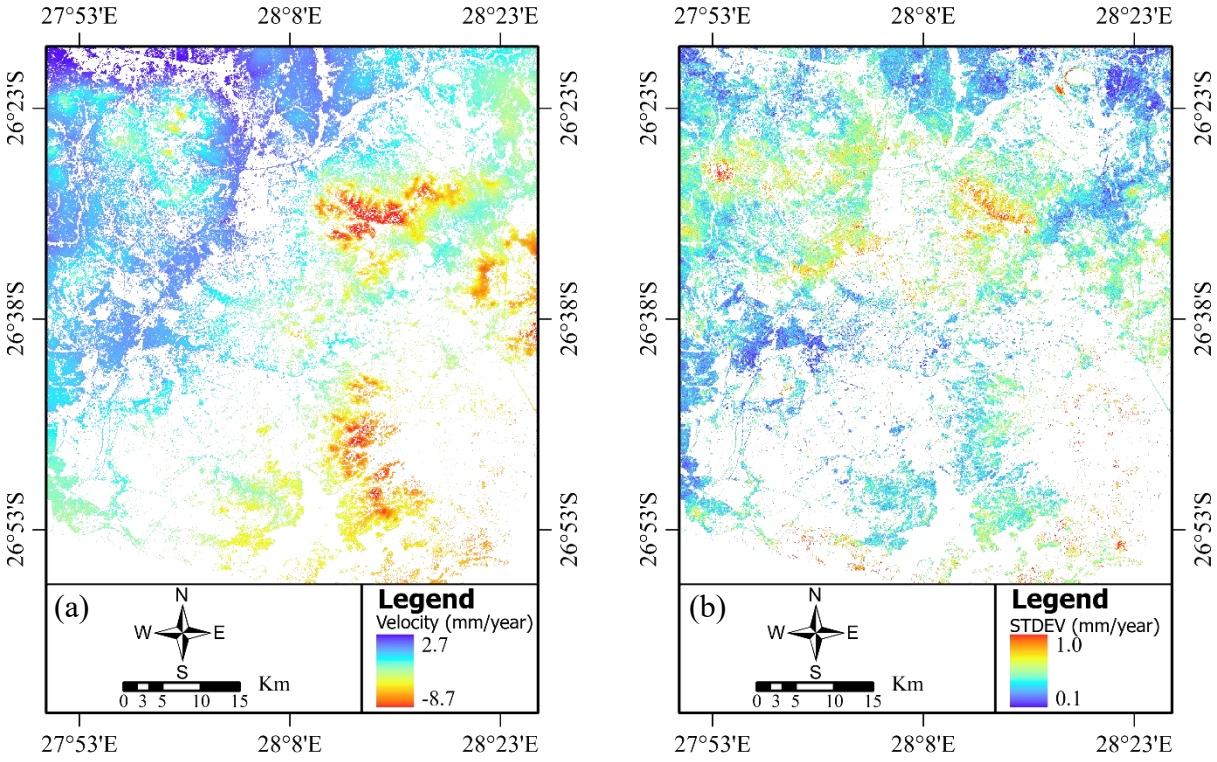
This methodology integrated comprehensive data preparation, statistical analyses, machine learning model implementation, and evaluation, with a specific focus on model explainability, ensuring a robust and interpretable assessment of ground subsidence susceptibility in the Midvaal region.

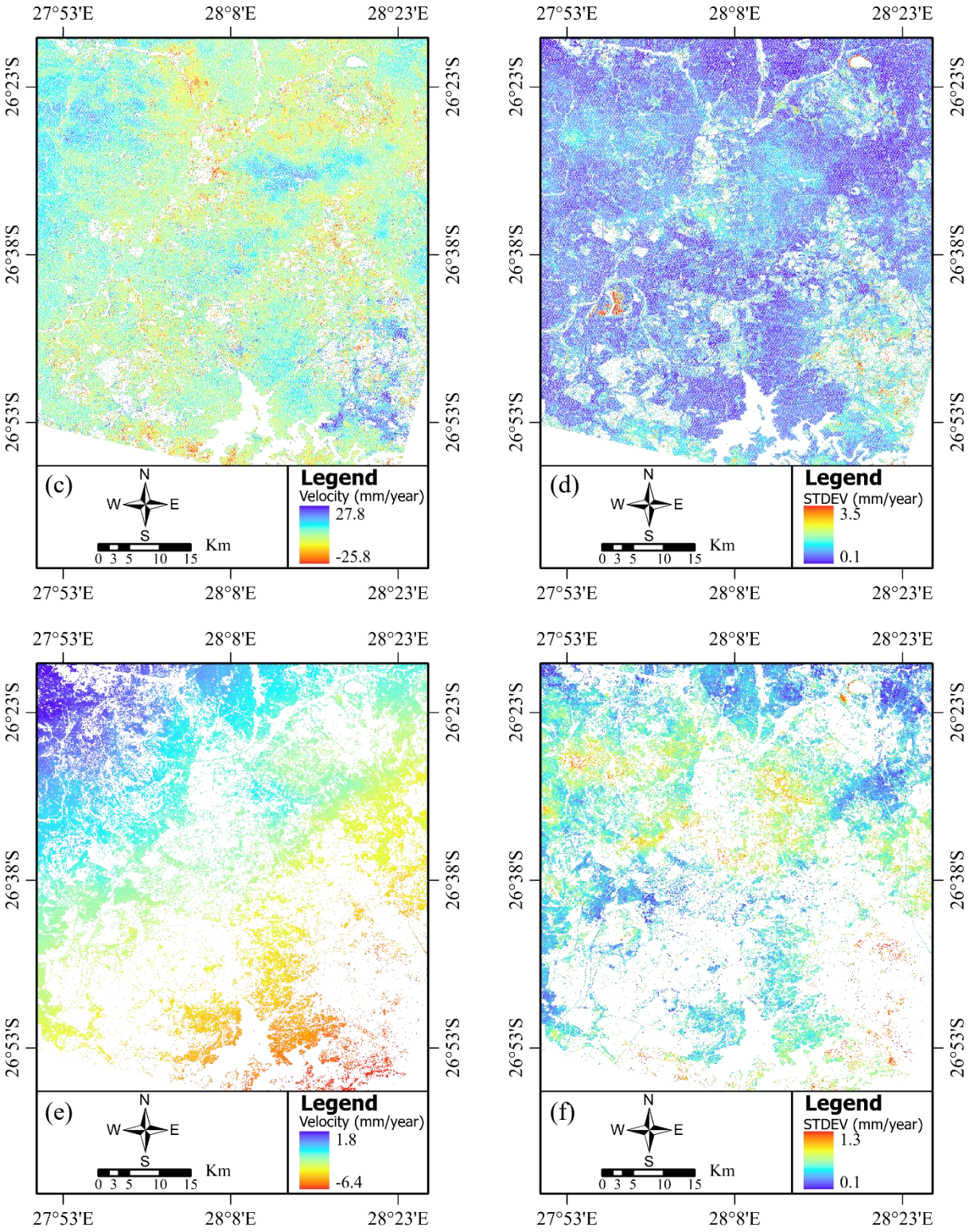
4 Results and Discussion.

4.1 Effects of InSAR Pre-processing and Processing Software, and Time series Analysis Method

An investigation was conducted to evaluate available InSAR time series analysis techniques and open-source processing tools, with a clear criterion of selecting open-source options. Based on this investigation, ISCE, SNAP, and ASF HyP3 software were chosen for preprocessing Sentinel-1 C-band data in descending interferometric wide mode. For InSAR time series analysis of multi-temporal Sentinel-1 data, StaMPS software was selected, employing both PS-InSAR and SBAS-InSAR techniques, MintPy was also selected to perform SBAS-InSAR. For PS-InSAR, combinations of ISCE and StaMPS, as well as SNAP and StaMPS were employed, while for SBAS-InSAR, combinations of ISCE and StaMPS, and HyP3 and MintPy were used.

The development of tools and software to complement these open-source preprocessing and processing tools led to the implementation of the StoSAP processor. This processor facilitated the implementation of SNAP PS-InSAR preprocessing and post-StaMPS processing, analysis, and visualisation. Through processing leveraging ISCE, SNAP, HyP3, StaMPS, and MintPy, velocity maps were obtained as shown in Figure 4.1. These velocity maps illustrate the spatial patterns of ground deformations, highlighting areas of ground subsidence (red / negative velocity) and uplift (blue / positive velocity) within the study region.




27°53'E
28°8'E
28°23'E
26°23'S
26°38'S
26°53'S
(c)
N
W
E
S
0 3 5 10 15
Km
Legend
Velocity (mm/year)
27.8
-25.8
(d)
Legend
STDEV (mm/year)
3.5
0.1
(e)
Velocity (mm/year)
1.8
-6.4
(f)
STDEV (mm/year)
1.3
0.1

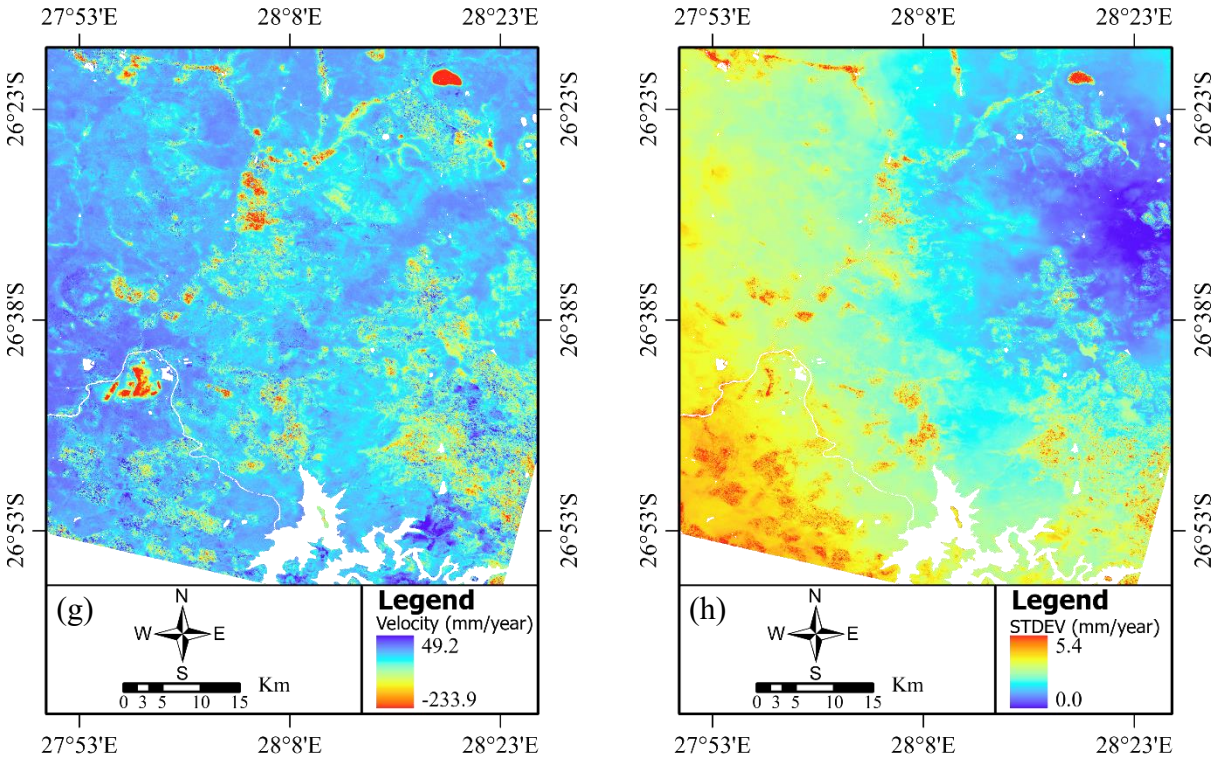


Figure 4.1: Ground deformation velocity and their corresponding velocity standard error maps obtained from, (a , b) ISCE-StaMPS PS-InSAR, (c , d) ISCE-StaMPS SBAS-InSAR, (e , f) SNAP-StaMPS PS-InSAR, (g , h) HyP3-MintPy SBAS-InSAR.

The velocity measurements obtained from ISCE PS-InSAR indicates significant ground subsidence in the north-eastern and south-eastern parts of the study area, while major uplift is observed in the northern to north-western part. The maximum rates of ground subsidence and uplift are -8.7 mm/year and 2.7 mm/year, respectively, with minimum and maximum standard errors of 0.1 mm/year and 1.0 mm/year. The small range of standard errors underscores the robustness and precision of ISCE PS-InSAR measurement results.

The velocity measurements obtained from ISCE SBAS-InSAR indicates significant ground subsidence in the northern, south-western, and eastern parts of the study area, while major uplift is observed in the south-eastern part. The areas not included as PS points in ISCE PS-InSAR are notably selected as SBAS/DS points and are subject to major ground subsidence. The maximum rates of ground subsidence and uplift are -25.8 mm/year and 27.8 mm/year, respectively, with minimum and maximum standard errors of 0.1 mm/year and 3.5 mm/year. The wider range of standard errors suggests less precision compared to ISCE PS-InSAR.

The velocity measurements obtained from SNAP PS-InSAR indicates major ground subsidence in the south-eastern parts of the study area, with medium ground subsidence in the eastern and south-western parts, while major uplift is observed in the north-western part, with medium uplifts in the western and north-eastern parts. The PS point distribution of SNAP PS-InSAR is highly similar to ISCE PS-InSAR but varies in spatial pattern based on velocity values. The maximum rates of ground subsidence and uplift are -6.4 mm/year and 1.8 mm/year,

respectively, with minimum and maximum standard errors of 0.1 mm/year and 1.3 mm/year. The small range of standard errors underscores the robustness and precision of SNAP PS-InSAR measurement results, though it is less precise compared to ISCE PS-InSAR and more precise than ISCE SBAS-InSAR.

Finally, the velocity measurements obtained from HyP3 SBAS-InSAR indicates significant ground subsidence in the south-eastern, north-eastern, and south-western parts of the study area, while major uplift is observed in the western part and some at the edge of the south-eastern part. The maximum rates of ground subsidence and uplift are -233.9 mm/year and 49.2 mm/year, respectively, with minimum and maximum standard errors of 0.0 mm/year and 5.4 mm/year. The large range of standard errors indicates reduced precision of HyP3 SBAS-InSAR measurement results compared to all other tools. However, given the wide range of velocities, the ratio between maximum standard error and maximum ground subsidence/uplift is better than that of other techniques, implying that HyP3 SBAS-InSAR has better relative uncertainty compared to other techniques. The standard error map indicates that velocities in the western part of the study area are subject to higher uncertainty than velocities in the eastern part, which has implications for validation stations in the western part.

4.1.1 InSAR and InSAR time series analysis Method Correlation

To further understand the relationship and correlation among the techniques, we leveraged the Pearson correlation among common pixels in all techniques as shown in Figure 4.2. The Pearson correlation results imply that there is a high correlation between ISCE and SNAP PS-InSAR of 0.7, while other combinations exhibited low correlation, underscoring the uniqueness of each technique and the significance of tool selection choices. The second-best correlation was between ISCE PS-InSAR and HyP3 SBAS-InSAR at 0.27.

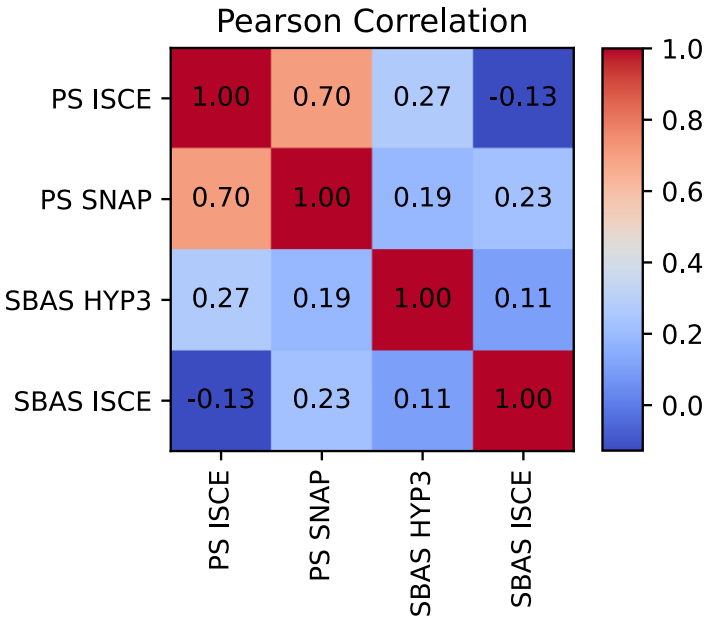


Figure 4.2: Techniques and Tools Pearson correlation.

Moreover, the scatter plots in Figure 4.3 illustrate variations among the techniques or tools used, as indicated by best-fit lines that are nearly horizontal and low Pearson correlation values, except for the relationship between ISCE and SNAP PS-InSAR.

- **ISCE and SNAP PS-InSAR**: This combination displays a strong positive correlation of 0.70 with a positive slope of 0.48. The data points form both linear and non-linear patterns, indicating a mixture of relationships.
- **ISCE PS-InSAR and HyP3 SBAS-InSAR**: This combination shows a weak positive correlation of 0.27 with a positive slope of 0.40. The points generally follow the best-fit line with some outliers, suggesting both linear and non-linear relationships.
- **ISCE PS-InSAR and SBAS-InSAR**: This combination shows a weak negative correlation of -0.13 with a negative slope of -0.15, underscoring contrasting PS and DS point characteristics. Additionally, the points form both linear and non-linear relationships.
- **SNAP PS-InSAR and HyP3 SBAS-InSAR**: This combination shows a weak positive correlation of 0.19 with a positive slope of 0.41. The points follow the best-fit line with some outliers, indicating both linear and non-linear relationships.
- **SNAP PS-InSAR and ISCE SBAS-InSAR**: This combination shows a weak positive correlation of 0.23 with a positive slope of 0.38. The points follow the best-fit line with some outliers, indicating both linear and non-linear relationships.

- **HyP3 and ISCE SBAS-InSAR**: This combination shows a weak positive correlation of 0.11 with a positive slope of 0.08. The points are symmetric about the best-fit line, but they do not form linear patterns. Instead, they form an oval shape, indicating strong non-linear patterns and differences in the tools employed.

Overall, all the techniques demonstrates uniform point clusters along the best-fit line. However, the clustering of points in HyP3 and ISCE SBAS-InSAR is concentrated between HyP3 SBAS-InSAR velocities of -20 mm/year to 20 mm/year. This underscores the differences in the capabilities of these techniques in capturing velocities outside this range. This observation is supported by the velocity maps, which show that HyP3 SBAS-InSAR can capture velocities beyond +/- 30 mm/year, whereas ISCE SBAS-InSAR does not reveal or capture these velocities measurements.

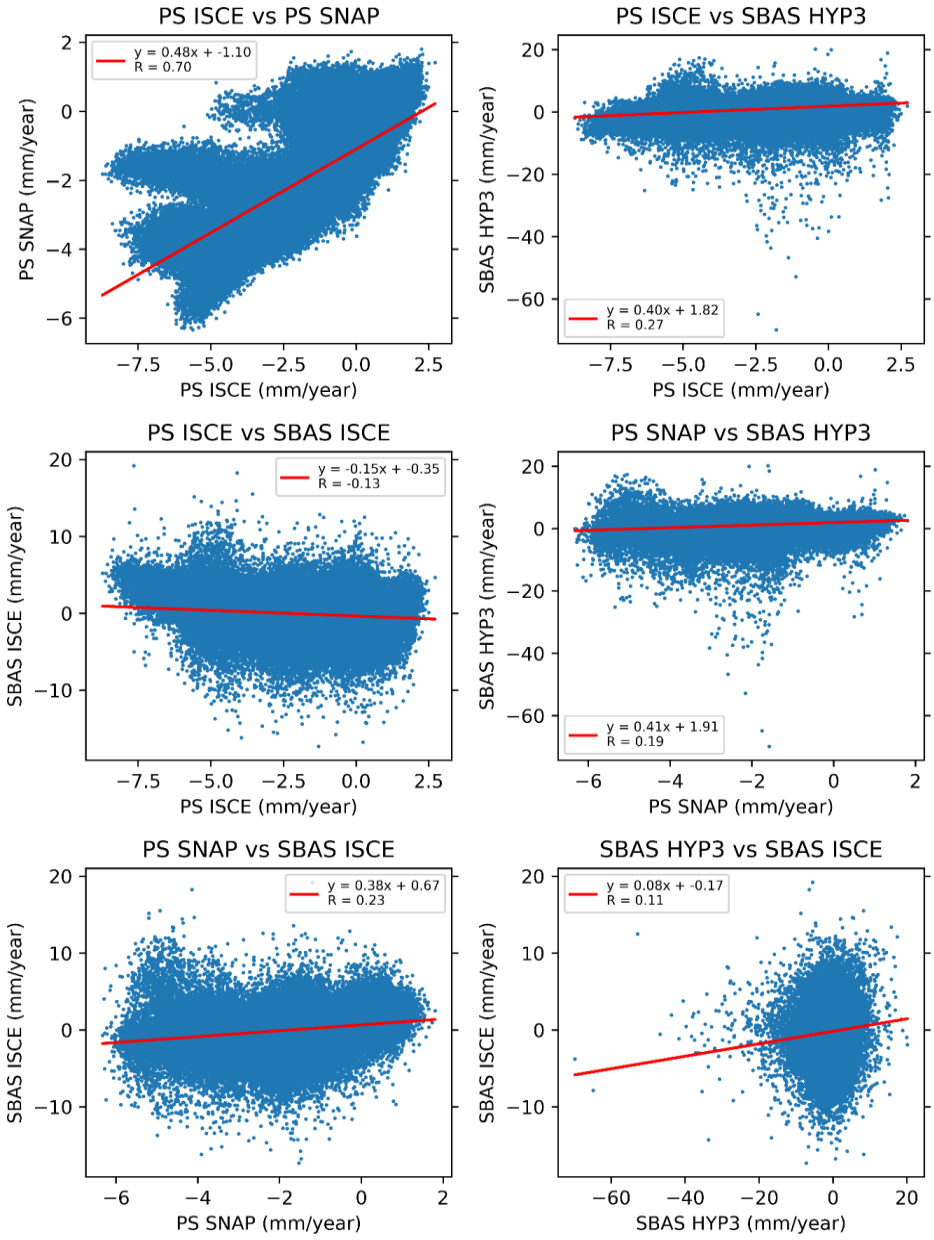


Figure 4.3: InSAR techniques and tools correlation scatter plots.

Finally, leveraging the common pixel points among all techniques, we plotted the frequency distribution of the velocities to understand their class patterns. ISCE PS-InSAR, HyP3 SBAS-InSAR, and ISCE SBAS-InSAR show higher velocity frequencies at 0 mm/year, except for SNAP PS-InSAR, which displays the highest frequency at around -2 mm/year. This indicates that velocities between ISCE PS-InSAR, HyP3 SBAS-InSAR, and ISCE SBAS-InSAR are similar and different from SNAP PS-InSAR, which may indicate errors in the SNAP PS-InSAR technique. All four techniques show positively skewed velocities, though ISCE and HyP3 SBAS-InSAR showed much best normal distribution patterns.

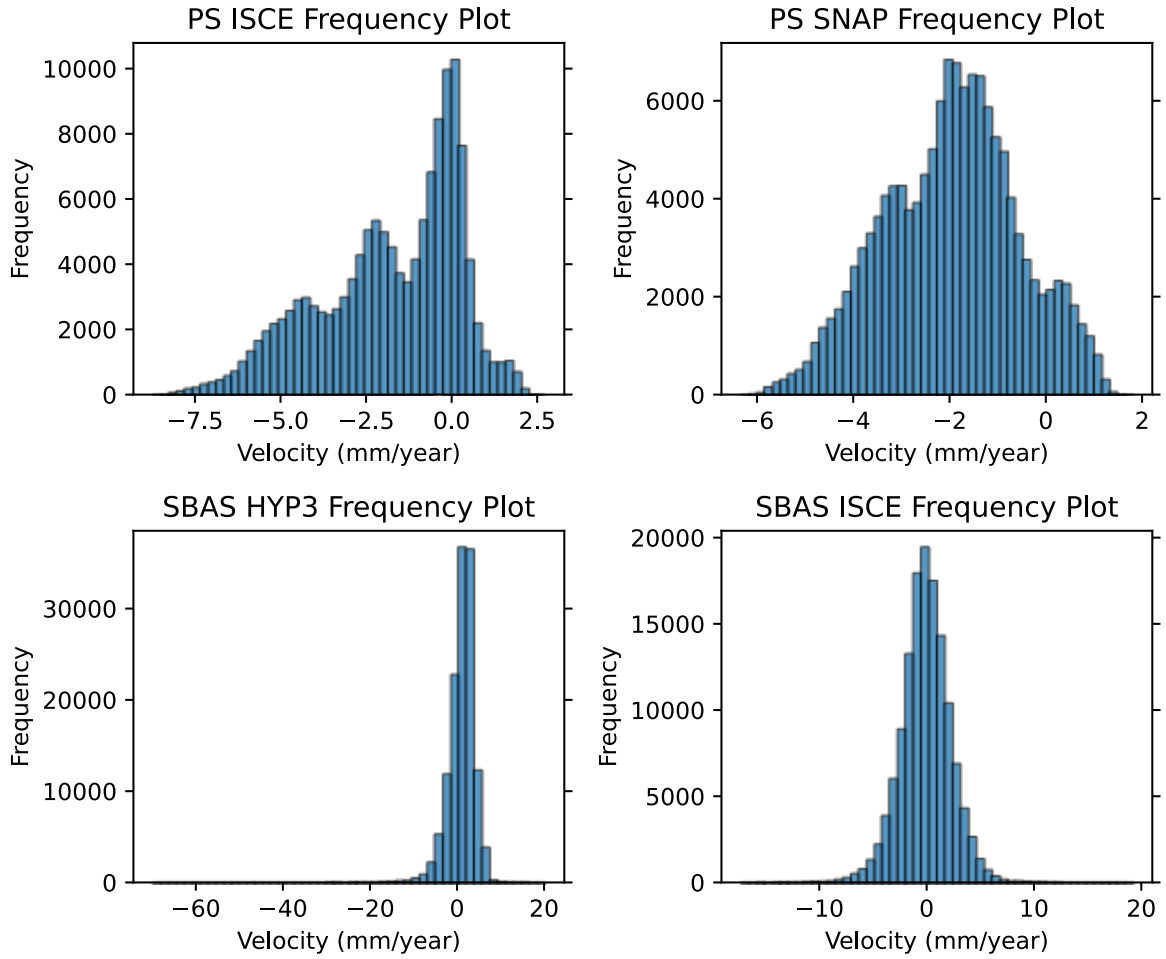


Figure 4.4: InSAR techniques and tools frequency distribution.

The results from the four processing techniques and tools reveal that each tool produces unique results, making the choice of tool selection a significant factor. ISCE PS-InSAR exhibits lower standard errors, making it the most precise technique among the four, followed by SNAP PS-InSAR, ISCE SBAS-InSAR, and lastly HyP3 SBAS-InSAR. While HyP3 SBAS-InSAR has larger standard errors, it captures a wide range of ground deformation velocities compared to the other tools, making it more robust for investigating localised ground subsidence spatio-temporal patterns. Nonetheless, the uniqueness and precision of each technique necessitated an

accuracy assessment to decide on the overall performance of each technique/tool, as discussed below.

4.1.2 InSAR and InSAR time series analysis Method Accuracy Evaluation

To assess the effects of InSAR pre-processing and processing software selection on the accuracy of ground subsidence monitoring, we leveraged continuous GNSS stations HEID and VERG as validation datasets. The study involved comparing the results from different software and processing techniques, specifically ASF HyP3, ISCE, SNAP, MintPy, and StaMPS, using PS and SBAS InSAR methods.

The results are shown in Figure 4.5 and Figure 4.6 in terms of relative and absolute (georeferenced) velocities, and the associated errors. These metrics are presented for each combination of software, method, and station.

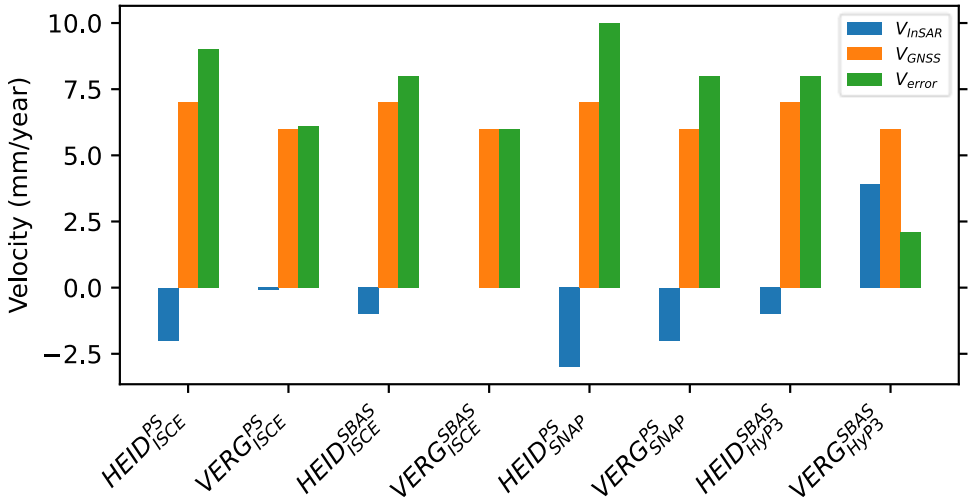


Figure 4.5: Relative LOS Velocity differences between cGNSS and InSAR implied.

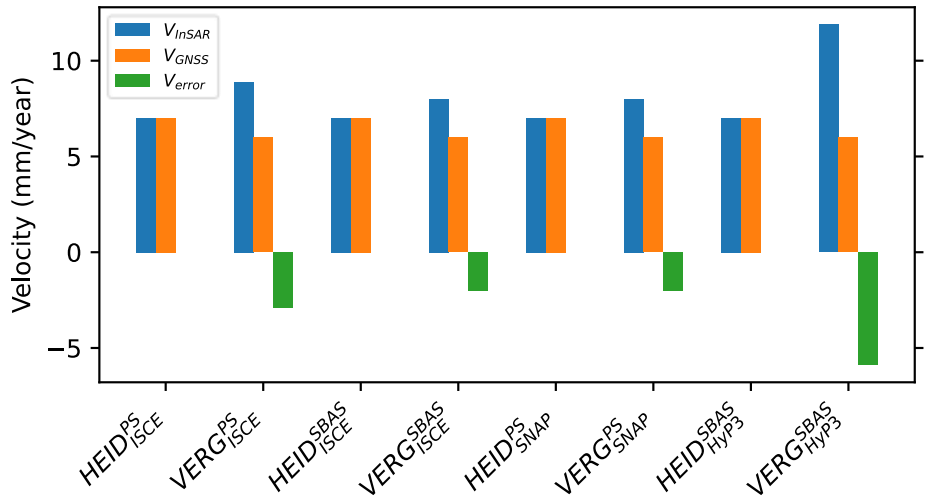


Figure 4.6: Absolute LOS Velocity differences between cGNSS and InSAR implied.

The results indicate that the selection of InSAR pre-processing and processing software significantly affects the accuracy of ground subsidence measurements. Notably, the relative velocity errors varied considerably across different methods and software. For instance, SNAP PS-InSAR at stations HEID and VERG showed high velocity errors of 10.0 mm/year and 8 mm/year respectively, followed by ISCE PS-InSAR and SBAS-InSAR, and HyP3 SBAS-InSAR with least errors. Generally, ISCE SBAS-InSAR demonstrated better consistency with GNSS data compared to ISCE PS-InSAR, likely due to the SBAS techniques effectiveness in mitigating phase unwrapping errors and atmospheric delays.

The absolute velocities were closer to cGNSS measurements than relative velocities, yet discrepancies were still evident. Since the HEID station was used for adjustment of velocities, consistent zero error across all techniques absolute velocity measurements is observed. However, at the station VERG, higher errors were observed, particularly with HyP3 SBAS-InSAR (−5.9 mm/year), suggesting potential issues with this processing method for absolute velocity at this location, however this large error is consistent with the HyP3 standard error map which showed larger standard error on the western part of the study area in the maximum order of up to ±5.4 mm/year. The second poor performing technique at station VERG is ISCE PS-InSAR, with noticeable discrepancies in both relative and absolute velocities. This could be attributed to site-specific factors such as local atmospheric conditions or ground characteristics affecting PS processing.

Further statistical analysis of relative velocity differences was conducted as shown in Table 2, The results indicated that HyP3 SBAS-InSAR had the lowest mean absolute error of 5.1 mm/year, followed by ISCE SBAS-InSAR, ISCE PS-InSAR, and lastly SNAP PS-InSAR. Moreover, ISCE SBAS-InSAR and SNAP PS-InSAR has the same RMSE and range of 1.4 mm/year and 2.0 mm/year respectively, followed by ISCE PS-InSAR, and lastly HyP3 SBA-InSAR. Although HyP3 SBAS-InSAR had the lowest mean absolute error, it also displayed a wider error range and higher RMSE of 5.9 mm/year and 4.2 mm/year, respectively, suggesting variability in performance.

Table 2: Relative Velocity differences statistics (units: mm/year)

STAT	**Persistent Scatterer**		**Small Baseline Subset**	
	ISCE$_{StaMPS}$	**SNAP$_{StaMPS}$**	**ISCE$_{StaMPS}$**	**HyP3$_{MintPy}$**
MAE	7.6	9.0	7.0	5.1
Range	2.9	2.0	2.0	5.9
RMSE	2.1	1.4	1.4	4.2

The absolute velocity differences statistics revealed similar trends, with ISCE SBAS-InSAR and SNAP PS-InSAR showing lower and equal values for mean absolute error, range, and RMSE of -1.0 mm/year, 2.0 mm/year, and 1.4 mm/year, respectively. This implies that ISCE SBAS-InSAR and SNAP PS-InSAR are more suitable for tectonic plate monitoring, where

precise measurements of smaller deformations are crucial, and techniques with lower absolute errors and realistic plate velocity ranges are essential.

Table 3: Absolute Velocity differences statistics (units: mm/year)

STAT	Persistent Scatterer		Small Baseline Subset	
	ISCE$_{StaMPS}$	**SNAP$_{StaMPS}$**	**ISCE$_{StaMPS}$**	**HyP3$_{MintPy}$**
MAE	-1.5	-1.0	-1.0	-3.0
Range	2.9	2.0	2.0	5.9
RMSE	2.1	1.4	1.4	4.2

In contrast, HyP3 SBAS-InSAR stands out with the highest RMSE, lowest relative uncertainty (0.0095) and a broad velocity range from 49.2 to -233.9 mm/year, as depicted in Figure 4.7, making it the least accurate, however, on the other making it the best choice for capturing significant localised ground deformation such as ground subsidence. For ground subsidence monitoring and susceptibility analysis, which often involves broader and more gradual movements, the method with the lowest relative uncertainty and a broader velocity range is preferable.

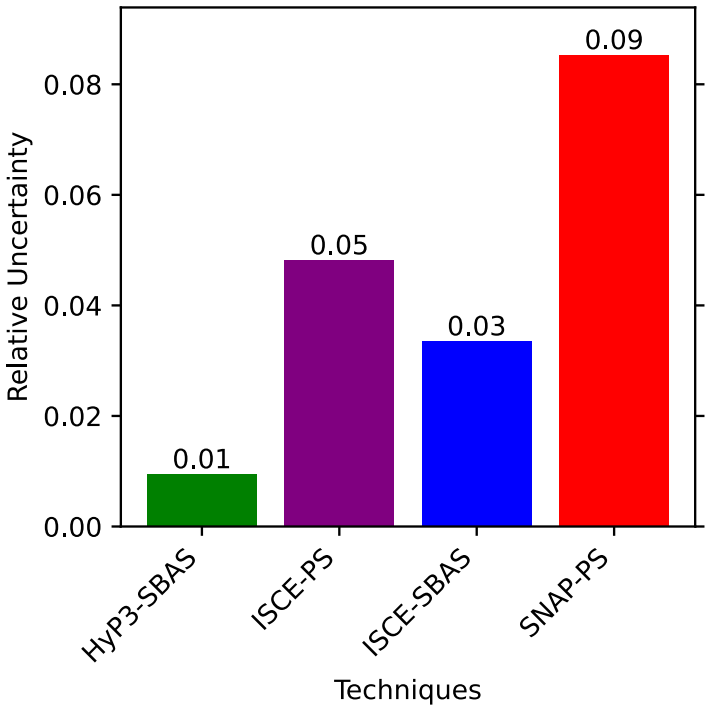


Figure 4.7: Technique and tools relative uncertainty.

Additionally, to assess whether InSAR and Sentinel-1 meet the stringent validation criteria of the NISAR mission, we examined the relative baseline error. When comparing the cGNSS implied velocity differences between HEID and VERG (-1 mm/year) with those of InSAR, we found a relative baseline error of ± 2.9 mm/year, ± 2.0 mm/year, ± 2.0 mm/year, and ± 5.9

mm/year over a distance of 49.73 kilometres for ISCE PS-InSAR, ISCE SBAS-InSAR, SNAP PS-InSAR, and HyP3 SBAS-InSAR respectively, as shown in Table 4.

Table 4: Baseline (HEID-VERG) LOS velocity differences between cGNSS and InSAR implied

Method ID	Baseline Error (mm/year)			Distance (km)	NISAR Error Threshold
	V_{InSAR}	V_{GNSS}	V_{error}		$3(1 + L^{0.5})$
ISCE-PS	-1.9	1.0	2.9	49.73	24.16
ISCE-SBAS	-1.0	1.0	2.0	49.73	24.16
SNAP-PS	-1.0	1.0	2.0	49.73	24.16
HyP3-SBAS	-4.9	1.0	5.9	49.73	24.16

This analysis shows that ISCE SBAS-InSAR and SNAP PS-InSAR meet the NISAR validation requirement of 2 mm/year for secular deformation, while ISCE PS-InSAR and HyP3 SBAS-InSAR do not. However, all methods meet the requirement of ± 24.16 mm/year for transient deformation over length scales from 0.1 km to 50 km baselines. It is noteworthy that these validation criteria are tailored to NISAR, which uses the L-band, while Sentinel-1 operates in the C-band. Nonetheless, these results demonstrate the suitability of InSAR and Sentinel-1 for effective ground deformation monitoring across the study area.

In summary, the evaluation of pre-processing and processing software for ground subsidence monitoring reveals distinct performance characteristics among the techniques. ISCE SBAS-InSAR exhibited the lowest RMSE, second lowest relative uncertainty and was the most precise, while SNAP PS-InSAR showed a higher frequency at around -2 mm/year, suggesting potential errors in its measurements. HyP3 SBAS-InSAR, despite having the highest RMSE, demonstrated the lowest relative uncertainty and a broader velocity range, making it particularly effective for capturing localised ground deformations such as ground subsidence. Due to its lowest relative uncertainty, HyP3 SBAS-InSAR was selected to conduct subsequent investigations.

4.2 Effects of Digital Elevation Models

In this section, a detailed investigation and statistical analysis was performed to evaluate the impact of Digital Elevation Models selection on the accuracy of ground subsidence susceptibility mapping. The analysis included the validation of DEMs using 247 ground levelling points within the study area, followed by the use of XGBoost supervised machine learning algorithm, HyP3 SBAS-InSAR descending orbit data, levelling data, and DEM datasets to predict velocities. The differences between levelling and DEM-implied velocities were statistically analysed to assess their accuracy and effects.

4.2.1 DEM Vertical Accuracy Assessment

The initial assessment involved comparing the elevations from levelling points with those from the original DEMs, and the statistics of these differences are summarised in Figure 4.8. The results reveal that the mean error for ASTER is 8.4 m, while mean error for AW3D30, COP30, and SRTMGL1 is 5.1 m. The standard deviation of errors for ASTER is 13.9 m, whereas the standard deviation for AW3D30, COP30, and SRTMGL1 is 11.8 m. The root mean square errors for ASTER, AW3D30, COP30, and SRTMGL1 are 16.2 m, 12.9 m, 14.3 m, and 13.1 m respectively. These results indicate that ASTER has the poorest performance among the DEMs, while AW3D30, COP30, and SRTMGL1 exhibit similar performance. The RMSE values reveal that AW3D30 has the best performance, followed by COP30, SRTMGL1, and finally ASTER.

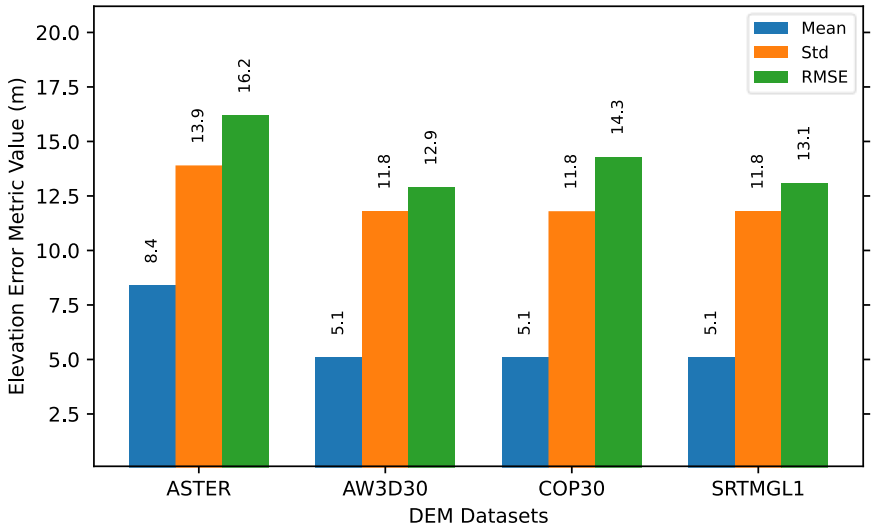


Figure 4.8: The statistics of the elevation errors before coregistration over Midvaal, South Africa.

Following this, the DEMs were co-registered to align their grid centroids, subsequently the elevation differences statistics are recomputed. The statistics of the differences between levelling points and co-registered DEM elevations are provided in Figure 4.9. Post co-registration, the results reveal that the mean errors for ASTER, AW3D30, COP30, and SRTMGL1 are 8.3 m, 5.1 m, 5.1 m, and 5.1 m respectively. The standard deviations of errors for ASTER is 13.8 m, and for AW3D30, COP30, and SRTMGL1 is 11.8 m. The RMSEs for ASTER, AW3D30, COP30, and SRTMGL1 are 16.1 m, 12.9 m, 14.3 m, and 12.9 m respectively. These results indicate a slight improvement for ASTER and SRTMGL1, with the RMSEs decreasing by 0.1 m and 0.2 m, respectively. The errors for AW3D30 and COP30 remained unchanged, this is due to the selection of AW3D30 as the reference grid, and the unchanged error of COP30 underscoring the coherence and robustness of the two datasets and the reference grid. The positive mean errors suggest that all DEMs tend to underestimate the

actual ground elevations. Overall, these results are consistent with studies conducted by Malindi and Odera (2022), and Okolie et. al. (2024).

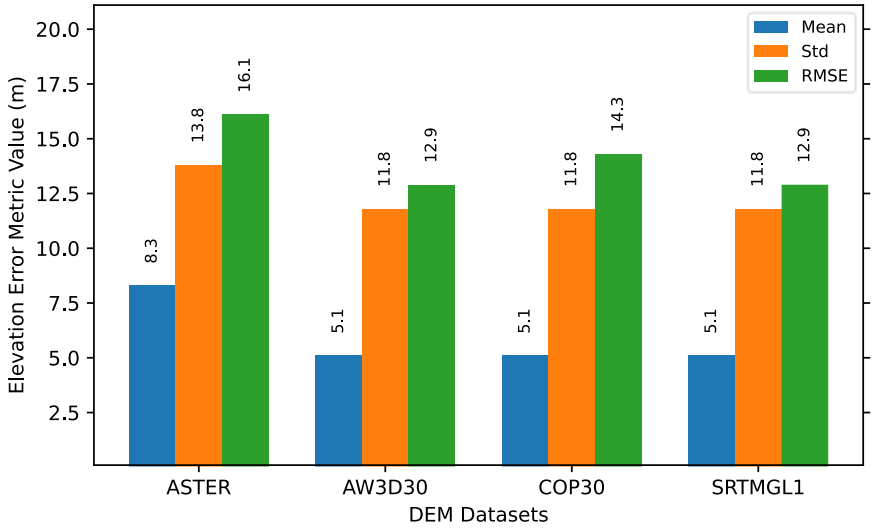


Figure 4.9: The statistics of the elevation errors after coregistration over Midvaal, South Africa.

To further visualise the spatial distribution of the elevation errors, the point elevation errors were interpolated over the study region using Inverse Distance Weighting interpolation algorithm in ArcPy, as shown in Figure 4.10. The elevation error ranges for ASTER, AW3D30, COP30, and SRTMGL1 are 134.8 m, 100.9 m, 100.3 m, and 100.5 m, respectively, indicating that ASTER has the worst performance. AW3D30, COP30, and SRTMGL1 show minimal elevation errors in most of the study area, with significant errors in the southwestern part, which corresponds to multi-storey buildings, mining sites, and highly vegetated ground. This can be attributed to challenges such as occlusions, variable surface reflectance, and rapid elevation changes, which complicate the InSAR, and photogrammetric methods used in DEM generation.

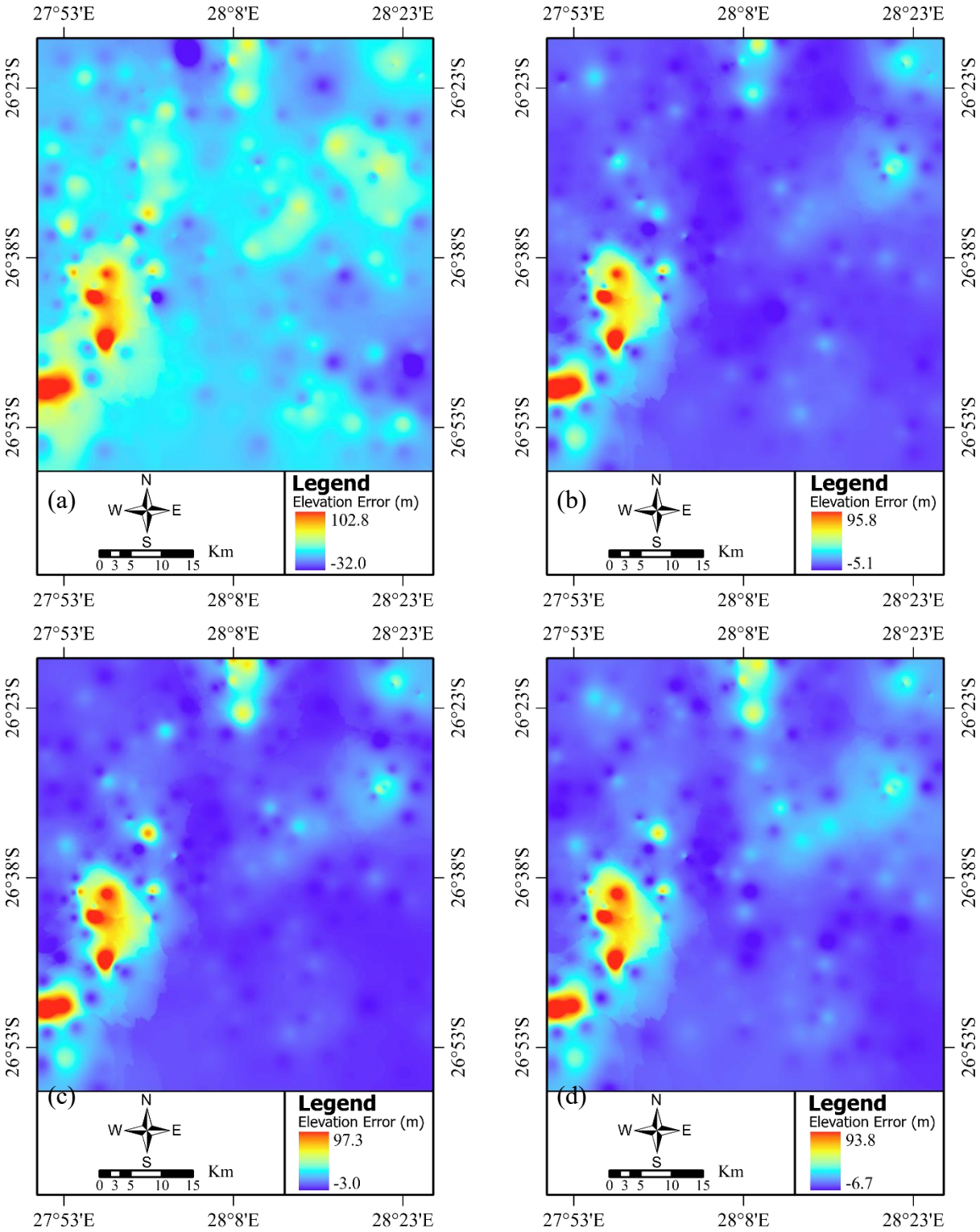


Figure 4.10: Spatial distribution of elevation errors over the study area, (a) ASTER, (b) AW3D30, (c) COP30, (d) SRTMGL1.

Subsequently, the distribution of elevation errors was further analysed using frequency and bar graphs, as depicted in Figure 4.11 and Figure 4.12 respectively. The results from Figure 4.11 indicates that COP30 has the most errors close to zero, followed by AW3D30, SRTMGL1, and ASTER. The AW3D30 DEM has the fewest errors greater than 40 m, followed by COP30,

SRTMGL1, and ASTER. These results highlight AW3D30's precision in estimating ground elevations with minimal significant errors.

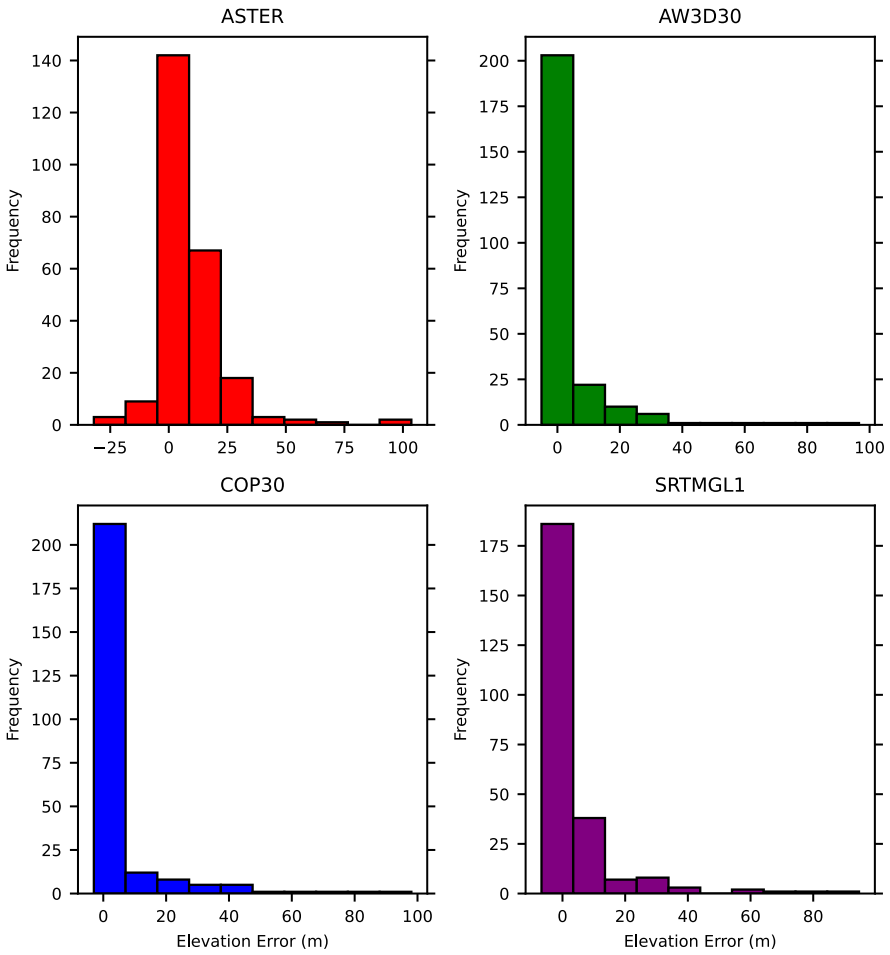


Figure 4.11: Elevation error frequency distribution over 247 levelling points in Midvaal.

The results from Figure 4.12 shows that AW3D30 has the most elevation errors in the bins between -7 m to 21 m and 21 m to 49 m, followed by SRTMGL1, COP30, and ASTER, highlighting AW3D30's precision in estimating ground elevations.

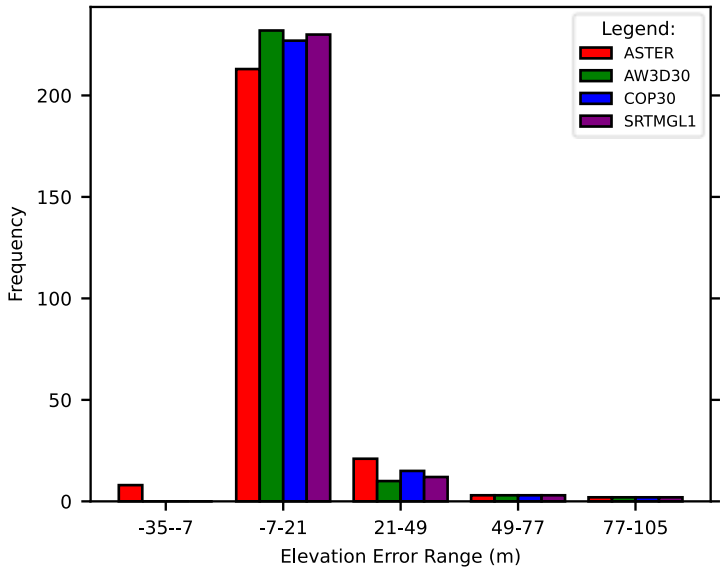


Figure 4.12: Elevation error range frequency over 247 levelling points in Midvaal.

Moreover, understanding the variation of the elevation errors across different elevations is crucial. Figure 4.13 depicts the relationship between DEM elevation errors and DEM-implied elevations. The results reveal that all DEMs show a negative slope, indicating that as elevation increases, the elevation error decreases. However, the regression coefficient of determination (R^2) for ASTER, AW3D30, COP30, and SRTMGL1 were 0.005, 0.013, 0.019, and 0.022, respectively, suggesting a weak relationship between elevation and elevation errors, implying other factors are at play.

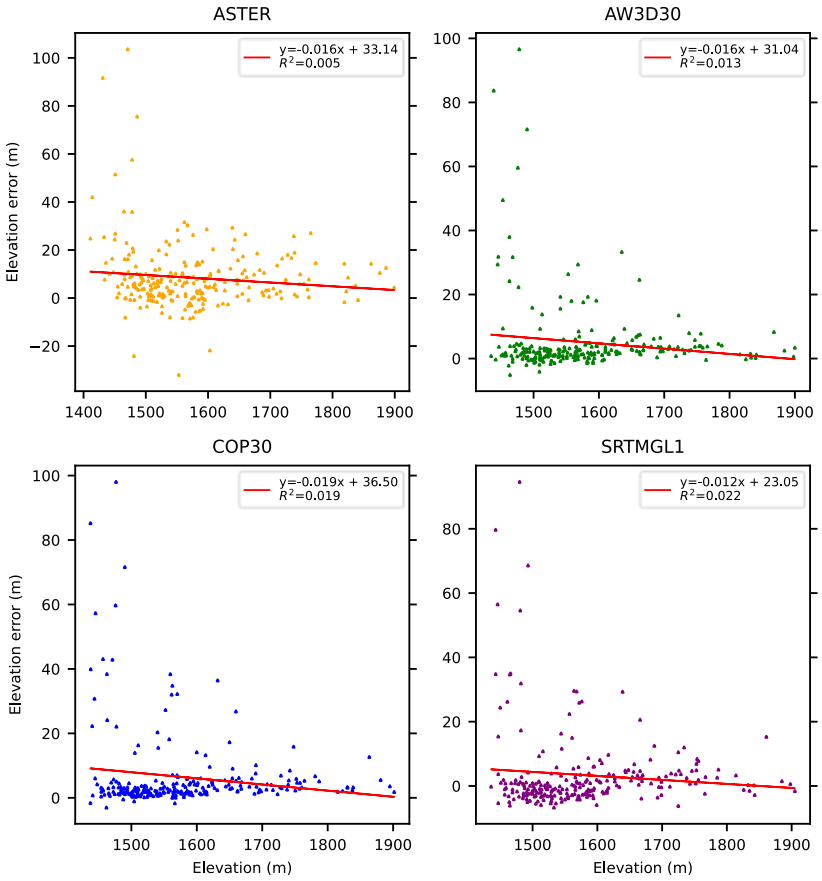


Figure 4.13: Relationship between elevation and elevation error over 247 levelling points in Midvaal, South Africa.

Overall, these findings highlight the complexities in DEM elevation errors and their spatial distribution. Moreover, AW3D30 demonstrates the best accuracy performance, followed by COP30, SRTMGL1, and ASTER. The weak correlation between elevation and elevation errors suggests that modelling these errors is complex, and their impact on ground subsidence analysis will be explored further in the subsequent section.

4.2.2 DEM Velocity Accuracy Assessment using Levelling dataset

To understand how elevation errors propagate into predicted velocity values in ground subsidence susceptibility mapping, the discrepancies between ground levelling point and DEM-implied velocities were statistically analysed, as shown in Figure 4.14. The results of the mean absolute errors for ASTER, AW3D30, COP30, and SRTMGL1 are 1.2 mm/year, 0.8 mm/year, 1.0 mm/year, and 1.1 mm/year, respectively. The RMSEs for ASTER, AW3D30, COP30, and SRTMGL1 are 2.0 mm/year, 1.1 mm/year, 1.4 mm/year, and 1.7 mm/year respectively. The R^2 scores are 52.9% for ASTER, 84.8% for AW3D30, 77.3% for COP30, and 66.4% for SRTMGL1. These results indicate that AW3D30 has the best performance in ground subsidence velocity modelling, followed by COP30, SRTMGL1, and ASTER. These results are consistent with the elevation error performance discussed earlier.

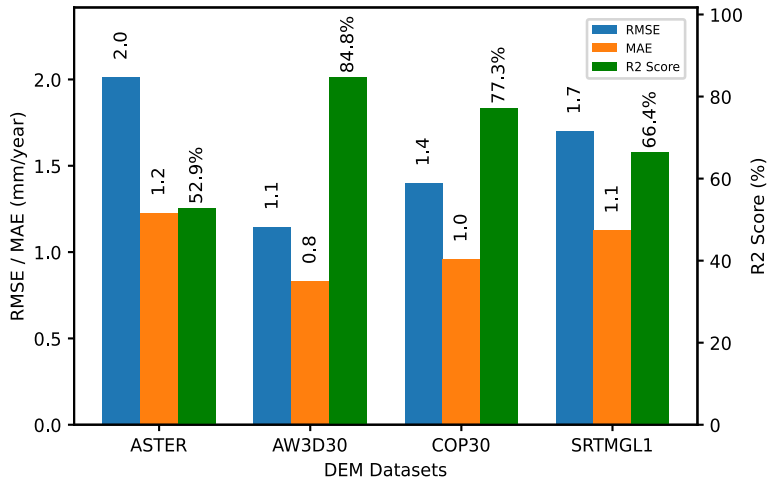
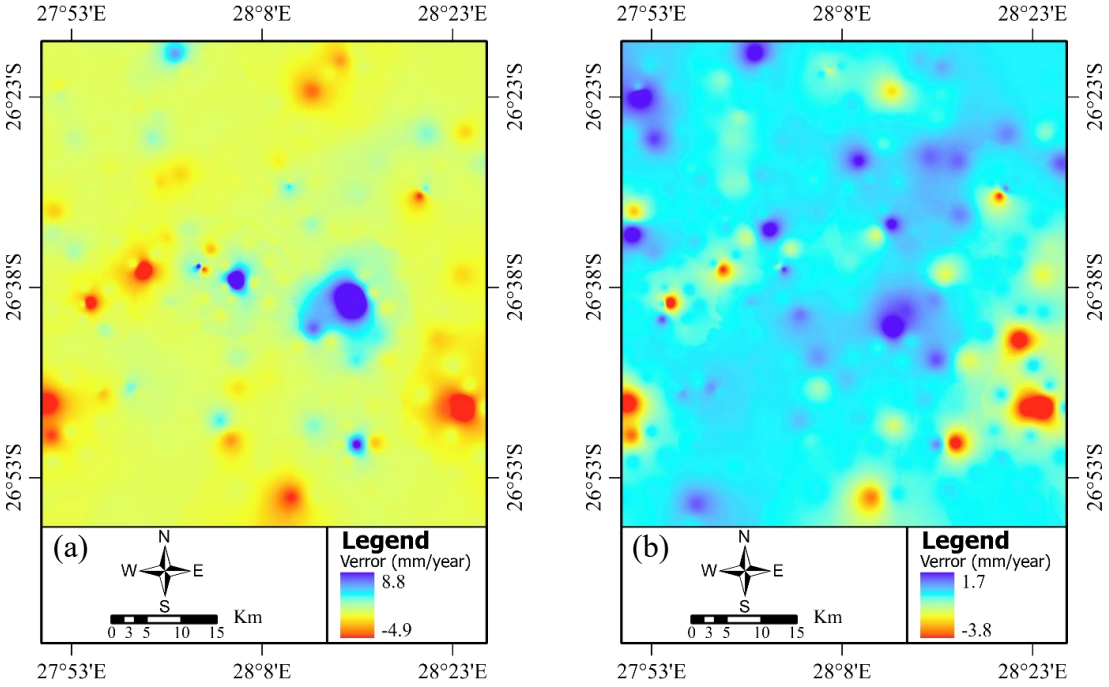


Figure 4.14: DEM velocity modelling statistics over 247 levelling points in Midvaal.

Subsequently, the spatial distribution of velocity errors was visualised using IDW interpolation in ArcPy, as shown in Figure 4.15. The results reveal that the velocity error ranges for ASTER, AW3D30, COP30, and SRTMGL1 are 13.7 mm/year, 5.5 mm/year, 7.8 mm/year, and 12.7 mm/year, respectively. AW3D30 exhibits the best performance, followed by COP30, SRTMGL1, and ASTER. Major velocity errors were observed in highly vegetated areas with trees and croplands.



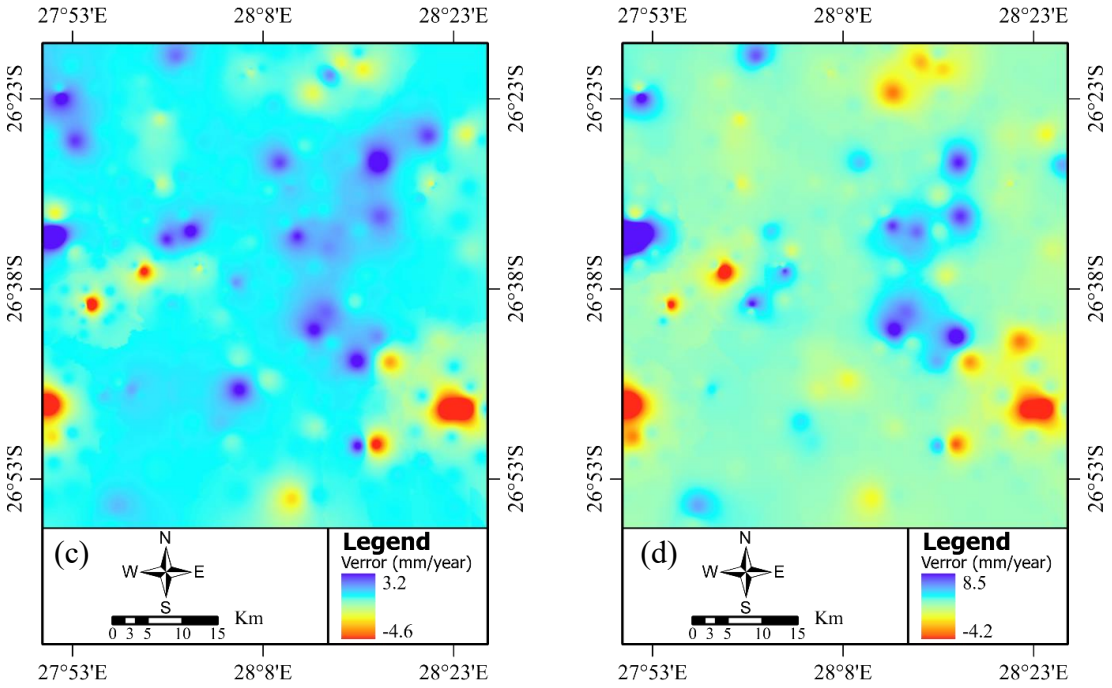


Figure 4.15: Velocity error spatial distribution over Midvaal, (a) ASTER, (b) AW3D30, (c) COP30, (d) SRTMGL1.

Subsequently, the distribution of velocity errors was further analysed using frequency distribution graphs, as shown in Figure 4.16. The results reveal that AW3D30 has the highest frequency of velocity errors within the -2.5 mm/year to 2 mm/year bin, followed by COP30, SRTMGL1, and ASTER. The ASTER DEM shows velocity errors extending up to -5 mm/year and 7.5 mm/year, indicating its poor performance.

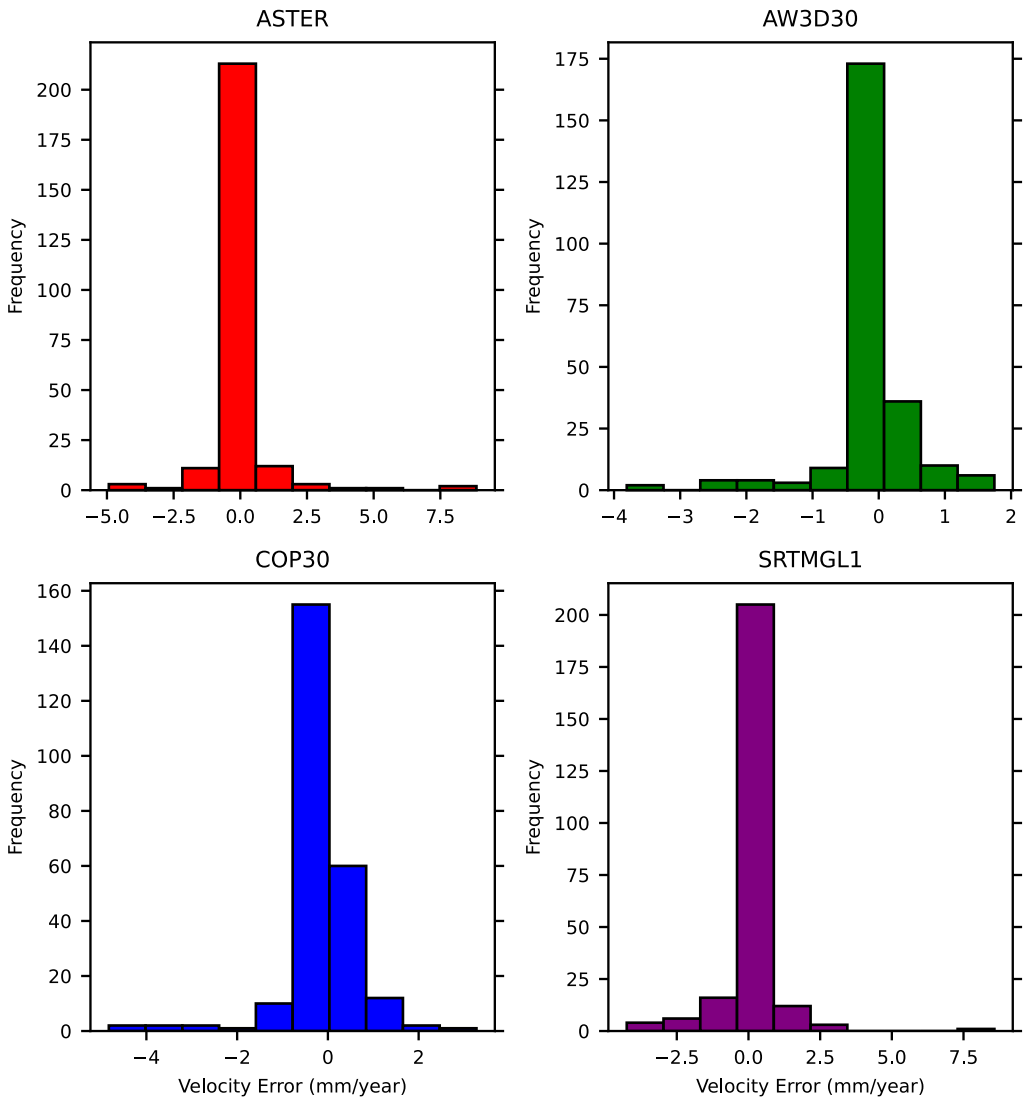


Figure 4.16: Velocity error frequency distribution over 247 levelling points in Midvaal.

Moreover, To understand the relationship between velocity errors and other variables, the relationships between velocity and velocity error, elevation and velocity error, and elevation error and velocity error were examined. The results from Figure 4.17 shows the relationship between velocity and velocity error, revealing positive slopes for AW3D30 and COP30, indicating that as velocity increases, velocity error also increases. In contrast, ASTER and SRTMGL1 show negative slopes. However, the coefficient of determination for ASTER, AW3D30, COP30, and SRTMGL1 are 0.045, 0.021, 0.012, and 0.003, respectively, indicating weak relationships.

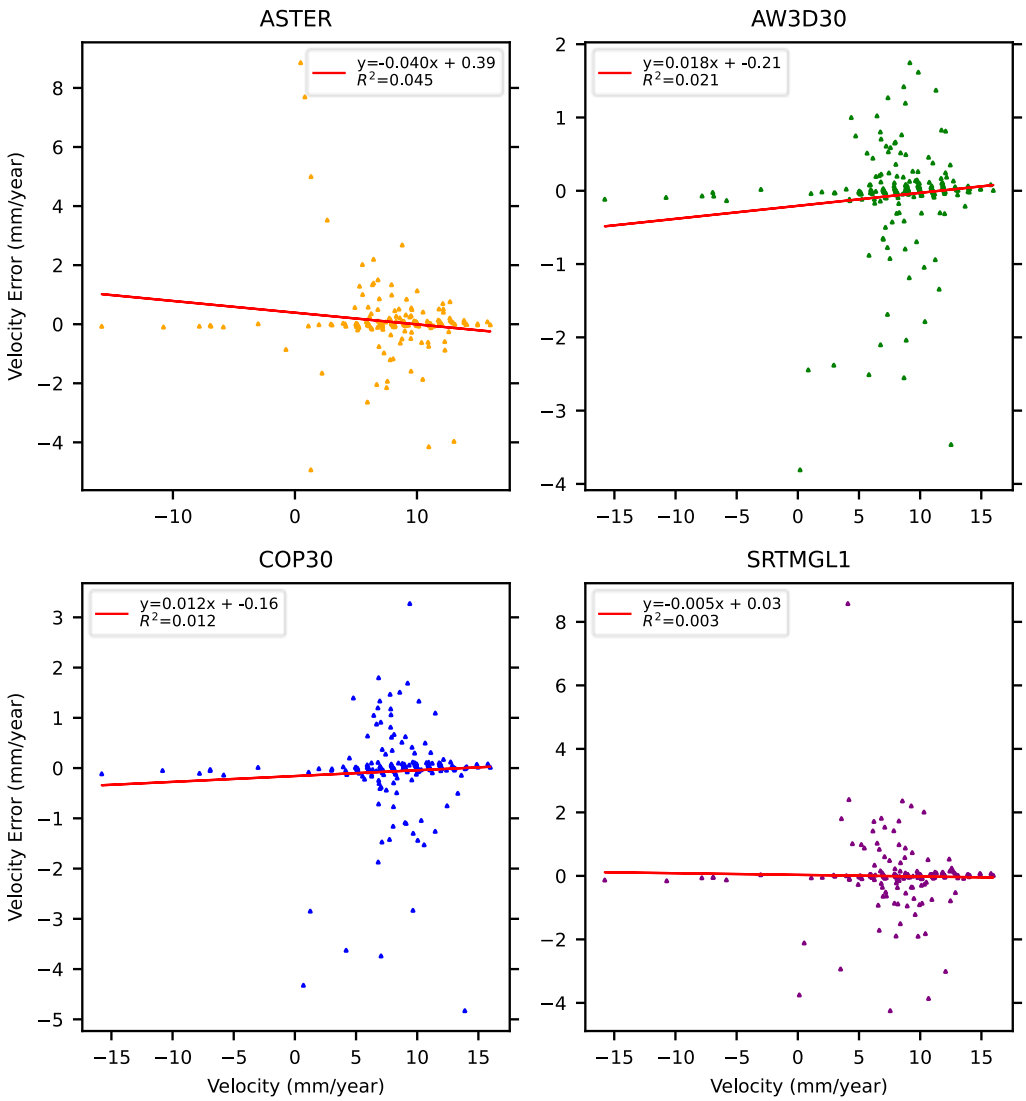


Figure 4.17: Relationship between velocity and velocity error over 247 levelling points in Midvaal, South Africa.

Moreover, Figure 4.18 depicts the relationship between elevation and velocity error over 247 levelling points in the study area. The results reveals that all the DEMs slopes between elevation and velocity errors are positive, equal, and close to zero (0.001 mm/year), which implies that there is no or minimal relationship between elevation and velocity error. Moreover, the results from the coefficient of determinations between velocity and velocity errors for ASTER, AW3D30, COP30, and SRTMGL1 are 0.002, 0.004, 0.004, and 0.012 respectively, indicating that these relationships are highly weak or poor, thus elevation should not solely be used in attempts to model velocity error. Additionally, high variations of velocities errors can be observed within the lower elevations, highlighting that in lower elevations the velocity errors are intricate.

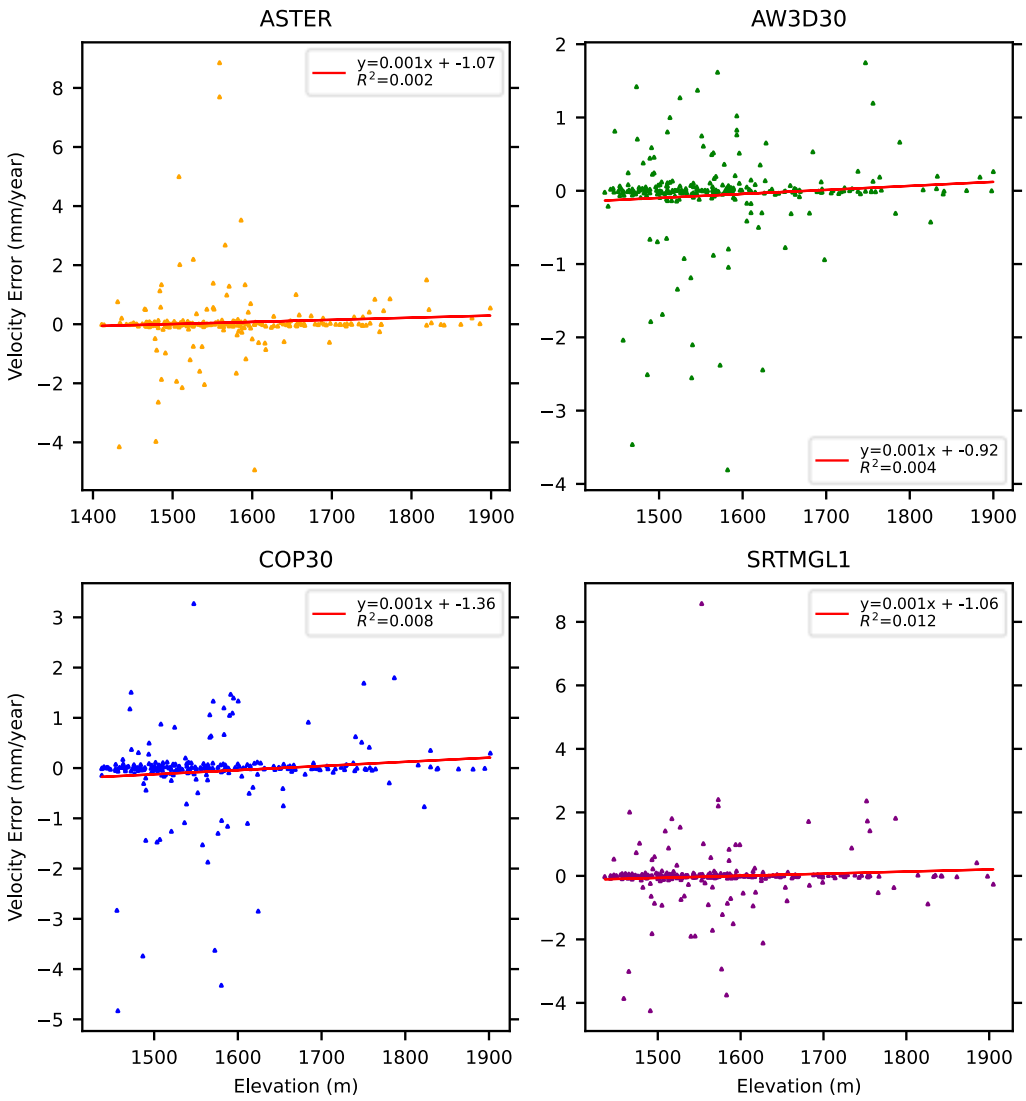


Figure 4.18: Relationship between elevation and velocity error over 247 levelling points in Midvaal, South Africa.

Finally, Figure 4.19 depicts the relationship between elevation and velocity error over 247 levelling points in the study area. The results indicate minimal correlation between velocity error and height error for all DEMs assessed, as indicated by the near-zero R^2 values. For ASTER, AW3D30, COP30, and SRTMGL1, the coefficient of determination values of 0.000 and 0.006 imply that height error does not significantly explain the variation in ground velocity error. The regression coefficients vary slightly, with negative coefficients for AW3D30, COP30, and SRTMGL1, suggesting a minor inverse relationship. However, the negligible R^2 values highlight the limitations of these models in accurately predicting ground velocity based on height error, indicating that other factors may play a more significant role in influencing velocity variations. Further research is needed to explore these additional factors and improve the predictive capability of DEM-based models for ground velocity estimation.

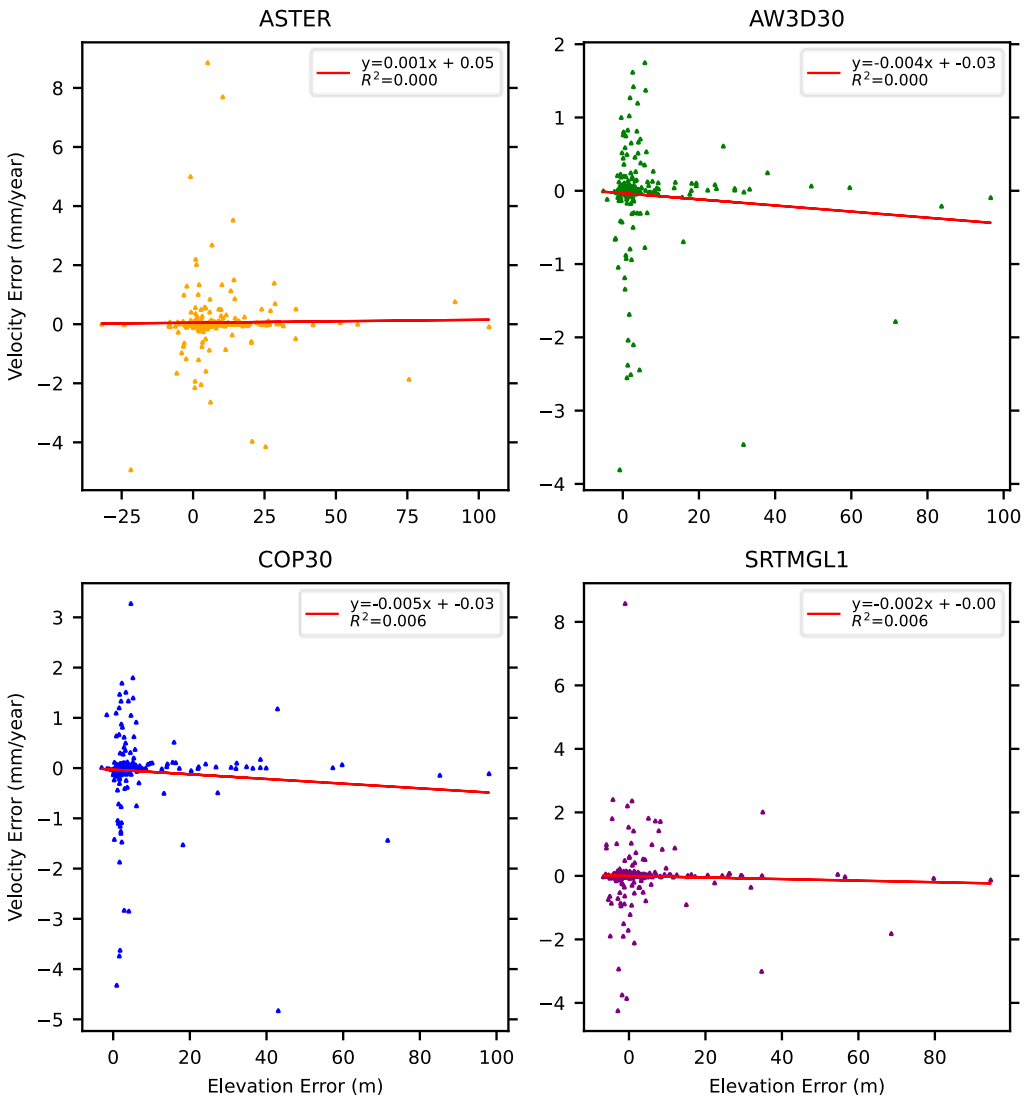


Figure 4.19: Relationship between elevation error and velocity error over 247 levelling points in Midvaal, South Africa.

Overall, these results highlight the complexities involved in modelling velocity and velocity errors using DEMs and the weak correlations between elevation and velocity errors. Despite AW3D30's superior performance in terms of accuracy, the relationships between variables indicate the need for more sophisticated modelling techniques to account for the various factors influencing ground subsidence susceptibility mapping.

4.3 Spatial and Temporal Patterns in Midvaal, South Africa

In this section, an investigation and statistical analysis were conducted to evaluate the rates and spatio-temporal patterns of ground subsidence in Midvaal. The analysis involved the use of HyP3 SBAS-InSAR velocity measurements, with a focus on comparing ground deformation patterns from both Sentinel-1 C-band ascending and descending orbits. This assessment included examining the maximum ground subsidence and uplift rates, as well as the spatial distribution of these phenomena. The study also explored temporal dynamics by selecting and

analysing ten random sample points across varying susceptibility levels. Statistical comparisons and correlations with ground water storage anomaly, precipitation, water indices, and vegetation indices were performed to assess their influence on ground deformation. The findings aim to provide a thorough understanding of ground subsidence dynamics and the contributing factors in the study area.

The HyP3 SBAS-InSAR velocity results revealed significant findings as shown in Figure 4.20. The results indicates pronounced ground subsidence in the south-eastern, north-eastern, and south-western regions of the study area, with maximum ground subsidence rates reaching -226.0 mm/year for descending orbit and -214.3 mm/year for ascending orbit. Conversely, notable uplift was observed in the western part and at the edges of the south-eastern region, with maximum uplift rates of 57.2 mm/year for descending orbit and 54.6 mm/year for ascending orbit.

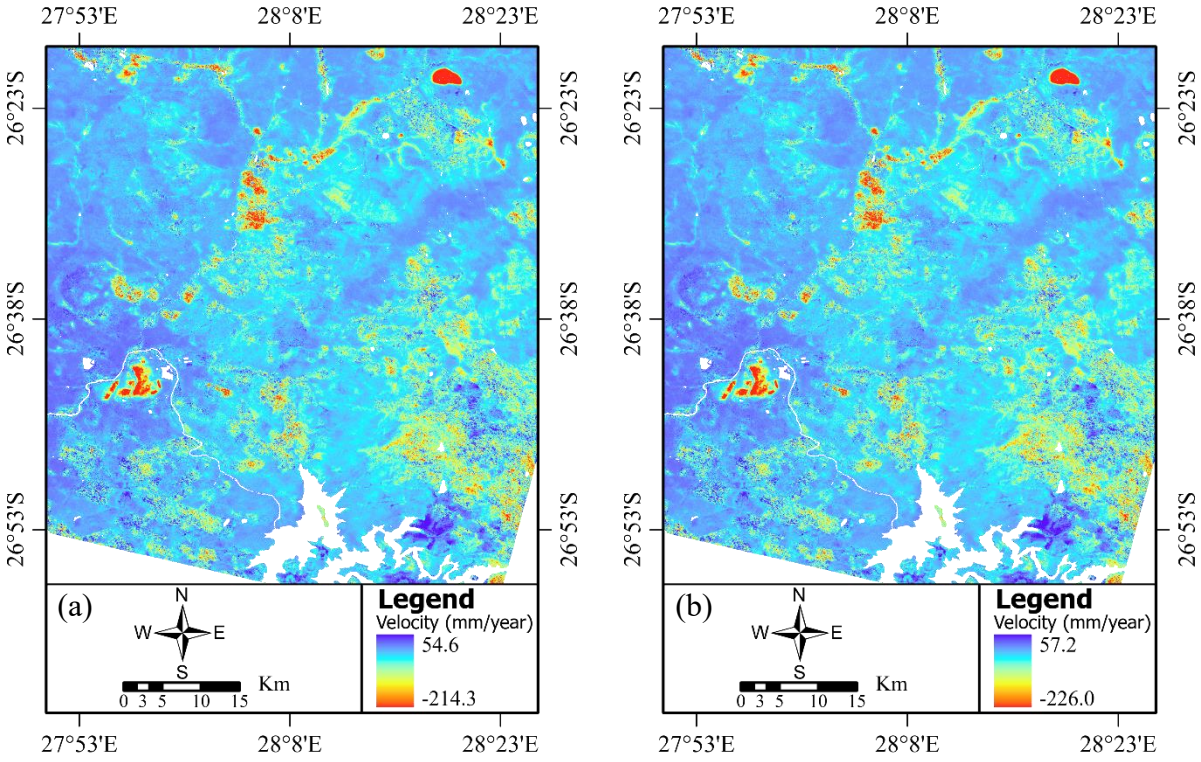


Figure 4.20: HyP3 SBAS-InSAR velocities in Midvaal, (a) Ascending orbit clipped to the footprint of descending track, (b) Descending orbit.

The spatial patterns of ground deformation identified from both ascending and descending orbits exhibit a high degree of similarity. This is supported by a high coefficient of determination of 1 between the ascending and descending measurements as shown in Figure 4.21, underscoring the consistency and reliability of the observed ground deformation patterns. Furthermore, the absolute difference statistics between ascending and descending measurements further emphasise this coherence, with a mean difference of 0.232 mm/year and a high absolute difference count of over 2 million, highlighting the robustness of Sentinel-1 data in capturing ground deformation.

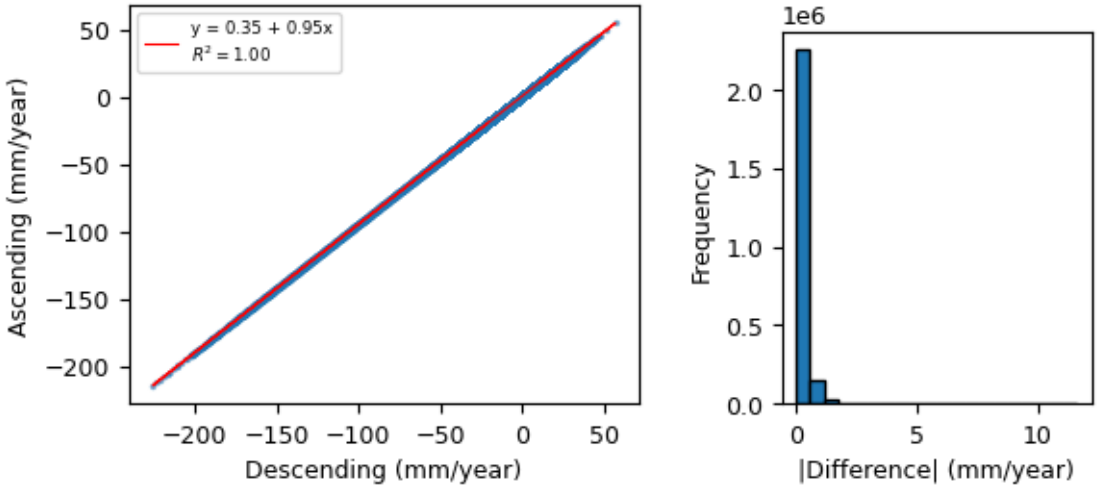


Figure 4.21: The correlation between ascending and descending orbit (a) scatter plot , (b) frequency plot of absolute differences.

The statistics of the absolute differences between ascending and descending measurements show a low mean difference (0.232 mm/year) and a standard deviation (0.337 mm/year), indicating that most of the differences are small and consistent. The wide range, from a minimum of 0.000 mm/year to a maximum of 11.653 mm/year, suggests that while most differences are minor, there are occasional outliers. The 25th, 50th, and 75th percentiles reflect a concentration of small differences, reinforcing the reliability and accuracy of the data.

Table 5: Statistics of the absolute difference between ascending and descending HyP3 SBAS-InSAR velocities

Statistic	**Value**
Count	2,441,780
Mean	0.232
Standard Deviation	0.337
Minimum	0.000
25th Percentile	0.076
50th Percentile	0.156
75th Percentile	0.267
Maximum	11.653

The similarity in ground deformation patterns across both orbits suggests that the observed ground subsidence and uplift are consistent and accurate, reflecting genuine geological and hydrological processes rather than observational discrepancies. The results not only reinforce the reliability of HyP3 SBAS-InSAR measurements but also provide valuable insights into the spatial distribution and intensity of ground deformation in Midvaal, contributing to a better understanding of ground subsidence dynamics in this region.

Subsequently, a comprehensive analysis was undertaken to explore the temporal dynamics of ground subsidence within the study area by selecting ten random points for focused

investigation. These points were deliberately chosen to represent a broad range of susceptibility levels as shown in Figure 4.22, providing a thorough understanding of ground subsidence patterns and trends over time. Given that the differences between ascending and descending velocities were minimal, the time series analysis used ascending velocity data only.

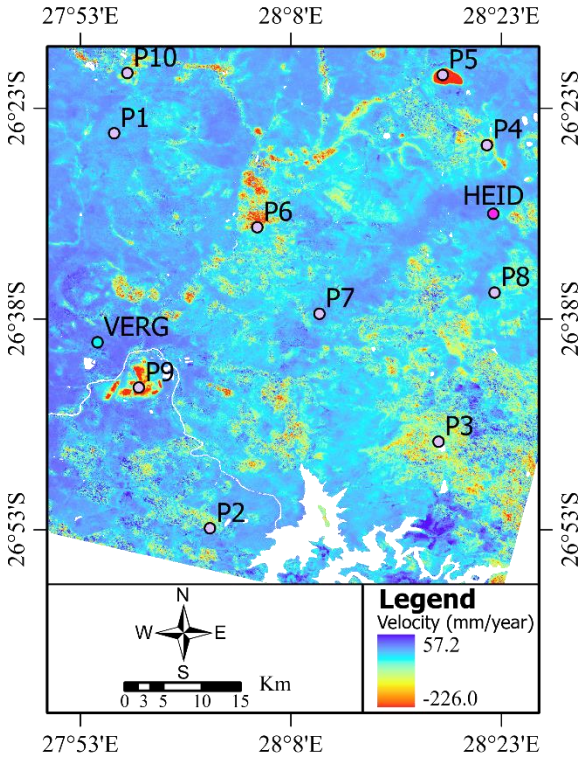


Figure 4.22: Points of interest overlaid on LOS deformation maps derived from HyP3 SBAS-InSAR and Sentinel C-Band descending mode.

Table 6: Ground deformation rates for cGNSS stations in the Study Area

Name	GNSS (mm/year)	Un-adjusted		Adjusted	
		Ascending (mm/year)	Descending (mm/year)	Ascending (mm/year)	Descending (mm/year)
HEID	7	-1	-1	7	7
VERG	6	4	4	12	12

Table 7: Ground deformation rates for Points of Interest in the Study Area

Name	Un-adjusted		Adjusted	
	Ascending (mm/year)	**Descending (mm/year)**	**Ascending (mm/year)**	**Descending (mm/year)**
P1	3	3	11	11
P2	-13	-13	-5	-5
P3	-7	-7	1	1
P4	-19	-20	-11	-12
P5	-192	-202	-184	-194
P6	-51	-53	-43	-45
P7	1	1	9	9
P8	-13	-14	-6	-6
P9	-87	-92	-79	-84
P10	-31	-33	-23	-25

4.3.1 Ground Water Storage and Precipitation

- **HEID**

The correlation between ground deformation and ground water storage anomaly, as derived from the GLDAS CLMS model, is shown in Figure 4.23(left). The analysis reveals that at station HEID, the ground is experiencing consistent, albeit minor, ground subsidence while the ground water storage anomaly shows an increasing trend. This observation suggests that the ground subsidence at station HEID is not primarily driven by changes in groundwater levels but is likely influenced by other unaccounted factors.

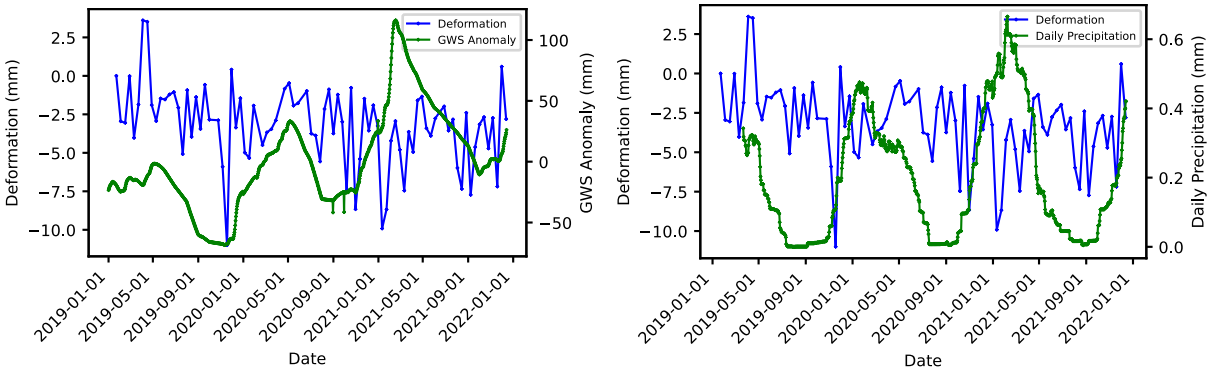


Figure 4.23: The comparison between ground subsidence and: (left) GWS anomaly, (right) daily precipitation at station HEID.

In a similar vein, the correlation between ground deformation and daily precipitation, as shown in Figure 4.23(right), further supports the conclusion that hydrological processes are not the primary drivers of ground subsidence at station HEID. The data indicates that ground subsidence at this station does not correlate with variations in daily precipitation, reinforcing the notion that other factors beyond hydrological influences are contributing to the observed ground subsidence.

- **VERG**

  The relationship between ground deformation and ground water storage anomaly, as derived from the GLDAS CLMS model, is shown in Figure 4.24(left). At station VERG, the data indicates that the ground is undergoing uplift concurrent with an increase in groundwater storage anomaly. Notably, periods of decreasing groundwater storage anomaly correspond with stabilising ground deformation, followed by a resumption of uplift. This pattern suggests that hydrological processes may induce or correlate with ground uplift at station VERG. However, the temporal patterns do not strongly align with the uplift trends, implying that other factors might also be influencing the observed uplift.

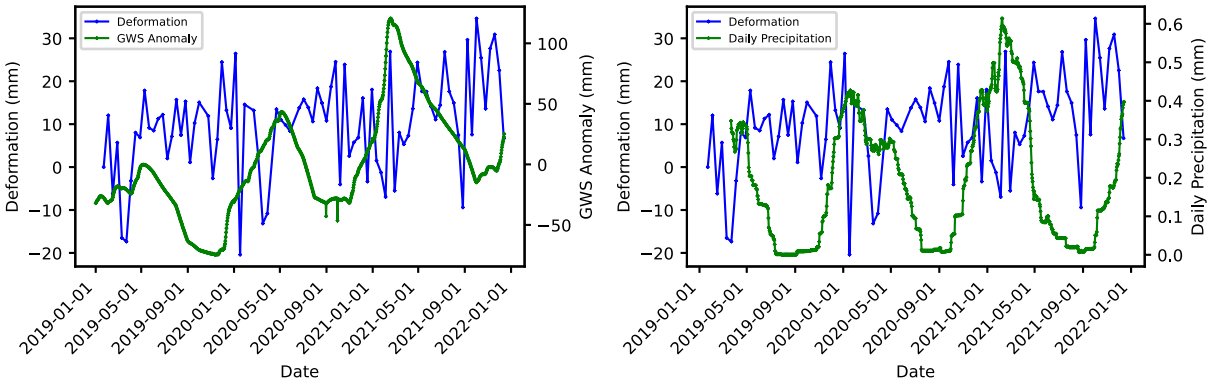


Figure 4.24: The comparison between ground subsidence and: (left) GWS anomaly, (right) daily precipitation at station VERG.

Similarly, the correlation between ground deformation and precipitation, as depicted in Figure 4.24(right), further supporting the analysis that hydrological processes are likely primary drivers of ground subsidence at station VERG. The data shows that ground uplift at this station does correlate with variations in daily precipitation. Nonetheless, some discrepancies in the temporal patterns suggest that other factors, beyond hydrological influences, are also contributing to the observed ground subsidence.

- **GWS Anomaly vs Precipitation**

  The relationship between ground water storage anomaly and 64 days averaged daily precipitation is shown in Figure 4.25. The results reveal a notable similarity in the temporal patterns of precipitation and groundwater storage anomaly, suggesting a strong influence of precipitation on groundwater storage. This indicates that precipitation plays a significant role in affecting groundwater levels.

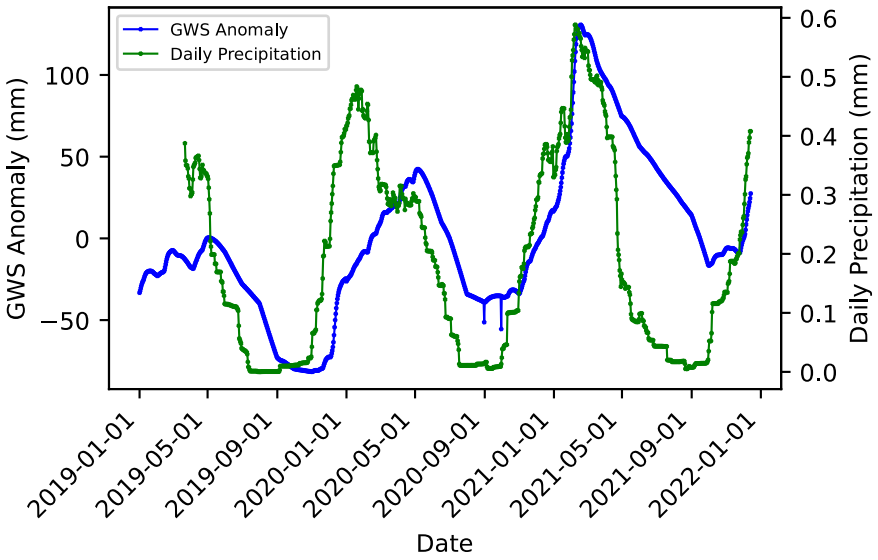


Figure 4.25: The comparison between GWS anomaly and daily precipitation.

However, a phase shift between the peaks of precipitation and groundwater storage anomaly is observed. This delay can be attributed to factors such as ground permeability, percolation rates, and infiltration processes. These factors suggest that groundwater recharge does not occur immediately following a precipitation event, as the water takes time to infiltrate the soil and reach the groundwater system.

Additionally, the observed patterns of precipitation and groundwater storage anomaly do not always show a perfect correlation. This discrepancy may be due to other influences on groundwater levels, such as underground water extraction for crop irrigation and household consumption. These activities can alter groundwater levels independently of precipitation events, thereby affecting the overall correlation between precipitation and groundwater storage anomaly.

Overall, while precipitation is a key driver of groundwater storage, other factors such as groundwater extraction and the time lag in recharge processes also play significant roles in shaping the observed patterns.

4.3.2 Time series: Sample Points

- **Point-1**

Factor	Value
GSSM	Very Low
Lithology	Shale, quartzite, conglomerate, breccia, diamictite
SANLC2018	Natural grassland
Elevation	1681 m
Aspect	South
Slope	0 - 2

The relationship between ground deformation and, NDVI and MNDWI is shown in Figure 4.26. The results reveal a strong correlation between ground deformation and NDVI patterns. Specifically, significant declines in NDVI are associated with stable ground conditions followed by uplift, while rapid increases in NDVI correspond to rapid ground subsidence. Conversely, the correlation between ground deformation and MNDWI indicates that an increase in MNDWI is associated with ground uplift, whereas a decrease in MNDWI corresponds to substantial ground subsidence. This suggests that an increase in vegetation cover can lead to reduced groundwater content, resulting in sediment consolidation and ground subsidence. Additionally, the findings indicate that ground deformation is influenced more by seasonal variations than by long-term fluctuations in NDVI and MNDWI. However, it is crucial to note that these correlations are not consistently observed, suggesting that other factors may also contribute to ground deformation.

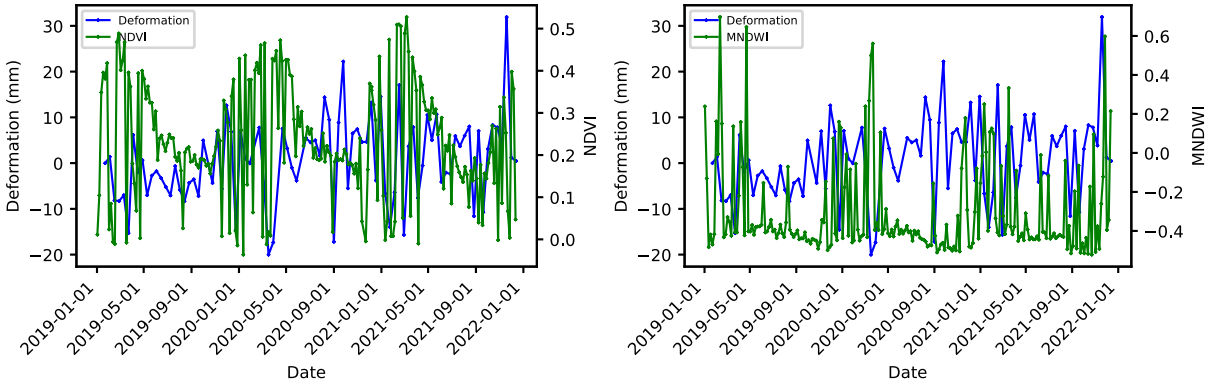


Figure 4.26: The comparison between ground subsidence and: (left) NDVI, (right) MNDWI at sample point-1.

- **Point-3**

Factor	Value
GSSM	Low
Lithology	Rhyolite with subordinate pyroclastic rocks
SANLC2018	Commercial annual crops rain-fed / dryland
Elevation	1518 m
Aspect	East
Slope	0 - 2

The relationship between ground deformation and, NDVI and MNDWI is shown in Figure 4.27. The results reveal a strong correlation between ground deformation and NDVI patterns. Specifically, significant declines in NDVI are associated with stable ground conditions followed by ground subsidence, whereas rapid increases in NDVI correspond to rapid ground uplift. The persistent decline in NDVI likely reflects reduced vegetation cover, which may lead to decreased soil stability and subsequent ground subsidence.

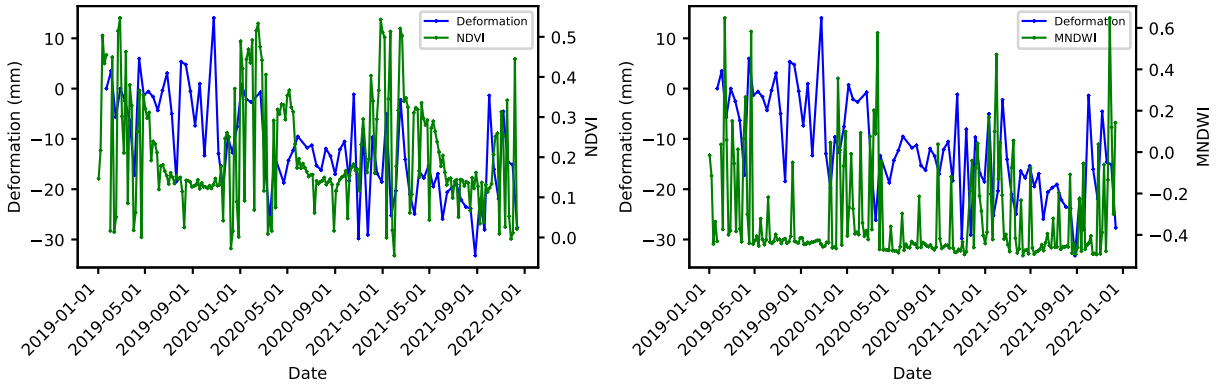


Figure 4.27: The comparison between ground subsidence and: (left) NDVI, (right) MNDWI at sample point-3.

Conversely, the correlation between ground deformation and MNDWI indicates that an increase in MNDWI is associated with ground subsidence, while a decrease in MNDWI corresponds to substantial ground uplift. This pattern suggests that a higher MNDWI, which indicates increased water content, promotes ground subsidence in areas lacking vegetation through processes such as soil erosion. Conversely, a decrease in MNDWI, indicating reduced water content, is linked to stable ground uplift. Therefore, ground subsidence is more closely associated with long-term, gradual declines in NDVI and increase in MNDWI rather than short-term fluctuations.

Overall, the results underscore that ground deformation is influenced by both NDVI and MNDWI, with long-term trends in these indices playing a more significant role

than short-term variations. Nonetheless, the observed patterns are not entirely consistent, suggesting that other factors may also contribute to ground deformation.

- **Point-2**

Factor	Value
GSSM	Moderate
Lithology	Dolerite sills, sheets and dykes
SANLC2018	Commercial annual crops rain-fed / dryland
Elevation	1493 m
Aspect	East
Slope	0 - 2

The relationship between ground deformation and, NDVI and MNDWI is shown in Figure 4.28. The results reveal a clear correlation between ground deformation and NDVI patterns. Specifically, a decline in NDVI is associated with ground uplift, while an increase in NDVI corresponds to ground subsidence. This suggests that short-term or seasonal NDVI fluctuations are the primary drivers of ground deformation, with seasonal changes in vegetation cover influencing soil stability and movement.

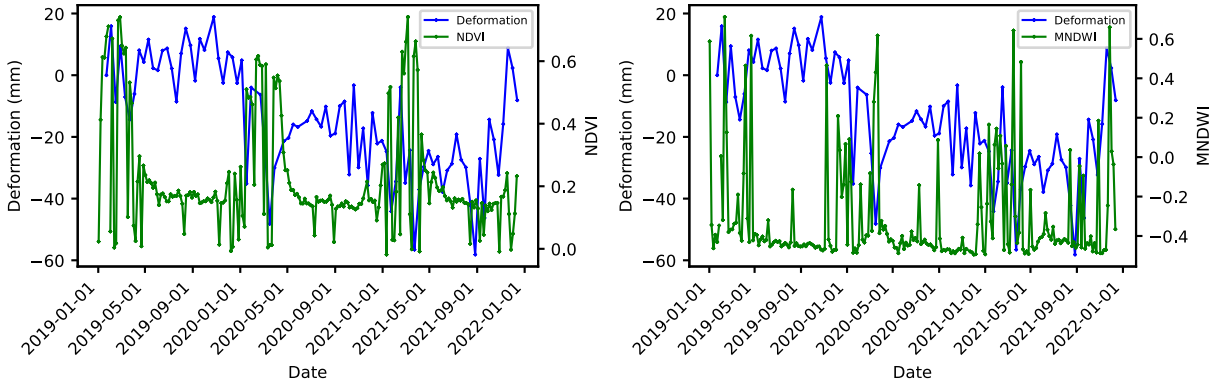


Figure 4.28: The comparison between ground subsidence and: (left) NDVI, (right) MNDWI at sample point-2.

In contrast, the correlation between ground deformation and MNDWI shows that a rapid decline in MNDWI is linked to ground uplift, whereas an increase in MNDWI is associated with ground subsidence. This indicates that ground subsidence is more closely related to long-term declines in MNDWI than to short-term variations. A sustained reduction in MNDWI, reflecting decreased water content, appears to drive long-term ground subsidence, while short-term fluctuations have less impact.

Interestingly, the simultaneous decline in both NDVI and MNDWI leading to ground uplift is unusual and suggests that other factors may be influencing these patterns. This

anomaly implies that additional variables or processes might be affecting the relationship between vegetation, water content, and ground deformation.

- **Point-4**

Factor	Value
GSSM	High
Lithology	Dolomite, subordinate chert, limestone and quartzite
SANLC2018	Herbaceous wetlands
Elevation	1550 m
Aspect	East
Slope	0 - 2

The relationship between ground deformation and, NDVI and MNDWI is shown in Figure 4.29. The results reveal that a decline in NDVI is marked by rapid ground subsidence, while a rapid increase in NDVI is associated with stable ground deformation. Similarly, a decline in MNDWI corresponds to rapid ground subsidence, whereas a rapid increase in MNDWI is linked to stable ground deformation. These findings indicate that ground subsidence is closely related to short-term fluctuations in both NDVI and MNDWI, rather than long-term trends. Despite this, the correlation is relatively low, as rapid ground subsidence remains consistent over time and is less influenced by changes in NDVI and MNDWI. This suggests that while short-term variations in vegetation cover and water content can drive ground changes, the overall ground subsidence trend is consistent and less affected by these fluctuations. The suspected persistence of the ground subsidence is attributed to the underlying lithology, consisting of dolomite and limestone rocks, which are particularly prone to ground subsidence.

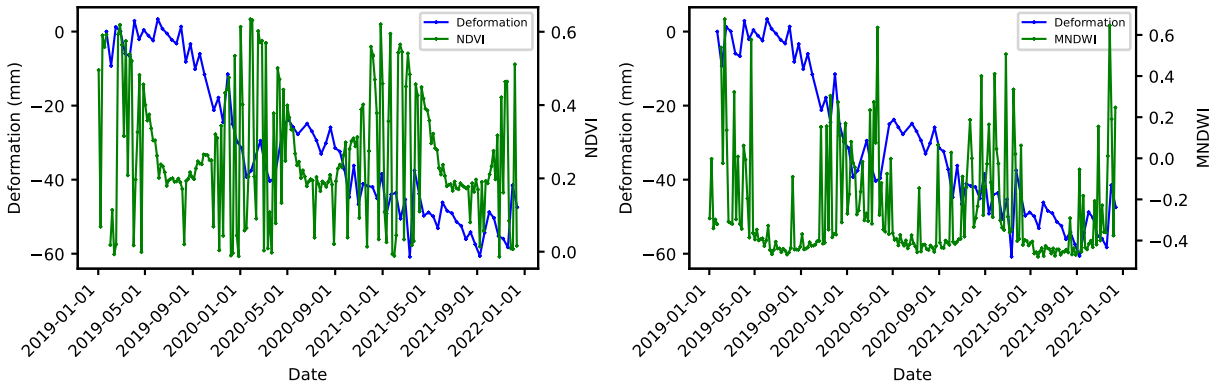


Figure 4.29: The comparison between ground subsidence and: (left) NDVI, (right) MNDWI at sample point-4.

- **Point-5**

Factor	Value
GSSM	Very High
Lithology	Dolerite sills, sheets and dykes
SANLC2018	Artificial flooded mine pits
Elevation	1644 m
Aspect	Flat
Slope	0 - 2

The relationship between ground deformation and, NDVI and MNDWI is shown in Figure 4.30. The results for Point-5, located in an artificial flooded mine with a lithology of Dolerite sills, sheets, and dykes, show that the ground is undergoing continuous rapid ground subsidence. Notably, this ground subsidence exhibits no significant correlation with NDVI and MNDWI. This lack of correlation suggests that the ongoing ground subsidence at Point-5 is influenced by factors other than vegetation cover and water content changes. Given the mine's artificial flooding and the presence of Dolerite formations, it is likely that the ground subsidence is driven by underlying geological and hydrological conditions specific to the mine environment, rather than by variations in NDVI and MNDWI.

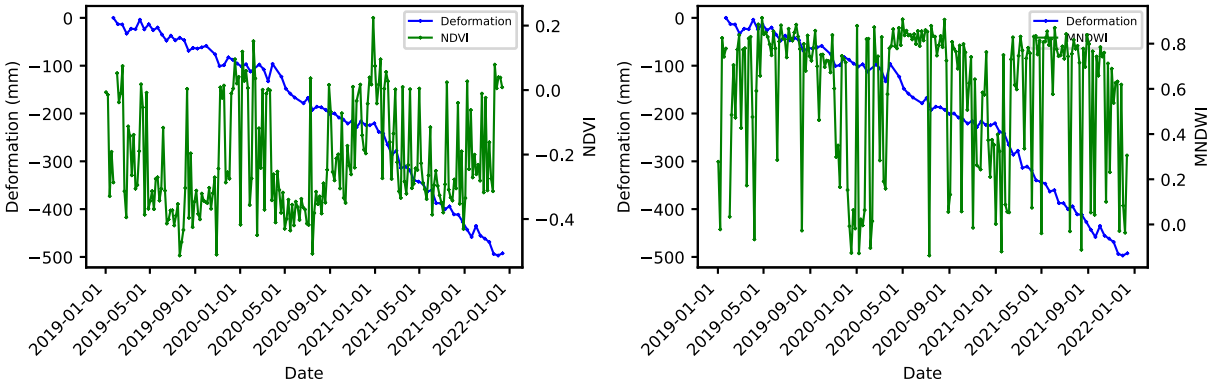


Figure 4.30: The comparison between ground subsidence and: (left) NDVI, (right) MNDWI at sample point-5.

Overall, the results reveal complex interactions between ground deformation, NDVI, and MNDWI. Short-term fluctuations in NDVI and MNDWI predominantly drive ground deformation, with rapid changes in NDVI and MNDWI often correlating with rapid ground deformation and stable conditions, respectively. However, ground subsidence at Point-5, located in an artificial flooded mine with Dolerite lithology, shows no correlation with these indices, suggesting that geological and hydrological factors play a significant role. While NDVI and MNDWI fluctuations offer insights into ground deformation dynamics, the observed inconsistencies highlight the need for further investigation into additional contributing factors.

4.4 Ground subsidence Susceptibility Mapping, and Inducing Factors in Midvaal, Gauteng

In this section, a detailed investigation and statistical analysis were conducted to determine the causes of ground subsidence and to model ground subsidence susceptibility in the Midvaal study region. Prior to the initial analysis, the ground subsidence data were georeferenced and adjusted using station HEID to ensure accuracy. The initial analysis included factor prioritisation and multicollinearity analysis. Factor prioritisation involved selecting effective ground subsidence-related factors using the Information Gain / mutual information algorithm, while multicollinearity analysis was performed to assess the interdependence of these factors using Pearson Correlation, Variance Inflation Factor, and Tolerance. This process identified and removed redundant factors that could decrease the modelling accuracy.

4.4.1 Ground Subsidence Velocity Adjustment

The ground subsidence inventory, obtained from HyP3 SBAS-InSAR ascending orbit velocity measurements, was chosen for its comprehensive coverage of the study area. The choice to select ascending orbit inventory has no significant implications as ascending and descending orbit measurements are highly correlated. The initial un-georeferenced inventory identified 1,190,550 pixels as experiencing ground subsidence. Following georeferencing with station HEID, this number was refined to 303,949 pixels, highlighting the crucial role of accurate georeferencing in correctly identifying subsiding and uplifting areas as shown in Figure 4.31. The georeferenced ground subsidence inventory was then used in the subsequent sections for detailed modelling and analysis.

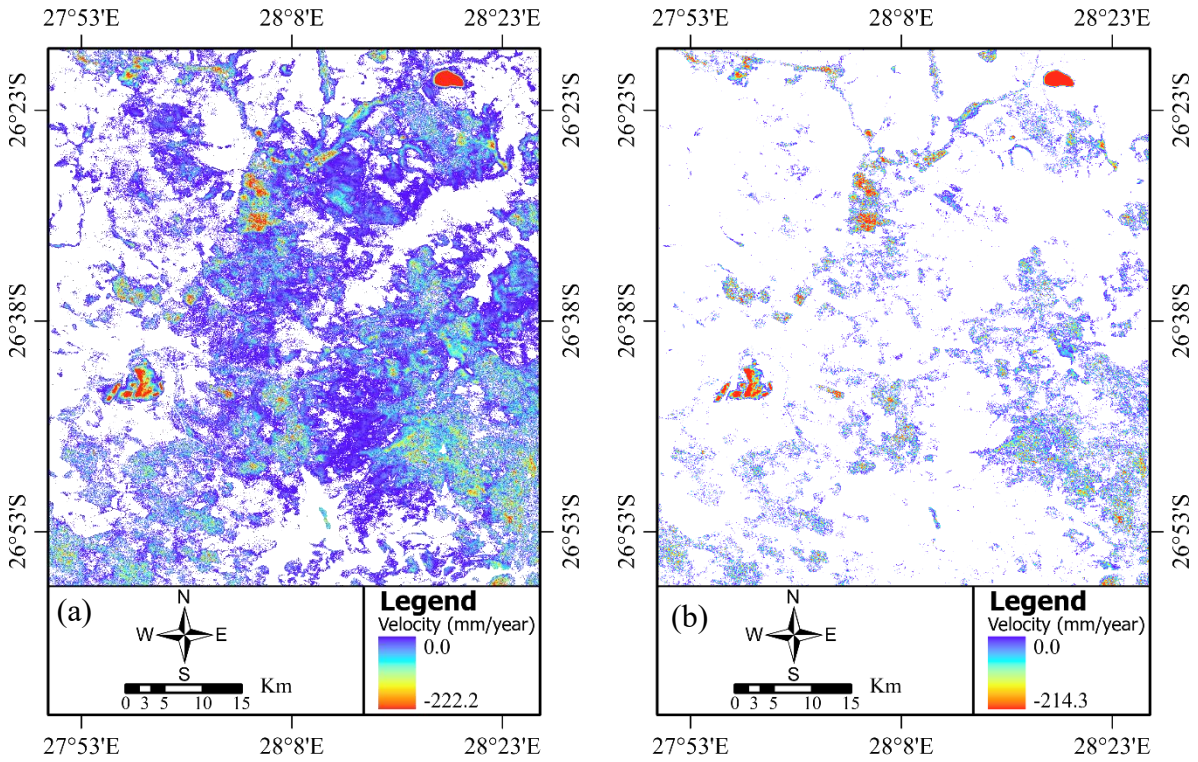


Figure 4.31: HyP3 SBAS-InSAR Ascending orbit ground subsidence inventory (a) before georeferencing, (b) after georeferencing.

4.4.2 Factor Prioritisation and Multicollinearity Analysis Before Removal of Factors

- **Information Gain Analysis**

  The relative importance of each ground subsidence related factor was assessed using information gain. Information gain computes the significance of each factor in providing optimal performance during model training. The results indicates a high correlation between ground subsidence and the South African National Land Cover 2018, followed by NDVI, SAVI, MNDWI, and NDWI. NDVI and SAVI are approximately equally correlated to ground subsidence, as are MNDWI and NDWI as shown in Figure 4.32. This is because SAVI and NDVI both quantify vegetation, and MNDWI and NDWI both quantify water content. These spectral indices are modified variations of each other, necessitating the removal of duplicates if high VIF and Pearson correlation is observed.

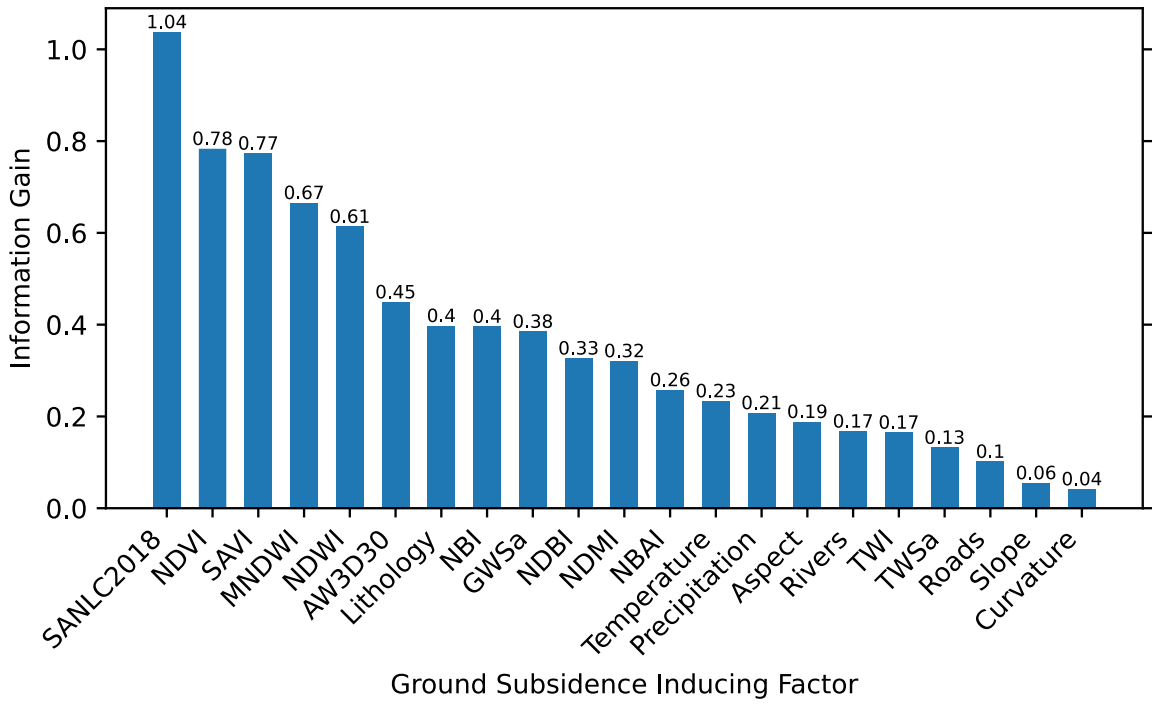


Figure 4.32: Information Gain analysis between ground subsidence and related factor.

- **VIF and Tolerance Analysis**

  Multicollinearity analysis was performed to detect and remove interdependent or redundant variables, using Variance Inflation Factor, and Tolerance. VIF values greater than 5 and Tolerance values below 0.1 indicate redundant or highly correlated factors that should be removed to avoid bias in subsequent modelling. The results indicates that NDVI, SAVI, and NDBI were highly correlated with VIF values of 1,289.2, 1,289.1, and 6.7, respectively as shown in Table 8. This was further supported by Tolerance

values for NDVI, SAVI, and NDBI of 0.0, 0.0, and 0.1, respectively, indicating high correlation.

Table 8: VIF and Tolerance analysis between ground subsidence related factors

Factor	VIF	Tolerance
Aspect	1.1	0.9
AW3D30	2.2	0.5
Curvature	1.2	0.8
GWSa	2.5	0.4
Lithology	1.2	0.8
MNDWI	3.0	0.3
NBAI	2.3	0.4
NBI	3.1	0.3
NDBI	6.7	0.1
NDMI	3.8	0.3
NDVI	1289.2	0.0
NDWI	3.0	0.3
Precipitation	2.8	0.4
Rivers	1.4	0.7
Roads	1.3	0.8
SANLC2018	1.3	0.8
SAVI	1289.1	0.0
Slope	1.3	0.8
Temperature	2.5	0.4
TWI	1.5	0.7
TWSa	2.3	0.4

- **Pearson Correlation Analysis**

  To further assess the interdependence of the factors, a multicollinearity analysis was performed using the Pearson correlation. Factors that show a high correlation indicate possible redundancy within the model, leading to bias. The results revealed a high correlation between NDVI and SAVI (1), while a moderate correlation of 0.5 was observed between MNDWI and NDWI, indicating that these two are unique to some extent as shown in Figure 4.33. Based on the information gain value and a selected Pearson threshold of 0.7, factors exceeding this threshold and having low IG values were removed. Based on our analysis of both Information Gain, VIF, Tolerance and Pearson threshold (high Pearson correlations), factors including Temperature, Terrestrial Water Storage anomaly, SAVI, NDMI, NDWI, NDBI, NBAI, Curvature, and TWI were removed from the sample of effective and uncorrelated ground subsidence related factors.

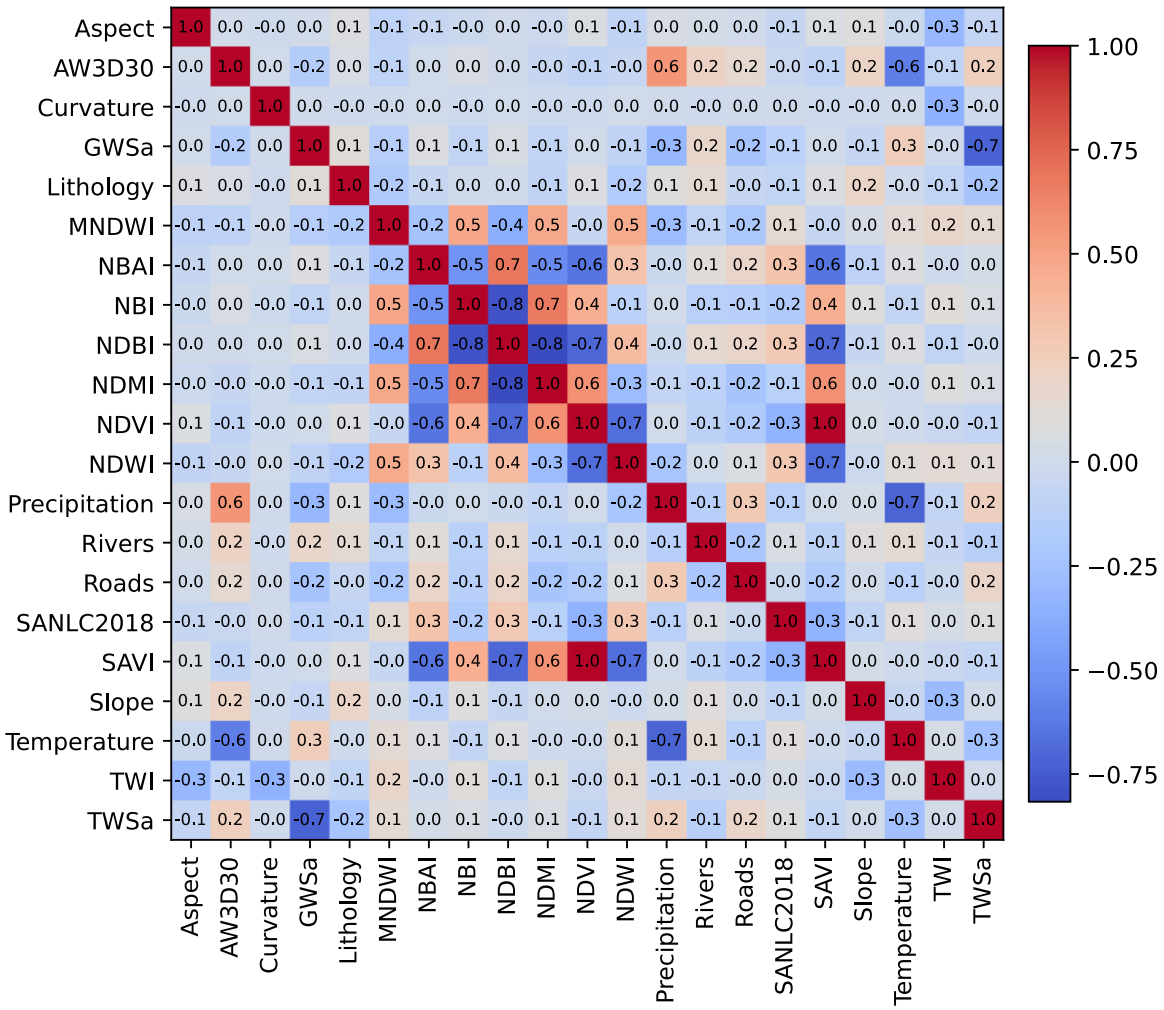


Figure 4.33: Pearson correlation analysis between ground subsidence related factors.

Subsequent to the removal of correlated factors, both factor prioritisation and multicollinearity analysis were recomputed using the remaining twelve factors. The results are discussed below.

4.4.3 Factor Prioritisation and Multicollinearity Analysis After Removal of Factors

- **Information Gain Analysis**

  After removing highly correlated factors, the information gain was recalculated using the twelve remaining factors as shown in Figure 4.34. The results indicate a high correlation between ground subsidence and Land Use and Land Cover (SANLC2018), suggesting that specific land uses exacerbate ground subsidence. The second highest correlation is observed between ground subsidence and NDVI, indicating that certain land uses, particularly where agricultural activities are prevalent, exacerbate ground subsidence. The high correlation between ground subsidence and SANLC2018 (LULC) occurs because specific land uses such as mining and agriculture lead to ground subsidence. NDVI is the second most influential factor, indicating that agricultural activities or a lack of vegetation cover induce ground subsidence. The third influential factor is MNDWI, which indicates that changes in water dynamics, particularly the

withdrawal of underground water for agricultural activities or mining operations, induce ground subsidence. A higher presence of water content could also indirectly indicate areas prone to ground subsidence. Elevation (AW3D30) is influential because most agricultural, mining and residential areas are favoured by specific altitudes. Additionally, lithology can indirectly induce ground subsidence as specific lithological formations, such as rhyolite, favour agricultural activities within the study area, or dolomite rock, which is prone to ground subsidence. A detailed assessment of these inducing factors will be discussed in a later section using Frequency Ratio analysis.

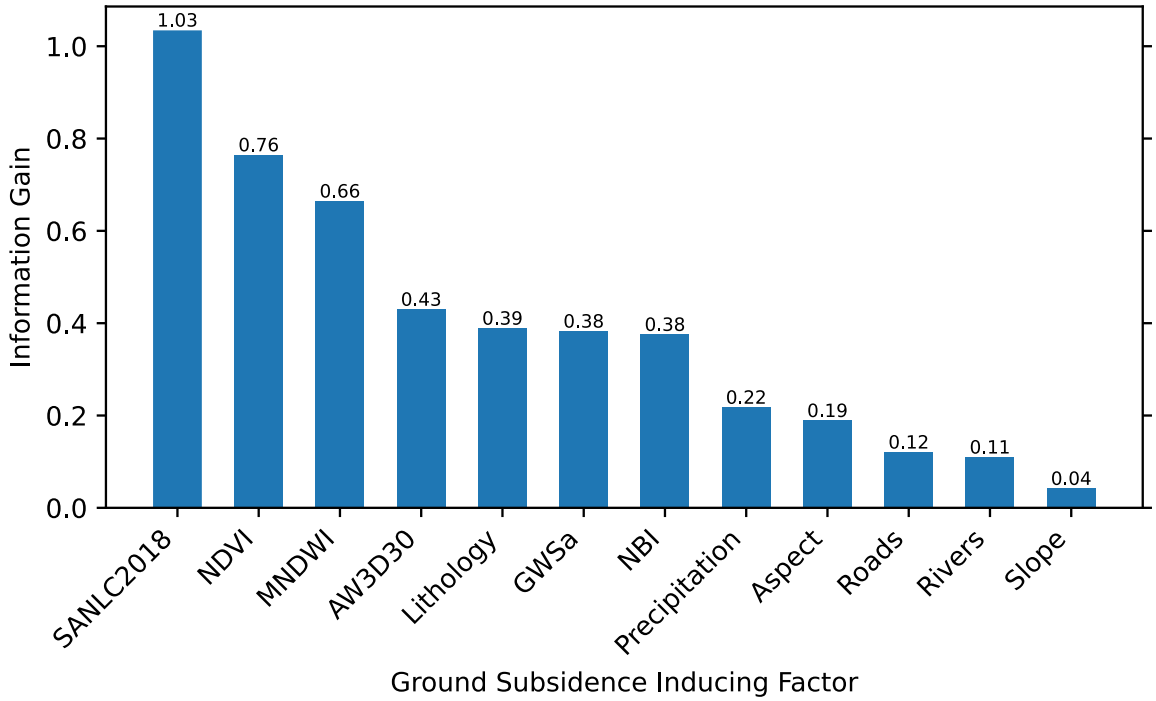


Figure 4.34: Information Gain analysis between ground subsidence and related factor.

- **VIF and Tolerance Analysis**

  To review the multicollinearity after removing redundant factors, the VIF and Tolerance were recalculated as shown in Table 9. The results indicate that all factors now have a VIF of less than 5 and Tolerance values greater than 0.1. A maximum VIF of 2.0 and a minimum Tolerance of 0.5 is observed for the precipitation variable, indicating that all these factors show a great degree of independence.

Table 9: VIF and Tolerance analysis between ground subsidence related factors

Factor	VIF	Tolerance
Aspect	1.0	1.0
AW3D30	1.8	0.6
GWSa	1.3	0.8
Lithology	1.1	0.9
MNDWI	1.9	0.5
NBI	1.9	0.5
NDVI	1.6	0.6
Precipitation	2.0	0.5
Rivers	1.3	0.8
Roads	1.3	0.8
SANLC2018	1.2	0.8
Slope	1.1	0.9

- **Pearson Correlation Analysis**

  To further validate the VIF and Tolerance multicollinearity analysis for assessing the interdependence between ground subsidence-related factors, the Pearson correlation was recalculated as shown in Figure 4.35. The results indicate that the majority of the factors have a Pearson correlation of less than 0.5. Two moderate Pearson correlations of 0.6 and 0.5 are observed between precipitation and elevation, and between NBI and MNDWI, respectively. The correlation between elevation and precipitation is attributed to higher elevation areas receiving more rainfall than lower elevations, while new built-up areas have lower water content (MNDWI) due to impervious surfaces compared to vegetated lands.

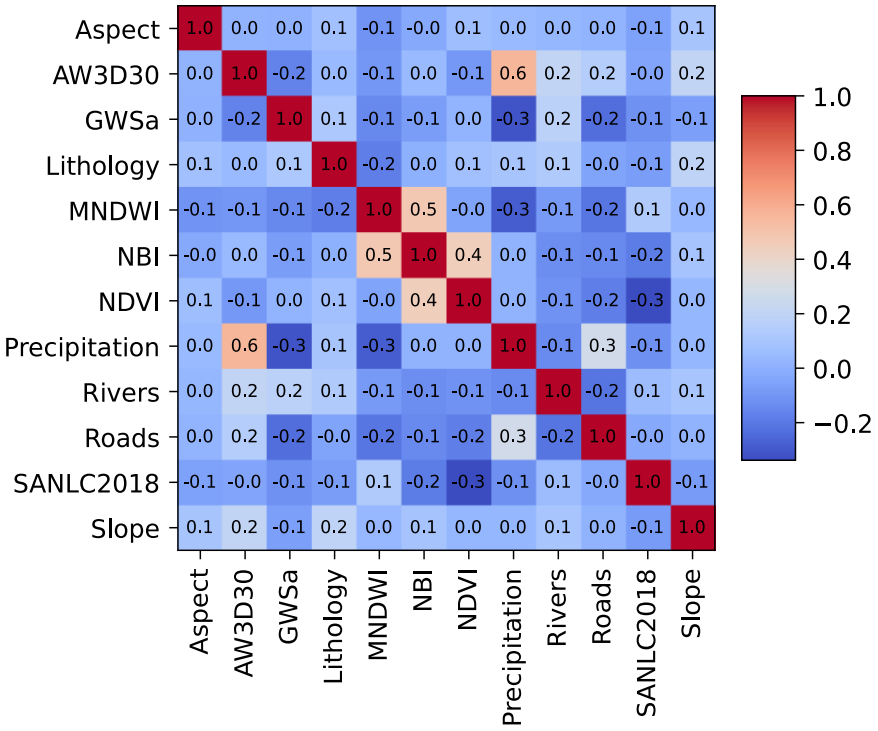


Figure 4.35: Pearson correlation analysis between ground subsidence related factors.

Having performed factor prioritisation and multicollinearity analysis, redundant variables were identified and removed, resulting in twelve remaining factors that are both effective and independent. The subsequent section discusses ground subsidence susceptibility mapping using machine learning techniques, leveraging these effective and independent factors to provide a comprehensive analysis.

4.4.4 Machine Learning Model

- **Data Preparation**

  The results from HyP3 SBAS-InSAR acquired in ascending mode, together with the twelve ground-subsidence related effective factors are used as inputs to the supervised ML algorithms namely CNN, LightGBM, Random Forest, and XGBoost to generate ground subsidence susceptibility map. Subsequently the accuracy of each model is assessed using mean absolute error, root mean square error, and coefficient of determination (R^2 score), based on the testing data for default model accuracy, and based on the training data for hyper-parameterisation accuracy.

- **Default Models Accuracy**

  The ground subsidence modelling using machine learning was conducted using supervised ML approach, and the generated model's accuracy was assessed using the mean absolute error, root mean square error, and coefficient of determination metrics between 30% testing data and corresponding model predicted values. The results shows that Random Forest algorithm had the best accuracy with R^2 score, RMSE, and MAE of 84.68%, 5.089 mm/year, and 3.651 mm/year respectively, followed by XGBoost, LightGBM, and lastly CNN.

Table 10: Default Machine Learning Model Accuracy (units: RMSE and MAE in mm/year, R^2 score in %)

Metric	CNN	LightGBM	Random Forest	XGBoost
MAE	5.054	4.302	3.651	4.064
RMSE	7.321	6.019	5.089	5.694
R^2 score	67.94	78.57	84.68	80.83

The default model's accuracy were subsequently improved by exploration of Hyperparameters using GridSearchCV algorithm implemented in scikit-learn python package, and these are discussed below.

- **Hyperparameters**
  - **CNN**

Hyperparameter optimization for the CNN model involved testing epochs (10, 20, 30), batch sizes (32, 64), kernel sizes (64, 128), and filter counts (2, 5) using GridSearchCV

as summarised in Table 11. The results revealed that increasing both the number of filters and kernel size improved accuracy. Additionally, a batch size of 64 and 20 epochs were found to be optimal.

Table 11: CNN Hyperparameters using GridSearchCV

Epochs	Batch Size	Kernel Size	Filters	Accuracy (%)
10	32	64	2	67.94
10	32	64	5	68.90
10	32	128	2	68.70
10	32	128	5	68.75
10	64	64	2	67.11
10	64	64	5	68.86
10	64	128	5	69.87
20	64	128	5	70.06
30	64	128	5	69.58
20	32	128	5	69.76

- **LightGBM**

Hyperparameter optimization for LightGBM model involved testing the number of estimators (800, 1000), learning rates (0.1, 0.2), and maximum depth (3, 5, 7, 9) using GridSearchCV as summarised in Table 12. The results reveal that increasing these parameters generally enhanced accuracy, with the best results achieved using 1000 estimators, a maximum depth of 9, and a learning rate of 0.2, resulting in an accuracy of 82.62%.

Table 12: LightGBM Hyperparameters using GridSearchCV

N estimators	Max Depth	Learning Rate	Accuracy (%)
1000	9	0.2	82.62
1000	7	0.2	81.94
800	7	0.2	81.82
1000	5	0.2	81.59
1000	7	0.1	81.52
800	5	0.2	81.40
1000	3	0.2	78.12
800	3	0.2	77.73
1000	3	0.1	76.71
800	3	0.1	76.26

- **Random Forest**

Hyperparameter optimization for Random Forest model involved testing the number of estimators (800, 1000), minimum samples split (2, 4, 5), and maximum depth (None, 5, 7, 9) using GridSearchCV as summarised in Table 13. The results reveal that increasing numbers of estimators and minimum samples splits improved accuracy, with

the maximum depth set to None yielding the highest performance. The optimal configuration was 1000 estimators, maximum depth of None, and minimum samples split of 4, achieving an accuracy of 83.92%.

Table 13: Random Forest Hyperparameters using GridSearchCV

N estimators	Max Depth	Min Samples Split	Accuracy (%)
1000	None	4	83.92
1000	None	2	83.81
1000	9	5	76.72
1000	9	2	76.72
800	9	2	76.71
1000	7	2	73.78
800	7	2	73.77
800	5	2	70.49
800	5	5	70.49
1000	5	2	70.47

- **XGBoost**

Hyperparameter optimization for XGBoost model involved testing the number of estimators (800, 1000), learning rates (0.1, 0.2), and maximum depths (2, 3, 5, 7) using GridSearchCV as summarised in Table 14. The results reveals that increasing the number of estimators and maximum depth generally improved accuracy. The best performance was obtained with 1000 estimators, a maximum depth of 7, and a learning rate of 0.1, resulting in an accuracy of 82.06%.

Table 14: XGBoost Hyperparameters using GridSearchCV

N estimators	Max Depth	Learning Rate	Accuracy (%)
1000	7	0.1	82.06
800	7	0.1	82.00
800	7	0.2	81.91
1000	7	0.2	81.80
1000	5	0.2	81.57
800	5	0.2	81.41
1000	2	0.2	78.15
800	3	0.2	77.76
1000	3	0.1	76.59
800	3	0.1	76.12

- **Final Hyper Models**

The final optimal parameters selected through GridSearchCV hyperparameter tuning are as follows. For the CNN model, the best configuration includes 20 epochs, a batch size of 64, a kernel size of 128, and 5 filters. For the LightGBM model, the optimal settings are 1000 estimators, a maximum depth of 9, and a learning rate of 0.2. The

Random Forest model performed best with 1000 estimators, a maximum depth of None, and minimum samples split of 4. Lastly, the ideal parameters for the XGBoost model are 1000 estimators, a maximum depth of 7, and a learning rate of 0.1.

4.4.5 Model Accuracy Metrics

The final hyperparameter-tuned prediction models were re-evaluated to assess their accuracy using the 30% test dataset. The results indicate that the Random Forest algorithm achieved the highest accuracy, with R^2 score, RMSE, and MAE values of 84.98%, 5.039 mm/year, and 3.621 mm/year, respectively as shown in Table 15. This was followed by XGBoost, LightGBM, and lastly CNN. All four models demonstrated enhanced accuracy compared to their default configurations, underscoring the significance of hyperparameter optimization in improving model performance.

Table 15: Final Machine Learning Model Accuracy (units: RMSE and MAE in mm/year, R^2 score in %)

Metric	CNN	LightGBM	Random Forest	XGBoost
MAE	4.796	3.839	3.621	3.806
RMSE	7.138	5.386	5.039	5.350
R^2 score	69.87	82.85	84.98	83.07

Subsequently, the accuracy of the ground subsidence susceptibility models was evaluated using ROC and AUC curve analysis as shown in Figure 4.36. The results demonstrate that Random Forest achieved the highest classification accuracy, with an AUC of 0.89, followed by XGBoost, LightGBM, and lastly CNN. These findings align with the results obtained from the R^2 score, MAE, and RMSE. Higher AUC scores, ranging between 0.8 and 0.9, indicate very good model prediction capabilities, surpassing those of models with unsatisfactory performance, which have an AUC less than 0.6.

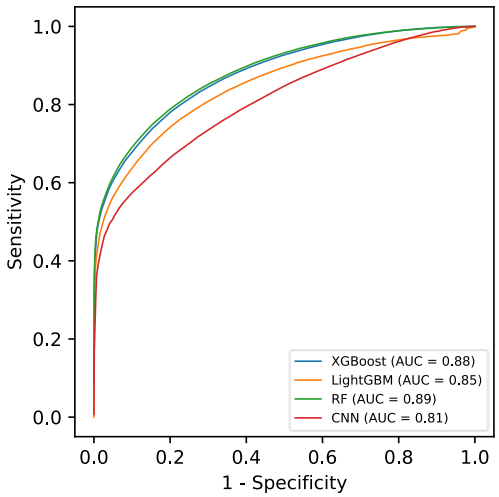
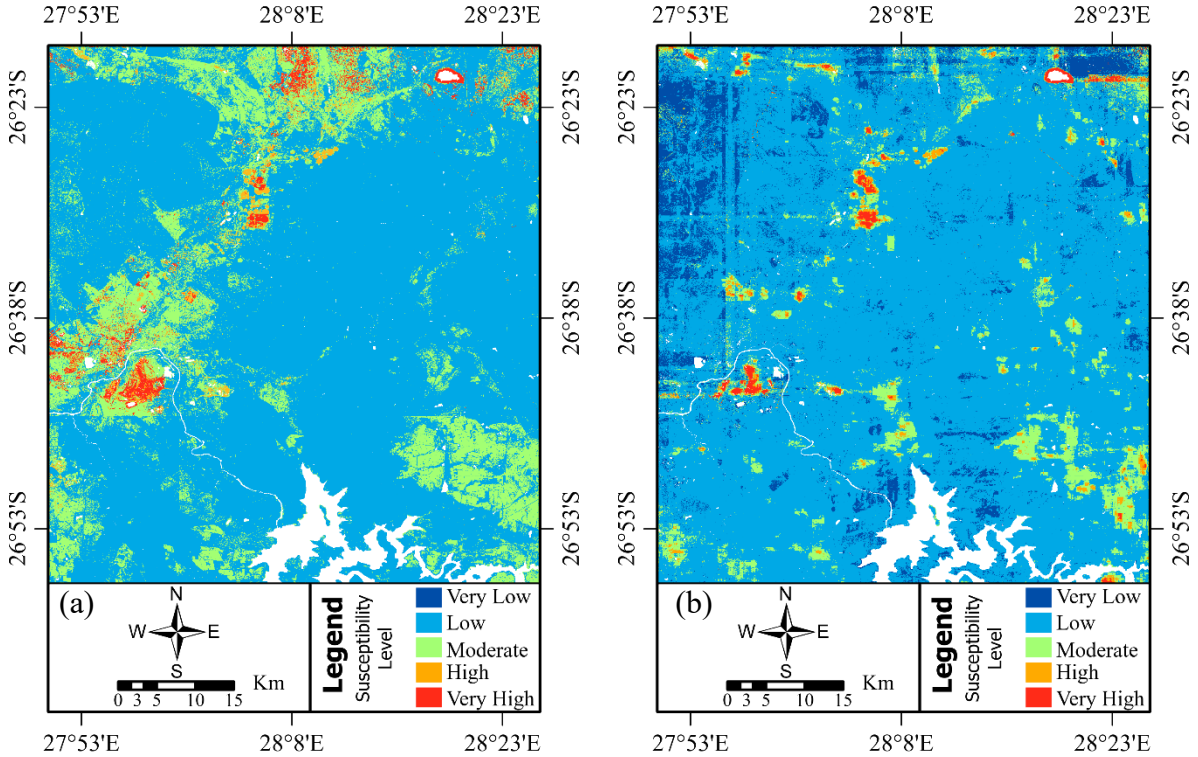


Figure 4.36: The comparison of ROC and AUC curve for each model.

Using these hyperparameter-tuned models with validated accuracy, the ground subsidence susceptibility maps (GSSM) were generated using CNN, LightGBM, Random Forest, and XGBoost, and categorised into five susceptibility classes: Very Low, Low, Moderate, High, and Very High, as shown in Figure 4.37. The results reveal that the ground subsidence susceptibility maps from the four algorithms exhibit distinct characteristics, particularly noticeable in the Moderate and Very Low classes. The CNN model identifies more Moderate classes compared to the other models, whereas both CNN and Random Forest show fewer instances of the Very Low class compared to LightGBM and XGBoost. Additionally, the results from CNN and Random Forest display smoother ground subsidence susceptibility maps, while the maps from LightGBM and XGBoost exhibit coarser patterns for the Very Low class, with visible line patterns in the northwestern part of the study region. These coarse patterns suggest that LightGBM and XGBoost may be overfitting certain features, failing to generalise the spatial pattern of ground subsidence effectively. This overfitting issue is likely due to the interlinked patterns of the Very Low and Low classes in the training dataset.



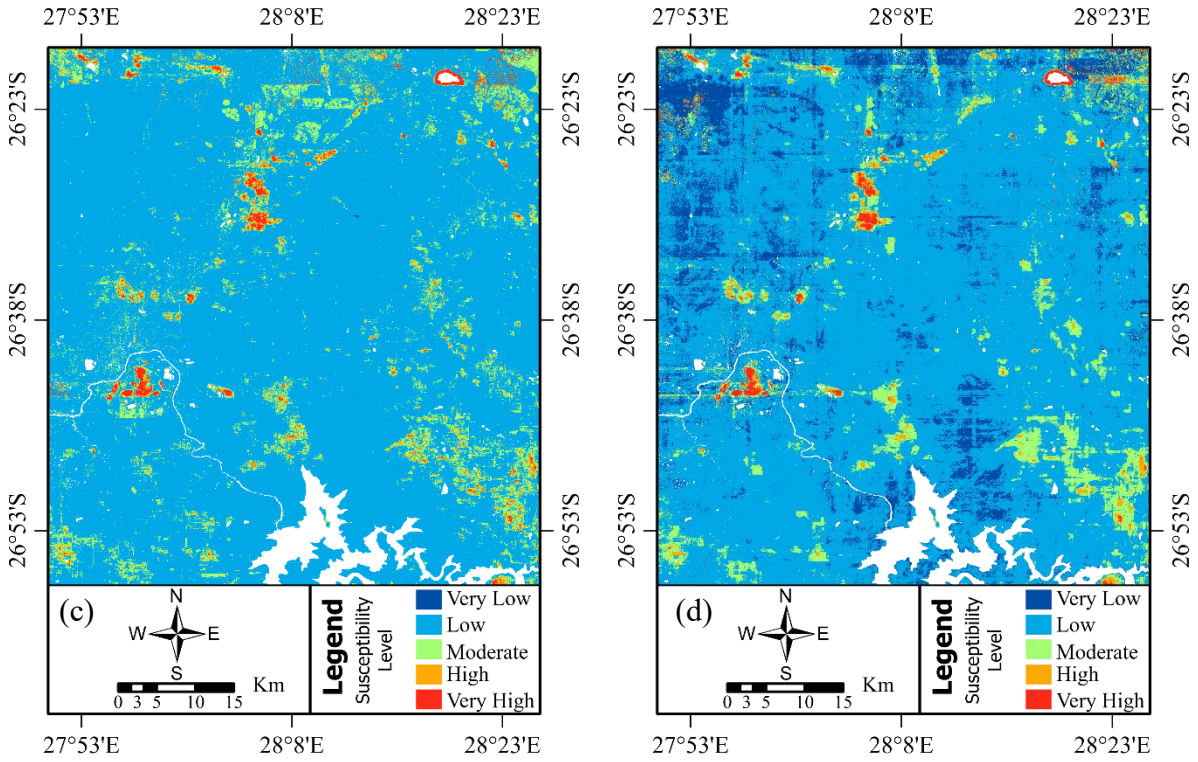


Figure 4.37: Ground subsidence susceptibility map (a) CNN, (b) LightGBM, (c) Random Forest, (d) XGBoost.

Subsequently, the percentage of each class predicted by the models is computed as shown in Figure 4.38. The CNN model predominantly predicts ground subsidence as “Low” (76.3%), with a notable proportion in the “Moderate” category (19.6%). The “Very Low” and “Very High” categories are minimally represented (0.1% and 1.8%, respectively). This distribution indicates that the CNN model tends to focus on moderate to low-risk areas, potentially underestimating high-risk zones.

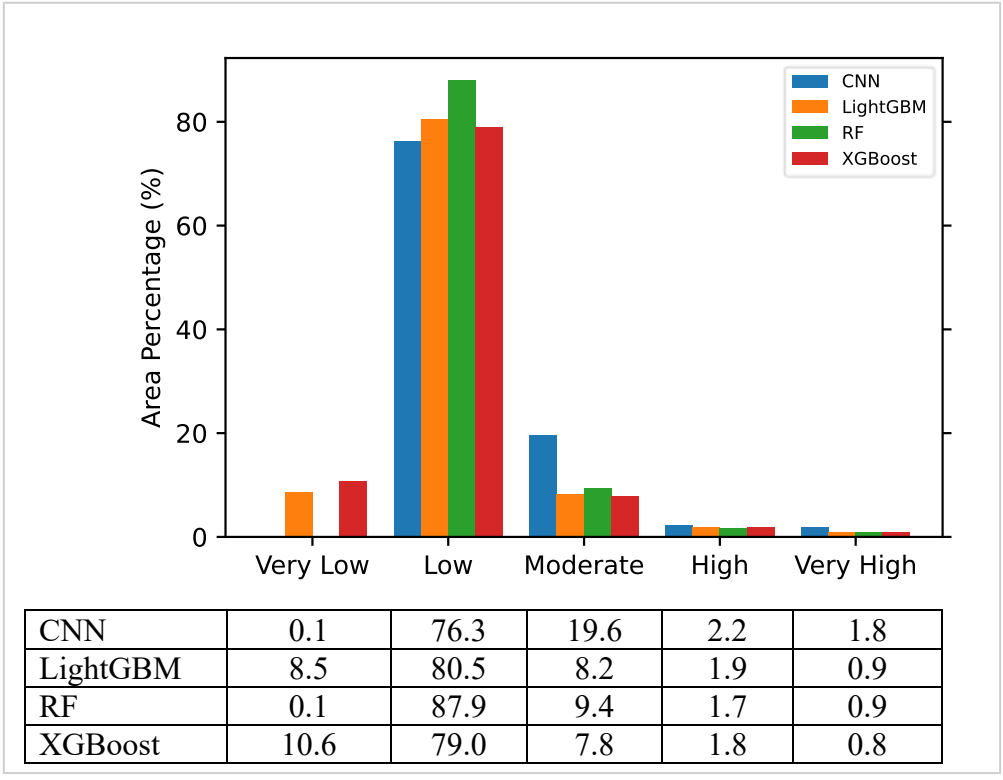
The LightGBM model displays a higher percentage in the “Very Low” category (8.5%) compared to the CNN model. It also predicts a high proportion as “Low” (80.5%), with relatively lower proportions in the “Moderate”, “High”, and “Very High” categories. This suggests that the LightGBM model is inclined to classify more areas as “Very Low” compared to other models.

The RF model demonstrates the highest accuracy in classifying areas as “Low” (87.9%) and a very low percentage in the “Very High” category (0.9%). The distribution is notably skewed towards lower susceptibility, suggesting that the RF model tends to predict more conservative ground subsidence susceptibility levels.

The XGBoost model predicts a significant proportion of “Very Low” susceptibility (10.6%), which is higher than both CNN and RF. It also classifies a substantial portion as “Low” (79.0%), with smaller percentages in the “Moderate”, “High”, and “Very High” categories. This suggests that XGBoost may be more inclined to classify areas as “Very Low” compared to CNN and RF.

Furthermore, while all models predominantly classify areas as having “Low” susceptibility, they exhibit varying tendencies in predicting “Very Low” and “Very High” categories. The CNN and RF models are conservative, with minimal predictions in the “Very High” category, whereas LightGBM and XGBoost offer a broader range of classifications, with LightGBM predicting a higher proportion of “Very Low” areas. Each model has its strengths and may be chosen based on the specific requirements of ground subsidence susceptibility mapping, considering the trade-offs between sensitivity and specificity.

Overall, the results indicate that the largest area in the study region is characterised by “Low” ground subsidence susceptibility. The next most prominent class is “Moderate”. Combined, the “Low” and “Moderate” classes account for at least 86.8% of the area across all models. This suggests that the majority of the ground in the study region does not pose a significant threat to the environment. However, the combined occurrence of “High” and “Very High” susceptibility covers at least 2.6% of the area, equivalent to approximately 107.28 square kilometres about the size of Midrand, Gauteng, South Africa (152.87 square kilometres). This highlights a potentially alarming impact, given the significant proportion of high-risk areas.



CNN	0.1	76.3	19.6	2.2	1.8
LightGBM	8.5	80.5	8.2	1.9	0.9
RF	0.1	87.9	9.4	1.7	0.9
XGBoost	10.6	79.0	7.8	1.8	0.8

Figure 4.38: Susceptibility class area percentage for CNN, LightGBM, Random Forest, and XGBoost models.

4.4.6 Model Explainability

The SHapley Additive exPlanations values revealed important insights into the contribution of various features to the predictions of ground subsidence susceptibility across different models, namely CNN, LightGBM, Random Forest, and XGBoost as shown in Figure 4.39.

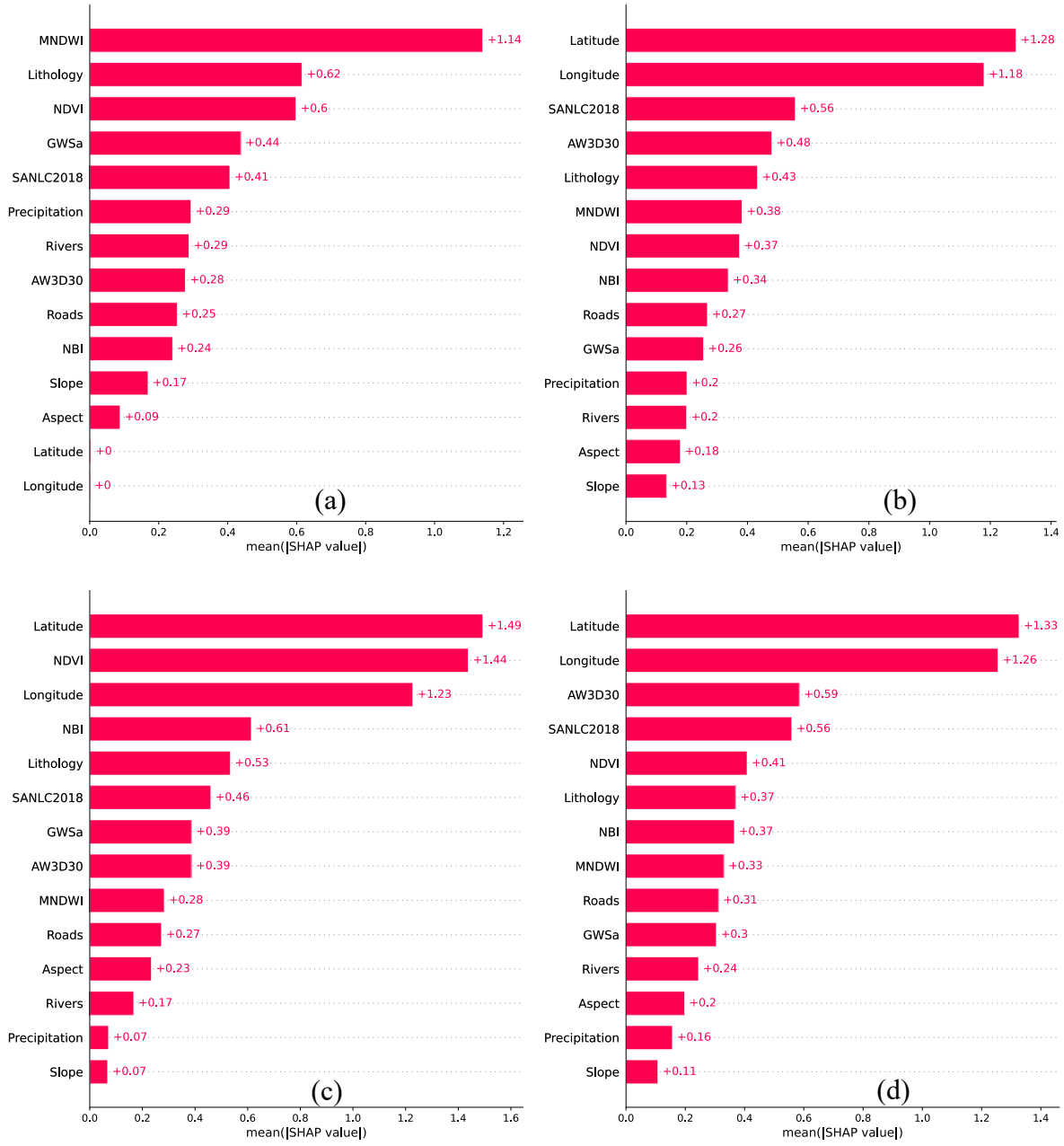


Figure 4.39: Feature importance using SHAP values, (a) CNN, (b) LightGBM, (c) Random Forest, (d) XGBoost.

The results reveal that MNDWI shows the highest mean absolute SHAP value for the CNN model, indicating that CNN relies heavily on this feature for its predictions. In contrast, LightGBM, Random Forest, and XGBoost place less emphasis on MNDWI, suggesting that CNN is more effective in capturing the influence of water related features. Lithology is consistently valued across all models, with CNN giving it the highest importance, underscoring

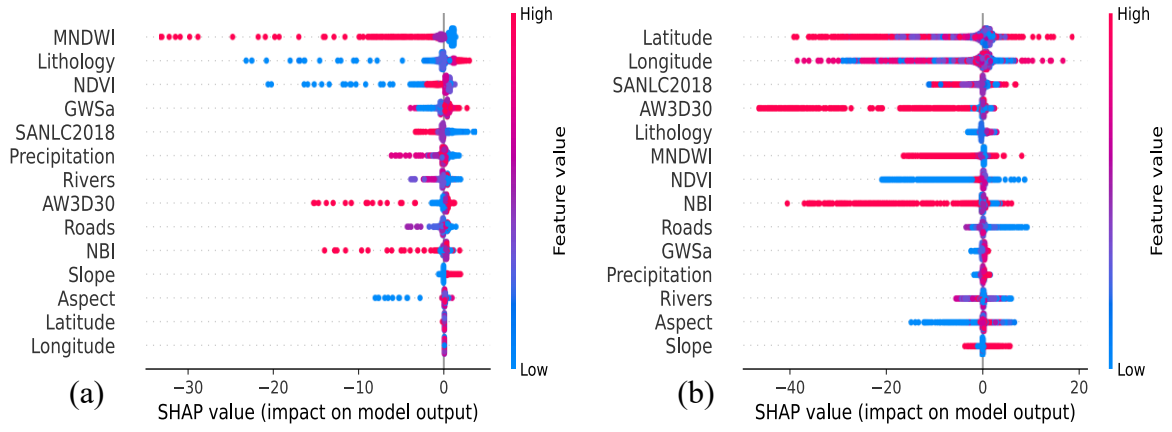
its crucial role in predicting ground subsidence susceptibility. The NDVI is most significant in Random Forest, which highlights its impact on ground subsidence predictions, although CNN, LightGBM, and XGBoost also consider it important to varying degrees.

Ground Water Storage anomaly holds more importance for CNN compared to the other models, reflecting CNN's greater sensitivity to this feature. Land Use and Land Cover (SANLC2018) shows consistent importance in LightGBM and XGBoost, indicating its significant role in predicting ground subsidence, whereas CNN values it less. Precipitation has relatively low SHAP values across all models, suggesting it has a minor impact on ground subsidence predictions. Proximity to Rivers also shows limited importance, though CNN places slightly more emphasis on it.

Elevation's influence is highest in XGBoost, indicating that this feature is more critical for XGBoost predictions, while CNN assigns it a lower importance. Proximity to Roads has moderate importance across all models, with XGBoost valuing it slightly more. The NBI is most significant in Random Forest, emphasizing its role in ground subsidence predictions within this model. Slope has the lowest SHAP values, reflecting its minimal impact on predictions across all models. Aspect shows higher SHAP values in Random Forest, suggesting a relatively greater role in its predictions, while CNN considers it the least important.

Latitude and Longitude are highly valued in LightGBM, Random Forest, and XGBoost, indicating their crucial role in these model's predictions, whereas CNN assigns minimal importance to these features. This disparity underscores the varying emphasis placed on geographic coordinates across different models. Overall, the SHAP analysis provides a comprehensive understanding of how different features contribute to ground subsidence susceptibility predictions, highlighting the nuances and varying importance of each feature across different model types.

Further analysis on the feature importance using SHAP values was conducted using the beeswarm plots, these plots not only details the feature importance, but also highlight the distribution and patterns of the point features as shown in Figure 4.40.



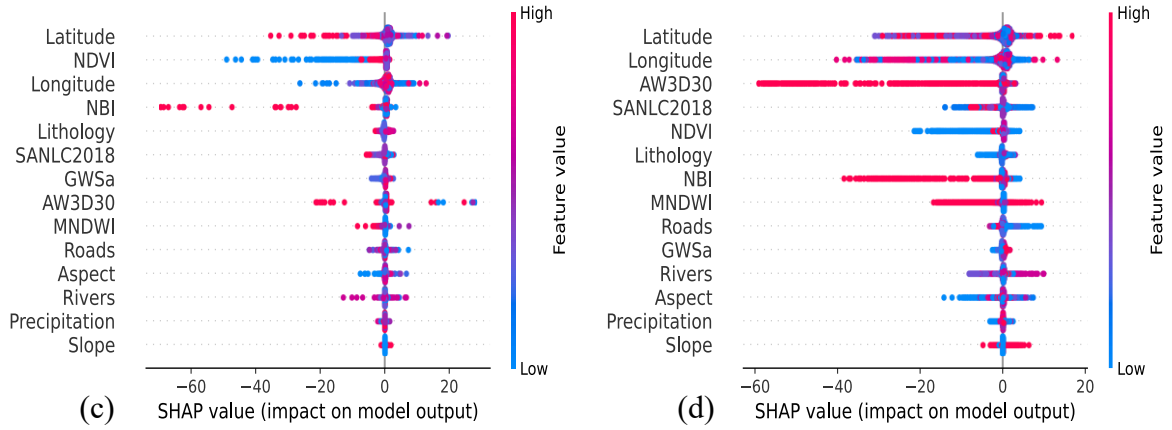


Figure 4.40: Feature and model sensitivity using SHAP values, (a) CNN, (b) LightGBM, (c) Random Forest, (d) XGBoost.

The beeswarm plot results reveals distinct patterns in how various features influence ground subsidence velocities across the different models, including CNN, LightGBM, Random Forest, and XGBoost. Starting with latitude, the impact on ground subsidence demonstrates a complex relationship. As the SHAP values become more negative, indicating lower ground subsidence, the feature value gradually shifts from purple to red. This observation suggests that ground subsidence tends to increase with latitude magnitude, particularly in regions where the latitude is higher, as indicated by the red shading. The presence of a cluster of purple points at a SHAP value of zero highlights insignificant ground subsidence in areas with intermediate latitude values.

The influence of longitude on ground subsidence introduces a nuanced pattern. The negative SHAP values, indicative of less ground subsidence, are more pronounced within specific longitude ranges. The transitions between purple, blue and red points reveal geographical variability in ground subsidence, suggesting that certain longitudes are more susceptible to ground subsidence than others.

Moving to the NDVI, the relationship becomes clearer. Large negative SHAP values, which reflect lower ground subsidence, are associated with lower NDVI values, shown in blue. On the other hand, positive SHAP values, indicating high ground subsidence, correspond to higher NDVI values, depicted in red. This pattern suggests that regions with higher vegetation cover are more susceptible to ground subsidence, whereas areas with lower NDVI are generally less prone to such effects. The higher vegetation cover may indirectly influence ground subsidence through groundwater extraction for crop irrigation, which could lead to soil sediment consolidation.

The MNDWI, as analysed using the CNN model, shows that higher ground subsidence is associated with low MNDWI values, represented by blue points. This suggests that areas with lower water content are more susceptible to ground subsidence, often due to agricultural and mining activities. In agricultural regions, extensive groundwater extraction for crop irrigation

decreases soil moisture and lowers the water table, leading to compaction and ground subsidence. Conversely, regions with high MNDWI values, indicating wet conditions, are generally more stable and less prone to ground subsidence.

Next, the NBI displays a clear correlation. Regions with higher NBI values, represented in red, are associated with higher negative SHAP values, indicating lower ground subsidence. This relationship suggests that areas with a high concentration of buildings, and thus more stable scatterers, experience less ground subsidence. In contrast, areas with low NBI values exhibit increased ground subsidence. The absence of points in some ranges suggests that the impact of NBI on ground subsidence is not uniform and may depend on particular thresholds or conditions.

The analysis of lithology underscores its significant influence on ground subsidence. Higher negative SHAP values are linked to specific lithological types, with a noticeable shift in colour from red to blue. This pattern indicates that certain geological formations are more prone to ground subsidence, emphasising the critical role of geological composition in determining ground subsidence susceptibility.

Finally, the analysis of elevation consistently shows that negative SHAP values, indicating lower ground subsidence, are associated with high elevations, represented in red. Nonetheless, both low and some high elevation values are associated with ground subsidence values highlighting the complexity of elevation impact on ground subsidence. The variation in dominant colours between CNN, LightGBM, Random Forest and XGBoost models highlights differences in how these models interpret the influence of elevation on ground subsidence.

Overall, the results indicate that ground subsidence is influenced by a combination of environmental and geographical factors. Low NDVI, specific lithological types, and high elevations are generally associated with lower ground subsidence, underscoring the complex interplay between natural features and ground subsidence risk.

4.4.7 Frequency Ratio Analysis

To further gain insight into the spatial relationship between ground subsidence and each ground subsidence related factor, a frequency ratio analysis was performed. A frequency ratio value greater than 1 suggests that the class has an effect on ground subsidence occurrence, while a frequency ratio less than 1 indicates that the class is not correlated with ground subsidence occurrence.

The frequency ratio analysis, using twelve effective and uncorrelated factors alongside the corrected HyP3 SBAS-InSAR ascending orbit ground subsidence inventory (adjusted using station HEID), provides important insights into the relationship between these factors and ground subsidence. The results reveals that ground subsidence is significantly associated with lower and mid-elevation zones, particularly in areas with elevations between 1,493 – 1,544

metres, 1,544 – 1,604 metres, and 1,422 – 1,493 metres, where the FR values are 1.3, 1.0, and 1.0, respectively. The slightly higher FR value at the lowest elevation suggests that geological, hydrological, and agricultural processes, more active in these zones, contribute to the observed ground subsidence patterns.

This tendency is further supported by the findings related to aspect and slope. Ground subsidence occurs across various slope aspects, with north, northwest, and northeast-facing slopes showing an FR value of 1.1, compared to 1.0 for other aspects. These slopes may experience differences in solar radiation and vegetation cover, which affect soil moisture and water dynamics, potentially making them more prone to ground subsidence. Additionally, flatter areas with slopes between 0 – 2 degrees exhibit an FR value of 1.3, indicating a higher susceptibility to ground subsidence. The accumulation of water and the impact of human activities, such as irrigation and mining, likely contribute to this increased risk.

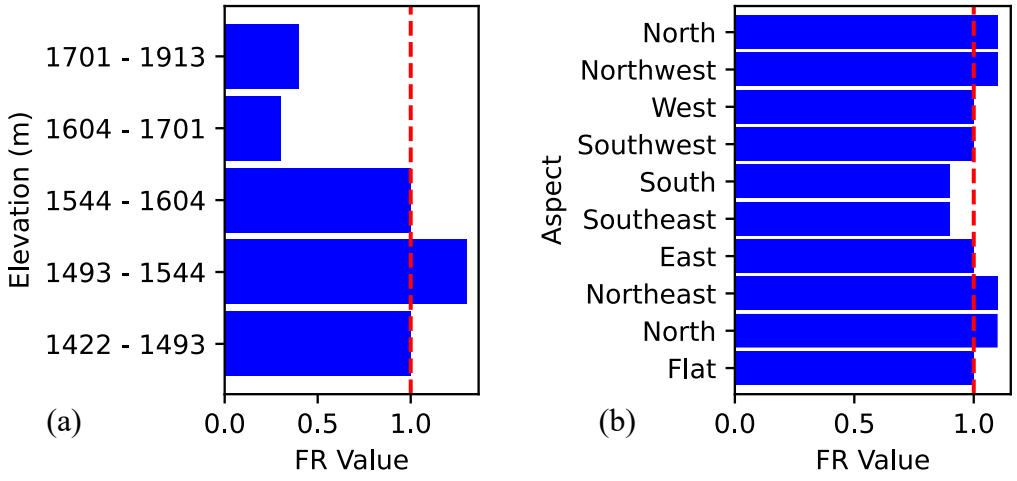
Water related factors also play a significant role in ground subsidence patterns. The relationship between mean ground water storage anomaly, precipitation, and ground subsidence reveals those areas with GWS anomalies between -0.032 – -0.029 mm/year and -0.029 – -0.016 mm/year show FR values of 1.5 and 1.1, respectively. This strong correlation suggests that ground subsidence is closely linked to reductions in groundwater storage, likely due to excessive water extraction. Similarly, regions with mean annual precipitation levels between 807 – 826 mm/year, 793 – 807 mm/year, and 757 – 772 mm/year have FR values of 1.8, 1.3, and 1.1, respectively, indicating that increased rainfall may enhance soil moisture and groundwater recharge processes, exacerbating ground subsidence risks. Furthermore, areas closer to rivers, with distances between 0 – 1,152 metres and 1,152 – 2,495 metres, exhibit FR values of 1.2 and 1.1, respectively, suggesting that proximity to rivers influences ground subsidence due to riverine effects on groundwater levels and soil stability.

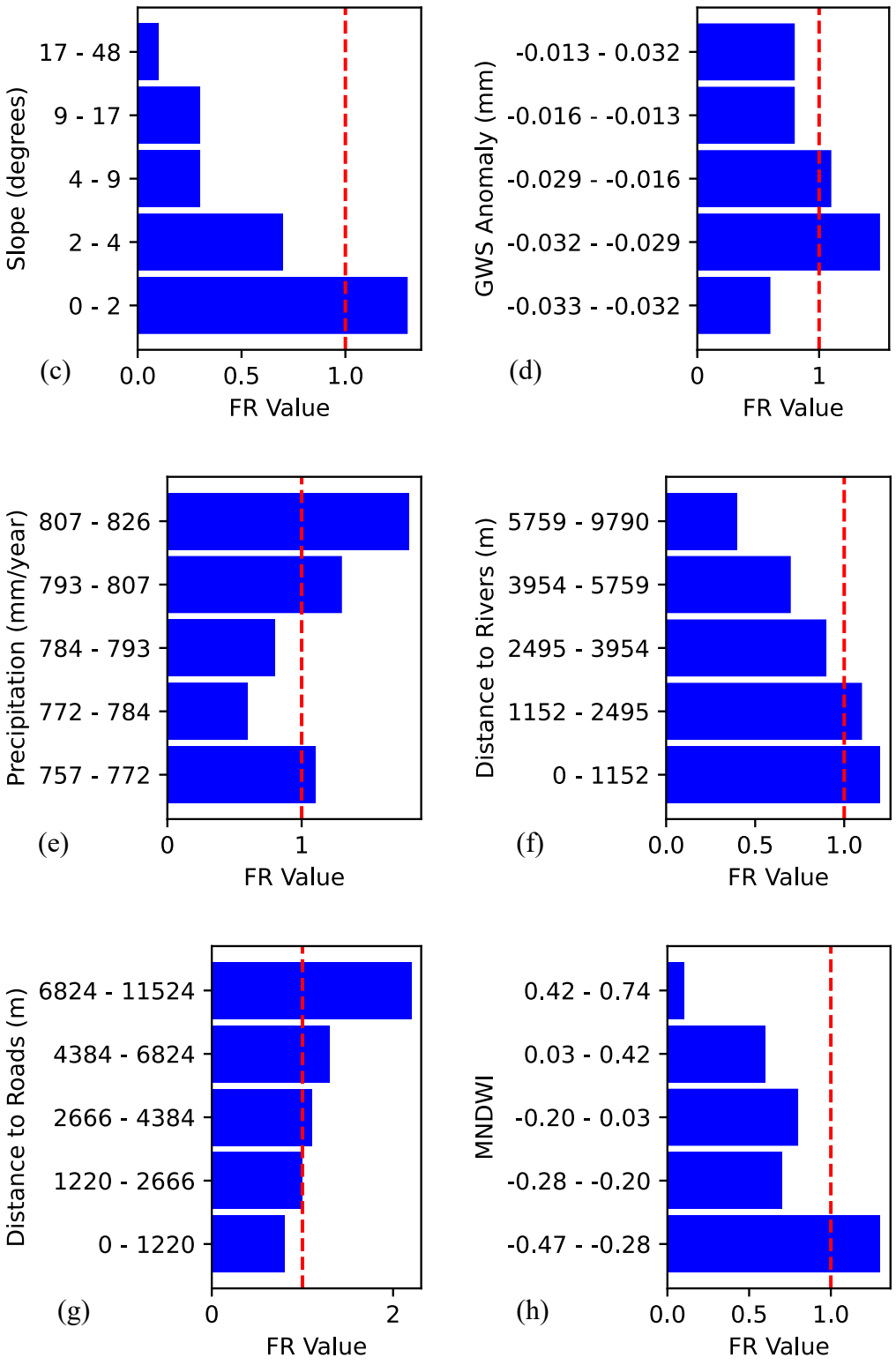
The integration of the MNDWI, NDVI, and NBI findings reveals a complex interaction between vegetation, water content, and land cover. Ground subsidence is more common in regions with MNDWI values between -0.47 – -0.28, reflected in an FR value of 1.3, which may be influenced by local hydrological conditions. Concurrently, areas with NDVI values between 0.29 – 0.58, 0.22 – 0.29, and 0.13 – 0.22 have FR values of 1.9, 1.0, and 1.0, respectively, indicating that denser vegetation is associated with ground subsidence. Moreover, regions with lower NBI values, specifically between -2.58 – -0.20 and -0.10 – 0.35, exhibit FR values of 1.6 and 1.0, respectively. This pattern suggests that ground subsidence may be driven by the interaction between vegetation and soil moisture, possibly due to groundwater extraction for crop irrigation. The low NBI values, reflecting less built-up land cover, indicate that areas with less urbanisation and more agricultural land use are more vulnerable to ground subsidence, highlighting the significant impact of human activities on ground subsidence patterns.

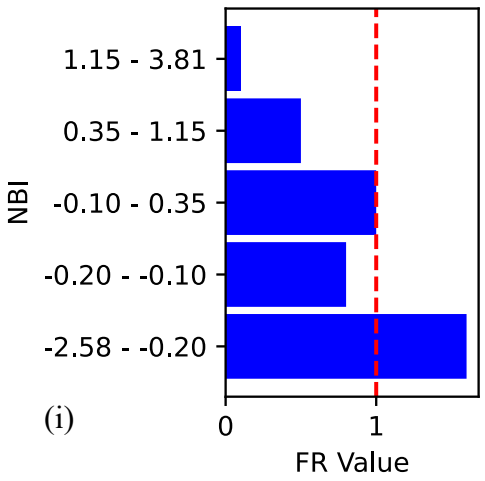
Lithology also plays a crucial role in ground subsidence, as certain rock types are more associated with this phenomenon. Areas with lithologies such as 'Rhyolite with subordinate pyroclastic rocks', 'Light grey, equigranular, unfoliated biotite granite', and 'Unfoliated, leucocratic and mesocratic granites' exhibit FR values of 1.7, 1.5, and 1.4, respectively. Other lithologies like 'Dolomite, subordinate chert, limestone and quartzite', 'Dolerite sills, sheets and dykes', and 'Quartzite, subordinate conglomerate and shale' have FR values of 1.1 each. These results indicate that the geological composition of an area contributes to varying levels of ground subsidence risk, with different rock types exhibiting different susceptibilities to ground subsidence.

The land use and land cover analysis further consolidates these findings by highlighting the significant impact of intensive agricultural and mining activities on ground subsidence. Areas characterised by LULC categories such as "commercial annual crops non-pivot irrigated", "commercial annual crops pivot irrigated", and "artificial flooded mine pits" show the highest FR values of 5.8, 4.8, and 3.7, respectively. Other categories like "subsistence/small-scale annual crops", "mine: tailings and resource dumps", and "mine extraction pits, quarries" have FR values of 3.1 each, suggesting that these activities significantly contribute to ground subsidence due to their effects on groundwater extraction and soil disturbance. These findings are consistent with the observed patterns of high NDVI and low MNDWI values, where groundwater extraction and soil disturbance from agricultural and mining activities exacerbate ground subsidence risks.

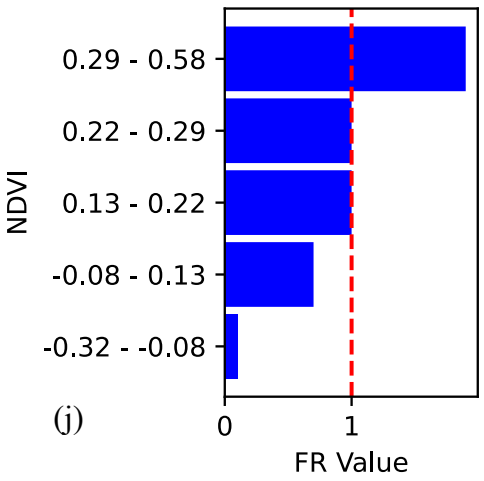
Overall, the frequency ratio analysis provides a comprehensive understanding of the factors influencing ground subsidence. The interconnectedness of elevation, aspect, slope, water related factors, lithology, and land use underscores the complexity of ground subsidence processes. This analysis highlights the need for integrated management strategies to mitigate ground subsidence risks in vulnerable areas, particularly those heavily influenced by human activities.



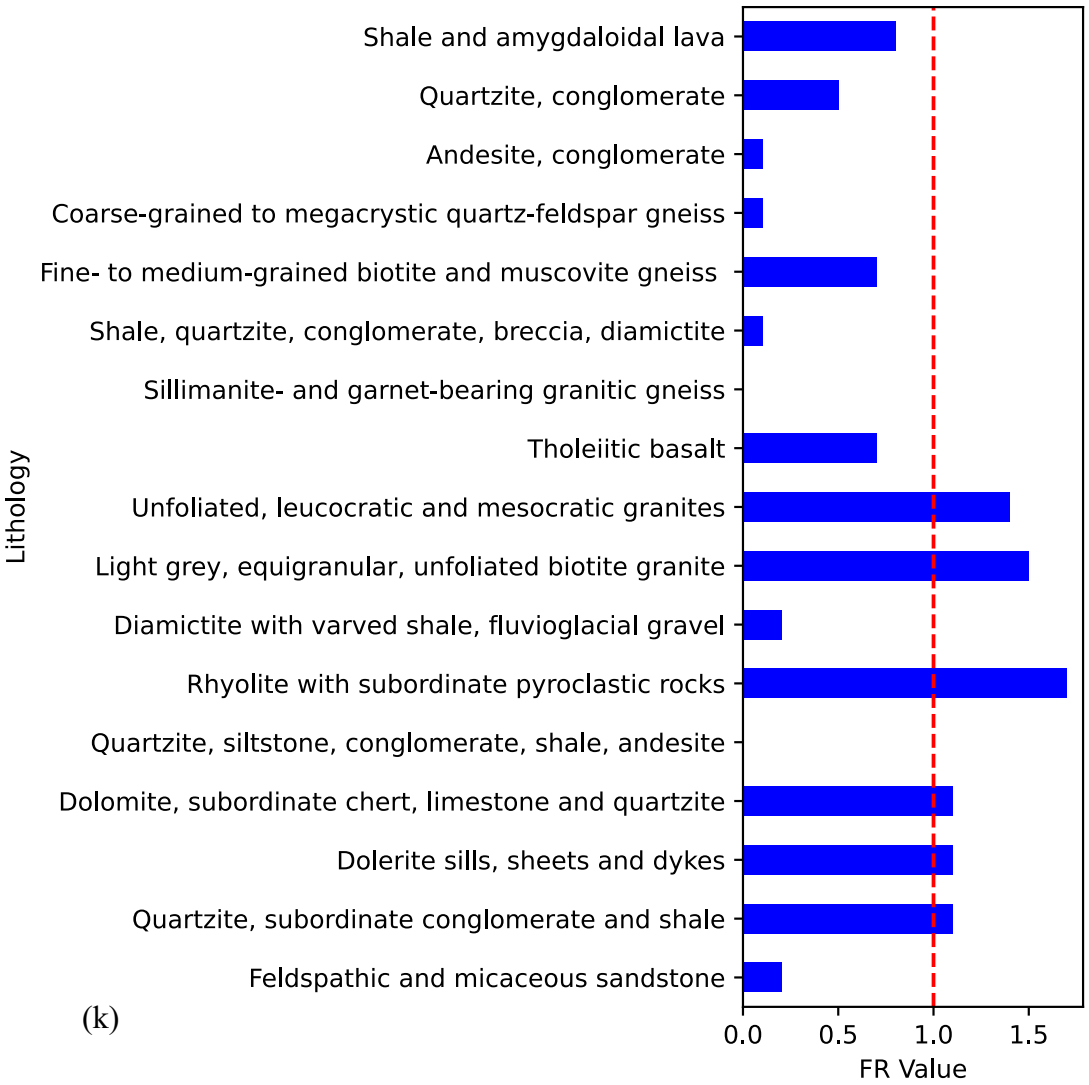

Slope (degrees)
17 - 48
9 - 17
4 - 9
2 - 4
0 - 2
0.0
0.5
1.0
FR Value
(c)
GWS Anomaly (mm)
-0.013 - 0.032
-0.016 - -0.013
-0.029 - -0.016
-0.032 - -0.029
-0.033 - -0.032
0
1
FR Value
(d)
Precipitation (mm/year)
807 - 826
793 - 807
784 - 793
772 - 784
757 - 772
0
1
FR Value
(e)
Distance to Rivers (m)
5759 - 9790
3954 - 5759
2495 - 3954
1152 - 2495
0 - 1152
0.0
0.5
1.0
FR Value
(f)
Distance to Roads (m)
6824 - 11524
4384 - 6824
2666 - 4384
1220 - 2666
0 - 1220
0
2
FR Value
(g)
MNDWI
0.42 - 0.74
0.03 - 0.42
-0.20 - 0.03
-0.28 - -0.20
-0.47 - -0.28
0.0
0.5
1.0
FR Value
(h)


NBI
1.15 - 3.81
0.35 - 1.15
-0.10 - 0.35
-0.20 - -0.10
-2.58 - -0.20
0
1
FR Value

(i)


NDVI
0.29 - 0.58
0.22 - 0.29
0.13 - 0.22
-0.08 - 0.13
-0.32 - -0.08
0
1
FR Value

(j)


Lithology
Shale and amygdaloidal lava
Quartzite, conglomerate
Andesite, conglomerate
Coarse-grained to megacrystic quartz-feldspar gneiss
Fine- to medium-grained biotite and muscovite gneiss
Shale, quartzite, conglomerate, breccia, diamictite
Sillimanite- and garnet-bearing granitic gneiss
Tholeiitic basalt
Unfoliated, leucocratic and mesocratic granites
Light grey, equigranular, unfoliated biotite granite
Diamictite with varved shale, fluvioglacial gravel
Rhyolite with subordinate pyroclastic rocks
Quartzite, siltstone, conglomerate, shale, andesite
Dolomite, subordinate chert, limestone and quartzite
Dolerite sills, sheets and dykes
Quartzite, subordinate conglomerate and shale
Feldspathic and micaceous sandstone
0.0
0.5
1.0
1.5
FR Value

(k)

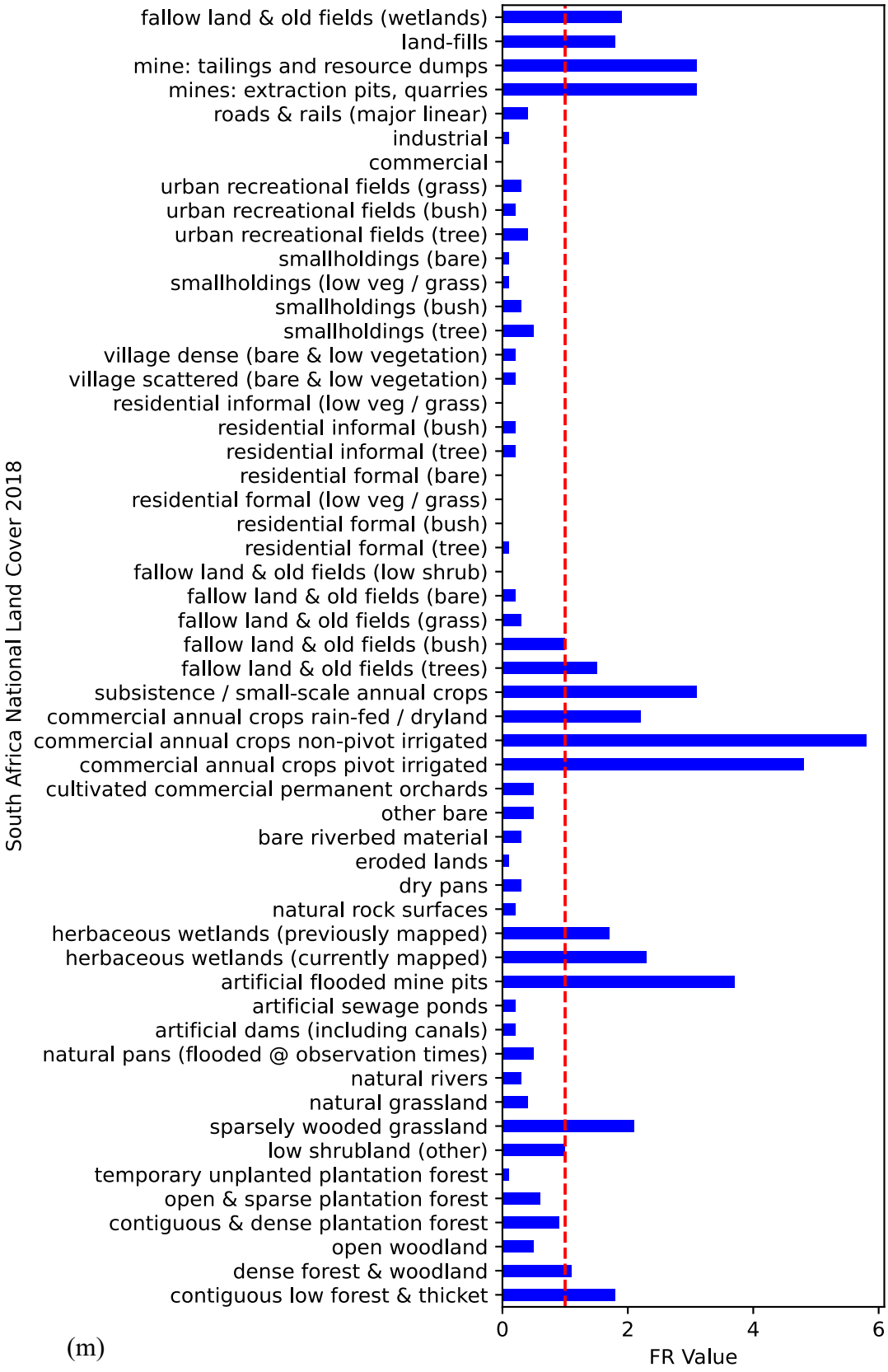


Figure 4.41: Frequency ratio analysis between ground subsidence and: (a) Elevation, (b) Aspect, (c) Slope, (d) GWS Anomaly, (e) Precipitation, (f) Distance to Rivers, (g) Distance to roads, (h) MNDWI, (i) NBI, (j) NDVI, (k) Lithology, (m) Land use and Land Cover.

5 Conclusion and Recommendations

This research successfully developed and applied an InSAR time series analysis framework integrated with machine learning to monitor ground subsidence and produce susceptibility maps in Midvaal, South Africa. The findings directly addressed the research objectives and provided significant contributions to ground deformation studies, emphasising the integration of geospatial techniques with predictive modelling for effective risk assessment.

5.1 Effects of Pre-Processing and Processing Software, and Time Series Analysis Method

The investigation into InSAR time series analysis techniques and open-source processing tools revealed SNAP, ISCE, and HyP3 as prevalent open-source InSAR preprocessing tools. While MintPy and StaMPS emerged as the dominant choices for InSAR time series analysis, based on the literature review. Both PS and SBAS InSAR approaches are well-supported by these open-source tools. However, the lack of automation in SNAP led to the development of StoSAP, which facilitates semi-automated PS and SBAS InSAR preprocessing.

Using these tools and methodologies, an assessment of the impact of InSAR pre-processing and InSAR time series analysis processing software selection on the accuracy of ground subsidence monitoring was conducted. The investigation revealed notable differences in the performance of ISCE, SNAP, HyP3, StaMPS, and MintPy. ISCE PS-InSAR demonstrated the highest precision, characterised by low standard errors and a narrow velocity range, making it the most suitable method for monitoring tectonic deformation. HyP3 SBAS-InSAR excelled in capturing localised ground deformations with the broadest velocity range and lower relative uncertainty, although it exhibited higher standard errors and variability. Meanwhile, SNAP PS-InSAR showed promise for identifying smaller ground deformation patterns but was more prone to errors, as indicated by anomalous velocity distributions. The findings highlight the trade-offs between precision, robustness, and the ability to detect localised ground deformations, underscoring the importance of selecting tools and methods suited to the specific ground deformation context.

5.2 Effects of Digital Elevation Models

The critical role of Digital Elevation Models in ground subsidence susceptibility mapping was highlighted through comparative analyses of AW3D30, COP30, SRTMGL1, and ASTER DEMs. AW3D30 consistently outperformed others in vertical accuracy and velocity modelling, proving particularly effective in non-complex terrains. While COP30 and SRTMGL1 performed similarly, ASTER produced the poorest results, particularly in areas with dense vegetation, urban infrastructure, or mining activities. These results emphasise that high-quality DEMs like AW3D30 enhance the accuracy of ground subsidence assessments. However, the limited improvement in elevation errors post-coregistration suggests that DEM accuracy alone cannot comprehensively address ground subsidence complexities. Future models must integrate additional geospatial and geotechnical factors to better capture the multifaceted nature of ground deformation dynamics.

5.3 Spatial and Temporal Patterns in Midvaal, South Africa

The investigation into rates and spatio-temporal patterns of ground deformation in Midvaal, South Africa, revealed by HyP3 SBAS-InSAR data, provide crucial insights into the dynamics of ground subsidence and uplift in the region. Sentinel-1 ascending and descending orbit data demonstrated consistent velocity measurements, confirming the reliability of the approach. Significant ground subsidence was observed in southeastern, northeastern, and southwestern areas, while ground uplift occurred predominantly in the western and southeastern edge zones. Maximum ground deformation rates reached -226.0 mm/year for ground subsidence and 57.2 mm/year for ground uplift, indicating severe localised impacts in specific regions. The analysis revealed that environmental factors such as NDVI and MNDWI inconsistently influenced ground deformation patterns, while geological and hydrological conditions, including dolomite formations and precipitation, emerged as dominant drivers of ground deformation susceptibility.

The interplay between precipitation, groundwater storage anomalies, and human activities such as groundwater extraction further underscores the complex interactions influencing ground subsidence. Temporal analyses showed that precipitation affects ground subsidence patterns with a lag, linked to groundwater infiltration and recharge processes. These findings demonstrate the importance of addressing both natural and anthropogenic factors in ground subsidence monitoring and management.

5.4 Ground Subsidence Susceptibility Mapping and Influencing Factors

The study developed a comprehensive methodology for ground subsidence susceptibility mapping by integrating georeferenced velocity data with influential factors while addressing challenges such as multicollinearity and factor redundancy. The use of machine learning algorithms, including Random Forest, XGBoost, LightGBM, and CNN, underscored their predictive power, with the Random Forest model achieving the highest performance metrics across various evaluation parameters. While CNN models underperformed due to their inability to capture critical geographic features like latitude and longitude, SHAP analysis highlighted the importance of these features in influencing spatial patterns of ground subsidence.

Ground subsidence susceptibility maps categorised the region into five risk zones: Very Low, Low, Moderate, High, and Very High. These maps, validated through SHAP analysis and frequency ratio evaluations, highlighted the significant roles of factors like vegetation indices, lithology, and hydrological parameters in shaping ground subsidence susceptibility. Key findings include the influence of medium to high vegetation density, groundwater storage anomalies, and specific lithological formations in driving ground subsidence patterns, especially in areas affected by intensive land use and mining.

5.5 Implications and Future Directions

The findings of this research have direct implications for policymaking, urban planning, and infrastructure management in regions susceptible to ground subsidence. The ground subsidence susceptibility maps provide actionable insights that can inform risk mitigation strategies, particularly in high susceptibility zones. The demonstrated integration of InSAR time series

analysis, DEM analysis, and machine learning methodologies serves as a scalable framework for broader geospatial applications, offering utility for real-time monitoring and early warning systems.

Future research directions are essential for enhancing and expanding the scope of ground subsidence monitoring and prediction. Several promising pathways are outlined below:

- Multi-Sensor Data Integration:
  Future studies should focus on integrating multi-sensor datasets, including L-band SAR imagery, UAV-based photogrammetry, optical imagery, and LiDAR data. Such integration can improve the spatial resolution and sensitivity to different types of ground deformation scatterers, capturing subtle and dynamic movements that may be missed by single-sensor approaches.
- Real-Time Environmental Monitoring Using AI and IoT:
  The development of dynamic predictive models that incorporate real-time environmental, geological, and anthropogenic data is critical. Leveraging AI techniques alongside Internet of Things (IoT) sensor networks can significantly enhance the temporal resolution of ground deformation monitoring, enabling near-instantaneous risk assessments and early warning systems.
- Refinement for Complex Terrains:
  Refining susceptibility models specifically for complex terrains such as mountainous, vegetated, and densely urbanised environments will improve model generalisability. Tailored calibration of DEMs and adjustment of feature selection strategies can increase accuracy in these challenging contexts.
- Extension to Other Geohazards:
  The proposed framework can also be adapted to monitor and predict other geohazards such as landslides and sinkholes. Testing the framework's adaptability across multiple hazards would strengthen its practical utility for disaster risk reduction initiatives.

In conclusion, this study advances the understanding and management of ground subsidence through innovative methodologies and practical insights. The integration of geospatial analysis, machine learning, and environmental modelling offers a robust foundation for future geohazard assessments and risk mitigation initiatives.

6 References

Abdikan, S., Hooper, A., Arikan, M., Balik Şanli, F., Cakir, F., Kemaldere, H., 2011. InSAR time series analysis of coal mining in Zonguldak city, Northwestern Turkey. Fringe Workshop 2011, European Space Agency (Special Publication) SP–696, Frascati, Italy, 19–23 September 2011.

Achour, Y., Saidani, Z., Touati, R., Pham, Q.B., Pal, S.C., Mustafa, F., Balik Sanli, F., 2021. Assessing landslide susceptibility using a machine learning-based approach to achieving land degradation neutrality. Environ Earth Sci 80, 575. https://doi.org/10.1007/s12665-021-09889-9

Ahmad, A., Chao, W., Yixian, T., Sultan, M., Falak, A., Wei, D., Jing, W., 2022. SAR-based Subsidence Monitoring and Assessment of the Factors Involved in the Occurrence of Subsidence, Lahore City. Journal of Resources and Ecology 13. https://doi.org/10.5814/j.issn.1674-764x.2022.05.007

Ai, Y., Song, R., Huang, C., Cui, C., Tian, B., Chen, L., 2024. A Real-Time Road Boundary Detection Approach in Surface Mine Based on Meta Random Forest. IEEE Trans. Intell. Veh. 9, 1989–2001. https://doi.org/10.1109/TIV.2023.3296767

Airbus, D., 2020. Copernicus DEM Copernicus Digital Elevation Model Validation Report. Airbus Defence and Space—Intelligence: Potsdam, Germany.

Ajayi, O.G., Yusuf, B., Opaluwa, Y.D., 2023. Monitoring ground displacement of Warri metropolis using persistent scatterer interferometry (PSI) and small baseline subset (SBAS) time series.

Alogayell, H.M., Al-Alola, S.S., Alkadi, I.I., Mohamed, S.A., Ismail, I.Y., El-Bukmi, F., 2021. Spatial modeling of ground subsidence susceptibility along Al-Shamal train pathway in Saudi Arabia. Open Geosciences 13, 1158–1173. https://doi.org/10.1515/geo-2020-0296

Argyrakis, P., Ganas, A., Valkaniotis, S., Tsioumas, V., Sagias, N., Psiloglou, B., 2020. Anthropogenically induced subsidence in Thessaly, central Greece: new evidence from GNSS data. Nat Hazards 102, 179–200. https://doi.org/10.1007/s11069-020-03917-w

Ashraf, H., Cawood, F., 2015. Geospatial subsidence hazard modelling at Sterkfontein Caves. SA J of Geomatics 4, 273. https://doi.org/10.4314/sajg.v4i3.8

Aslan, G., Cakir, Z., 2019. Monitoring of Surface Deformation in NorthWest Turkey from High Resolution InSAR: Focus on Tectonic Aseismic Slip and Subsidence.

Auriol, A., Tourain, C., 2010. DORIS system: The new age. Advances in Space Research 46, 1484–1496. https://doi.org/10.1016/j.asr.2010.05.015

Avci, C., Budak, M., Yağmur, N., Balçik, F., 2023. Comparison between random forest and support vector machine algorithms for LULC classification. International Journal of Engineering and Geosciences 8, 1–10. https://doi.org/10.26833/ijeg.987605

Ayani, R.H., Ching, K.E., Anjasmara, I.M., Lin, Y.N., 2021. Crustal Deformation of the Kendeng Fault Branches Area from GNSS and InSAR Data in Surabaya City, Indonesia. IOP Conf. Ser.: Earth Environ. Sci. 936, 012019. https://doi.org/10.1088/1755-1315/936/1/012019

Ayati, A.H., Haghighi, A., Ghafouri, H.R., 2022. Machine learning approach to transient-based leak detection of pressurized pipelines: Classification vs Regression. J Civil Struct Health Monit 12, 611–628. https://doi.org/10.1007/s13349-022-00568-2

Bagheri, A., Sadeghfam, S., Karimzadeh, S., Nadiri, A.A., 2024. Subsidence vulnerability indexing using convolutional neural networks based on clustering and regression modeling strategies. Groundwater for Sustainable Development 25, 101180. https://doi.org/10.1016/j.gsd.2024.101180

Bagheri-Gavkosh, M., Hosseini, S.M., Ataie-Ashtiani, B., Sohani, Y., Ebrahimian, H., Morovat, F., Ashrafi, S., 2021. Land subsidence: A global challenge. Science of The Total Environment 778, 146193. https://doi.org/10.1016/j.scitotenv.2021.146193

Behera, A., Singh Rawat, K., 2023. A brief review paper on mining subsidence and its geo-environmental impact. Materials Today: Proceedings S2214785323020795. https://doi.org/10.1016/j.matpr.2023.04.183

Bianchini, S., Solari, L., Del Soldato, M., Raspini, F., Montalti, R., Ciampalini, A., Casagli, N., 2019. Ground Subsidence Susceptibility (GSS) Mapping in Grosseto Plain (Tuscany, Italy) Based on Satellite InSAR Data Using Frequency Ratio and Fuzzy Logic. Remote Sensing 11, 2015. https://doi.org/10.3390/rs11172015

Blewitt, G., Hammond, W., Kreemer, C., 2018. Harnessing the GPS Data Explosion for Interdisciplinary Science. Eos 99. https://doi.org/10.1029/2018EO104623

Blöthe, D.J., 2018. Analysing the Capabilities and Limitations of InSAR using Sentinel-1 data for Landslide Detection and Monitoring.

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brempong, E.K., Almar, R., Angnuureng, D.B., Mattah, P.A.D., Avornyo, S.Y., Jayson-Quashigah, P.N., Addo, K.A., Minderhoud, P., Teatini, P., 2023. Future flooding of the Volta Delta caused by sea level rise and land subsidence. J Coast Conserv 27, 24. https://doi.org/10.1007/s11852-023-00952-0

Bushenkova, A., Soares, P.M.M., Johannsen, F., Lima, D.C.A., 2024. Towards an improved representation of the urban heat island effect : A multi-scale application of XGBoost for madrid. Urban Climate 55, 101982. https://doi.org/10.1016/j.uclim.2024.101982

Cai, Y., Jin, Y., Wang, Z., Chen, T., Wang, Y., Kong, W., Xiao, W., Li, X., Lian, X., Hu, H., 2023. A review of monitoring, calculation, and simulation methods for ground subsidence induced by coal mining. Int J Coal Sci Technol 10, 32. https://doi.org/10.1007/s40789-023-00595-4

Chai, L., Xie, X., Wang, C., Tang, G., Song, Z., 2024. Ground subsidence risk assessment method using PS-InSAR and LightGBM: a case study of Shanghai metro network. International Journal of Digital Earth 17, 2297842. https://doi.org/10.1080/17538947.2023.2297842

Chaussard, E., Amelung, F., Abidin, H., Hong, S.-H., 2013. Sinking cities in Indonesia: ALOS PALSAR detects rapid subsidence due to groundwater and gas extraction. Remote Sensing of Environment 128, 150–161. https://doi.org/10.1016/j.rse.2012.10.015

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Presented at the KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco California USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785

Choi, J.-K., Kim, K.-D., Lee, S., Won, J.-S., 2010. Application of a fuzzy operator to susceptibility estimations of coal mine subsidence in Taebaek City, Korea. Environ Earth Sci 59, 1009–1022. https://doi.org/10.1007/s12665-009-0093-6

Cian, F., Blasco, J., Carrera, L., 2019. Sentinel-1 for Monitoring Land Subsidence of Coastal Cities in Africa Using PSInSAR: A Methodology Based on the Integration of SNAP and StaMPS. Geosciences 9, 124. https://doi.org/10.3390/geosciences9030124

Cigna, F., Esquivel Ramírez, R., Tapete, D., 2021. Accuracy of Sentinel-1 PSI and SBAS InSAR Displacement Velocities against GNSS and Geodetic Leveling Monitoring Data. Remote Sensing 13, 4800. https://doi.org/10.3390/rs13234800

Cohen, M.H., 1973. Introduction to very-long-baseline interferometry. Proc. IEEE 61, 1192–1197. https://doi.org/10.1109/PROC.1973.9244

Constantinou, S., Van Rooy, J., 2018. Sinkhole and subsidence size distribution across dolomitic land in Gauteng. J. S. Afr. Inst. Civ. Eng. 60, 2–8. https://doi.org/10.17159/2309-8775/2018/v60n2a1

Costache, R., Ali, S.A., Parvin, F., Pham, Q.B., Arabameri, A., Nguyen, H., Crăciun, A., Anh, D.T., 2022. Detection of areas prone to flood-induced landslides risk using certainty factor and its hybridization with FAHP, XGBoost and deep learning neural network. Geocarto International 37, 7303–7338. https://doi.org/10.1080/10106049.2021.1973115

Dauteuil, O., Jolivet, M., Gaudaré, L., Pastier, A., 2023. Rainfall-induced ground deformation in southern Africa. Terra Nova 35, 260–266. https://doi.org/10.1111/ter.12650

Dawod, G.M., Mohamed, H.F., Amin, A.M., 2023. Monitoring Vertical Land Movements in the Nile Delta, Egypt by GNSS Datasets Over 2012-2023.

Dehne, M., Cervantes, F.G., Sheard, B., Heinzel, G., Danzmann, K., 2009. Laser interferometer for spaceborne mapping of the Earth's gravity field. J. Phys.: Conf. Ser. 154, 012023. https://doi.org/10.1088/1742-6596/154/1/012023

Di Stefano, F., Cuevas-Gonzalez, M., Luzi, G., Malinverni, E.S., 2022. Assessing levelling and DInSAR for deformation monitoring in seismic region. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLIII-B3-2022, 263–270. https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-263-2022

Ding, X., Huang, W., 2011. D-InSAR monitoring of crustal deformation in the eastern segment of the Altyn Tagh Fault. International Journal of Remote Sensing 32, 1797–1806. https://doi.org/10.1080/01431160903490190

Dong, J., Mei, Y., Jiang, H., Gao, K., 2023. InSAR monitoring and evaluation study on the safety and stability of high-altitude limestone dumps. https://doi.org/10.21203/rs.3.rs-2636149/v1

Edmonds, C., 2018. Five decades of settlement and subsidence. QJEGH 51, 403–416. https://doi.org/10.1144/qjegh2018-089

Eghrari, Z., Delavar, M.R., Zare, M., Beitollahi, A., Nazari, B., 2023. Land subsidence susceptibility mapping using machine learning algorithms. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. X-4/W1-2022, 129–136. https://doi.org/10.5194/isprs-annals-X-4-W1-2022-129-2023

Emery, W.J., Camps, A., 2017. Introduction to satellite remote sensing: atmosphere, ocean, land and cryosphere applications. Elsevier, Amsterdam, Netherlands ; Cambridge, MA.

Engelbrecht, J., Inggs, M., 2013. Differential Interferometry Techniques on L-Band Data Employed for the Monitoring of Surface Subsidence Due to Mining. South African Journal of Geomatics 2.

Fabris, M., Achilli, V., Menin, A., 2014. Estimation of Subsidence in Po Delta Area (Northern Italy) by Integration of GPS Data, High-Precision Leveling and Archival Orthometric Elevations. IJG 05, 571–585. https://doi.org/10.4236/ijg.2014.56052

Fadhillah, M.F., Achmad, A.R., Lee, C.-W., 2020. Integration of InSAR Time-Series Data and GIS to Assess Land Subsidence along Subway Lines in the Seoul Metropolitan Area, South Korea. Remote Sensing 12, 3505. https://doi.org/10.3390/rs12213505

Fadhillah, M.F., Hakim, W.L., Park, S., Kim, D., Park, Y.-C., Kim, C.-H., Lee, C.-W., 2022. Surface deformation simulation for InSAR detection using a machine learning approach on the hantangang river volcanic field: A case study on the orisan mountain. Front. Environ. Sci. 10, 968120. https://doi.org/10.3389/fenvs.2022.968120

Fan, H., Lian, X., Yang, W., Ge, L., Hu, H., Du, Z., 2021. Mining large-gradient subsidence monitoring using D-InSAR optimized by GNSS. The Imaging Science Journal 69, 207–218. https://doi.org/10.1080/13682199.2022.2156667

Fattahi, H., Agram, P., Simons, M., 2017. A Network-Based Enhanced Spectral Diversity Approach for TOPS Time-Series Analysis. IEEE Trans. Geosci. Remote Sensing 55, 777–786. https://doi.org/10.1109/TGRS.2016.2614925

Ferretti, A., 2014. Satellite InSAR Data: Reservoir Monotoring from Space, Education Tour Series. EAGE, DB Houten.

Flores, A., Herndon, K., Thapa, R., Cherrington, E., 2019. Synthetic Aperture Radar (SAR) Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation. https://doi.org/10.25966/NR2C-S697

Foumelis, M., Delgado Blasco, J.M., Desnos, Y.-L., Engdahl, M., Fernandez, D., Veci, L., Lu, J., Wong, C., 2018. Esa Snap - Stamps Integrated Processing for Sentinel-1 Persistent Scatterer Interferometry, in: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Valencia, pp. 1364–1367. https://doi.org/10.1109/IGARSS.2018.8519545

Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., Wolff, E., 2018. Very High Resolution Object-Based Land Use–Land Cover Urban Classification Using Extreme Gradient Boosting. IEEE Geosci. Remote Sensing Lett. 15, 607–611. https://doi.org/10.1109/LGRS.2018.2803259

Gharechaee, H., Samani, A.N., Sigaroodi, S.K., Baloochiyan, A., Moosavi, M.S., Hubbart, J.A., Sadeghi, S.M.M., 2023. Land Subsidence Susceptibility Mapping Using Interferometric Synthetic Aperture Radar (InSAR) and Machine Learning Models in a Semiarid Region of Iran. Land 12, 843. https://doi.org/10.3390/land12040843

Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random Forests for land cover classification. Pattern Recognition Letters 27, 294–300. https://doi.org/10.1016/j.patrec.2005.08.011

Gómez, M.F., Lencinas, J.D., Siebert, A., Díaz, G.M., 2012. Accuracy Assessment of ASTER and SRTM DEMs: A Case Study in Andean Patagonia. GIScience & Remote Sensing 49, 71–91. https://doi.org/10.2747/1548-1603.49.1.71

Guth, P.L., Van Niekerk, A., Grohmann, C.H., Muller, J.-P., Hawker, L., Florinsky, I.V., Gesch, D., Reuter, H.I., Herrera-Cruz, V., Riazanoff, S., López-Vázquez, C., Carabajal, C.C., Albinet, C., Strobl, P., 2021. Digital Elevation Models: Terminology and Definitions. Remote Sensing 13, 3581. https://doi.org/10.3390/rs13183581

Hakim, W.L., Fadhillah, M.F., Kim, B., Park, S., Lee, C.-W., 2023a. Integrating ICOPS Time-Series InSAR Measurement with the Convolutional Neural Network (CNN) and Optimized Hot Spot Analysis (OHSA) to Monitor Land Subsidence in Pekalongan, Indonesia, in: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Pasadena, CA, USA, pp. 1724–1727. https://doi.org/10.1109/IGARSS52108.2023.10282286

Hakim, W.L., Fadhillah, M.F., Park, S., Pradhan, B., Won, J.-S., Lee, C.-W., 2023b. InSAR time-series analysis and susceptibility mapping for land subsidence in Semarang, Indonesia using convolutional neural network and support vector regression. Remote Sensing of Environment 287, 113453. https://doi.org/10.1016/j.rse.2023.113453

Han, D., Currell, M.J., Cao, G., Hall, B., 2017. Alterations to groundwater recharge due to anthropogenic landscape change. Journal of Hydrology 554, 545–557. https://doi.org/10.1016/j.jhydrol.2017.09.018

Han, Y., Liu, G., Liu, J., Yang, J., Xie, X., Yan, W., Zhang, W., 2023. Monitoring and Analysis of Land Subsidence in Jiaozuo City (China) Based on SBAS-InSAR Technology. Sustainability 15, 11737. https://doi.org/10.3390/su151511737

Ho Tong Minh, D., Hanssen, R., Rocca, F., 2020. Radar Interferometry: 20 Years of Development in Time Series Techniques and Future Perspectives. Remote Sensing 12, 1364. https://doi.org/10.3390/rs12091364

Hofmann-Wellenhof, B., Lichtenegger, H., Wasle, E., 2008. GNSS - Global Navigation Satellite Systems: GPS, GLONASS, Galileo and more. Springer, Wien New York.

Hooper, A., Bekaert, D., Spaans, K., Arıkan, M., 2012. Recent advances in SAR interferometry time series analysis for measuring crustal deformation. Tectonophysics 514–517, 1–13. https://doi.org/10.1016/j.tecto.2011.10.013

Hosseinzadeh, E., Anamaghi, S., Behboudian, M., Kalantari, Z., 2024. Evaluating Machine Learning-Based Approaches in Land Subsidence Susceptibility Mapping. Land 13, 322. https://doi.org/10.3390/land13030322

Hrysiewicz, A., Wang, X., Holohan, E.P., 2023. EZ-InSAR: An easy-to-use open-source toolbox for mapping ground surface deformation using satellite interferometric synthetic aperture radar. Earth Sci Inform 16, 1929–1945. https://doi.org/10.1007/s12145-023-00973-1

hyp3-docs.asf.alaska.edu, n.d. HyP3. [online] Available at: https://hyp3-docs.asf.alaska.edu/ [Accessed 28 July. 2023]. .

Ignjatović Stupar, D., Rošer, J., Vulić, M., 2020. Investigation of Unmanned Aerial Vehicles-Based Photogrammetry for Large Mine Subsidence Monitoring. Minerals 10, 196. https://doi.org/10.3390/min10020196

Isiaka, A.I., Durrheim, R.J., Manzi, M.S.D., 2019. High-Resolution Seismic Reflection Investigation of Subsidence and Sinkholes at an Abandoned Coal Mine Site in South Africa. Pure Appl. Geophys. 176, 1531–1548. https://doi.org/10.1007/s00024-018-2026-3

Jing, X., Li, Z., Chen, H., Zhang, C., 2024. “Is What We See Always Real?” A Comparative Study of Two-Dimensional and Three-Dimensional Urban Green Spaces: The Case of Shenzhen’s Central District. Forests 15, 983. https://doi.org/10.3390/f15060983

Josset, L., Lall, U., Prakash, D., Dinar, A., 2024. Public Health, Socioeconomic and Environmental Impacts of Urban Land Subsidence. https://doi.org/10.22541/essoar.170808438.89045530/v2

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc.

Khan, S.D., Huang, Z., Karacay, A., 2014. Study of ground subsidence in northwest Harris county using GPS, LiDAR, and InSAR techniques. Nat Hazards 73, 1143–1173. https://doi.org/10.1007/s11069-014-1067-x

Kirui, P., Oiro, S., Waithaka, H., Odera, P., Riedel, B., Gerke, M., 2022. Detection, characterization, and analysis of land subsidence in Nairobi using InSAR. Nat Hazards 113, 213–236. https://doi.org/10.1007/s11069-022-05296-w

Kok, S., Costa, A.L., 2021. Framework for economic cost assessment of land subsidence. Nat Hazards 106, 1931–1949. https://doi.org/10.1007/s11069-021-04520-3

Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. Geoscience Frontiers 7, 3–10. https://doi.org/10.1016/j.gsf.2015.07.003

Le, T.-T.-H., Hwang, Y., Choi, C., Wardhani, R.W., Putranto, D.S.C., Kim, H., 2024. Enhancing Structured Query Language Injection Detection with Trustworthy Ensemble Learning and Boosting Models Using Local Explanation Techniques. Electronics 13, 4350. https://doi.org/10.3390/electronics13224350

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

Lee, S., Kang, J., Kim, J., 2023. Machine Learning-Based Predictive Model of Ground Subsidence Risk Using Characteristics of Underground Pipelines in Urban Areas. IEEE Access 11, 69326–69336. https://doi.org/10.1109/ACCESS.2023.3293089

Lee, S., Park, I., Choi, J.-K., 2012. Spatial Prediction of Ground Subsidence Susceptibility Using an Artificial Neural Network. Environmental Management 49, 347–358. https://doi.org/10.1007/s00267-011-9766-5

Li, F., Liu, G., Tao, Q., Zhai, M., 2023. Land subsidence prediction model based on its influencing factors and machine learning methods. Nat Hazards 116, 3015–3041. https://doi.org/10.1007/s11069-022-05796-9

Li, P., Wu, J., Zhou, W., LaMoreaux, J.W., 2023. Land Subsidence and Groundwater Seepage, in: Hazard Hydrogeology, Environmental Earth Sciences. Springer International Publishing, Cham, pp. 45–74. https://doi.org/10.1007/978-3-031-48427-8_3

Li, X., Zhang, F., Li, Y., Guo, Q., Wan, Y., Bu, X., Liu, Y., Liang, X., 2021. An Elevation Ambiguity Resolution Method Based on Segmentation and Reorganization of TomoSAR Point Cloud in 3D Mountain Reconstruction. Remote Sensing 13, 5118. https://doi.org/10.3390/rs13245118

Li, Z., Luo, S., Tan, X., Wang, J., 2024. Trend Analysis of High-Resolution Soil Moisture Data Based on GAN in the Three-River-Source Region During the 21st Century. Remote Sensing 16, 4367. https://doi.org/10.3390/rs16234367

Linder, W., 2003. Digital Photogrammetry. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-06725-3

Liu, R., Kuffer, M., Persello, C., 2019. The Temporal Dynamics of Slums Employing a CNN-Based Change Detection Approach. Remote Sensing 11, 2844. https://doi.org/10.3390/rs11232844

Maaß, A.-L., Schüttrumpf, H., 2018. Long-term effects of mining-induced subsidence on the trapping efficiency of floodplains. Anthropocene 24, 1–13. https://doi.org/10.1016/j.ancene.2018.10.001

Martone, M., n.d. Onboard Quantization for Interferometric and Multichannel Synthetic Aperture Radar (SAR) Systems.

Maune, D.F., 2007. Digital elevation model technologies and applications: the DEM users manual. American Society for Photogrammetry & Remote Sensing.

Memon, N., Parikh, H., Patel, S.B., Patel, D., Patel, V.D., 2021. Automatic land cover classification of multi-resolution dualpol data using convolutional neural network (CNN). Remote Sensing Applications: Society and Environment 22, 100491. https://doi.org/10.1016/j.rsase.2021.100491

Modoni, G., Darini, G., Spacagna, R.L., Saroli, M., Russo, G., Croce, P., 2013. Spatial analysis of land subsidence induced by groundwater withdrawal. Engineering Geology 167, 59–71. https://doi.org/10.1016/j.enggeo.2013.10.014

Mohammady, M., Pourghasemi, H.R., Amiri, M., 2019. Land subsidence susceptibility assessment using random forest machine learning algorithm. Environ Earth Sci 78, 503. https://doi.org/10.1007/s12665-019-8518-3

Monterroso, F., Bonano, M., Luca, C.D., Lanari, R., Manunta, M., Manzo, M., Onorato, G., Zinno, I., Casu, F., 2020. A Global Archive of Coseismic DInSAR Products Obtained Through Unsupervised Sentinel-1 Data Processing. Remote Sensing 12, 3189. https://doi.org/10.3390/rs12193189

Montjane, A.K., Tshibubudze, A., Woldai, T., Heath, L., 2020. The influence of geological fractures on karstic sinkhole development in eastern areas of Centurion, South Africa. Environ Earth Sci 79, 488. https://doi.org/10.1007/s12665-020-09234-6

Na, T., Kawamura, Y., Kang, S., Utsuki, S., 2021. Hazard mapping of ground subsidence in east area of Sapporo using frequency ratio model and GIS. Geomatics, Natural Hazards and Risk 12, 347–362. https://doi.org/10.1080/19475705.2021.1873198

Naghibi, S.A., Khodaei, B., Hashemi, H., 2022. An integrated InSAR-machine learning approach for ground deformation rate modeling in arid areas. Journal of Hydrology 608, 127627. https://doi.org/10.1016/j.jhydrol.2022.127627

Najafi, Z., Pourghasemi, H.R., Ghanbarian, G., Fallah Shamsi, S.R., 2020. Land-subsidence susceptibility zonation using remote sensing, GIS, and probability models in a Google Earth Engine platform. Environ Earth Sci 79, 491. https://doi.org/10.1007/s12665-020-09238-2

National Planning Commission, 2011. National Development Plan 2030: Our Future-make it work. Sherino Printers

Nur, A.S., Nam, B.H., Choi, S., Kim, Y.J., 2024. Monitoring of ground subsidence using PS-InSAR technique in the Southeast Texas (SETX) Region. Geo-Engineering 15, 13. https://doi.org/10.1186/s40703-024-00215-3

Ogundare, J.O., 2015. Precision Surveying: Principles and Geomatics Practice, The, 1st ed. Wiley. https://doi.org/10.1002/9781119147770

Orhan, O., Yakar, M., Ekercin, S., 2020. An application on sinkhole susceptibility mapping by integrating remote sensing and geographic information systems. Arab J Geosci 13, 886. https://doi.org/10.1007/s12517-020-05841-6

Osmanoğlu, B., Sunar, F., Wdowinski, S., Cabral-Cano, E., 2016. Time series analysis of InSAR data: Methods and trends. ISPRS Journal of Photogrammetry and Remote Sensing 115, 90–102. https://doi.org/10.1016/j.isprsjprs.2015.10.003

Pacheco-Martínez, J., Cabral-Cano, E., Wdowinski, S., Hernández-Marín, M., Ortiz-Lozano, J., Zermeño-de-León, M., 2015. Application of InSAR and Gravimetry for Land Subsidence Hazard Zoning in Aguascalientes, Mexico. Remote Sensing 7, 17035–17050. https://doi.org/10.3390/rs71215868

Pan, B., Tian, L., 2016. Advanced video extensometer for non-contact, real-time, high-accuracy strain measurement. Opt. Express 24, 19082. https://doi.org/10.1364/OE.24.019082

Papoutsis, I., Kontoes, C., Alatza, S., Apostolakis, A., Loupasakis, C., 2020. InSAR Greece with Parallelized Persistent Scatterer Interferometry: A National Ground Motion Service for Big Copernicus Sentinel-1 Data. Remote Sensing 12, 3207. https://doi.org/10.3390/rs12193207

Petro, L., Palvo, L., 2019. Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS. https://doi.org/10.48550/ARXIV.1912.06059

Qiao, X., Chu, T., Krell, E., Tissot, P., Holland, S., Ahmed, M., Smilovsky, D., 2024. Interpretation and Attribution of Coastal Land Subsidence: An InSAR and Machine Learning Perspective. IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing 17, 4768–4783. https://doi.org/10.1109/JSTARS.2024.3361391

Ranjgar, B., Razavi-Termeh, S.V., Foroughnia, F., Sadeghi-Niaraki, A., Perissin, D., 2021. Land Subsidence Susceptibility Mapping Using Persistent Scatterer SAR Interferometry Technique and Optimized Hybrid Machine Learning Algorithms. Remote Sensing 13, 1326. https://doi.org/10.3390/rs13071326

Ravi, R., Baranidharan, Dr.B., 2020. Crop Yield Prediction using XG Boost Algorithm. IJRTE 8, 3516–3520. https://doi.org/10.35940/ijrte.D9547.018520

Reuter, H.I., Hengl, T., Gessler, P., Soille, P., 2009. Chapter 4 Preparation of DEMs for Geomorphometric Analysis, in: Developments in Soil Science. Elsevier, pp. 87–120. https://doi.org/10.1016/S0166-2481(08)00004-4

Richardson, S., 2013. Sinkhole and subsidence record in the chuniespoort group dolomite, Gauteng, South Africa, Master's thesis. Pretoria: University of Pretoria, Department of Geology.

Rosen, P.A., Gurrola, E., Sacco, G.F., Zebker, H., 2012. The InSAR scientific computing environment, paper presented at EUSAR 2012, 23-26 April 2012.

Saether, O.M., Caritat, P. de (Eds.), 2020. Geochemical processes, weathering and groundwater recharge in catchments. CRC Press, Boca Raton.

Saha, A., Tripathi, L., Villuri, V.G.K., Bhardwaj, A., 2024. Exploring machine learning and statistical approach techniques for landslide susceptibility mapping in Siwalik Himalayan Region using geospatial technology. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-023-31670-7

Salehi Moteahd, F., Hafezi Moghaddas, N., Lashkaripour, G.R., Dehghani, M., 2019. Geological parameters affected land subsidence in Mashhad plain, north-east of Iran. Environ Earth Sci 78, 405. https://doi.org/10.1007/s12665-019-8413-y

Sekkeravani, M.A., Bazrafshan, O., Pourghasemi, H.R., Holisaz, A., 2022. Spatial modeling of land subsidence using machine learning models and statistical methods. Environ Sci Pollut Res 29, 28866–28883. https://doi.org/10.1007/s11356-021-18037-6

Sheikh, S., 2020. Navigation Using Pulsars and Other Variable Celestial Sources, in: Morton, Y.T.J., Diggelen, F., Spilker, J.J., Parkinson, B.W., Lo, S., Gao, G. (Eds.), Position, Navigation, and Timing Technologies in the 21st Century. Wiley, pp. 1635–1667. https://doi.org/10.1002/9781119458555.ch52

Shi, L., Gong, H., Chen, B., Zhou, C., 2020. Land Subsidence Prediction Induced by Multiple Factors Using Machine Learning Method. Remote Sensing 12, 4044. https://doi.org/10.3390/rs12244044

Sośnica, K., Thaller, D., Dach, R., Steigenberger, P., Beutler, G., Arnold, D., Jäggi, A., 2015. Satellite laser ranging to GPS and GLONASS. J Geod 89, 725–743. https://doi.org/10.1007/s00190-015-0810-8

Stouthamer, E., Erkens, G., Cohen, K., Hegger, D., Driessen, P., Weikard, H.P., Hefting, M., Hanssen, R., Fokker, P., Van Den Akker, J., Groothuijse, F., Van Rijswick, M., 2020. Dutch national scientific research program on land subsidence: Living on soft soils – subsidence and society. Proc. IAHS 382, 815–819. https://doi.org/10.5194/piahs-382-815-2020

Sun, H., Peng, H., Zeng, M., Wang, S., Pan, Y., Pi, P., Xue, Z., Zhao, X., Zhang, A., Liu, F., 2023. Land Subsidence in a Coastal City Based on SBAS-InSAR Monitoring: A Case Study of Zhuhai, China. Remote Sensing 15, 2424. https://doi.org/10.3390/rs15092424

Sun, M., Du, Y., Liu, Q., Feng, G., Peng, X., Liao, C., 2023. Understanding the Spatial-Temporal Characteristics of Land Subsidence in Shenzhen under Rapid Urbanization Based on MT-InSAR. IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing 16, 4153–4166. https://doi.org/10.1109/JSTARS.2023.3264652

Tadono, T., Ishida, H., Oda, F., Naito, S., Minakawa, K., Iwamoto, H., 2014. Precise Global DEM Generation by ALOS PRISM. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. II–4, 71–76. https://doi.org/10.5194/isprsannals-II-4-71-2014

Takaku, J., Tadono, T., Tsutsui, K., 2014. Generation of High Resolution Global DSM from ALOS PRISM. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XL–4, 243–248. https://doi.org/10.5194/isprsarchives-XL-4-243-2014

Takaku, J., Tadono, T., Tsutsui, K., Ichikawa, M., 2016. Validation of AW3D global DSM generated from ALOS PRISM. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. III–4, 25–31. https://doi.org/10.5194/isprsannals-III-4-25-2016

Tampubolon, W., Reinhardt, W., Behr, F.-J., 2023. Investigations on InSAR data processing standard for volcano island monitoring in Indonesia. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLVIII-1/W2-2023, 1281–1287. https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-1281-2023

The Global Goals, 2023. Goal 15: Life on land [WWW Document]. Goal 15: Life on land. URL https://www.globalgoals.org/goals/15-life-on-land/ (accessed 6.4.23).

Theron, A., Engelbrecht, J., Kemp, J., Kleynhans, W., Turnbull, T., 2017. Detection of Sinkhole Precursors Through SAR Interferometry: Radar and Geological Considerations. IEEE Geosci. Remote Sensing Lett. 14, 871–875. https://doi.org/10.1109/LGRS.2017.2684905

Tien Bui, D., Ho, T.-C., Pradhan, B., Pham, B.-T., Nhu, V.-H., Revhaug, I., 2016. GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. Environ Earth Sci 75, 1101. https://doi.org/10.1007/s12665-016-5919-4

Toomanian, A., Kakroodi, A.A., Etemadi, M.A., 2022. Spatial Modeling of Land Subsidence Using GIS-Based Machine Learning Algorithms. SSRN Journal. https://doi.org/10.2139/ssrn.4024231

Touboul, P., Metris, S., Le Traon, O., Bresson, A., Zahzam, N., Christophe, B., Rodrigues, M., 2016. Gravitation and Geodesy with Inertial Sensors, from Ground to Space. AerospaceLab Journal Issue 12, 16 pages. https://doi.org/10.12762/2016.AL12-11

Wahyudi, E.J., Hafidza, M.H., Tahta, M.A., 2021. Simple design to estimate time-lapse microgravity response due to shallow subsurface density redistribution caused by land subsidence. IOP Conf. Ser.: Earth Environ. Sci. 873, 012040. https://doi.org/10.1088/1755-1315/873/1/012040

Wang, H., Gong, H., Chen, B., Zhou, C., Yang, Y., Sun, X., 2024. Research on land subsidence-rebound affected by dualistic water cycle driven by climate change and human activities in Dezhou City, China. Journal of Hydrology 636, 131327. https://doi.org/10.1016/j.jhydrol.2024.131327

Wang, S., Sun, H., Wei, L., Pi, P., Zeng, M., Pan, Y., Xue, Z., Jiang, X., 2024. Integrated Assessment of Coastal Subsidence in Nansha District, Guangzhou City, China: Insights from SBAS-InSAR Monitoring and Risk Evaluation. Remote Sensing 16, 248. https://doi.org/10.3390/rs16020248

Wang, X., Yang, Y., Xia, Y., Chen, S., She, Y., 2023. Integrating SAR and Geographic Information Data Revealing Land Subsidence and Geological Risks of Shanghai City. Applied Sciences 13, 12091. https://doi.org/10.3390/app132112091

Wegnüller, U., Werner, C., Strozzi, T., Wiesmann, A., Frey, O., Santoro, M., 2016. Sentinel-1 Support in the GAMMA Software. Procedia Computer Science 100, 1305–1312. https://doi.org/10.1016/j.procs.2016.09.246

Werner, C., Wegmüller, U., Strozzi, T., Wiesmann, A., 2000. Gamma SAR and Interferometric Processing Software.

Willis, P., Jayles, C., Bar-Sever, Y., 2006. DORIS: From orbit determination for altimeter missions to geodesy. Comptes Rendus. Géoscience 338, 968–979. https://doi.org/10.1016/j.crte.2005.11.013

Woodhouse, I.H., 2006. Introduction to microwave remote sensing. Taylor&Francis, Boca Raton.

Wu, S., Zhang, B., Ding, X., Zhang, L., Zhang, Zhijie, Zhang, Zeyu, 2023. Radar Interferometry for Urban Infrastructure Stability Monitoring: From Techniques to Applications. Sustainability 15, 14654. https://doi.org/10.3390/su151914654

Wu, Z., Ma, P., Zheng, Y., Gu, F., Liu, L., Lin, H., 2023. Automatic detection and classification of land subsidence in deltaic metropolitan areas using distributed scatterer InSAR and Oriented R-CNN. Remote Sensing of Environment 290, 113545. https://doi.org/10.1016/j.rse.2023.113545

Xu, K., Han, Z., Xu, H., Bin, L., 2023. Rapid Prediction Model for Urban Floods Based on a Light Gradient Boosting Machine Approach and Hydrological–Hydraulic Model. Int J Disaster Risk Sci 14, 79–97. https://doi.org/10.1007/s13753-023-00465-2

Yalcin, A., Reis, S., Aydinoglu, A.C., Yomralioglu, T., 2011. A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. CATENA 85, 274–287. https://doi.org/10.1016/j.catena.2011.01.014

Yalvac, S., 2020. Validating InSAR-SBAS results by means of different GNSS analysis techniques in medium- and high-grade deformation areas. Environ Monit Assess 192, 120. https://doi.org/10.1007/s10661-019-8009-8

Yang, L., Meng, X., Zhang, X., 2011. SRTM DEM and its application advances. International Journal of Remote Sensing 32, 3875–3896. https://doi.org/10.1080/01431161003786016

Yang, M., Zou, L., Cai, H., Qiang, Y., Lin, B., Zhou, B., Abedin, J., Mandal, D., 2022. Spatial–Temporal Land Loss Modeling and Simulation in a Vulnerable Coast: A Case Study in Coastal Louisiana. Remote Sensing 14, 896. https://doi.org/10.3390/rs14040896

Yoon, H.I., Lee, H., Yang, J.-S., Choi, J.-H., Jung, D.-H., Park, Y.J., Park, J.-E., Kim, S.M., Park, S.H., 2023. Predicting Models for Plant Metabolites Based on PLSR, AdaBoost, XGBoost, and LightGBM Algorithms Using Hyperspectral Imaging of Brassica juncea. Agriculture 13, 1477. https://doi.org/10.3390/agriculture13081477

Yuan, Y., Zhang, D., Cui, J., Zeng, T., Zhang, G., Zhou, W., Wang, J., Chen, F., Guo, J., Chen, Z., Guo, H., 2023. Land subsidence prediction in Zhengzhou's main urban area using the GTWR and LSTM models combined with the Attention Mechanism. Science of The Total Environment 907, 167482. https://doi.org/10.1016/j.scitotenv.2023.167482

Yunjun, Z., Fattahi, H., Amelung, F., 2019. Small baseline InSAR time series analysis: Unwrapping error correction and noise reduction. Computers & Geosciences 133, 104331. https://doi.org/10.1016/j.cageo.2019.104331

Zaki, A., Chang, L., Manzella, I., Meijde, M.V.D., Girgin, S., Tanyas, H., Fadel, I., 2024. Automated Python workflow for generating Sentinel-1 PSI and SBAS interferometric stacks using SNAP on Geospatial Computing Platform. Environmental Modelling & Software 178, 106075. https://doi.org/10.1016/j.envsoft.2024.106075

Zhang, P., Guo, Z., Guo, S., Xia, J., 2022. Land Subsidence Monitoring Method in Regions of Variable Radar Reflection Characteristics by Integrating PS-InSAR and SBAS-InSAR Techniques. Remote Sensing 14, 3265. https://doi.org/10.3390/rs14143265

Zhang, Z., Zhu, L., 2023. A Review on Unmanned Aerial Vehicle Remote Sensing: Platforms, Sensors, Data Processing Methods, and Applications. Drones 7, 398. https://doi.org/10.3390/drones7060398

Zhao, F., Gong, W., Tang, H., Pudasaini, S.P., Ren, T., Cheng, Z., 2023. An integrated approach for risk assessment of land subsidence in Xi'an, China using optical and radar satellite images. Engineering Geology 314, 106983. https://doi.org/10.1016/j.enggeo.2022.106983

Zhao, F., Miao, F., Wu, Y., Xiong, Y., Gong, S., Sun, D., 2024. Land subsidence susceptibility mapping in urban settlements using time-series PS-InSAR and random forest model. Gondwana Research 125, 406–424. https://doi.org/10.1016/j.gr.2023.09.019

Zhiqiang, W., Jun, L., 2017. A review of object detection based on convolutional neural network, in: 2017 36th Chinese Control Conference (CCC). Presented at the 2017 36th Chinese Control Conference (CCC), IEEE, Dalian, China, pp. 11104–11109. https://doi.org/10.23919/ChiCC.2017.8029130

Appendix

Appendix A

The tables showing the data sources of the datasets used in this study.

Table A1:Sentinel-2 L2A datasets spatio-temporal resolution, and sources.

Dataset	**Spatial Resolution**	**Temporal Resolution**	**Source**
Sentinel-2 L2A	10 metres	10 days	https://dataspace.copernicus.eu/browser/

Table A2: Precipitation and Temperature datasets and sources.

Dataset	**Source**
Precipitation	https://cds.climate.copernicus.eu/datasets/reanalysis-era5-land-monthly-means?tab=download
Temperature	

Table A3: DEM datasets spatial resolution, sources and vertical accuracy.

Dataset	**Spatial Resolution**	**Source**	**Vertical Accuracy**
AW3D30	30 m	https://portal.opentopography.org/raster?opentopoID=OTALOS.112016.4326.2	< ± 05 m
ASTER	30 m	https://search.earthdata.nasa.gov/search	< ± 16 m
COP30	30 m	https://portal.opentopography.org/raster?opentopoID=OTSDEM.032021.4326.3	< ± 10 m
SRTMGL1	30 m	https://portal.opentopography.org/raster?opentopoID=OTSRTM.082015.4326.1	< ± 16 m

Appendix B

The number of triplets with non-zero integer ambiguity for ascending orbit employing HyP3 and MintPy.

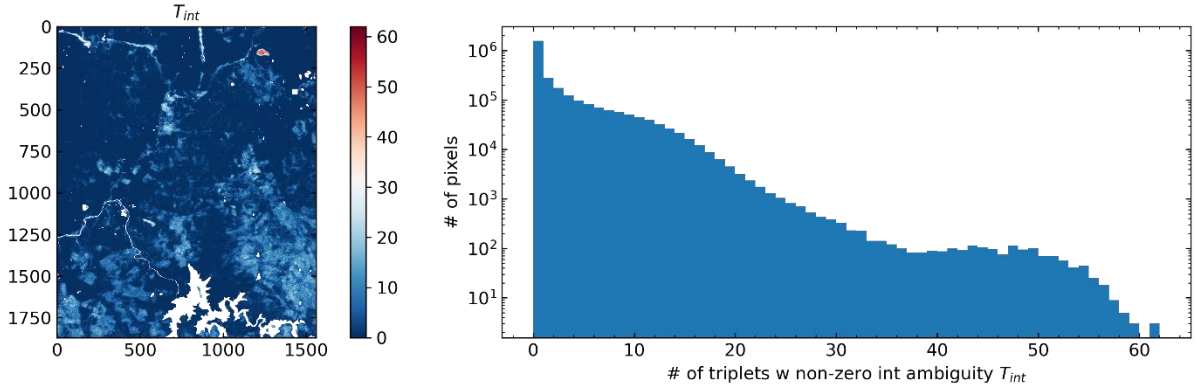


Figure B1: Number of triplets with non-zero integer ambiguity.

Appendix C

The SBAS-InSAR network for the ascending orbit, illustrating the interferometric connections used for InSAR time-series analysis employing HyP3 and MintPy.

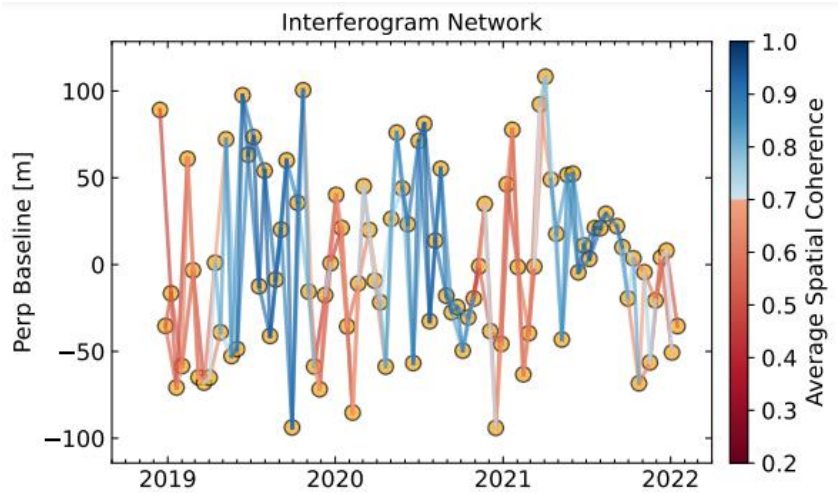


Figure C1: HyP3 SBAS-InSAR network for the ascending orbit.

Appendix D

The SBAS-InSAR RMS time-series residual ramp for the ascending orbit employing HyP3 and MintPy.

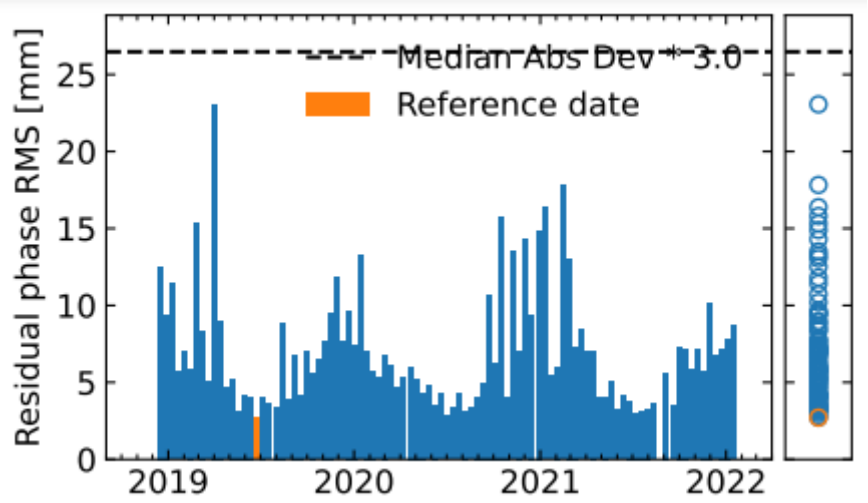


Figure D1: Residual phase RMS for ascending acquisitions.

Appendix E

The SBAS-InSAR Coherence Matrix and History for the ascending orbit employing HyP3 and MintPy.

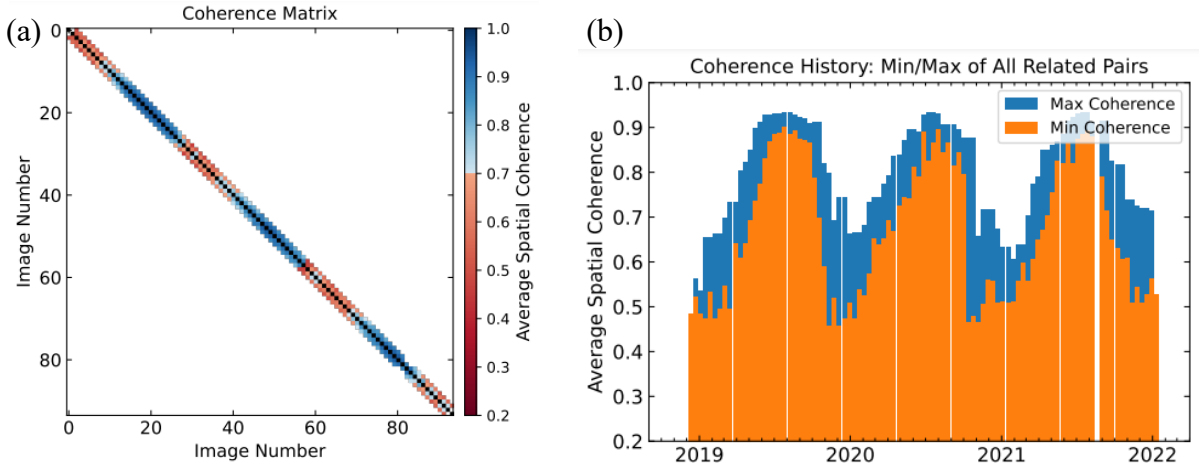


Figure E1: (a) Coherence matrix, and (b) Coherence history for the ascending orbit.

Appendix F

The SBAS-InSAR Spatial and Temporal Coherence for the ascending orbit network shown in Appendix A employing HyP3 and MintPy.

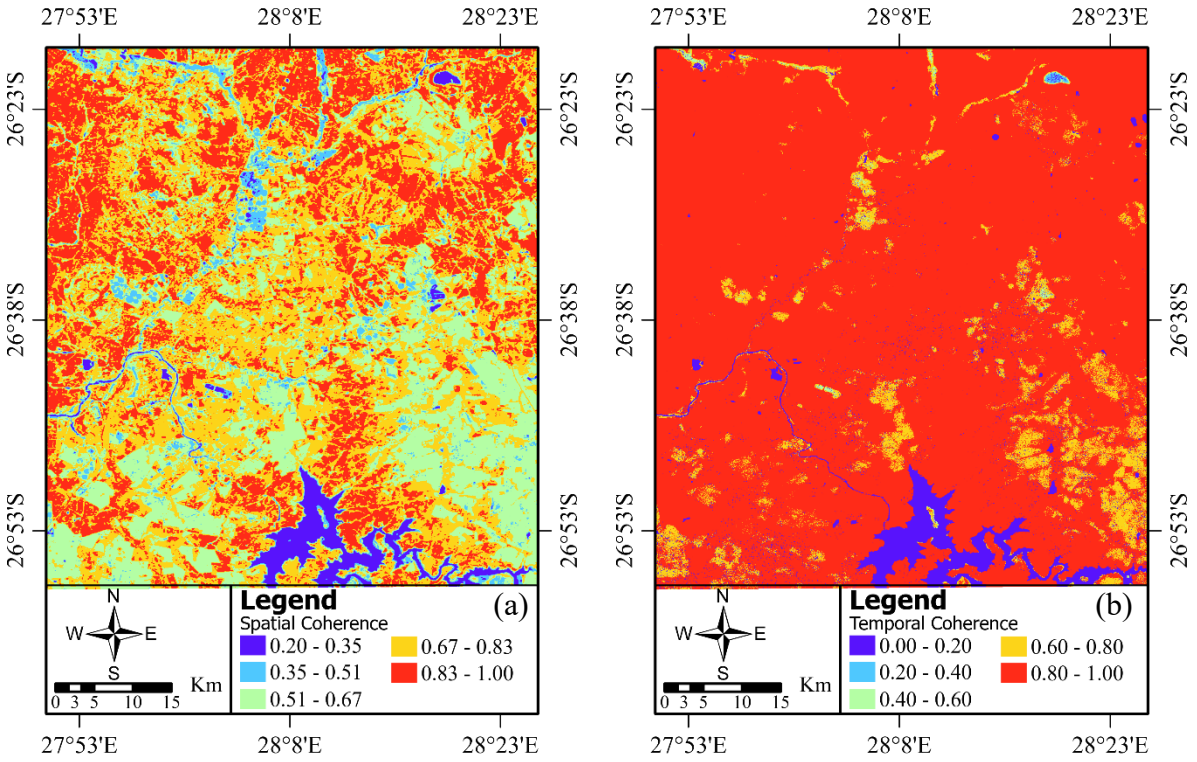


Figure F1: SBAS-InSAR (a) Spatial, and (b) Temporal Coherence for the ascending orbit network.

Appendix G

The SBAS-InSAR Relative and Absolute LOS Velocity differences between cGNSS and InSAR ascending orbit implied, employing ISCE, SNAP, HyP3, MintPy, and StaMPS.

GNSS ID	Relative			Absolute		
	V_{InSAR}	V_{GNSS}	V_{error}	V_{InSAR}	V_{GNSS}	V_{error}
$HEID_{ISCE}^{PS}$	-2.0	7.0	9.0	7.0	7.0	0.0
$VERG_{ISCE}^{PS}$	-0.1	6.0	6.1	8.9	6.0	-2.9
$HEID_{ISCE}^{SBAS}$	-1.0	7.0	8.0	7.0	7.0	0.0
$VERG_{ISCE}^{SBAS}$	0.0	6.0	6.0	8.0	6.0	-2.0
$HEID_{SNAP}^{PS}$	-3.0	7.0	10.0	7.0	7.0	0.0
$VERG_{SNAP}^{PS}$	-2.0	6.0	8.0	8.0	6.0	-2.0
$HEID_{HyP3}^{SBAS}$	-1.0	7.0	8.0	7.0	7.0	0.0
$VERG_{HyP3}^{SBAS}$	3.9	6.0	2.1	11.9	6.0	-5.9

Appendix H

The elevation ranges, DEM implied velocity, and DEM implied velocity ranges frequency distribution over 247 levelling points in Midvaal, South Africa. And the relationship between elevation and velocity over 247 levelling points in Midvaal, South Africa.

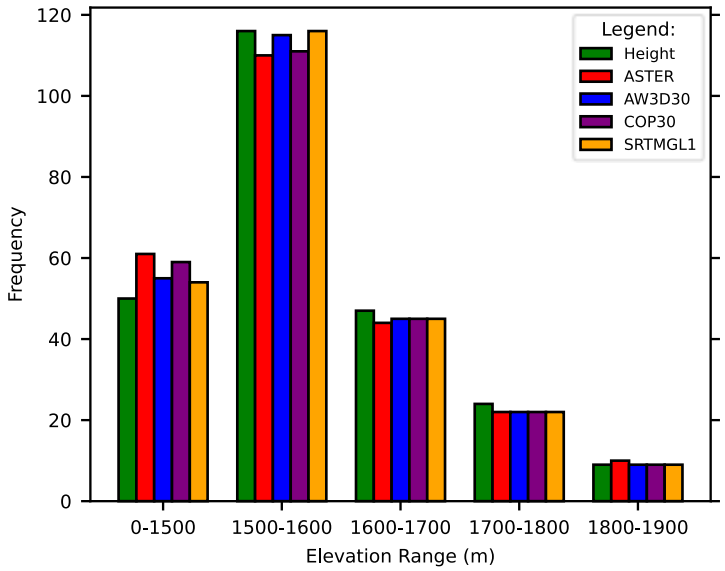


Figure H1: Frequency Distribution of Elevation Ranges in Midvaal, South Africa.

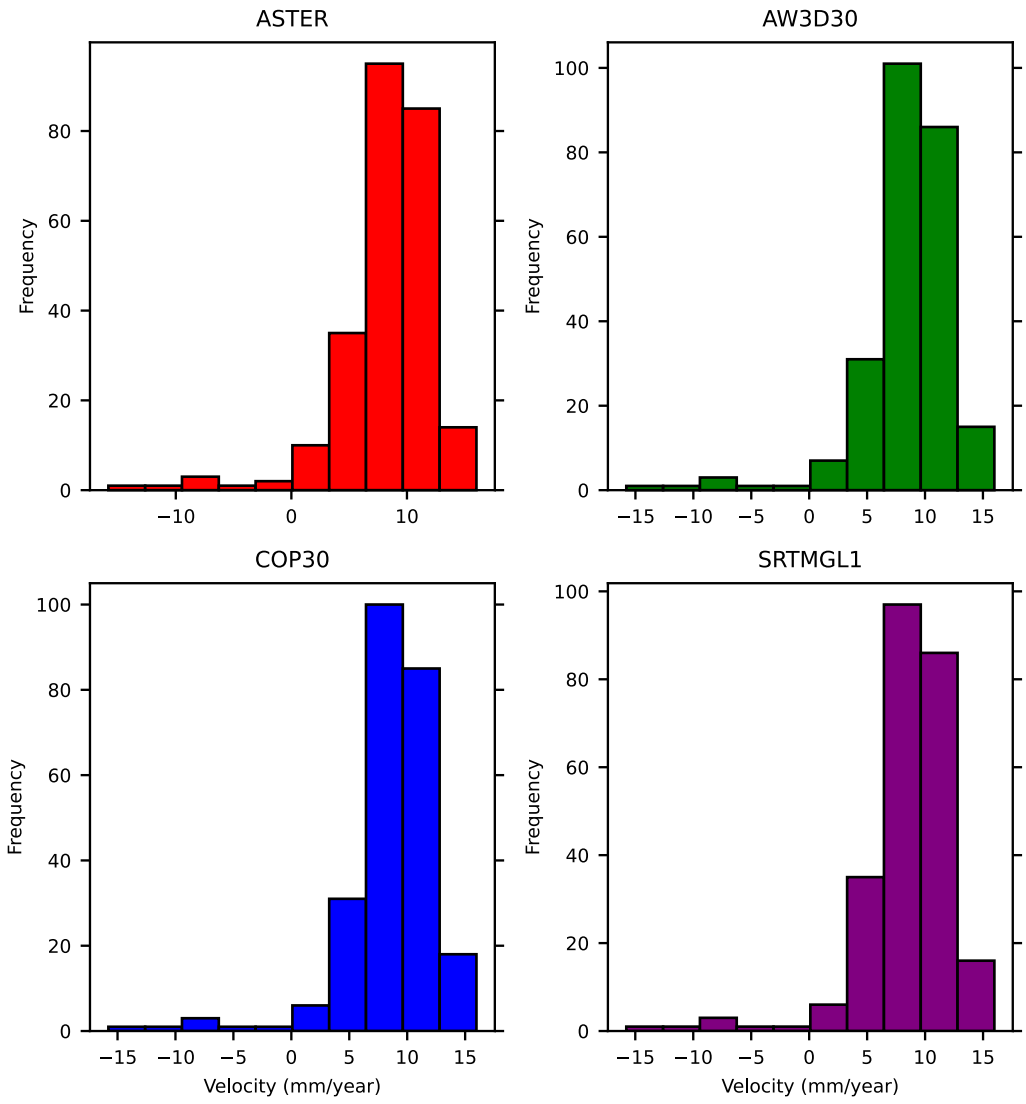


Figure H2: DEM velocity frequency distribution over 247 levelling points in Midvaal, South Africa.

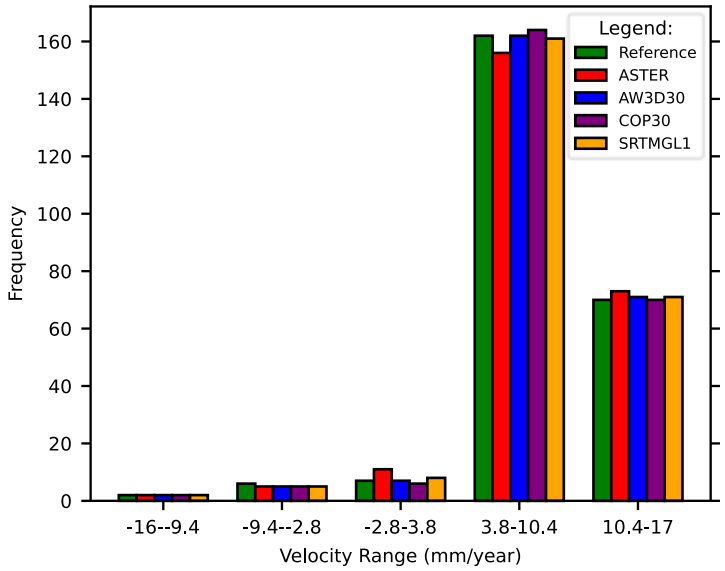


Figure H3: DEM velocity range frequency distribution over 247 levelling points in Midvaal, South Africa.

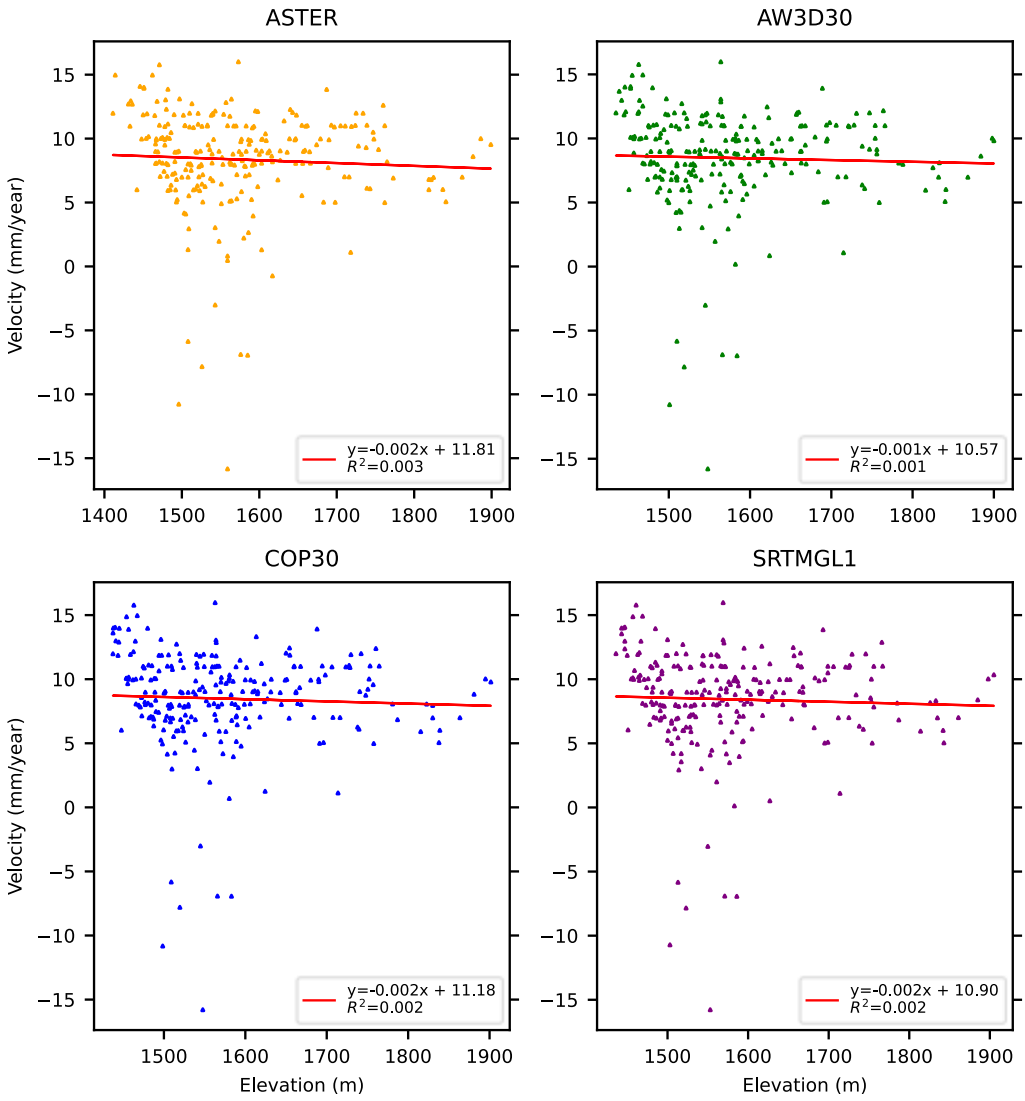
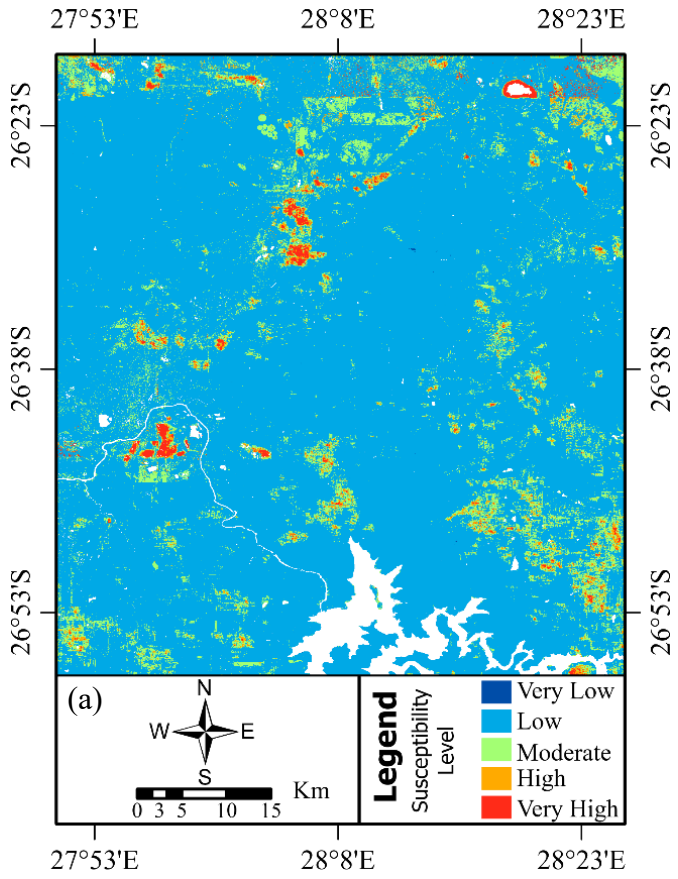


Figure H4: Relationship between elevation and velocity over 247 levelling points in Midvaal, South Africa.

Appendix J

Midvaal ground subsidence susceptibility map implied by Random Forest machine learning algorithm, susceptibility class statistics, and map accuracy statistics.



(b)

Metric	Value
MAE	3.621 mm/year
RMSE	5.039 mm/year
R2 score	84.98 %
AUC	0.89 %

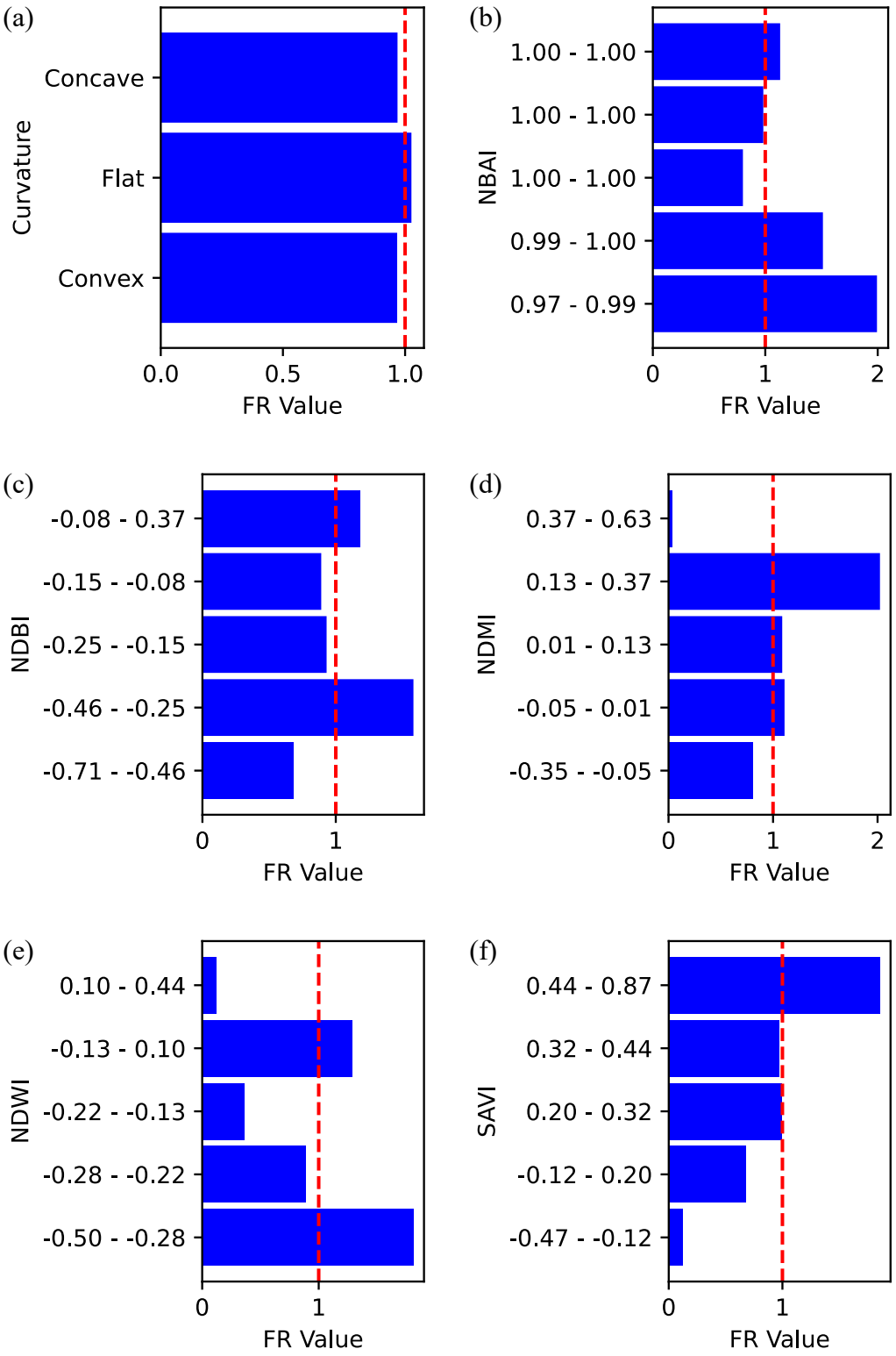
(c)

Class	Percentage (%)	Area (sq. km)
Very Low	0.1	4.13
Low	87.9	3,627.05
Moderate	9.4	387.88
High	1.7	70.15
Very High	0.9	37.13

Figure J1: Random Forest implied (a) ground subsidence susceptibility map, (b) map accuracy statistics, and (c) susceptibility class statistics.

Appendix K

The frequency ratio analysis results for removed correlated and ineffective redundant factors.


(a)
Curvature
Concave
Flat
Convex
0.0
0.5
1.0
FR Value
(b)
NBAI
1.00 - 1.00
1.00 - 1.00
1.00 - 1.00
0.99 - 1.00
0.97 - 0.99
0
1
2
FR Value
(c)
NDBI
-0.08 - 0.37
-0.15 - -0.08
-0.25 - -0.15
-0.46 - -0.25
-0.71 - -0.46
0
1
FR Value
(d)
NDMI
0.37 - 0.63
0.13 - 0.37
0.01 - 0.13
-0.05 - 0.01
-0.35 - -0.05
0
1
2
FR Value
(e)
NDWI
0.10 - 0.44
-0.13 - 0.10
-0.22 - -0.13
-0.28 - -0.22
-0.50 - -0.28
0
1
FR Value
(f)
SAVI
0.44 - 0.87
0.32 - 0.44
0.20 - 0.32
-0.12 - 0.20
-0.47 - -0.12
0
1
FR Value

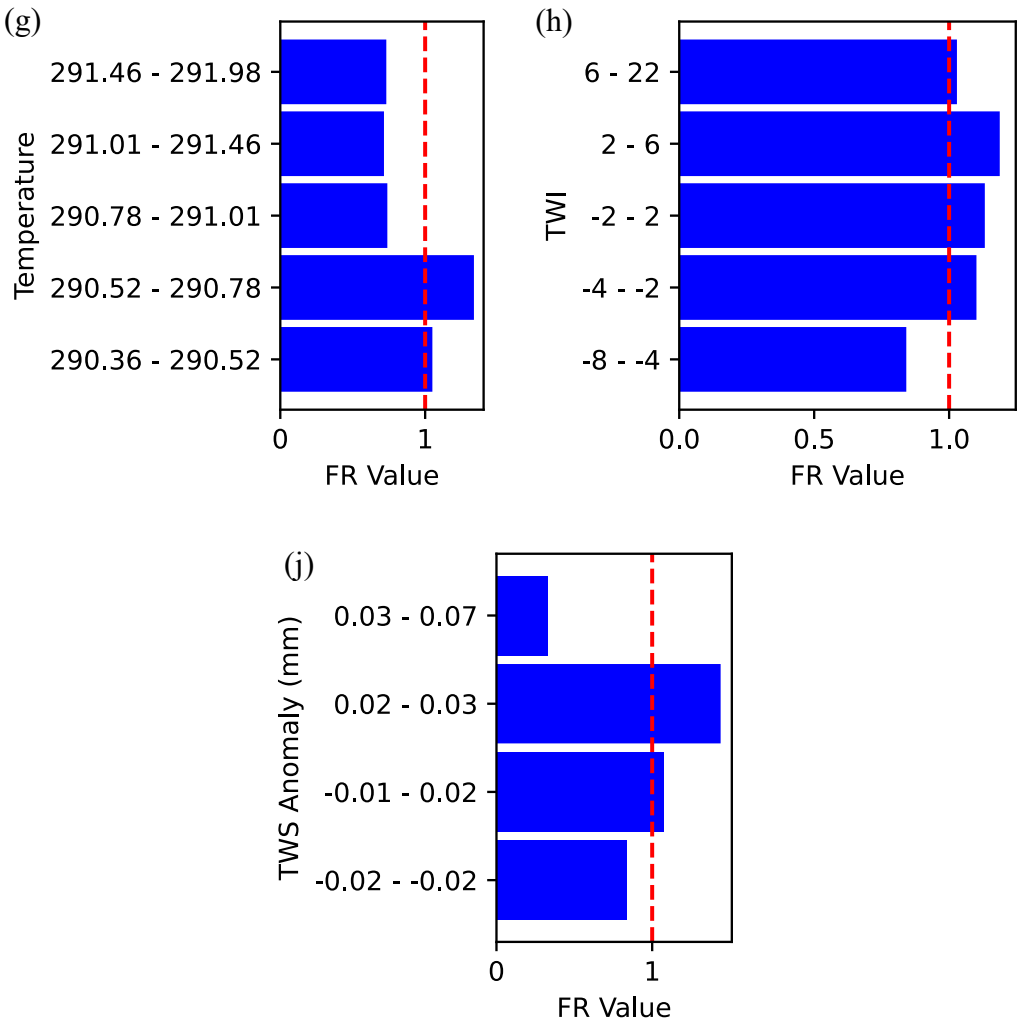


Figure K1: Frequency ratio analysis between ground subsidence and: (a) Curvature, (b) NBAI, (c) NDBI, (d) NDMI, (e) NDWI, (f) SAVI, (g) Temperature, (h) TWI, (j) TWS Anomaly.

Appendix L

The *InSAR LOS-ENU conversion* has two approaches: The first one being the robust approach requiring a minimum of three SAR acquisition geometries, each with distinct look angles (thus, unique incident angles).

$$\begin{bmatrix} D_{LOS,p} \\ D_{LOS,q} \\ D_{LOS,r} \end{bmatrix} = \begin{bmatrix} -sin\theta_p cos\beta_p & sin\theta_p sin\beta_p & cos\theta_p \\ -sin\theta_q cos\beta_q & sin\theta_q sin\beta_q & cos\theta_q \\ -sin\theta_r cos\beta_r & sin\theta_r sin\beta_r & cos\theta_r \end{bmatrix} \begin{bmatrix} D_E \\ D_N \\ D_U \end{bmatrix} + e \qquad \text{(L1)}$$

Where p, q, r represents SAR acquisition 1, 2 and 3 respectively. SAR acquisition 1 and 2 can be sentinel-1 ascending and descending acquisition respectively, and SAR acquisition 3 can be any other satellite acquisition such as ALOS PALSAR, Radarsat, TerraSAR-X etc.

The second approach requiring a minimum of two SAR acquisition geometries, and by neglecting the North-South deformation due to InSAR insensitivity in that direction as shown in Figure 1, the East-West and Up-Down velocities can be computed from two SAR acquisition (usually the ascending and descending acquisitions).

$$\begin{bmatrix} D_{LOS,p} \\ D_{LOS,q} \end{bmatrix} = \begin{bmatrix} -sin\theta_p cos\beta_p & cos\theta_p \\ -sin\theta_q cos\beta_q & cos\theta_q \end{bmatrix} \begin{bmatrix} D_E \\ D_U \end{bmatrix} + e \qquad \text{(L2)}$$

However, both the first and second approach for the LOS to ENU velocity conversion technique has limitations and challenges. One significant challenge is that the required three different SAR images cannot be captured at the same time; Thus, this relies on the assumption that between the three SAR image acquisition the change in deformation will be zero (no deformation between three different SAR acquisition). Another limitation in the second approach, is by neglecting North-South deformation which are significant and relevant to precise geodetic and geophysical research. Thus, for a comprehensive and robust comparison the conversion of ENU to LOS is employed for validation in this research.

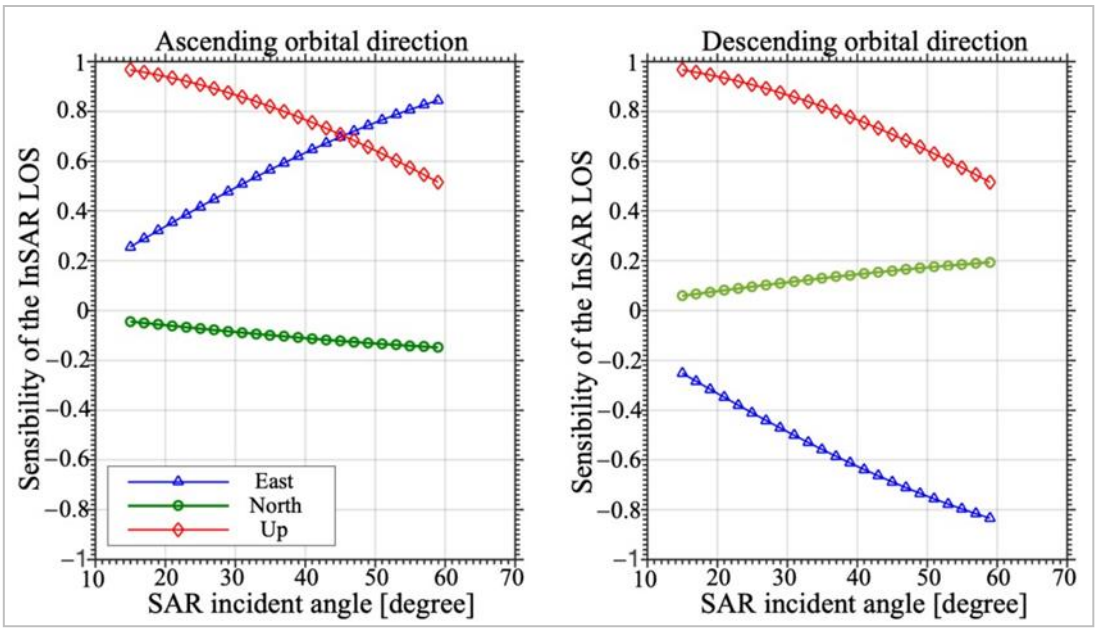


Figure L1: Sensibility of the radar LOS ground deformation to the SAR incident angles.

Uncertainty analysis helps in evaluating the robustness of the results, identifying potential biases, and assessing the limitations of InSAR time series analysis. It offers valuable information for decision-making processes and allows researchers to make more informed interpretations of the observed ground surface movements.

Appendix M

The goal of *horizontal datum unification* is to co-register or align the pixels of Digital Elevation Models, where misalignment in DEM pixels is suspected to arise from shifts in DEM origin due to differences in DEM origin definition during retrieval. Ideally, a DEM is defined with an

origin at x_0, y_0, for simplicity we can assume that all DEMs have complete global coverage spanning from 0 to 360 in absolute latitude and longitude. Then, the pixel index is an integer greater or equal to zero, computed by multiplying the latitude or longitude decimal degree value with their respective spatial resolution in decimal degrees, as shown by the equations below:

$$n_{col} = int\{(longitude - x_0) \div x_{resolution}\}$$
$$n_{row} = int\{(latitude - y_0) \div y_{resolution}\} \quad \text{(M1)}$$

Where n_{row} , n_{col} are the row and column index of a particular pixel in the raster array respectively, $y_{resolution}$, $x_{resolution}$ are the row and column spatial resolution of the raster respectively, and int{} is the integer operator that returns only the integer part of the expression.

From the above equations, we observe that changing the values of x_0 and y_0 implies changing the index of the point of interest. This means that one point on the ground could have two different indexes on two or more rasters with different raster origin coordinates. This introduces a horizontal shift, causing a horizontal datum defect and changing the absolute position of the pixels. The horizontal datum defect comprises both North-South and East-West shifts.

In the exploration of four DEMs, it was noted that AW3D30 uses zero-based remainder values for x_0 and y_0, while ASTER, COP30, and SRTMGL1 use 0.5- remainder based values for x_0 , y_0. This implies that ASTER, COP30, and SRTMGL1 would always be co-registered if the same coordinates are used to download the DEM. However, AW3D30 would be mis-shifted by Δx_0 and Δy_0 due to the 0.5 differences in x_0 and y_0.

To unify or co-register the DEMs, we used the 'raster_align' module implemented in the PyGEOS python library. This module requires ground elevation validation points and the rasters input datasets for processing. By applying the equations discussed earlier, the module recursively aligns DEMs with respect to each other, generating multiple combinations. Subsequently, all generated combinations are validated using ground elevation data, and the reference image that provides the best combination accuracy for all DEMs is selected as the reference grid or coregistered stack. The coregistration process involves sampling reference raster pixel coordinates from the secondary raster pixel arrays rather than shifting the pixels by some integer, as the observed differences were found to be half a pixel. In this study, selecting AW3D30 as the reference raster produced the best overall accuracy for the aligned stack.

Appendix N

The *stamps_nearby* module, implemented within the StoSAP processor, designed to extract point velocity data from the velocity CSV file exported by StaMPS.

```
# Importing Standard Library
import math as mt
import numpy as np
import pandas as pd
```

```
# Function to calculate the Haversine distance
def haversine(lat1, lon1, lat2, lon2):
    R = 6371000  # Radius of the Earth in meters
    lat1, lon1, lat2, lon2 = map(np.radians, [lat1, lon1, lat2, lon2])
    dlat = lat2 - lat1
    dlon = lon2 - lon1
    a = np.sin(dlat / 2)**2 + np.cos(lat1) * np.cos(lat2) * np.sin(dlon /
2)**2
    c = 2 * np.arctan2(np.sqrt(a), np.sqrt(1 - a))
    return R * c


# IDW interpolation
def idw_interpolation(df, val_key, target_lat, target_lon, power=2):
    # Calculate weights based on inverse distance
    df['weight'] = 1 / (df['distance'] ** power)

    # Calculate the weighted average value
    weighted_velocity = (df['weight'] * df[val_key]).sum() /
df['weight'].sum()
    return weighted_velocity


def stamps_nearby(latitude=None, longitude=None, df=None,
val_key='velocity', n_points=6):
    # Specified point
    target_lat = latitude
    target_lon = longitude

    df = df.copy()

    # Calculate the Euclidean distance
    df.loc[:,'temp_distance'] = np.sqrt((df['latitude'] - target_lat)**2 +
(df['longitude'] - target_lon)**2)

    # Sort the DataFrame by distance
    df_sorted = df.sort_values(by='temp_distance')

    # Get the six closest points
    closest_n_points = df_sorted.head(n_points)

    # Calculate the Haversine distance from selected points
    closest_n_points = closest_n_points.copy()
    closest_n_points.loc[:,'distance'] = closest_n_points.apply(lambda row:
haversine(target_lat, target_lon, row['latitude'], row['longitude']),
axis=1)

    # Sort the DataFrame by harvesine distance
    closest_n_points = closest_n_points.sort_values(by='distance')

    # Remove records containing nan/null values for required key field
(i.e., velocity key)
    closest_n_points = closest_n_points.dropna(subset=[val_key])

    # Compute the interpolated value for the target point
    interpolated_value = idw_interpolation(closest_n_points, val_key,
target_lat, target_lon)

    return interpolated_value
```

```
# Example Usage at Station HEID:
target_lat, target_lon = -26.508495, 28.373852
vHEID = stamps_nearby(latitude=target_lat, longitude=target_lon,
df=nd_data_sb_isce, n_points=25)
```

Appendix P

The *stamps_to_raster* module, implemented within the StoSAP processor, designed to convert StaMPS csv point velocity data to .tiff raster.

```
# Importing Standard Library
import numpy as np
from osgeo import gdal, osr
from collections import defaultdict


# Function to convert StaMPS csv to tiff raster.
def stamps_to_raster(df=None, output_raster='ps_isce.tif',
pixel_size=0.0002777):
    # Set the output raster file name
    output_raster = output_raster

    # Define the resolution of the raster
    pixel_size = pixel_size

    # Get the min and max coordinates
    min_lon, max_lon = df['longitude'].min(), df['longitude'].max()
    min_lat, max_lat = df['latitude'].min(), df['latitude'].max()

    # Define the raster size
    x_res = int((max_lon - min_lon) / pixel_size) + 1
    y_res = int((max_lat - min_lat) / pixel_size) + 1

    # Create the raster file
    driver = gdal.GetDriverByName('GTiff')
    raster = driver.Create(output_raster, x_res, y_res, 1,
gdal.GDT_Float32)

    # Set the geotransform
    geotransform = (min_lon, pixel_size, 0, max_lat, 0, -pixel_size)
    raster.SetGeoTransform(geotransform)

    # Create a spatial reference
    srs = osr.SpatialReference()
    srs.ImportFromEPSG(4326)  # EPS4326 for WGS84 GCS
    raster.SetProjection(srs.ExportToWkt())

    # Create a dictionary to hold the sum of velocities and count of
occurrences for each grid cell
    velocity_sum = defaultdict(float)
    velocity_count = defaultdict(int)

    # Populate the dictionaries with your data
    for _, row in df.iterrows():
        x = int((row['longitude'] - min_lon) / pixel_size)
        y = int((max_lat - row['latitude']) / pixel_size)
        velocity_sum[(x, y)] += row['velocity']
        velocity_count[(x, y)] += 1

    # Create an array with the data
```

```
    data = np.full((y_res, x_res), np.nan)  # Initialize with NaNs

    # Calculate the average velocity for each grid cell
    for (x, y), sum_velocity in velocity_sum.items():
        avg_velocity = sum_velocity / velocity_count[(x, y)]
        data[y, x] = avg_velocity

    # Write the array to the raster band
    band = raster.GetRasterBand(1)
    band.WriteArray(data)
    band.SetNoDataValue(np.nan)
    band.FlushCache()

    # Close the raster dataset
    raster = None
```

Appendix Q

The *h5_to_raster* module, implemented within the StoSAP processor, designed to convert MintPy velocity and standard error data to .tiff raster

```
# Importing Standard Library
import os
import h5py
import numpy as np
from osgeo import gdal, osr

# Function to Convert h5 file velocity and standard error into raster
(.tiff) format.
def h5_to_raster(file_path_grid=None, h5_file_path=None,
output_velocity_path=None, output_velocityStd_path=None):
    #file_path_grid is the raster with the desire coordinates sysytem,
commonly in EPSG4326
    dataset = gdal.Open(file_path_grid)

    # Check if the file exists
    if os.path.exists(h5_file_path):
        with h5py.File(h5_file_path, 'r') as h5_file:
            # Get a summary of the file's contents
            file_summary = {key: h5_file[key].shape for key in
h5_file.keys()}

            # Optionally, read a specific dataset (e.g., 'velocity')
            if 'velocity' in h5_file:
                velocity_data = h5_file['velocity'][:]

                # Get the geotransform and projection from the dataset
                geotransform = dataset.GetGeoTransform()
                projection = dataset.GetProjection()

                # Get the dimensions of the velocity data
                rows, cols = velocity_data.shape

                # Create a new raster dataset
                driver = gdal.GetDriverByName('GTiff')
                out_raster = driver.Create(output_velocity_path, cols,
rows, 1, gdal.GDT_Float32)

                # Set the geotransform and projection
                out_raster.SetGeoTransform(geotransform)
```

```
                out_raster.SetProjection(projection)

                # Write the velocity data to the raster band
                out_band = out_raster.GetRasterBand(1)
                out_band.WriteArray(velocity_data)

                # Flush the cache to ensure all data is written
                out_band.FlushCache()

                print(f"Velocity data saved as raster at
{output_velocity_path}")


                    # Optionally, read a specific dataset (e.g.,
'velocity')
            if 'velocityStd' in h5_file:
                velocityStd_data = h5_file['velocityStd'][:]

                # Get the geotransform and projection from the dataset
                geotransform = dataset.GetGeoTransform()
                projection = dataset.GetProjection()

                # Get the dimensions of the velocity data
                rows, cols = velocityStd_data.shape

                # Create a new raster dataset
                driver = gdal.GetDriverByName('GTiff')
                out_raster = driver.Create(output_velocityStd_path, cols,
rows, 1, gdal.GDT_Float32)

                # Set the geotransform and projection
                out_raster.SetGeoTransform(geotransform)
                out_raster.SetProjection(projection)

                # Write the velocity data to the raster band
                out_band = out_raster.GetRasterBand(1)
                out_band.WriteArray(velocityStd_data)

                # Flush the cache to ensure all data is written
                out_band.FlushCache()

                print(f"Velocity data saved as raster at
{output_velocityStd_path}")
    else:
        print("File does not exist.")
```