The influence of floods-recharged soil moisture on tree biochemical and biophysical features in Mbire and Muzarabani semi-arid lands in northern Zimbabwe

by

Ezra Pedzisai (218085751)

Submitted in fulfilment of the academic requirements of

Doctor of Philosophy

in Environmental Science

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg

South Africa

Supervised by
Professor Onisimo Mutanga
Professor John Odindi

January 2025

PREFACE

The research contained in this thesis was completed by the candidate while based in the Discipline of Geography, School of Agricultural, Earth and Environmental Sciences, of the College of Agriculture, Engineering and Science, University of KwaZulu Natal, Pietermaritzburg, South Africa, from June 2018 to December 2024, under the supervision of Professor Onisimo Mutanga and Professor John Odindi. The research was financially supported by the South Africa Government through the National Research Foundation of South Africa (NRF) Research Chair in Land Use Planning and Management (Grant Number: 84157) provided by Profesor O. Mutanga.

I declare that the work reported in this thesis has never been submitted in any form to any other institution and that results reported are mine except where due acknowledgments are made.

[redacted]

Signed: Ezra Pedzisai

(Student)

Date: 19/01/2025

As the candidate's supervisors, we certify the above statement and have approved this thesis for submission.

[redacted]

Signed: Professor Onisimo Mutanga

(Supervisor)

Date: 19/01/2025

[redacted]

Signed: Professor John Odindi

(Supervisor)

Date: 19/01/2025

DECLARATION 1 – PLAGIARISM

I, Ezra Pedzisai, declare that:

1. The research reported in this thesis, except where otherwise indicated, is my original research.
2. This thesis has not been submitted for any degree or examination at any other institution.
3. This thesis does not contain other people's data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted:
   a. Their words have been rewritten and the general information attributed to them has been referenced.
   b. Where their exact words have been used, their writing has been placed in italics inside quotation marks, and referenced.
5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the references section.

Date: 19/01/2025

DECLARATION 2 – PUBLICATIONS

I was the principal investigator and corresponding author in all the papers listed below. I conceptualised the research, designed field protocols, collected field data, downloaded and preprocessed satellite images, built and executed machine/deep learning algorithms, undertook data analysis, evaluated model performance, presented results and wrote all draft manuscripts with guidance from supervisors and co-authors as expressly stated in each publication.

CHAPTER 2: LITERATURE REVIEW ON THE USE OF REMOTE SENSING INDICES TO UNDERSTAND FLOOD-RECHARGED SOIL MOISTURE INFLUENCE ON TREES IN SEMI-ARIDS LANDS

Pedzisai, E., Mutanga, O., Odindi, J. and Mushore T.D., 2022. The use of remote sensing indices to understand flood-recharged soil moisture impacts on trees in semi-arid floodplains: A review. *Ecohydrology***.** *15*(7), e2460. **https://doi.org/10.1002/eco.2460**

CHAPTER 3: CHANGE DETECTION AND THRESHOLD-BASED ENSEMBLE OF SCENARIOS PYRAMID FOR FLOOD EXTENT MAPPING USING RADAR DATA

Pedzisai, E., Mutanga, O., Odindi, J., and Bangira, T., 2023. A novel change detection and threshold-based ensemble of scenarios pyramid for flood extent mapping using Sentinel-1 data. *Heliyon,* 9, e13332, 1-13. **https://doi.org/10.1016/j.heliyon.2023.e13332**.

CHAPTER 4: THE USE OF HYBRID DEEP LEARNING FRAMEWORK TO MODEL FLOOD-RECHARGED SOIL MOISTURE FEATURES IN SEMI-ARID LANDS

Pedzisai, E., Odindi, J. and Mutanga, O. Use of long short-term memory autoencoder deep learning to model flood-recharged soil moisture in a semi-arid floodplain. *Under prepararti*on.

CHAPTER 5: SHORT-TERM INFLUENCE OF FLOOD-RECHARGED SOIL MOISTURE ON TREE BIOCHEMICAL CHARACTERISTICS IN SEMI-ARID LANDS

Pedzisai, E., Odindi, J. and Mutanga, O. Machine learning modelling influence of floods-recharged soil moisture on leaf chlorophyll content of *Ziziphus mauritiana* tree in semi-arid floodplain of Northern Zimbabwe. *Under prepararti*on.

CHAPTER 6: LONG-TERM INFLUENCE OF FLOOD-RECHARGED SOIL MOISTURE ON TREE BIOPHYSICAL CHARACTERISTICS IN SEMI-ARID LANDS

Pedzisai, E., Mutanga, O., and Odindi, J. Long-term influence of floods-recharged soil moisture on tree biophysical characteristics in semi-arid lands. *Under prepararation.*

Signed: Ezra Pedzisai

Date: 19 /01/2025

ABSTRACT

Ecological influence of floods-recharged soil moisture (FRSM) on tree growth in tropical semi-arid lands remains poorly understood, hence, the study focused on tree biochemical and biophysical characteristics in northern Zimbabwe. It modelled flood extent using a novel Ensemble of Scenarios Pyramid (ESP) to facilitate adoption of an experimental design with flooded (experiment) and on-flooded (control) set-ups enabling assessment of FRSM anomaly, lag and memory applying a novel hybrid deep learning long short-term memory autoencoder algorithm (LSTMAE) on Sentinel-1 SAR data. Study assessed short-term influence of FRSM on biochemical leaf chlorophyll content (LCC) and evaluated long-term biophysical characteristics using diameter at breast height (DBH), canopy diameter, and tree height for multipurpose *Z. mauritiana*. Field data included flood extent boundary, tree LCC, tree height, DBH, and canopy diameter. Since floods occur during cloudy conditions rendering passive remote sensing data inapplicable, Sentinel-1 SAR temporal data (2017 - 2019) were downloaded. The ESP produced a more accurate flood extent map than the six base scenarios with overall accuracy (93.204%), F1-score (0.927), Matthews' correlation coefficient (0.871), Recall (0.870), Intersect over Union (0.865), and Kappa (0.864). The LSTMAE detected positive (wet) FRSM anomaly, depicting a 1-week lag that occurred starting 2017/02/02 and initiating a 70-day FRSM memory inside the flooded unlike non-flooded zones. On the short-term, Random Forest machine learning indicated three flood-related spatial variables namely distance to river, distance to floodplain and floodplain location ranked as the most important predictors to determine the short-term FRSM influence on *Z. mauritiana* LCC. Thus, flooded zone (hence FRSM) influenced significantly higher LCC as compared to the non-flooded zone. On the long-term, flooded trees were significantly bigger compared to those in non-flooded areas in canopy diameter ($p<0.001$), tree height ($p<0.05$) and DBH ($p<0.001$) using the Levene's test. Similarly, the Mann-Whitney-Wilcoxn test also showed significant differences in canopy size ($p<0.001$), tree height ($p<0.001$) and DBH ($p<0.01$). Conclusively, the FRSM positively influences soil water recharge in semi-arid lands supporting flooded tree growth to confirm deep infiltration. Results inform soil moisture and *Z. mauritiana* management plans in semi-arid lands in view of important deforestation triggered by overexploitation owing to its multiple uses.

EXTENDED ABSTRACT

In dry lands characterised by scarce precipitation, floods uniquely facilitate soil moisture recharge through deep infiltration. However, the ecological influence of flood-recharged soil moisture (FRSM) on tree growth in semi-arid lands remains poorly understood. To address this knowledge gap, the aim of the study was to assess the influence of floods-recharged soil moisture on tree biochemical and biophysical characteristics in a semi-arid landscape of northern Zimbabwe which was split into the following specific objectives. Firstly, the study sought to assess the utility of remote sensing indices in understanding the nexus between FRSM with tree biochemical and biophysical features in tropical semi-arid floodplains. Secondly, the study sought to model flood extent mapping in semi-arid floodplains in northern Zimbabwe using Sentinel-1 SAR data. Thirdly, the study sought to assess the utility of deep learning long short-term memory autoencoder algorithm applied on Sentinel-1 SAR data to model FRSM features in semi-arid floodplains in northern Zimbabwe. Fourthly, the study sought to assess the influence of FRSM on short-term biochemical properties of *Z. mauritiana* tree in semi-arid floodplain in northern Zimbabwe. Finally, the study sought to evaluate the long-term influence of FRSM on *Z. mauritiana* tree biophysical characteristics in semi-arid lands in northern Zimbabwe. Therefore, to understand the influence of FRSM on trees, this study compared a multipurpose Musawu (Shona) or Jujube (English) (*Ziziphus mauritiana*) tree species located inside against outside of Mbire and Muzarabani flood-prone semi-arid lands of northern Zimbabwe. The study used an experimental design, flooded constituted the experimental set-up while non-flooded the control set-up for soil moisture, leaf chlorophyll content and tree size variables to enable respective hypotheses of influence of FRSM were tested.

Primary and secondary data were collected through fieldwork and downloaded from online repository respectively. Field data measurements included flood extent boundary, tree leaf chlorophyll content measured using chlorophyll meter, tree height measured using a Haglöf Vertex Laser Geo hypsometer, diameter at breast height (DBH) measured using diameter tape and canopy diameter estimated using tape measure respectively. Since floods often occur during cloudy conditions, passive remotely sensed secondary data are commonly inapplicable, hence this study undertook flood extent mapping that used temporal synthetic arperture radar (SAR) data. Firstly, to accurately map flood extent, we innovated an Ensemble of Scenarios Pyramid which is based on change detection and thresholding, utilising a normalized difference flood index (NDFI) framework with Sentinel-1 SAR data. The flood extent map was necessary to spatially

discriminate flooded from non-flooded soil moisture, short-term leaf chlorophyll content and long-term tree biophysical characteristics. Secondly, the study innovated a hybrid deep learning Long Short-Term Memory-Autoencoder framework that used temporal Sentinel-1 SAR data to model soil moisture anomaly, lag and memory features compared between flooded and non-flooded locations. Thirdly, short-term post-flood leaf chlorophyll content for flooded tree samples were compared with non-flooded samples using machine learning. Finally, the long-term influence of FRSM on biophysical characteristics was evaluated using canopy diameter, tree height and diameter at breast height (DBH) as proxies of tree growth.

The novel Ensemble of Scenarios Pyramid produced a more accurate flood extent map as compared to all the base NDFI scenarios using six metrics (overall accuracy = 93.204%; F1-score = 0.927; Matthews' correlation coefficient = 0.871; Recall = 0.870; Intersect over Union = 0.865; Kappa = 0.864). The Long Short-Term - AutoEncoder detected positive (wet) FRSM anomaly, with one week lag that occurred starting from the second of February 2017, and initiating a 70-day soil moisture memory inside the flooded as compared to the non-flooded zones. On the short-term, Random Forest machine learning rejected 43 as unimportant and accepted 16 variables as important to determine leaf chlorophyll content. The flood-related three spatial variables namely distance to river, distance to floodplain and floodplain location were ranked as the most important predictors to determine the short-term FRSM influence on *Z. mauritiana* leaf chlorophyll content. Thus, on the short-term, the flooded zone (hence FRSM) influenced higher leaf chlorophyll content as compared to the non-flooded zone. On the long-term, flooded trees inside the floodplain were significantly bigger compared to those in non-flooded areas in canopy diameter ($p<0.001$), tree height ($p <0.05$) and DBH ($p<0.001$) using the Levene's test. Similarly, the Mann-Whitney-Wilcoxn test also showed significant differences in canopy size ($p < 0.001$), tree height ($p < 0.001$) and DBH ($p < 0.01$).

Inside flooded locations mapped with the novel Ensemble of Scenarios Pyramid, the innovated Long Short-Term-AutoEncoder modelled that FRSM clearly depicted a minimum of two months longer soil moisture memory than outside. Ultimately, this longer FRSM memory supply soil water for longer to flooded trees than in non-flooded areas, hence the higher leaf chlorophyll content for the former rather than the latter. Consequently, in the long-term, flooded trees inside the floodplain grew bigger than outside as the three selected growth biophysical proxies confirmed using non-parametric tests of differences. In conclusion, the FRSM positively influences soil water recharge in semi-arid lands, which ultimately supports flooded tree growth, thereby confirming the deep infiltration concept noted in the literature review. These results uniquely inform soil

moisture and related *Z. mauritiana* tree management plans in semi-arid lands in view of the current deforestation triggered by its overexploitation by both humans and animals owing to its multiple uses.

Keywords

Ensemble of Scenarios Pyramid; Flood extent map; Flood-recharged soil moisture; Leaf chlorophyll content; Long Short-Term Memory-Autoencoder; Machine learning; *Z. mauritiana* tree.

ACKNOWLEDGEMENTS

I am greatly indebted to my supervisors; Professor Onisimo Mutanga and Professor John Odindi; for their dedicated professional guidance and scientific grooming that guided me immensely in this research. They helped me to conceptualize, analyze, model, and revise all chapters in this thesis. Their guidance informed my data collection protocols, analytical framework and results with noted innovations that contribute to science. I am, also indebted to my co-authors, Professor Terrence D. Mushore, and Dr. Tsitsi Bangira for Chapters 2 and 3, respectively. I am also grateful to the editors and anonymous reviewers for their constructive comments in the published chapters.

I acknowledge the National Research Foundation Chair of Landuse Management of South Africa (Grant No: 84157), Professor Mutanga, for financial assistance. I appreciate the Government of the Republic of South Africa for granting me a study permit and fee remission for my postgraduate studies tenable at the UKZN, Pietermaritzburg Campus. I am grateful to Marsha Manjoo and staff in the School of Agricultural, Earth and Environmental Sciences and Shanita Ramiroop and staff in the Discipline of Geography that provided laboratory resources, office, and field equipment. Thanks to Tamylin Skye for administrative and registration assistance. Administrative and welfare assistance provided by Andile Mshengu is greatly appreciated. I acknowledge technical support by Mr Brice Gijsbertsen and Mr Donovan deVos.

I acknowledge the developers of libraries in Anaconda (Python) relevant Keras models and layers in TensorFlow backend executed in Spyder IDE whose packages where useful in this research. I acknowledge the online user communities, specifically tutorials by Dr Snrivanus Bhattiprolu which helped me to innovate the dualized soil moisture anomaly detection, lag and memory modelling that informed the foundation of the model applied in this research. I acknowledge the generosity of the QGIS, and R-Project developers for free software and their user communities for guidance and notes. I thank the European Space Agency Copernicus Open Access Hub for Sentinel-1 data and Sentinel Application Platform (SNAP) software, and NASA for elevation data.

I am grateful to friends in PhD Room 34, Scottsville, Pietermaritzburg, for cross-fertilising ideas, experiences, and lighter moments during research. They were my pillar of strength on this academic journey. Thanks for the advice from Dr Pedzisai Kowe, and Dr Mbulisi Sibanda for reviewing my initial research objectives, Dr Odebiri Israel Omasalewa, Dr Royimani Lwando, Dr Mtembisi Mngadi, Dr Samuel Kumbula, and Dr Collins Matiza for the encouragement. Special mention goes to Dr Trylee N. Matongera for arranging my accommodation and facilitating my

communication including during the COVID-19 pandemic. I acknowledge feedback from all participants in the UKZN-facilitated online seminars co-ordinated by Professor O. Mutanga, including contributions from Professor Timothy Dube.

I am greatly indebted to the Mbire and Muzarabani communities and leadership, and World Vision during field guidance. I also thank my colleagues at Bindura University of Science Education (BUSE) and staff in the Faculty of Science and Engineering, School of Geosciences, Disasters and Development who took extra duties during my stay in South Africa. Thank you to Zion Apostolic Faith Mission Church (ZAFMC) for supporting me financially, spiritually, morally, and providing the car I used to during data collection into the study area. Many heart-felt thanks for all leadership in ZAFMC for leading the organization during my studies. In the ZAFMC, SHP Sam T.K. Mawisire, Engineer Culverwell and his spouse Pride Dube for the hospitality, financial support and logistical arrangements. I acknowledge my family for supporting my studies especially my spouse Sununguravi Charamba-Pedzisai, and the children. I am indebted to Silas Pedzisai for driving me during fieldwork. Many thanks to SkyComp (Pvt) Ltd directors; O. Mhande, E. Magodhi and S. Dzipai for the financial, and moral support. Special mention to the Mr S. and Mrs Dzipai as well as Mr J. and Mrs Temezani for providing me the two laptops that I used to complete my research. Above all, thanks to everyone who contributed to my studies in any way.

Most of all, I would like to give all thanks to the Almighty God. I give thanks and glorify the Lord Jesus Christ for providing all the above-noted persons and resources that enabled me to complete this research studies.

DEDICATION

This thesis is dedicated to my;

spouse - Sununguravi Charamba-Pedzisai,

mother - Sarah Matongo-Pedzisai,

in memory of my late father – Archbishop Dorias P.A.J. Pedzisayi,

sons - Tawananyasha Philip and Takudzwanashe,

daughters - Tatenda, Rutendo, Rudairo, and Rukudzo and,

the entire Pedzisai family.

TABLE OF CONTENTS

CONTENTS **PAGE**

PREFACE ..... **i**

DECLARATION 1 – PLAGIARISM ..... **ii**

DECLARATION 2 – PUBLICATIONS ..... **iii**

ABSTRACT ..... **v**

EXTENDED ABSTRACT ..... **vi**

ACKNOWLEDGEMENTS ..... **ix**

DEDICATION ..... **xi**

TABLE OF CONTENTS ..... **xii**

LIST OF TABLES ..... **xvii**

LIST OF FIGURES ..... **xviii**

LIST OF APPENDICES ..... **xxii**

LIST OF ACRONYMS ..... **xxiii**

CHAPTER 1: GENERAL INTRODUCTION **1**

1.1 Background **1**

1.2 Aim **6**

1.2.1 Specific Objectives 6

1.3 Description of the study area **6**

1.4 Organization of the thesis **9**

1.5 References **11**

CHAPTER 2: THE USE OF REMOTE SENSING INDICES TO UNDERSTAND FLOOD-RECHARGED SOIL MOISTURE INFLUENCE ON TREES IN SEMI-ARID LANDS: A REVIEW **18**

Abstract **19**

2.1. Introduction **19**

2.2. Role of remote sensing indices for FRSM and trees in semi-arid lands **24**

2.2.1 Application of remote sensing indices for FRSM retrieval 24

2.2.1.1 Passive optical and thermal remote sensing indices applications for FRSM 25

2.2.1.2 Microwave remote sensing indices applications for FRSM retrieval 28

2.2.2 Application of remote sensing indices for estimating tree parameters 30

2.2.3. Linking floods and FRSM to tree biophysical parameters using remote sensing indices ... 34

2.2.3.1 Short-term impacts of FRSM on trees species biochemical characteristics 36

2.3. Challenges in application of remote sensing indices on the impact of FRSM on trees . 40

2.4 The future of remote sensing indices to assess impacts of FRSM on riparian trees.....42
2.5. Conclusions.....47
2.6 References.....47
CHAPTER 3: CHANGE DETECTION AND THRESHOLD-BASED ENSEMBLE OF SCENARIOS PYRAMID FOR FLOOD EXTENT MAPPING USING SYNTHETIC APERTURE RADAR (SAR) DATA.....65
Abstract.....66
Keywords.....66
3.1 Introduction.....66
3.2 Materials and methods.....70
3.2.1 Study Area.....70
3.2.2 Data.....71
***3.2.2.1 In situ flood extent data collection*.....71**
***3.2.2.2 Satellite image acquisition and processing*.....72**
3.2.3 Overview of the Ensemble of Scenarios Pyramid modeling.....72
***3.2.3.2 The Ensemble of Scenarios Pyramid algorithm*.....74**
3.2.4 The Ensemble of Scenarios Pyramid (ESP) performance validation metrics.....75
3.3 Results.....76
3.3.1 Comparing VV with VH polarized SAR data for flood extent monitoring.....76
3.3.2 The ESP flood extent classification.....77
3.3.3 The ESP flood extent map validation.....79
***3.3.3.1 The Scenario Ensemble Agreement using IoU and MCC metrics*.....80**
3.4 Discussion.....85
3.4.1 The ESP flood extent map.....85
3.4.2 The ESP Evaluation.....86
3.4.3 Single polarization SAR utility for monitoring flood extent.....87
3.5 Conclusion.....87
3.6 References.....88
CHAPTER 4: THE USE OF HYBRID DEEP LEARNING FRAMEWORK TO MODEL FLOOD-RECHARGED SOIL MOISTURE FEATURES IN SEMI-ARID LANDS.....95
Abstract.....96
4.1 Introduction.....96
4.2 Materials and methods.....100
4.2.1 The study area.....100
4.2.2 Data collection.....101
***4.2.2.1 Field and remotely sensed data*.....101**

4.2.3 Methods ... 102
4.2.3.1 The modelling approach ... 102
- *Description of the LSTMAE model structure* ... 103

4. 2.3.2 Model building and parameterization ... 104
4.2.3.3 Model implementation ... 105
4.2.3.4 LSTMAE model evaluation ... 108
4.3 Results ... 109
4.3.1 The LSTMAE model ... 109
4.3.2 FRSM anomaly, lag and memory detection using LSTMAE ... 111
4.3.3 Evaluation of the LSTMAE performance ... 111
4.3.4 The exploratory data analysis results ... 114
4.4 Discussion ... 115
4.4.1 Anomaly detection using LSTMAE algorithm ... 116
4.4.2 Estimating soil moisture lag and memory using LSTMAE algorithm ... 117
4.4.4 Inferring FRSM using LSTMAE algorithm ... 118
4.4.5 The utility of polarization mode to infer FRSM ... 119
4.5 Conclusions and prospects ... 120
4.6 References ... 120
CHAPTER 5: SHORT-TERM INFLUENCE OF FLOOD-RECHARGED SOIL MOISTURE ON LEAF CHLOROPHYLL CONTENT OF *ZIZIPHUS MAURITIANA* TREE IN SEMI-ARID LANDS ... 129
Abstract ... 130
5.1 Introduction ... 130
5.2 Materials and Methods ... 134
5.2.1 Study area ... 134
5.2.2 Data ... 134
5.2.3 Data Collection ... 135
5.2.4 Preprocessing ... 136
5.2.5 Modelling ... 137
5.2.5.1 Binary Logistic Regression machine learning ... 137
5.2.5.2 Boruta function variable importance projection ... 137
5.2.5.3 Random Forest machine learning classification ... 138
5.2.5.4 Linear Regression Modelling ... 138
5.2.5.5 Variability of LCC using ANOVA ... 139
5.2.6 Tree aggregated LCC against spatial variables correlation matrix ... 140
5.2.7 Model performance evaluation metrics ... 140

5.3 Results **140**
5.3.1 Explanatory variables of LCC using floodplain binary class 140
5.3.2 Variable importance projection for LCC modelling 140
5.3.3 Variability of LCC between inside and outside floodplain using ANOVA 145
5.3.4 The effect of data augmentation using statistical derivation 148
5.3.5 Model performance evaluation 148
5.4 Discussion **150**
5.4.1 Variables influencing LCC in a semi-arid floodplain 150
5.4.2 Variable importance relating to LCC in semi-arid floodplain 150
5.4.3 Higher LCC inside than outside the semi-arid floodplain 151
5.4.4 Statistical data augmentation of LCC in semi-arid floodplains 152
5.5 Conclusion **154**
5.6 References **155**
CHAPTER 6: LONG-TERM INFLUENCE OF FLOOD-RECHARGED SOIL MOISTURE ON TREE BIOPHYSICAL CHARACTERISTICS IN SEMI-ARID LANDS 168
Abstract **169**
6.1 Introduction **170**
6.2. Materials and methods **173**
6.2.1 Study sites 173
6.2.2 Field Data 173
6.2.4 Difference of biophysical characteristics between flooded and non-flooded trees 176
6.2.5 Multiple linear regression model of *Z. mauritiana* tree growth on spatial predictors 178
6.3. Results **178**
6.3.1 Diagnostic plots for the three biophysical characteristics 179
6.3.2 Comparison of *Z. mauritiana* tree biophysical characteristics by flood status 181
6.3.2.1 Canopy diameter differences between flooded and non-flooded Z. mauritiana trees 181
6.3.2.2 The tree height differences between flooded and non-flooded *Z. mauritiana* trees 182
6.3.3.1 Pairwise linearity test 186
6.4 Discussion **188**
6.4.1 Overall influence of FRSM on tree biophysical characteristics in semi-arid areas 188
6.4.1.1 The FRSM influence on canopy diameter 189
6.4.1.2 Influence of FRSM on tree height in semi-arid areas 190
6.4.1.3 The influence of FRSM on DBH in semi-arid areas 191
6.4.2 Relationship between tree biophysical characteristics and soil moisture predictors 191
6.5 Conclusion **192**

6.6 Recommendations **192**
6.7 References **192**
CHAPTER 7: SYNTHESIS, CONCLUSION AND RECOMMENDATIONS **202**
7.1 Synthesis **202**
7.2 Conclusion **203**
7.3 Recommendations **205**
8. APPENDICES **206**

LIST OF TABLES

Table **Caption** **Page**

Table 2.1: Passive remote sensing indices for soil moisture characterization ……..…26

Table 2.2: Specifications of selected global satellite remote sensing-based soil moisture retrieval missions. …….28

Table 3.1: Scenario dates, and respective temporal stacks based on vertical-horizontal (VH) and vertical-vertical (VV) polarization modes. …….……...…..74

Table 3.2: The Ensemble of Scenarios Pyramid (ESP) summary of validation metrics and the respective mean by level. …….80

Table *3*.3: The level-based intersect over union (IoU) and Matthews Correlation Coefficient (MCC) validation metrics at the base (E1), centre (E2) and pinnacle (E3). ……... 83

Table 4.1: Thresholds derived from train data to determine anomalies on test data ……...… 110

Table 4.2: The LSTM-AE model MAE and RMSE by polarisation and location. ……... 112

Table 4.3: LSTM-AE model evaluation ME metric (train: n = 61, test: n = 30) ……...…. 113

Table 5.1: Data collection method, data, variable(s) and description of the data (n = 119) …. 134

Table 5.2: LCC derived on AOC, ETL and HOC for locations in and out of the floodplain .. 145

Table 5.3: The effect of ETL effect on leaf chlorophyll content …….….. 147

Table 5.4: Significant test results from linear regression using four spatial variables …….. 149

Table 6.1: Comparing flooded (195) with non-flooded (171) Z. mauritiana trees …… 181

Table 6.2: Three predictors' model compared with bootstrapping coefficients ……. 185

Table 6.3: Statistical significance test for heteroscedasticity, normality, linearity using single predictor linear regression models ……....…….. 187

Table 6.4: Linear regression models of biophysical features on distance for flooded, non-flooded and both (global) for Z. mauritiana tree measurements (n = 366) …… 188

LIST OF FIGURES

Figure: **Caption** **Page**

Figure 1.1: Conceptual framework of the the study……3

Figure 1.2: Map locating Mashonaland Central Province (a), zoomed-in with stripe-shaded Mbire and Muzarabani (b) constituting Site 1 (c) and Site 2 (d) respectively showing flooded and non-flooded tree samples, floodplain and drainage in semi-arid northern Zimbabwe. ……9

Figure 2.1: Conceptualisation of soil moisture (SM) depth for locations outside flood zone (A) and inside flood zone (A + B) recharged through shallow infiltration by precipitation only and precipitation plus deep infiltration, respectively ……35

Figure 2.2: Conceptualisation of short-term precipitation only (a) versus flood plus precipitation (b) recharged soil moisture ……36

Figure 2.3: Conceptualisation of long-term effect of floods on structure of vegetation height, canopy size and diameter at breast height ……39

Figure 3.1: The inserts for Africa (a), Zimbabwe (b) and the Mbire study area (c) showing elevation and drainage, with flooded and not flooded field sample points used as ground truth for validation ……71

Figure 3.2: Methodological framework showing, a) calculation of six normalized difference flood index (NDFI) scenarios using VV and VH polarized 2 (bi-temporal), 5 (penta-temporal) and 10 (deca-temporal) images and, b) the Ensemble of Scenarios Pyramid (ESP) pairwise union of six VV and VH (base) into three centre scenarios (i.e. 7-9) and pinnacle (Scenario10). ……72

Figure 3.3: Pre-flood (09-01-2017), flood (02-02-2017) and post-flood (26-02-2017) extent maps with panels vertical-horizontal (VH) (a, b, c) and vertical-vertical (VV) (d, e, f) ……77

Figure 3.4: Flood extent maps at the base (E1), centre (E2) and pinnacle (E3) levels of the Ensemble Scenario Pyramid (ESP) with converging arrows showing union function calculation. ……78

Figure 3.5: Results of the intersect (a, b, c) and union (d, e, f) of the VV and VH maps from E1 used to calculate the intersect over union (IoU) (g, h, i) ……81

Figure 3.6: Maps showing the comparison of flood detection using vertical-vertical (VV) versus vertical-horizontal (VH) polarization for bi-temporal (a), penta-temporal (b) and deca-temporal (c) scenarios respectively …….....82

Figure 3.7: The Ensemble of Scenarios Pyramid (ESP) minimum, mean and maximum by level (left) and change (right). The Y-axis shows a standardized metric scale (left) and the magnitude of change (right). …….……84

Figure 4.1: Map of Africa (a), Zimbabwe (b) and location of Chitsungo study area (c) showing elevation and drainage network ……...101

Figure 4.2: An anomaly detection LSTMAE model with observed sequence (input) into dimensionality reduction two LSTM layers (encoder), a latent storage (code), and decompressing (decoder) that regenerates output sequence in two LSTM layers while the number of units at each part is shown respectively ……...…104

Figure 4.3: Graphical definition of three soil moisture features namely anomaly, lag and memory (a) based on test mean error metric and the empirical determination of the respective thresholds (b) based on the trainME and testME distribution …….……...105

Figure 4.4: The translation of the ME (a) into positive, normal and negative thresholds (b), determination of dates of anomalies, lag and memory (c) features, respectively …108

Figure 4.5: The thresholds for VV (a and b) versus VH (e and f), respective anomaly detection, lag and memory calculation for VV (c and d) as compared to VH (g and h) for points inside and outside the flood zone in Mbire, Zimbabwe ……...……...110

Figure 4.6: The training and validation loss showing MAE and MSE for VV (a and b) versus VH polarisation (c and d) inside (left) and outside (right) flood area in Mbire, Zimbabwe …….……...112

Figure 4.7: Evaluation of the LSTM-AE algorithm for training used VV (a & b), and VH (c & d), the testing used VV (e & f) and VH (g & h) with left and right columns showing locations inside and outside the flood zone, respectively ……....113

Figure 4.8: Two co-polarisation (VV) and two cross-polarisation (VH) for flood and non-flood zone time series (a), moving average (b) and rolling difference (c), and the respective boxplot of their backscatter values in Chitsungo area in Mbire, Zimbabwe. …..…114

Figure 5.1: Mbire (a), and Muzarabani (b) study sites showing tree sample locations inside and outside the floodplain with Hunyani and Musengezi river networks……...135

Figure 5.2: Sequenced machine learning method to assess leaf chlorophyll content in R …...139

Figure 5.3: The Boruta variable importance projection (BVIP) history plot showing the respective confirmed (green) rejected (red), and tentative (yellow) while the three shadow statistical derivatives (minimum, mean and maximum) metrics are shown as blue lines ……..……141

Figure 5.4: Boxplots showing Boruta variable importance projection model with 16 confirmed important (green), 43 rejected (red), and 2 tentative (yellow) with 3 randomly shuffled shadow (blue) features from musawu or jujube (*Ziziphus mauritiana*) leaf chlorophyll content (aspect on canopy, exposure to sunlight, height on canopy), statistical chlorophyll derivative and spatial locational features (n = 62) in semi-arid Mbire and Muzarabani floodplain in Zimbabwe (Software: R-Project, Library: Boruta) 142

Figure 5.5: The Random Forest variable importance projection for the 16 BVIP selected variables using the Mean Decrease Impurity or Gini Impurity Index (a) and Mean Decrease Accuracy or Permutation Importance ……..…….....…...…143

Figure 5.6: Box-and-whisker plots of 18 samples per tree field leaf chlorophyll content (LCC) (a), derived aspect on canopy (AOC) (b) and the four summary LCC from 18 samples per tree (c) with confirmed in green, rejected in red, and tentative in yellow boxplots respectively while the shuffled shadow minimum, mean, and maximum are presented in blue ……...……...144

Figure 5.7: Comparison of leaf chlorophyll content for samples exposed against not exposed to sunlight distinguished by location based on descriptive statistical metrics …...….145

Figure 5.8: Pairwise scatterplot matrix for D2FP and D2R for samples located inside floodplain (orange) and outside floodplain (blue) using GChlMax (a), GChlMean (b) GChlMin (c) and GChlStdev (d) for Z. mauritiana tree ……..……...……..147

Figure 6.1: Location of study area in Zimbabwe (a), elevation and drainage in study area (b), with zoomed-in Mbire, Site 1 (c) and Muzarabani, Site 2 (d) study sites showing drainage, sample distribution for flooded (green) and non-flooded (black) Ziziphus mauritiana trees ……..…….....175

Figure 6.2: Framework to assess the difference influenced of FRSM on Ziziphus mauritiana tree biophysical characteristics in semi-arid floodplain, northern Zimbabwe. …….. 177

Figure 6.3: Differences of canopy diameter, tree height and DBH Z. mauritiana between flooded (inside) and not flooded (outside) floodplains in semi-arid northern Zimbabwe …179

Figure 6.4: Diagnostic Studentised residuals (a-c), QQ (d - f), outlier and leverage (g - i) plot for canopy diameter, tree height and DBH in semi-arid floodplain, Zimbabwe. ……..180

Figure 6.5: The boxviolin plots showin flooded and non-flooded Z. mauritiana tree canopy diameter, northern Zimbabwe …..……...…….182

Figure 6.6: The boxviolin plots showing flooded and non-flooded Z. mauritiana tree height, northern Zimbabwe……..……183

Figure 6.7: The boxviolin plots showing flooded and non-flooded Ziziphus mauritiana tree DBH, northern Zimbabwe. ……...184

Figure 6.8: Scatterplots showing linearity of canopy diameter (a-c), tree height (d-f) and diameter at breast height (g-i) plotted against distance to floodplain (left), distance to river (middle) and flood status (right), depicted for flooded and non-flooded Z. mauritiana trees inside (black) and outside (red) floodplains of northern Zimbabwe. ……...…… 186

LIST OF APPENDICES

Appendix **Caption** **Page**

Appendix 3.1: SNAP Sentinel-1 Data preprocessing ……206

Appendix 4.1: Hybrid Deep Learning (LSTMAE) FRSM modelling in Spyder Python ……..216

Appendix 5.1: List of acronymns used in the Boruta variable importance projection …….....233

Appendix 5.2: Script used to model *Z. mautitiana* Leaf Chlorophyll Content (LCC) in R ….....236

Appendix 6.1: Script for analysis of long-term tree biophysical characteristics in R programming language ……...248

LIST OF ACRONYMS

AE	AutoEncoder
AGB	Aboveground Biomass
BVIP	Boruta variable importance projection
CD	Canopy Diameter
DBH	Diameter at Breast Height
ECDT	Ensemble of Change Detection and Thresholding
ESP	Ensemble of Scenarios Pyramid
FEM	Flood Extent Mapping
FRSM	Flood-Recharged Soil Moisture
GEE	Google Earth Engine
GPS	Global Positioning Systems
GRDH	Ground Range Detected High Resolution
IPCC	Intergovernmental Panel on Climate Change
LCC	Leaf Chlorophyll Content
IoU	Intersect over Union
LSTM	Long Short-Term Memory
LSTMAE	Long Short-Term Memory-AutoEncoder
MAE	Mean Absolute Error
ME	Mean Error
MSE	Mean Squared Error
NDFI	Normalized Difference Flood Index
NDFVI	Normalized Difference Flood (in low) Vegetation Index
NDVI	Normalized Difference Vegetation Index
ReLU	Rectified Linear Unit
RMSE	Root Mean Squared Error
RTM	Radiative Transfer Model
SAR	Synthetic Aperture Radar
SNAP	Sentinel Application Platform
UKZN	University of KwaZulu Natal
TH	Tree Height
VH	Vertical-transmit Horizontal-receipt
VV	Vertical-transmit Vertical-receipt

CHAPTER 1: GENERAL INTRODUCTION

1.1 Background

Floods are catastrophic events that affect millions of people, cause large number of fatalities, and lead to huge conomic losses, hence considered one of the biggest disasters globally (Merz et al., 2021; Jonkman 2005). As such, these disasters trigger multi-scale disaster management initiatives at various spatial scales of operation (Alfieri et al., 2016; Raikes et al., 2019). Internationally, the United Nations has initiated the Disaster Charter to provide data access to institutions during notable disasters to expedite efficient decision-making for disaster management response actions to mitigate adverse impacts (Bessis et al., 2004; Mahmood et al., 2002). In Zimbabwe, the Civil Potection Unit is responsible for disaster management (Mutizwa, 2021). While the disaster-related effects of floods are well documented, the cross-cutting ecological influence, including the provision of ecosystem goods, functions and services are poorly understood. In semi-arid areas, the ecological function of floods is initiated by recharging soil moisture by facilitating deep infiltration (Dai et al., 2019; Jiao et al., 2019). The crucial connection of floods to soil moisture and ultimately to tree biochemical and biophysical features has important implications for carbon storage necessary for climate change regulation, hence warrants attention. In semi-arid areas, though unquantified, floods play an important function associated with proliferation of vegetation. In semi-arid areas, the resulting vegetation hotspots comprise trees which contribute 26% tree carbon stock (Tucker et al., 2023) that is important to mitigate the ongoing global warming. Trees metabolic processes involving respiration, emission, and photosynthetic sequestration of carbon dioxide (Harris et al., 2021). It is thus imperative to undertake cross-cutting research that investigates how floods influence soil moisture, which in turn influences the short-term leaf chlorophyll content biochemical and ultimately long-term canopy diameter, diameter at breast height and tree height biophysical characteristics on trees in semi-arid middle Zambezi floodplains of northern Zimbabwe. A subsistence and commercial multi-purpose keystone jujube (*Ziziphus mauritiana*) tree located both inside and outside the floodplain was used to address this important knowledge gap.

Musawu (Shona) or jJujube (*Z. mauritiana*) tree species is found in many parts of the world (Alsuwayt 2025; O'Brien et al., 2023). It is a vascular angiosperm tree with seed inside an edible fruit (Huang et al., 2017) that thrives in arid and semi-arid conditions e.g., Pakistan and India

(Usman et al., 2024), China (Li et al., 2024, Duan et al., 2023; Li et al., 2015) and Zimbabwe (Chanza and Musakwa, 2021). It is a multi-purpose fruit tree species (Arndt et al., 2001) with multiple other functions (Guo et al., 2024). In Zimbabwe, the jujube tree, locally known as Musawu, has many subsistence functions and produces fruit with commercial products including herbal, and medicinal use (Bhardwaj et al., 2025; Keihanian et al., 2025; Alghamdi et al., 2024; El-Sawaf et al., 2024) and fruit juice extraction (Bussmann et al., 2024). Additionally, the fruit is important for food security since it is eaten as a snack, or desert while fresh and dried (Yadav et at., 2025). Some communities brew beer out of fermented *Z. mauritiana* fruit (Ashaolu, 2024; Bussman et al., 2024). Recently, there has been an increasing commercialisation with rising consumption as a beverage (Ashaolu, 2024). The local people have established a new line of income generation as the urban-based beverage production come to middle Zambezi Valley to purchase and transport the fruit to produce beverages (Nemapare et al., 2024). Musawu fruits are eaten both fresh and dry (Maruza et al., 2025; Gurjar, et al., 2025; Nyanga et al., 2008) hence are sold to urban markets (Chingombe et al., 2015). The fruit is consumed for its high nutritional value and can also be fermented (Nyanga et al., 2013). Thus, the fruit is an important source of yeast, lactic acid bacteria and antioxidants (Nyanga et al., 2007). The Musawu leaves can be used as livestock feed. Furthermore, the shade under Musawu tree is a convenient for family, social, cultural, and religious gatherings.

The Musawu tree is highly adaptable to different climatic conditions (Usman et al., 2023; Singh et al., 2021). In high humidity areas, it attains evergreen status (Asif et al., 2023; Meghwal et al., 2007). However, in arid and semi-arid ecosystems, higher soil moisture is attained largely during rainy season which declines drastically during arid cool dry winters to depict deciduous character (Usman et al., 2023). In semi-arid areas, there are local differences between soil moisture inside floodplains (due to floods) as opposed to outside due to precipitation alone which is important to assess the role of floods-recharged soil moisture (FRSM) on tree short-term biochemical and and long-term biophysical characteristics (Pedzisai et al., 2022). This local soil moisture disparity influence on trees facilitates evaluation of FRSM on trees in semi-arid areas (Oyedeji, 2023).

In tropical semi-arid floodplains, floods recharge soil moisture through deep infiltration. However, the influence of flood-recharged soil moisture (FRSM) on trees in the tropical semi-arid lands remains poorly understood. Yet, trees provide multiple benefits to society (Brancalion et al., 2025). For example, trees constitute large stock of carbon in biomass compared to other forms of

vegetation (Augusto et al., 2025; Dangulla et al., 2025; Islam, 2025) which is important for climate regulation (Watson and Bai, 2025). Economicaly, trees provide various timber and non-timber products for human development (Aleru et al., 2025; Nabaloum et al., 2025; Soumya et al., 2025; Teo et al., 2025). Ecologically, trees provide feed for browsing domestic and wild animals (Wang et al., 2025), habitat for dirvese species (Hendel et al., 2025). As such, assessing the short-term influence of FRSM on leaf chlorophyll content and subsequent long-term tree size based on canopy diameter, diameter at breast height, and height was imperative. However, the complex link between FRSM and trees requires robust models (Figure 1.1).

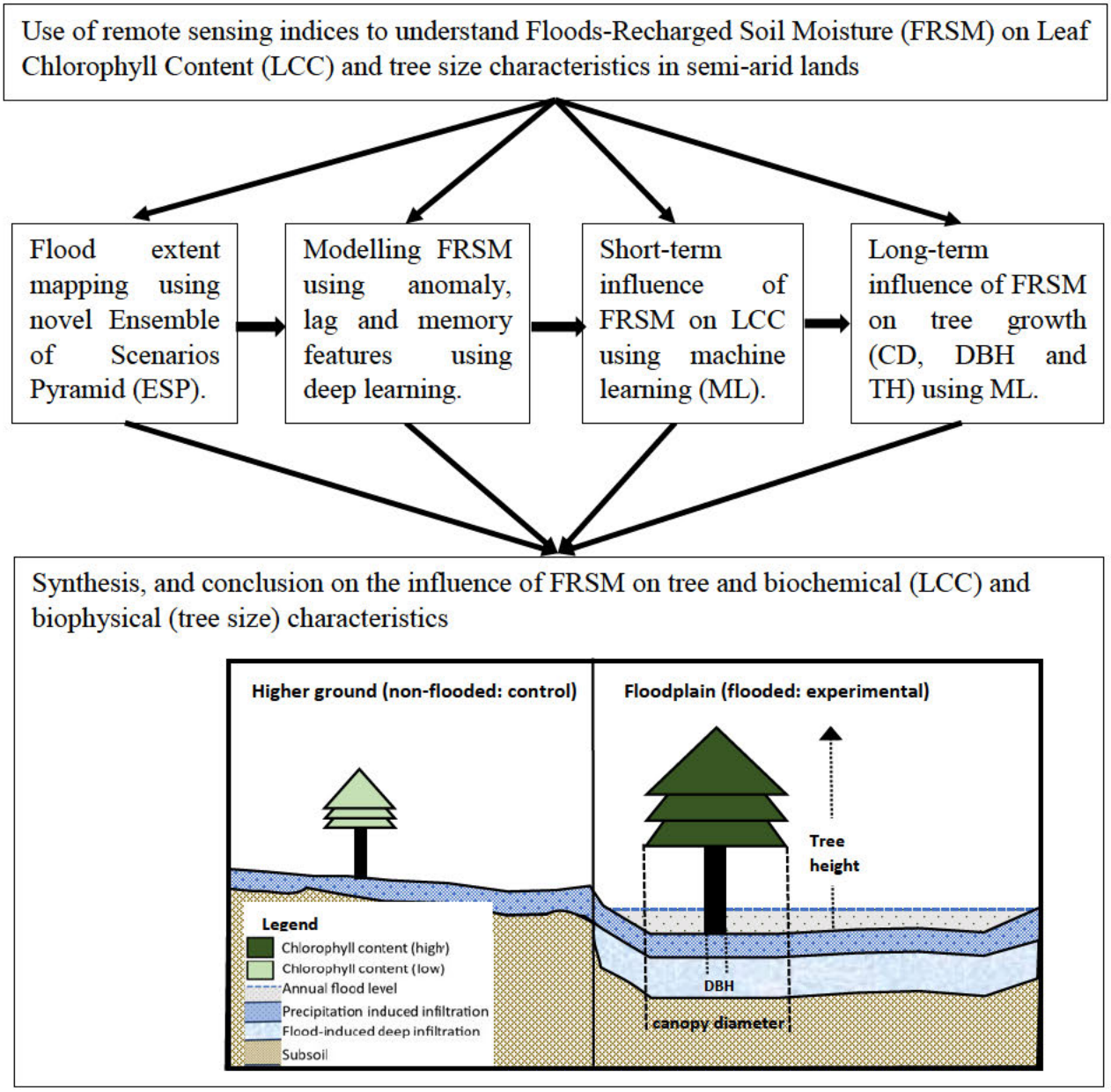


Figure 1.1: Conceptual framework of the the study

Figure 1.1 shows coherent links from literature review to; 1) flood extent mapping, 2) floods-recharged soil moisture (FRSM), 3) leaf chlorophyll content (LCC) and 4) tree size. These four in turn are linked with a synthesis and conclusions chapter that depicts a schematic presentation of the soil moisture, LCC and tree size characteristics compared between flood and non-flooded areas. Flood extent map achieved through a novel Ensemble of Scenarios Pyramid (ESP) is useful to spatially dermacate flooded zone that recharges soil moisture through deep infiltration. Floods recharge soil moisture through deep infiltration process, which was characterized by positive (wet) anomaly, distinct lag and longer memory features unlike for those in non-flooded locations. The deeper and larger in flooded zone soil moisture layer, modelled using anomaly, lag and memory within a hybrid deep learning Long Short-Term Memory AutoEncoder (LSTMAE) framework, sustains vegetation in productive 'green state' for longer a season than on non-flooded. Consequently, higher LCC modelled through machine learning than in non-flooded trees was modelled through machine learning. Therefore, he short-term biochemical LCC, is ultimately linked to the long-term biophysical tree size characteristics. Thus, on the short-term physiological and long-term growth responses of trees to flooding is essential to understand how FRSM dynamics drive tree productivity and enhance their resilience semi-arid hydrological conditions. Over several years of recurrent flooding, sustaining longer green-up season inside flooded zone thus promotes more tree growth as modelled using ML than on non-flooded trees. The flooded areas provided the experimental conditions while non-flooded ones were the control setup. Therefore, Chapters 2 through 7 were informed by this conceptual framework (Figure 1.1).

Floods are an important input for soil moisture balance modelling in semi-arid floodplains (Pedzisai et al., 2022; Kool et al., 2017). High soil moisture is important to sustain vegetation. However, there is lack of soil moisture observations in semi-arid floodplains. Alternatively, remote sensing plays a significant role in soil moisture assessment. Specifically, detecting soil moisture anomaly and monitoring its subsequent memory is important for vegetation growth and health. Thus, mapping soil moisture can adopt remote sensing indices. An example is the use of soil moisture index (SMI) propounded by Potic et al. (2017). Meanwhile, optical remote sensing is constrained by night-time and cloudy weather, hence utility of SAR data. The Synthetic Aperture Radar (SAR) data can be used to derive soil moisture since there is a distinct dielectric property (ε) between water ($\varepsilon \approx 80$) and dry soil ($2 \leq \varepsilon \leq 3$). Wet soil depicts high backscatter signal (Wagner et al., 2008). However, freely available remote sensing soil moisture products (e.g., Soil Moisture

Active Passive (SMAP) mission and Soil Moisture and Ocean Salinity (SMOS)) show huge spatial-temporal resolution trade-off (≤3 days vs 3 - 50km). Such shortcomings have prompted derivation of indices such as the ERS-2 SAR normalized radar Backscatter soil Moisture Index (NBMI) proposed to quantify soil moisture amount (Shoshany et al., 2000). These coarse spatial resolution data are inapplicable at local scale thus, testing Sentinel-1 SAR C-band data to understand FRSM in this study.

Trees can be mapped using the traditional e.g., normalized difference vegetation index (NDVI) (Rouse et al., 1974), soil adjusted vegetation index (SAVI) (Huete, 1988), and enhanced vegetation index (EVI) (Huete et al., 2002), to the more recent red edge inflection point index (REIPI) (Herrmann et al., 2010) remote sensing indices. In tree mapping, different biophysical canopy diameter, DBH and tree height characteristics relating to phenology such as green-up phase characterized by high chlorophyll content and high leaf area index (LAI) are useful proxies. These phenological phases are also influenced by leaf water content, which also preserves the health of leaves before senescence distinguishable using remote sensing indices. However, floods may occur at night or under cloudy conditions, hence passive optical remote sensing becomes less relevant leading to the preference for active SAR. Tree biophysical characteristics including canopy diameter, tree height, and DBH were thus quantified to evaluate the influence of FRSM on trees in semi-arid lands.

Apparently, floods have influence on FRSM (Pedzisai et al., 2022). Therefore, to understand the influence of FRSM on trees in semi-arid lands, firstly, flood extent mapping is important. Secondly, this flood extent map is useful to model the temporal FRSM dynamics. Thirdly, modelling of FRSM has implications on the length of the season with green leaves on trees which informs the leaf chlorophyll content. Finally, repeated longer seasons of higher soil omoisture in flooded areas with longer productivity window due to LCC, would also lead to higher tree biophysical charactersitics expressed in tree size. Thus, delimited landscape scale study separating soil moisture, LCC and tree size variables were requisite to distinguish the flooded areas form non-flooded so as to facilitate evaluation of the influence of FRSM on trees in sei-arid areas as outlined in the objectives.

1.2 Aim

The aim of study was to assess the influence of FRSM on tree biochemical and biophysical characteristics in a semi-arid landscape of northern Zimbabwe.

1.2.1 Specific Objectives

The objectives of the study are to;

1 assess the utility of remote sensing indices in understanding the nexus between FRSM with tree biochemical and biophysical features in tropical semi-arid floodplains.
2 model flood extent mapping using Ensemble of Scenarios Pyramid in semi-arid floodplains in northern Zimbabwe using Sentinel-1 SAR data.
3 assess the utility of deep learning long short-term memory autoencoder algorithm applied on Sentinel-1 SAR data to model FRSM features in semi-arid floodplains in northern Zimbabwe.
4 assess the influence of FRSM on short-term biochemical properties of *Z. mauritiana* tree in semi-arid floodplain in northern Zimbabwe.
5 evaluate the long-term influence of FRSM on *Z. mauritiana* tree biophysical characteristsics in semi-arid lands in northern Zimbabwe.

1.3 Description of the study area

The study area is Middle Zambezi Valley, located 29°E - 31°E and 15°S - 17°S in the tropical semi-arid area of northern Zimbabwe. The area is predominantly rural, partly covering Mbire District on the west and Muzarabani District on the east. It is a predominantly sparsely populated rural area with agrarian economy that grows crops and rears livestock. The farming activities are predominantly rain-fed as well as post-flood riparian farming activities (Chingombe et al., 2015). The local consumption and sale of Masawu or jujube fruits to urban centres (Maruza et al., 2017; Nynga et al., 2008) and related products is predominant in the area. The soils are largely depositional silts (Chimweta et al., 2018), and rich in nutrients, prompting the post-flood farming that occurs during the moderately warm winter season (Mavhura, 2019). Infrastructure comprises settlements, primary and secondary schools, small-scale shops, clinics and hospitals, roads and bridges.

The area is located in semi-arid tropical region with average annual summer rainfall (< 1 000 mm), high summer and moderate winter temperatures (Chingombe et al., 2015). The rain season which stretches between November through March, is characterized by torrential downpours commonly

due to tropical cyclonic activity pronounced in the first three months of the year. The hydro-meteorologic extremes make the area uniquely susceptible to extremes, namely floods and drought (Mavhura, 2019). It has average elevation around 400 m above sea level, gently sloping and highly drained by three large rivers that ultimately discharge into the Zambezi River. The Hunyani or Manyame River in the western Mbire District and Musengezi River in the eastern Muzarabani District are associated with floods in the study area. The Hunyani River has Monozi, and Ambi tributaries, while the Musengezi River has Hoya, and Nzoumvunda tributaries, which contribute to flooding in the study area. The Hunyani River originates from high elevation in Goromonzi passing nearby Harare urban, while Musengezi also originates from distant upstream source. The terrain, drainage, and seasonal heavy rainfalls often associated with tropical cyclones originating from the Indian Ocean carried in-land by easterlies and associated with flood episodes make the area susceptible to flooding.

Floods occur in two forms, due to heavy precipitation including that resulting from tropical cyclone and dam-induced backflow (Muhonda et al., 2014). Firstly, there heavy precipitation-induced annual floods occur between December and March (Nharo et al., 2019; Nyamwanza, 2018). These floods often persist for duration of weeks depending on the cause of heavy rainfall (Muhonda et al., 2014). The intertropical convergence zone is known for initiating heavy precipitation that often leads to flooding on the Zambezi catchment (Clarence-Smith, 2022). A common trigger of these annual floods are the annual tropical cyclones that originate from the Indian Ocean on the eastern direction (Dube and Manyani, 2022; Mudavanhu et al., 2020), The second type of floods are the winter season backflow floods (Muhonda et al., 2014). These are often triggered by higher water levels downstream due to the damming at Cahora Bassa that holds the water which floods the main tributaries namely Musengezi and Hoya (Gumindoga et al., 2016). These backflow floods often affect areas in proximity to the dam during winter months of May to July (Gumindoga et al., 2016).

The area experiences summer season precipitation-recharged soil moisture, while floodplains exclusively have additional FRSM, firstly during summer and secondly, as winter backflow floods, promoting deep infiltration (Colloff and Baldwin, 2010; Mohammadi et al., 2017; Chingombe et al., 2015). As such, flood occurrence in the floodplains in the area are associated with a second agricultural season benefiting from deep infiltration recharge. This soil moisture is necessary to support vegetation, particularly tree productivity, albeit not yet quantified. There are tree species that exist both inside as well as outside the floodplains. To assess the influence of floods recharged

soil moisture, a candidate tree species becomes relevant within the same locality, distinguished only by floodplain extent.

Vegetation in the study area includes grasslands, isolated trees and deciduous forests inside and outside the semi-arid floodplains. The deciduous woodland comprises several species but the predominant is the mopane-terminalia (*Colophospermum mopane* and *Terminalia stulhmani*) and mopane-combretum (*Combretum apiculatum)* in the plains and riverine landscapes (Chanza and Musakwa, 2021). Other tree species include Fig (*Ficus ingens*), and Baobab (*Adonsia digitata*). Trees constitute an important resource as a source of food security through wild fruits consumption, income generation through sales to drink and beverage producers, and leaves providing fodder for both wild and domestic animals as well as cheap health options providing herbs and medicinal products. Large trees have spacious shade areas which are often convenient venues for social and religious functions. Trees also support other life forms as habitat for other organisms, pollinating insects including honey-producing insects, and store carbon. The location of the study sites; Mbire and Muzarabani, and drainage are shown on Figure 1.2.

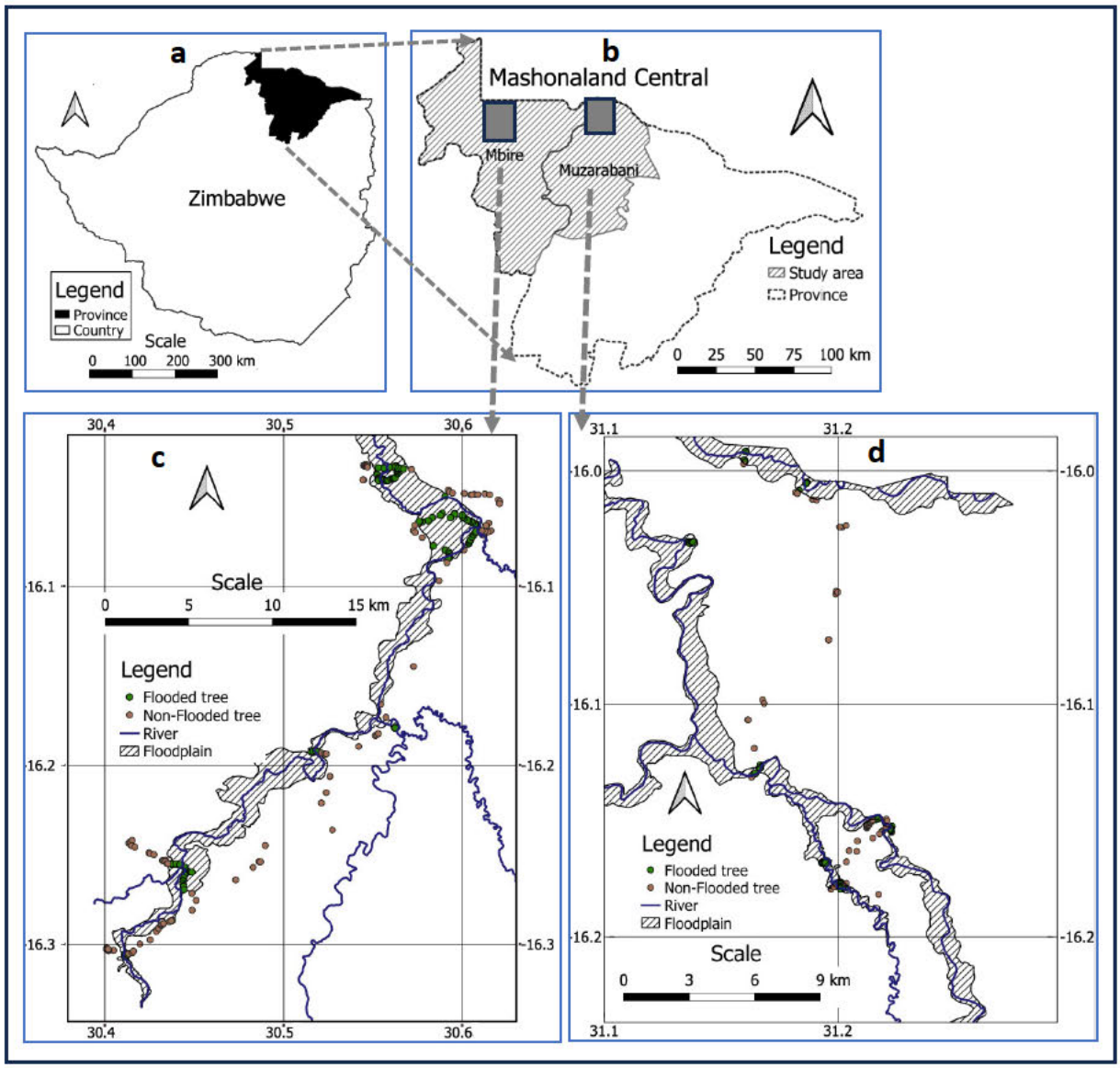


Figure 1.2: Map locating Mashonaland Central Province in Zimbabwe (a), zoomed-in with Mbire and Muzarabani (b) constituting Site 1 (c) and Site 2 (d) respectively showing both previously flooded and non-flooded tree samples, floodplain and drainage in semi-arid area.

1.4 Organization of the thesis

This thesis was organised in seven chapters as follows.

Chapter 1 is the general introduction of the study. It briefly presents a background to the study, highlighting the soil moisture shortages in semi-arid lands and the role of floods in recharging soil moisture. Further to highlighting the multiple importance of trees, jujube (*Z. mauritiana*) used as a proxy to assess the influence of FRSM on trees was introduced. Jujube adapts to aridity

conditions thus its wide existence in diverse ecological environs including arid and semi-arid areas. The chapter highlights the aim and specific objectives of the study, describes the two study sites and outlines the thesis chapters.

Chapter 2 reviewed related literature on the use of remote sensing indices to detect, map, model, and monitor characteristics such as flood extent, soil moisture and *Z. mauritiana* tree species, particularly focusing on trees. In the chapter, the missing links between floods, FRSM and trees in the short-term biochemical and long-term biophysical characteristics were reviewed. The gaps in literature were noted, defining future research directions with emphasis on the importance of applying the robust machine and deep learning approaches using modern Earth Observation data.

Chapter 3 innovated a flood extent mapping method. The mapping of food extent was an important determinant of spatial extent of soil moisture recharge through deep infiltration. A novel method in flood extent mapping based on time and weather independent SAR Sentinel-1 data termed Ensemble of Scenarios Pyramid was introduced and its utility demonstrated with a specific use-case. The Ensemble of Scenarios Pyramid is based on normalized difference flood index (NDFI), which applies a change detection and thresholding principle.

Chapter 4 was informed by the flood extent map to sample soil moisture inside versus outside the floodplain. The chapter modelled soil moisture features namely anomaly, lag and memory using a hybrid deep learning framework. The Long Short-Term Memory versatile in time series data analyses, imbedded as layers into Autoencoder framework facilitated dualized positive (wet) and negative (dry) anomaly detection while defining lag and temporally monitoring its memory.

Chapter 5 presents the short-term seasonal post-flood biochemical influence of FRSM on leaf chlorophyll content comparing floodplain with non-floodplain trees using a connected machine learning pipeline. The spatial data and descriptive statistical derivatives were screened using Boruta variable importance projection and the Random Forest algorithms. The Gini and Permutation indices were subsequently used to rank the selected important predictors.

Chapter 6 assessed the long-term FRSM influence on tree biophysical characteristics based on canopy diameter, tree height and DBH. It compared these characteristics between flooded and non-flooded trees. Since the area is small scale, the study assumed that edaphic, rainfall and temperature

variables were invariant. Therefore, distance to river, distance to floodplain and flood status were used to explain the observed difference between trees within and outside floodplains.

Chapter 7 provides a synthesis of the thesis, that shows novelties in flood extent mapping using the ESP, FRSM modelling using deep learning LSTMAE, leaf chlorophyll content assessment and tree size characteristics. The chapter proffers conclusions drawn and recommendations emanating from the assessment of FRSM on trees in semi-arid lands.

1.5 References

Aleru, K.K., Agbugba, I.K. and Taremwa, N.K., 2025. The bioeconomic profitability of non-timber forest products in the national building: A Review. *Rwanda Journal of Agricultural Sciences*, 4(1), pp.169 - 202.

Alfieri, L., Cohen, S., Galantowicz, J., Schumann, G.J., Trigg, M.A., Zsoter, E., Prudhomme, C., Kruczkiewicz, A., de Perez, E.C., Flamig, Z. and Rudari, R., 2018. A global network for operational flood risk reduction. *Environmental Science and Policy*, 84, pp.149 - 158.

Alghamdi, S.S., Alghashem, S.A., Ali, R., Alsubait, A., Suliman, R.S., Mohammed, A.E., Alehaideb, Z., Alshafi, R.A., Alturki, A.Y. and Rahman, I., 2024. Exploring the potential of Ziziphus nummularia and luteolin-7-O-glucoside as tubulin inhibitors in cancer therapy and survival. *Scientific Reports* 14(1), p.7202.

Alsuwayt, B., 2025. Evaluation of antioxidant, anti-inflammatory, and analgesic potentials of combined polyphenol-rich fractions from Ziziphus mauritiana and Ziziphus spina-christi leaves through modulation of inflammatory and oxidative stress markers in Sprague Dawley rats model. *Inflammopharmacology*, pp.1 - 15.

Arndt, S.K., Clifford, S.C. and Popp, M., 2001. Ziziphus - A multipurpose fruit tree for arid regions. In *Sustainable land use in deserts* (pp.388 - 399). Springer Berlin Heidelberg.

Ashaolu, T.J., 2024. Fermented fruits and legumes in Africa: production factors and health benefits. *International Journal of Food Science and Technology* 59(10), pp.6866 - 6874.

Asif, M., Naseer, J., Turabi, T.H., Khan, R.I. and Maryam, A., 2023. Drought induced molecular and physiological adaptations in Ziziphus mauritiana and Salvadora oleoides under climate change scenario. *Global Forests*, *2*(01), pp.1-17.

Augusto, L., Borelle, R., Boča, A., Bon, L., Orazio, C., Arias-González, A., Bakker, M.R., Gartzia-Bengoetxea, N., Auge, H., Bernier, F. and Cantero, A., 2025. Widespread slow growth of acquisitive tree species. *Nature* 640(8058), pp.395 - 401.

Bessis, J.L., Bequignon, J. and Mahmood, A., 2004. The international charter on "space and major disasters" initiative. *Acta Astronautica* 54(3), pp.183 - 190.

Bhardwaj, S., Chauhan, R., Chaudhary, M., Kaushik, D. and Singh, R., 2025. Phytochemical screening, antioxidant and antidiabetic potential of Ziziphus nummularia. *Comparative Clinical Pathology*, pp.1 - 14.

Brancalion, P.H., Hua, F., Joyce, F.H., Antonelli, A. and Holl, K.D., 2025. Moving biodiversity from an afterthought to a key outcome of forest restoration. *Nature Reviews Biodiversity*, pp.1 - 14.

Bussmann, R.W., Paniagua-Zambrana, N.Y., Kikvidze, Z., Batsatsashvili, K., Khutsishvili, M., Maisaia, I., Sikharulidze, S., Tchelidze, D., Mehdiyeva, N.P., Mursal, N. and Fayvush, G., 2024. Ziziphus jujuba Mill. Rhamnaceae. In *Ethnobotany of the Caucasus* (pp. 1-16). Springer, Cham.

Chanza, N. and Musakwa, W., 2021. "Trees are our relatives": Local perceptions on forestry resources and implications for climate change mitigation. *Sustainability* 13(11), p.5885. 10.3390/su13115885.

Clarence-Smith, W.G., 2022. Rainfall and Floods in the Upper Zambezi Basin, 1680s to 1910s. In *Droughts, Floods, and Global Climatic Anomalies in the Indian Ocean World* (pp. 127-163). Cham: Springer International Publishing.

Colloff, M.J. and Baldwin, D.S., 2010. Resilience of floodplain ecosystems in a semi-arid environment. *The Rangeland Journal* 32(3), pp.305 - 314.

Dai, L., Guo, X., Zhang, F., Du, Y., Ke, X., Li, Y., Cao, G., Li, Q., Lin, L., Shu, K., and Peng, C., 2019. Seasonal dynamics and controls of deep soil water infiltration in the seasonally-frozen region of the Qinghai-Tibet plateau. *Journal of Hydrology* 571, pp.740 - 748.

Dangulla, M., Abd Manaf, L. and Aliero, M.M., 2025. The contribution of small and medium diameter trees in West African Savanna to regional biomass and carbon pools. *Academia Environmental Sciences and Sustainability,* 2(1), doi.org/10.20935/AcadEnvSci7473

Duan, Y., Liu, S., Zhu, Y., Wang, Y., Yan, F., Liu, Z., Shi, X., Liu, P. and Liu, M., 2023. The influences of soil and meteorological factors on the growth and fruit quality of Chinese Jujube (Ziziphus jujuba Mill.). *Plants* 12(24), p.4107.

El-Sawaf, A.K., El-Moslamy, S.H., Kamoun, E.A. and Hossain, K., 2024. Green synthesis of trimetallic CuO/Ag/ZnO nanocomposite using Ziziphus spina-christi plant extract:

characterization, statistically experimental designs, and antimicrobial assessment. *Scientific Reports* 14(1), p.19718.

Gumindoga, W., Makurira, H., Phiri, M. and Nhapi, I., 2016. Estimating runoff from ungauged catchments for reservoir water balance in the Lower Middle Zambezi Basin. *Water SA* 42(4), pp.641 - 649.

Guo, M., Bi, G., Wang, H., Ren, H., Chen, J., Lian, Q., Wang, X., Fang, W., Zhang, J., Dong, Z. and Pang, Y., 2024. Genomes of autotetraploid wild and cultivated Ziziphus mauritiana reveal polyploid evolution and crop domestication. *Plant Physiology* 196(4), pp.2701 - 2720.

Gurjar, P.S., Berwal, M.K., Kumar, R., Sarolia, D.K. and Choudhary, M.K., 2025. Post-harvest Melatonin Treatment Maintains Freshness and Delayed Peel Browning by Sustaining Higher Antioxidants in Indian Jujube (Ziziphus mauritiana Lamk) Fruits. *Applied Fruit Science* 67(1), p.21.

Harris, N.L., Gibbs, D.A., Baccini, A., Birdsey, R.A., De Bruin, S., Farina, M., Fatoyinbo, L., Hansen, M.C., Herold, M., Houghton, R.A. and Potapov, P.V., 2021. Global maps of twenty-first century forest carbon fluxes. *Nature Climate Change* 11 (3), pp.234 - 240.

Hendel, A.L., Douma, J.C., Klingenfuß, S., Pereira, J.M., Ruppert, L., Spînu, A.P., Frey, J., Denter, M., Liu, X., Storch, I. and Klein, A.M., 2025. Disentangling direct and indirect effects of forest structure on biodiversity: Bottom-up and top-down effects between forestry, bats and their insect prey. *Journal of Applied Ecology* 62(1), pp.93-105.

Herrmann, I., Pimstein, A., Karnieli, A., Cohen, Y., Alchanatis, V. and Bonfil, D., 2010, March. Assessment of leaf area index by the red-edge inflection point derived from VENμS bands. In *Proceedings of the ESA hyperspectral workshop, Frascati, Italy* 683, pp. 1-7).

Huang, J., Chen, R. and Li, X., 2017. Comparative analysis of the complete chloroplast genome of four known Ziziphus species. *Genes*, *8*(12), p.340.

Huete, A. R., 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* 25, pp.295 - 309. doi:10.1016/0034-4257(88)90106-X.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., and Ferreira, L. G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of the Environment* 83, pp.195–213. doi:10.1016/S0034-4257(02)00096-2.

Intergovernmental Panel on Climate Change (IPCC)., 2019. 2019: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems in Shukla,

P.R. Skea, J. Calvo Buendia, E. Masson-Delmotte, V. Pörtner, H.-O. Roberts, D. C. Zhai, P. Slade, R. Connors, S. van Diemen, R. Ferrat, M. Haughey, E. Luz, S. Neogi, S. Pathak, M. Petzold, J. Portugal Pereira, J. Vyas, P. Huntley, E. Kissick, K. Belkacemi, M. Malley, J. (eds). https://www.ipcc.ch/site/assets/uploads/2019/11/SRCCL-Full-Report-Compiled-191128.pdf.

Islam, F.S., 2025. The Convergence of AI and Nature: Advancing Carbon Dioxide Capture, Removal, and Storage Technologies through Integrated Ecosystem-Based Strategies. *International Journal of Applied and Natural Sciences* 3(1), pp.90 - 130.

Jiao, W., Tian, C., Chang, Q., Novick, K. A., and Wang, L., 2019. A new multi-sensor integrated index for drought monitoring. *Agricultural and Forest Meteorology* 268, pp.74 - 85.

Jonkman, S.N., 2005. Global perspectives on loss of human life caused by floods. *Natural hazards* 34(2), pp.151 - 175.

Keihanian, F., Maleknejad, S., Saeidinia, A., Soltanipour, S. and Pirooz, A., 2025. Comparison of Ziziphus jujube Mill. Syrup versus polyethylene glycol in children with functional constipation: a randomized clinical trial. *Scientific Reports* 15(1), p.1674.

Kool, M., van Steenbergen, F., Mehari Haile, A., Abbas, Y.M., and Hagos, E., 2017. The promise of flood-based farming systems in arid and semi-arid areas. In *Rainwater-Smart Agriculture in Arid and Semi-Arid Areas: Fostering the Use of Rainwater for Food Security, Poverty Alleviation, Landscape Restoration and Climate Resilience* (pp. 77 - 94). Cham: Springer International Publishing.

Li, X., Li, Y., Zhang, Z. and Li, X., 2015. Influences of environmental factors on leaf morphology of Chinese jujubes. *PloS one* 10(5), p.e0127825.

Li, Y., Zhou, X., Zhao, K., Liu, J., Chen, G., Zhang, Y., Ma, J., Sun, N. and Li, X., 2024. Cultivation and morphology of jujube (Ziziphus Jujuba Mill.) in the Qi River basin of Northern China during the Neolithic period. *Scientific Reports* 14(1), p.2305.

Mahmood, A., Bessis, J.L., Bequignon, J., Lauritson, L. and Venkatachary, K.V., 2002, June. An overview of the international charter space and major disasters'. In *IEEE International Geoscience and Remote Sensing Symposium* 2, pp.771-773.

Maruza, M.I., Gandiwa, E., Muboko, N., Sango, I., Tarakini, T. and Mukomberanwa, N.T., 2024. An Analysis of Land Use and Land Cover Changes, and Implications for Conservation in Mukumbura (Ward 2), Mt Darwin, Zimbabwe, 2002-2022. *Open Journal of Ecology* 14(9), pp.706 - 730.

Maruza, I.M., Musemwa, L., Mapurazi, S., Matsika, P., Munyati, V.T. and Ndhleve, S., 2017. Future prospects of Ziziphus mauritiana in alleviating household food insecurity and illnesses in arid and semi-arid areas: A review. *World Development Perspectives* 5, pp.1 - 6.

Meghwal, P.R., Khan, M.A. and Tewari, J.C., 2007. Growing Ber (Ziziphus mauritiana Lam.) for Sustainable Income and Employment in Arid and Semi-Arid Regions. *ICAR—Central Arid Zone Research Institute (CAZRI): Jodhpur, India*.

Merz, B., Blöschl, G., Vorogushyn, S., Dottori, F., Aerts, J.C., Bates, P., Bertola, M., Kemter, M., Kreibich, H., Lall, U. and Macdonald, E., 2021. Causes, impacts and patterns of disastrous river floods. *Nature Reviews Earth and Environment* 2(9), pp.592 - 609.

Mohammadi, A., Costelloe, J.F., and Ryu, D., 2017. Application of time series of remotely sensed normalized difference water, vegetation and moisture indices in characterizing flood dynamics of large-scale arid zone floodplains. *Remote Sensing of the Environment 190*, pp.70 - 82.

Mudavanhu, C., Manyangadze, T., Mavhura, E., Pedzisai, E., and Manatsa, D., 2020. Rural households' vulnerability and risk of flooding in Mbire District, Zimbabwe. *Natural Hazards* 103(3), pp.3591 - 3608.

Mutizwa, B., 2021. An investigation into organisation capacity for national disaster management in Zimbabwe: The case of the Department of Civil Protection. *International Journal of Humanities, Management and Social Science (IJ-HuMaSS)* 4(1), pp.23 - 34.

Nabaloum, A., Sabo, P., Goetze, D., Sehoubo, Y.J., Ouédraogo, A., and Thiombiano, A., 2025. Availability and Trade Values of Some Non-timber Forest Products in Local Markets: Implications for the Conservation of Savanna Ecosystems in Burkina Faso. *Economic Botany* pp.1 - 21.

Nemapare, P., Mugadza, D.T., Gadaga, T.H. and Jombo, T.Z., 2024. Indigenous knowledge and marketing of edible wild fruits in Zimbabwe: A case study of Shurugwi, Gokwe south, Chirumhanzu, and Chivi districts. *Journal of Food Technology Research* 11(3), pp.62 - 81.

Nharo, T., Makurira, H. and Gumindoga, W., 2019. Mapping floods in the middle Zambezi Basin using earth observation and hydrological modeling techniques. *Physics and Chemistry of the Earth, Parts A/B/C* 114, p.102787.

Nyanga, L.K., Gadaga, T.H., Nout, M.J., Smid, E.J., Boekhout, T. and Zwietering, M.H., 2013. Nutritive value of masau (Ziziphus mauritiana) fruits from Zambezi Valley in Zimbabwe. *Food Chemistry* 138(1), pp.168 - 172.

Nyanga, L.K., Nout, M.J., Gadaga, T.H., Boekhout, T. and Zwietering, M.H., 2008. Traditional processing of masau fruits (Ziziphus mauritiana) in Zimbabwe. *Ecology of Food and Nutrition* 47(1), pp.95 - 107.

Nyanga, L.K., Nout, M.J., Gadaga, T.H., Theelen, B., Boekhout, T. and Zwietering, M.H., 2007. Yeasts and lactic acid bacteria microbiota from masau (Ziziphus mauritiana) fruits and their fermented fruit pulp in Zimbabwe. *International Journal of Food Microbiology* 120(1-2), pp.159 - 166.

O'Brien, C.J., Campbell, S., Young, A., Vogler, W. and Galea, V.J., 2023. Chinee apple (Ziziphus mauritiana): A comprehensive review of its weediness, ecological impacts and management approaches. *Plants* 12(18), p.3213.

Oyedeji, S., 2023. Plant adaptations in dry tropical biomes: an ecophysiological perspective. In *Ecophysiology of Tropical Plants* pp. 3 - 14.

Paudel, M.R., Poudeyal, M.R. and Devkota, H.P., 2023. Ziziphus spp.(Ziziphus jujuba mill., Ziziphus mauritiana lam.). In *Himalayan Fruits and Berries* pp. 491 - 497. Academic Press.

Pedzisai, E., Mutanga, O., Odindi, J. and Mushore, T.D., 2022. The use of remote sensing indices to understand flood-recharged soil moisture impacts on trees in semi-arid floodplains: A review. *Ecohydrology* 15(7), p.e2460.

Potic, I., Bugarski, M., and Matic-Varinica, J., 2017. Soil moisture determination using remote sensing data for the property protection and increase of agricultural production. Presentation 2017 World Bank Conference on Land and Poverty. The World Bank - Washington DC, March 20 - 24, DOI:10.13140/RG.2.2.30426.59845.

Raikes, J., Smith, T.F., Jacobson, C. and Baldwin, C., 2019. Pre-disaster planning and preparedness for floods and droughts: A systematic review. *International Journal of Disaster Risk Reduction* 38, p.101207.

Roberts, J., Dhileepan, K. and Florentine, S., 2024. A review of the biology, distribution, and management challenges posed by the invasive weed Ziziphus mauritiana L., with special reference to its invasion in Australia. *Weed Research* 64(1), pp.8 - 18.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., and Harlan, J. C., 1974. Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation. NASA/GSFC, Greenbelt, MD, USA, pp.1 - 137.

Shoshany, M., Svoray. T., Curran, P J., Perevolotsky, A., 2000. The relationship between ERS-2 SAR backscatter and soil moisture: Generalization from a humid to semi-arid transect. *International Journal of Remote Sensing* 2(11), pp.2337 - 2343.

Singh, A., Singh, R.K., Kumar, A., Kumar, A., Kumar, R., Kumar, N., Sheoran, P., Yadav, R.K. and Sharma, D.K., 2021. Adaptation to social-ecological stressors: a case study with Indian jujube (Ziziphus mauritiana Lam.) growers of north-western India. *Environment, Development and Sustainability* 23, pp.3265 - 3288.

Soumya, K.V., Shackleton, C.M. and Setty, S.R., 2025. Consumer Preferences and Markets for a Cultural Non-Timber Forest Product (Boswellia serrata Roxb.) Around Hindu Temples in Southwestern India. *Forests* 16(6), p.911.

Teo, H.C., Lamba, A., Ng, S.J., Nguyen, A.T., Dwiputra, A., Lim, A.J.Y., Nguyen, M.N., Torngern, P., Zeng, Y., Dewi, S. and Koh, L.P., 2025. Reduction of deforestation by agroforestry in high carbon stock forests of Southeast Asia. *Nature Sustainability* pp.1 - 5.

Tucker, C., Brandt, M., Hiernaux, P., Kariryaa, A., Rasmussen, K., Small, J., Igel, C., Reiner, F., Melocik, K., Meyer, J. and Sinno, S., 2023. Sub-continental-scale carbon stocks of individual trees in African drylands. *Nature* 615(7950), pp.80 - 86.

Usman, M., Mirza, S.A. and Fatima, B., 2023. Climate resilience in Indian jujube (Ziziphus mauritiana Lam.) and other jujube species. In *Cultivation for Climate Change Resilience* 1 (pp. 241 - 271). CRC Press.

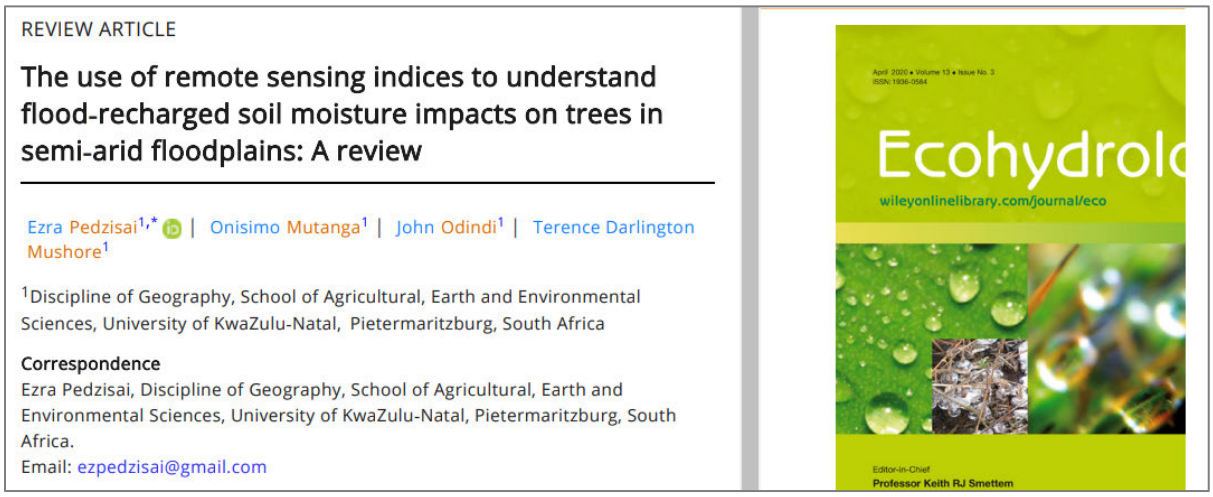
Wagner, W., Pathe, C., Doubkova, M., Sabel, D., Bartsch, A., Hasenauer, S., Bloschl, G., Scipal, K., Martinez-Fenandez, J., and Low, A., 2008. Temporal stability of soil moisture and radar backscatter observed by the Advanced Synthetic Aperture Radar (ASAR). *Sensors (Basel)* 8(2), pp.1174 - 1197.

Wang, L., Wei, F., Tagesson, T., Fang, Z. and Svenning, J.C., 2025. Transforming forest management through rewilding: Enhancing biodiversity, resilience, and biosphere sustainability under global change. *One Earth* 8(3).

Watson, A.S. and Bai, R.S., 2025. Studies on ecosystem services and air-pollution mitigation in tropical urban vegetation using i-Tree Eco Model. *Environment, Development and Sustainability*, pp.1 - 18.

Yadav, A., Kumar, M., Bangar, S.P. and Siroha, A.K., 2025. Response of pre-harvest application of chemicals on physico-chemical properties of ber (Ziziphus mauritiana Lamk.) cv. Gola during storage. *Journal of Food Measurement and Characterization* 19(5), pp.3600-3612.

CHAPTER 2: THE USE OF REMOTE SENSING INDICES TO UNDERSTAND FLOOD-RECHARGED SOIL MOISTURE INFLUENCE ON TREES IN SEMI-ARID LANDS: A REVIEW



REVIEW ARTICLE

The use of remote sensing indices to understand flood-recharged soil moisture impacts on trees in semi-arid floodplains: A review

Ezra Pedzisai[1,*] | Onisimo Mutanga[1] | John Odindi[1] | Terence Darlington Mushore[1]

[1]Discipline of Geography, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa

Correspondence
Ezra Pedzisai, Discipline of Geography, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa.
Email: ezpedzisai@gmail.com

The chapter is based on: **Pedzisai, E.,** Mutanga, O., Odindi, J. and Mushore, T.D., 2022. The use of remote sensing indices to understand flood-recharged soil moisture impacts on trees in semi-arid floodplains: A review. *Ecohydrology* 15(7), p.e2460. https://doi.org/10.1002/eco.2460.

Abstract

Remote sensing indices are essential for detection, mapping and monitoring of floods, soil moisture and tree characteristics. An understanding of the nexus between flood-recharged soil moisture (FRSM) and endangered biodiversity tree hotspots in semi-arid floodplains is critical. Specifically, the influence of FRSM on biophysical characteristics of keystone tree species with social, economic, ecological and environmental functions is crucial. However, the use of remote sensing indices to understand the link between episodic floods, FRSM and trees, as well as their respective spatiotemporal dynamics within the semi-arid floodplains, remains largely unexplored. Hence, this study reviewed literature on the adoption of remote sensing indices in understanding short-term phenological and long-term tree structural changes arising from flood-related infiltration. The reviewed literature shows the predominance of the normalized difference water index, the land surface water index and normalized difference vegetation index for flood, FRSM and tree mapping. The review also shows the emergence of SAR-based normalized difference flood index for understanding the impacts of FRSM on riparian trees. Although the review identified an increase in data availability, there are high costs associated with active commercial remotely sensed data of superior quality and computing resources. Furthermore, available free data are incapable of penetrating canopies and have coarse spatial resolution, constraining the use of remote sensing indices to understand the impacts of FRSM on trees. In view of increasing free SAR data and cloud-based computing resources, application of emerging deep learning algorithms offers prospects for the use of remote sensing indices to understand the impacts of FRSM on riparian trees.

Keywords

Flood-recharged soil moisture, floodplain, machine learning, remote sensing indices, semi-arid, synthetic aperture radar, tree biophysical characteristicses.

2.1. Introduction

Floods are known for their adverse socio-economic and environmental impacts (Zhang, et al., 2021). However, their occurrence also significantly influences floodplain ecosystem integrity and function (Wallace et al., 2021). Generally, episodic floods recharge shallow alluvial aquifers and enrich deep soil water content (Dahan et al., 2007) while pre-flood soil moisture determines

flooding in a bidirectional feedback link (Nied et al., 2013). Therefore, floods have a beneficial soil moisture recharge ecological function.

Floods play a critical role on floodplain soil moisture and riparian vegetation. Commonly, floodplains are ecologically rich, especially in areas where flooding is the predominant episodic abiotic perturbation that supplies flood-recharged soil moisture (FRSM) through deep infiltration (Colloff and Baldwin, 2010; Mohammadi et al., 2017). Generally, floods influence species diversity and distribution, while riverine vegetation is an important nutrient sink promoting flood retention and recharges groundwater and soil moisture (Fernando et al., 2021). Hence, Sera et al. (2001) note that episodic floods alter post-flood plant communities, while Hukusima and Yoshikawa (1997) conclude that differential flood deposition influences vegetation species distribution with unique biophysical characteristics related to age, form and species composition. As such, understanding the influence of episodic floods on riparian vegetation, in comparison to non-flooded areas is critical to a range of landscape management decisions.

In arid and semi-arid landscapes, trees are keystone vegetation with significant socio-economic, ecological and environmental benefits (Wallace et al., 2021). For instance, Mausse et al. (2021) noted that wood fuel and several non-timber products that include fruit, fibre and tree shade, in addition to being important sites of assembly, provide human and livestock medicinal value. Furthermore, Dlamini and Geldenhuys (2009) noted that trees are an important source of timber. Floodplain trees also protect the soil from erosion (Fischer et al., 2021), assimilate carbon and help to mitigate climate change (Fernando et al., 2021). Therefore, trees in the floodplain are important since they serve multiple functions.

Traditionally, mapping the interaction between floods, FRSM and floodplain tree characteristics has relied on surveys and other field-based approaches (Ety et al., 2021). Other studies have adopted empirical, conceptual and physics-based engineering 1D/2D/3D, hydraulic/hydrodynamic models or computer-based models (Nied et al., 2013). Among the traditional approaches, the gravimetric method that uses an auger, core tubes and open-drive samplers, e.g., Pomona and Johnson, has been the most popular (Wallace et al., 2021). Other methods utilise electrical resistance, heat diffusion, absorption, tensiometric, penetration and radiative transfer models (Schmugge, 1983). Tree biophysical characteristics have mainly been determined by measuring diameter at breast height (DBH) using diameter tape (Wallace et al., 2021), as well as crown size

and height using laser hypsometer within the sampling quadrats or along a transect (Suarez et al., 2015). Generally, these approaches are tedious, costly, time consuming and limited in spatial coverage and require skilled expertise (Ety et al., 2021; Jiang et al., 2021). Whereas post-flood field surveys are also widely adopted, they are commonly constrained by inadequate access that typifies flooded zones (Liang and Liu, 2020).

To address these challenges, geo-information approaches have become valuable in understanding the relationship between floods, FRSM and tree characteristics. Commonly, these approaches are used to complement rather than to replace ground- or laboratory-based measurements (Tangdamrongsub et al., 2021). Specifically, remote sensing approaches have become valuable in understanding floods, FRSM and tree characteristics connection. Some of the pioneer studies that adopted remote sensing in flood mapping include Hallberg et al. (1973), Morrison and White (1976), Gale and Bainbridge (1990) and Rasid and Pramanik (1993). To date, the use optical remote sensing with varied spatio-temporal and spectral characteristics in understanding floods, FRSM and tree characteristics has become popular. Recently, freely available synthetic aperture radar (SAR) sensors, e.g., Sentinel-1, have facilitated the use of SAR data in floods, FRSM and tree characteristics (Zhang et al., 2021). Other geo-information-based approaches include participatory geographic information systems mapping (Feng et al., 2020) and social media data mining (Shoyama et al., 2021), while the use of unmanned aerial vehicles imagery to determine local inundation is still at infancy stage (Hashemi-Beni and Gebrehiwot, 2021).

Currently, commonly used remote sensing techniques to determine flood extents include histogram thresholding to generate binary maps (di Baldassarre et al., 2011), change detection (Huang and Jin, 2020), change detection and thresholding (Long et al., 2014), principal component analysis and statistical and hybrid methods (Akay, 2021). Determination of flood depth involves the fusion of surface extent with elevation data (di Baldassarre et al., 2011; Schumann et al., 2007), while flow velocity is a function of slope: a derivative of elevation data. Soil moisture measurement involves microwave remote sensing to discriminate wet from dry soil using very high as opposed to very low dielectric constants, respectively (Li et al., 2018). Tree parameters can be retrieved from remote sensing data. There are many applications of the use of remote sensing in riparian trees in semi-arid areas, for example, analyzing annual tree growth cycle that constitute pigments, phenology, structure and biomass (Yan et al., 2015). Congalton et al. (2002) used air photos and Landsat TM images to map tree location and structure. Dufour et al. (2013) found radar data to be

relevant for phenology extraction. Johansen and Phinn (2006) used IKONOS and Landsat ETM+ data to map structural parameters and species composition of riparian trees in Australia. Mohamed (2017) used normalized difference vegetation index (NDVI) from MODIS and Landsat to derive canopy cover in a semi-arid environment. Cho and Ramoelo (2019) used MODIS NDVI and Google Earth Image tree cover mapping using time series analysis for 18 years. Adam et al. (2017) mapped individual tree species using satellite images and machine learning in semi-arid area in South Africa. Yamaç et al. (2020) tested the use of machine learning and deep learning for soil moisture that affects trees in semi-arid area using soil samples. Burchard-Levine et al. (2022) evapotranspiration in semi-arid tree-grass ecosystems. Paz-Kagan et al. (2021) assessed tree species distribution and diversity on a woodland-to-shrubland climatic gradient in semi-arid Mediterranean area. Muthoka et al. (2021) used Sentinel-2 data and ensemble machine learning classifiers to map *Opuntia stricta* in the semi-arid Kenya. Chen et al. (2022) ensembled machine learning model over the semi-arid areas of northwest China. Although both optical and SAR approaches are useful in flood, soil moisture and tree mapping, the former applications are often constrained by clouds and night-time (Ety et al., 2021), while the latter are associated with complex scattering, especially by trees (Das and Ghosh, 2016).

Traditional image classification algorithms such as pixel-based (Landuyt et al., 2017), object-based and hybrid have been widely used in understanding floods, FRSM and tree characteristics nexus (Zhang et al., 2021). However, such algorithms are computationally costly, require manual tuning and generate inferior mapping quality (Jiang et al., 2021); hence, the emergence of modern data-driven machine learning analyses methods that include Random Forest (Breiman, 2001), Support Vector Machine (Dhara et al., 2020) and Artificial Neural Network (Dong et al., 2021). Recently, more robust approaches utilising complex architecture of the Convolutional Neural Networks to form Deep Neural Network have demonstrated superiority over existing machine learning approaches in handling remotely sensed data (Ma et al., 2019) with some applications in floods (Zhou et al., 2021). Machine learning can characterize local features such as topography, edaphic properties and spectral indices (Rad et al., 2021), radar backscatter coefficient (Li et al., 2019; Zheng et al., 2021) and temporal dynamics (Rad et al., 2021) necessary for flood, FRSM and tree mapping, while deep learning approaches can be used to analyse ‘big’ floods, FRSM and tree remotely sensed data (Fischer et al., 2021).

To date, several studies have exclusively reviewed literature on floods, FRSM, or riparian trees in general. For instance, Tockner and Stanford (2002) reviewed literature only on the use of RS to determine floodplain trends, while Bijeesh and Narasimhamurthy (2020) focused on water detection, delineation methods and algorithms. Both Huang and Jin (2020) and Liang and Liu (2020) reviewed literature on specifically the use of SAR in flood mapping, while di Baldassarre et al. (2011) concentrated on approaches and uncertainty in flood modelling (di Baldassarre et al., 2011). Other studies, e.g., Petropoulos et al. (2015) and Vereecken et al. (2014), dwelt on soil moisture measurement techniques, while Kuenzer et al. (2011), Wang et al. (2019) and Dai et al. (2019) reviewed literature on wetlands, mangrove ecosystems, mangrove forests, tree response to precipitation and snow melt and use of Convolution Neural Network in tree mapping, respectively. On the influence of FRSM on tree characteristics, Casanova (2015) reviewed literature on water necessities for floodplain gum tree species, while Wallace et al. (2021) studied the connection between tree physiology and FRSM specifically in Australian floodplains. Eisfelder et al. (2012) reviewed literature on aboveground biomass. Whereas the reviews noted above have highlighted the value of remote sensing in understanding floodplain physical and ecological characteristics, it is evident that these phenomena (i.e., flooding, FRSM, and trees) were reviewed in isolation and in a fragmented approach rather than in tandem. Hence, there is lack of knowledge on the use of remote sensing indices on flood-FRSM-trees nexus.

Typically, flooding, FRSM and tree characteristics are derived from water, soil moisture and tree's wet and dry matter within the three distinct zones of the electromagnetic spectrum namely, passive optical (350–2,500 nm), thermal (3,500–14,000 nm) and microwave (5–1,000 mm) (Wijewardana et al., 2019). Optical remote sensing relies on solar domain to detect flooding, FRSM and tree characteristics within the visible (blue, green and red), near infra-red (NIR) and short-wave infrared spectra (Yue et al., 2019), while thermal remote sensing measures flooding, FRSM and tree characteristics based on land surface temperatures or thermal inertia (Sadeghi et al., 2015). The microwave remote sensing of flooding, FRSM and tree characteristics relies on backscatter intensity of energy pulse (Liu et al., 2021). Typically, such delineation is based on a combination of at least two bands of either active (Wessels, 2014) or passive sensor data (Dorigo et al., 2007). Hence, combining spectral bands that belong to different portions of the electromagnetic spectrum (i.e., indices) has become valuable in understanding landscape characteristics (Pasqualotto et al.,

2018), making a review of use of remote sensing indices to assess flood and soil related biophysical characteristics, including key tree types necessary.

There are many passive remote sensing indices that are applicable to study floods, FRSM and trees biophysical characteristics at different spatial scales ranging from leaf, canopy, stand, up to ecosystem scales. However, it is important to note that broadband multispectral remote sensing indices can inform what target materials are, e.g., the NDVI distinguishes vegetation from non-vegetation landcover, while the recently developed hyperspectral remote sensing indices can discriminate more on the physiological characteristics such as plant light use efficiency, photosynthetic and stomatal behaviour (Coops et al., 2010; Dorigo et al., 2007; Marino, 2014). The application of SAR based indices is also emerging, especially with Sentinel-1 SAR open data policy (Zhao et al., 2022).

2.2. Role of remote sensing indices for FRSM and trees in semi-arid lands

Remote sensing indices are useful to detect, map and monitor floods (Cian et al., 2018a; 2018b), soil moisture (Potic et al., 2017; Wang and Qu, 2007; Yue et al., 2019) and trees (Croft et al., 2014; Wessels, 2014). The most common passive remote sensing indices using red and near infra-red channels include simple ratio (SR) (Jordan 1969), normalized difference vegetation index (NDVI) (Rouse et al., 1974), enhanced vegetation index (EVI) (Huete et al., 2002), soil-adjusted vegetation index (SAVI) (Huete, 1988) and the normalized difference water index (NDWI) (McFeeters, 1996) or its modified versions e.g. mNDWI (Xu, 2006). Remote sensing indices are also computed using arithmetic operations and modelled using associated auxiliary data. The polarimetric transformed radar backscatter coefficients for microwave data including the quotient product (Forkuor et al., 2020), cross ratio (Vreugdenhil et al., 2018) and decomposed metrics, e.g., entropy and phase differencing (Wessels, 2014) are important remote sensing indices. The microwave remote sensing indices (Dong and Crow, 2019) are used for FRSM retrieval while EVI, NDVI and NDWI are widely used to understand post-flood tree biophysical parameters using change detection (Michener and Houhoulis, 1997) and phenological metrics (phenometrics) (Yan et al., 2015).

2.2.1 Application of remote sensing indices for FRSM retrieval

There are many applications of remote sensing indices to estimate soil moisture. Images from various remote sensing platforms largely in the microwave spectrum are useful to formulate the remote sensing indices for soil moisture estimation. For example, remote sensing data from

METEOSAT are useful to understand FRSM in semi-arid areas (Jarlan et al., 2003, in Eisfelder et al., 2012). The Gravity Recovery and Climate Experiment (GRACE), which operated from 2002 till 2017, provides water and soil moisture data (Sadeghi et al., 2020), hence useful for FRSM retrieval in semi-arid floodplains. Tian et al. (2017) used GRACE data assimilation to improve soil moisture retrieval in Australia. Remote sensing data from optical, thermal and especially microwave regions of the electromagnetic spectrum are useful to detect, map and monitor FRSM. The respective remote sensing indices applied in retrieval of FRSM are outlined below.

2.2.1.1 Passive optical and thermal remote sensing indices applications for FRSM

Soil moisture is commonly influenced spatially and temporally by precipitation, irrigation, soil physical properties, topography, land cover and evapotranspiration (Malbéteau et al., 2016). Kerr (2007) defined soil moisture as the total amount of water in unsaturated zone, which is classified into surface and root zone expressed in either gravimetric or volumetric units. Soil moisture saturation arising from flooding can be estimated using optical and thermal remote sensing data (Dorigo et al., 2007). Yue et al. (2019) proposed three normalized difference SWIR soil moisture indices (NSDSI) for bare soil transformed mathematically in Equations 2.1–2.3:

$$NSDSI_1 = (\lambda 1 - \lambda 2) / \lambda 1 \quad (2.1)$$

$$NSDSI_2 = (\lambda 1 - \lambda 2) / \lambda 2 \quad (2.2)$$

$$NSDSI_3 = (\lambda 1 - \lambda 2) / (\lambda 1 + \lambda 2) \quad (2.3)$$

Where: $NSDSI_1$, $NSDSI_2$ and $NSDSI_3$ are the 1st, 2nd and 3rd normalized difference SWIR indices respectively, while λ1 is SWIR1 (1,550 – 1,750μm) and λ2 is SWIR2 (2,100 – 2,300μm) bands. The evaluation of the three SMIs showed good performance for unsaturated (soil moisture<50%) and unsatisfactory for moderate to saturated (soil moisture > 50%) bare soil. Lobell and Asner (2002) showed that after soil moisture reaches a critical point, soil reflectance at band ($\rho_{S(\lambda)}$) increases exponentially with soil moisture as presented in Equation 2.4:

$$\rho_{S(\lambda)} = \rho_{satS(\lambda)} + (\rho_{dryS((\lambda)} - \rho_{satS(\lambda)}) \exp(-\alpha \times SM) \quad (2.4)$$

where $\rho_{satS(\lambda)}$ and $\rho_{dryS(\lambda)}$ are saturated and dry soil reflectance at band λ respectively, and α is the rate at which the soil moisture affects reflectance.

Reflective remote sensing measurement of soil parameters can be achieved with remote sensing indices using data in the blue, green, red (BGR) and NIR electromagnetic spectrum ranges. For instance, Hounkpatin et al. (2018) used RapidEye and Landsat 8 OLI data to derive remote sensing indices for soil, including brightness, coloration, and saturation for assessing soil reflectance

magnitude, soil colour, and spectral slope, respectively. The coloration index is important for FRSM since soil becomes darker when wet (Kerr, 2007). Remote sensing of soil moisture depends on its organic matter content, mineralogy, surface roughness and water content (Yue et al., 2019). Floods positively influence FRSM in low lying terrain by promoting deep infiltration (Dai et al., 2019; Jiao et al., 2019), while inland flooding is linked to increased FRSM regimes (Boni et al., 2016). Therefore, after flood recession, the retrieval of FRSM depicts a dark spectral signature on moist rich flooded zones. The estimation of soil moisture can be achieved using several remote sensing indices (Table 2.1).

Table 2.1: Passive remote sensing indices for soil moisture characterization

Soil Moisture Index	Application	Study
	Optical (350 -2500nm)	
Water Deficit Index (WDI)	Actual and potential ET rates in partially vegetated surfaces	Moran et al. (1994)
Shortwave Infrared Water Stress Index (SIWSI)	leaf water is highly absorptive in SWIR	Fensholt and Sandholt (2003)
Perpendicular Drought Index (PDI)	Estimate soil moisture on bare soil and low vegetation based on NIR-R space	Ghulam et al. (2006)
Modified Perpendicular Drought Index (MPDI)	Introduced fractional vegetation cover to PDI to estimate SM in moderate vegetation cover and bare areas	Ghulam et al. (2007)
Surface Water Capacity Index (SWCI)	Estimate SM with bands 6 and 7 relationship between NDVI and normalized MODIS	Du et al. (2007)
Shortwave Infrared SM Index (SIMI)	Estimate SM using relationship between MODIS bands 6 and 7	Yao et al. (2011)
Shortwave Infrared Drought Index (VSDI)	Estimate SM using relationship using SWIR, B, and R channels	Zhang et al. (2013)
Shortwave Infrared Water Stress Index (MSIWSI)	Estimate SM using MODIS bands 6, 7 and SWIR reflectance and fractional vegetation cover	Tin (2015)

Land surface water index (LSWI)	Estimate SM using NIR and SWIR bands	Chandrasekar et al. (2010)
Thermal (3500-14000nm)		
Vegetation Supply Water Index (VSWI)	SM in moderate to dense vegetation	Nemani et al. (1993)
Temperature Vegetation Dryness Index (TVDI)	Big differences between bare and water based on LST-NDVI feature space	Sandholt et al. (2002)
Normalized Multiband Index (NMDI)	Estimate SM using NDVI and LST	Wang and Qu (2007)
Soil Moisture Index (SMI)	Red, near infra-red and thermal bands	Potic et al. (2017)
Surface Water Content Temperature Index (SWCTI)	Combines water sensitive SWCI with modified LST to estimate surface SM status for sandy loam soils	Hong et al. (2018)
Temperature condition index (TCI)	Drought monitoring using RS indices	Jiao et al. (2019)

Both optical and thermal bands are useful for determining soil moisture as shown on Table 2.1. The visible spectrum and SWIR are useful channels in many optical remote sensing indices for FRSM. All the optical remote sensing indices for FRSM utilise either red-NIR or NDVI space except shortwave infrared soil moisture index (SIMI), which combines red, SWIR and blue bands. The thermal remote sensing indices combine red, NIR or NDVI with land surface temperature (LST) to retrieve FRSM. Surface water content index (SWCI) combines the optical SWCI with LST to determine FRSM. Both optical and thermal remote sensing indices for soil moisture retrieval make use of red-NIR or NDVI. Dorigo et al. (2007), for instance, showed that thermal remote sensing is useful for estimating soil moisture, while the Temperature Vegetation Dryness Index (TVDI) assumes that the main source of variation in surface temperature is soil moisture. However, some sensors such as Sentinel-2 MSI do not have thermal bands required to compute TVDI (Sandholt et al., 2002). Although Li et al. (2018) noted that optical/thermal retrieval of soil moisture produces high resolution products, these remote sensing indices are constrained by bad weather and time of the day (di Baldassarre et al., 2011; Rahman and Thakur, 2018; Schumann and Moller, 2015; Shen et al., 2019). In that view, active and passive microwave remote sensing are essential.

2.2.1.2 Microwave remote sensing indices applications for FRSM retrieval

The most prominent remote sensing data for soil moisture is in the microwave region which uses radar, scatterometers and radiometers to measure the water influenced dielectric permittivity (ε) of soil (Kerr, 2007). The difference between water (ε=80) and dry soil (ε<3) (Li et al., 2018; Sadeghi et al., 2015) is useful in mapping FRSM. The very coarse spatial resolution soil moisture missions include ESA's Advanced SCATterometer (ASCAT) (Kerr, 2007), NASA's Soil Moisture Active Passive (SMAP) (Entekhabi et al., 2010), ESA's Soil Moisture and Ocean Salinity (SMOS) (Kerr et al., 2012), Advanced Microwave Scanning Radiometer–EOS (AMSRE-1) (Schumann and Moller, 2015) and AMSR2 (Chen et al., 2017). The GRACE mission provides global water and soil moisture data (Sadeghi et al., 2020). A summary of soil moisture products based on radar is shown on Table 2.2.

Table 2.2: Specifications of selected global satellite remote sensing-based soil moisture retrieval missions**.**

Sensor	Launch Date	Active/ Passive (A/P)	Depth (cm)	Resolution		Band	Study
				Spatial (km)	Temporal (days)		
AMSR-E	May 2002	P	1-2	50<	1-2	X/C	Kawanishi et al. (2003)
GRACE	May 2002			3	30	K, Ko	Tian et al. (2017)
ASCAT	Oct 2006	A	1-2	25-50	1-2	C	Figa-Saldaña et al. (2002)
ASCAT	Sep 2012	A	1-2	12,5-25	1-3	C	Steele-Dunne et al. (2019)
ASCAT	Nov 2018	A	1-2	25-50	1-3	C	Steele-Dunne et al. (2019)
SMOS	Nov 2009	A	3-5	35	1-3	L	Kerr et al. (2010)
AMSR2	May 2012	P	1	15-2170	0.5	C/X	Maeda and Taniguchi (2013)
SMAP	Jan 2015	A + P	5	3-40	3	L	Entekhabi et al. (2010)
Sentinel-1A	Apr 2014	A	5	0.01	12	C	Greifeneder et al. (2019)
Sentinel-1B	Apr 2016	A	5	0.01	12	C	Greifeneder et al. (2019)

There is an evidently a trade-off between the very coarse spatial versus the very high temporal resolutions as shown on Table 2.2. Shen et al. (2019) noted that Sentinel-1A and -1B have 6-day or better temporal resolution only in Europe and Hawaii. Except for Sentinel-1 based FRSM retrieval, which is still in infancy, the available soil moisture products are applicable at catchment to global rather than at field scale (Vereecken et al., 2014). Local scale remote sensing of FRSM including field point measurements such as time domain reflectometry, capacitance and wireless soil moisture sensors (Vereecken et al., 2014) require upscaling. Additionally, multi-point hydrogeophysical methods including ground penetrating radar, electromagnetic induction and electrical resistivity tomography characterized by expensive installation (Vereecken et al., 2014) often need to be upscaled for local applications. Tian et al. (2017) used GRACE data assimilation to improve soil moisture retrieval in Australia. GRACE data are useful to monitor soil moisture. It is an important satellite that provides soil water information; hence it was also widely used in numerous studies to evaluate soil water storage changes. Therefore, remote sensing methods, although indirect, are promising at field scale FRSM retrieval but require high resolution images for fusion with coarse microwave and field data to calibrate and validate the models (Vereecken et al., 2014). While increasing the antennae size of very coarse remote sensing-derived soil moisture products improves spatial resolution, it is technically challenging and reduces sensitivity of sensing, hence the addition of other frequencies as proposed by Kerr (2007). Therefore, achieving high spatiotemporal resolution depends on either upscaling of field observations that employ advanced statistical methods (Kang et al., 2021), or downscaling of very coarse spatial resolution microwave soil moisture data (Li et al., 2018). Li et al. (2018) for instance downscaled SMOS data with Sentinel-1 C-band data to 1.25 and 2.5km resolution using wavelet transform fusion. Recently, Sun and Cui (2021) applied machine learning technique to evaluate downscaling factors for coarse microwave soil moisture. Fusing microwave, thermal and optical data is beneficial in soil moisture estimation (Anusha and Barathi, 2020). The passive soil moisture sensors challenge is the low spatial resolution while that for active sensors is the low temporal resolution, vegetation cover and surface roughness, with AMSR2 and SMOS suffering from radio frequency interference (Kerr, 2007), making the retrieval of FRSM using remote sensing indices complex.

Use of active microwave remote sensing in soil moisture retrievals utilizes the more superior L-band and P-band (Holtgrave et al., 2018) as well as the less effective X- and C-bands (Dong and

Crow, 2019). It relies on interpretation of the radar backscatter echo (sigma naught) which is influenced by dielectric permittivity, roughness and vegetation. Generally, it is more effective on bare soil. Single channel algorithm using VV polarization retrieves soil moisture better than HH and cross polarised VH (Ma et al., 2020). Choker et al. (2017) noted that Sentinel-1 SAR C-band VH backscatter coefficient is sensitive to vegetation, hence inferior in determining soil moisture in forested flooded areas. Although soil moisture estimation in vegetated areas is difficult to retrieve, the application of the water cloud model minimises error. Backscatter-based soil moisture retrieval uses either the empirical Oh (Oh et al., 1992) and Dubois (Dubois et al., 1995) models or the physical Integral Equation Model (IEM) and advanced IEM models (Choker et al., 2017).

2.2.2 Application of remote sensing indices for estimating tree parameters

In floodplains, FRSM is an important source of vegetation water content, hence, its impact on trees is detectable using remote sensing indices as vegetation absorbs more of blue and red energy fluxes for photosynthesis than green (Dorigo et al., 2007). The low red to high NIR reflectance yields the red-edge phenomenon that uniquely distinguishes vegetation from other materials (Clevers et al., 2002). Tree spectral measurements are based on physical and chemical components that influence the interaction of solar energy with the material's properties. Colour, shape, size and orientation of leaves and canopy background material properties are important factors. Cho and Ramoelo (2019) used MODIS NDVI and Google Earth Image for tree cover mapping using time series analysis for period 2001–2018. Norman et al. (2014) indicated that the NDVI is biomass proxy, as noted studying phenology at the Mexico and US border. Eisfelder et al. (2012) noted that low-resolution optical data are useful to understand phenology, using a combination of optical and radar data applying the NDVI and modelling of tree parameters indicating a need for similar studies in semi-arid areas. Apparently, mapping of AGB (total NDVI), woody, foliage, shrub and green leaf volumetric density (GVD) uses remote sensing indices, e.g., multispectral (NDVI) (Eisfelder et al., 2012) and hyperspectral (photochemical reflectance index (PRI), SIPI, NDVI)) data to determine the greenness, cover and leaf area index (LAI). Barnes et al. (2017) used hyperspectral data to deduce the photosynthetic capacity. The PRI; originally developed to estimate rapid changes in the epoxidation state of xanthophyll pigments (Sims and Gamon, 2002) which is a major component of non-photochemical quenching which was used by Garmon et al. (1992) as a derivative of hyperspectral reflectance measurements. Chou et al. (2017) calculated the PRI for plants (Equation 2.5):

$$PRI = (\lambda 1 - \lambda 2) / (\lambda 1 + \lambda 2) \quad (2.5)$$

where: $\lambda 1$ and $\lambda 2$ are the reflectance at wavelength 531 nm and 570 nm respectively.

Combining leaf, canopy and soil models enables calculation of top-of-canopy reflectance (Dorigo et al., 2007). The use of PRI has been extended to detect plant biochemistry, photosynthetic activity and productivity at different scales from leaf to the canopy level, particularly for plants experiencing drought stress (Sun et al., 2019; Yang, 2019). Recently, PRI was viewed as the most promising index to evaluate vegetation physiology status for global scales (Coops et al., 2010), owing to a fact that satellite based hyperspectral remote sensing data is available, e.g., NASA's 60 m resolution, 3 week temporal resolution, 10 nm spectral band HyspIRI, Germany EnMAP with 30 m, 4 days resolution (Coops et al., 2010), and sound relationship between SIF (sun induced fluorescence) and PRI had been revealed (e.g., Wang et al., 2020; Yan et al., 2018). Therefore, PRI is suitable for studying tree responses to flooding and drought characteristic of semi-arid areas. Data-driven models parameterise soil reflectance variations due to soil related factors including soil moisture (Baret et al., 2005). A robust retrieval of tree parameters requires multi-spectral remote sensing indices.

Some of the commonly used optical satellite sensors include coarse AVHRR, VEGETATION (SPOT-VGT-5 and SPOT-VGT-7), moderate MODIS-NDVI, medium Landsat (TM, ETM+ and OLI), high spatial resolution space-borne (IKONOS and QucikBird), air-borne (HyMap) and ground-based spectral measurements (Eisfelder et al., 2012). Space-borne radar used to determine AGB (C and L bands) and airborne multipolarimetric. SAR C, and L bands ERS1/2 and Japanese Earth Resources Satellite (JERS-1; JERS-2), advanced synthetic aperture radar (ASAR) (e.g., ENVISAT C-Band), phased array type L-band SAR (PALSAR; ALOS L band), passive SSM/I and AIRSAR (airborne C, L and P bands) used to determine AGB with various polarisation modes namely HH, HV, VV and VH in different availability and combinations by sensor from one to quad-polarization. Mitchard et al. (2009) predicted tree AGB for Africa found that large trees strongly correlated with cross polarised data (HV) using ALOS PALSAR data with logarithmic fitting models.

The fusion of both optical with microwave data is also essential in tree parameter estimation using remote sensing indices. For instance, Cao et al. (2018) derived eight texture variables in the Grey Level Co-occurrence Matrix and computed four topographic derivations from LiDAR with RVI,

NDVI, SAVI and MSAVI from optical satellite (ZiYuan-3) to model forest AGB using Random Forest, Support Vector Machine, k-nearest neighbours (k-NN), back propagation NN and generalized linear mixed model (GLMM) machine learning algorithms in semi-arid Chinese forest. Herrmann et al. (2011) proved the utility of superspectral Vegetation and Environmental New micro-Spacecraft (VENμS) satellite with Sentinel-2 to predict LAI using NDVI, and the Red Edge Inflection Point (REIP).

The adoption of remote sensing indices in tree mapping is pervasive. Remote sensing indices based on multi-temporal optical data are often constrained by weather, especially cloud cover (Ety et al., 2021), which may distort the remote sensing index expressions in multi-temporal studies. Croft et al. (2014) for instance provided a comprehensive list of remote sensing indices, grouping them into ratios, normalized differences and red-edge. These remote sensing indices are derived from the distinct wavelength specific interactions and response of the leaf foliar pigments, namely, chlorophylls (C_{ab}), carotenoids (C_{x+c}) and anthocyanins.

In temporal vegetation analyses linked to FRSM time series using remote sensing data, phenology is an important measurable variable. Plant phenology is the timing of cyclic developmental events (Chen et al., 2020), which Kaspar et al. (2014) qualified as the temporal physiologic dynamics of plants as marked by seasonality that constitute bud-burst (BB), green-up, flowering and senescence. Phenology is a stable tree feature useful to identify various tree characteristics such as growth, shape and colour (Kou et al., 2017). Trees phenological pattern includes a stable feature such as leaf-on and leaf-off, depicts adaptation to physical environment and climate (Kou et al., 2017). At field scale, temperature, precipitation, and soil factors can be held constant, making flood extent a principal factor for soil moisture differentiation. Since FRSM may last over a season, phenology is essential to trace the influence of floods on trees on the floodplain using remote sensing indices.

Phenology is often studied using remote sensing indices that include NDVI, EVI and SAVI (Croft et al., 2014; Kou et al., 2017; Lu et al., 2015; Yan et al., 2015). Lu et al. (2015) for instance, applied NDVI, EVI and SAVI using MODIS data to determine phenology characterising start of season and end of season based on a 30% threshold of left and right minimum of the amplitude of the time series, while Kou et al. (2017) mapped tropical forest using phenology-based method after integrating MODIS and Landsat data using NDVI, EVI and land surface water index (LSWI)

(Chandrasekar et al., 2010). Yan et al. (2015) distinguished forest phenology based on NDVI and EVI time series using decision tree classifier on MODIS data by quantifying phenometrics. The spring green-up and senescence are principal indicators of phenology based on the onset and offset of photosynthetic processes while the most popular phases include flowering, rolling out of leaves, start of season, mid of season and end of season. According to Yan et al. (2015), different growth trajectories can be characterized through harmonic components namely, mean annual, amplitude, and phase of a Fourier Transform, which show productivity and are distinguishable using analyses of satellite imagery. Fourier Transform harmonic analyses transform growth trajectories of different plants through amplitude and phase images to find components of different frequencies (Yan et al., 2015). The Fourier Transform is defined by Yan et al. (2015) in Equation 2.6;

$$x_t = a_0 + \sum_{i=0}^{N=1}(a_i \cos 2\pi f_i t + b_i \sin 2\pi f_i t) \quad (2.6)$$

where x_t is the time series of remote sensing index, f_i is the frequency, t is the time, N is the length of the sequence, and a_0 is the residuals, a white noise sequence and equal to the average of the time series, x_t. Parameters a_i and b_i are estimated using least squares (Yan et al., 2015). The assumption is that different tree types have different growth trajectories, showing temporal variations, hence different harmonic components (Liu et al., 2006 in Yan et al., 2015). One species inside and outside flooded areas depict different growth trajectories due to FRSM influence. However, NDVI time series is noisy (Hird and McDermid, 2009), hence, Viovy et al. (1992) suggested the use of a best index slope extraction (BISE) method to maximum value composite. Principal component analysis based on spectral and NDVI data is more sensitive to flood affected trees than those based purely on spectral-temporal change and NDVI change detection data while, NDVI image differencing proved superior in quantifying post-flood tree changes (Michener and Houhoulis, 1997). Both short and long-term dynamics are important to denote post-flood phenological phases. For instance, Chaber and Badeck (2002) studied the timing of both short-term variations and long-term changes on the various phases of onset of growing season resulting from climate change.

Phenological differences are useful for digital image classification (Kou et al., 2017). Yan et al. (2015) used phenometrics applying decision tree classification of image to distinguish tree species based on remote sensing indices. The remote sensing indices can use either traditional or modern machine learning methods (Camps-Valls, 2009). Image regression enables an understanding of spatiotemporal tree dynamics after a flood event. For biophysical parameters extraction,

reflectance is first mathematically computed into remote sensing indices (Myneni et al., 1995) before analysis. The phenology of deciduous trees is easier to compare between the flooded and non-flooded zones since green tissue is present during part of the year. Green-up and senescence are critical to understand changes in trees, monitoring of individual species and comparing measurements before and post-flood events inside and outside the floodplain. The use of TIMESAT statistical tool for time series analysis is important using NDVI and WDRVI remote sensing indices (Jönsson et al., 2010). Phenology is an important long-term indicator that can be useful to understand the influence of FRSM in semi-arid floodplains, which is presented in Section 2.2.3.2.

Once a remote sensing index is computed, image analyses such as classification can be undertaken (Yang et al., 1993). Efficient data-driven machine learning approaches are applied to analyse big data (Lary et al., 2016). Machine learning maps high dimensional data into a lower dimension, while preserving the main features of the original data for better analysis and visualization (Camps-Valls, 2009). These machine learning techniques offer accurate semantic interpretation of the data (Hamida et al., 2018). Furthermore, the advent of computer technology, with progressive improvements in computing, has also influenced use of more complex methods of image classification from traditional to more complex machine learning like Neural Networks, Decision Trees, Support Vector Machine, Artificial Neural Network and more recently Deep Neural Network.

2.2.3. Linking floods and FRSM to tree biophysical parameters using remote sensing indices

Floods, FRSM and floodplain trees are intricately linked, with association ranging from short- to long-term time frames (Mohammadi et al., 2017; Sera et al., 2001). Both short- and long-term effects of FRSM on trees are detailed in Sections 2.2.3.1 and 2.2.3.2, respectively. Mavhura (2019) noted that floods recharge soil moisture, which persists over an entire growing season, while Dai et al. (2019) and Jiao et al. (2019) note that precipitation recharge depletes quickly resulting in drought. Using Sentinel-1 and COSMO-SkyMed SAR data, post-flood image showed saturated soil using the σ^0_VV polarisation backscatter signal analyses (Boni et al., 2016). Rood et al. (2019) deduced rhythmic post-flood colonisation patterns using mapping. Therefore, the elevated FRSM resulting from the inundation recharge is useful to determine FRSM impact on trees (Fischer et al., 2021). However, pre-flood soil moisture also influences flooding. For instance, El Saadani et

al. (2021) noted that soil moisture determines the likelihood of flooding while Wasko and Nathan (2019) concluded that changes in soil moisture and rainfall influence flood trend in Australia. Figure 2.1 depicts the conceptual comparison of the post-flood results between precipitation-induced infiltration soil moisture and precipitation plus deep infiltration FRSM on depth for semi-arid local area south of the Equator.

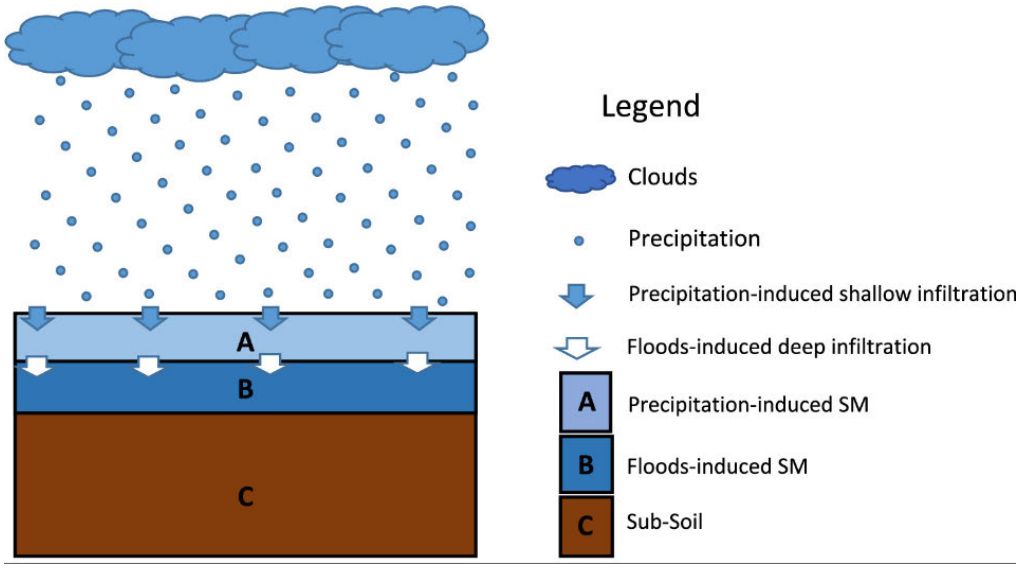


Figure 2.1: Conceptualization of soil moisture (SM) depth for locations outside flood zone (A) and inside flood zone (A + B) recharged through shallow infiltration by precipitation only and precipitation plus deep infiltration, respectively

Precipitation-recharge penetrates to shallower depth through only infiltration as compared to flood influenced combining both precipitation and flood effects. The flood influences deep infiltration effect which increases the depth of soil moisture in the flood zone.

A universal remote sensing index to monitor water, soil moisture and vegetation that simultaneously combines the NDWI, normalized difference snow index (NDSI) and the NDVI; named NDXI was proposed by Takeuchi and Yasuoka (2004). The NDXI is defined as a definite integral function with NIR and red bands as upper and lower bounds respectively as in Equation 2.7:

$$\acute{\sigma} = \int_{\rho_1}^{\rho_2} \Phi(\rho)\, \sigma(\rho) d\rho \qquad (2.7)$$

Where: $\acute{\sigma}$ is the NDXI, ρ_1 and ρ_2 are red (630 nm) and NIR (860 nm) bands $\Phi(\rho)$ and $\sigma(\rho)$ are the spectral response of sensor and target reflectance respectively. For the detection of water, soil and vegetation, the NDWI, NDSI and NDVI have the highest values, respectively.

2.2.3.1 Short-term impacts of FRSM on trees species biochemical characteristics

In a semi-arid region, a soil moisture experiment showed differential influence on trees during post dry season (Lima et al., 2021), while tree species that tolerate seasonal floods tend to dominate the floodplain ecosystem (Lopez and Kursar, 1999). To show the importance of floods to trees, Wallace et al. (2021) noted that in areas of changed flood regimes, tree die-back was widespread. In a semi-arid floodplain study in Australia, Akeroyd et al. (1998) showed that *Eucalytptus* depicted a short-term decline in post-flood water stress, while recently, Fernando et al. (2021) noted that floods influence FRSM and endemic *Eucalyptus* species. The FRSM prolongs for several months promoting vegetation productivity longer than by rainfall which is quickly depleted by high evapotranspiration in semi-arid regions (Dai et al., 2019; Jiao et al., 2019; Parsons and Thoms, 2013). Powell et al. (2014) deduced a significant post-flood positive NDVI feedback response for over 1 month, while Mohammadi et al. (2017) and Croft et al. (2014) concluded that FRSM and trees response persisted for over 6 months. Sims and Colloff (2012) found that floods trigger positive ecological feedback in semi-arid floodplains showing almost 20% post-flood increase in NDVI observable for over 10 months. Therefore, the short-term link between FRSM and trees can be understood based on a seasonal time scale. The short-term persistence of both precipitation-only versus precipitation plus FRSM is shown on Figure 2.2.

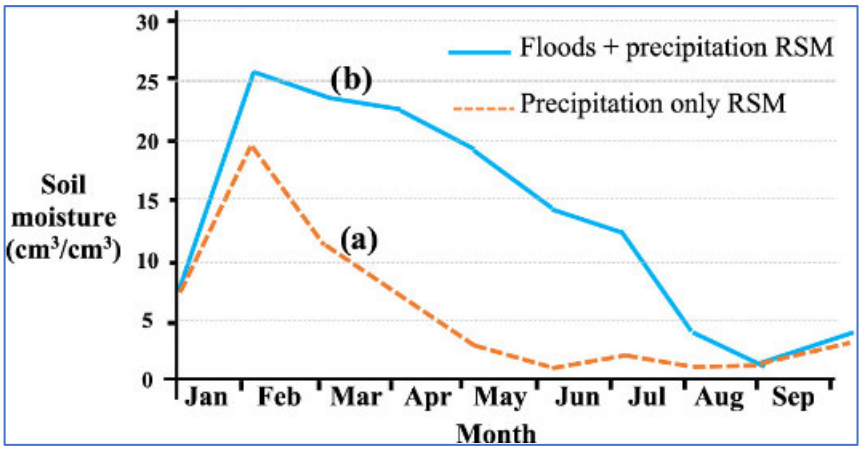


Figure 2.2: Conceptualization of short-term precipitation only (a) versus flood plus precipitation (b) recharged soil moisture

As shown on Figure 2.2, the short-term temporal persistence of high amount of soil moisture from flooded area takes much longer than that of precipitation only recharged soil moisture adjacent area.

As aforementioned, a comparison of the phenological behaviour of trees inside versus outside a flood zone enables detection of FRSM impact on trees. Mohammadi et al. (2017) for instance, presented a time series plot that showed floods of 2004 in Cooper Creek floodplain being correlated with a matching sharp rise of both the mNDWI and LSWI that corresponds with onset of flooding. Croft et al. (2014) used a flood-event based analysis to understand the impacts on soil moisture on trees, while Parsons and Thoms (2013) used NDVI to distinguish vegetation greenness during flood, rain and dry episodes to establish that floods were more influential to floodplain greenness than rainfall. Hence, on the short term, comparison of the chlorophyll and LAI inside as opposed to outside the floodplain using remote sensing indices gives an indication on seasonal influence of FRSM on trees.

Since soil moisture has been established to influence tree growth (Kerr, 2007), the use of time series analyses of species data to understand seasonal progression of trees is common (Lu et al., 2015). To evaluate the influence of site characteristics on trees, Jönsson et al. (2010) used TIMESAT, a statistical tool for time series data analysis of remote sensing indices by using annual amplitude of NDVI and WDRVI. The green amplitude was defined as the difference between the summer maximum and the eight-day period index values before budburst and results were expressed as the mean value for each vegetation zone (Jönsson et al., 2010). Zhang et al. (2006) fitted piecewise logistic models on MODIS time series data to conclude that key transition dates in annual vegetation growth cycle can be estimated in an ecologically realistic manner.

Additionally, short-term temporal chlorophyll content monitoring is also important to show the effect of FRSM on trees on the floodplain (Akeroyd et al., 1998). Therefore, temporal seasonal chlorophyll content of trees in the flood zone is higher and takes longer than that of trees outside the flood zone (Lima et al., 2021; Lopez and Karsar, 1999; Wallace et al., 2021). The deciduous trees outside the flood zone suffer from soil moisture deficit earlier than that of trees inside the floodplain and hence shed leaves earlier. The use of remote sensing indices can therefore be useful to monitor phenology of trees inside as opposed to those outside the floodplain. The green-up season is shorter outside the floodplain as compared to that of that inside.

2.2.3.2 Long-term impact of FRSM on trees species biophysical characteristics

There are many examples of the long-term impacts of FRSM. Yang et al. (2012) who used NDVI to determine tree cover dynamics in drylands defined long-term as a period of more than 5 years. Norman et al. (2014) used long-term period of nearly three decades of Landsat TM temporal that applied NDVI computed from dry-season months data. Floods facilitate colonisation of semi-arid floodplains by new species which later establish to become forests in the long term as noted in South Africa (Muthuri et al., 2005). Again, in South Africa, floods promoted the recruitment of seedlings of *Accacia karoo* tree that eventually increase forest size (Child et al., 2010). Either time series analyses or change detection of remote sensing indices maps can therefore be used to understand this long-term impact of FRSM on riparian trees. Pettit and Naiman (2006) concluded that wood deposited by floods promotes regeneration of trees and nutrient deposition. While a flood regime is important to determine tree establishment, however, where trees that tolerate frequent flooding exist, those favouring prolonged flooding may not (Hughes, 1988).

Previous floods depict a background effect promoting residue floodplain plant vitality and biomass (Croft et al., 2014). Fu and Burgher (2015) used two decades NDVI time series to show that floods are an important indicator that maintain floodplain trees integrity over time. Mohammadi et al. (2017) also use the NDVI, mNDWI and LSWI as vegetation, water, and soil moisture proxies for a 13-year window daily time series to show that flooding, FRSM and vegetation response are significantly correlated on a long term. The biophysical parameters of trees such as DBH, canopy height and spread are ideal indicators of long-term impact of recurrent FRSM within the floodplains as opposed to soil moisture deficits outside the floodplains (Li et al., 2018; Suarez et al., 2015). To understand image spectral response on which remote sensing indices are derived, radiative transfer models (RTMs) are essential. Hence, Casanova (2015) noted that a species-specific flood frequency ranging from 1 to 10 years is ideal for sustenance of different gum trees in Northern Australian floodplains, although permanent flooding kills trees on the floodplain. Precipitation triggered floods tend to be repetitive (Cassanova, 2015). Therefore, over a longer time-frame, the repeated replenishment and longer soil moisture regimes in the flood prone area influences noticeable differences in tree biophysical characteristics as compared to the shallower and lower precipitation induced soil moisture areas. Figure 2.3 illustrates the effect of flood on riparian trees as compared to the adjacent non-flooded floodplain area in a semi-arid area.

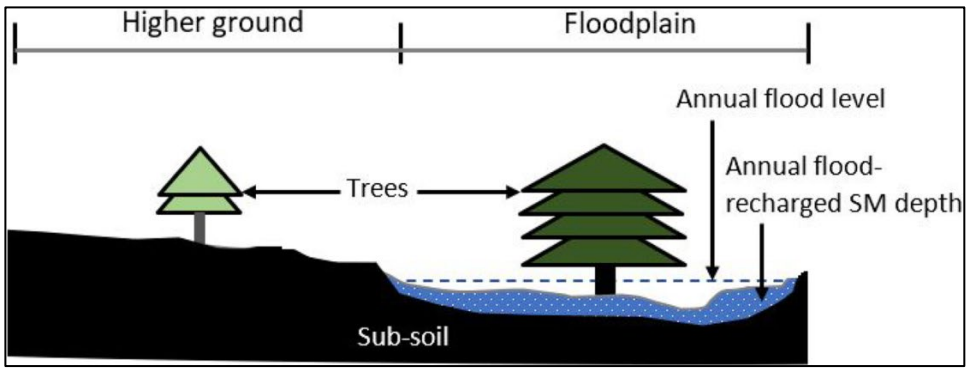


Figure 2.3: Conceptualization of long-term effect of floods on structure of vegetation based on height, canopy area and diameter at breast height.

Figure 2.3 shows that trees in frequently flooded area benefit more prolonged higher soil moisture (SM) regimes in comparison to those immediately beyond the flooded. In regularly flooded and well drained floodplains, mature evergreen trees dominate either higher micro-topography (e.g., anthills) or those where the water-table is closer to the surface, since forests largely depend on ground water (Hughes, 1988), which floods replenish through promoting deep infiltration (Colloff and Baldwin, 2010; Dahan et al., 2007; Dai et al., 2019; Jiao et al., 2019; Mohammadi et al., 2017). Hughes (1988) depicted that floodplains sustain evergreen trees closer to levees that may become forests that stretch for several kilometres sustained by the high water-table while, in poorly drained soils, floods support thorn trees. Additionally, in irregularly flooded areas, specific species e.g., *Ficus*, *Kigelia* and *Acacia*, were observed in Zambia (Michelmore, 1939, cited in Hughes, 1988). The duration, frequency and magnitude of inundation, climate, soil parameters (depth, porosity and bulk density), sediment deposition and local gently sloping terrain promote high retention of floodwaters influence the distribution of trees in semi-arid African floodplains (Hughes, 1988). Due to flooding, tall and dense tree assemblages exist even in the most arid-Sahel region such as the Tana floodplain with low rainfall (averages ~400/year) on levees (Hughes, 1988). However, biophysical characteristics are species dependent. For instance, *Ficus* species tend to be strictly riparian, unlike *Tamarindus indica* that is sustained by higher rainfall hence independent from FRSM (Hughes, 1988). The repeated soil moisture recharge in floodplain benefits trees in the area, while the precipitation recharged area suffer moisture deficiency in the semi-arid areas; hence, trees depict stunted growth thus have smaller canopies, shorter height, and smaller DBH. Therefore, the impact of FRSM on trees is also evident on the long-term.

Important remote sensing indices that inform understanding of long-term impacts of recurrent flood utilise RTMs. Some RTMs simulate canopy and atmospheric response to radiation (Verrelst et al., 2016). Models include leaf, for instance PROSPECT (Feret et al., 2008), canopy, such as SEMIDISCRETE (Gobron et al., 1997) and SAIL (Verhoef, 1984). There are also some coupled leaf-canopy models such as PROSAIL that combines PROSPECT with SAIL (Jacquemoud et al., 2009). Lu et al. (2021) examined the most widely used leaf and canopy level PROSPECT and PROSAIL RTMs to estimate the biophysical parameters of vegetation. However, the challenge with RTMs is the high computational cost (Gu et al., 2019; Verrelst et al., 2017). Therefore, emulators which are surrogate models are gaining acceptance and becoming more preferred than RTMs (Verrelst et al., 2016). Such emulators convert biophysical input variables and output into spectral data (Verrelst et al., 2016). They are easy to execute, computationally efficient and accurately approximate RTMs with less input and output pairs, yet with nearly similar results at a fraction of computational cost (Gómez-Dans et al., 2016; Johnson et al., 2020). Examples of emulators include maximum likelihood estimation (Gu et al., 2019), Bayesian (Hankin, 2005), Robust Gaussian Stochastic Process (RobustGaSP) (Gu et al., 2019) and parallel partial Gaussian stochastic process (PPGaSP) (Gu and Berger, 2016). The use of RobustGaSP package is an impressive scientific breakthrough in reducing computational demands while preserving modelling efficiency (Gu et al., 2019). In brief, machine learning regression algorithms can be useful emulators for vegetation biophysical parameters such as LAI and chlorophyll (Verrelst et al., 2017). Floods promote FRSM which influences vegetation vigour, whose leaf and canopy properties are modelled using either RTMs or the emulators. Trees respond to floods through FRSM is shown in Croft et al. (2014). Emulation stability and efficiency in prediction is, however, dependent on parameters and the model applied. Furthermore, emulators can be unstable when correlation is either very weak or very strong (Gu et al., 2019); hence, their usage should be precautious.

2.3. Challenges in application of remote sensing indices on the impact of FRSM on trees

There are several challenges associated with application of remote sensing indices from the growing remote sensing data (Dorigo et al., 2007). While the use of higher spatial resolution SAR data such as COSMO-SkyMED (Pierdicca et al., 2013) and TerraSAR-X (Pradhan et al., 2016) with 3 m resolution have proved the robustness of the normalized difference flood index (NDFI) (Cian et al., 2018a), such validation is constrained by the prohibitive data costs from commercial

systems. However, use of tools such as BISE outliers (Lange et al., 2017) may be of interest in small area FRSM. Whereas dense time series analyses are useful in understanding high temporal resolution soil moisture dynamics, they are computationally expensive, hence the need for emulation especially using free data.

Tree canopy obstructs flood and FRSM detection beneath. While Cian et al. (2018a) mapped flood under trees using normalized difference flood in vegetated areas index (NDFVI), Shen et al. (2019) concluded that dense tree canopies obscure flood and FRSM below. Based on a review by Tsyganskaya et al. (2018), the C-band data is less suitable than L-band to detect flood below trees due to technical wavelength effect. Lange et al. (2017) noted two challenges when using satellite data. Firstly, the coarse spatial resolution leading to mixed pixel problem where the individual pixel represents a complex mixture of different land use land cover types. Second, non-tree materials such as atmospheric components, soil moisture, viewing geometry and illumination of sensor alters spectral signature of the coarse pixel. Despite the development of robust algorithms and improved workflows architecture (Rahman et al., 2019; Li, Wang et al., 2018), remote sensing methods still suffer from either under- or over-prediction of inundated areas, thereby raising accuracy concerns (Shen et al., 2019). Many broadband remote sensing indices such as NDVI use red and NIR bands, which diminish in sensitivity as LAI saturates with tree density (Haboudane et al., 2004). Hence, hyper-spectral remote sensing indices has become one of the viable options (Dorigo et al., 2007). Several red-edge indices have also emerged, especially with the launch of Sentinel-2 (Evangelides and Nobajas, 2020). However, each remote sensing index is formulated for particular purpose, hence cannot be compared to another (Dorigo et al., 2007). Generally, the performance of remote sensing indices depends on many factors including trees, view geometry and other variables (Dorigo et al., 2007). Therefore, the delineation of flood extent, useful to demarcate the spatial FRSM with impacts on woody trees' biophysical parameters including LAI, chlorophyll, DBH, canopy height and spread requires an appropriate choice of remote sensing data type of adequate resolution or polarization and relevant remote sensing index. Since remote sensing indices are derived from sensors with different properties, including number of optical bands or active polarisation modes from optical or active sensors respectively, resolutions or polarisation influence the quality of the output. For example, calculating chlorophyll content using MODIS, Landsat, and SPOT basing on one index calculated using such different sensors results in inconsistencies in both spatial and temporal scales.

Notably, there is increasing application of machine learning algorithms on data to derive remote sensing indices to understand complex scientific challenges. The machine learning algorithms can efficiently handle big and high dimensional spatial data to parsimoniously model complex patterns with much higher accuracy than traditional methods (Maxwell et al., 2018). For example, Zhao et al. (2022) noted that Random Forest is a popular machine learning classification scheme widely used for remote sensing image processing. Stojanova et al. (2010) determined height and canopy size of individual trees using machine learning with LiDAR data applying 3D extraction. Liu et al. (2018) highlighted the progress and prospects of machine learning implementation on trees using three commonly used methods namely, decision tree learning, Artificial Neural Nework and Support Vector Machine, and their applications in five different aspects of forest ecology over the last decade. Zeraatpisheh et al. (2019) used machine learning of soil organic carbon, calcium carbonate equivalent and clay content using digital soil mapping approaches in central Iran. Brandt et al. (2020) used remote sensing-based NDVI implemented through supervised deep learning to count individual tree canopies which depicted high NDVI values during the dry season. Deval and Joshi (2022) tested three supervised machine learning algorithms to map tree species in a semi-arid Indian forest. Although these benefits are evident, there are also challenges in using both machine learning and deep learning to characterize tree parameters.

There are several challenges associated with the use of remote sensing indices to understand the flood-FRSM-trees nexus applied using machine learning algorithms. For instance, Maxwell et al. (2018) reviewed Support Vector Machine, k-Nearest Neighbour, Random Forest, Artificial Neural Netwok, single and boosted decision trees to conclude that their implementation is not straight-forward, yet literature is conflicted on critical issues including choice of best algorithm, calibration and validation data requirements, computational costs, user-defined parameterization and optimization affecting accuracies. Furthermore, ensembles are more accurate yet deal with weak classifier assemblages (Maxwell et al., 2018).

2.4 The future of remote sensing indices to assess impacts of FRSM on riparian trees

The developments in Earth Observation (EO) technologies aim to improve instrumentation specifications to achieve higher data quality (Zhao et al., 2022). Typical examples include the ESA space roadmap (Atzei et al., 1996), the Germany hyperspectral Environmental Mapping and

Analysis Program (EnMAP) (Foerster et al., 2016; Guanter et al., 2015) and NASA's Landsat 9 missions (Zhao et al., 2022). Accordingly, as of October 2020, the Committee on EO Satellites recorded 327 satellite missions, 845 instruments either operational or to be launched in less than 2 decades from only 30 affiliate national space agencies (Zhao et al., 2022). The recently launched Landsat 9 OLI-2 is a continuity mission for Landsat 8 OLI (Markham et al., 2020; Zhao et al., 2022). The new sensor technologies carry recent instruments, e.g., microwave radiometers, space-borne radar and SAR scatterometers and laser scanners, with improved data quality (Zhao et al., 2022) in addition to the upgraded payload (Markham et al., 2020), communications and powering in planned missions as noted for ESA (Atzei et al., 1996; Rosenqvist et al., 2003). The respective potential uses of remote sensing technologies to study the impacts of FRSM on riparian trees are outlined and exemplified.

First, there is an increase in remote sensing technology including platforms (ground-based, airborne and space-borne), sensor detector array systems and scanning mechanisms (Markham et al., 2020; Zhao et al., 2022). Remote sensing platforms have evolved from the ancient bird and kite platforms to the current development of miniature; micro- and nano-satellites carrying payloads of <10 and 10–100 kg, respectively (Esper et al., 2000), were predicted to shape future EO (Zhao et al., 2022). As such, the NASA's versatile instrumentation in the ultraviolet, visible and infrared hyperspectral spectrometers (Marzioli et al., 2020) is promising to study including FRSM impact on trees in semi-arid areas. Additionally, the current boom in drone data for local application is noticeable (Dufour et al., 2013). Since the first air photograph captured from a balloon in 1858, the sensor technologies have continually developed to further improve the data quality through improved resolutions (Zhao et al., 2022) hence can enable detection of individual tree canopy parameters, such as height, size, pigment levels and at the shortest time interval of data collection. The military–civilian partnerships as well as government-to-government help to expand the horizon of applications (Atzei et al., 1996). Notably, the increase in both operational and planned EO satellite missions in the private, public and commercial services domains in passive optical and thermal and active microwave regions, especially as planned by leading data providers (Foerster et al., 2016; Masek et al., 2020) is promising for future use of remote sensing indices, e.g., NDVI, PRI, SIPI and SAVI, to understand FRSM influence on trees. For example, increases in the multispectral Landsat 9 (Masek et al., 2020), superspectral WorldView 3 and COSMO-SkyMed (Zhao et al., 2022), the Vegetation and Environmental New micro Spacecraft

(VENμS) (Herrmann et al., 2011) and hyperspectral (AVIRIS) (Koenig and Gueguen, 2016) have been noted over the years. Zhao et al. (2022) calculated a remote sensing impact factor (RSIF) to conclude that Landsat, Sentinel and MODIS data were the most widely used satellites to date, with Sentinel predicted to become the highest by 2025 while AVHRR, SPOT and TerraSAR data showed a declining trend. The spatiotemporal ecological applications for detection, mapping, monitoring and evaluation of tree parameters, e.g., AGB, structure, phenology and forest damage, can be tenable using Landsat missions since 1972, including the longest operating sensor, Landsat 5 (TM) (Markham et al., 2020), to the most recently launched Landsat 9 (OLI-2) (Zhao et al., 2022; Zhu et al., 2019) or in fusion with other sensors due to mission continuity aspect to understand FRSM impact on riparian trees using remote sensing indices.

Second, the above technological improvements result in data deluge which coincides with development of efficient computer science data processing and analyses algorithms. The big Earth data (Gorelick et al., 2017; Zhao et al., 2022) is regarded as the new engine to generate Earth science knowledge provoking commensurate technological and methodological developments (Guo, 2017) with challenges and opportunities (Guo et al., 2017). The synergistic coincidence of data explosion and super-computing environment, especially the quantum computers help to expedite data access and processing as provided by the Google Earth Engine (GEE) including free access to customizable codes (Gorelick et al., 2017). Furthermore, the internet of things has availed an enabling environment with growing online users interacting virtually through resources such as Zoom and Youtube, which enrich the sharing of data, access resources, (pre)processing and analyses codes as exemplified in Khan and Memon (2021). Henceforth, the policy of free data especially by Landsat (Zhu et al., 2019) and Sentinel (Showstack, 2014.) to wider user community coupled with the emergence and expansion of artificial intelligence especially machine learning and deep learning algorithms shows good prospects for EO. These algorithms have a paucity to efficiently analyse big spatial data ranging from optical, thermal, microwave radar and SAR bands with single- (VV or HH), to quad-polarizations (HH + HV + VV + VH), and LiDAR (Rosenqvist et al., 2003). Additionally, there are increases in passive and active data fusion methods (Rosenqvist et al., 2003; Shao et al., 2020) executed at pixel, feature and decision levels (Joshi et al., 2016). Furthermore, there are algorithm ensemble methods, new remote sensing indices based on the new vegetation bands, e.g., the red edge; RE_1, RE_2 and RE_3 to estimate tree parameters including LAI, NDVI, MSR, Chlorophyll indices and WDRVI were used after the launch of

Sentinel-2 (A and B) (Sun et al., 2019) and researches using the new data and developing new algorithms are also promising (Guo, 2017; Zhao et al., 2022) to study FRSM impacts on trees. Therefore, the increase in free remote sensing data is promising for EO applications on tree parameters. The online open big data policy, data access and analyses algorithms especially on the GEE (Gorelick et al., 2017; Zhao et al., 2022) are positive developments to help understanding of FRSM on riparian trees. Recently, some applications were done.

There are some examples to inform improvements indicative of the trajectory on studying riparian trees using remote sensing indices data including the use of PlanetScope, WorldView 3, and UAV data. Baloloy et al. (2018) used optical multispectral RapidEye, PlanetScope and Sentinel-2 data, to derive NDVI, Green-NDVI (GNDVI), SAVI, SR, Red Edge SR to model LAI, fractional vegetation cover, leaf chlorophyll content (C_{ab}), and fraction of absorbed photosynthetically active radiation (FAPAR) biophysical variables of mangrove tree AGB. Recently, PlanetScope data were used to derive phenology (Cheng et al., 2020; Francini et al., 2020). Bradshaw et al. (2019) related remote sensing indices to drone and PlanetScope data in southern African semi-arid areas, while Wu et al. (2021) monitored tree-crown scale leaf phenology using PlanetScope and drone data based on remote sensing indices. Collin et al. (2018) used superspectral WorldView 3 data to map vegetation height and composition using remote sensing index. Furthermore, then on-going researches (e.g., Atzei et al., 1996) helped to study the impacts of FRSM on riparian trees.

Huylenbroeck et al. (2020) noted that platform and sensor characteristics determine the type and scale of application with airborne sensor data often having local applications while space-borne have local to regional coverage. The other factors influencing sensor application include areal extent, indicators of species, with LiDAR data being ideal for structure (e.g., height, DBH, density, biomass and roughness) and physiology (e.g., evapotranspiration, health and phenology) (Huylenbroeck et al., 2020). On the study of physiology, Dufour et al. (2013) used data from technologies including LiDAR, radar and UAV to extract 3D canopy structure and height. Specifically, optical data, e.g., IKONOS and SPOT-5 WorldView3, can be useful to determine tree species composition, while AVHRR is ideal for tree phenology (Huylenbroeck et al., 2020). The radar (<2%) application is decreasing and being replaced by LiDAR technology, and the recent increase of drone data is noticeable (Huylenbroeck et al., 2020).

However, the exorbitant investment costs of space technology development and deployment remain a challenge in this regard (Zhao et al., 2022). Furthermore, it is difficult to study trees due to their biogeochemical functionality and complexity in their spatial structure (e.g., shade, height, roughness and linearity of corridors) (Huylenbroeck et al., 2020). Overall, the implementation of the new missions adds to data quality and quantity, thereby facilitating the continuity needed in environmental change studies (Zhao et al., 2022).

Specifically, technological innovation reducing the cost, weight and size of devices towards nanotechnology such as wireless smartphone spectrometer with wireless connectivity and graphical output visualization is promising (Lange et al., 2017). Super-computing in connection with increasing satellite missions point towards Earth Observation big data that require heavy pre-processing, especially for SAR data which is more applicable to cloud-contaminated data or night-occurring floods is expected (Giachetta, 2015; Ma et al., 2015). Hence, the coupling of the GIS visual analytics toolbox with remote sensing indices derived from big data will be necessary (Li et al., 2016). In particular, the deployment of geoprocessing frameworks e.g., MapReduce (Giachetta, 2015) and GEE with free access and efficient processing and analyses benefits, shows that technology is increasingly becoming cheaper, portable and affordable (Das and Ghosh, 2016) and hence will avail more opportunities for remote sensing index-based methods for flood, FRSM and trees related applications. Such frameworks make it imperative to enhance accuracy and cost-effective modelling of flood characteristics through application of remote sensing indices (Rahman and Thakur, 2018). The deployment of operational high spatial resolution SAR data by ESA's twin constellations with Sentinels-1A and-1B offer opportunities to investigate complex challenges such as mapping cloud obscured or night-time occurring flood events which is not possible with optical remote sensing indices.

Emulation, with computation economy as reported by Gómez-Dans et al. (2016) is relevant to synthesise floods, FRSM and trees tasks. Comparison of machine learning outputs is attractive in view of the benefits, especially in regression and classification studies (Lary et al., 2016). Partial least squares regression and Artificial Neural Networks are promising alternatives with data fusion across the active/passive remote sensing divide promising (Petropoulos et al., 2015). Kattenborn et al. (2021) noted that, although Convolution Neural Network is new in vegetation studies, it has high agility and adaptable architectures, hence applicable on remote sensing indices. Furthermore, deep learning as the recent best benchmark in machine learning (Alzubaidi et al., 2021) has a huge

potential to further revolutionise human knowledge on flood impacts on riparian trees in semi-arid floodplains if combined with remote sensing indices.

2.5. Conclusions

Floodplain trees are highly productive and rich in biomass, which provides ecosystem services that include carbon sequestration, soil erosion control, removal of pollutants and mining minerals eroded upstream. These trees are also habitat for diverse wildlife. Ecologically, floods promote soil moisture recharge that provide a suitable condition for germination, regeneration, growth, preservation and sustainability of the trees on the floodplain. Floods impact on biophysical vegetation characteristics indirectly by influencing deep infiltration, which promotes prolonged FRSM. Such effects can be studied using remote sensing indices in both passive and active modes.

In this review, the connection of floods to FRSM and trees was presented. There is an increase in literature on the use of optical remote sensing indices in flood studies, soil moisture and trees, however, SAR application is constrained by limited data availability from dual polarized e.g. Sentinel missions. While many backscatter decomposition metrics exist, these often require full polarimetric data while the common dual polarized metrics seem to be under-explored. Notably, employing SAR backscatter decomposed metrics could be incorporated into the application of emerging deep learning approaches to enable spatiotemporal characterization of the biophysical effects influenced by FRSM. Increase in access to available cloud-based computing resources especially GEE offers cheap, efficient and fast data access, access to storage, processing and analyses which offers a valuable platform for easy dissemination of research results. New remote sensing indices and ensembles can be formulated based on current and planned satellite missions to understand FRSM impacts on trees. The free availability of specific canopy penetrating SAR data e.g. in L-band remains a challenge to be addressed by data providers in view of the active remote sensing. Future studies should also seek to address the challenge of floods occurring without satellite overpass as well as use of FRSM to monitor trees biophysical parameters in semi-arid floodplains taking advantage of machine and deep learning algorithms on cloud computing resources and data.

2.6 References

Akay, H., 2021. Flood hazards susceptibility mapping using statistical, fuzzy logic, and MCDM methods. *Soft Computing* 25, pp.9325–9346. doi:10.1007/s00500-021-05903-1.

Akeroyd, M.D., Tyerman, S.D., Walker, G.R., and Jolly. I.D., 1998. Impact of flooding on the water use of semi-arid riparian eucalypts. *Journal of Hydrology* 206, pp.104-117. doi:10.1016/S0022-1694(98)00092-4.

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O.,…, and Farhan, L., 2021 Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data* 8(53). doi:10.1186/s40537-021-00444-8.

Anusha, N., and Barathi, B., 2020. Flood detection and flood mapping using multi-temporal synthetic aperture radar and optical data. *Egyptian Journal of Remote Sensing and Space Science* 23, pp.207-219. doi:10.1016/j.ejrs.2019.01.001.

Baret, F., Pavageau, K., Béal, D., Weiss, M., Berthelot, B., and Regner, P., 2005. Algorithm theoretical basis document for MERIS Top of Canopy Land Products (TOC_VEG). Contract ESA AO/1-4233/02/I-LG, INRA & Noveltis, Avignon.

Bijeesh, T.V., and Narasimhamurthy, K.N., 2020. Surface water detection and delineation using remote sensing images: a review of methods and algorithms. *Sustainable Water Resources Management* 6(68). doi:10.1007/s40899-020-00425-4.

Boni, G., Ferraris, L., Pulvirenti, L., Squicciarino, G., Pierdicca, N., Candela, L., …, and Pagliara, P., 2016. A prototype system for flood monitoring based on flood forecast combined with COSMO-SkyMed and Sentinel-1 data. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing* 9(6), pp.2794-2805. DOI:10.1109/JSTARS.2016.2514402.

Breiman, L., 2001. Random forests. *Machine Learning* 45 (1), pp.5-32. doi:10.1023/A:1010933404324.

Camps-Valls, G., 2009. Machine learning in remote sensing data processing. Paper presented at the Proceedings of the 2009 IEEE International Workshop on Machine Learning for Signal Processing, Grenoble, France, September 2 - 4.

Casanova, M.T., 2015. Review of water requirements for key floodplain vegetation for the northern basin: Literature review and expert knowledge assessment. Report to the Murray–Darling Basin Authority, Charophyte Services, Lake Bolac.

Chaber, J., and Badeck, F. -W., 2002. Combining physiological time series and outlier detection. *Tree Physiology* 22, pp.973 - 982. DOI:10.1093/treephys/22.14.973.

Chandrasekar, K., Sai, M.V.R.S., Roy, P.S., and Dwevedi, R.S., 2010. Land surface water index (LSWI) response to rainfall and NDVI using the MODIS vegetation index. *International Journal of Remote Sensing* 31(15), pp.3987-4005. doi:10.1080/01431160802575653.

Chen, L., Hänninen, H., Rossi, S., Smith, N.G., Pau, S., Liu, Z., Feng, G., Gao, J, and Liu, J., 2020. Leaf senescence exhibits stronger climatic responses during warm than during cold autumns. *Nature Climate Change* 10, pp.777–780. doi:10.1038/s41558-020-0820-2.

Chen, Y., Yang, K., Qin, J., Cui, Q., Lu, H., La, Z., Han, M., and Tang, W., 2017. Evaluation of SMAP, SMOS and AMSR2 soil moisture retrievals against observations from two networks on the Tibetan Plateau. *Journal Geophysical Research: Atmospheres* 122 (11), pp.5780–5792. doi:10.1002/2016JD026388.

Choker, M., Baghdadi, N., Zribi, M., El Hajj, M., Paloscia, S., Verhoest, N.E., Lievens, H., and Mattia, F., 2017. Evaluation of the Oh, Dubois and IEM backscatter models using a large dataset of SAR data and experimental soil measurements. *Water* 9(1), 38. doi:10.3390/w9010038.

Chou, S., Chen, J. M., Yu, H., Chen, B., Zhang, X., Croft, H., Khalid, S., Li, M., and Shi, Q., 2017. Canopy-level photochemical reflectance index from hyperspectral remote sensing and leaf-level non-photochemical quenching as early indicators of water stress in maize. *Remote Sensing* 9(8), 794. doi:10.3390/rs9080794.

Cian, F., Marconcini, M., and Ceccato, P., 2018a. Normalized difference flood index for rapid flood mapping: Taking advantage of EO big data. *Remote Sensing of Environment* 209, pp.712 - 730. doi:10.1016/j.rse.2018.03.006.

Cian, F., Marconcini, M., Ceccato, P., and Giupponi, C., 2018b. Flood depth estimation by means of high-resolution SAR images and lidar data. *Natural Hazards Earth System Science* 18, pp.3063 - 3084. doi:10.5194/nhess-18-3063-2018, 2018.

Clevers, J.G.P.W., de Jong, S.M., Epema, G.F., van der Meer, F.D., Bakker, W.H., Skidmore, A. K., and Scholte, K. H., 2002. Derivation of the red edge index using the MERIS standard band setting. *International Journal of Remote Sensing* 23(16), pp.3169-3184. doi:10.1080/01431160110104647.

Colloff, M.J., and Baldwin, D.S., 2010. Resilience of floodplain ecosystems in a semi-arid environment. *Rangeland Journal* 32(3), pp.305 - 314. DOI:10.1071/RJ10015.

Croft, H., Chen, J.M., and Zhang, Y., 2014. The applicability of empirical vegetation indices for determining leaf chlorophyll content over different leaf and canopy structures. *Ecological Complexity* 17, pp.119 - 130. doi:10.1016/j.ecocom.2013.11.005.

Dahan, O., Shani, Y., Enzel, Y., Yechieli, Y., and Yakirevich, A., 2007. Direct measurements of floodwater infiltration into shallow alluvial aquifers. *Journal of Hydrology* 344, pp.157– 170, DOI:10.1016/j.jhydrol.2007.06.033.

Dai, L., Guo, X., Zhang, F., Du, Y., Ke, X., Li, Y., Cao, G., Li, Q., Lin, L., Shu, K., and Peng, C., 2019. Seasonal dynamics and controls of deep soil water infiltration in the seasonally-frozen region of the Qinghai-Tibet plateau. *Journal of. Hydrology* 571, pp.740 - 748, doi:10.1016/j.jhydrol.2019.02.021.

Das, M., and Ghosh, S.K., 2016. Deep-STEP: A deep learning approach for spatiotemporal prediction of remote sensing data. *IEEE Geoscience and Remote Sensing Letters* 13(12), pp.1984 - 1988, doi:10.1109/LGRS.2016.2619984.

Dhara, S., Dang, T., Parial, K. and Lu, X.X., 2020. Accounting for uncertainty and reconstruction of flooding patterns based on multi-satellite imagery and support vector machine technique: a case study of can Tho City, Vietnam. *Water* 12, 1543. doi:10.3390/w12061543.

Di Baldassarre, G., Schumann, G., Brandimarte, L., and Bates, P., 2011. Timely low resolution SAR imagery to support floodplain modeling: A case study review. *Surveys in Geophysics* 32, pp.255-269, DOI:10.1007/s10712-011-9111-9.

Dlamini, C., and Geldenhuys, C.J., 2009. The socioeconomic status of the non-timber forest product subsector in Swaziland. *Southern Forests* 71(4), DOI:10.2989/SF.2009.71.4.9.1036.

Dong, J., and Crow, W.T., 2019. L-band remote-sensing increases sampled levels of global soil moisture-air temperature coupling strength. *Remote Sensing of Environment* 220, pp.51 - 58, DOI:10.1016/j.rse.2018.10.024.

Dong, Z., Wang, G., Amankwah, S.O.Y., Wei, X., Hu, Y., and Feng, A., 2021. Monitoring the summer flooding in the Poyang Lake area of China in 2020 based on Sentinel-1 data and multiple convolutional neural networks. *International Journal of Applied Earth Observation and Geoinformation* 102, 102400, doi:10.1016/j.jag.2021.102400.

Dorigo, W.A., Zurita-Milla, R., W. de Wit, A.J., Brazile, J., Singh, R., and Schaepman, M. E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modelling. *International Journal of Applied Earth Observation and Geoinformation* 9, pp.165–193. doi:10.1016/j.jag.2006.05.003.

Du, X., Wang, S., Zhou, Y., and Wei, H., 2007. Construction and validation of a new model for unified surface water capacity based on MODIS data. *Geomatics and Information Science of Wuhan University* 32(3), pp.205 - 207.

Dubois, P.C., van Zyl, J., and Engman, T., 1995. Measuring soil moisture with imaging radars. *IEEE Transactions on Geoscience and Remote Sensing* 33(4), pp.915–926.

El Saadani, M., Habib, E., Abdelhameed, A.M., and Bayoumi, M., 2021. Assessment of spatiotemporal deep learning approach for soil moisture prediction and filling the gaps in between soil moisture observations. *Frontiers in Artificial Intelligence* 4(11), doi:10.3389/frai.2021.636234. Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J., Kimball, J., Piepmeier, J.R., Koster, R.D., Martin, N., McDonald, K.C., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., Spencer, M.W., Thurman, S.W., Tsang, L., and van Zyl, J., 2010. The Soil Moisture Active Passive (SMAP) mission." *Proceedings of the IEEE* 98(5), 704 - 716. doi:10.1109/jproc.2010.2043918.

Ety, N.J., Chu, Z., and Masum, S.M., 2021. Monitoring of flood water propagation based on microwave and optical imagery. *Quaternary International* 574, pp.137-145. doi:10.1016/j.quaint.2020.10.038.

Evangelides, C., and Nobajas, A., 2020. Red-edge normalised difference vegetation index (NDVI705) from Sentinel-2 imagery to assess post-fire regeneration. *Remote Sensing Applications: Society and Environment.* doi:10.1016/j.rsase.2019.100283.

Feng, Y., Brenner, C., and Sester, M., 2020. Flood severity mapping from volunteered geographic information by interpreting water level from images containing people: a case study of Hurricane Harvey. *ISPRS Journal of Photogrammetry and Remote Sensing* 169, pp.301-319. DOI:10.1016/j.isprsjprs.2020.09.011.

Fensholt, R., and Sandholt, I., 2003. Derivation of a shortwave infrared water stress index from MODIS near- and shortwave infrared data in a semiarid environment. *Remote Sensing of Environment* 87, pp.111-121. DOI:10.1016/j.rse.2003.07.002.

Feret, J.B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P.R., Ustin, S.L., Le Maire, G., and Jacquemoud S., 2008. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sensing of Environment* 112, 3030 - 3043. doi:10.1016/j.rse.2008.02.012.

Fernando, D.R., Fernando, A.E., Koerber, A.E., and Doody, T.M., 2021. Tree-soil interactions through water release to a floodplain ecosystem: a case study of black box (*Eucalyptus largiflorens*) on loamy sands. *Wetlands* 41(17), doi:10.1007/s13157-021-01419-4.

Figa-Saldaña, J., Wilson, J.J., Attema, E., Gelsthorpe, R., Drinkwater, M.R., and Stoffelen, A., 2002. The Advanced SCATterometer (ASCAT) on the meteorological operational (MetOp) platform: a follow on for European wind scatterometers. *Canadian Journal of Remote Sensing* 28 (3), pp.404 – 412, DOI:10.5589/m02-035.

Fischer, S., Greet, J., Walsh, C.J., and Catford, J.A., 2021. Restored river-floodplain connectivity promotes woody plant establishment. *Forest Ecology and Management* 493, 119264. doi:10.1016/j.foreco.2021.119264.

Forkuor, G., Zoungrana, J-B.B., Dimobe, K., Ouattara, B., Vadrevu, K.P., and Tondoh, J.E., 2020. Above-ground biomass mapping in West African dryland forest using Sentinel-1 and -2 datasets - A case study. *Remote Sensing of Environment* 236, 111496. doi:10.1016/j.rse.2019.111496.

Fu, H., and Burgher, I., 2015. Riparian vegetation NDVI dynamics and its relationship with climate, surface water and groundwater. *Journal of Arid Environments* 113, pp.59-68. doi:10.1016/j.jaridenv.2014.09.010.

Gale, S.J., and Bainbridge, S., 1990. The floods in eastern Australia. *Nature* 345, 767. doi:10.1038/345767a0.

Ghulam, A., Qin, Q., and Zhan, Z., 2006. Designing of the perpendicular drought index. *Environmental Geology 52*, pp.1045-1052. doi:10.1007/s00254-006-0544-2.

Ghulam, A., Qin, Q., Teyip, T., and Li, Z.-L., 2007. Modified perpendicular drought index (MDPI): A real-time drought monitoring method. *ISPRS Journal of Photogrammetry and Remote Sensing* 62, pp.150 - 164, doi:10.1016/j.isprsjprs.2007.03.002.

Giachetta, R., 2015. A framework for processing large scale geospatial and remote sensing data in MapReduce environment. *Computer Graphics* 49, pp.37-46. doi:10.1016/j.cag.2015.03.003

Gobron, N., Pinty, B., Verstraete, M.M., and Govaerts, Y., 1997. A semidiscrete model for the scattering of light by vegetation. *Journal of Geophysical Research* 102, 9431–9446. DOI:10.1029/96JD04013.

Gómez-Dans, J.L., Lewis, P.E., and Disney, M., 2016. Efficient emulation of radiative transfer codes using gaussian processes and application to land surface parameter inferences. *Remote Sensing* 8, 119. doi:10.3390/rs8020119.

Greifeneder, F., Khamala, E. Sendabo, D. Wagner, W. Zebisch, M. Farah, H. and Notarnicola, C., 2019. Detection of soil moisture anomalies based on Sentinel-1. *Physics and Chemistry of the Earth* Parts A/B/C 112, pp.75 – 82, doi:10.1016/j.pce.2018.11.009.

Gu, M., and Berger, J. O., 2016. Parallel partial Gaussian process emulation for computer models with massive output. *Annals of Applied Statistics* 10(3), pp.1317 – 1347, DOI:10.1214/16-AOAS934.

Gu, M., Palomo, J., and Berger, J. O., 2019. RobustGaSP: robust gaussian stochastic process emulation in R. *R Journal* 11(01), pp.1 - 24, doi:10.32614/RJ-2019-011.

Haboudane, D., Miller, J.R. Pattey, E., Zarco-Tejada, P.J., and Strachan, I.B., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modeling and validation in the context of precision agriculture. *Remote Sensing of Environment* 90(3), pp.337 - 352, doi:10.1016/j.rse.2003.12.013.

Hallberg, G.R., Hoyer, B.E., and Rango, A., 1973. Application of ERTS-1 imagery to flood inundation mapping. NASA Special Pub. No. 327, Symposium on significant results obtained from the Earth Resources Satellite-1, 1. Technical Presentations, Section A, 745 - 753.

Hamida, A.B., Benoit, A. Lambert, P., and Amar, C.B., 2018. Three-dimensional deep learning approach for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing* 56(8), pp.4420 - 4434, doi:⟨10.1109/TGRS.2018.2818945⟩.⟨hal-01814542⟩.

Hankin, R. K. S., 2005. Introducing BACCO, an R bundle for bayesian analysis of computer code output. *Journal of Statistical Software* 14(16), DOI:10.18637/jss.v014.i16.

Hashemi-Beni, L., and Gebrehiwot, A., 2021. Flood extent mapping: an integrated method using deep learning and region growing using UAV optical data. *IEEE Selected Topics in Applied Earth Observations and Remote Sensing* 14, pp.2027-2135, DOI:10.1109/JSTARS.2021.3051873.

Hird, J.N., and McDermid, G.J., 2009. Noise reduction of NDVI time series: An empirical comparison of selected techniques 113, pp.248 - 258, doi:10.1016/j.rse.2008.09.003.

Holtgrave, A-K., Förster, M., Greifeneder, F., Notarnicola, C., and Kleinschmit, B., 2018. Estimation of soil moisture in vegetation-covered floodplains with sentinel-1 SAR data using support vector regression. *Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 86, pp.85 - 101, doi:10.1007/s41064-018-0045-4.

Hong, Z., Zhang, W., Yu, C., Zhang, D., Li, L., and Meng, L., 2018. SWCTI: surface water content temperature index for assessment of surface soil moisture status. *Sensors* 18(9), 2875. doi:10.3390/s18092875.

Hounkpatin, K.O.L., Schmidt, K., Stumpf, F., Forkuor, G., Behrens, T., Scholten, T., Amelung, W., and Welp, G., 2018. Predicting reference soil groups using legacy data: A data pruning

and random forest approach for tropical environment (Dano Catchment, Burkina Faso). *Scientific Reports* 8, 9959. DOI:10.1038/s41598-018-28244-w

Huang, M., and Jin, S., 2020. Rapid flood mapping and evaluation with a supervised classifier and change detection in Shouguang using Sentinel-1 SAR and Sentinel-2 optical data. *Remote Sensing 12*, 2073. doi:10.3390/rs12132073.

Huete, A. R., 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* 25, pp.295–309. doi:10.1016/0034-4257(88)90106-X.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., and Ferreira, L. G,. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment* 83, pp.195–213. doi:10.1016/S0034-4257(02)00096-2.

Hukusima, T., and Yoshikawa, M., 1997. The impact of extreme run-off events from the Sakasagawa River on the Senjogahara Ecosystem, Nikko National Park. IV. Changes in tree and understory vegetation distribution patterns from 1982 to 1992. *Ecological Research* 12, pp.27 - 38.

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., and Ustin, S. L., 2009. PROSPECT+ SAIL models: A review of use for vegetation characterization. *Remote Sensing of Environment* 13, pp.56 - 66, doi:10.1016/j.rse.2008.01.026.

Jiang, X., Liang, S., He, X., Ziegler, A.D., Lin, P., Pan, M., Wang, D., Zou, J., Hao, D., Mao, G., and Zeng, Z., 2021. Rapid and Large-scale mapping of flood inundation via integrating spaceborne synthetic aperture radar imagery with unsupervised deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing* 178, pp.36 - 50, doi:10.1016/j.isprsjprs.2021.05.019.

Jiao, W., Tian, C., Chang, Q., Novick, K. A., and Wang, L., 2019. A new multi-sensor integrated index for drought monitoring. *Agricultural and Forest Meteorology* 268, 74-85, doi:10.1016/j.agrformet.2019.01.008.

Johnson, J.E., Laparraz, V., and Camps-Valls, G., 2020. Accounting for input noise in gaussian process parameter retrieval, DOI:10.1109/LGRS.2019.2921476.

Jönsson, A.M., Eklundh, L., Hellström, M., Bärring, L., and Jönsson, P., 2010. Annual changes in MODIS vegetation indices of Swedish coniferous forests in relation to snow dynamics and tree phenology. *Remote Sensing of Environ* 114, pp.2719 - 2730, doi:10.1016/j.rse.2010.06.005.

Jordan, C.F., 1969. Derivation of leaf area index from quality of light on the forest floor. *Ecology.* 50, pp.663 - 666, doi:10.2307/1936256.

Kang, J., Jin, R., Li, X., and Zhang, Y., 2021. Mapping high spatiotemporal resolution soil moisture by upscaling sparse ground-based observations using a bayesian linear regression method for comparison with microwave remote sensed soil moisture products. *Remote Sensing* 13(2), 228, DOI:10.3390/rs13020228.

Kattenborn, T., Leitloff, J., Schiefer, F., and Hinz, S., 2021. Review on convolutional neural networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 173, pp.24 - 49, doi:10.1016/j.isprsjprs.2020.12.010.

Kaspar, F., Zimmermann, K., and Polte-Rudolf, C., 2014. An overview of the phenological observation network and the phenological database of Germany's national meteorological service (Deutscher Wetterdienst. *Advances in Science and Research* 11(1), pp.93 - 99. doi:10.5194/asr-11-93-2014.

Kerr, Y. H., 2007. Soil moisture from space: where are we? *Hydrogeology Journal* 15, pp.117 - 120, doi:10.1007/s10040-006-0095-3.

Kerr, Y., Waldteufel, P., Richaume, P., Wigneron, J. P., Ferrazzoli, P., Mahmoodi, A., Al Bitar, A., Cabot, F., Gruhier, C., Juglea, S.E., and Delwart, S., 2012. The SMOS soil moisture retrieval algorithm. *IEEE Transactions on Geoscience and Remote Sensing* 50, pp.1384 – 1403, DOI:10.1109/TGRS.2012.2184548.

Kerr, Y.H., Waldteufel, P., Wigneron, J-P., Delwart, S. Cabot, F. Boutin, J., Escorihuela, M.J., Font, J., Reul, N., Gruhier, C., and Mecklenburg, S., 2010. The SMOS mission: New tool for monitoring key elements of the global water cycle. *Proceedings of the IEEE* 98(5), pp.666–687, DOI:10.1109/JPROC.2010.2043032.

Kou, W., Liang, C., Alexander, L., Hernandez, J., and Yang, X., 2017. Phenology-based method for mapping tropical evergreen forests by integrating of MODIS and Landsat imagery. *Forests,* 8 (34), DOI:10.3390/F8020034.

Kuenzer, C., Bluemel, A., Gebhardt, S., Quoc, T.V., and Dech, S., 2011. Remote sensing of mangrove ecosystems: a review. *Remote Sensing* 3, pp.878 - 928, doi:10.3390/rs3050878.

Kawanishi, T., Sezai, T., Ito, Y., Imaoka, K. Takeshima, T. Ishido, Y. Shibata, A., Miura, M., Inahata, H., and Spencer, R. W., 2003. The Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSRE), NASDA's contribution to the EOS for global energy and water cycle studies. *IEEE Transactions on Geoscience and Remote Sensing* 41, pp.184 - 194. DOI:10.1109/TGRS.2002.808331.

Landuyt, L., van Wesemael, A., van Coillie, F. M. B., and Verhoest, E. C., 2017. Pixel-based flood mapping from SAR imagery: A comparison of approaches. 19th EGU General Assembly, EGU2017, Proceedings from the conference held 23-28 April, 2017in Vienna, Austria., 14060.

Lange, M., Dechant, B., Rebmann, C., Vohland, M., Cuntz, M., and Doktor, D., 2017. Validating MODIS and Sentinel-2 NDVI products at a temperate deciduous forest site using two independent ground-based sensors. *Sensors* 17, 1855, doi:10.3390/s17081855.

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

Li, D., Cheng, T., Zhou, K., Zheng, H., Yao, X., Tian. Y., Zhu, Y., and Cao, W., 2017. WREP: a wavelet-based technique for extracting the red edge position from reflectance spectra for estimating leaf and canopy chlorophyll contents of cereal crops. *ISPRS Journal of Photogrammetry and Remote Sensing* 129, pp.103 - 117, doi:10.1016/j.isprsjprs.2017.04.024.

Li, S., Dragicevic, S., Castro, F. A., Sester, A., Winter, S. Coltekin, A., Pettit, C., Jiang, B., Haworth, J., Stein, A., and Cheng, T., 2016. Geospatial big data handling theory and methods: a review and research challenges. *ISPRS Journal of Photogrammetry and Remote Sensing* 115, pp.119 - 133, doi:10.1016/j.isprsjprs.2015.10.012.

Li, Y., Martinis, S., Plank, S., and Ludwig, R., 2018. An automatic change detection approach for rapid flood mapping in Sentinel-1 SAR data. *International Journal of Applied Earth Observation and Geoinformation* 73: pp.123-135. https://doi.org/10.1016/j.jag.2018.05.023.

Li, Y., Martinis, S., and Wieland, M., 2019. Urban flood mapping with an active self-learning convolutional neural network based on TerraSAR-X intensity and interferrometric coherence. *ISPRS Journal of Photogrammetry and Remote Sensing* 152, pp.178-191, https://doi.org/10.1016/j.isprsjprs.2019.04.014.

Liang, J., and Liu, D., 2020. A local thresholding approach to flood water delineation using Sentinel-1 SAR imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 159, pp.53 - 62, doi:10.1016/j.isprsjprs.2019.10.017.

Lima, A. L. A., Rodal, M. J. N., Castro, C. C., Antonio, A. C. D., Melo, A. L., Gonçalves-Souza, T., and Sampaio, E. V. S. B., 2021. Phenology of high- and low-density wood deciduous species responds differently to water supply in tropical semiarid regions. *Journal of Arid Environments* 193, 104594, doi:10.1016/j.jaridenv.2021.104594.

Liu, J., Chai L., Dong J., Zheng, D., Wigneron, J.-P. Liu, S., Zhou, J., Xu, T., Yang, S., Song, Y., and Lu, Z., 2021. Uncertainty analysis of eleven multisource soil moisture products in the

third pole environment based on the three-corned hat method. *Remote Sensing of Environment* 255. doi:10.1016/j.rse.2020.112225.

Lobell, D. B., and Asner, G. P., 2002. Moisture effects on soil reflectance. *Soil Science Society America Journal* 66(3), pp.722-727.

Long, S., Fatoyinbo, T. E., and Policelli, F., 2014. Flood extent mapping for Namibia using change detection and thresholding with SAR. *Environmental Research Letters* 9(3), 035002, doi:10.1088/1748-9326/9/3/035002.

Lopez, O. R., and Kursar, T. A., 1999. Flood tolerance of four tropical tree species. *Tree Physiology* 19(14), pp.925 - 932, DOI:10.1093/treephys/19.14.925.

Lu, L., Kuenzer, C., Wang, C., Guo, H., and Li, Q., 2015. Evaluation of three MODIS-derived vegetation index time series for dryland vegetation dynamics monitoring. *Remote Sensing* 7, pp.7597 - 7614, doi:10.3390/rs70607597.

Lu, B., Proctor, C., and He, Y., 2021. Investigating different versions of PROSPECT and PROSAIL for estimating spectral and biophysical properties of photosynthetic and non-photosynthetic vegetation in mixed grasslands. *GIScience and Remote Sensing* doi:10.1080/15481603.2021.1877435.

Ma, C., Li, X., and McCabe, M. F., 2020. Retrieval of high-resolution soil moisture through combination of Sentinel-1 and Sentinel-2 data. *Remote Sensing* 12(14), 2303, doi:10.3390/rs12142303.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., and Johnson, B.A., 2019. Deep learning in remote sensing applications: a meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing* 152, pp.166 - 177, doi:10.1016/j.isprsjprs.2019.04.015.

Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., and Jie, W., 2015. Remote sensing big data computing: challenges and opportunities. *Future Generation Computer Systems* 51, pp.47 - 60, doi:10.1016/j.future.2014.10.029.

Maeda, T., and Taniguchi, Y., 2013. Descriptions of GCOM-W1 AMSR2 level 1R and level 2 algorithms. Japan Aerospace Exploration Agency Earth Observation Research Center, Ibaraki, Japan, suzaku.eorc.jaxa.jp/GCOM_W/data/doc/NDX-120015A.pdf.

Malbéteau, Y., Merlin, O., Molero, B., Rüdiger, C., and Bacon, S., 2016. DisPATCh as a tool to evaluate coarse-scale remotely sensed soil moisture using localized in situ measurements: Application to SMOS and AMSR-E data in South eastern Australia. *International. Journal Applied Earth Observation and Geoinformation* 45, pp.221 - 234.

Mausse, B.J., Munyemana, F., Uamusse, A., and Manjate, A., 2021. Determination of total phenols and evaluation of the antioxidant activity of pulps and fruit derivatives of Vangueria infausta and Strychnos spinosa. *Journal of Medicinal Plant Studies* 9(3), pp.06 - 13.

Mavhura, E., 2019. Systems analysis of vulnerability to hydrometeorological threats: an exploratory study of vulnerability drivers in northern Zimbabwe. *International Journal of Disaster Risk Science* 10, pp.204 - 219, doi:10.1007/s13753-019-0217-x.

McFeeters, S.K., 1996. The use of the normalized difference water index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing* 17(7), pp.1425 - 1432, doi:10.1080/01431169608948714.

Michener, W.K., and Houhoulis, P.F., 1997. Detection of vegetation changes associated with extensive flooding in a forested ecosystem. *Photogrammetric Engineering and Remote Sensing* 63(12), pp.1363 - 1374.

Mohammadi, A., Costelloe, J.F., and Ryu, D., 2017. Application of time series of remotely sensed normalized difference water, vegetation and moisture indices in characterizing flood dynamics of large-scale arid zone floodplains. *Remote Sensing of Environment* 190, pp.70 - 82, doi:10.1016/j.rse.2016.12.003.

Moran, M.S., Clarke, T.R., Inoue, Y., and Vidal, A., 1994. Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation Index. *Remote Sensing of Environment* 49, pp.246 - 263, doi:10.1016/0034-4257(94)90020-5.

Morrison, R.B., and White, P. G., 1976. Monitoring of flood inundation. *US Geological Survey* Prof Paper 929 ERTS-1, A new window on our planet, pp.196 - 208, doi:/10.3133/pp929.

Myneni, R.B, Hall, F.G., Sellers, P.J., and Marshak, A.L., 1995. The interpretation of vegetation spectral indexes. *IEEE Transactions on Geoscience and Remote Sensing* 33(2), pp.481 - 486. DOI:10.1109/36.377948.

Nemani, R., Pierce, L., Running, S., and Goward, S., 1993. Developing satellite-derived estimates of surface moisture status. *Journal of Applied Meteorology* 32, pp.548 - 557, DOI: https://doi.org/10.1175/1520-0450(1993)032<0548:DSDEOS>2.0.CO;2.

Nied, M., Hundecha, Y., and Merz, B., 2013. Flood-initiating catchment conditions: a spatio-temporal analysis of large-scale soil moisture patterns in the Elbe River basin. *Hydrology and Earth System Sciences* 17, pp.1401 - 1414, doi:10.5194/hess-17-1401-2013.

Oh, Y., Sarabandi, K., and Ulaby, F.T., 1992. An empirical model and an inversion technique for radar scattering from bare soil surfaces. *IEEE Transactions on Geoscience Remote Sensing* 30, pp.370 - 381, DOI:10.1109/36.134086.

Parsons, M., and Thoms, M.C., 2013. Patterns of vegetation greenness during flood, rain and dry resource states in a large, unconfined floodplain landscape. *Journal Arid Environments* 88, pp.24 - 38, doi:10.1016/j.jaridenv.2012.07.023.

Pasqualotto, N., Delegido, J., Wittenberghe, van S., Verrelst, J., Rivera, J.P., and Moreno, J., 2018. Retrieval of canopy water content of different crop types with two new hyperspectral indices: water absorption area index and depth water index. *International. Journal of Applied Earth Observation and Geoinformation* 67, pp.69 - 78, doi:10.1016/j.jag.2018.01.002.

Petropoulos, G.P., Ireland, G., and Barrett, B., 2015. Surface soil moisture retrievals from remote sensing: Current status, products and future trends. *Physics and Chemistry of the Earth* (83–84), pp.36 - 56, doi:10.1016/j.pce.2015.02.009.

Pierdicca, N., Pulvirenti, L., Chini, M., Guerriero, L., and Candela, L., 2013. Observing floods from space: experience gained from COSMO-SKTMed observations. *Arcta Astranomica* 84, pp.122 - 133, doi:10.1016/j.arctaastro.2012.10.034.

Potic, I., Bugarski, M., and Matic-Varinica, J., 2017. Soil moisture determination using remote sensing data for the property protection and increase of agricultural production. Presentation 2017 World Bank Conference on Land and Poverty. The World Bank - Washington DC, March pp.20 - 24, DOI:10.13140/RG.2.2.30426.59845.

Powell, S.J., Jakeman, A., and Croke, B., 2014. Can NDVI response indicate the effective flood extent in macrophyte dominated floodplain wetlands? *Ecological Indicators* 45, pp.486 - 493, doi:10.1016/j.ecolind.2014.05.009.

Pradhan, B., Tehrany, M. S., and Jebur, M. N., 2016. A new semiautomatic detection mapping of flood extent from TerrarSAR-X satellite image using rule-based classification and Taguchi optimization techniques. *IEEE Transactions on Geoscience and Remote Sensing* 54(7), pp.4331- 4342.

Rad, A.M., Kreitler, J., and Sadegh, M., 2021. Augmented normalized difference water index for improved surface water monitoring. *Environmental Modelling and Soft*ware 105030. doi:10.1016/j.envsoft.2021.105030.

Rahman, M.S., Di, L., Yu, E., Lin, L., Zhang. C., and Tang, J., 2019. Rapid flood progress monitoring in cropland with NASA SMAP. *Remote Sensing* 11(2), 191, doi:10.3390/rs11020191.

Rahman, M.R., and Thakur, P.K., 2018. Detecting, mapping and analyzing of flood water propagation using synthetic aperture radar (SAR) satellite data and GIS: a case study from the Kendrapara District of Orissa State of India. *Egyptian Journal of Remote Sensing and Space Science* 21, pp.S37 - S41, doi:10.1016/j.ejrs.2017.10.002.

Rasid, H., and Pramanik, M.A.H., 1993. Areal extent of the 1988 flood in Bangladesh: how much did the satellite imagery show? *Natural Hazards 8*, pp.189-200. doi:10.1007/BF00605441.

Rood, S.B., Kaluthota, S., Philipsen, L.J., Slaney, J., Jones, E., Chasmer, L., and Hopkinson, C., 2019. Camo-maps: an efficient method to assess and project riparian vegetation colonization after a major river flood. *Ecological Engineering* 141, 105610, doi:10.1016/j.ecoleng.2019.105610.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., and Harlan, J. C., 1974. Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation. NASA/GSFC, Greenbelt, MD, USA, pp.1 - 137.

Sadeghi, M., Jones, S.B., and Philpot, W.D., 2015. A linear physically-based model for remote sensing of soil moisture using short wave infrared bands. *Remote Sensing of Environment* 164, pp.66 - 76, DOI:10.1016/j.rse.2015.04.007.

Sandholt, I., Rasmussen, K., and Andersen, J.A., 2002. Simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sensing of Environment* 79, pp.213 - 224, doi:10.1016/S0034-4257(01)00274-7.

Schmugge, T.J., 1983. Remote sensing of soil moisture: recent advances. *IEEE Transactions on Geoscience and Remote Sensing GE*-21(3), pp.336 - 344, DOI:10.1109/TGRS.1983.350563.

Schumann, G., Hostache, R., Puech, C., Hoffmann, L., Matgen, P., Pappenberger, F., and Pfister, L., 2007. High resolution 3D flood information from radar imagery for flood hazard management. *IEEE Transactions on Geoscience and Remote Sensing* 45(6), pp.1715 - 1725, DOI:10.1109/TGRS.2006.888103.

Schumann, G.J.-P., and Moller, D.K., 2015. Microwave remote sensing of flood inundation. *Physics and Chemistry of the Earth* 83 - 84, pp.84 - 95, doi:10.1016/j.pce.2015.05.002.

Sera B., Cudlin, P., and Budejovice, C., 2001. Flood impact on vegetation communities. *International Journal of Ecological Problems of the Biosphere* 20(1), pp.38 - 46.

Shen, X., Anagnostou, E.N., Allen, G.H., Brakenridge, G.R., and Kettner, A.J., 2019. Near-real-time non-obstructed flood inundation mapping using synthetic aperture radar. *Remote Sensing of Environment* 221, pp.302 - 315, doi:10.1016/j.rse.2018.11.008.

Shoyama, K., Cui, Q., Hanashima, M., Sano, H., and Usuda, Y., 2021. Emergency flood detection using multiple information sources: Integrated analysis of natural hazard monitoring and social media data. *Science of the Total Environment 767*, 144371. doi:10.1016/j.scitotenv.2020.144371.

Sims, N.C., and Colloff, M.J., 2012. Remote sensing of vegetation responses to flooding of a semi-arid floodplain: implications for monitoring ecological effects of environmental flows. *Ecological Indicators* 18, pp.387 - 391, doi:10.1016/j.ecolind.2011.12.007.

Steele-Dunne, S.C., Hahn, S., Wagner, W., and Vreugdenhil, M., 2019. Investigating vegetation water dynamics and drought using MetOp ASCAT over the North American Grasslands. *Remote Sensing of Environment* 224, pp.219 - 235, doi:10.1016/j.rse.2019.01.004.

Suarez, L., Restrepo-Coupe, N., Hueni, A., and Chisholm, L.A., 2015. Vegetation spectroscopy, in Held, A., Phinn, S., Soto-Berelov, M., Jones, S., (eds.), *AusCover good practice guidelines: A technical handbook supporting calibration and validation activities of remotely sensed data product 221-233. Version 1.1*. TERN AusCover, ISBN 978-0-646- 94137-0.

Sun, H., and Cui, Y., 2021. Evaluating downscaling of microwave satellite soil moisture based on machine learning method. *Remote Sensing* 13, 133, doi:10.3390/rs13010133.

Takeuchi, W., and Yasuoka, Y., 2004. Development of normalized vegetation, soil and water indices derived from satellite remote sensing data. *Journal of Japan Society of Photogrammetry and Remote Sensing* 43(6), pp.7 - 19, doi:10.4287/jsprs.43.6_7.

Tangdamrongsub, N., Forgotson, C., Gangodagamage, C., and Forgotson, J., 2021. The analysis of using satellite soil moisture observations for flood detection, evaluating over the Thailand's Great Flood of 2011. *Natural Hazards* pp.1 - 26, doi:10.1007/s11069-021-04804-8.

Tin, D., 2015. Analysis of the application of MODIS shortwave infrared water stress index in monitoring agricultural drought. *Journal of remote sensing*. https://www.researchgate.net/publication/220822142_Application_of_Two_Shortwave_Infrared_Water_Stress_Indices_to_Drought_Monitoring_over_Northwestern_China.

Tockner, K., and Stanford, J.A., 2002. Review of riverine flood plains: present state and future trends, *Biological Sciences Faculty Publications* 166, https://scholarworks.umt.edu/biosci_pubs/166.

Tsyganskaya, V., Martinis, S., Marzahn, P., and Ludwig, R., 2018. SAR-based detection of flooded vegetation - a review of characteristics and approaches. *International Journal of Remote Sensing* 39(8), pp.2255 - 2293, DOI:10.1080/01431161.2017.1420938.

Vereecken, H., Huisman, J.A.Y. Pachepsky, Y. Montzka, C. van der Kruk, J. Bogena, H., …, and Vanderborght, J., 2014 On the spatio-temporal dynamics of soil moisture at the field scale. *Journal of Hydrology* 516, pp.76 - 96, doi:10.1016/j.jhydrol.2013.11.061.

Verhoef, W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sensing of Environment* 16, pp.125 - 141, doi:10.1016/0034-4257(84)90057-9.

Verrelst, J., Rivera, J.P., and Moreno, J., 2017. Emulation of radiative transfer models, Presentation: Annual OPTIMIZE Workshop and MC Meeting 22 February 2017, DOI:10.13140/RG.2.2.34109.26083.

Verrelst, J., Sabater, N., Rivera, J.P., Muñoz-Marí, J. Vicent, J. Camps-Valls, G., and Moreno, J., 2016. Emulation of leaf, canopy and atmosphere radiative transfer models for fast global sensitivity analysis. *Remote Sensing* 8, 673, doi:10.3390/rs8080673.

Viovy, N., Arino, O., and Belward, A. S., 1992. The best index slope extraction (BISE): A method for reducing noise in NDVI time-series. *International Journal of Remote Sensing* 13(8), pp.1585 - 1590, doi:10.1080/01431169208904212.

Vreugdenhil, M., Wagner, W., Bauer-Marschallinger, B., Pfeil, I., Teubner, I., Rüdiger, C., and Strauss, P., 2018. Sensitivity of Sentinel-1 backscatter to vegetation dynamics: an Austrian case study. *Remote Sensing* 10, 1396, doi:10.3390/rs10091396.

Wallace, T.A., Gehrig, S. L, Doody, T.M., Davies, M., Walsh, R., Fulton, C. , Cullen, R., and Nolan, M., 2021. A multiple-lines-of-evidence approach for prioritising environmental watering of wetland and floodplain trees. *Ecohydrology*. doi:10.1002/eco.2272.

Wang, L., Jia, M., Yin, D., and Tian, J., 2019. A review of remote sensing for mangrove forests: 1956–2018. *Remote Sensing of Environment* 231, 111223. doi:10.1016/j.rse.2019.111223.

Wang, L., and Qu, J. J., 2007. NMDI- A normalised multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing, *Geophysical Research Letters* 34(20), doi:10.1029/2007GL031021.

Wasko, C., and Nathan, R., 2019. Influence of changes in rainfall and soil moisture on trends in flooding. *Journal of Hydrology* 575, pp.432 - 441, doi:10.1016/j.jhydrol.2019.05.054.

Wessels, J., 2014. SAR polarimetry for flooded vegetation, Video Presentation: LIDAR/SAR Wetland and water monitoring workshop, 27 June 2014, University of Lethbridge.

Wijewardana, C., Alsajri, F.A., Irby, J.T., Krutz, L.J., Golden, B., Henry, W.B., Gao, W., and Reddy, K.R., 2019. Physiological assessment of water deficit in soybean using midday leaf water potential and spectral features. *Journal of Plant Interaction.* 14(1), pp.533 - 543, doi:10.1080/17429145.2019.1662499.

Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing* 27 (14), pp.3025–3033. doi:10.1080/01431160600589179.

Yan, E., Wang, G., Lin, H., Xia, C., and Sun, H., 2015. Phenology-based classification of vegetation cover types in northeast China using MODIS NDVI and EVI time series. *International. Journal of Remote Sensing* 36(2), pp.489-512, DOI:10.1080/01431161.2014.999167.

Yang, Y-K., Nam, H-O., Kim, K-O., and Cho, S-I., 1993. Development of supercomputer image processing software with x-window user-interface for the processing of the remotely sensed data, pp.235 - 239.

Yao, Y., Qin, Q. Zhao, S., and Yuan, W., 2011. Retrieval of soil moisture based on MODIS shortwave infrared spectral feature. *Journal Infrared Millimeter Waves* 30 (1), pp.61 - 67, doi:10.3724/SP.J.2008.00009.

Yue, J., Tian, J., Tian, Q., Xu, K., and Xu, N., 2019. Development of soil moisture indices from differences in water absorption between shortwave-infrared bands. *ISPRS Journal of Photogrammetry and Remote Sensing* 154, 216 - 230, doi:10.1016/j.isprsjprs.2019.06.012.

Zhang, X., Chan, N.W., Pan, B., Ge, X., and Yang, H., 2021. Mapping flood by the object-based method using backscattering coefficient and interference coherence of Sentinel-1 time series. *Science of the Total Environment* 794. 148388, doi:10.1016/j.scitotenv.2021.148388.

Zhang, X., Friedl, M. A., and Schaaf, C. B., 2006. Global vegetation phenology from Moderate Resolution Imaging Spectroradiometer (MODIS): evaluation of global patterns and comparison with *in situ* measurements. *Journal of Geographical Research* 111, G04017, doi:10.1029/2006JG000217.

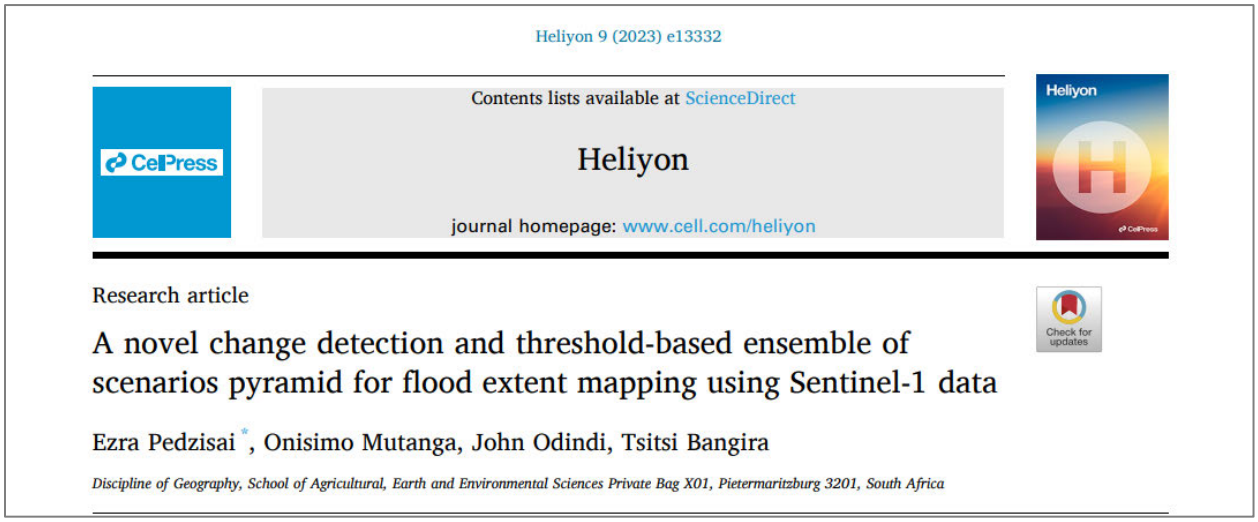
Zhang, N., Hong, Y., Qin Q., and Liu, L., 2013. VSDI: a visible and shortwave infrared drought index for monitoring soil and vegetation moisture based on optical remote sensing.

International. Journal of Remote Sensing 34(13), pp.4585-4609, DOI:10.1080/01431161.2013.779046.

Zheng, X., Feng, Z., Li, L., Li, B., Jiang, T., Li, X., …, and Chen, S., 2021. Simultaneously estimating surface soil moisture and roughness of bare soils by combining optical and radar data. *International Journal of Applied Earth Observation and Geoinformation* 100, 102345, doi:10.1016/j.jag.2021.102345.

Zhou, Y., Wu, W., Nathan, R., and Wang, Q.J., 2021. A rapid flood inundation modelling framework using deep learning with spatial reduction and reconstruction. *Environmental Modelling and Software* 143, 105112, doi:10.1016/j.envsoft.2021.105112.

CHAPTER 3: CHANGE DETECTION AND THRESHOLD-BASED ENSEMBLE OF SCENARIOS PYRAMID FOR FLOOD EXTENT MAPPING USING SYNTHETIC APERTURE RADAR (SAR) DATA



Heliyon 9 (2023) e13332

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

A novel change detection and threshold-based ensemble of scenarios pyramid for flood extent mapping using Sentinel-1 data

Ezra Pedzisai *, Onisimo Mutanga, John Odindi, Tsitsi Bangira

Discipline of Geography, School of Agricultural, Earth and Environmental Sciences Private Bag X01, Pietermaritzburg 3201, South Africa

The chapter is based on: **Pedzisai, E.**, Mutanga, O., Odindi, J., and Bangira, T., 2023. A novel change detection and threshold-based Ensemble of Scenarios Pyramid for flood extent mapping using Sentinel-1 data. *Heliyon* 9; e13332

Abstract

Flood disasters destroy infrastructure, disrupt ecosystem processes, adversely affect social and economic activities and cause human fatalities. As such, flood extent mapping (FEM) is critical to mitigate these impacts. Specifically, FEM is essential to mitigate adverse impacts through early warning, efficient response during evacuation, search, rescue and recovery. Furthermore, accurate FEM is crucial for policy formulation, planning and management, rehabilitation, and promoting community resilience for sustainable occupation and use of floodplains. Recently, remote sensing has become valuable in flood studies. However, whereas free passive remote sensing images have been common input into predictive models, damage assessment and FEM, their utility is constrained by clouds during flooding events. Conversely, microwave-based data is unconstrained by clouds, hence is important for FEM. Hence, to increase the reliability and accuracy of FEM using Sentinel-1 radar data, the study proposed a three-step process that builds an ensemble of scenarios pyramid (ESP) based on change detection and thresholding technique. The study deployed the ESP technique and tested it on a use-case based on two, five and 10 images. The use-case calculated three co-polarized Vertical-Vertical (VV) and three cross-polarized Vertical-Horizontal (VH) normalized difference flood index scenarios to form six binary-classified FEMs at the base. We ensembled the base scenarios to three dual-polarized centre FEMs. Similarly, the study ensembled centre scenarios to a final pinnacle flood extent map. The base, centre, and pinnacle scenarios were validated using six binary classification performance metrics. The results show that the ESP increased the base-to-pinnacle minimum classification performance metrics with overall accuracy, Cohen's Kappa, intersect over union, recall, F1-score, and Matthews Correlation coefficient of 93.204%, 0.864, 0.865, 0.870, 0.927, and 0.871, respectively. The study also established that the VV channels were superior in FEM than VH at the ESP base. Overall, this study demonstrates the efficacy of the ESP for operational flood disaster management.

Keywords

Change detection and thresholding; Ensemble of Scenarios Pyramid; flood extent mapping; SAR; Sentinel-1.

3.1 Introduction

Flood disasters, influenced by a complex interaction between physical and anthropogenic factors, are detrimental to among others local livelihoods, local and national economies, and the

environment (Tabari, 2021). Between 2000 and 2019 for instance, Centre for Research on the Epidemiology of Disasters (CRED, 2020b) notes that globally, 7348 disasters that affected 4.03 billion people, caused 1.23 million deaths, and led to losses worth US$2.97 billion. In Africa, during the same period, flood-related disasters accounted for 64% of the 1143 recorded natural disasters that affected 16% of 337 million people and caused 32% of 46078 deaths (CRED, 2020a). Sub-Saharan Africa is already experiencing more frequent and intense climate extremes especially tropical cyclones (Chatiza, 2019). In Southern Africa for instance, the 2019 Cyclone Idai caused 602 and 340 deaths and losses estimated at US$3 billion and $274 million in Mozambique and Zimbabwe, respectively (Nhundu et al., 2021; Chatiza, 2019). Generally, vulnerability to flooding and associated impacts are influenced by floodplain pull factors that include the availability of highly productive fertile depositional soils with post-flood residual soil moisture, social and recreational value, water security, fishing, and ecological integrity (Carozza and Boudreault, 2021; Mazzoleni et al., 2021; Ndehedehe et al., 2021; Perosa et al., 2021). Due to the changing climate, the Intergovernmental Panel on Climate Change (IPCC, 2013) predicts increased frequency and severity of floods and associated impacts. Such increases will increase the number of new flooding hot spots, with communities in developing countries particularly vulnerable due to their low coping and resilience capacities. Hence, reliable determination of flood extents is valuable for optimal decision-making and minimization of flood-related losses (Vanama et al., 2020). Specifically, there is an urgent need for accurate flood mapping and prediction techniques to abate, manage and mitigate flood-related losses.

Field surveys (Lamichhane and Sharma, 2017) and modelling (Zhang and Yusop, 2021) approaches have commonly been used for flood mapping. Typically, these approaches rely on field data and the often sparsely distributed weather stations, hence subjective, tedious, costly and prone to errors, especially in the developing countries with limited resources (Ogashawara et al., 2013; Bangira et al., 2021). On the other hand, advances in remote sensing technology with freely available satellite images have opened up new opportunities for its adoption for flood mapping and monitoring over large spatial extents (Domeneghetti et al., 2019; Bangira et al., 2021). Floods can be monitored using either optical (Ogashawara et al., 2013) or synthetic aperture radar (SAR) (Zhang et al., 2020) remotely sensed datasets. Optical imagery is easily interpretable and the extraction of open water is relatively straightforward (Huang et al., 2018). However, flood affected regions are often covered with clouds during inundation, making it difficult to map peak flooding

using commonly used optical imagery such as Sentinel-2, MODIS and Landsat. Conversely, SAR, with its characteristic cloud penetrating imaging capabilities, can be used to map inundated areas regardless of meteorological conditions (Vanama, Rao and Bhatt 2021). The freely available Sentinel-1 SAR sensor with a 10 m spatial resolution and a 400 km swath width has short revisit times (i.e. approximately 6 days) valuable for flood monitoring (Cao et al., 2017; Liang and Liu, 2020; Vanama et al., 2021). The sensor's flood mapping is premised on the principle that smooth open water, as opposed to non-water surfaces are specular reflectors, hence characterized by low SAR backscatter values (Pelich et al., 2017; Manjusree et al., 2012). SAR-based flood mapping techniques include manual or automatic thresholding (Manjusree et al., 2012; Martinis et al., 2015), clustering algorithms (Tsyganskaya et al., 2018), rule-based classification and decision trees (Long et al., 2014). In addition, more advanced approaches based on machine learning (Hao et al., 2021), fuzzy logic (Pierdicca et al., 2010; Pulvirenti, Pierdicca et al., 2011; Cao et al., 2017) and Markov Random Field modelling (Martinis et al. 2010) have been used to segment SAR imagery for flood mapping. Generally, the accuracy of these approaches depends on the polarization modes, phase information, observation angle, and pixel size of the data used (Pulvirenti, Chini et al., 2011; Amitrano et al., 2018; Clement et al., 2018).

Due to limitations of coarse temporal resolution of many open-source data available, most flood mapping approaches are based on a single image commonly acquired during inundation (Gan et al., 2012; Grimaldi et al., 2020). Single images can be analyzed with low computing power, hence valuable for the often-necessary quick turnarounds (Martinis et al., 2009; Twele et al., 2016). Generally, automation is commonly preferred in single image mapping. For example, Grimaldi et al. (2020) proposed an automatic algorithm using one SAR image to map flood and attained 80% accuracy, while Twele et al. (2016) used an automated processing chain to rapidly map flood, with a 95% accuracy. However, single images only show instantaneous flooding at a satellite overpass (Giustarini et al., 2015), hence may not be ideal where satellite overpass is asynchronous to flooding or short-lived flood occurrences. Alternatively, SAR based classification techniques using time-series images that allow for the detection of inherent changes in backscatter values over extended periods could be valuable in flood mapping (Tsyganskaya et al., 2018; Huang and Jin, 2020). Such imagery could be used to derive backscatter measurements for year-round surface mapping under different conditions (i.e. change detection). During flooding, abrupt changes in backscattered signal on the wet pixel are observed (Pulvirenti, Chini et al., 2011), which in turn

improves delineation of flooded areas. This approach is particularly useful in determining flooded areas in a heterogeneous landscape as it captures seasonal or annual fluctuations of the backscatter (Long, Fatoyinbo and Policelli, 2014; Tsyganskaya et al., 2018). Furthermore, numerous images of the same area allow for multiple observations, which in turn improves flood-mapping accuracy. Meanwhile, including time series images of the same flood event ensures that the time-dependent changes between the wet and dry periods are not missed by arbitrary selection of any two minimum images which may occur within the same season.

The ambiguity resulting from low backscatter values over dry areas in the reference image can be minimized by combining change detection with other approaches (Giustarini et al., 2013). For example, Long et al. (2014) combined change detection with thresholding of SAR data to map flood extent in the Caprivi floodplain, Namibia. Recently, studies have shown that mapping accuracies can be improved by incorporating backscatter indices in the change detection approach. For instance, Ulloa et al. (2019) showed that the Normalized Difference Scattering Index (NDSI) based on both VV- and VH-polarized data was useful for quantifying the total flooded area due to Cyclone Idai in Mozambique. Similarly, Cian et al. (2018) obtained high accuracies using the Normalized Difference Flood Index (NDFI).

Studies have shown that the use of model or algorithm 'ensembles' improves the efficacy in FEM (Costache and Bui 2019; Panahi et al., 2021; Saha et al., 2021). Whereas Thangavel and Sreevalsan-Nair (2021) proposed innovation based on ensemble of methods that used consensus voting, the novelty of this study lies in the pyramidal ensemble starting with scenarios with different numbers of temporal image combinations to bring an improved FEM. Therefore, to benefit from the merits of complementarity of these scenarios, this study involves a novel approach of an ensemble of change detection and thresholding (ECDT) framework on the various NDFI outputs. Specifically, we applied the ECDT framework named Ensemble of Scenarios Pyramid (ESP) on Sentinel-1 data to determine a flooded area's spatiotemporal variability. This was achieved by describing the scenario modelling procedure, determining ESP framework and assessing the performance evaluation metrics.

3.2 Materials and methods

3.2.1 Study Area

This study was conducted within the Mbire District, a key natural floodplain of the Zambezi River in northern Zimbabwe located from 30º 26' to 30º 30' E and 16º 13' to 16º 16' S. The area falls within a semi-arid climate with low and fluctuating annual rainfall of about 650−700 mm. It is characterized by hot and wet summers and cool and dry winters, with temperatures ranging between 18°C and 26°C. The area is drained by six major rivers within approximately 40 km^2 (Figure 3.1). The Hunyani River is the main hydrological channel associated with floods in the area. Tropical cyclone-induced heavy rains commonly result in floods and the downstream high levels at the Lake Cahora Bassa in Mozambique cause backflow floods on the floodplain. The gently sloping terrain, with low elevation ranging from 360 to 420 m above sea level is a major cause of extended floods. Depositional silt soils within the area support riparian vegetation associated with predominant deciduous trees, bushes, and grasslands. Whereas rain-fed subsistence agriculture, supported by depositional silt soils is the major economic activity (Mavhura et al., 2020), poverty in the area is prevalent and resilience is low (Mudavanhu et al., 2020). The district population is projected to reach 110,000 by the year 2022 (Zimstats, 2020), further increasing settlement on floodplains and population vulnerability.

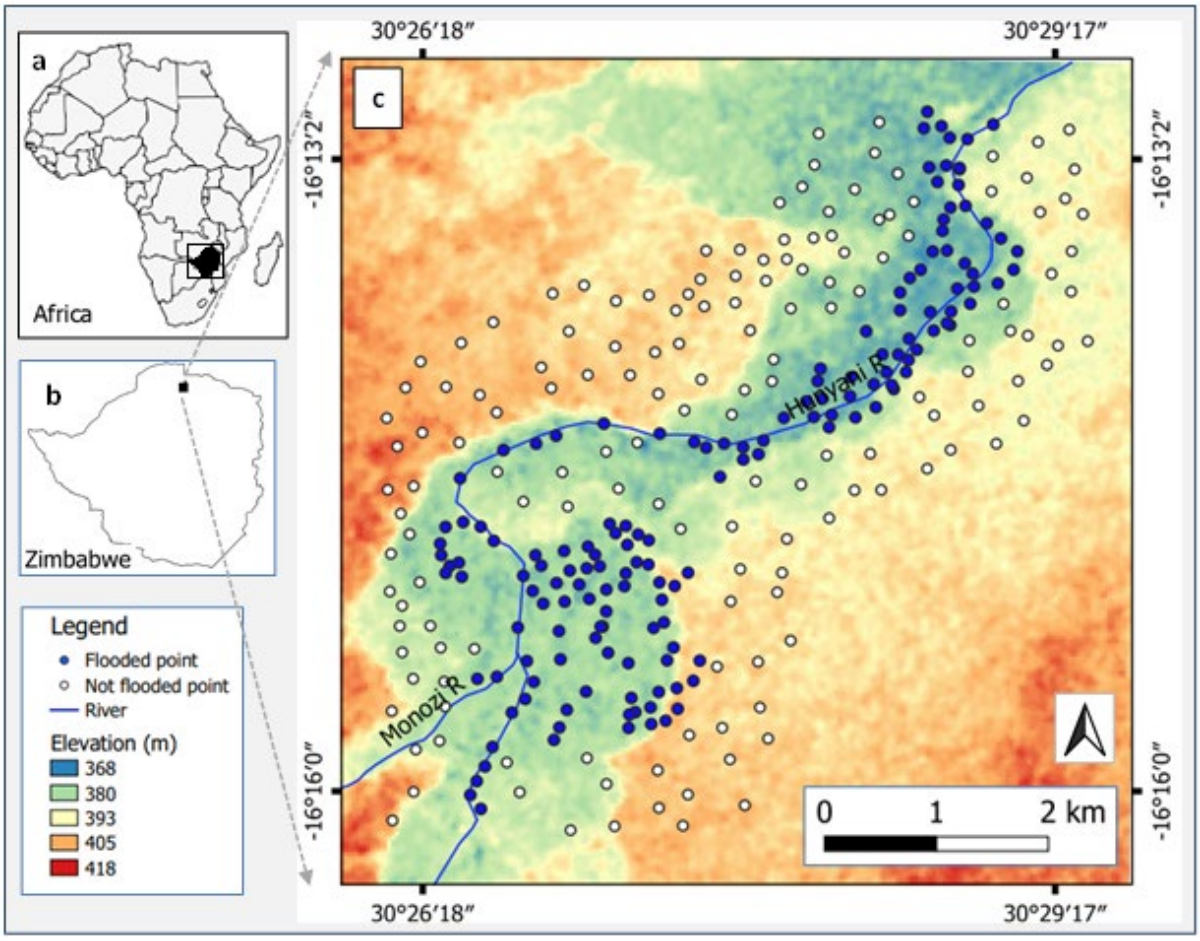


Figure 3.1: The inserts for Africa (a), Zimbabwe (b) and the Mbire study area (c) showing elevation and drainage, with flooded and not flooded field sample points used as ground truth for validation.

3.2.2 Data

3.2.2.1 In situ flood extent data collection

Data for the most recent historical flood event (2nd of February 2017) was used in this study. The post-rain season period data was deemed convenient to collect validation data as local communities, rescue and disaster management personnel and local non-governmental organization field staff had a better recollection of flood extents. Ground truth points for the flooded and non-flooded classes were collected during a guided field survey between 15th and 25th of April 2017. A hand-held global positioning systems (GPS) receiver was used to randomly collect a sum of 309 field points of which 155 were from inside and 155 outside the flooded (155) areas that were used to classify the flooded extent on satellite images.

3.2.2.2 Satellite image acquisition and processing

Sentinel-1 remotely sensed images were acquired and used to map flood extent. The mission provides data from a dual-polarization (VH+VV), C-band SAR instrument at 5.405GHz that is freely available for download at the Sentinels Scientific Data Hub website (https://scihub.copernicus.eu/). Thirty Sentinel-1B Ground Range Detected High Resolution (GRDH) imagery from January 2017 to December 2017 were used in this study. The images were preprocessed using a Sentinel-1 Toolbox (S1TBX 8.0), an inbuilt algorithm within the Sentinel Application Platform (SNAP) tool. The preprocessing steps involved orbital file application, subset, thermal noise removal, radiometric calibration, and terrain correction (Appendix 3.1).

3.2.3 Overview of the Ensemble of Scenarios Pyramid modeling

The Ensemble of Scenarios Pyramid (ESP) is a framework based on ensemble, change detection, and thresholding (ECDT). In the ESP, the Normalized Difference Flood Index (NDFI) comprising a different number of images using VV and VH polarizations is first calculated to form scenarios, which are then incorporated as input into the pyramid architecture (Figure 3.2).

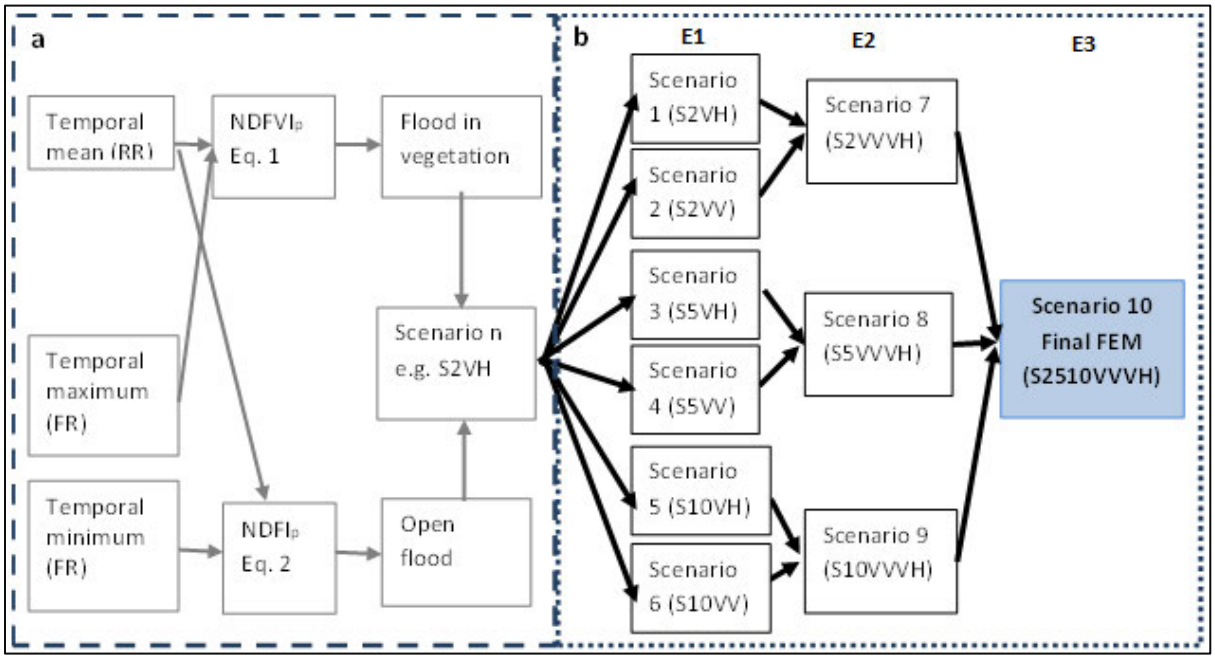


Figure 3.2: Methodological framework showing, a) calculation of six normalized difference flood index (NDFI) scenarios using VV and VH polarized 2 (bi-temporal), 5 (penta-temporal) and 10 (deca-temporal) images and, b) the Ensemble of Scenarios Pyramid (ESP) pairwise union of six VV and VH (base) into three centre scenarios (i.e. 7-9) and pinnacle (Scenario10).

3.2.3.1 Development of the normalized difference flood index scenarios

The NDFI was used to generate a single output denoted as Scenario *n* (see Figure 3.2). Six scenarios with bi-temporal (S2), penta-temporal (S5,) and deca-temporal (S10) imagery were generated using both the VV and VH polarization modes as shown in Figure 3.2. The scenarios comprised two image stacks with the first stack compiled with the non-flood or references only images (RR) and the second stack built on flood (F) plus reference images (R), hence FR. The two stacks were used to compute the NDFI and NDFVI maps and sliced using a threshold of 0.7 as described by Cian et al. (2018) (Equations 3.1 – 3.2);

$$\mathrm{NDFVI_P} = (\mathrm{FR_{max}} - \mathrm{RR_{mean}})/(\mathrm{FR_{max}} + \mathrm{RR_{mean}}) \quad (3.1)$$

$$\mathrm{NDFI_P} = (\mathrm{RR_{mean}} - \mathrm{FR_{min}})/(\mathrm{RR_{mean}} + \mathrm{FR_{min}}) \quad (3.2)$$

Where*;* $NDFI_P$ and $NDFVI_P$ are the respective polarized indices, FR_{max}, FR_{min} and RR_{mean} are the maximum, minimum and mean maps for two respective stacks of single-polarization respectively. As outlined in Cian et al. (2018), the NDFI discriminates flood from non-flood in non-vegetated pixels while the NDFVI identifies flood in vegetated pixels, hence the differences between the two thresholds.

The thresholds were used to slice the NDFVI and NDFI into flood in vegetation and open flood extents, respectively by polarization and scenarios as shown in Table 3.1. The two maps were converted into flooded and non-flooded classes using raster calculator in ArcGIS Pro version 2.9. The union of the flood in vegetation and the open flood maps was used to produce binary classified (1 = flooded; 0 = not flooded) pixels, respectively.

Table 3.1: Scenario dates, and respective temporal stacks based on vertical-horizontal (VH) and vertical-vertical (VV) polarization modes.

	Temporal images stack			
	Flood + Reference (FR)		References only (RR)	
Scenario code (name)	Date		Date	
S2 (bi-temporal)	2017/02/02	2017/08/25	2017/08/13	2017/08/25
S5 (penta-temporal)	2017/02/02	2017/08/13	2017/07/08	2017/08/13
	2017/02/14	2017/08/25	2017/07/20	2017/08/25
	2017/02/26		2017/08/01	
S10 (deca-temporal)	2017/01/21	2017/07/08	2017/06/14	2017/08/13
	2017/02/02	2017/07/20	2017/06/26	2017/08/25
	2017/02/14	2017/08/01	2017/07/08	2017/09/06
	2017/02/26	2017/08/13	2017/07/20	2017/09/18
	2017/03/10	2017/08/25	2017/08/01	2017/09/30

3.2.3.2 The Ensemble of Scenarios Pyramid algorithm

As shown in Figure 3.2, the Ensemble of Scenarios Pyramid (ESP) progressively combines temporal ECDT of NDFI flood extent maps on both VV and VH channels at three levels namely, E1, E2, and E3 to improve the accuracy of the flood extent. Firstly, the ESP base (E1) used SAR data to create six scenarios comprising of time series built on flood and post-flood images using three pairs of stacks of bi-temporal (S2), penta-temporal (S5) and deca-temporal (S10) time-series images of similar polarization flood extent maps. Although a number of stack sizes were tested, S2, S5 and S10 were representative enough to test the effect of the different sets of images within a year of flood occurrence. The logic behind merging the co-polarized (VV) with cross-polarized (VH) scenarios at E1 into those at E2 is that, some of the pixels classified as flooded in the co-polarized (VV) were not flooded in cross-polarized (VH). Furthermore, flood extent maps in S2 differed from those in S5 and S10, respectively. In disaster management, a false alarm is better than a missed disaster. Therefore, to cater for such suspiciously flooded pixels, we merged the scenarios by a union calculation which effectively improved the performance as also noted using all the six, evaluation metrics in the results section. Secondly, the ESP centre (E2) is a dual ensemble that applies a union function of the VV+VH pairs of E1 scenarios into three dual-polarized ESP centre flood extent maps that benefit from their complementarity. Thirdly, the ESP

pinnacle (E3) combines the three E2 scenarios by union function into one final flood extent map. A flood event case study is then presented.

3.2.4 The Ensemble of Scenarios Pyramid (ESP) performance validation metrics

In evaluating image classes, reliance on single validation metric can be erroneous. For instance, although the overall accuracy, F1-score and precision-recall are commonly used validation approaches, they can be misleading for imbalanced datasets (Halimu et al., 2019; Chicco and Jurman, 2020). Hence, in this study, the accuracy of the ESP model was assessed using six performance metrics namely, one-aspect-based metrics (i.e. intersect over union - IoU and recall - Rc) and total-summary-based metrics (overall accuracy - OA, Cohen's Kappa - CK and F1-score). The Matthew's correlation coefficient (MCC) was used to evaluate the models. To achieve this, respective metrics pixel values for 309 (N) ground truth points were classified based on guided fieldwork with flooded (154) or non-flooded (155) locations. The model predicted class were extracted using ArcGIS Pro 3.9 Spatial Analyst 'Extract Values to Points' Tool. Based on the comparison of the model versus the fieldwork classification, an error matrix was generated to give the four respective parts. Specifically, 1) true positive (TP) denoted that the model (NDFI) and the fieldwork (ground truth) concurred on the identified flooded pixels, 2) false positive (FP) occurred when the model discordantly classified as flooded yet the fieldwork indicated that it was not flooded, 3) false negative (FN) occurred when the model identified a pixel as not flooded when in fact the fieldwork confirmed that it was flooded and 4) true negative (TN) was defined as a pixel which both the model and fieldwork concurred to confirm that it was not flooded. Based on the aggregated error matrix, the six metrics were computed through Equation (3.3) to Equation (3.8). The binary classification accuracy assessment metrics were used to report the performance of the classification.

$$\text{Recall (Rc)} = (TP)/(TP+FN) \tag{3.3}$$

$$\text{F1-Score (F1)} = 2*[(TP/(TP+FP))* (TP/(TP+FN))]/[(TP/(TP+FP)) + (TP/(TP+FN))] \tag{3.4}$$

$$\text{Intersect over Union (IoU)} = (TP)/(TP+FP+FN) \tag{3.5}$$

$$\text{Overall Accuracy (OA)} = (TP+TN)/N \tag{3.6}$$

$$\text{Cohen's Kappa (CK)} = (P_o - P_e)/(N - P_e) \tag{3.7}$$

$$\text{Matthews Correlation Coefficient (MCC)} = [(TP * TN) – (FP * FN)]/ \sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)} \tag{3.8}$$

Where; N=309 is the grand total (TP+FP+FN+TN); TP, FP, FN, and TN are the true positive, false positive, false negative and true negative respectively, P_o is OA and P_e is the expected accuracy. The metrics range from the worst (0) to the perfect (1) while the MCC ranges from the worst (-1), random (0), to the perfect (+1). The F1-score, IoU, Rc, and CK range from 0 (weakest) to 1 (strongest) while the OA is usually expressed as a percentage. Intersect, and union maps are computed at E2 scenarios of the ESP to enable the calculation of the 3 E2 IoU maps. The map-based IoU was calculated using the intersect and the union binary maps (Equations (3.9 - 3.10);

$$IoU = (Sn_{VV} \cap Sn_{VH})/(Sn_{VV} \cup Sn_{VH}) \quad (3.9)$$

$$If\ [(Sn_{VV} \cap Sn_{VH})/ (Sn_{VV} \cup Sn_{VH}) = 1\ then\ 1\ else\ 0 \quad (3.10)$$

where: $Sn_{VV} \cap Sn_{VH}$ and $Sn_{VV} \cup Sn_{VH}$ show intersect and union of a pair of Scenarios (*Sn*) respectively with 1 denoting agreement between the two intersect and union maps, while 0 denotes disagreement between the respective ensembled scenarios. The agreement between the flooded pixels on the E1 VV versus VH FEM was tested using their intersection while their union was used to determine the total inundation as both polarizations had different FEMs, making the IoU an effective evaluation metric.

3.3 Results

3.3.1 Comparing VV with VH polarized SAR data for flood extent monitoring

Figure 3.3 shows the evolution of the flood extents derived from the VV and VH polarization modes during six weeks of observation (i.e. from 9th of January to 26th of March 2017). As shown in the figure, flooding peaked in early February. Before the flooding, the backscatter intensity on flooded and non-flooded areas for both polarization modes area was generally similar, while post-flood images show a significant difference in backscatter between flooded and non-flooded areas.

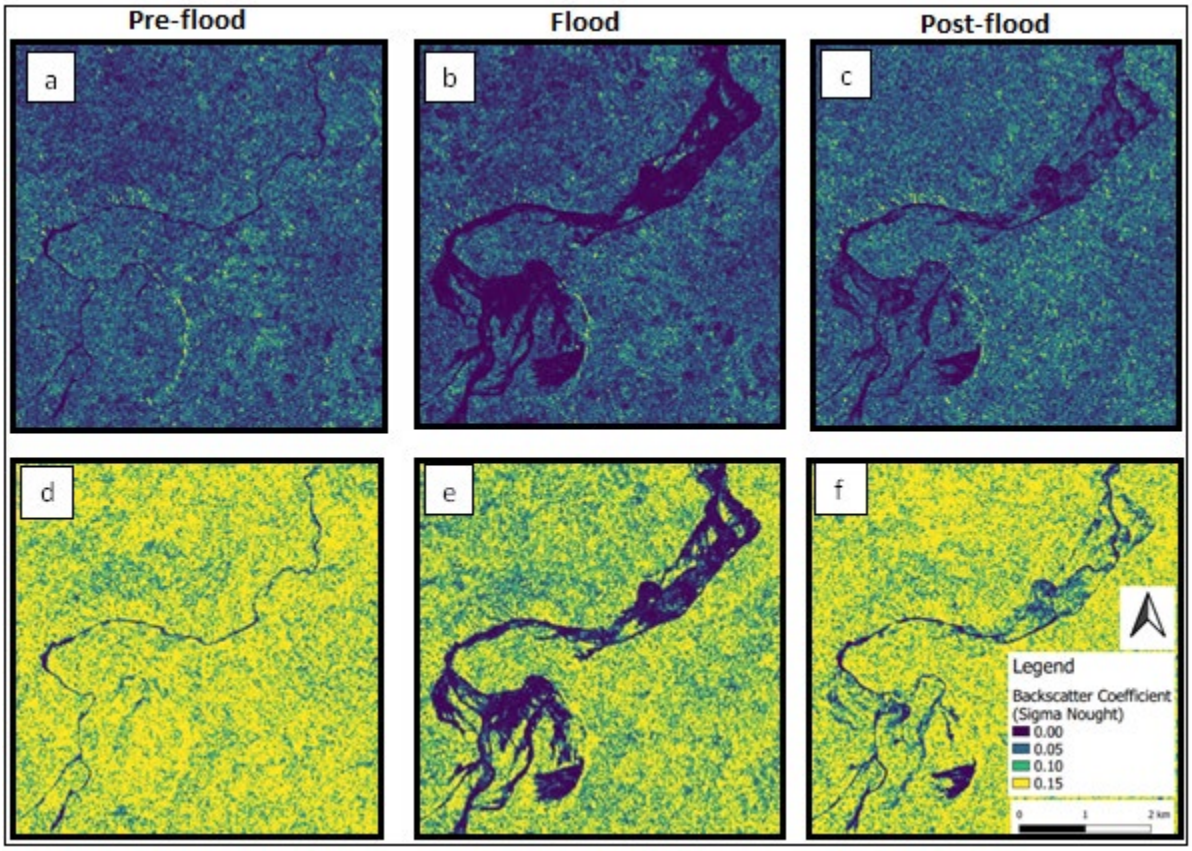


Figure 3.3: Pre-flood (2017/01/09), flood (2017/02/02) and post-flood (2017/02/06) extent maps with panels vertical-horizontal (VH) (a, b, c) and vertical-vertical (VV) (d, e, f).

3.3.2 The ESP flood extent classification

Figure 3.4 shows the flood extent maps based on the ESP approach. The six E1 flood extent maps on the ESP algorithm demonstrate the inferiority of S2VH and S10VH polarized scenarios, while the VV scenarios classified more flooded pixels than VH at the E1. Flooded areas in the E2 classifications were more consistent than the E1. The global ESP E3 (S2510) shows a final flood extent.

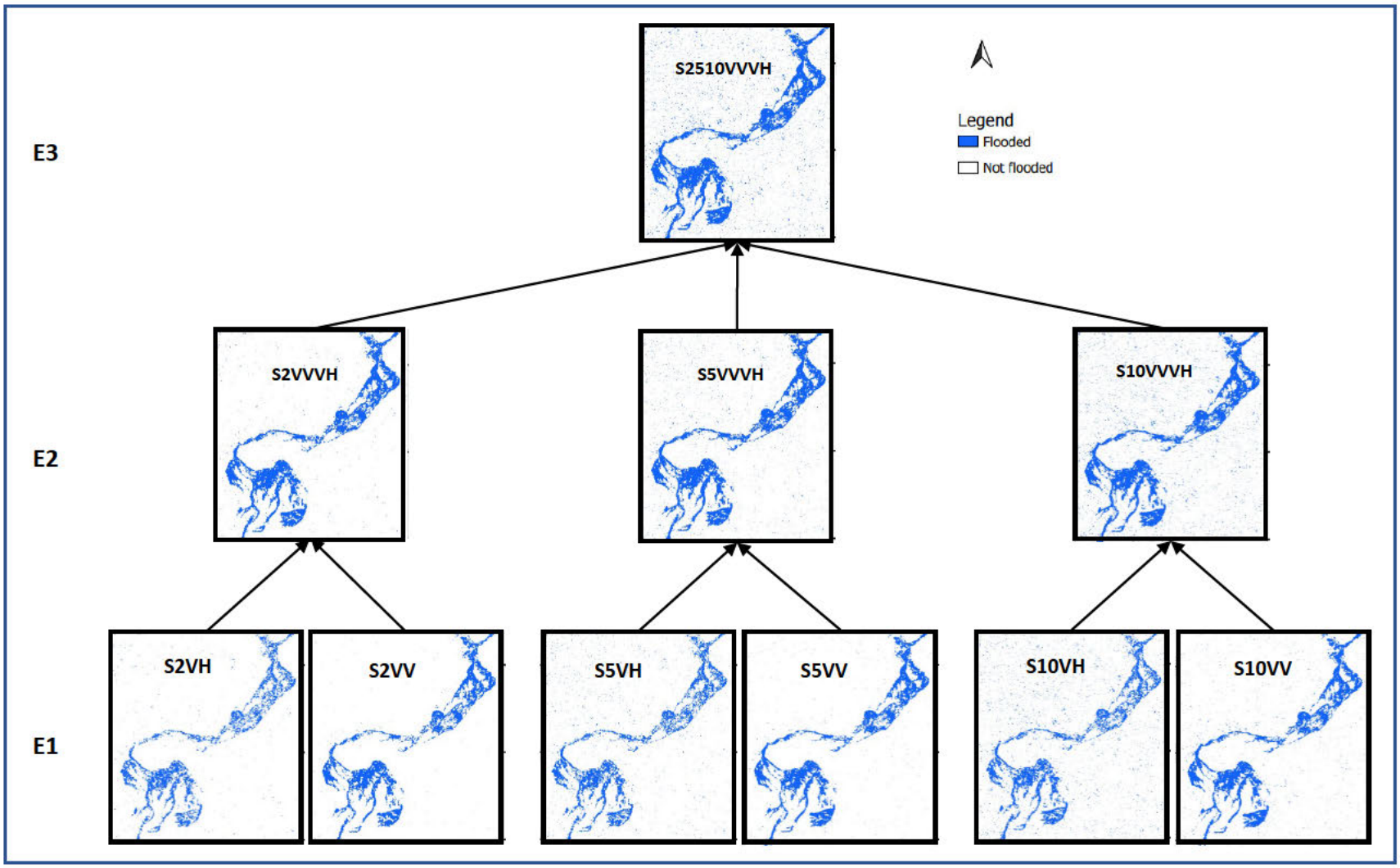


Figure 3.4: Flood extent maps at the base (E1), centre (E2) and pinnacle (E3) levels of the Ensemble Scenario Pyramid (ESP) with converging arrows showing union function calculation.

3.3.3 The ESP flood extent map validation

Table 3.2 shows the classification accuracy when mapping flood extent using ESP. The FEMs shown on Figure 3.4 validates the ESP scenarios. The E1 scenarios show variations across the metrics. The best performance at E1 was achieved using S10VV, while the least accuracy was achieved using S5VH. The VV scenarios were superior to VH for each of the three pairs. Across all metrics, the highest difference in overall accuracy (OA) between the VV versus VH polarized pairs of scenarios were 7.385, 11.65 and 14.24 percent for S2, S5 and S10, respectively. The results for other metrics were the F1 (0.137, 0.162, 0.198), IoU (0.187, 0.222, 0.267), CK (0.188 0.233, 0.285), Rc (0.188, 0.208, 0.247) and MCC (0.151, 0.203, 0.252) for S2, S5 and S10, respectively. Notably, S10 had the best performance compared to S2 across all the selected metrics.

The FEMs at E1 were useful in deriving the IoU using intersect and union maps calculated in Raster Calculator for the bi-temporal, penta-temporal and deca-temporal scenario pairs using the intersect and union, hence the three IoU maps (Figure 3.5). To evaluate the use of VV versus VH in deriving NDFI scenarios, spatial presentations of the IoU are presented for the three pairs of E1 to validate the ESP scenarios (Figure 3.5). At E2, the S2, S5 and S10 had 91.909%, 90.615% and 89.320% OA, respectively. Results from the rest of the metrics were generally high (i.e., 0.79 to 0.91), demonstrating an increased accuracy in comparison to E1. The E3 classified the flood extent with highest accuracy in all metrics. The level showed the highest accuracy (OA = 93.204%) in comparison to both mean of all metrics at E2 (OA = 90.615% and E1 (OA = 82.470%). The five other metrics were highest (i.e., CK = 0.864, IoU = 0.865, Rc = 0.870, MCC = 0.871 and F1 = 0.927). There was a noticeable increase in metrics performance from E2 to E3 (Table 3.2). Based on Table 3.2, the E1 to E3 absolute increase for the metrics were determined as F1 = 0.142, MCC = 0.184, Rc = 0.209, IoU = 0.210, CK = 0.215 and OA = 10.734%.

Table 3.2: The Ensemble of Scenarios Pyramid (ESP) summary of validation metrics and the respective mean by level.

Level	Scenario	OA	Rc	F1	IoU	CK	MCC
E1	S2VH	77.994	0.565	0.719	0.561	0.559	0.619
E1	S2VV	87.379	0.753	0.856	0.748	0.747	0.770
E1	S5VH	77.023	0.565	0.710	0.551	0.540	0.591
E1	S5VV	88.673	0.773	0.872	0.773	0.773	0.794
E1	S10VH	74.757	0.532	0.678	0.513	0.494	0.547
E1	S10VV	88.997	0.779	0.876	0.779	0.780	0.799
Mean	E1	82.470	0.661	0.785	0.654	0.649	0.687
E2	S2VVVH	91.909	0.851	0.913	0.840	0.838	0.846
E2	S5VVVH	90.615	0.838	0.899	0.816	0.812	0.820
E2	S10VVVH	89.320	0.825	0.885	0.794	0.786	0.794
Mean	E2	90.615	0.838	0.899	0.817	0.812	0.820
E3	S2510VVVH	93.204	0.870	0.927	0.865	0.864	0.871

OA = overall accuracy, Rc = recall, F1 = F1-Score, IoU = Intersect over Union, CK = Cohen's Kappa, MCC = Matthews Correlation Coefficient.

3.3.3.1 The Scenario Ensemble Agreement using IoU and MCC metrics

Figure 3.5 shows FEMs based on the IoU metric on E2. The figure shows that the least flood extent map classification accuracy evaluated using IoU was the S2 scenario while the best performance was achieved at the S5 scenario.

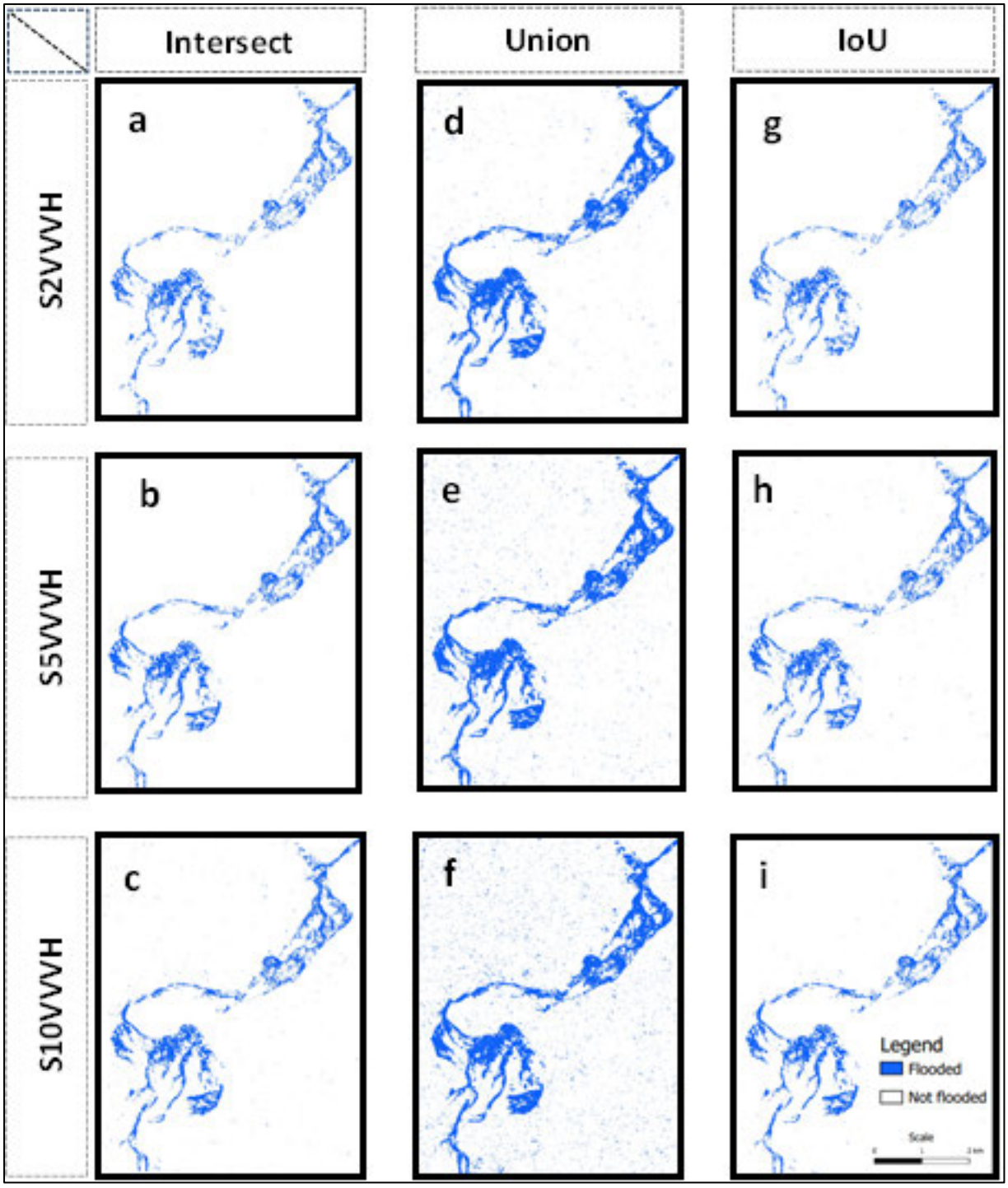


Figure 3.5: Results of the intersect (a, b, c) and union (d, e, f) of the VV and VH maps from E1 used to calculate the intersect over union (IoU) (g, h, i).

The IoU test showed high agreement of classification of the inundated pixels (Figure 3.5) from all the six predecessor scenarios at E1 in comparison to MCC as quantified in Table 3.3. As shown in Figure 3.6, there is a reasonable agreement between VV and VH polarization shown as blue pixels in detecting flooded pixels. However, there are also pixels classified as flooded

exclusively VV and VH polarization detected flood pixels in red and green colours, respectively. The flood extent has a comparatively bigger agreement than disagreement pixels.

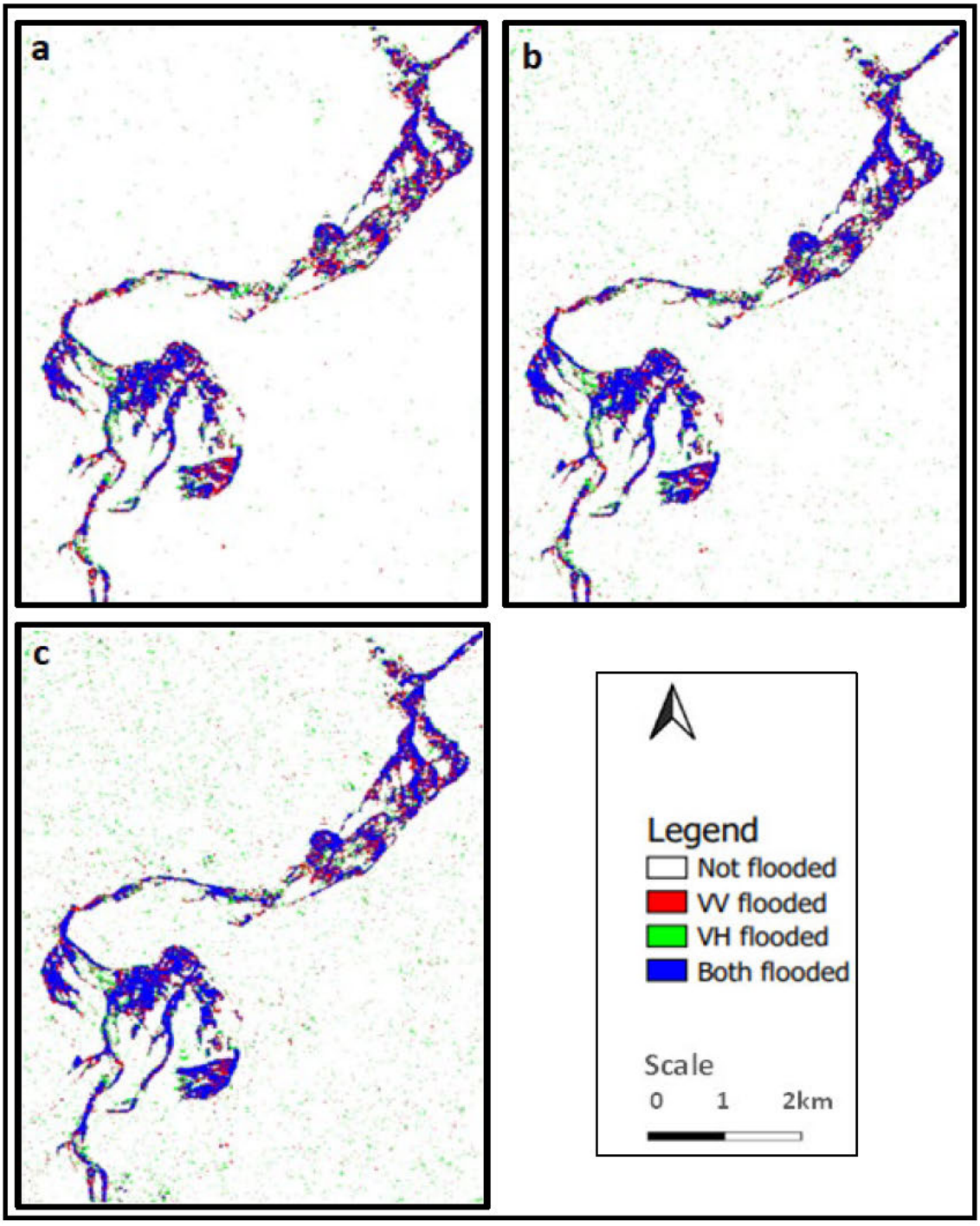


Figure 3.6: The comparison of flood detection using vertical-vertical (VV) versus vertical-horizontal (VH) polarization for bi-temporal (a), penta-temporal (b) and deca-temporal (c) scenarios respectively.

This result shows that both VV and VH polarizations can be used for FEM, however, their union would improve the flood spatial extent. Results in Table 3.3 further confirm the superiority of S10VV over S10VH at E1 for FEM, also shown on Figure 3.6 using both IoU and MCC. These results increased at E2 and ultimately at E3.

Table 3.3: The level-based intersect over union (IoU) and Matthews Correlation Coefficient (MCC) validation metrics at the base (E1), centre (E2) and pinnacle (E3).

E1	IoU	MCC	E2	IoU	MCC	E3	IoU	MCC
S2VV	0.748	0.770						
S2VH	0.561	0.619	S2VVVH	0.840	0.846			
S5VV	0.773	0.794						
S5VH	0.551	0.591	S5VVVH	0.816	0.820	S2510	0.865	0.871
S10VV	0.779	0.799						
S10VH	0.513	0.547	S10VVVH	0.794	0.794			

E1 = ESP Base level, E2 = ESP Centre level, E3 = ESP Pinnacle level, IoU = Intersect over Union, MCC = Matthews Correlation Coefficient.

The metrics-based classification accuracy of the ESP is shown using a level-to-level assessment based on the minimum, mean and maximum (Figure 3.7). Level-to-level change was calculated as the difference between E1 and E2, E2 and E3, and the E1 and E3. The E1 to E3 was regarded as the global change from the base to pinnacle of the ESP to evaluate the overall effect of the ESP. The highest increase was noted at minimum classification from E1 to E3, while lower increases were noted at mean and maximum. Overall, there was an improvement from E1 to E2 and ultimately to E3.

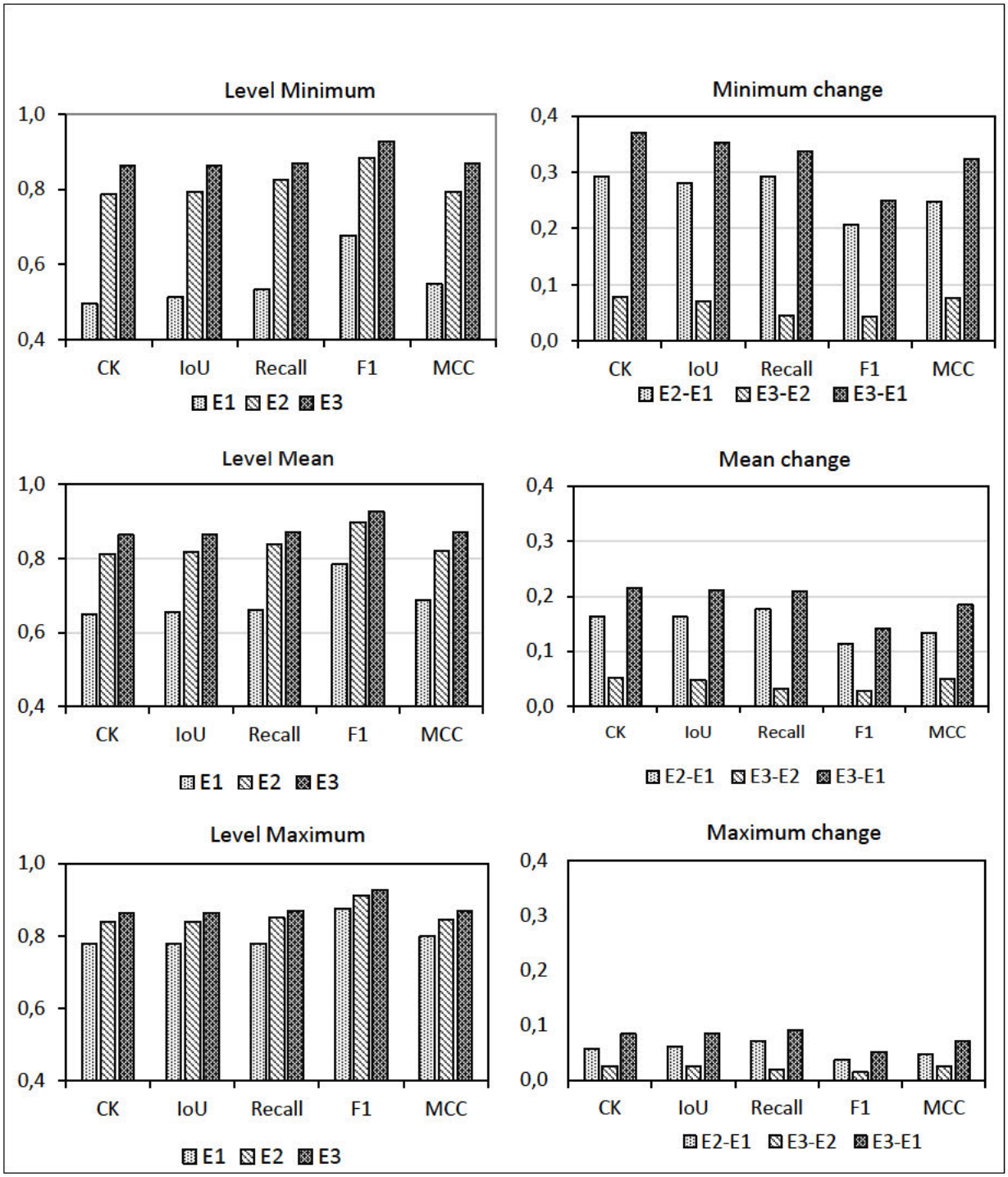


Figure 3.7: The Ensemble of Scenarios Pyramid (ESP) minimum, mean and maximum by level (left) and change (right). The Y-axis shows a standardized metric scale (left) and the magnitude of change (right).

3.4 Discussion

3.4.1 The ESP flood extent map

This study sought to explore the value of combining change detection, thresholding and NDFI scenarios ensemble in mapping flood extent in the Chitsungo area of Mbire District, Zimbabwe. Specifically, the study investigated the potential of increasing the number of images, i.e. bi-temporal, penta-temporal and deca-temporal to improve flood mapping accuracy within one calendar year. The study demonstrated the utility of the ESP algorithm in mapping flood extent, with ESP's classification accuracies increasing from E1 to E3. Overall classification results obtained from the S2 were comparable to those obtained using the S5, while S10 were superior. The performance of the S10 in the overall classification is primarily attributed to more balanced set of wet and dry images as well as longer temporal interval between the flooded and the driest reference. The latter effect provides a stronger flood signal, while identifying backscatter changes due to specular reflectance of flooded areas. Results of the ESP achieved high classification accuracies (>90%) for F1-score and OA metrics, which is comparable to some ensembles deployed in complex machine learning algorithms (Hosseini et al., 2020; Luu et al., 2021; Ngo et al., 2021).

The ESP combined predictions through an ensemble of single polarization E1 scenarios to dual-polarized channels (E2) and ultimately E3, which improved flood mapping accuracy. The mean level-based prediction accuracy increased significantly from E1 to E2 and slightly from E2 to E3. The accuracy of the results compares with those by Costache and Bui (2019) who ensembled machine learning and statistical algorithms with 14 input, 29 hidden and two output layers using eight combined machine learning models. In this study, the F1-score and OA for instance were comparable to the ensemble that used 64 machine learning algorithms on passive optical data to achieve a recall of 81.7% (Sarker et al., 2019). However, the ESP was less accurate in flood classification accuracy than some deep learning frameworks with area under curve prediction accuracy of 0.967 in Pham et al. (2021) and 99% detected inundated pixels in Zhou et al. (2021). Since this study was based on Sentinel-1 SAR data, the non-flooded smooth surfaces are known to depict low backscatter (Zhao et al., 2021) that may be ambiguous to flooded pixels, which may in turn lead to false flood alarms.

The ESP demonstrate the utility of both co- and cross-polarized channels to extract flood extent at the E1, albeit with different accuracies. This is because inundation tends to have low backscatter feedback in both polarizations. However, the ESP E1 detection results (Figure 3.4)

showed that VV predicted flood extent more accurately than VH scenarios in all validation metrics. This finding concurs with Manjusree et al. (2012) who established that the VV has high accuracy, further confirming that cross-polarized channels are noisy due to speckle (Rana and Suryanarayana 2019). However, whereas VV showed more flood extent coverage than VH, the ensemble of the two from E1 to E2 using union function predicted a larger flood extent map and improved mapping accuracy.

Results in this study are consistent with Zhao et al. (2020) who reported a reduction in backscatter in maps using the change detection method. Consistent with existing literature, the use of multiple imagery is useful in improving flood mapping accuracy. For instance, Palomba et al. (2020) delineated flood extent using Sentinel-1 SAR temporal data while Cian et al. (2018) noted that the classification accuracy increases with the number of images. However, processing many SAR images is often impeded by computing power, hence processing multiple SAR data requires high speed processers and super-computing environments (Gorelick et al., 2017). In this regard, the utility of cloud-computing resources such as GEE (Gorelick et al., 2017; Mullissa et al., 2021) for data access, preprocessing, analyses, and visualization becomes handy.

3.4.2 The ESP Evaluation

The proposed ESP approach improved FEM accuracy as evaluated using six classification metrics. The metrics used show very high binary classification accuracy with five metrics performing above 92% (Table 3.2), while the level based IoU for E2 was generally high (Table 3.3). The IoU (Figure 3.5) showed an increase in the minimum level-based classification, demonstrating that the ESP approach improved classification accuracy. The OA results compare with literature that augmented data in their deep neural network architectures (Tiwari et al., 2020; Hashemi-Beni and Gebrehiwot, 2021). The ESP mean OA compares with the 85-87% found in Bangira et al. (2021) using change detection and thresholding based on combined optical and SAR data using a machine learning algorithm. Existing methods of FEM that include operationally challenging fieldwork (Bangira et al., 2021), data intensive physical (Tu et al., 2020) and statistical-based models (Cian et al., 2018) and modern machine learning algorithms tend to have complex architectures and data challenges (Hashemi-Beni and Gebrehiwot, 2021). To attain effective flood management objectives that include early warning, preparedness, evacuations and post flood recovery, detection and mapping of floods requires simple, fast, and accurate approaches (Jiang et al., 2021; Lei et al., 2021; Zhou et al.,

2021). Overall, the level-based minimum classification accuracy increased across all metrics while the mean and maximum increased, albeit with varying proportions from E1 to E3. The VV data classified more flooded pixels than VH, concurring with literature (Berezowski et al., 2020). Hence, the ensemble enhanced the classification accuracy by combining the two polarization maps up the pyramid. For instance, the CK showed an increase in classification accuracy from high to very high level, comparable to that found using automated change detection and thresholding algorithm (Li et al., 2018). As such, the ESP algorithm classified floods with very high accuracy and showed a much higher classification fidelity than a single NDFI scenario (Long et al., 2014; Cian et al., 2018).

3.4.3 Single polarization SAR utility for monitoring flood extent

The SAR VV- and VH- polarized scenarios can be used to form the base of the ESP scenarios and qualitatively monitor floods. The flood extent maps across the flood event have shown the utility of single image bands, especially the VV polarized, to qualitatively detect the occurrence and monitor the evolution of a flood event (Figure 3.3). Both the VH- and VV-polarized temporal images were useful in monitoring spatio-temporal flood extent, with the latter showing better visible distinctions. Such flood extent scenarios can be used to visually delineate the extent of inundation, imperative in operational crisis management. However, the visual interpretation of flood extent maps produced from single polarizations is dependent on training, skill, and experience for meaningful image information mining (Mullissa et al., 2021). Furthermore, using SAR data, smooth calm water has low backscatter feedback due to specular reflectance while turbulent water, characterized by rough surface gives high feedback in both VH- and VV-polarizations (Pelich et al., 2017; Mason et al., 2021) and may lead to either missed hits or false alarms, affecting flood extent mapping accuracy. Therefore, in view of possible false alarms, single polarization-based flood extent maps need to be interpreted with some caution.

3.5 Conclusion

This study examined the potential of the proposed Ensemble of Scenarios Pyramid (ESP) approach in mapping flood extent. The ESP was proved to be a simple and reliable method for FEM applicable for operational flood disaster management. Based on our findings, we conclude that:

1) the ESP can be used to reliably classify the Sentinel-1 images to map flood extent.

2) the ESP improved mapping accuracy from using a single index as comprehensively validated using six accuracy metrics, which showed increases in the minimum and mean flood extent classification from the base to pinnacle of the framework.

Overall, findings from this study have demonstrated the potential of the Sentinel-1 as an invaluable primary data source for monitoring floods regardless of cloud cover; a challenge using optical sensors. This Sentinel-1 data characteristic, combined with performance of the ESP, offer a reliable approach for flood mapping. However, speckle remains a challenge in the quality of the classification. The selection of the respective temporal images is important yet constrained by the period of data record since Sentinel-1 constellations have a short historic record which misses earlier flood events. The method is data-driven; hence the determination of thresholds remains a challenge. Furthermore, the method needs further testing in other flood events. Therefore, further investigations are required to determine the temporal image dates differences optimal to maximize the accuracy in flood extent mapping.

3.6 References

Amitrano, D., Di Martino, G., Iodice, A., Riccio, D. and Ruello, G., 2018. Unsupervised rapid flood mapping using Sentinel-1 GRD SAR images. *IEEE Transactions on Geoscience and Remote Sensing* 56, pp.3290 - 3299.

Bangira, T., Iannini, L., Menenti, M., Van Niekerk, A. and Vekerdy, Z., 2021. Flood extent mapping in the Caprivi Floodplain using Sentinel-1 time series. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14*, pp.5667 - 5683.

Berezowski, T., Bieliński, T. and Osowicki, J., 2020. Flooding extent mapping for synthetic aperture radar time series using river gauge observations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13, pp.2626 - 2638.

Cao, W., Martinis, S. and Plank, S., 2017. Automatic SAR-based flood detection using hierarchical tile-ranking thresholding and fuzzy logic. *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017 Jul 23 (pp. 5697-5700).

Cao, H., Zhang, H., Wang, C., and Zhang, B. (2019) Operational flood detection using Sentinel-1 SAR data over large areas. *Water* 11(786), Doi:10.3390/w11040786.

Carozza, D.A. and Boudreault, M., 2021. A global flood risk modeling framework built with climate models and machine learning. *Journal of Advances in Modeling Earth Systems* 13, e2020MS002221.

Centre for Research on the Epidemiology of Disasters, 2020a. Disasters in Africa: 20 year review (2000 - 2019). UCLouvain. Online: URL: https://drmims.sadc.int/sites/default/files/document/2020-03/2019_SADC_CRED_USAID_UCLouvain_Disasters_Africa_20yearreview.pdf

Centre for Research on the Epidemiology of Disasters, 2020b. Human Cost of Disasters. An Overview of the last 20 years: 2000–2019. Online: URL: https://www.undrr.org/publication/human-cost-disasters-overview-last-20-years-2000-2019.

Chatiza, K., 2019. Cyclone Idai in Zimbabwe: An analysis of policy implications for post-disaster institutional development to strengthen disaster risk management. Oxfam Great Britain, Oxford.

Chicco, D. and Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 21, pp.1 - 13.

Cian, F., Marconcini, M. and Ceccato, P., 2018. Normalized difference flood index for rapid flood mapping: Taking advantage of EO big data. *Remote Sensing of Environment* 209, pp.712 - 730.

Clement, M.A., Kilsby, C.G. and Moore, P., 2018. Multi-temporal synthetic aperture radar flood mapping using change detection. *Journal of Flood Risk Management* 11, pp.152 - 168.

Costache, R. and Bui, D.T., 2019. Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: A case study at the Putna river catchment of Romania. *Science of The Total Environment* 691, pp.1098 - 1118.

Domeneghetti, A., Schumann, G.J.P. and Tarpanelli, A., 2019. Preface: Remote Sensing for flood mapping and monitoring of flood dynamics. *Remote Sensing* 11. https://www.mdpi.com/journal/remotesensing/special_issues/rs_flood.

Gan, T.Y., Zunic, F., Kuo, C.-C. and Strobl, T., 2012. Flood mapping of Danube River at Romania using single and multi-date ERS2-SAR images. *International Journal of Applied Earth Observation and Geoinformation* 18, 69 - 81.

Giustarini, L., Hostache, R., Matgen, P., Schumann, G.J.P., Bates, P.D. and Mason, D.C., 2013. A change detection approach to flood mapping in urban areas using TerraSAR-X. *IEEE Transactions on Geoscience and Remote Sensing* 51, 2417 - 2430.

Giustarini, L., Vernieuwe, H., Verwaeren, J., Chini, M., Hostache, R., Matgen, P., Verhoest, N.E. and De Baets, B., 2015. Accounting for image uncertainty in SAR-based flood

mapping. *International Journal of Applied Earth Observation and Geoinformation* 34, pp.70 - 77.

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

Grimaldi, S., Xu, J., Li, Y., Pauwels, V.R. and Walker, J.P., 2020. Flood mapping under vegetation using single SAR acquisitions. *Remote Sensing of Environment* 237, 111582.

Halimu, C., Kasem, A. and Newaz, S.S., 2019. Empirical comparison of area under ROC curve (AUC) and Mathew correlation coefficient (MCC) for evaluating machine learning algorithms on imbalanced datasets for binary classification. Proceedings of the 3rd international conference on machine learning and soft computing (pp. 1-6)..

Hao, C., Yunus, A.P., Subramanian, S.S. and Avtar, R., 2021. Basin-wide flood depth and exposure mapping from SAR images and machine learning models. *Journal of Environmental Management* 297, 113367.

Hashemi-Beni, L. and Gebrehiwot, A.A., 2021. Flood extent mapping: an integrated method using deep learning and region growing using UAV optical data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14, pp.2127 - 2135.

Hosseini, F.S., Choubin, B., Mosavi, A., Nabipour, N., Shamshirband, S., Darabi, H. and Haghighi, A.T., 2020. Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: Application of the simulated annealing feature selection method. *Scince of the Total Environment* 711, 135161.

https://scihub.copernicus.eu Copernicus Open Access Hub [online]. Available from https://scihub.copernicus.eu

Huang, C., Chen, Y., Zhang, S. and Wu, J., 2018. Detecting, extracting, and monitoring surface water from space using optical sensors: A review. *Reviews of Geophysics* 56, 333 - 360.

Huang, M. and Jin, S., 2020. Rapid Flood Mapping and Evaluation with a Supervised Classifier and Change Detection in Shouguang Using Sentinel-1 SAR and Sentinel-2 Optical Data. *Remote Sensing* 12.

Intergovernmental Panel on Climate Change (IPCC), 2013. Climate change 2013: The physical science basis. Cambridge University Press, Cambridge.

Jiang, X., Liang, S., He, X., Ziegler, A.D., Lin, P., Pan, M., Wang, D., Zou, J., Hao, D, Mao, G., Zeng, Y., Yin, J., Feng, L., Miao, C., Wood, E.F. and Zeng, Z., 2021. Rapid and large-scale mapping of flood inundation via integrating spaceborne synthetic aperture

radar imagery with unsupervised deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing* 178, pp.36 - 50.

Lamichhane, N. and Sharma, S., 2017. Development of flood warning system and flood inundation mapping using field survey and LiDAR data for the Grand River near the city of Painesville, Ohio. *Hydrology* 4, 24.

Lei, X., Chen, W., Panahi, M., Falah, F., Rahmati, O., Uuemaa, E., Kalantari, Z., Ferreira, C.S.S., Rezaie, F., Tiefenbacher, J.P., Lee, S. and Bian, H., 2021. Urban flood modeling using deep-learning approaches in Seoul, South Korea. *Journal of Hydrology* 601.

Li, Y., Martinis, S., Plank, S. and Ludwig, R., 2018. An automatic change detection approach for rapid flood mapping in Sentinel-1 SAR data. *International Journal of Applied Earth Observation and Geoinformation* 73, pp.123 - 135.

Liang, J. and Liu, D., 2020. A local thresholding approach to flood water delineation using Sentinel-1 SAR imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 159, pp.53 - 62.

Long, S., Fatoyinbo, T.E. and Policelli, F., 2014. Flood extent mapping for Namibia using change detection and thresholding with SAR. *Environmental Research Letters* 9.

Luu, C., Pham, B.T., Phong, T.V., Costache, R., Nguyen, H.D., Amiri, M., Bui, Q.D., Nguyen, L.T., Le, H.V., Prakash, I. and Trinh, P.T., 2021. GIS-based ensemble computational models for flood susceptibility prediction in the Quang Binh Province, Vietnam. *Journal of Hydrology* 599.

Manjusree, P., Kumar, L.P., Bhatt, C.M., Rao, G.S. and Bhanumurthy, V., 2012. Optimization of threshold ranges for rapid flood inundation mapping by evaluating backscatter profiles of high incidence angle SAR images. *International Journal of Disaster Risk Science* 3, pp.113 - 122.

Martinis, S., Kersten, J. and Twele, A., 2015. A fully automated TerraSAR-X based flood service. *ISPRS Journal of Photogrammetry and Remote Sensing* 104, pp.203 - 212.

Martinis, S., Twele, A. and Voigt, S., 2009. Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution TerraSAR-X data. *Natural Hazards and Earth System Sciences* 9, pp.303 - 314.

Martinis, S., Twele, A. and Voigt, S., 2010. Unsupervised extraction of flood-induced backscatter changes in SAR data using Markov image modeling on irregular graphs. *IEEE Transactions on Geoscience and Remote Sensing* 49, pp.251 - 263.

Mason, D.C., Dance, S.L. and Cloke, H.L., 2021. Floodwater detection in urban areas using Sentinel-1 and WorldDEM data. *Journal of Applied Remote Sensing* 15.

Mavhura, E., Manyangadze, T., Mudavanhu, C. and Pedzisai, E., 2020. An assessment of riparian communities' preparedness to flood risk: the case of Mbire communities in Zimbabwe. *GeoJournal*: pp.1-23.

Mazzoleni, M., Mård, J., Rusca, M., Odongo, V., Lindersson, S. and Di Baldassarre, G., 2021. Floodplains in the Anthropocene: A global analysis of the interplay between human population, built environment, and flood severity. *Water Resources Research* 57, e2020WR027744.

Mudavanhu, C., Manyangadze, T., Mavhura, E., Pedzisai, E. and Manatsa, D., 2020. Rural households' vulnerability and risk of flooding in Mbire District, Zimbabwe. *Natural Hazards* 103: pp.3591 - 3608.

Mullissa, A., Vollrath, A., Odongo-Braun, C., Slagter, B., Balling, J., Gou, Y., Gorelick, N. and Reiche, J., 2021. Sentinel-1 SAR backscatter analysis ready data preparation in Google Earth Engine. *Remote Sensing* 13.(10), p.1954.

Ndehedehe, C.E., Onojeghuo, A.O., Stewart-Koster, B., Bunn, S.E. and Ferreira, V.G., 2021. Upstream flows drive the productivity of floodplain ecosystems in tropical Queensland. *Ecological Indicators* 125, 107546.

Ngo, P.T, Pham, T.D., Hoang, N.D., Tran, D.A., Amiri, M., Le, T.T., Hoa, P.V., Bui, P.V., Nhu, V.H. and Bui, D.T., 2021. A new hybrid equilibrium optimized SysFor based geospatial data mining for tropical storm-induced flash flood susceptible mapping. *Journal of Environmental Management* 280, 111858.

Nhundu, K., Sibanda, M. and Chaminuka, P., 2021. Economic losses from cyclones Idai and Kenneth and floods in Southern Africa: implications on Sustainable Development Goals. *Cyclones in Southern Africa*, pp.289 - 303, Springer.

Ogashawara, I., Curtarelli, M.P. and Ferreira, C.M., 2013. The use of optical remote sensing for mapping flooded areas. *International Journal of Engineering Researsch Applications* 3, pp.1956 - 1960.

Palomba, G., Farasin, A. and Rossi, C., 2020. Sentinel-1 flood delineation with supervised machine learning. Proceedings of the 17th ISCRAM Conference In *ISCRAM* (pp. 1072-1083).

Panahi, M., Dodangeh, E., Rezaie, F., Khosravi, K., Van, Le, H., Lee, M.-J., Lee, S. and Thai, Pham, B., 2021. Flood spatial prediction modeling using a hybrid of meta-optimization and support vector regression modeling. *Catena* 199, p.105114.

Pelich. R., Chini, M., Hostache, R., Matgen, P., Delgado, J.M. and Sabatino, G., 2017. Towards a global flood frequency map from SAR data. *IGARSS*, pp.4024 - 4027.

Perosa, F., Fanger, S., Zingraff-Hamed, A. and Disse, M 2021. A meta-analysis of the value of ecosystem services of floodplains for the Danube River Basin. *Science of the Total Environment* 777, 146062.

Pham, B.T., Luu, C., Phong, T.V., Trinh, P.T., Shirzadi, A., Renoud, S., Asadi, S., Le, H.V., von Meding, J. and Clague, J.J., 2021. Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modeling? *Journal of Hydrology* 592, p.125615.

Pierdicca, N., Pulvirenti, L., Chini, M., Guerriero, L. and Ferrazzoli, P., 2010. A fuzzy-logic-based approach for flood detection from Cosmo-SkyMed data. 2010 IEEE International Geoscience and Remote Sensing Symposium (pp. 4796 - 4798).

Pulvirenti, L., Chini, M., Pierdicca, N., Guerriero, L. and Ferrazzoli, P., 2011. Flood monitoring using multi-temporal COSMO-SkyMed data: Image segmentation and signature interpretation. *Remote Sensing of Environment* 115, pp.990 - 1002.

Pulvirenti, L., Pierdicca, N., Chini, M. and Guerriero, L., 2011. An algorithm for operational flood mapping from Synthetic Aperture Radar (SAR) data using fuzzy logic. *Natural Hazards and Earth System Sciences* 11, pp.529 - 540.

Rana, V.K. and Suryanarayana, T., 2019. Evaluation of SAR speckle filter technique for inundation mapping. *Remote Sensing Applications: Society and Environment* 16, 100271.

Saha, A., Pal, S.C., Arabameri, A., Blaschke, T., Panahi, S., Chowdhuri, I., Chakrabortty, R., Costache, R. and Arora, A., 2021. Flood susceptibility assessment using novel ensemble of hyperpipes and support vector regression algorithms. *Water* 13, 241.

Sarker, C., Mejias, L., Maire, F. and Woodley, A., 2019. Flood mapping with convolutional neural networks using spatio-contextual pixel information. *Remote Sensing* 11, 2331.

Tabari, H., 2021. Extreme value analysis dilemma for climate change impact assessment on global flood and extreme precipitation. *Journal of Hydrology* 593, 125932.

Thangavel, R. and Sreevalsan-Nair, J., 2021. CV4FEE: Flood extent estimation using consensus voting in ensemble of methods for change detection in Sentinel-1 GRD SAR images. *7th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*, pp.1 - 6, doi: 10.1109/APSAR52370.2021.9688390.

Tiwari, V., Kumar, V., Matin, M.A., Thapa, A., Ellenburg, W.L., Gupta, N. and Thapa, S., 2020. Flood inundation mapping- Kerala 2018; Harnessing the power of SAR, automatic threshold detection method and Google Earth Engine. *PLoS One* 15, e0237324.

Tsyganskaya, V., Martinis, S., Marzahn, P. and Ludwig, R., 2018. SAR-based detection of flooded vegetation–a review of characteristics and approaches. *International Journal of Remote Sensing* 39, pp.2255 - 2293.

Tu, H., Wang, X., Zhang, W., Peng, H., Ke, Q. and Chen, X., 2020. Flash flood early warning coupled with hydrological simulation and the rising rate of the flood stage in a mountainous small watershed in Sichuan province, China. *Water* 12, 255.

Twele, A., Cao, W., Plank, S. and Martinis, S., 2016. Sentinel-1-based flood mapping: a fully automated processing chain. *International Journal of Remote Sensing* 37, pp.2990 - 3004.

Ulloa, N.I., Chiang, S. and Yun, S.-H., 2019. Normalized difference scattering index for flood mapping. AGU fall meeting abstracts (Vol. 2019, pp. H13F-02).

Vanama, V.S.K., Mandal, D. and Rao, Y.S., 2020. GEE4FLOOD: rapid mapping of flood areas using temporal Sentinel-1 SAR images with Google Earth Engine cloud platform. *Journal of Applied Remote Sensing* 14, 034505.

Vanama, V.S.K., Rao, Y.S. and Bhatt, C.M., 2021. Change detection based flood mapping using multi-temporal Earth Observation satellite images: 2018 flood event of Kerala, India. *European Journal of Remote Sensing* 54, pp.42 - 58.

Zhang, M., Chen, F., Liang, D., Tian, B. and Yang, A., 2020. Use of Sentinel-1 GRD SAR images to delineate flood extent in Pakistan. *Sustainability* 12, 5784.

Zhang, Q. and Yusop, Z., 2021. Flood catastrophes in a changing environment. *Hydrology Research* 52, pp.1 - 3.

Zhao, J., Chini, M., Pelich, R., Matgen, P., Hostache, R., Cao, S. and Wagner, W., 2020. Deriving exclusion maps from C-Band SAR time-series: An additional information layer for SAR-based flood extent mapping. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* V-1-2020, pp.395 - 400.

Zhao, J., Pelich, R., Hostache, R., Matgen, P., Cao, S., Wagner, W. and Chini, M., 2021. Deriving exclusion maps from C-band SAR time-series in support of floodwater mapping. *Remote Sensing of the Environment* 265, 112668.

Zhou, Y., Wu, W., Nathan, R. and Wang, Q.J., 2021. A rapid flood inundation modelling framework using deep learning with spatial reduction and reconstruction. *Environmental Modelling and Software* 143.

Zimstat, 2020. Mashonaland central province district population projections report. In Zimstat (ed.) Harare: Government of Zimbabwe.

CHAPTER 4: THE USE OF HYBRID DEEP LEARNING FRAMEWORK TO MODEL FLOOD-RECHARGED SOIL MOISTURE FEATURES IN SEMI-ARID LANDS

The chapter is based on: **Pedzisai, E.**, Odindi, J., and Mutanga, O. Use of long short-term memory autoencoder deep learning to model flood-recharged soil moisture in a semi-arid floodplain, under review.

Abstract

Soil moisture is a fundamental climate variable sustaining the terrestrial biosphere. Whereas flood-recharged soil moisture (FRSM) is an important input flux in semi-arid floodplain ecosystems, its spatio-temporal dynamics is not well understood due to lack of field data. While existing active remotely sensing data are valuable to understand soil moisture, a trade-off between high temporal against coarse spatial resolutions limit their utility at local scales hence Sentinel-1 is promising. The study study extracted linear backscatter coefficient *sigma nought* (σ_0) and time series data from 91 pre-processed 10 m multi-temporal dual Sentinel-1 images. The data was collected from both inside and outside the floodplain in a semi-arid area in northern Zimbabwe. To characterize FRSM anomaly, lag and memory, the study built a hybrid deep learning long short-term memory autoencoder (LSTMAE) model based on a recent flood event, which was subsequently evaluated using mean absolute error (MAE) and root mean squared error (RMSE) loss metrics using an independent validation dataset. Validation results showed that both VV and VH-polarized data effectively detected FRSM positive anomaly with very small MAE ($0.0799\sigma_0$; $0.0191\sigma_0$) and small RMSE ($0.0967\sigma_0$, $0.0250\sigma_0$), respectively. In the floodplain, the LSTMAE model detected three positive anomalies for both polarizations. Also, this study study established that the VV LSTMAE model was effective in detecting subtle positive anomalies while VH depicted the longest lag and memory at a local scale. The study concludes that the extraction of σ_0 on Sentinel-1 time series data offers a good understanding of localized FRSM characteristics within semi-arid floodplains.

Keywords

Deep learning; Flood-recharged soil moisture; Long short-term memory autoencoder; Remote sensing; Semi-arid floodplain; Sentinel-1.

4.1 Introduction

Soil moisture is a dynamic key state climate variable influencing climatic, ecologic, economic and hydrologic systems (Seneviratne et al., 2010). It constitutes ~0.05% of the global water cycle (Robinson et al., 2008), hence a critical element in the local, regional and global hydrological budget. Globally, soil moisture sustains vegetation, which in turn facilitates more carbon sinking (~2500 gigatons/year) by photosynthesis than its emission (~60 gigatons/year) through respiration (Lal et al., 2007). High soil moisture sequesters soil organic carbon while low soil moisture increases inorganic carbon storage; essential for climate change mitigation (Lal, 2009; 2004). Soil moisture is an important variable in global circulation models

(Joussaume et al., 1984). Thus, studies have identified soil moisture as critical in maintaining global carbon balance, bio-physical and chemical cycling and mitigating climate change (Berg and Sheffield, 2018; Green et al., 2019; McColl et al., 2017b; Orth and Seneviratne, 2012).

At local scales, soil moisture dissolves nutrients and may leach them if in excess (Ben-Shahar, 1990). In hydrology, soil moisture enables the detection and monitoring of evapotranspiration (Dolman and de Jeu, 2010; Jung et al., 2010), precipitation (Brocca et al., 2020; Yang et al., 2018; Guillod et al., 2015) and discharge (Rezaei et al., 2019). Hence, soil moisture can be used to understand local meteorological characteristics useful in drought assessment (Liu et al., 2020; Seneviratne, 2012), flood potential and flood-related disaster mitigation (Wasko and Nathan, 2019; McColl et al., 2017a) and veld fires (Sungmin et al., 2020). Soil moisture is also important in sustaining crop systems, irrigation optimization (Schmugge et al., 1980) and crop yield estimation (Zhuo et al., 2019). In semi-arid floodplains, flood-recharged soil moisture (FRSM) is important for sustaining natural ecosystems and creates a valuable food security safety net through post-flood farming (Chingombe et al., 2015; Wajih, 2008). In northern Zimbabwe, an important semi-arid middle Zambezi floodplain is periodically flooded hence understanding its soil moisture spatio-temporal dynamics was imperative.

In semi-arid floodplains, soil moisture is recharged through either precipitation (Deshmukh et al., 2022) or flooding (Morales et al., 2021). The latter is known to replenish soil moisture storage that sustain ecosystem productivity (Mohammadi et al., 2017). The soil moisture excess and deficit result in positive and negative anomalies, respectively (Li, et al., 2022). Unlike the precipitation-recharged soil moisture that is quickly depleted by excessive evapotranspiration (Dintwe et al., 2022; Humphrey et al., 2021; McColl et al., 2017a), flooding promotes deep infiltration that triggers a positive (wet) anomaly (Nketia et al., 2022; Sadeghi et al., 2020; Ghazavi et al., 2010), thereby initiating soil moisture memory (Gxokwe et al., 2022; Khorrami et al., 2022). Soil moisture memory is the time ranging from weeks to months before the soil 'forgets' either an extremely wet or dry anomalous event (McColl et al., 2017b; Orth and Seneviratne, 2012; Koster and Suarez, 2001). It is the period that water remains in the soil or the time-lapse for the soil moisture storage anomaly to disappear (McColl et al., 2017a). Seneviratne et al. (2006) noted that soil moisture memory drives the climate system. Therefore, soil moisture memory is vital in understanding terrestrial-meteorological nexus and its role in landscape productivity and vulnerability. As such, soil moisture memory is ideal to assess flood and post-flood dynamics (Mohammadi et al., 2017). Generally, understanding soil moisture memory is model-dependent as influential soil properties insignificantly vary at a local scale

(Martínez-Fernández et al., 2021; Orth and Seneviratne, 2012). In semi-arid areas, floods can trigger positive soil moisture anomalies with distinct memory and lag features. A soil moisture lag proposed in this study is defined as the time window between the anomaly onset and offset dates, i.e., the duration of the anomaly. In that regard, FRSM memory is an important proxy for vegetation productivity in semi-arid floodplains and can be used to predict, detect, monitor, and manage vegetation, agricultural productivity, hydrologic models, and disastrous weather phenomena (McColl et al., 2017b).

Generally, soil moisture can be assessed using models based on water balance equations that subtract output from input parameters (McColl et al., 2017a, 2017b). These models correct bias when simulating physical processes based on laws of physics that govern fluid motion and storage (Luo et al., 2022). Typically, models use balance equations with either precipitation or irrigation as input (McColl et al., 2017a), while flood input remains poorly understood (Morales et al., 2021). For localized inundations, rainfall adds to flood input while evaporation and transpiration could be modelled as constants. McColl et al. (2017a) for instance computed soil moisture based on an integral function on a balance equation with time using time series precipitation, backscatter and sampling frequency to model stored rain fraction. However, mass balance equations that depend on parameterisation are commonly constrained by a lack of observations (Kang et al., 2021). Alternatively, FRSM could be modelled as a net water flux that uses time, an arbitrary value, and an estimate of the initial soil moisture profile (Sadeghi et al., 2020). Sadeghi et al. (2020) derived components including the irrigation term which logically could be replaced by a flood term. Another model captures changes by multiplying the second-order derivative of soil moisture with depth and a diffusive term that subtracts the product of the first-order derivative of the soil moisture with depth and a gravitational water flow term in an unsaturated soil layer (Sadeghi et al., 2020). While the above models exemplify input-output parameterisation, they are designed for either precipitation or irrigation, not FRSM. There are also models that use a standardized approach to assess soil moisture. For instance, Sadeghi et al. (2020) derived a standardized relative soil moisture content expressed in volumetric units. However, model output and remote sensing signatures showed discrepancies (Tian et al., 2019). Thus, to reduce such differences, Tian et al. (2019) proposed a metric to resolve FRSM content input in a floodplain characterized by complex interaction between floods and precipitation.

Model assumptions, lack of expertise, parameterisation, high computational costs and data quantity and quality affect model performance, resulting in high uncertainty (Li et al., 2022;

Koster et al., 2010). Accurate field soil moisture measurements are scarce (Seneviratne et al., 2006), hence the use of active remote sensing products in models have become increasingly appealing (Kang et al., 2021; Kim et al., 2019). At regional scale, remote sensing offers new methods to estimate soil moisture (Zhang and Zhou, 2016). Whereas studies like Tao et al. (2019) combined optical and Synthetic Aperture Radar (SAR) data to retrieve soil moisture at a regional scale, remote sensing products e.g., Advanced Microwave Scanning Radiometer -2 (AMSRE-2) (Okuyama and Imaoka, 2015), Soil Moisture and Ocean Salinity (SMOS) (Kerr et al., 2012) and Soil Moisture Active Passive (SMAP) (Entekhabi et al., 2010) have coarse spatial resolution (Adab et al., 2020; Ma et al., 2019) that limits their utility at local scale. Zhang and Zhou (2016) noted limited applicability of modelling soil moisture at local scales. To overcome such a limitation, Sun and Cui (2021) used a machine-learning method to downscale soil moisture data. The recent launch of Sentinel-1A and 1B satellites by European Space Agency (ESA) offers new prospects to infer local soil moisture spatial and temporal variability.

Soil moisture anomaly, and memory assessment can adopt either machine or deep learning. Machine learning methods are simple but less robust than deep learning algorithms (Ranjan, 2020). Long short-term memory (LSTM) algorithm, an advanced recurrent neural network can robustly model phenomena with temporal dependencies (Hochreiter and Schmidhuber, 1997). It effectively handles non-linearity in complex temporal correlation with unstructured and unlabelled data (Bengio et al., 2021; Park and Kellis, 2015), thus catering for exploding and vanishing gradients (Farahani, 2020; Ranjan, 2020; LeCun et al., 2015). Exploding gradients are truncated through either backpropagation or gated recurrent units (Cho et al., 2014) while vanishing gradients are resolved by rectified linear unit (ReLU) activation (Glorot et al., 2011**)**. The non-linear ReLU activation allows successive layers to disregard irrelevant variations while linear hidden ReLU unit simulates covariance between data values (Hinton, 2012). The LSTM resolves constant error carousel challenge with input and output weights-conflict (Hochreiter and Schmidhuber, 1997). An LSTM has linked cell states with input, forget and output gates which open and close the constant error carousel (Hochreiter and Schmidhuber, 1997). Sungmin and Orth (2021) simulated global daily soil moisture variability using LSTM.

An autoencoder is more robust in reducing high data dimensionality than principal component analysis while reducing reconstruction loss (Farahani, 2020). An anomaly loses information when mapped onto lower dimensional space since it cannot be reconstructed, hence a high reconstruction loss (Farahani, 2020). Since a flood can result in a large reconstruction loss of

soil moisture time series, a hybrid deep learning framework merging an LSTM with an autoencoder that forms a long short-term memory autoencoder (LSTMAE) becomes applicable. Specifically, robust anomaly detection applies an autoencoder while the soil moisture memory extraction requires an LSTM (Farahani, 2020). Using sequential data, Sungmin and Orth (2021) confirmed that the LSTMAE effectively analyses long-term dependencies across time steps while Farahani (2020) proved its superiority in anomaly detection. On that merit, we simulated FRSM using LSTMAE that embedded LSTM layers into the encoder and the decoder parts of the autoencoder. The scarcity of high spatial resolution L-band soil moisture products (Entekhabi et al., 2010) provides a need to evaluate the utility of free Sentinel-1 data to detect FRSM characteristics. Thus, the LSTMAE algorithm was applied on a 3-year window (2017/01/01 - 2019/12/31) in Mbire semi-arid floodplain, Zimbabwe. FRSM anomaly, lag and memory were modelled based on *sigma nought* (σ_0) inside and outside the flood zone to facilitate the evaluation of floods on floodplain soil moisture. Existing remote sensing-based soil moisture data products e.g., AMSRE-2, SMAP, and SMOS, suffer from a trade-off between very low spatial and high temporal resolutions. Clearly, these data products are less applicable at local spatial scales. Therefore, the novelty of this study is in modelling FRSM using moderate spatial resolution Sentinel-1 data to address local spatial coverage. Specific objectives to this study were to, firstly, detect FRSM anomaly using the LSTMAE algorithm secondly, estimate the temporal post-flood FRSM lag and memory in the semi-arid floodplain and finally, compare the utility of Sentinel-1 co-polarized (VV) against cross-polarized (VH) data to detect FRSM anomaly, lag and memory.

4.2 Materials and methods

4.2.1 The study area

The study area is located in Chitsungo in the flood-prone rural Mbire District, northern Zimbabwe (Figure 4.1). The area is at an average altitude of 400 m above sea level (Gumindoga et al., 2020). It is in semi-arid area with low annual precipitation (~500 mm), with high summer (>30ºC) and moderate winter (15ºC) temperatures (World Food Programme, 2016). The area has a population of about 83 724 (Zimbabwe National Statistics Agency, 2022). Land use in the area is predominantly subsistence agriculture, including in the floodplain where the rural community benefits from the silt soils and the FRSM (Chingombe et al., 2015). Land cover in the area includes partly forested floodplain that experiences seasonal floods and droughts (Gumindoga et al., 2020), while savannah grasslands and thorn bushes occur outside the

flooded zone. The area has physical infrastructure that include rural homes, clinics, schools, shops and roads.

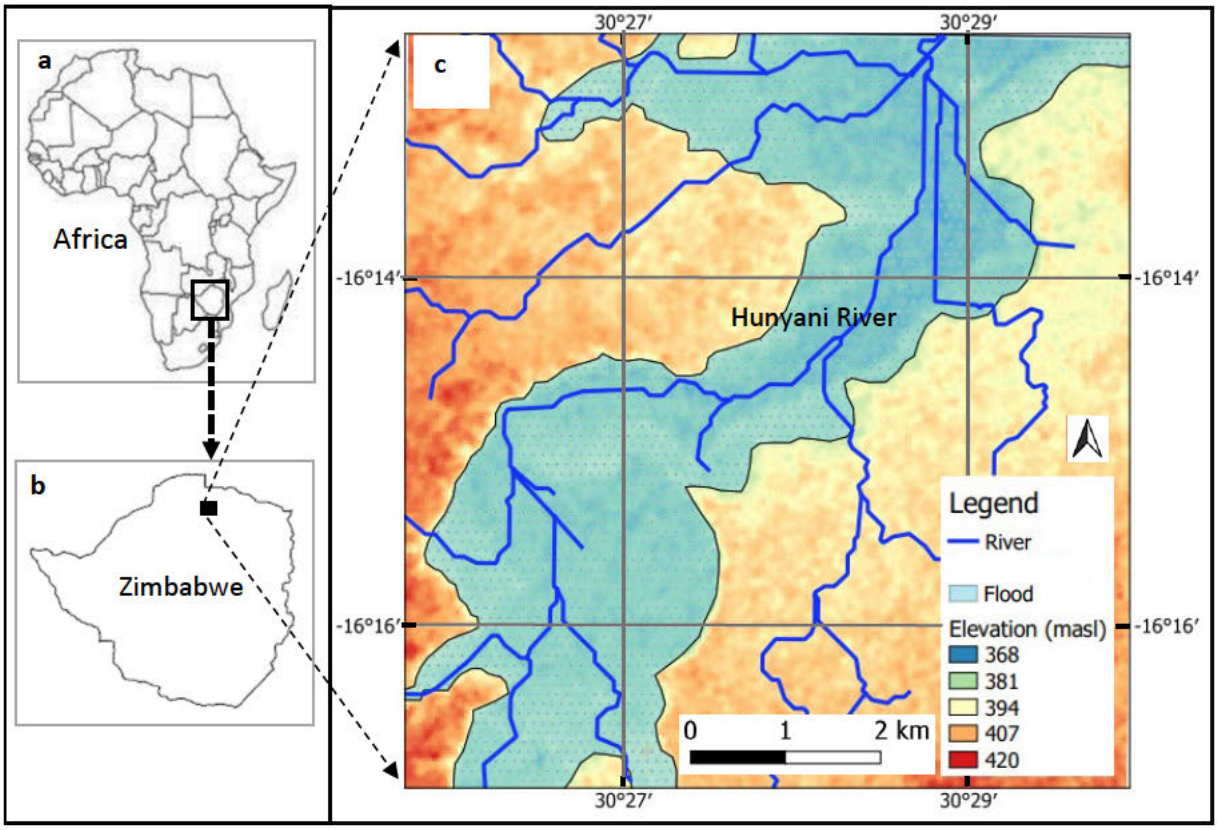


Figure 4.1: Map of Africa (a), Zimbabwe (b) and location of Chitsungo study area (c) showing elevation and drainage network.

4.2.2 Data collection

4.2.2.1 Field and remotely sensed data

During a community-guided fieldwork, field points were extracted from inside and outside the 2017/02/02 flood zone in the floodplain. Fieldwork was guided by the community who indicated whether a randomly selected location point captured using a Garmin eTrex 10 GPS was inside or outside the flood extent. Ninety-one Sentinel-1 dual polarized (VV+VH) C-band SAR interferometric wide swath captured on ascending and Ground Range Detected High Resolution (GRDH) models for the period from 2017/01/01 to 2019/12/31 were downloaded from the Copernicus Open Access Hub provided by the ESA (https://scihub.copernicus.eu). The images are 10 m spatial and 12-day temporal resolutions.

The Sentinel-1B data was pre-processed using Sentinel-1 Toolbox with an in-built Graph Builder Tool that creates a visualisation using five connected step-to-step architecture while building an executable code in the backend within the SNAP (8.0.0) software environment.

The script was used to undertake orbital file application, thermal noise removal, radiometric calibration, terrain correction to linear scale sigma nought (σ_0) on the image and sub-setting to spatially delimit the study area. The image preprocessing was outlined earlier in Appendix 3.1.

4.2.3 Methods

The first stage partitioned 91 extracted backscatter values from time series data into training (61) and testing (30) using 'train_test_split' in Scikit-Learn (1.1.2) Library by polarization for a location inside and another outside the floodplain. The data were not normalized since the variances were sufficiently low. The second stage implemented an advanced neural network in Python (3.8.8) within Spyder (4.2.5) using Keras (2.4.2) deep learning library on a TensorFlow (2.3.0) backend to build, train and test a hybrid LSTMAE model. Two LSTM hidden state layers were embedded into encoder and decoder parts within an autoencoder architecture that was compiled by activation and two loss metrics. Sensitivity analysis tuned a sequential Keras model parameterized by batch size, number of epochs, and validation split ratio. The third stage evaluated model performance using mean absolute error (MAE) and mean square error (MSE) loss metrics on both training and validation. The fourth stage derived thresholds of 0.75 of the maximum and minimum of the train mean error (trainME) of sigma nought value, respectively. The fifth stage modelled anomalies based on empirical distribution of test mean error (testME) as compared to trainME values. Anomaly onset and offset dates were extracted to define a proposed feature, FRSM lag, while from onset to mean defined the post-anomaly memory. The sixth stage compared the positive anomaly, lag and memory features against non-flood zones to deduce the influence of on floodplain soil moisture. The model was executed twice inside and outside the flood zone using Sentinel-1 data. A script was used in the study to implement the hybrid deep learning LSTMAE model. The modelling steps resulted in the detection of FRSM anomaly, estimation of lag and memory features. The final stage applied descriptive statistics to evaluate the utility of polarization mode to understand the influence of floods on soil moisture using dual polarization backscatter multi-temporal sequences for pixels inside and outside the floodplain. The exploratory data analysis stacked four time series curves based on polarization to compare flood with non-flood soil moisture in the study area.

4.2.3.1 The modelling approach

The modelling of anomaly, lag and memory on the time series involved defining structure, building, parameterization, implementation and evaluation of the model using loss metrics. A simulation comprised a single time series into the LSTMAE architecture based on Keras

Library in TensorFlow backend. Based on VV and VH, the model was run four times; two inside to detect FRSM anomaly, lag and memory and two outside the floodplain for comparison. Pixel sigma nought values were extracted in ArcGIS Pro (2.9) from images.

- *Description of the LSTMAE model structure*

The LSTM is based on connected neurons. A neuron is a segment of the LSTM that constitutes three gates and memory (cell state) which can be incorporated into a neural network with input, hidden and output layers. The input gate specifies how input data updates the memory, a forget gate determines prior memory values to keep or discard from previous cell state, a cell state keeps the long-term memory, and the output gate sets the extent of effect of input and memory to feed to the next cell (Abduljabbar et al., 2021; Olah, 2015). These four components are calculated on the instant cell state within an LSTM architecture. An unrolled neural network can form an LSTM learning architecture with connected cells namely the prior slow state (long-term memory) and fast state (short-term memory) that run together with input through the forget, input and output gates with sigmoid activation to produce updated time-specific cell states (Ranjan, 2020). Input from a previous hidden cell state is fed through input gate into the current observation (cell state). Then, the forget gate applies a sigmoid activation on the input to either keep or forget. The memory is updated to current cell state, which finally, in combination with stored weights are forwarded to the next cell state through the output gate.

The algorithm comprised an autoencoder architecture with three connected parts namely encoder, code and decoder (Figure 4.2). The encoder is an inferential function defined as the probability Z given X, P(Z|X). The code (Z) is a bottleneck of compressed representation in latent space storing learned dependences in the sequential data. The decoder is a reverse function to the encoder that reconstructed the time series (X) by computing the probability of X given the representation Z, P(X|Z). The architecture fitted two LSTM encoder layers, one code and two decoder layers into an autoencoder structure specifying units, defining activation and sequence preservation. Anomaly detection was based on the decoder function of regenerating the sequence signified by big magnitude of differences between training and testing errors. Appendix 4.1 shows all modelling steps in Spyder (Python) used in the chapter.

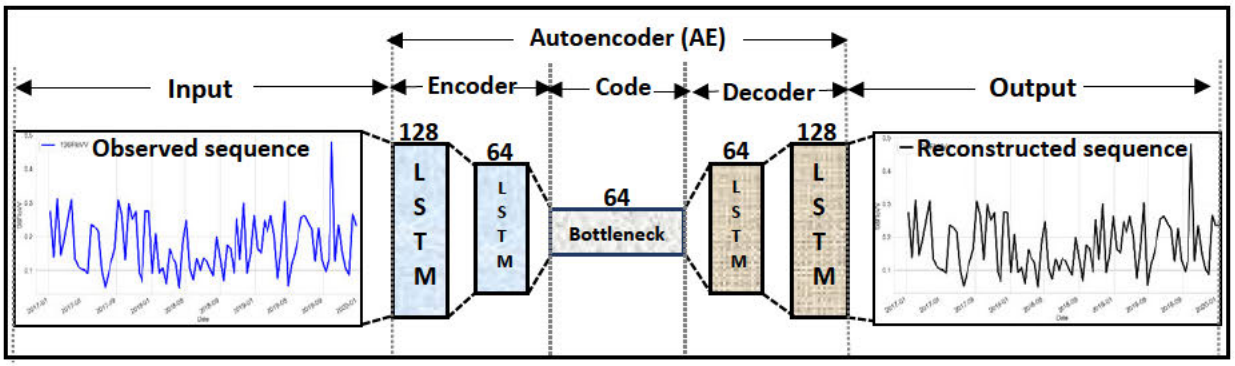


Figure 4.2: An anomaly detection LSTMAE model with observed sequence (input) into dimensionality reduction two LSTM layers (encoder), a latent storage (code), and decompressing (decoder) that regenerates output sequence in two LSTM layers while the number of units at each part is shown, respectively.

4. 2.3.2 Model building and parameterization

Model pre-processing involved partitioning the Sentinel-1B data. Modelling splits data into train and test sets for calibration and validation to evaluate its ability to generalize the learned patterns (LeCun et al., 2015). A data partitioning into train (67%) and test (33%) based on two non-flood years (2018, 2019) and one flood year (2017), respectively for the LSTMAE model was used. We trained the model on normal non-flood data (n=61) to enable detection of any anomalies during the flood year (n=30) test dataset. Observed sequential data were input into two LSTM inferential hidden encoder layers with 128 and 64 units. Stacking LSTM layers in an autoencoder facilitates efficient feature detection (Bengio et al., 2006), hence the LSTMAE architecture allowed for sequence preservation. The output from encoder was represented in latency in a code with 64 units which enabled reconstruction of the initial sequential data by the decoder with 64 and 128 units. A single *TimeDistributed* dense layer with 128 units facilitated a reconstructed time series comparable to the observed layer. The LSTMAE training iteration sequence used epoch, batch and sample to derive a sequential time-step *(*Ranjan, 2020*)*. Within an epoch, a batch was activated and a sample extracted to constitute a single time-step comprising four sequenced stages. The model was trained using the fast-learning *ReLU* activation noted by Glorot et al. (2011).

The LSTMAE compilation applied adam optimization (Kingma and Ba, 2014), which is robust as it combines the efficient handling of sparse gradient optimizer AdaGrad (Duchi et al., 2011) with the RMSProp appropriate for sequential and dynamic data (Tieleman and Hinton, 2012). The adam optimization was selected because it is efficient in both calculations and memory, is

less sensitive to hyper-parameters, solves gradient explosion and decay challenges affecting machine learning algorithms (Ranjan, 2020). The model was also compiled for loss using the MAE and MSE metrics which appropriately evaluate model performance by comparing the observed against predicted instant values in the sequence. The model was parameterized with 1000 epochs, batch_size of 2, validation split ratio 0.33 in a sequential model that generated 247,937 trainable parameters to simulate sigma nought since the flood date.

4.2.3.3 Model implementation

An extraction of the anomaly, lag and memory using the LSTMAE is presented on Figure 4.3a on a mean error plot that resembles a wave form.

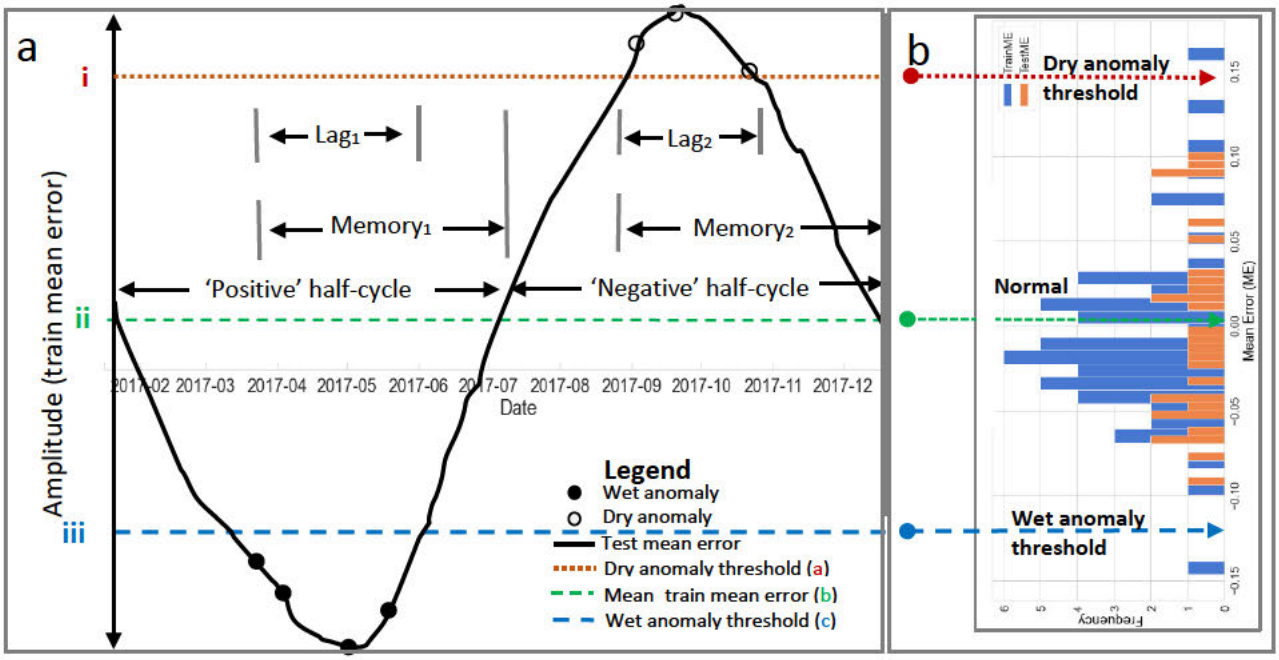


Figure 4.3: Graphical definition of three soil moisture features namely anomaly, lag and memory (a) based on test mean error metric and the empirical determination of the respective thresholds (b) based on the trainME and testME distribution.

The critical point in the model lies in the application of the testME amplitude to model anomaly within an annual cycle, half-cycle memory and lag for the negative and positive phases plotted on date as x-axis, against trainME as y-axis (Figure 4.3a). The y-axis locates three trainME parameters namely maximum (max_trainME), mean (mean_trainME) and minimum (min_trainME) that are required to compute and plot anomaly, lag and memory while time (Date) is the x-axis. The trainME is used to delimit both the negative (max_trainME) and positive thresholds (min_trainME) as shown on Figure 4.3b.

A mean error (ME) distribution on train data was computed based on an objective function depicting distance between the observed and the predicted while keeping the sign (+/-) of the differences to enable distinction between positive and negative anomalies. Two boundary thresholds namely train MEs max_trainME*0.75 and min_trainME*0.75 were required thresholds to model negative and positive anomalies, respectively. We thus determined empirical positive, and negative thresholds using the train data histogram distribution of backscatter. To accurately model wet anomalies either inside (flood + precipitation) or outside (precipitation only) the flood zone and dry anomalies due to seasonal drought, we stacked trainME with testME histogram distributions (Figure 4.3b).

The study proposed Equations 4.1 - 4.5 to derive the third threshold; the mean_trainME which benchmarks the 'normal' as computed. Specifically, the mean of observations (Equation 4.2) and mean of predictions (Equation 4.3) on train data are arithmetically combined to model the optimum which was considered to be mid-point between the observed and predicted (mean_trainME), hence Equation (4.4), while the testME was derived in Equation (4.5);

$$\text{trainME} = \frac{1}{n}\sum_{1}^{n}\left(\text{train}_{\text{pred}} - \text{train}_{\text{obs}}\right) \quad (4.1)$$

$$\text{meantrain}_{\text{obs}} = \frac{1}{n}\sum_{1}^{n}\left(\text{train}_{\text{obs}}\right) \quad (4.2)$$

$$\text{meantrain}_{\text{pred}} = \frac{1}{n}\sum_{1}^{n}\left(\text{train}_{\text{pred}}\right) \quad (4.3)$$

$$\text{mean\_trainME} = 0.5 * \left(\text{meantrain}_{\text{obs}} + \text{meantrain}_{\text{pred}}\right) \quad (4.4)$$

$$\text{testME} = \frac{1}{n}\sum_{1}^{n}\left(\text{test}_{\text{pred}} - \text{test}_{\text{obs}}\right) \quad (4.5)$$

where: trainME and testME are the empirical train and test mean error, $\text{train}_{\text{pred}}$ and $\text{test}_{\text{pred}}$ are the model results for train and test data, while $\text{trainME}_{\text{obs}}$ and $\text{testME}_{\text{obs}}$ are the observed train and test datasets, respectively. With the objective to detect flood and aridity anomalies, the independent identical distribution of train and test samples (Bengio et al., 2021) was not upheld in the model since train data was regarded as normal non-flood years (2018 and 2019), while anomaly in flood year (2017) was the test dataset, hence flooding and drought defined with anomalous distribution benchmarked against extremes in the train set (trainME). The negative and positive thresholds were extracted from the histograms as drawn using Spyder (Anaconda Python 3.8.8) (Figure 4.3b). This fluctuating testME is plotted on the trainME (y-axis) and Date (x-axis). The 'intersect' of the testME with the max_trainME and min_trainME, may define the positive and negative anomalies which may be influenced by prolonged aridity and flood, hence the determination of memory. The mean_trainME is the mean of the train and test datasets.

Once we defined the three thresholds (i.e., min_trainME, mean_trainME and max_trainME) (Figure 4.3b), we translated them into positive, normal, and negative thresholds, respectively. (Equations 4.6 - 4.8). Specifically, where Boolean logic held a True value in each case, a combination of comparative operators defined positive and negative anomalies (Equations 4.6 and 4.7) while the normal was as defined in Equation (4.8);

$$\text{Positive Anomaly} = \text{iff}(\text{min\_trainME} > \text{testME}, \text{True}, \text{False}) \quad (4.6)$$

$$\text{Negative Anomaly} = \text{iff}(\text{max\_trainME} < \text{testME}, \text{True}, \text{False}) \quad (4.7)$$

$$\text{Normal} = \text{iff}(\text{mean\_trainME} = \text{testME}, \text{True}, \text{False}) \quad (4.8)$$

where: min_trainME, max_trainME, and mean_trainME are as defined above. Therefore, the specific dates of positive and negative anomalies were detected, enabling the ultimate calculation of temporal lag and memory features executed by applying the minimum and maximum mean error thresholds of the train mean error. Figure 4.4 shows how positive and negative anomaly thresholds were extracted using the min_trainME and max_trainME, respectively. The study used histograms (Figure 4a) to translate the three thresholds into positive and negative anomaly detection (Figure 4.4b). These anomalies enabled the estimation of onset and offset dates useful to calculate negative (dry) and positive (wet) features on lag and memory (Figure 4.4c), respectively. A truncated positive (wet) anomaly around October 2017 could be a result of a random dryness period that could have been followed by a wetting rainfall event.

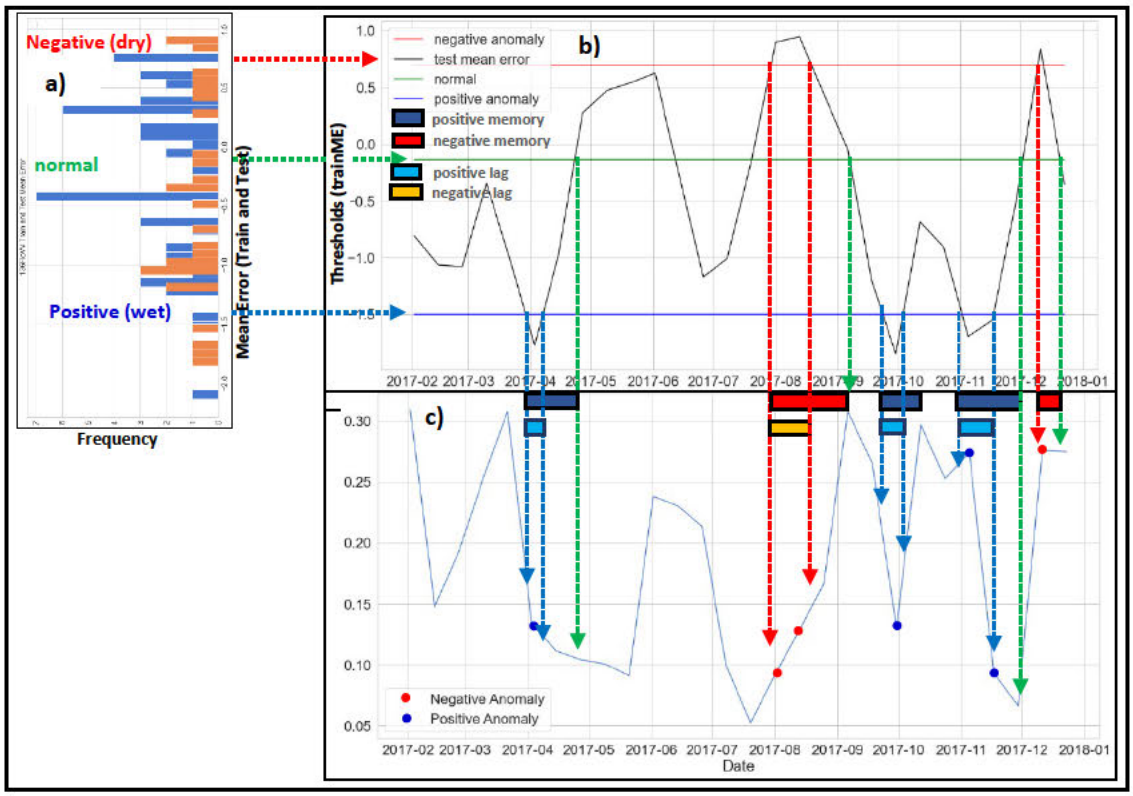


Figure 4.4: The translation of the ME (a) into positive, normal and negative thresholds (b), determination of dates of anomalies, lag and memory (c) features, respectively.

A calculation of a lag differenced the first date detected as anomaly from the last date of appearance of the same anomaly. A memory was defined as the difference between the first date of appearance of the positive anomaly from the date of detection of the normal.

4.2.3.4 LSTMAE model evaluation

McColl et al. (2017a) identified several approaches based on either models or point-scale observations to validate soil moisture anomaly and memory. Examples include timescale-based methods which check the degree of correlation between anomalies over a fixed time window since the anomaly is defined as a deviation from a reference state, e.g., seasonal mean, and time-varying trend. We evaluated each individual polarization run by comparing training with validation loss. Moderate spatial resolution Sentinel-1 SAR sequential data was applicable to detect and monitor temporal backscatter values inside and outside the flood zone so as to infer FRSM features. Positive and negative anomalies can be inferred to onset of flood wetting and aridity due to drought, respectively.

The LSTMAE algorithm evaluation compared the observed from predicted for the training as well as testing. The LSTMAE model evaluation applied the MAE and RMSE loss metrics computed as shown in Equations 4.9 and 4.10 (Li et al., 2022), respectively.

$$\mathrm{MAE} = \frac{1}{\mathrm{n}}\left(\sum_{\mathrm{i}=1}^{\mathrm{n}} |\mathrm{X}_{\mathrm{obs(i)}} - \mathrm{X}_{\mathrm{pred(i)}}|\right) \tag{4.9}$$

$$\mathrm{RMSE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{X}_{\mathrm{obs(i)}} - \mathrm{X}_{\mathrm{pred(i)}}\right)^2 \tag{4.10}$$

where: $\mathrm{X}_{\mathrm{obs(i)}}$ is the i^{th} observation of the test sequence, and $\mathrm{X}_{\mathrm{pred(i)}}$ is the corresponding i^{th} prediction of the test SAR backscatter signal Sentinel-1B dates. The MAE computes the summation of the magnitude of the differences between the observed versus the predicted values in the sequence, while the RMSE squares differences the data.

Once the study evaluated the model using the two metrics, a third metric; mean error (ME) that allowed for detecting both positive and negative soil moisture extremes was necessary to distinguish the wet flood-recharged from the dry drought-depleted phases, respectively. Therefore, we computed the time-specific trainME data used to define thresholds and compared to testME to plot anomaly and the subsequent temporal lag and memory components. While some of the metrics explicitly consider either positive, or negative anomaly (e.g., Hou and Orth, 2020; Brubaker and Entekhabi, 1996), this study innovated the detection of both anomalies in one model. Therefore, the study computed a ME to detect both positive and negative anomalies (Equation 4.11);

$$\mathrm{ME} = \frac{1}{(\mathrm{n}-\mathrm{bs})}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{x}_{\mathrm{pred(i)}} - \mathrm{x}_{\mathrm{obs(i)}}\right) \tag{4.11}$$

where: n is the number of observations, bs is batch size, $\mathrm{x}_{\mathrm{obs(i)}}$ is the observed backscatter value at instance (i) and $\mathrm{x}_{\mathrm{pred(i)}}$is the predicted value at the same instance as observed. The study computed the ME on training and evaluated it on testing datasets.

4.3 Results

4.3.1 The LSTMAE model

Empirical thresholds to determine soil moisture anomalies using the test data are shown in Table 4.1. The maximum metric was consistent by polarization mode with VV higher than VH, while the mean and minimum varied in the negative between the flooded against the non-floodplain areas. Applying these computed thresholds in anomaly translated into an illustration as shown in Figures 4.5a, 4.5b, 4.5e and 4.5f.

Table 4.1: Thresholds derived from train data to determine anomalies on test data.

Threshold	Flood Zone		Not Flood Zone	
	VV	VH	VV	VH
Negative	0.130	0.012	0.120	0.020
Normal	-0.020	-0.002	-0.015	-0.004
Positive	-0.140	-0.025	-0.110	-0.040

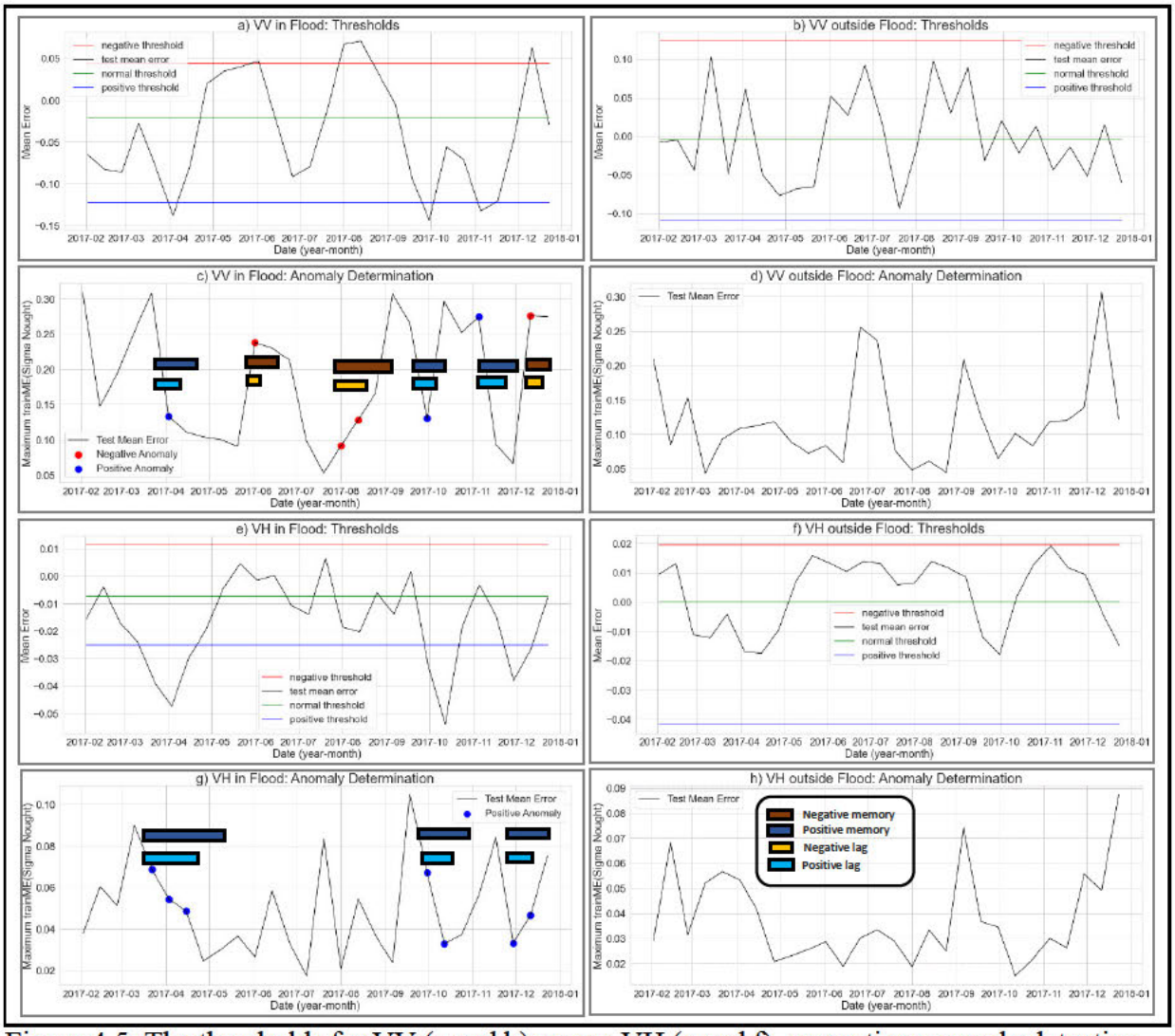


Figure 4.5: The thresholds for VV (a and b) versus VH (e and f), respective anomaly detection, lag and memory calculation for VV (c and d) as compared to VH (g and h) for points inside and outside the flood zone in Mbire, Zimbabwe.

The excess of testME on the bottom and top threshold lines (Figure 4.5a and 4.5e) indicates occurrence of the positive and negative anomalies, respectively. In the flood zone, VV polarization detected both negative and positive anomalies (Figure 4.5a), while VH detected only positive anomalies in flooded Outside the flood zone, both polarizations did not detect any anomalies (Figure 4.5d and 4.5h). There are three positive anomalies beyond the wet threshold detected on post-flood and dry season (Figure 4.5). The first positive anomaly was

evident at the beginning of the year, for both VV and VH polarizations on the flood zone (Figures 4.5c and 4.5g). On the other hand, three negative anomalies occurred only on flood zone, detected only by VV polarization data during the dry months of the year (Figure 4.5c).

4.3.2 FRSM anomaly, lag and memory detection using LSTMAE

We detected a positive FRSM anomaly on 03/04/2017 with more than two months post-flood memory in the floodplain, unlike outside. Figure 4.5 shows the anomalies, lag and memory for both polarizations inside (left) as opposed to none outside the floodplain. The VV-polarized mode in floodplain depicts both negative and positive anomalies (Figure 4.5a and 4.5c). The VH showed only three positive anomalies, with the first linked to flood having a long memory of approximately 70 days. Outside floodplain, both VV and VH polarized data did not show any anomaly (Figure 4.5b, 4.5d, 4.5f and 4.5h). In the floodplain, the VV model detected both anomalies unlike the VH which depict only the positive in the floodplain. The detected early positive anomaly in both VV and VH polarizations is probably linked to the February flood-wetting event. Figure 4.5c, modelled using VV polarized data, shows both wet and dry anomalies within floodplain, unlike Figure 4.5d modelled using VH polarized data outside floodplain. Specifically, a post-flood anomaly in the first quarter of the year 2017 on Figure 4.5g clearly shows that exclusively inside the floodplain, a longer memory and lag features associated with positive anomaly are evident. The duration of the features corresponds to the length of the bars connecting date of onset to date of offset for both lag and memory features, respectively. The length of the bars is evident that the estimations of the time by noting changes from dates of onset to offset hence the 70 days window of memory stated above. This is not true for unlike Figure 4.5h outside the floodplain.

4.3.3 Evaluation of the LSTMAE performance

Modelling studies emphasise on evaluation based on the calibration and validation datasets. In this study, we used the MAE and RMSE loss metrics to evaluate the LSTMAE model. Sensitivity analysis assessed the number of epochs, and layers in the architecture. The sensitivity analysis showed the ideal parameterisation of 1000 epochs, batch size of two, and the train/test (67/33) split was used to model the temporal dependencies in the sequential data based on the three-year period. The relatively small gain in training MAE loss at 1000 epoch barely compensates for the relatively higher testing MAE over 400 epoch. The MAE and RMSE loss metrics are shown on Figure 4.6.

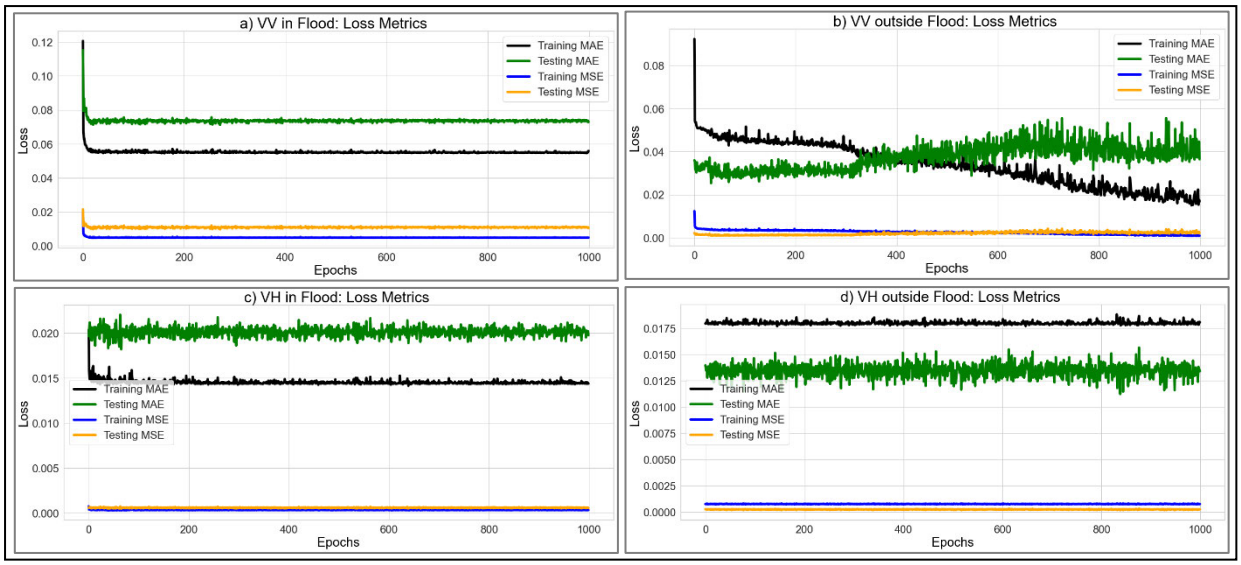


Figure 4.6: The training and validation loss showing MAE and MSE for VV (a and b) versus VH polarisation (c and d) inside (left) and outside (right) flood area in Mbire, Zimbabwe

The MAE for validation was higher than for training on the flood zone. The RMSE was comparatively higher than MAE both inside and outside the flood zone, which is a square of the root mean square error. There was a sharp drop of loss on both metrics in the initial epochs. The model was less stable, probably due to the short length of the time series. The loss metric values for the train and test samples are presented by location and polarization mode in Table 4.2.

Table 4.2: The LSTM-AE model MAE and RMSE by polarisation and location.

Metric	Train (n = 61)				Test (n = 30)			
	Flood		Non-Flood		Flood		Non-Flood	
	VV	VH	VV	VH	VV	VH	VV	VH
MAE	0.0799	0.0159	0.0221	0.0160	0.0799	0.0191	0.0636	0.0137
RMSE	0.0835	0.0192	0.0358	0.0243	0.0967	0.0250	0.0745	0.0179

MAE is the mean absolute error, and RMSE is the root mean squared error.

As shown on Table 4.2, the model computed the MAE, and RMSE loss functions on train and test simulations. These metrics are comparably similar by polarization in floodplain between training and testing datasets. The same applies to the non-flooded zone, hence the model was consistent between train and test by polarization. A dropout architecture was built to check if the model was overfitting. A single random VH dropout (rate = 0.2) scenario inside the floodplain was run. However, results showed a similarity in the two model structures that considered the cross polarization in the floodplain.

In Table 4.3, both model training and testing had high performance characterized by an extremely low ME proving the utility of LSTMAE to model complex influence of floods on soil moisture. Further evaluation of train and test datasets by location and polarization is shown in Figure 4.7**.**

Table 4.3: LSTMAE model evaluation ME metric (train: n = 61, test: n = 30)

Polarisation	Class	Mean error (ME) metric	
		Train	Test
VV	Flood	9.3250E-08	-4.8197E-08
VH	Flood	-6.0371E-08	-6.8084E-07
VV	Non-flood	3.7465E-08	-2.1228E-07
VH	Non-flood	-1.7228E-07	2.8774E-08

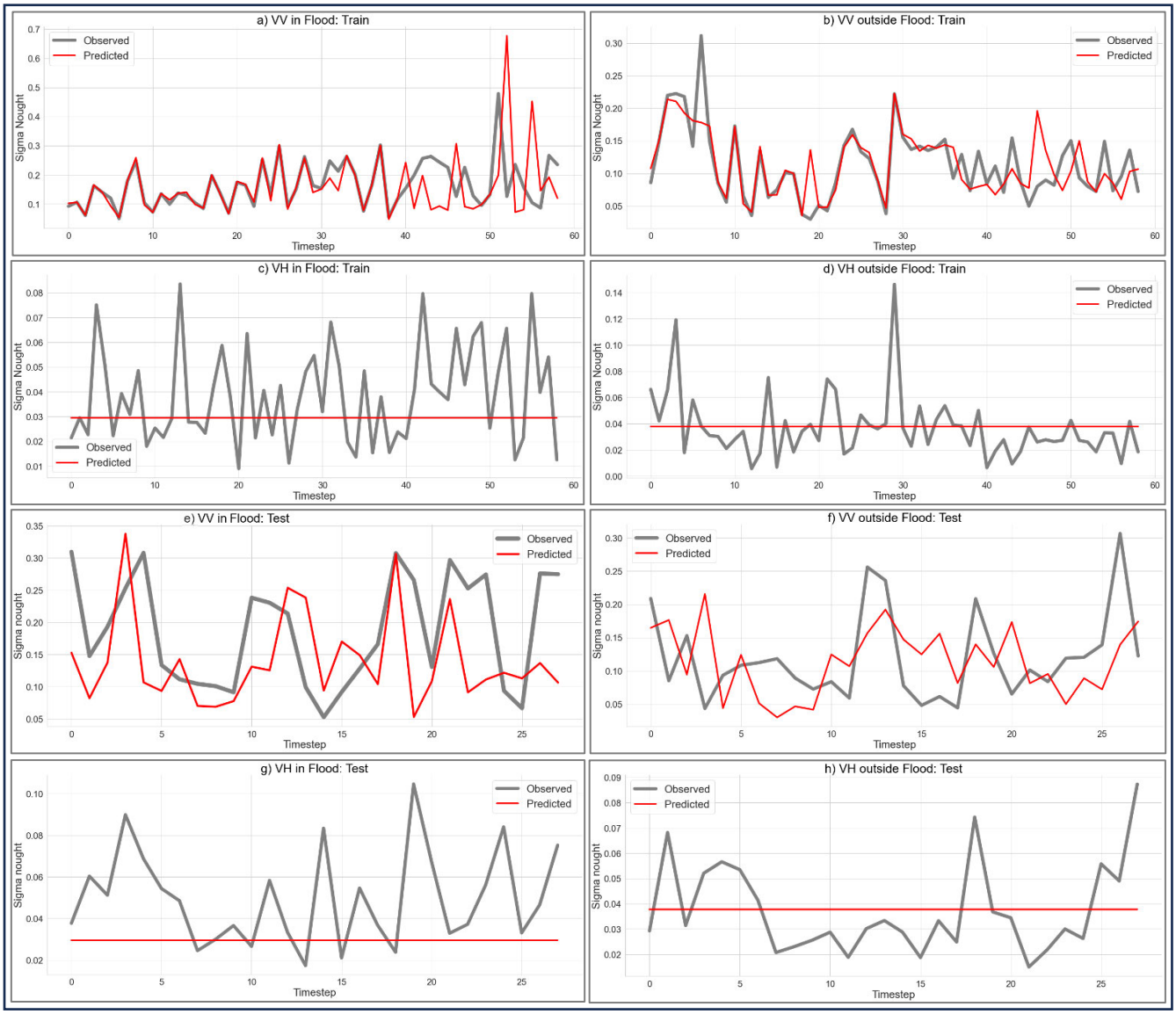


Figure 4.7: Evaluation of the LSTMAE algorithm for training used VV (a & b), and VH (c & d), the testing used VV (e & f) and VH (g & h) with left and right columns showing locations inside and outside the flood zone, respectively.

Results in this study showed that LSTMAE algorithm was useful to model FRSM features with small difference in all the four training and testing samples. Varying of the Y-axis seems unable to model the minimal changes in the predictions based on VH polarized data on both training and testing inside and outside the floodplain This unusual result shows that VH data maybe less useful for extracting the soil moisture features both for wet and dry soils. However, in this case, the algorithm was trained on 61 Sentinel-1B data for the two years and tested on one year to reduce over-fitting challenge.

4.3.4 The exploratory data analysis results

The exploratory data analysis tested the existence of potential anomalies by plotting the four backscatter signal time series, moving average and difference by location and polarization (Figure 4.8). The VV-polarized backscatter was higher than VH for both inside and outside the floodplain. The moving average result showed that VV was higher inside than outside the floodplain, exceptionally during the year 2018 that had no floods. Although the year 2019 also had not floods, the long-lasting effect on the post flood season with deep soil moisture recharge could have influenced that temporal bahaviour. The difference plot showed sporadic fluctuations and sharp spikes that are evident, especially on VV both inside and outside the floodplain. However, distinguishing VV backscatter signal inside from outside the flood zone remains a challenge. The same applied to the VH between the floodplain and non-floodplain.

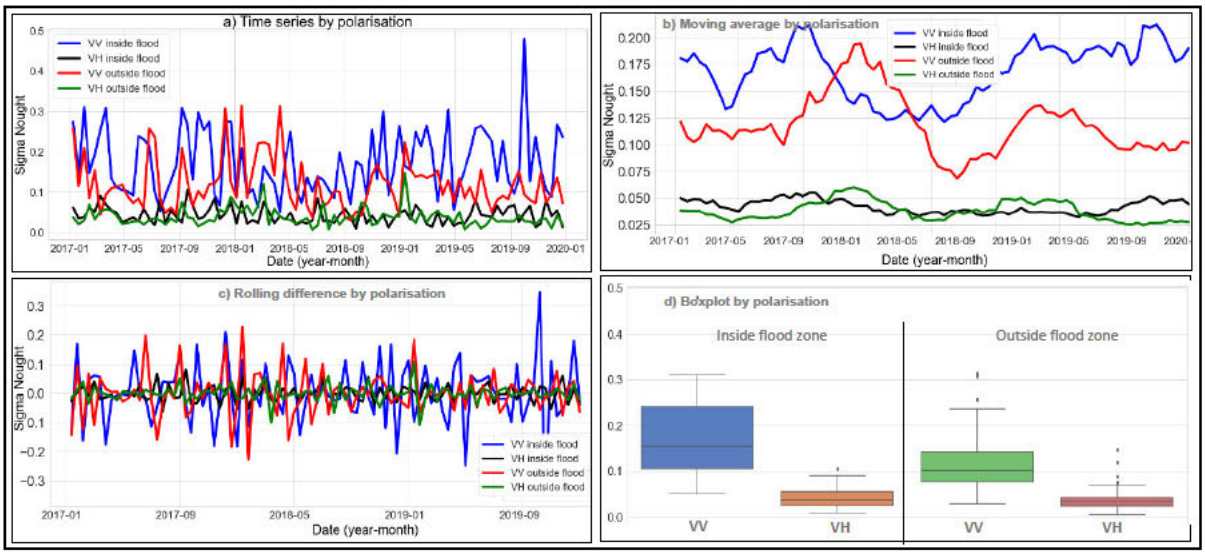


Figure 4.8: Two co-polarisation (VV) and two cross-polarisation (VH) for flood and non-flood zone time series (a), moving average (b) and rolling difference (c), and the respective boxplot of their backscatter values in Chitsungo area in Mbire, Zimbabwe.

Further, a pairwise polarization-based comparison was done (Figure 4.8). Two stacked time series of backscatter coefficient (Figure 4.8a), moving average window (Figure 4.8b) and difference (Figure 4.8c) were plotted for floodplain versus non-floodplain, respectively. The comparison of VV showed haphazard behaviour in time series between flood and non-flood areas. The same applies to VH. As such, temporal behaviour is evident on the moving average (Figure 4.8b).

Notably, the moving average showed a cyclic behaviour with evident amplitude and half-cycles for both VV and VH pairs. However, VV is clearly higher in flood than non-flood with an overlap occurring during 2018. The VH comparison showed two peaks in amplitude for non-flood VH, with overshooting occurring in 2018 and 2019, which were coincidentally non-flood years. The differences plot showed discordantly fluctuating VV and VH plots with irregular spiking. The difference plot discordant spiking behaviour inside and outside the floodplain suggest potential anomalies in the time series. Hence, a comparison of the effect of flooding on polarization backscatter was the basis of the second comparison between VV and VH inside and outside the floodplain.

The polarization-based comparison does not show distinct differences in time series. However, a cross locational-based comparison (Figure 4.8) showed that the time series depict a persisting difference between both VV and VH in the floodplain as well as non-floodplain. The moving average for flood as compared to non-flood clearly distinguished higher VV as compared to lower VH in both floodplain and non-floodplain. There is more variability and evidence of spiking in VV as compared to the VH for both inside and outside the floodplain. Figure **4.**8(a-c) shows lower backscatter for VV-polarized time series, moving average and less fluctuations than VH-polarized. The spiking fluctuations on the amplitude and cyclic dynamic features suggest existence of backscatter anomalies, and temporal dynamics. As such, a model that efficiently handles temporal dependences to simulate anomalies based on multi-date data was crucial.

4.4 Discussion

This study detected FRSM anomaly, lag and memory features by applying a hybrid deep learning LSTMAE algorithm on Sentinel-1 SAR data based on a February 2017 flood event. The LSTMAE was applicable to infer backscatter values to FRSM to understand the influence of post-flood soil moisture temporal dynamics in Mbire floodplain, Zimbabwe. This model was partly based on machine and deep learning on LSTM method used by Li et al. (2022) to

conclude that soil moisture memory was conspicuous for both dry and wet anomalies. The three FRSM features were useful to infer FRSM effect on floodplains (Figure 4.5) after Zhang and Zhou (2016) had deduced a close relationship between backscatter coefficient and soil moisture model results.

4.4.1 Anomaly detection using LSTMAE algorithm

The first objective sought to detect FRSM anomaly using thresholds derived from the distribution of the trainME extremes in comparison to testME (Figure 4.6). This was evident using the mean error computation which showed the potential anomalies by the trainME versus the testME. The positive and negative anomalies were identified by the condition where the extremes (minimum and maximum) of the testME was beyond those of the trainME (Equations 46 and 4.7). Using these values as thresholds, the mean error was used to slice the test data to demarcate the dates of the onset and offset of the respective anomalies. On Figure 4.5, a positive anomaly was shown on floodplain on both polarizations. The first anomaly was positive and occurred at the beginning of the year. Evidently, it was the wettest as a result of flood since it was solely in the floodplain, confirming the deep infiltration feature identified by Pedzisai et al. (2022). On the contrary, the second anomaly was a negative (dry) and occurred in both floodplain and non-floodplain in September 2017, coinciding with the driest season. This shows that by that date, soil moisture was at the lowest for both floodplain and for non-floodplain; as semi-arid areas are characterized by moisture deficits (Humphrey et al., 2021, McColl et al., 2017a). This finding concurs with Sadeghi et al. (2020) who noted that excess input over output leads to positive anomaly.

Once the anomaly dates were extracted, the temporal duration between the onset to offset of the anomaly (lag) and time since the onset of the anomaly to the normal based on the mean of the train and predicted values (memory) (Figure 4.5) were determined. As suggested by Khorrami et al. (2022), flooding was expected to pull the testME to the minimum, hence to predominate in the floodplain while aridity or drought-related depletion of soil moisture to be evident outside the floodplain (Figures 4.5). Morales et al. (2021) found an almost 50% increase in gravimetric soil moisture resulting from rapid initial infiltration into the surface layer due to flooding in the opening and closing clays. Morales et al. (2021) noted that flooding is an important source of initial recharge on shrinking-swelling vertic clay soils. However, that result differs from silt floodplain soil in the study area that is likely influenced by different advection and diffusion processes. Furthermore, the short period of the sequence may constrain

results as noted by McColl et al. (2017a). The method used to model anomalies was based on mean backscatter, which depends on uncertain mean state as noted by McColl et al. (2017a).

4.4.2 Estimating soil moisture lag and memory using LSTMAE algorithm

The second objective sought to characterize the FRSM lag and memory temporal features. The lag was defined by the dates from the onset-to-offset of the anomaly while the memory was the period from onset to normal or half cycle between onset of first anomaly and offset of the opposite anomaly. Unlike the non-flood zone, the flood zone showed a wet flood-related lag and memory. This confirmed that flood-induced deep infiltration promotes more soil moisture recharge than from shallow infiltration through precipitation (Pedzisai et al., 2022). It further established that excessive evaporation quickly depleted precipitation induced soil moisture outside the floodplain as the study area is semi-arid. This finding confirms Seneviratne et al. (2006) and Koster and Suarez (2001) findings that soil moisture memory is influenced by evaporation rate. Whereas precipitation recharges soil moisture (Liu et al., 2011), <15% lasts for more than a few days (McColl et al., 2017a). This was true with respect to the VV rather than VH, while outside the floodplain, anomalies were not evident except for VV (Figure 4.5d).

4.4.3 LSTMAE performance evaluation

The LSTMAE model was evaluated based on the calibration and validation datasets. Therefore, we used the MAE and RMSE metrics to evaluate the LSTMAE model. Sensitivity analysis assessed number of epochs, and number of layers in the architecture. The evaluation of the model architecture to test for a dropout to check for over-fitting using the MAE also showed consistency in the model built. There is a similarity in the two model structures considering the VH in the floodplain. Therefore, the backscatter shown on the model is consistent with literature. Whereas dry soil has low dielectric constant, hence depicts lower backscatter (de Jeu et al., 2008), using C-band SAR data in arid areas, Wagner et al. (2022) noted that dry soil may also depict anomalously higher backscatter resulting from underlying materials since drier soils may allow deeper penetration of SAR pulse. Therefore, backscatter in the non-floodplains have lower MAE since changes resulting from the arid-wet phenological phases make the modelling more dynamic than in the nearly consistent floodplain. Where vegetation canopy is healthy, the vegetation canopy tends to have higher backscatter, particularly for the VH polarization, which is known to respond better than VV (Schmugge, 1983). In that case, vegetation in the floodplains may remain healthy, hence for VH, the influence of surface roughness may contribute to higher modelling error on floodplain than non-floodplain areas. This seems to support the exploratory data analysis result on the time series, and moving average since in dry

season, the VH exceeds the VV (Figure 4.8), most likely due to high backscatter from healthy vegetation and soil moisture from floods. Similarly, Tao et al. (2019) used a modified vegetation backscattering model using RADARSAT-2 C-band data to model soil moisture with MAE of 1.207m^3m^{-3} and consequent vegetation in a floodplain. Inside the floodplain, the dual polarized bands showed that the training loss is higher than validation loss on the non-floodplain while the reverse was true for VH (Figure 4.6). Dubois et al. (1995) noted that in soil moisture measurement, co-polarized data is less affected by system noise than cross-polarized. This showed that there are characteristic differences in the backscatter expression between the floodplain and non-floodplain.

Results in this study showed that LSTMAE algorithm was useful to model FRSM features with small difference in all the four training and testing samples. Since deep learning efficiently handles complex computational intelligence (Abduljabbar et al., 2021), the LSTMAE handled the FRSM anomaly detection and extract temporal feature modelling without losing both long and short-term temporal dependences. Furthermore, the observed versus predicted plots depict a more accurate relationship for training than testing in both locations for VV polarization (Figure 4.7). The dropout model defined regularization and learning rate to confirm that there was no over-fitting. Therefore, over-fitting was less likely since the algorithm was trained on 61 Sentinel-1B two years non-flood temporal images and tested on one-year with 30 image dates. Notably, the performance in LSTMAE algorithm was less probable to suffer from over-fitting due to the train-test ratio. The testing enabled the detection and evaluation of the flood event that occurred in early February 2017, with temporal soil moisture effects evaluated using the positive anomaly, lag and memory (Figure 4.3a). This finding is consistent with Luo et al. (2022) who noted that LSTM is precise and improves model performance for time series temporal dependencies. Moreso, by incorporating LSTM layers into an autoencoder architecture, Zhu and Laptev (2017) improved their prediction accuracy by nearly 40%. Therefore, combining the LSTM with the autoencoder enabled the algorithm relevant to model temporal dependences and detect anomaly, lag and memory influenced by floods.

4.4.4 Inferring FRSM using LSTMAE algorithm

Post-flood backscatter was higher in floodeplain than outside (Figure 4.8), showing that FRSM contributed to this as the study area is semi-arid. A high dielectric property of wet soil was found to be consistent with high backscatter signal (de Jeu et al., 2008). Therefore, this dielectric property can be used to infer FRSM anomalies, lag and memory features derived using LSTMAE model. The VH-polarized comparison between the observed and predicted

was poor (Figures 4.7c, 4.7d, 4.7g and 4.7h). This was probably due to the small variability in sigma nought as shown on the boxplots (Figure 4.8d). However, besides soil moisture, there are other factors that affect SAR backscatter including vegetation height and shape, soil type, surface roughness, and slope (Schmugge et al., 1980). As such, these results have uncertainty, hence need be treated with caution.

4.4.5 The utility of polarization mode to infer FRSM

This section focused on the application LSTMAE model to detect anomaly, lag and memory features based on the third objective that sought to compare the utility of cross-polarized (VH) versus co-polarized (VV) Sentinel-1 SAR C-band data to link with FRSM. The time series results (Figure 4.8) for the two polarizations showed a peculiar pattern. The VV-polarized backscatter was higher than VH-polarized within and across the floodplain. Consistent with findings by Dubois et al. (1995), the VH-polarized backscatter was more sensitive to vegetation health than the VV-polarized. The difference by location showed that the flood influenced the backscatter with positive anomaly detected in the floodplain. Ulaby and Batlivala (1976) noted that co-polarized was more sensitive to soil moisture with a value of 0.25dB/0.01g/cm^3 than cross-polarized**.** Figure 4**.**5 showed that VV polarized data identified both positive and negative anomalies within the floodplain. However, both the co-polarized and cross-polarized did not show anomalies outside the floodplain. In the floodplain, VH polarized detected only a positive anomaly (Figure 4.5). Since Moran et al. (2000) and Dubois et al. (1995) concurred that backscatter is also influenced by roughness, floodplain healthy vegetation showed higher sigma nought values on cross-polarized data. The results showed that flooding had a noticeable effect for both polarization modes inside as compared to outside the floodplain. Therefore, vegetation with adequate soil moisture within floodplain had higher backscatter in the VH-polarized band during the flood year than that outside as shown on Figure 4.8b.

The following limitations were noted in this study. Firstly, the study period was short; covering three years at a 12-day temporal resolution Sentinel-1B data. Furthermore, FRSM dynamics could occur at a shorter timeframe. Secondly, the diverse vegetation cover may have resulted in measurement noise, temperature, and roughness. Thirdly, the lack of sequential fieldwork FRSM measured data led to the utilization of Sentinel-1 SAR data to model and infer soil moisture characteristics based on the LSTMAE that utilized sequential backscatter data, with extracted values being sufficient to enable this study.

4.5 Conclusions and prospects

This study sought to understand the influence of floods on soil moisture. We therefore, modelled FRSM anomaly, lag and memory features using a hybrid LSTMAE algorithm in semi-arid Mbire District floodplain in Zimbabwe. The LSTM layers that handle the long-term dependences were embedded into the dimensionality reduction and anomaly detection autoencoder architecture using SAR data for locations inside and outside floodplain for post 2017 flood event. Using Sentinel-1 multi-temporal image values extracted inside and outside the floodplain, four polarization-based LSTMAE scenarios were run, thresholds determined to compute anomaly, lag and memory computed. In view of the three research objectives, the following results were found;

- Flood triggered both positive and negative soil moisture anomalies
- Noticeable FRSM lag and memory were detected inside the floodplain unlike outside
- Sentinel-1 VV-polarized data was useful to detect both positive (wet) and negative (dry) anomalies unlike the VH-polarized that was sensitive to only positive FRSM anomalies.

The findings showed that during post-flood in 2017, anomaly inside the floodplain was influenced by the flood, unlike outside. The influence of floods on the SAR backscatter inferred to FRSM was conspicuous for the floodplain as opposed to non-floodplain. Overall, the LSTMAE model detected anomalies inside as opposed to outside the floodplain.

Future research could compare higher spatial resolution L-band which provides more robust soil moisture characteristics with LSTMAE model results. The L-band also produces less vegetation affected backscatter, hence represent better soil moisture than C-band. Such data may become useful at local scale to study floods on floodplain soil moisture for various eco-hydrological and climatological applications.

4.6 References

Abduljabbar, R.L., Dia, H., and Tsai, P.W., 2021. Development and evaluation of bidirectional LSTM freeway traffic forecasting models using simulation data. *Scientific Reports,* 11, 23899. https://doi.org/10.1038/s41598-021-03282-z.

Adab, H., Morbidelli, R., Saltalippi, C., Moradian, M., and Ghalhari, G.A.F., 2020. Machine learning to estimate surface soil moisture from remote sensing data. *Water 12*(11), doi:10.3390/w12113223.

Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H., 2006. Greedy layer-wise training of deep networks. *Advances in Neural Information Processing Systems*, *19*.

Bengio, Y., LeCun, Y., and Hinton, G., 2021. Deep learning for AI. *Communications of the ACM*, 64(7), pp.58 - 65, DOI: 10.1145/3448250.

Bengio, Y., Simard, P., and Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Networks,* 5, pp.157 - 166.

Ben-Shahar, R., 1990. Soil nutrients distribution and moisture dynamics on upper catena in a semi-arid nature reserve. *Vegetation*, 89(1), pp.69 - 77. https://www.jstor.org/stable/20038663.

Berg, A., and Sheffield, J., 2018. Climate change and drought: the soil moisture perspective. *Current Climate Change Reports*, 4(2), pp.180 – 191, doi: 10.1007/s40641-0095-0.

Bhattiprolu, S. 2021. 180-LSTM Autoencoders for Anomaly detection. online: youtube.com/watch?v=6S2v7G-OupA&t=40s.

Brocca, L., Massari, C., Pellarin, T., Filippucci, P., Ciabatta, L., Camici, S., Kerr, Y.H., and Fernández-Prieto, D., 2020. River flow prediction in data scarce regions: soil moisture integrated satellite rainfall products outperform rain gauge observations in West Africa. *Scientific Reports* 10(1), pp.1 - 14, doi: 10.1038/s41598-020-69343-x.

Brubaker, K.L., and Entekhabi, D., 1996. Asymmetric recovery from wet versus dry soil moisture anomalies. *Journal of Applied Meteorology and Climatology* 35(1), pp.94 - 109. doi: 10.1175/1520-0450(1996)035<0094:ARFWVD>2.0.CO;2.

Chingombe, W., Pedzisai, E., Manatsa, D., Mukwada, G., and Taru, P., 2015. A participatory approach in GIS data collection for flood risk management, Muzarabani District, Zimbabwe. *Arabian Journal of Geosciences* 8(2), pp.1029 - 1040. doi: 10.1007/s12517-014-1265-6.

De Jeu, R.A., Wagner, W., Holmes, T.R.H., Dolman, A.J., van de Giesen, N.C., and Friesen, J., 2008. Global soil moisture patterns observed by space borne microwave radiometers and scatterometers. *Surveys in Geophysics* 29(4), pp.399 - 420, doi: 10.1007/s10712-008-9044-0.

Deshmukh, M.M., Elbeltagi, A., and Kouadri, S., 2022. Chapter 2: Climate change impact on groundwater resources in semi-arid regions. *Climate Change Impact on Groundwater Resources* pp.9 - 23, doi: 10.1007/978-3-031-04707-7_2.

Dintwe, R., Patrick, C., and Gaboutloeloe, G.K., 2022. Effects of common tillage practices on soil moisture retention within a crop rooting zone in a sandy loam soil, DOI: doi: 10.30574/wjaets.2022.5.2.0076.

Dolman, A.J., and de Jeu, R.A.M., 2010. Evaporation in focus. *Nature Geoscience* 3(5), pp.296 - 296, doi: 10.1038/ngeo849.

Dube, E. and Manyani, A., 2022. Indigenous Knowledge Systems for Building-back-better Flood-Impacted Communities in Zimbabwe. *Re-imagining Indigenous Knowledge and Practices in 21st Century Africa: Debunking Myths and Misconceptions for Conviviality*, pp.209-225.

Dubois, P.C., van Zyl, J., and Engman, T., 1995. Measuring soil moisture with imaging radars. *IEEE Transactions on Geoscience & Remote Sensing* 33(4), pp.915 - 926.

Duchi, J., Hazan, E., and Singer, Y., 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research* 12, 2121 - 2159.

Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T. J., Johnson, J., and Kimball, J., 2010. The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE 98*(5), pp.704 - 716. doi: 10.1109/jproc.2010.2043918.

European Space Agency, 2022. Sentinel-1 SAR Technical Guide, https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-1-sar.

Farahani, M., 2020. Anomaly detection on gas turbine time-series' data using Deep-LSTM-autoencoder. Master of Science degree in computational and engineering submitted to UMEA University.

Ghazavi, R., Vali, A., and Eslamian, S., 2010. Impact of flood spreading on infiltration rate and soil properties in an arid environment. *Water Resources Management* 24(11), pp.2781 - 2793, DOI: 10.1007/s11269-010-9579-y.

Glorot, X., Bordes, A., and Bengio. Y., 2011. Deep sparse rectifier neural networks. *Proc. 14th International Conference on Artificial Intelligence and Statistics* pp.315 - 323.

Green, J.K., Seneviratne, S.I., Berg, A.M., Findell, K.L., Hagemann, S., Lawrence, D.M., and Gentine, P., 2019. Large influence of soil moisture on long-term terrestrial carbon uptake. *Nature* 565(7740), pp.476 - 479, doi: 10.1038/s41586-018-0848-x.

Guillod, B.P., Orlowsky, B., Miralles, D.G., Teuling, A.J., and Seneviratne, S.I., 2015. Reconciling spatial and temporal soil moisture effects on afternoon rainfall. *Nature Communications* 6(1), pp.1 - 6, doi: 10.1038/ncomms7443.

Gumindoga, W., Murwira, A., Rwasoka, D.T., Jahure, F.B., and Chikwiramakomo, L., 2020. The spatio-temporal soil moisture variation along the major tributaries of Zambezi River in the Mbire District, Zimbabwe. *Journal of Hydrology: Regional Studies 32*, doi: 10.1016/j.ejrh.2020.100753.

Gxokwe, S., Dube, T., Mazvimavi, D., and Grenfell, M., 2022. Using cloud computing techniques to monitor long-term variations in ecohydrological dynamics of small

seasonally-flooded wetlands in semi-arid South Africa. *Journal of Hydrology 612*, 128080, doi: 10.1016/j.jhydrol.2022.128080.

Hinton, G.E., 2012. Tutorial on deep learning. *IPAM Graduate Summer School: Deep Learning, Feature Learning* 262.

Hochreiter, S., and Schmidhuber, J., 1997. Long short-term memory. *Neural Computation,* 9(8), pp.1735 - 1780, doi: 10.1162/neco.1997.9.8.1735.

Humphrey, V., Berg, A., Ciais, P., Gentine, P., Jung, M., Reichstein, M., Seneviratne, S.I., and Frankenberg, C., 2021. Soil moisture–atmosphere feedback dominates land carbon uptake variability. *Nature* 592(7852), pp.65 – 69, doi: 10.1038/s41586-021-03325-5.

Joussaume, S., Sadourny, R., and Jouzel, J., 1984. A general circulation model of water isotope cycles in the atmosphere. *Nature* 311(5981), pp.24-29, doi: 10.1038/311024a0.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S.I., Sheffield, J., Goulden, M.L., Bonan, G., Cescatti, A., Chen, J., De Jeu, R., and Dolman, A.J., 2010. Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature* 467(7318), pp.951 - 954, doi: 10.1038/nature09396.

Kang, J., Jin, R., Li, X., and Zhang, Y., 2021. Mapping high spatiotemporal-resolution soil moisture by upscaling sparse ground-based observations using a Bayesian linear regression method for comparison with microwave remotely sensed soil moisture products. *Remote Sensing* 13, 228, doi: 10.3390/rs13020228.

Kerr, Y.H., Waldteufel, P., Wigneron, J.P., Martinuzzi, J.A.M.J., Font, J., and Berger, M., 2001. Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE transactions on Geoscience and remote sensing 39*(8), pp.1729 - 1735, DOI**:** 10.1109/36.942551.

Khorrami, B., Fistikoglu, O., and Gunduz, O., 2022. A systematic assessment of flooding potential in a semi-arid watershed using GRACE gravity estimates and large-scale hydrological modeling. *Geocarto International* pp.1 - 22. doi: 10.1080/10106049.2022.2045365.

Kim, S., Zhang, R., Pham, H., and Sharma, A., 2019. A review of satellite-derived soil moisture and its usage for flood estimation. *Remote Sensing in Earth Systems Sciences*, 2(4), pp.225 - 246, doi:10.1007/s41976-019-00025-7.

Kingma, D.P., and Ba, J., 2014. Adam: A method for stochastic optimization. In *arXiv preprint arXiv:1412.6980*.

Koster, R.D., Mahanama, S.P., Livneh, B., Lettenmaier, D.P., and Reichle, R.H., 2010. Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow. *Nature Geoscience* 3(9), pp.613 - 616, DOI: 10.1038/NGEO944.

Koster, R.D., and Suarez, M.J., 2001. Soil moisture memory in climate models. *Journal of Hydrometeorology* 2(6), pp.558 - 570, doi: 10.1175/1525-7541(2001)002<0558:SMMICM>2.0.CO;2.

Lal, R., 2004. Soil carbon sequestration to mitigate climate change. *Geoderma* 123(1-2), pp.1 - 22, doi: 10.1016/j.geoderma.2004.01.032.

Lal, R., 2009. Sequestering carbon in soils of arid ecosystems. *Land Degradation & Development* 20(4), pp.441 - 454, DOI: 10.1002/ldr.

Lal, R., Follett, R.F., Stewart, B.A., and Kimble, J.M., 2007. Soil carbon sequestration to mitigate climate change and advance food security. *Soil Science* 172(12), pp.943 - 956.

Li, Q., Zhu, Y., Shangguan, W., Wang, X., Li, L., and Yu, F., 2022. An attention-aware LSTM model for soil moisture and soil temperature prediction. *Geoderma* 409, 115651. doi:10.1016/j.geoderma.2021.115651.

Liu, L., Gudmundsson, L., Hauser, M., Qin, D., Li, S., and Seneviratne, S.I., 2020. Soil moisture dominates dryness stress on ecosystem production globally. *Nature Communications*, 11(1), pp.1 – 9, doi:10.1038/s41467-020-18631-1.

Liu, Y., Xu, Z., Duffy, R., Chen, W., An, S., Liu, S., and Liu, F., 2011. Analyzing relationships among water uptake patterns, rootlet biomass distribution and soil water content profile in a subalpine shrubland using water isotopes. *European Journal of Soil Biology 47*(6), pp.380 - 386, doi:10.1016/j.ejsobi.2011.07.012.

Luo, X., Liu, P., Dong, Q., Zhang, Y., Xie, K., and Han, D., 2022. Exploring the role of the long short-term memory model in improving multi-step ahead reservoir inflow forecasting. *Journal of Flood Risk Management* / Early View / e12854. doi: 10.1111/jfr3.12854.

Ma, H., Zeng, J., Chen, N., Zhang, X., Cosh, M.H., and Wang, W., 2019. Satellite surface soil moisture from SMAP, SMOS, AMSR2 and ESA CCI: A comprehensive assessment using global ground-based observations. *Remote Sensing of Environment* 231, 111215.

Martínez-Fernández, J., González-Zamora, A., and Almendra-Martín, L., 2021. Soil moisture memory and soil properties: An analysis with the stored precipitation fraction. *Journal of Hydrology* 593, 125622.

McColl, K.A., Alemohammad, S.H., Akbar, R., Konings, A.G., Yueh, S., and Entekhabi, D., 2017a. The global distribution and dynamics of surface soil moisture. *Nature Geoscience* 10(2), 100 – 104, doi.: 10.1038/ngeo2868.

McColl, K.A., Wang, W., Peng, B., Akbar, R., Short Gianotti, D.J., Lu, H., Pan, M., and Entekhabi, D. 2017b**.** Global characterization of surface soil moisture drydowns. *Geophysical Research Letters* 44, pp.3682 – 3690, doi:10.1002/2017GL072819.

Mohammadi, A., Costelloe, J.F., and Ryu, D., 2017. Application of time series of remotely sensed normalized difference water, vegetation and moisture indices in characterizing flood dynamics of large-scale arid zone floodplains. *Remote Sensing of Environment 190*, pp.70 – 82, doi: 10.1016/j.rse.2016.12.003.

Morales, S. R., Lemon, M. G. T., Stewart, R. D., and Keim, R. F., 2021. Flood-induced recharge of matrix water in a vertic forest soil. *Water Resources Research* 57(7), e2021WR029675, doi:10.1029/2021WR029675.

Moran, M.S., Hymer, D.C., Qi, J., and Sano, E.E., 2000. Soil moisture evaluation using multi-temporal synthetic aperture radar (SAR) in semiarid rangeland. *Agricultural and Forest Meteorology* 105(1-3), pp.69 – 80, doi:10.1016/S0168-1923(00)00189-1.

Nketia, K.A., Asabere, S.B., Ramcharan, A., Herbold, S., Erasmi, S., and Sauer, D., 2022. Spatio-temporal mapping of soil water storage in a semi-arid landscape of northern Ghana–A multi-tasked ensemble machine-learning approach. *Geoderma* 410, 115691. doi: 10.1016/j.geoderma.2021.115691.

Okuyama, A., and Imaoka, K., 2015. Intercalibration of advanced microwave scanning radiometer-2 (AMSR2) brightness temperature. *IEEE Transactions on Geoscience and Remote Sensing 53*(8), pp.4568 - 4577, DOI: 10.1109/TGRS.2015.2402204.

Olah, C., 2015. Understanding LSTM networks. *Understanding LSTM Networks-Colah's Blog. Github*. *https://colah. github. io/posts/2015-08-Understanding-LSTMs.*

Orth, R., and Seneviratne, S. I., 2012. Analysis of soil moisture memory from observations in Europe. *Journal of Geophysical Research: Atmospheres* 117(D15), doi: 10.1029/2011JD017366.

Park, Y., and Kellis, M., 2015. Deep learning for regulatory genomics. *News and Views; Nature Biotechnology* 33 (8), pp.825 - 826.

Ranjan, C., 2020. Understanding deep learning: Application in rare event prediction. Connaissance Publishing: Atlanta, GA, USA.

Rezaei, A.R., Ismail, Z.B., Niksokhan, M.H., Ramli, A.H., Sidek, L.M., and Dayarian, M.A., 2019. Investigating the effective factors influencing surface runoff generation in urban catchments–A review. *Desalination Water Treatment* 164, pp.276 - 292. doi: 10.5004/dwt.2019.24359.

Robinson, D.A., Campbell, C.S., Hopmans, J.W., Hornbuckle, B.K., Jones, S.B., Knight, R., Ogden, F., Selker, J., and Wendroth, O., 2008. Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review. *Vadose Zone Journal* 7(1), pp.358 - 389, doi:10.2136/vzj2007.0143.

Sadeghi, M., Gao, L., Ebtehaj, A., Wigneron, J-P., Crow, W.T., Reager, J.T., and Warrick, A. W., 2020. Retrieving global surface soil moisture from GRACE satellite gravity data. *Journal of Hydrology* 584, 124717, doi: 10.1016/j.jhydrol.2020.124717.

Schmugge, T.J., 1983. Remote sensing of soil moisture: Recent advances. *IEEE Transactions on Geoscience & Remote Sensing* (3), pp.336 – 344, DOI**:** 10.1109/TGRS.1983.350563.

Schmugge, T.J., Jackson, T.J., and McKim, H.L., 1980. Survey of methods for soil moisture determination. *Water Resources Research* 16(6), pp.961 - 979. doi: 10.1029/WR016i006p00961.

Seneviratne, S.I., 2012. Historical drought trends revisited. *Nature*, 491(7424), pp.338 – 339, doi.: 10.1038/491338a.

Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B., and Teuling, A.J., 2010. Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews* 99(3 - 4), pp.125 - 161. doi: 10.1016/j.earscirev.2010.02.004.

Seneviratne, S.I., Koster, R.D., Guo, Z., Dirmeyer, P.A., Kowalczyk, E., Lawrence, D., Liu, P., Mocko, D., Lu, C.H., Oleson, K.W., and Verseghy, D., 2006. Soil moisture memory in AGCM simulations: analysis of global land–atmosphere coupling experiment (GLACE) data. *Journal of Hydrometeorology*, 7(5), pp.1090 – 1112, doi: 10.1175/JHM533.1.

Sims, N.C., and Colloff, M.J., 2012. Remote sensing of vegetation responses to flooding of a semi-arid floodplain: Implications for monitoring ecological effects of environmental flows. *Ecological Indicators* 18, pp.387 - 391, doi: 10.1016/j.ecolind.2011.12.007.

Srivastava, P.K., Pandey, V., Suman, S., Gupta, M., and Islam, T., 2016. Available data sets and satellites for terrestrial soil moisture estimation. *Satellite Soil Moisture Retrieval* pp.29 – 44, doi.: 10.1016/B978-0-12-803388-3.00002-4.

Sun, H., and Cui, Y., 2021. Evaluating downscaling factors of microwave satellite soil moisture based on machine learning method. *Remote Sensing* 13(1), 133. doi: 10.3390/rs13010133.

Sungmin, O., and Orth, R., 2021. Global soil moisture data derived through machine learning trained with in-situ measurements. *Scientific Data* 8(170), pp.1 - 14. doi: 10.1038/s41597-021-00964-1.

Sungmin, O., Hou, X., and Orth, R., 2020. Observational evidence of wildfire-promoting soil moisture anomalies. *Scientific Reports* 10(1), pp.1 – 8, doi:10.1038/s41598-020-67530-4.

Tao, L., Wang, G., Chen, W., Chen, X., Li, J., and Cai, Q., 2019. Soil moisture retrieval from SAR and optical data using a combined model. *Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12(2), pp.637 – 647. DOI:10.1109/JSTARS.2019.2891583.

Tian, S., Renzullo, L.J., van Dijk, A.I., Tregoning, P., and Walker, J.P., 2019. Global joint assimilation of GRACE and SMOS for improved estimation of root-zone soil moisture and vegetation response. *Hydrology and Earth System Sciences* 23(2), pp.1067 - 1081. doi: 10.5194/hess-23-1067-2019.

Tieleman, T., and Hinton, G., 2012. Lecture 6.5-rmsprop, Coursera: Neural networks for machine learning. *Technical Report* 6. University of Toronto.

Ulaby, F.T., and Batlivala, P.P., 1976. Optimum radar parameters for mapping soil moisture. *IEEE Transactions on Geoscience Electronics* 14(2), pp.81 - 93. DOI:10.1109/TGE.1976.294414.

Wagner, W., Lindorfer, R., Melzer, T., Hahn, S., Bauer-Marschallinger, B., Morrison, K., Calvet, J.C., Hobbs, S., Quast, R., Greimeister-Pfeil, I., and Vreugdenhil, M., 2022. Widespread occurrence of anomalous C-band backscatter signals in arid environments caused by subsurface scattering. *Remote Sensing of Environment* 276, 113025. doi: 10.1016/j.rse.2022.113025.

Wajih, S.A., 2008. Adaptive agriculture in flood affected areas. *LEISA Magazine* 24(4), pp.24 - 25.

Wasko, C., and Nathan, R., 2019. Influence of changes in rainfall and soil moisture on trends in flooding. *Journal of Hydrology* 575, pp.432 - 441.

Western, A.W., Grayson, R.B., and Blöschl, G., 2002. Scaling of soil moisture: A hydrologic perspective. *Annual Review of Earth & Planetary Sciences* 30(1), pp.149 - 180. doi: 10.1146/30.091201.140434.

World Food Programme, 2016. Zimbabwe monthly food security monitoring Report. https://documents.wfp.org/stellent/groups/public/documents/ena/wfp287989.pdf?_ga=2.235987458.2087756766.1571042887-108065798.1571042887.

Yang, L., Sun, G., Zhi, L., and Zhao, J., 2018. Negative soil moisture-precipitation feedback in dry and wet regions. *Scientific Reports* 8(1), pp.1 - 9. DOI:10.1038/s41598-018-22394-7

Zhang, D., and Zhou, G., 2016. Estimation of soil moisture from optical and thermal remote sensing: A review. *Sensors* 16, 1308, doi:10.3390/s16081308.

Zhu, L., and Laptev, L., 2017. Deep and confident prediction for time series at Uber. arXirv:1709.01907v1 [stat.ML] 6 Sep 2017.

Zhuo, W., Huang, J., Li, L., Zhang, X., Ma, H., Gao, X., Huang, H., Xu, B., and Xiao, X., 2019. Assimilating soil moisture retrieved from Sentinel-1 and Sentinel-2 data into WOFOST model to improve winter wheat yield estimation. *Remote Sensing* 11(13), 1618. doi:10.3390/rs11131618.

Zimbabwe National Statistics Agency, 2022. 2022 Population and Housing Census. Preliminary Report on Population Figures. Harare, Zimstat. https://zimbabwe.unfpa.org/.

CHAPTER 5: SHORT-TERM INFLUENCE OF FLOOD-RECHARGED SOIL MOISTURE ON LEAF CHLOROPHYLL CONTENT OF *ZIZIPHUS MAURITIANA* TREE IN SEMI-ARID LANDS

This chapter is based on: **Pedzisai**, **E.,** John Odindi, J. and Mutanga, O. Modelling the influence of flood-recharged soil moisture on leaf chlorophyll content of *Ziziphus mauritiana* tree in semi-arid floodplains of northern Zimbabwe. *Under preparation*.

Abstract

Trees mitigate climate change impacts, contribute to terrestrial hydrology, and provide ecosystem goods and services that benefit society and wildlife. In semi-arid landscapes, floods input soil moisture that support living trees. However, the seasonal influence of flood-recharged soil moisture (FRSM) on leaf chlorophyll content (LCC), an important photosynthetic pigment that determine tree growth and biomass productivity is not adequately understood. In this study, measurements of chlorophyll from Indian jujube (*Ziziphus mauritiana*) tree leaves, locally known as Musawu, a multipurpose tree with significant economic, cultural, environmental, and hydrologic benefits within and out of the floodplains in northern Zimbabwe. The study augmented LCC measurements with statistical derivatives based on aspect on canopy (AOC), exposure to sunlight (ETL), and height on canopy (HOC) and either flooded or non-flooded tree location spatial data. Firstly, the study performed Binary Logistic Regression of floodplain class on spatial, and LCC variables. Secondly, the study selected important predictors using Boruta variable importance projection (BVIP) and Random Forest model, extracting classification metrics, Gini, and Permutation indices. Thirdly, the study regressed LCC on spatial variables, and finally, evaluated models using two-way ANOVA on three variables. The binary logistic regression indicated statistical insignificance for all predictors, while the BVIP confirmed 16 as important and rejected 43. Two remained tentative. The Random Forest classification showed high accuracy (accuracy>90%; kappa>0.89). The LCC inside the floodplain was significantly higher than outside. This was confirmed using two-way ANOVA of linear regression with spatial, and selected variables ($p<0.001$), ETL ($p<0.01$), and HOC ($p<0.1$). Spatial variables alone were statistically significant ($p<0.01$), to show that FRSM inside the floodplain influenced higher LCC than outside. Also, the ANOVA showed significant LCC differences for leaves in-shade than sunlit. Therefore, FRSM positively influences LCC for trees inside the floodplain of semi-arid landscapes.

Keywords

Aspect on canopy; exposure to sunlight; floodplain; height on canopy; leaf chlorophyll content; machine learning; modelling; variables

5.1 Introduction

Trees drive life-supporting atmosphere-biosphere systems (Boonman et al., 2024; Gumbs et al., 2024; Mori et al., 2021; Gamfeldt et al., 2013). In semi-arid areas, floods contribute to soil moisture recharge (Pedzisai et al., 2022), supporting trees that provide myriad socio-ecological

and environmental benefits. Trees are characterized by green biomass that sustain fauna (Jiang et al., 2020; Bonan, 2008), and help circulate soil microbes (Anthony et al., 2024). As such, trees contribute the largest proportion of global carbon budget (Crowther et al., 2015), important for climate change mitigation and support ~80% of terrestrial biodiversity (Global Forest Watch, 2002). Meanwhile, carbon accumulation is directly proportional to tree size (Stephenson et al., 2014). Trees are also recognized as a biotic 'pump' that promotes localized precipitation (Makariev and Gorshkov, 2007) and trees inside and outside forests help maintain ecological stability (Wang et al., 2025). Generally, trees provide ecosystem goods, functions and services important for human existence and livelihoods (Steel et al., 2024; Zhang, Liu et al., 2024; Zhang, Xiong, et al., 2024;). Examples include provision of goods (food, medicines, cultural, economic, recreational, genetic biodiversity resources, raw materials, timber, and ecological habitat to organisms), functions (water storage, habitat, biogeochemical cycling), and services (water, and food production, regulating climate and natural fluxes) (Boonman et al., 2024; Neugarten et al., 2024; Reiner et al., 2023; Brandt et al., 2018; Vieilledent et al., 2018; Betts et al., 2017; Hansen et al., 2013; Sandel and Svenning, 2013). According to Costanza et al., (1997), trees contribute an estimated average global value of US$33 trillion per year, while Li, Wang et al. (2024) notes that in Tonglu County, China, a gross ecosystem product value contributed CYN 30 billion (~US4.21 billion), ~50% in 2021. Such gains justify and enable prioritisation of ecosystem protection, reasonable utilisation, and preservation (Zhang, Rong, et al., 2024).

Semi-arid landscapes are characterized by trees that cover nearly 2.4 billion hectares (Rotenberg and Yakir, 2010), of which ~21.4% is in Africa (FAO, 2020). However, approximately one third of trees exist outside forests, which makes the above statistics an underestimation at canopy scale (Reiner et al. 2023). This implies that, collectively, all trees store a larger proportion of the global terrestrial carbon than officially recorded. Trees store carbon in form of live and dead biomass both above- and belowground (Mo et al., 2023; Yang, Ciais et al., 2023; Harris et al., 2021; Spawn et al., 2020; Betts et al., 2017; Bonan, 2008). Specifically, tree foliage contain chlorophyll which is a biochemical and physiological indicator (Xu, Fu et al., 2024) that sequesters carbon during photosynthesis. Individual leaves with chlorophyll play a pivotal carbon sequestration function (Soysal, 2024). However, for carbon accounting, deforestation assessment, and biomass estimation, studies commonly focus on either forest (Duncanson et al., 2023; Tucker et al., 2023; Green and Keenan, 2022; Jin et al., 2020; Holtmann et al., 2019) or single tree (Scogings, 2023; Stephenson et al., 2014) than

on leaf (Dusenge et al., 2019; Wang et al., 2017). Generally, in semi-arid landscapes, studies that consider species are scarce, let alone those at leaf scale.

Remote sensing offers several advantages over other methods for productivity-based chlorophyll content assessments (Chaurasia et al., 2024). Generally, studies are based on forest scale using 250 m MODIS (Liu et al., 2024; Qin et al., 2019; Yin et al., 2017; Li, Sulla-Menashe, et al., 2017; Qin et al., 2016), and 30 m Landsat (Li, Guo et al., 2023; Ma et al., 2023; Hansen et al., 2022), tree scale 10 m Sentinel (Verhegghen et al., 2022; Waldeland et al., 2022), 3 m PlanetScope (Reiner et al., 2023) and sub-canopy scale 61cm QuickBird (Liu, Liu et al., 2024). Generally, such image data lack sufficient spatial detail for leaf and chlorophyll content analysis. Whereas remotely sensed data can be used to derive spectral indices to study chlorophyll content, the outputs can be affected by factors that include spatial characteristics and species type (Inoue et al., 2016). Thus, efforts to determine chlorophyll content remains paramount for leaf scale studies.

Leaf chlorophyll is regarded as an antenna pigment (Herritt et al., 2020), which is central to photosynthesis that sequesters carbon dioxide, hence promoting the necessary environmental cooling effect (Rotenberg and Yakir, 2010). Studies have assessed chlorophyll content from global (Xu, He et al., 2024; Li et al., 2023; Ma et al., 2023; Mo et al., 2023; Scogings, 2023), continental (Anthony et al., 2024; Reiner et al., 2023; Verhegghen et al., 2022), national (Mugabowindekwe et al., 2023, Wagner et al., 2023), forest (Green and Keenan, 2022; Jiang et al., 2020; Doughty et al., 2015) to canopy (Xu, Fu et al., 2024) levels to account for carbon, which informs climate change studies. Therefore, understanding chlorophyll content is essential, especially in view of the anticipated climate change impacts.

Leaves contain chlorophyll, which is used to convert water and carbon dioxide into sugars and oxygen through photosynthesis (Zhang, Huang et al., 2024). As such, solar induced chlorophyll fluorescence measurements have increasingly become useful to determine LCC (Sun et al., 2017; Maxwell and Johnson, 2000). While sunlight availability and quantity differ spatially, exposure to sunlight and shade conditions influence LCC (Lo Piccolo et al., 2023).

Field leaf chlorophyll content can be derived using either destructive sampling followed by laboratory analysis or instantaneous non-destructive sampling (Ciganda et al., 2009). Specifically, in destructive sampling, samples are collected by plucking leaves off the tree for chlorophyll content measurement. Destructive samples must be stored correctly and be assessed within an acceptable time window (e.g., Iqbal et al., 2015).. Alternatively, LCC can

be non-destructively sampled using field-based chlorophyll meter and multispectral imaging (Li, Cui et al., 2023; Pan et al., 2015). Non-destructive field-based measurements are particularly essential for model calibration and validation in both classification and regression algorithm development and deployment (Angel and McCabe, 2022). Hence, to augment field data for the purpose of assessing tree LCC inside against outside a floodplain so as to evaluate effect of FRSM, field measurements are necessary. Once collected, the quantitative data require robust analyses to generate information.

Improvements in computing resources in concert with advent of robust artificial intelligence and machine learning (Yamashita et al., 2020; Shah et al., 2019; Li et al., 2018) has improved the study of variables in empirical studies. Specifically, feature selection is valuable in understanding important variables that can help explain interactions of dependent with independent variables. Decision tree-based machine learning e.g., Random Forests (Breiman, 2001), and Boruta (Kursa and Rudnicki, 2022) can perform variable importance projection, efficiently discriminating between important and unimportant variables. These machine learning methods can also handle big data (Adadi, 2021; Emmauel et al., 2021; Rathore, et al., 2021). As such, deriving statistical features from field measurements in addition to GIS based data derivation is promising to semi-arid floodplain tree LCC studies. The machine learning algorithms are robust for modelling phenomena using both regression and classification. However, feature selection using machine learning methods remains a challenge, hence, approaches such as neural networks, parameterized with regularisation schemes, train-test data partitioning, drop-out, and learning rate are adopted to reduce model overfitting (Bejani and Ghatee, 2021; Yamada et al., 2020; Lengerich et al., 2020).

Therefore, using machine learning modelling, this study analyzed field LCC within and outside a floodplain. Specifically, the study adopted spatial data augmented by statistically derived LCC based on tree structural canopy factors to model the influence of flooding on LCC. With a specific focus on musawu or jujube (*Z. mauritiana)* tree in Mbire and Muzarabani semi-arid floodplain, the study sought to answer the following questions;

1. Which variables are important in predicting LCC?
2. How does floodplain LCC compare with that on the outside location?
3. What is the value of statistical data augmentation in modelling LCC?

5.2 Materials and Methods

5.2.1 Study area

This study was carried out in- and out-side floodplains in the Mbire and Muzarabani districts in Mashonaland Central, northern Zimbabwe (Figure 5.1). The study area is characterized by semi-arid condition with both drought and floods (Mavhura et al., 2022; Chanza and Musakwa, 2021; Mudavanhu et al., 2020). Summers are hot and wet, while winters are dry (Mavhura et al., 2022). The area is highly drained by the Hunyani and Musengezi rivers that receive discharge from several tributaries. Floods are common due to frequent seasonal tropical cyclones and backflow from Cahora Bassa Dam downstream (Mavhura, 2019). The generally low elevation exacerbates the floods experienced in this area. The area comprises of depositional silts, with deciduous trees, bushes and grasses, while human activities include rural settlements, rain-fed crop and animal agriculture (Mavhura et al., 2022), though poverty-stricken (Mudavanhu et al., 2020).

5.2.2 Data

The study utilized primary data including altitude, locational tree co-ordinates, floodplain class and LCC. The LCC measurements were augmented by deriving descriptive statistical values. A total of 119 chlorophyll tree samples constituting 29 outside and 90 inside the floodplain were collected. The outside floodplain samples were the control while those inside the floodplain constituted the variables for evaluation. Table 5.1 outlines the data collection method, data, and variables.

Table 5.1: Data collection method, data, variable(s) and description of the data (n = 119)

Data source	Data	Variable	Description of variables	Data type
Field	Chlorophyll	18	3 by AOC, 2 by ETL, and 3 by HOC	Numeric
Field	Spatial	01	Altitude	Numeric
Field + GIS	Location	01	2 classes: Floodplain (1), Non-floodplain (0)	Binary
Field + GIS	Coordinates	02	distance to river, and distance to floodplain	Numeric
Derived	Chlorophyll	40	4 statistical derivatives by AOC, HOC, and ETL	Numeric
Total		**62**		

AOC - Aspect on canopy with north, east, south and west directions, ETL - Exposure to light with binary: i.e., in-Light and in-Shade; HOC - Height on canopy with bottom, middle and top levels.

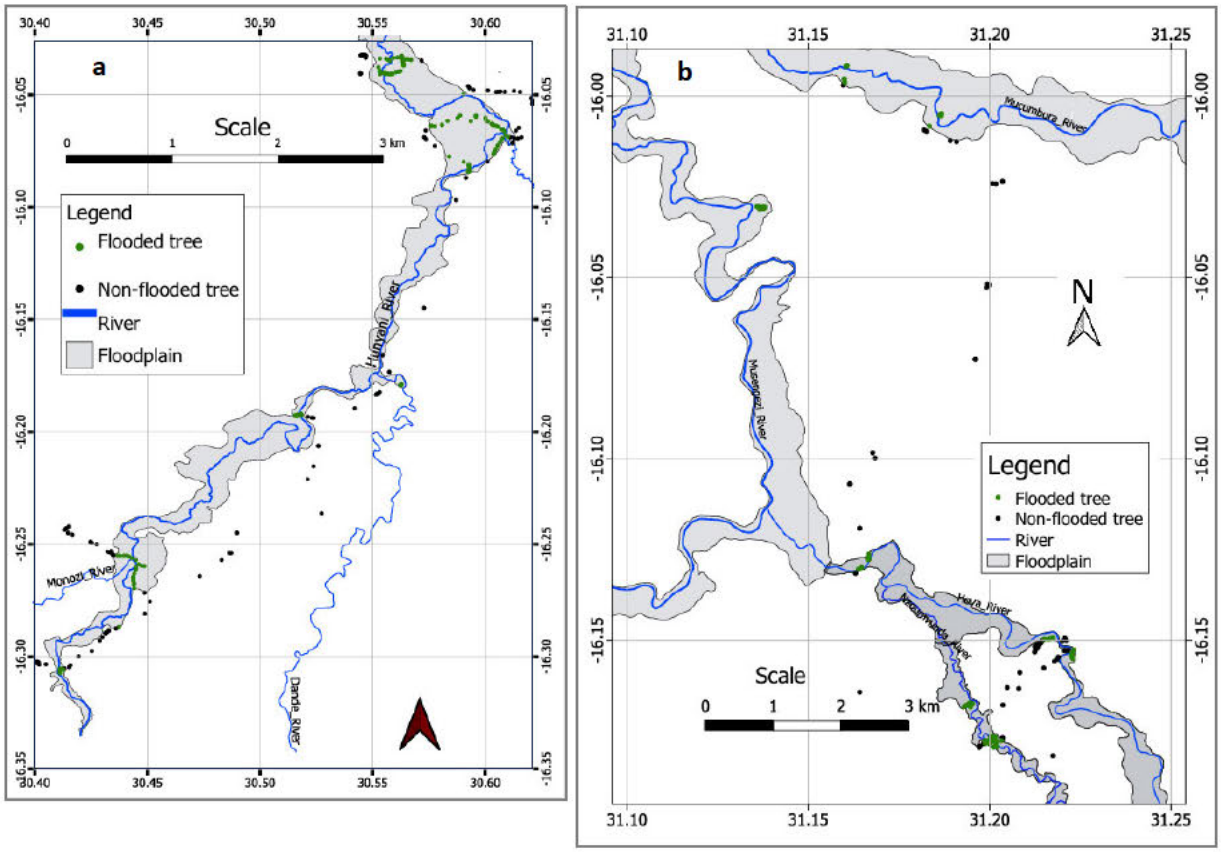


Figure 5.1: Mbire (a), and Muzarabani (b) study sites showing tree sample locations inside and outside the floodplain with Hunyani and Musengezi river networks.

5.2.3 Data Collection

A set-up that discriminated LCC inside (experiment) from outside (control) floodplain was essential. Community-guided fieldwork facilitated transect across the each recently flooded area, also taking field observations, and GPS locational coordinates were undertaken in the study area from 2019/10/31 to 2019/11/06, where LCC data were collected using a handheld SPAD 502 (Konika Minolta, Japan) leaf chlorophyll meter while a hand-held Garmin eTrex 12-channel GPS (Garmin Ltd, USA) was used to collect both flooded and non-flooded multipurpose *Ziziphus mauritiana* tree locations. In some cases, destructive leaf sampling was necessary since tree height constrained in-situ measurements. Statistically derived (maximum, mean, minimum, and standard deviation) variables per tree were computed. A total of 40 statistically derived variables based on height on canopy, aspect on canopy and exposure to light augmented field LCC samples (Table 5.1). Mature leaves are often preferred for sampling (e.g., Lo Piccolo et al., 2023) because young leaves contain half the amount of Chlorophyll a+b per unit area than mature ones (Bertamini and Nedunchezhian, 2003). Therefore, we sampled mature leaves before they started to senesce (Li, Xu et al., 2025) from trees located both inside and outside floodplain. The sampling protocol targeted 18 mature leaves per tree taken from

four campus directions (aspect on canopy) at three height levels (height on canopy) and in pairs of in-shadow and on sunlit (exposure to light) locations to measure LCC features for 119 trees.

Detailed LCC measurements were defined firstly by aspect on canopy using four cardinal compass directions (East, South, West, and North), exposure to light (in-Light and in-Shade) and height on canopy (Bottom, Middle and Top). A leaf sample was selected based on absence of signs of damage, independent of pest attack and no sign of disease (Xu, Fu et al., 2023). Every LCC measurement, except top, was coded with three letters of the alphabet. The first code represented height on canopy with either Bottom (B) or Middle (M). The second code was for aspect on canopy with East (E), South (S), West (W) and North (N) as demonstrated by Xu, Fu et al. (2023). The third code was for exposure to light with either in-Light (L) or in-Shade (S). The top pair of leaf samples dropped the middle aspect on canopy code, hence either top in-Light (TL) or top in-Shade (TS). For instance, on a LCC sample coded BEL, B stands for bottom on the height on canopy, E for east-side aspect on canopy and L for a leaf sampled on sunlit exposure to light. Based on one aspect on canopy e.g., east, two leaves at the bottom; one on sunlit and one in-shade were collected. Using a chlorophyll meter, the LCC was then instantaneously measured three times and average noted as suggested by Xu, Fu et al. (2023) to avoid alterations of the pigmentation. The same procedure was repeated at the middle on height on canopy but for the same aspect on canopy. The leaf measurement of LCC was replicated for the remaining three aspect on canopy attributes. Finally, one pair with in-Light (TL) and top in-Shade (TS) LCC were collected. Thus, a total of 18 LCC measurements were taken per tree for the entire 119 sample of trees. The fieldwork identified tree locations as either inside or outside the floodplain with a binary class variable delimitation. Appendix 5.1 shows all the 65 (including the three shadow) variables modelled using Boruta Variable Importance Projection (BVIP).

5.2.4 Preprocessing

The LCC measurements were statistically aggregated to derive additional LCC augmentation as defined in Table 5.1. Specifically, the maximum (max), mean, minimum (min) and standard deviation (stdev) by aspect on canopy, height on canopy, and exposure to light derived 40 variables. Altitude was recorded using a global positioning systems device, while coordinates were overlaid on a floodplain polygon and river centreline multipoint vector layers to extract distance-to-floodplain (D2FP) for outside floodplain samples and distance-to-river (D2R), respectively, for flood- and non-floodplain locations. Statistically aggregated chlorophyll content for east-facing leaf (aspect on canopy) samples gave EChlMax, EChlMean, EChlMin,

EChlStdev data. The same was applied for the remaining three (South, West, North) compass directions. All sunlit leaf samples were aggregated using the same descriptive statistics, hence LEccLmax, LEccLmean, LEccLmin, and LEccLstdev. The same applied to in-shade samples. The 18-leaves per tree chlorophyll aggregated samples also had four variables (GChl) hence 65, including three Boruta-generated shadow (min, mean and max) variables for machine learning in R.

5.2.5 Modelling

Modelling was executed using a code (Appendix 5.2) in R (4.3.3). R-packages used include Boruta (Kursa and Rudnicki, 2010), caret for model prediction (Kuhn et al., 2023, Kuhn, 2008), Random Forest for ranking the variable importance projection and classification (Breiman, 2001), DescTools (Signorell et al., 2024), caTools, manipulate for binary logistic regression analysis (Allaire, 2014), GGally for scatterplot matrix, and ggplot2 for graphics plotting (Schloerke et al., 2024). Spatial modelling of LCC applied machine learning which involved binary logistic regression, BVIP, Random Forest classification, and linear regression model evaluation using ANOVA (Figure 5.2).

5.2.5.1 Binary Logistic Regression machine learning

The study held a linear regression assumption. Therefore, the study built a binary logistic regression machine learning model on floodplain class on 61 variables by calling a glm () function from 'glm Library' within R. The study used 58 chlorophyll, plus three flood-related (D2FP, D2R and altitude) explanatory variables, to fit the binary logistic linear regression model within 'manipulate Library'. Further hypotheses testing was necessary to safeguard against either Type I error (Forstmeier and Schielzeth 2011; Anderson et al., 2000) or the Hauck-Donner Effect that affects the Wald test (Yee, 2024; 2022). Since binary linear logistic regression showed no statistical significance, selecting important variables became imminent, thus, the study used Boruta, a decision-tree algorithm (Kursa and Rudnicki, 2010).

5.2.5.2 Boruta function variable importance projection

Boruta algorithm is a wrapper technique capable of selecting important explanatory variables and present them in order of importance, useful when there are too many variables (Kursa and Rudnicki, 2010). The BVIP adds randomness by creating shadow features which are shuffled copies of all features (Kursa and Rudnicki, 2010). The algorithm trains the model on Random Forest classifier and uses mean decrease in accuracy (Breiman, 2001). Every computation checked if real feature has higher importance (higher z-value) than the best (maximum z-value)

of its shadow features, while also removing confirmed unimportant ones (Breiman, 2001). The algorithm was executed in R-Project using Boruta Library. The study parameterized the response variable (binary floodplain class), and 61 explanatory variables with '*doTrace = 2*' and '*maxRuns = 1000*', *holdHistory* = TRUE. The BVIP has three outcomes namely, 1) confirmed (important), 2) confirmed (unimportant), and 3) unconfirmed (tentative). The third outcome may require increasing the iterations until all variables are determined. However, this also implies that weak predictors are being forcedly determined. The BVIP screens bad predictors from important ones (Bain et al., 2024; Lyu et al., 2024), hence, the BVIP confirmed 16 (important), and rejected 43 (unimportant) while 2 remained tentative at the 1000th iteration.

5.2.5.3 Random Forest machine learning classification

The Random Forest is a robust type of ensemble machine learning classification and regression model based on bootstrap aggregation or bagging (Breiman, 2001). A 70/30 train/test data split ratio was performed using dplyr and CaTools libraries**.** The study converted the target variable class to binary factor level and optimized the Random Forest model to number of runs (mtry = 1000), number of trees (ntree = 4), stepFactor = 1.2, improve = 0.01, trace = T, and plot = T. The study trained the Random Forest model and tested it on independent data, noting accuracy, kappa, sensitivity, and specificity (Equations 5.1 - 5.4);

$$\text{Accuracy} = (TP+TN)/(TP+FP+TN+FN) \quad (5.1)$$

$$\text{Kappa} = (P_o - P_e)/(N - P_e) \quad (5.2)$$

$$\text{Sensitivity} = TP/(TP+FN) \quad (5.3)$$

$$\text{Specificity} = TN/(TN+FP) \quad (5.4)$$

were: P_o is observed and P_e is the expected accuracy. and N is TP+TN+FP+FN.

The model performance metrics automatically computed in the algorithm include out-of-bag (OOB) error, and confusion matrix derived model performance metrics namely accuracy, kappa, a 95% confidence interval, no information rate (NIR), p-Value (i.e., accuracy>NIR), kappa, McNemar's test p-Value, sensitivity, specificity, positive predictive value, negative predictive value, prevalence, detection rate, detection prevalence, and balanced accuracy during both model training and testing (Breiman, 2001).

5.2.5.4 Linear Regression Modelling

The study regressed the LCC maximum per tree on 16 BVIP explanatory variables comprising field and statistically aggregated chlorophyll (maximum, mean, minimum, and standard deviation) on altitude, distance (D2R, D2FP), and floodplain class using Gaussian method.

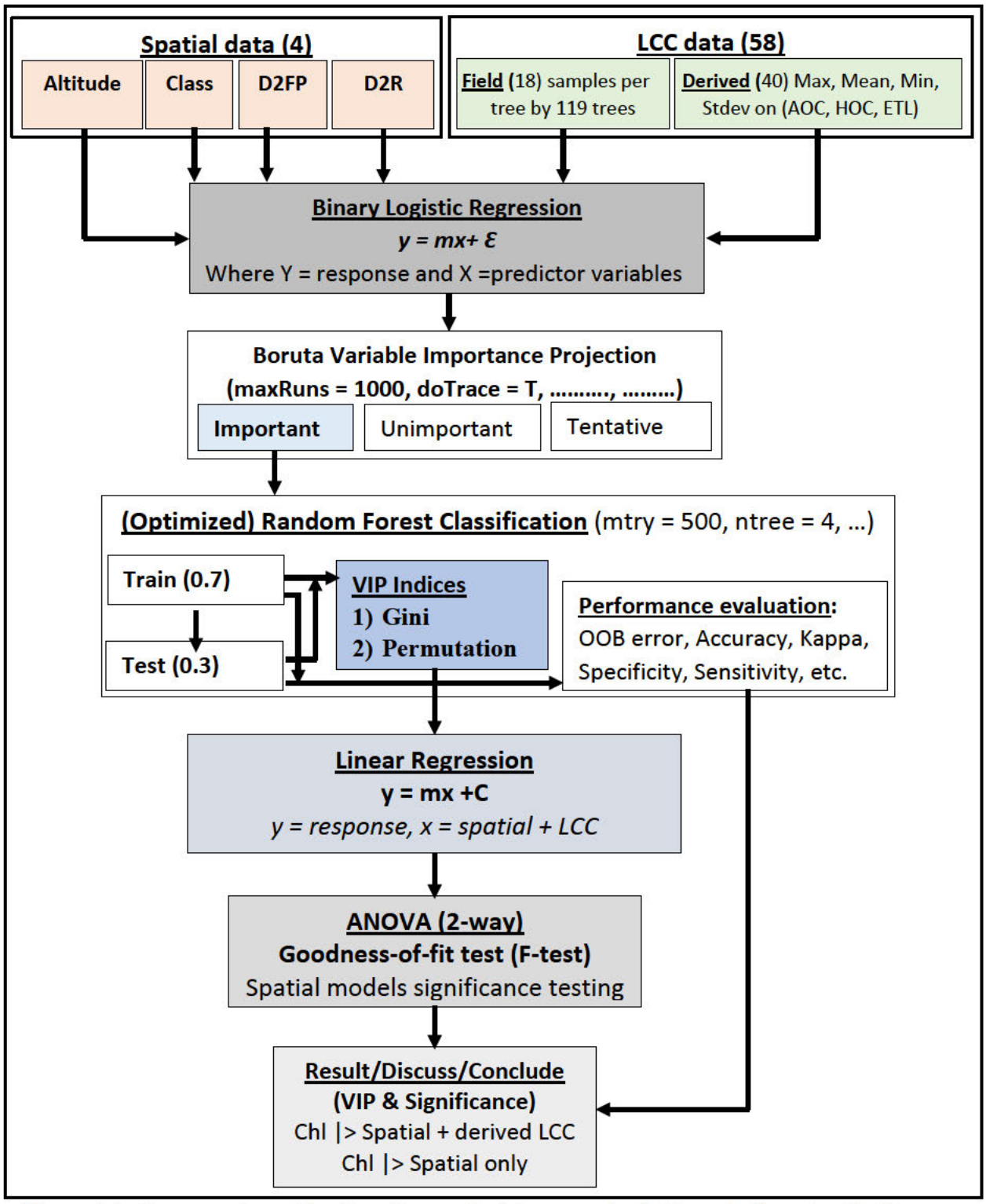


Figure 5.2: Sequenced machine learning method to assess leaf chlorophyll content in R.

5.2.5.5 Variability of LCC using ANOVA

The variability of LCC was based on three clusters of variables namely HOC, AOC and ETL as defined in Section 5.2.3. The variability was tested based on ANOVA from the linear regression model generated after selecting the confirmed important variables using firstly the

BVIP and secondly Random Forest machine learning classification. Firstly, a linear model based on all the bottom leaves was compared to the one from middle height and ultimately the one on top level leaves. Secondly, variability based on linear regression model of AOC based on East canopy side LCC was compared with the south, west and north, respectively. Four compus directions were considered namely east, south, west and north**.** Thirdly, linear model of LCC based on ETL compared the effect of sunlight or shade to LCC. Leaves not exposed to sunlight were compared to those in the shade.

5.2.6 Tree aggregated LCC against spatial variables correlation matrix

The study run a pairwise correlation matrix in GGAlly Library (2.2.1) using the ggpairs() function (Schloerke et al., 2024) to test the four-tree canopy chlorophyll against four spatial variables. The correlation used tree canopy LCC maximum (GChlMax), mean (GChlMean), minimum (GChlMin) and standard deviation (GChlStdev). This pairwise correlation matrices between tree chlorophyll mean with D2FP, D2R, and the binary floodplain class were computed to test the spatial FRSM related variables.

5.2.7 Model performance evaluation metrics

Model performance evaluation metrics were based on Random Forest machine learning and linear regression computed using likelihood ratio test based on ANOVA test statistic. The Random Forest machine learning metrics derived variable importance was depicted using Gini and Permutation indices. Factor-based linear models were evaluated using two-way ANOVA to assess their variability. Specifically, to evaluate the noted differences between sun and shade leaves, likelihood ratio test based on ANOVA, and p-Values for statistical significance of the explanatory variables were also determined (e.g., Lo Piccolo et al., 2023).

5.3 Results

5.3.1 Explanatory variables of LCC using floodplain binary class

The binary logistic regression output did not find any variable significant. Instead, it gave two warning messages; firstly that glm.fit: algorithm did not converge and secondly, glm.fit: fitted probabilities numerically 0 or 1 occurred. The second warning is also true for perfectly separated classes. Since all explanatory variables were concluded not significant, further hypotheses testing became necessary to avoid Type I error.

5.3.2 Variable importance projection for LCC modelling

Where multiple predictor variables were involved, it was important to determine their relative magnitude of influence. Decision trees specifically derived using the Boruta model was

efficient. Therefore, variable importance projection was undertaken in Boruta package executed in R. The model confirmed 16 variables, shown as green lines on the history plot on Figure 5.3. The BVIP output including three shadow metrics is shown using boxplots (Figure 5.4). Then, the Random Forest machine learning model derived the order of importance depicted using a bar graph presented on Figure 5.5.

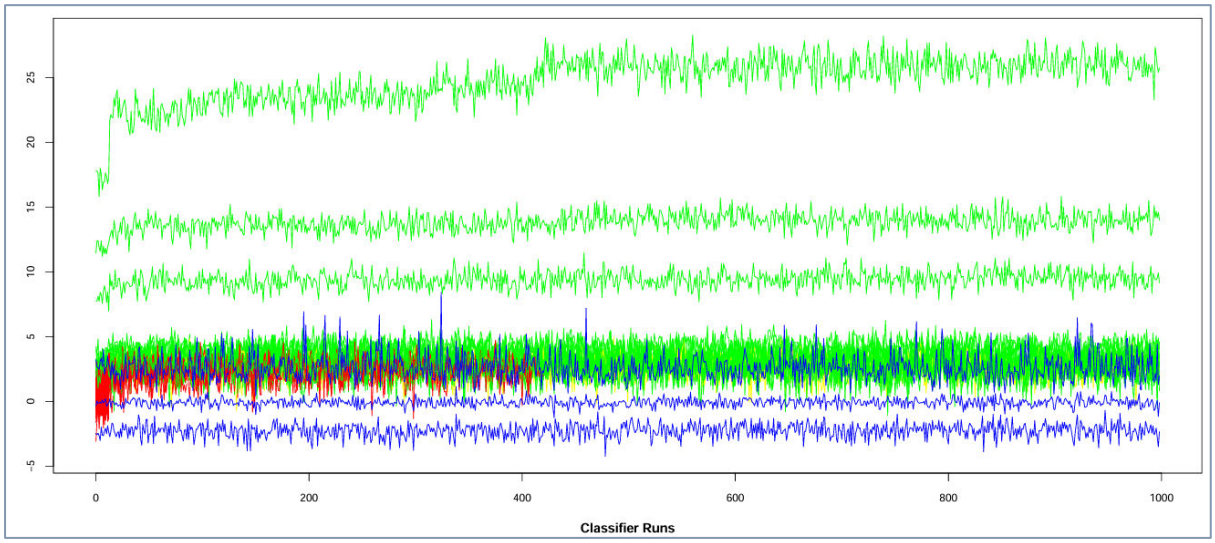


Figure 5.3: The Boruta variable importance projection (BVIP) history plot showing the respective confirmed (green) rejected (red), and tentative (yellow) while the three shadow statistical derivatives (minimum, mean and maximum) metrics are shown as blue lines.

The explanatory variables in order of importance using the Gini Impurity Index are shown on Figure 5.5. For both indices, D2FP, D2R and altitude (last green boxplots) were the most important explanatory variables of LCC. Spatial variables within GIS modelling were important to determine influence of floods on natural systems and the environment. For instance, altitude and distance to rivers were important explanatory variables in characterizing flood susceptibility using five machine learning models (Janizadeh et al., 2019). The D2FP, though rarely characterized, is crucial for non-flooded locations (D2FP > 0), which becomes important control measure useful to discriminate LCC outside against that inside the floodplain (D2FP = 0). Six variables were statistically significant, though to different levels of confidence. The LCC derived one AOC, two ETL, one TH, two distance based and elevation were statistically significant to model the tree LCC maximum measurement.

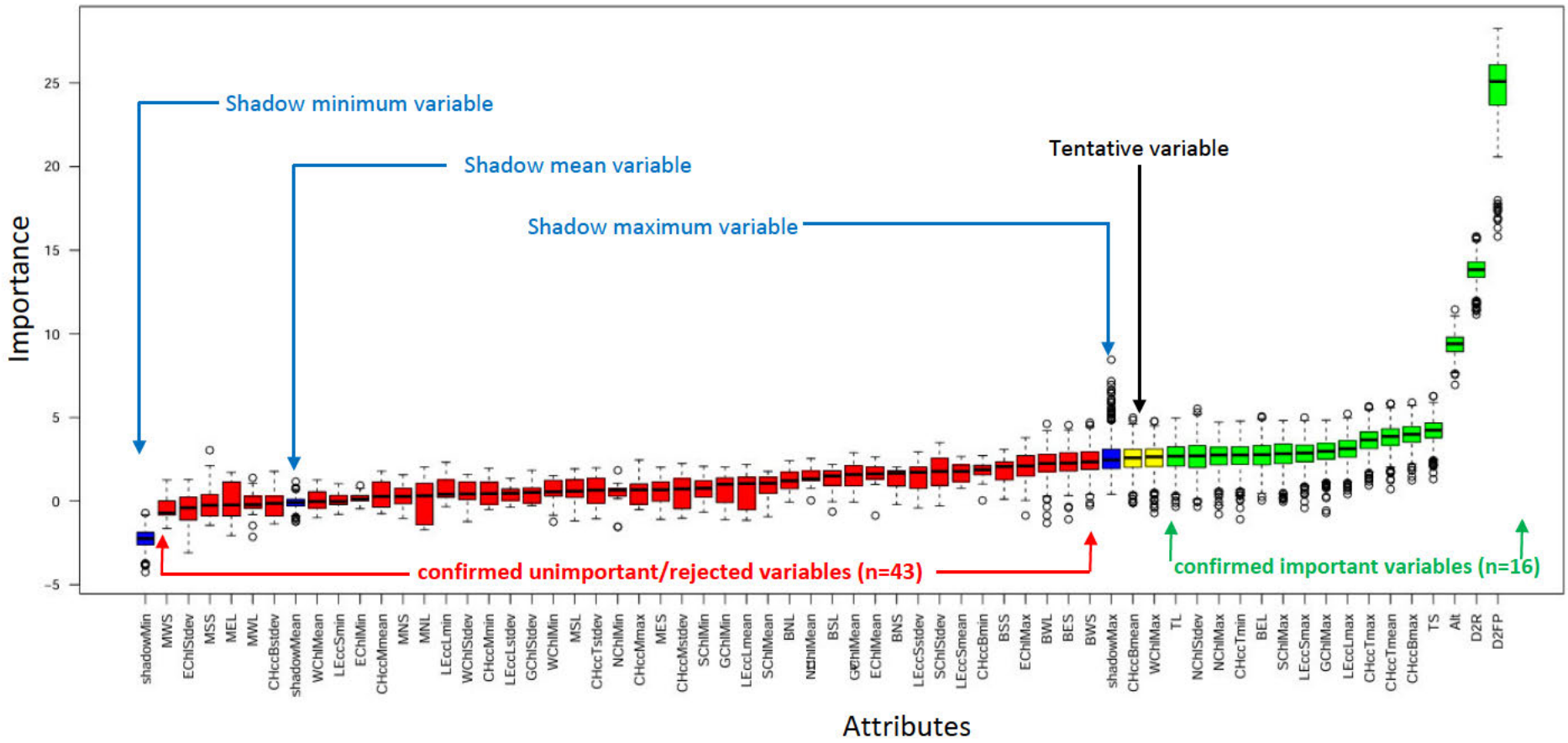


Figure 5.4: Boxplots showing Boruta variable importance projection model with 16 confirmed important (green), 43 rejected (red), and 2 tentative (yellow) with 3 randomly shuffled shadow (blue) features from musawu or jujube (*Ziziphus mauritiana*) leaf chlorophyll content (aspect on canopy, exposure to sunlight, height on canopy), statistical chlorophyll derivative and spatial locational features (n = 65) in semi-arid Mbire and Muzarabani Floodplain in Zimbabwe (Software: R-Project, Library: Boruta). Names of respective abbreviated attributes are presented in full on Appendix 5.1.

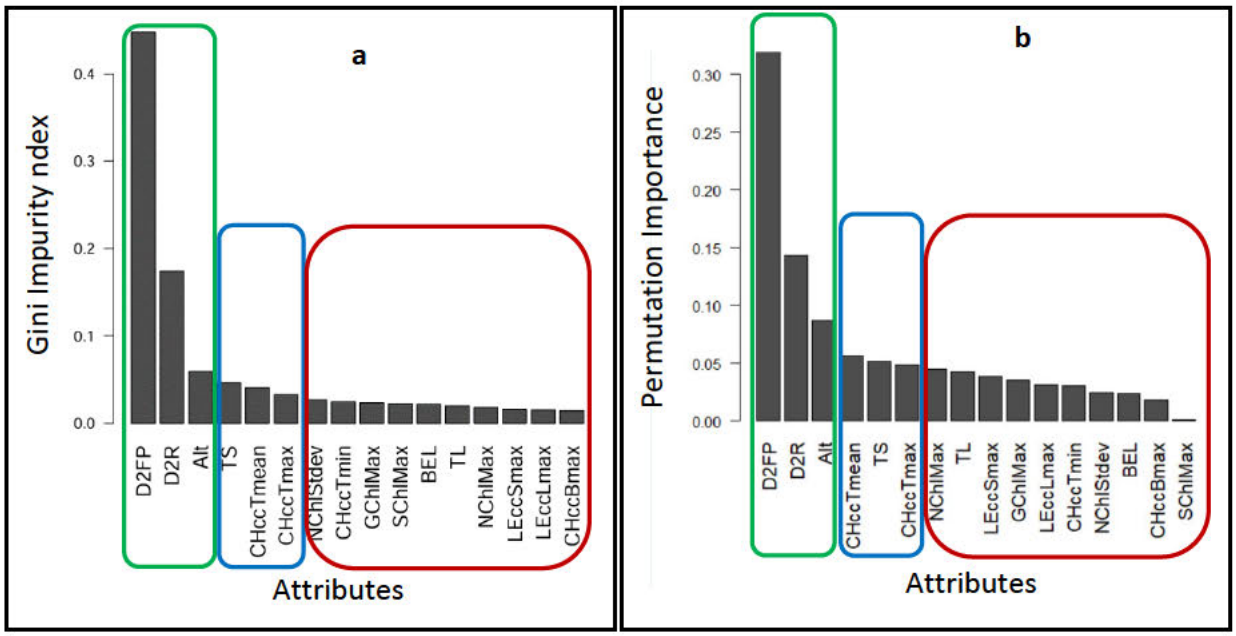


Figure 5.5: The Random Forest variable importance projection for the 16 BVIP selected variables using the Mean Decrease Impurity or Gini Impurity Index (a) and Mean Decrease Accuracy or Permutation Importance. Names of the respective abbreviated attributes are presented in full on Appendix 5.1.

However, the ETL showed the largest level of significance. Using the Random Forest machine learning model, the training set had a small OOB error (2.41%) than testing (0.00%) for train and test for assessing the influence of tree canopy architecture factors (aspect, height, and exposure to sunlight) and spatial (D2FP and D2R) on LCC. Spatial explanatory variables were the most important compared to those non-spatial (Figures 5.4 and 5.5). Notably, TS, CHccTmean, and NChlMax (a) (blue box) are the second most important explanatory variables which exchange positions in the Permutation importance (b). The rest of the explanatory variables (c) (red boxplots) are consistently noted but haphazardly interchanged between the two indices. For instance, TL is promoted from the 12th position in Gini Index (Figure 5.5a) to 8th position in Permutation Index (Figure 5.5b). This behaviour is also true for the other remaining explanatory variables. These exchanges show instability in the performance of these variables unlike the most important three spatial locational variables in both indices. Therefore, D2FP, and D2R as well as altitude are important explanatory variables in both indices. These indices confirm that floodplain FRSM has significant influence on the LCC. Figure 5.6 shows variables run in Boruta model.

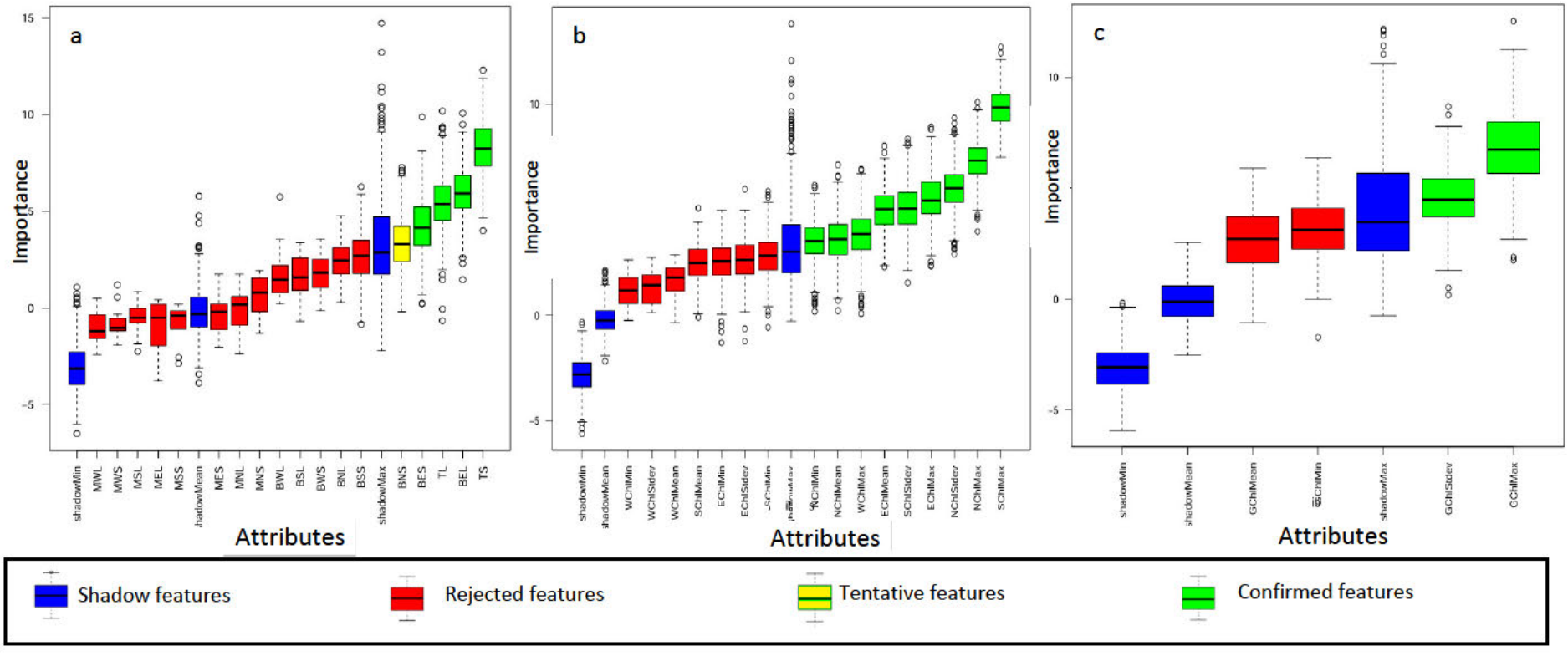


Figure 5.6: Box-and-whisker plots of 18 samples per tree field leaf chlorophyll content (LCC) (a), derived aspect on canopy (AOC) (b) and the four summary LCC from 18 samples per tree (c) with confirmed, rejected, tentative and shadow boxplots respectively. Names of the respective abbreviated attributes are presented in full on Appendix 5.1.

5.3.3 Variability of LCC between inside and outside floodplain using ANOVA

The effect of sunlight on leaf chlorophyll content was significant. Leaves in the shade depicted higher maximum, mean, minimum and standard deviation (Table 5.2).

Table 5.2: LCC derived on AOC, ETL and HOC for locations in and out of the floodplain

Class	Leaf Chlorophyll Content (LCC) Metrics by exposure to sunlight							
	Maximum		Mean		Minimum		Stdev	
	exposed	obstructed	exposed	obstructed	exposed	obstructed	exposed	obstructed
Flooded	139.6	194.3	121.7	125.1	101.1	108.3	61.1	69.1
Non-flooded	121.1	136.2	106.5	110.1	098.2	099.0	46.3	47.9
Difference (FP-NFP)	018.5	058.1	005.2	015.0	002.9	009.3	14.8	21.2

Comparatively, the LCC for samples in-shade (obstructed) was higher than for those in sunlight (exposed) for all the four statistical metrics (Table 5.2). Similarly, the LCC inside the floodplain was higher than outside as shown on computed difference on Table 5.2. This confirms a combined influence of location and ETL on LCC also in Figure 5.7. The combined influence of both ETL and floodplain location of tree LCC is evident. Therefore, LCC for in-shade leaf samples located inside floodplain was the highest while in sunlight outside floodplain was the least.

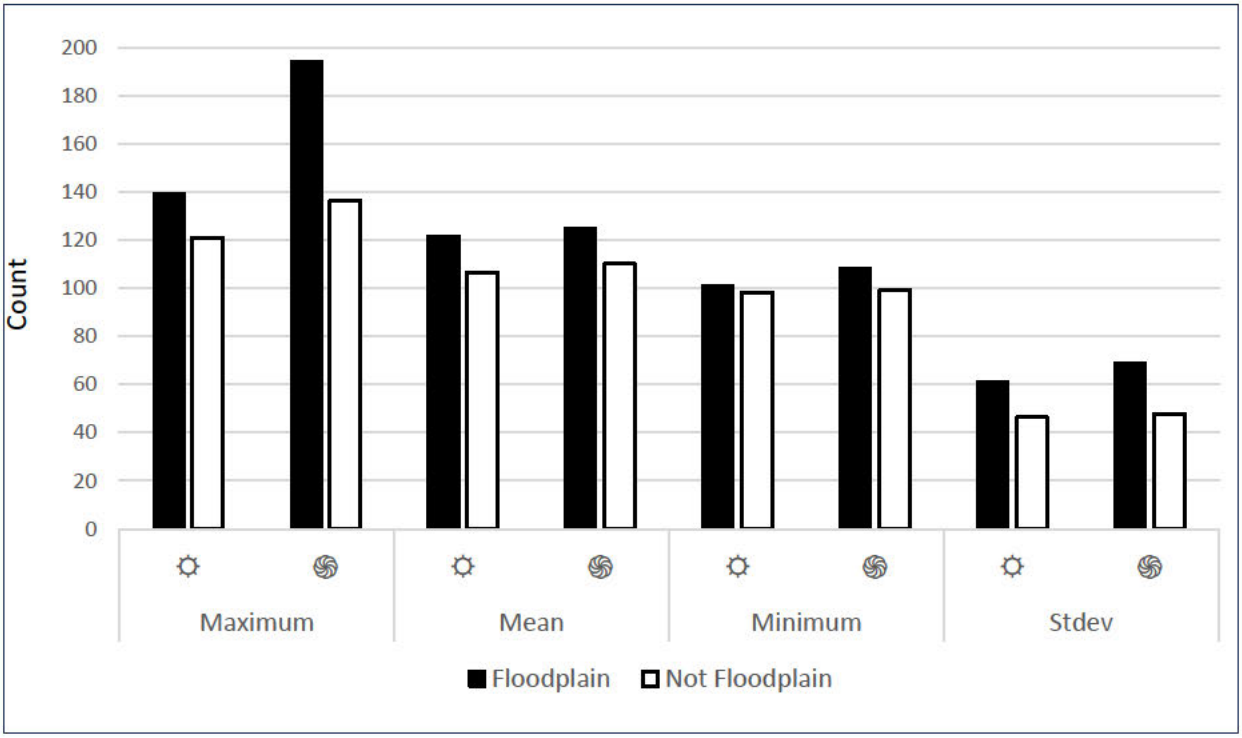


Figure 5.7: Comparison of leaf chlorophyll content for samples exposed against not exposed to sunlight distinguished by location based on descriptive statistical metrics.

All metrics on Table 5.2 show higher LCC inside the flood-influenced floodplain as compared to the outside which is also confirmed on Figure 5.7. Using two-way ANOVA, significant differences ($p<0.001$) were observed in LCC between the sunlit (ETL) as opposed to shade (not ETL) leaves inside and outside floodplain locations. This result confirms the indirect importance on floods on the LCC of floodplain trees. Therefore, the importance of FRSM as compared to the leaves not influenced by flooding is apparent. Since spatial variables were the highest in variable importance after Random Forest machine learning, further comparison using pairwise correlation was performed.

Figure 5.8. presents results of the pairwise correlation matrix. These correlation matrices related tree sample LCC to D2FP and D2R. Since flood extent has spatial coverage, the D2FP, and D2R were important to determine the influence of FRSM on LCC and hence on trees.

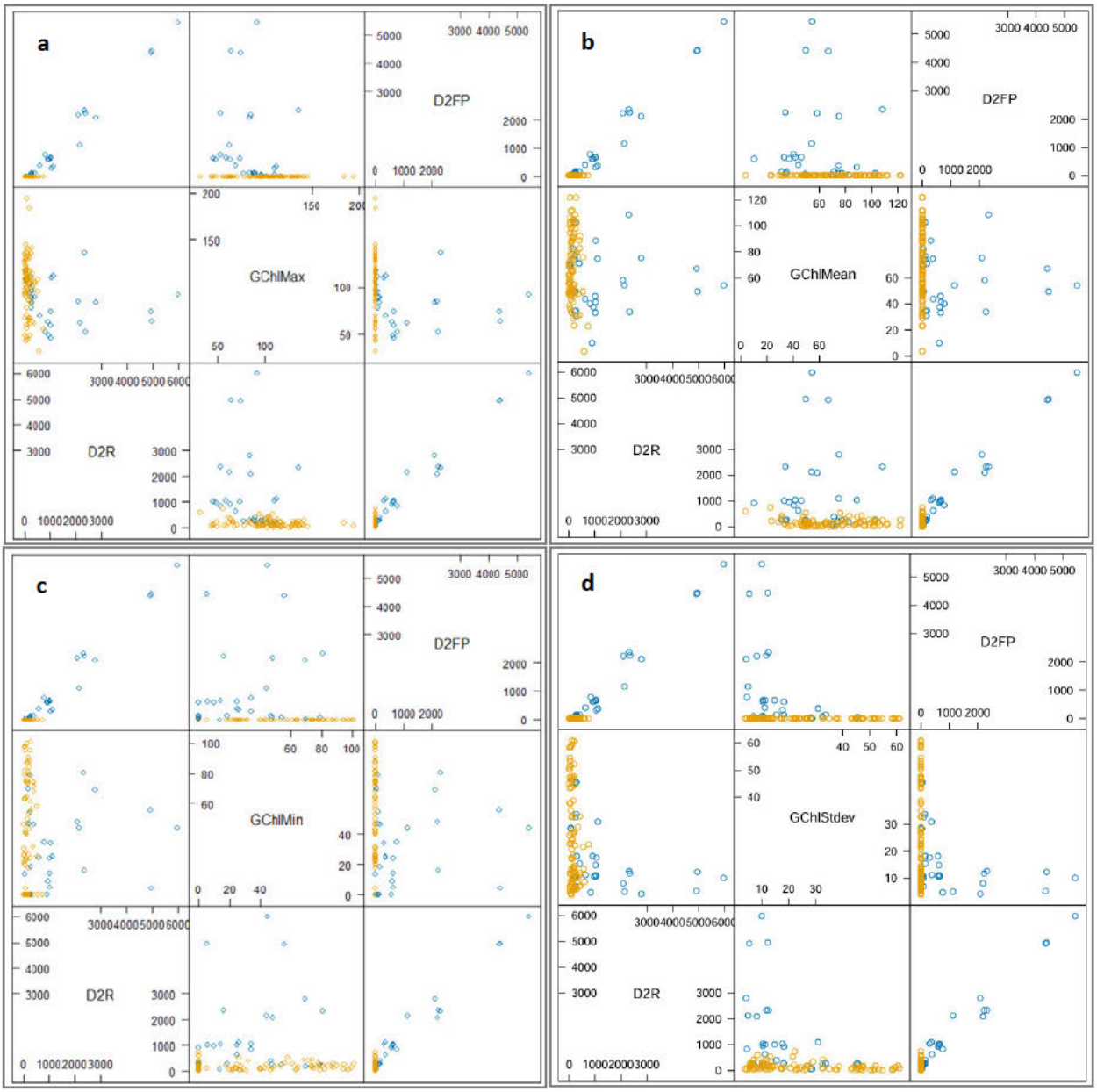


Figure 5.8: Pairwise scatterplot matrix for distance to floodplain (D2FP) and river (D2R) for samples inside floodplain (orange) and outside floodplain (blue) using GChlMax (a), GChlMean (b) GChlMin (c) and GChlStdev (d) for LCC of the *Z. mauritiana* tree.

Table 5.3: Evaluating the effect of exposure to light (ETL) effect on leaf chlorophyll content (LCC) using linear models based on descriptive metrics data.

Predictor Variables	**Model**							
	a) Maximum		**b) Mean**		**c) Minimum**		**d) Stdev**	
(ETL)	**F**	**Pr(>F)**	**F**	**Pr(>F)**	**F**	**Pr(>F)**	**F**	**Pr(>F)**
LEccL	10.609	0.00148**	4.220	0.0422*	0.103	0.7492	4.950	0.028*
LEccS	0.120	0.72955	0.566	0.4535	0.027	0.86954	0.068	0.795
LEccL:LEccS	1.538	0.21740	0.039	0.8431	9.467	0.00262**	0.740	0.391

a) Floodplain Class ~ LEccLMax+LEccSMax+LEccLMax:LEccSMax

b) Floodplain Class ~ LEccLMean+LEccSMean+LEccLMean:LEccSMean

c) Floodplain Class ~ LEccLMin+LEccSMin+LEccLMin:LEccSMin

d) Floodplain Class ~ LEccLStdev+LEccSStdev+LEccLStdev:LEccSStdev

*Significant codes: 0 '***' 0.001'**' 0.01 '*' 0.05 '•' 0.1 ' ' 1.0*

5.3.4 The effect of data augmentation using statistical derivation

The variables were useful in predicting influence of LCC aggregation by AOC, ETL and HOC. The results of the respective linear models are summarized on Table 5.3. Exposure to sunlight is an important chlorophyll determinant. The ANOVA was crucial to delimit Chlorophyll content by the descriptive attributes. While all top and middle HOC LCC were confirmed important using the BVIP, the ANOVA test showed that these were not statistically significant. The TS was also noted to be important using Gini and Permutation indices. The aggregation was important since augmented data significantly showed essential factors.

5.3.5 Model performance evaluation

The Random Forest classification metrics, and two-way ANOVA were used to evaluate classification and linear regression models, respectively. The Random Forest variable selection and classification evaluation showed high performance with accuracy (train = 0.8846, test = 1.000), kappa (train = 0.7234, test = 1.0000), sensitivity (train = 0.6667, test = 1.0000), and specificity (train = 0.1000, test = 1.0000) based on floodplain classification accuracy assessment using confusion matrix. Distance related factors (D2FP, D2R) and floodplain location (floodplain class) were significant ($p<0.001$) determinants of LCC. The variability using ANOVA confirmed that D2FP and D2R are significant determinants of the noted differences between floodplain and non-floodplain LCC. Also, the ETL showed significant differences ($p<0.01$), though lower than distance-based between LCC for trees exposed as compared to those obstructed to sunlight. In-shade leaves contained higher LCC than those exposed to sunlight, both inside and outside the floodplain. The HOC was also significant at a much lower level ($p<0.1$) than that for ETL. The AOC was not significant in determining chlorophyll on trees inside and outside floodplain, although north-facing slope leaves maximum metric had some weak signal. Clearly, floodplain located leaves contain higher LCC as compared to those outside the floodplain. The same was true for leaf ETL. Table 5.4 presents model coefficients and significance level for the spatial explanaory variables.

Table 5.4: Significant test results from linear regression using four spatial variables

	Tree maximum chlorophyll content by four spatial variables			
Variable	**Maximum**	**Mean**	**Minimum**	**Stdev**
Coefficient	**105.2200***	**94.9706***	75.9872	24.8914
Altitude	-0.0301	-0.0957	-0.1480	0.0012
Distance to river	**-0.0319***	-0.0103	0.01786	**-0.02147***
Distance to floodplain	**0.0311•**	0.0112	-0.0146	**0.0193***
Class	**12.5143•**	8.9856	7.3412	2.7135
	Canopy Height Chlorophyll content at the bottom of canopy (CHccB)			
Coefficient	**107.2044***	**145.0130****	**161.5119****	**-21.8807•**
Altitude	-0.0448	-0.1906	**-0.2768***	**0.0906****
Distance to river	**-0.0336***	**-0.0332***	**-0.0403****	0.0024
Distance to floodplain	**0.0312***	**0.0325***	**0.0422****	-0.0041
Class	11.2068	8.1671	9.0690	0.5609
	Canopy Height Chlorophyll content at the middle (CHccM)			
Coefficient	12.9815	42.7054	89.3020	-26.4709
Altitude	0.1246	0.0049	-0.1655	**0.1027***
Distance to river	0.0106	0.0157	0.0211	0.0034
Distance to floodplain	-0.0072	-0.0133	-0.0195	0.0037
Class	7.3196	6.4498	4.0300	1.0958
	Canopy Height Chlorophyll content at the top (CHccT)			
Coefficient	**95.8621***	**103.7528***	**111.6435***	-1.116e+01
Altitude	-0.0867	-0.1176	-0.1484	4.363e-02
Distance to river	**-0.0242•**	**-0.0234•**	-0.0226	-1.077e-03
Distance to floodplain	**0.0254•**	**0.0250•**	0.0246	6.123e-04
Class	**23.2598****	**22.3669****	**21.4741****	1.263e+00
	Aspect on canopy Chlorophyll content for East-facing canopy sides			
Coefficient	**141.1178****	**124.1360****	**118.8753•**	17.6535
Altitude	-0.1554	-0.1705	-0.2204	0.01804
Distance to river	**-0.0249•**	-0.0077	0.0027	-0.0154
Distance to floodplain	0.0233	0.0087	0.0016	0.0121
Class	7.1083	6.3620	6.8530	0.4138
	Aspect on canopy Chlorophyll content for South-facing canopy sides			
Coefficient	**117.0091***	**93.7043***	76.4832	26.6432
Altitude	-0.1030	-0.0906	-0.0940	-0.0186
Distance to river	-0.0215	-0.0046	0.0100	-0.0177
Distance to floodplain	0.0198	0.0042	-0.0097	0.0164
Class	**12.7690•**	6.6351	1.8997	5.5897
	Aspect on canopy Chlorophyll content for West-facing canopy sides			
Coefficient	**98.6911***	**84.3044•**	67.6309	26.1359
Altitude	-0.0396	-0.0587	-0.0722	-0.0097
Distance to river	**-0.0404****	-0.0190	0.0021	-0.0238
Distance to floodplain	**0.0417***	0.0204	0.0005	0.0229
Class	12.1001	7.1202	4.0450	4.2001
	Aspect on canopy Chlorophyll content for North-facing canopy sides			
Coefficient	**103.4583**	**73.32264•**	51.5148	33.5890
Altitude	-0.0772	-0.0490	-0.0472	-0.0327
Distance to river	**-0.0264•**	-0.0036	0.0222	**-0.0276***
Distance to floodplain	**-0.0264•**	0.0050	-0.0196	**0.0261***
Class	**16.7892***	9.1050	2.7897	8.0797
	Light exposure Chlorophyll content			
Coefficient	**97.2354***	**89.4447***	58.3022	22.1639
Altitude	-0.0221	-0.0848	-0.0949	0.0098
Distance to river	**-0.0270***	-0.0093	0.0161	**-0.0215***
Distance to floodplain	**-0.0239•**	0.0101	-0.0109	**0.0185***
Class	9.9853	8.7855	8.1621	2.0320
	In-shade (light prohibited) Chlorophyll content			
Coefficient	**118.7788***	**100.4723***	71.7893	28.7658
Altitude	-0.0712	-0.1064	-0.1257	-0.0077
Distance to river	**-0.0301***	-0.0114	0.0235	**-0.0229***
Distance to floodplain	**0.0290•**	0.0125	-0.0212	**0.0211***
Class	12.3862	9.17778	5.3074	3.4463

*Significant codes: 0 '***' 0.001'**' 0.01 '*' 0.05 '•' 0.1 ' ' 1.0*

The four spatial explanatory variables of LCC interacted with AOC, ETL and HOC show inconsistent significance levels by model. The maximum followed by mean are the leading metrics depicting significance of spatial variables. However, little contribution by minimum and standard deviation was noted. Overall, the data augmentation helped to incorporate some of the variables that otherwise would have remained unexplained.

5.4 Discussion

This study sought to assess the influence of FRSM on LCC based on comparison of leaves in-versus out-side the floodplain, exposure to sunlight, distance to river and foodplain and augmented statistical derivatives. To achieve this, a method based on a sequenced machine learning pipeline executed in R by calling relevant functions from particular libraries was adopted due to the complex link and subtlety of the relationships.

5.4.1 Variables influencing LCC in a semi-arid floodplain

Firstly, the study sought to assess the variables that related floodplain FRSM with LCC. The fieldwork was conducted when the trees outside the floodplain had shed their leaves while those inside floodplain had high chlorophyll content. This informs an understanding on the role of floods in sustaining trees through soil moisture supply, hence promote net primary productivity, important for climate regulation (Duncanson et al., 2023; Mo et al., 2023; Reiner et al., 2023). With a linear regression assumption, we fitted a Binary Logistic Regression on floodplain/non-floodplain class on LCC and spatial data to test whether LCC inside the floodplain was higher than outside. If that was the case, then leaves inside the floodplain should have significantly different values from that on the control experiment. Surprisingly, the binary logistics regression did not show statistical significance on the explanatory variables to floodplain class based on chlorophyll and distance variables. Instead, the Binary Logistic Regression gave warnings on model non-convergence and probability. Meanwhile, there was no statistical significance. However, the two warnings given were not necessarily error flags. Correia et al. (2019) indicated that such could be a result of perfect separation between classes. The lack of statistical significance was thus inconclusive, hence a need to assess variable importance with decision tree-based methods.

5.4.2 Variable importance relating to LCC in semi-arid floodplain

Variable importance projection was essential since the binary logistic regression did not find statistically significant relationships explaining the floodplain class using LCC and other spatial explanatory variables. The BVIP function executed in Boruta package was efficient for explanatory variables importance selection visualized distinctively using explicit boxplots for

confirmed important, confirmed unimportant and those remaining tentative against the maximum, mean and minimum shadow variables (Kursa and Rudnicki, 2010) at the end of the 'maxRuns' (1000 iterations) parameter ceiling. Therefore, executing BVIP and then running a Random Forest classification using selected important variables further assessed variable importance using the Gini and Permutation indices (Breiman 2001. Random Forest is a robust classification algorithm in machine learning capable of assessing variable importance (Breiman, 2001). Therefore, Random Forest classification was ideal to classify floodplain versus non-floodplain to model LCC variability between LCC in the two locations. Gini and Permutation indices noted D2FP, D2R and altitude as the most important variables determining LCC in the study area for the species studied. Notably, spatial flood-related distance (D2FP and D2R) and altitude were the most important. Janizadeh et al. (2019) also used distance to river to model flood susceptibility. There was a higher LCC inside the floodplain compared to outside since higher soil moisture positively influences chlorophyll content (Liu, Gudmundsson et al., 2020).

5.4.3 Higher LCC inside than outside the semi-arid floodplain

Overall, the study found that trees inside the floodplain depicted higher LCC using maximum, mean and minimum compared to those outside. This shows the important influence of FRSM on LCC and consequently tree chlorophyll. The spatial factors were shown to be more important using the Gini, and Permutation Importance as compared to tree-based chlorophyll factors and derivative data. Results of fitting spatial variables to test whether floodplain LCC were explained using spatial variables. We found that spatial variables were statistically significant ($p<0.001$). As such, an ANOVA test of the noted variability between inside and outside floodplain confirmed existence of significant differences between inside versus outside floodplain. The floodplain LCC was higher than that outside ($p<0.001$). Therefore, flooding is essential to enrich soil moisture through deep infiltration (Pedzisai et al., 2022). While samples inside the floodplain had higher minimum, mean and maximum LCC, they also had the highest variability as shown with higher standard deviation. This resulted in wider LCC distribution as compared to a narrower LCC for tree leaves outside the floodplain. Kulkarni et al. (2010) noted that short term drought affects the leaf anatomy and response hence, the observed differences in LCC between flooded and non-flooded trees.

Floodplain benefit soil moisture through deep infiltration as noted by Pedzisai et al. (2022). Doughty et al. (2015) also found that drought impacted on forest carbon balance in the Amazon. As such, trees inside floodplain benefit from more FRSM unlike those outside, which are affected largely by the drought in the study area (Mavhura et al., 2022; Mudavanhu et al., 2020). Studies

have confirmed that soil moisture is directly linked to LCC (González-Espíndola et al., 2024; Wang et al 2021; Zhang, Duan et al., 2007). Meanwhile, LCC is positively related to primary productivity (Liu et al., 2020). In semi-arid lands, floodplains have FRSM, which provides more soil water as compared to the outside (Pedzisai et al., 2022). Consistently, an evaluation of noted LCC difference between LCC for leaves inside as compared to outside the floodplain was attained using ANOVA. The study found that floods have a significant effect on LCC. Floodplain tree leaves had higher LCC as compared to those outside as confirmed using ANOVA. This was true for the maximum, mean, and minimum metrics between flooded and non-flooded samples. This finding is confirmed by an earlier study by Pedzisai et al., (2022) who noted that floods positively influence soil moisture content.

The determination of effect of floods on LCC required robust prediction of LCC based on spatial and tree canopy-based factors. While initial binary logistic regression depicted no significant explanatory variables, after Boruta-selected variables were used to classify floodplain versus non-floodplain LCC, Random Forest training and testing had high accuracy, kappa and very low OOB errors, specificity, and sensitivity. The BVIP selected 16 explanatory variables from which the leading explanatory variables were explicitly spatial namely D2FP, D2R, altitude and floodplain class followed by ETL samples. Janizadeh et al. (2019) also utilized D2R in predicting flood susceptibility, thus showing that D2R is an important variable in evaluating riverine flood-related phenomena. In this study, unique extraction of samples inside floodplain to exclusively discriminate LCC from that entirely dependent on D2FP for non-floodplain samples proved important to discriminate further the noted higher LCC inside floodplain influenced by flooding. Hence, Random Forest variable selection machine learning models selected important explanatory variables established that LCC was significantly influenced by D2FP, D2R and floodplain location.

5.4.4 Statistical data augmentation of LCC in semi-arid floodplains

Amongst the three HOC, TS and TL were important while north-facing maximum LCC AOC were significant. The two categories of tree-based explanatory variables were insignificant for determining LCC both inside and outside the floodplain. Leaf data augmentation enabled addition of variables that were found to be useful to explain variation of chlorophyll by location, height and aspect. On the contrary, ETL was found to be a significant factor discriminating LCC. Specifically, ETL (sunlit) sample LCC were lower than that of non-ETL (shade) leaves. This could be a result of photoinhibition (Takahashi et al., 2002) with lower LCC during mid-day as compared to higher at either dawn or dusk (Hoel and Solhaug, 1998). The ETL influences tree

photosynthetic performance (Lo Piccolo et al., 2023). Related to ETL, Adams et al. (2013) noted that photoinhibition occurs in leaves in response to excess light, thus decreasing photosynthetic efficiency regarded as photodamage that in turn reduces both plant productivity and growth. Similarly, Gardiner and Krauss (2001) tested short-term flooding effect on cherrybark oak seedlings to conclude that in-shade leaves had a 15% greater capacity on photosynthesis than those exposed to sunlight. Unlike in Kitao et al. (2024) study that noted mature leaves reaching senescence, in this study, some of the trees outside floodplain had completely shed their leaves, while those in the floodplain were still green.

In agreement with Lo Piccolo et al. (2023), this study results show that leaves in the shade have higher chlorophyll content as compared to those exposed to sun. This disparity is due to low photosynthesis during mid-day as stomata close to reduce water loss (Fu et al., 2022). In-Shade leaves optimise light absorption by increasing chlorophyll density per unit leaf mass (e.g., Dai et al., 2009). Takahashi et al. (2002) noted a high light sensitivity in *Ficus microcarpa*, where sun leaves are yellow while shade-leaves are green, to show that chlorophyll remain high in the latter. Kitao et al. (2024) proposed holocanopy hypothesis that argues that outer-canopy leaves protect inner-canopy leaves from photoinhibition, hence the difference in LCC noted in this study. Kitao et al. (2024) also noted that inner leaves may even remain green for longer than outside leaves, hence delayed senescence. Moore (1965) indicated that within the maturity-to-senescence phenology window, there are associated changes in leaf constituents including chlorophylls, carotene, and sugar in tree leaves. This could be the case in the study, since outside floodplain, some of the trees had either shed leaves, or had lower chlorophyll as revealed by descriptive statistical results. Individual leaf non-destructive temporal measurements showed that autumn senescence for deciduous trees is characterized by chlorophyll degradation and synthesis of flavonoids (Mattila et al., 2018). The individual tree leaf can retain its photosynthetic activity until a rapid phase of chlorophyll degradation and flavonoid synthesis begins (Mattila et al., 2018). The *Z. mauritiana* is a senescing deciduous tree characterized by shedding of leaves that starts with the degradation of chlorophyll before carotenoids and proteins (Mattila et al., 2018). However, in this study, inside floodplain, fieldwork showed that floodplain trees had high LCC while outside, they had already shed all leaves. Muller et al. (2001) noted that extreme light can lead to photooxidative damage resulting in damaging of the casing, pigment bleaching, and death. This non-photochemical quenching, which is a thermal dissipation process, can explain the observed differences between lower chlorophyll content in leaves exposed to sunlight as compared to higher chlorophyll in leaves in the shade. When light energy absorption exceeds

capacity for utilization, plants regulate and protect photosynthesis (Muller et al. 2001). The HOC differed from Zeng et al. (2024) who found lower leaves having higher chlorophyll than higher ones.

Some limitations of the study were noted. The Random Forest model tuning was optimized on the limited available field data. More data could possibly improve on model performance (Sheil, 2018). Chlorophyll content leaf sampling occurred during different times of day due to a short fieldwork window. This could also affect the results obtained, hence compromising the comparison, since Lo Piccolo et al. (2023) observed the highest LCC during either early morning or late evening, while it was minimal at midday.

Future research directions should aim to collect LCC for heterogeneous tree species to give a more realistic model of the effects of FRSM on different trees. Furthermore, more spontaneous sampling for more fieldwork to collect more samples is necessary. However, this calls for more efficient data sampling that requires use of automated LCC measuring instruments. The combination of machine learning with more data could give more insights into hidden effects. The ongoing data deluge especially in view of increasing spatial resolution could help to detect chlorophyll content using remote sensing-based indices. Up-scaling algorithms based on field measurements and artificial intelligence could facilitate up-scaling assessments.

5.5 Conclusion

This study sought to address a gap in literature on the influence of FRSM on short-term tree green-up phenology by assessing LCC. Sequenced ML modelling with chlorophyll and spatial flood-related variables was undertaken. Regression modelling on floodplain response variable against explanatory variables showed statistical significance, warning of model non-convergence and probability, hence variable importance selection applied Boruta algorithm. Subsequently, the selected explanatory variables were input into Random Forest classification modelling computing Gini and Permutation importance indices as well as accuracy, and kappa performance metrics, which confirmed the spatial (D2FP, D2R, altitude, and floodplain class) followed by exposure (no-ETL, and ETL) as important and statistically significant explanatory variables, respectively. Two-way ANOVA confirmed the existence of significant differences in LCC with higher values inside floodplain as compared to outside apparent in the minimum, mean and maximum LCC metrics. Furthermore, inside and outside the floodplain, sunlit leaves depicted lower LCC as compared to those in the shade. Since the spatial variables of LCC are flood-related, we conclude that FRSM has significant influence on tree LCC in the study area. The

study demonstrated that floods increase LCC in semi-arid areas. Future research could focus on heterogeneous species to evaluate inter-species effects of floods on LCC. Chlorophyll quenching could be measured using a commercial fluorometer and assessed with robust ML. Chlorophyll seasonality was different between inside and outside of the floodplain, confirming the influence of FRSM in semi-arid areas.

5.6 References

Adadi, A., 2021. A survey on data-efficient algorithms in big data era. *Journal of Big Data*, 8(1), 24.

Adams, W.W., Muller, O., Cohu, C.M. and Demmig-Adams, B., 2013. May photoinhibition be a consequence, rather than a cause, of limited plant productivity? *Photosynthesis Research* 117, pp.31 - 44.

Allaire, J.J., 24-12-2014 manipulate: Interactive Plots for RStudio. DOI: 10.32614/CRAN.package.manipulate.

Anderson, D.R., Burnham, K.P. and Thompson, W.L., 2000. Null hypothesis testing: problems, prevalence, and an alternative. *The Journal of Wildlife Management* pp.912 - 923.

Andrianto, H. and Faizal, A., 2017, October. Measurement of chlorophyll content to determine nutrition deficiency in plants: A systematic literature review. *2017 International Conference on Information Technology Systems and Innovation (ICITSI),* pp.392-397.

Angel, Y. and McCabe, M.F., 2022. Machine learning strategies for the retrieval of leaf-chlorophyll dynamics: model choice, sequential versus retraining learning, and hyperspectral predictors. *Frontiers of Plant Science* 13, 722442.

Anthony, M.A., Tedersoo, L., De Vos, B., Croisé, L., Meesenburg, H., Wagner, M., Andreae, H., Jacob, F., Lech, P., Kowalska, A. and Greve, M., 2024. Fungal community composition predicts forest carbon storage at a continental scale. *Nature Communications* 15(1), 2385.

Baciu, G.E., Dobrotă, C.E. and Apostol, E.N., 2021. Valuing forest ecosystem services. Why is an integrative approach needed? *Forests* 12(6), 677.

Bain, C.M., Shi, D., Ethridge, L.E., Norris, J.E. and Loeffelman, J.E., 2024. A Tutorial on supervised machine learning variable selection methods for the social and health Sciences in R, doi: 10.21203/rs.3.rs-4425792/v1.

Bejani, M.M. and Ghatee, M., 2021. A systematic review on overfitting control in shallow and deep neural networks. *Artificial Intelligence Rev*iews 54(8), pp.6391 - 6438.

Bennett, M.G., Lee, S.S., Schofield, K.A., Ridley, C.E., Washington, B.J. and Gibbs, D.A., 2021. Response of chlorophyll a to total nitrogen and total phosphorus concentrations in lotic ecosystems: a systematic review. *Environmental Evidence* 10, pp.1 - 25.

Bertamini, M. and Nedunchezhian, N., 2003. Photoinhibition of photosynthesis in mature and young leaves of grapevine (Vitis vinifera L.). *Plant Science* 164(4), pp.635 - 644.

Betts, M.G., Wolf, C., Ripple, W.J., Phalan, B., Millers, K.A., Duarte, A., Butchart, S.H. and Levi, T., 2017. Global forest loss disproportionately erodes biodiversity in intact landscapes. *Nature* 547 (7664), pp.441 - 444.

Bonan, G.B., 2008. Forests and Climate change: Forcings, Feedbacks, and climate benefits of forests. *Science* 320, pp.1444 - 1449.

Boonman, C.C., Serra-Diaz, J.M., Hoeks, S., Guo, W.Y., Enquist, B.J., Maitner, B., Malhi, Y., Merow, C., Buitenwerf, R. and Svenning, J.C., 2024. More than 17,000 tree species are at risk from rapid global change. *Nature Communications* 15(1), 166, doi: 10.1038/s41467-023-44321-9.

Brandt, M., Rasmussen, K., Hiernaux, P., Herrmann, S., Tucker, C.J., Tong, X., Tian, F., Mertz, O., Kergoat, L., Mbow, C. and David, J.L., 2018. Reduction of tree cover in West African woodlands and promotion in semi-arid farmlands. *Nature Geoscience* 11(5), pp.328 - 333.

Breiman, L., 2001. Random forests. *Machine Learning* 45, pp.5 - 32.

Chanza, N. and Musakwa, W., 2021. “Trees Are Our Relatives”: Local perceptions on forestry resources and implications for climate change mitigation. *Sustainability* 13, 5885.

Chaurasia, A.N., Parmar, R.M., Dave, M.G. and Krishnayya, N.S.R., 2024. Integrating field-and remote sensing data to perceive species heterogeneity across a climate gradient. *Scientific Reports* 14(1), 42.

Ciganda, V., Gitelson, A. and Schepers, J., 2009. Non-destructive determination of maize leaf and canopy chlorophyll content. *Journal of Plant Physiology* 166(2), pp.157 - 167.

Correia, S., Guimarães, P. and Zylkin, T., 2019. Verifying the existence of maximum likelihood estimates for generalized linear models. *arXiv preprint arXiv:1903.01633*.

Costanza, R. 2020. Valuing natural capital and ecosystem services toward the goals of efficiency, fairness, and sustainability. *Ecosystem Services* 43, 101096.

Costanza, R., d'Arge, R., De Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'neill, R.V., Paruelo, J. and Raskin, R.G., 1997. The value of the world's ecosystem services and natural capital. *Nature* 387 (6630), pp.253 - 260.

Crowther, T.W., Glick, H.B., Covey, K.R., Bettigole, C., Maynard, D.S., Thomas, S.M., Smith, J.R., Hintler, G., Duguid, M.C., Amatulli, G. and Tuanmu, M.N., 2015. Mapping tree density at a global scale. *Nature 525*(7568), pp.201 - 205.

De Foresta, H., Somarriba, E., Temu, A., Boulanger, D., Feuilly H., and Gauthier, M., 2013. Towards the assessment of trees outside forests. Resource Assessment Working Paper 183. 336**.**

Dezeure, R., Buhlmann, P., Meier, L. and Meinshausen, N., 2015. High-dimensional inference: Confidence intervals, p-Values and R-Software hdi. *Statistical Science* 30(4) pp.533 - 558.

Doughty, C.E., Metcalfe, D.B., Girardin, C.A.J., Amézquita, F.F., Cabrera, D.G., Huasco, W.H., Silva-Espejo, J.E., Araujo-Murakami, A., Da Costa, M.C., Rocha, W. and Feldpausch, T.R., 2015. Drought impact on forest carbon dynamics and fluxes in Amazonia. *Nature* 519 (7541), pp.78 - 82.

Duncanson, L., Liang, M., Leitold, V., Armston, J., Krishna Moorthy, S.M., Dubayah, R., Costedoat, S., Enquist, B.J., Fatoyinbo, L., Goetz, S.J. and Gonzalez-Roglich, M., 2023. The effectiveness of global protected areas for climate change mitigation. *Nature Communications* 14(1), 2908.

Dusenge, M.E., Duarte, A.G. and Way, D.A., 2019. Plant carbon metabolism and climate change: elevated CO 2 and temperature impacts on photosynthesis, photorespiration and respiration. *New Phytologist* 221(1), pp.32 - 49.

Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B. and Tabona, O., 2021. A survey on missing data in machine learning. *Journal of Big Data* 8, pp.1 - 37.

Food and Agriculture Organization of the United Nations, FRA platform, 2020 (accessed 03/01/2024). https://fra-data.fao.org/AF/fra2020/extentOfForest/.

Forstmeier, W. and Schielzeth, H., 2011. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology 65*, pp.47 - 55.

Fu, Z., Ciais, P., Prentice, I.C., Gentine, P., Makowski, D., Bastos, A., Luo, X., Green, J.K., Stoy, P.C., Yang, H. and Hajima, T., 2022. Atmospheric dryness reduces photosynthesis along a large range of soil water deficits. *Nature Communications* 13(1), p.989.

Gamfeldt, L., Snäll, T., Bagchi, R., Jonsson, M., Gustafsson, L., Kjellander, P., Ruiz-Jaen, M.C., Fröberg, M., Stendahl, J., Philipson, C.D. and Mikusiński, G., 2013. Higher levels of multiple ecosystem services are found in forests with more tree species. *Nature Communications* 4(1), 1340.

Gardiner, E.S., and Krauss, K.W., 2001. Photosynthetic light response of flooded cherrybark Oak (q*uercus* p*agoda*) seedlings grown in two light regimes. *Tree Physiology* 21, pp.1103 - 1111.

Watch, G.F., 2002. Global forest watch. *World Resources Institute, Washington, DC Available from http://www. globalforestwatch. org (accessed 12 April 2024)*.

González-Espíndola, L.Á., Pedroza-Sandoval, A., Trejo-Calzada, R., Jacobo-Salcedo, M.D.R., García de los Santos, G. and Quezada-Rivera, J.J., 2024. Relative water content, chlorophyll index, and photosynthetic pigments on *Lotus corniculatus* L. in response to water deficit. *Plants* 13(7), 961.

Green, J.K. and Keenan, T.F., 2022. The limits of forest carbon sequestration. *Science*, 376(6594), pp.692 - 693.

Gumbs, R., Scott, O., Bates, R., Böhm, M., Forest, F., Gray, C.L., Hoffmann, M., Kane, D., Low, C., Pearse, W.D. and Pipins, S., 2024. Global conservation status of the jawed vertebrate tree of life. *Nature Communications* 15(1), 1101.

Gumindoga, W., Makurira, H., Phiri, M. and Nhapi, I., 2016. Estimating runoff from ungauged catchments for reservoir water balance in the Lower Middle Zambezi Basin. *Water SA*, *42*(4), pp.641-649

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R. and Kommareddy, A., 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342(6160), pp.850 - 853.

Hansen, M.C., Potapov, P.V., Pickens, A.H., Tyukavina, A., Hernandez-Serna, A., Zalles, V., Turubanova, S., Kommareddy, I., Stehman, S.V., Song, X.P. and Kommareddy, A., 2022. Global land use extent and dispersion within natural land cover using Landsat data. *Environmental Research Letters* 17(3), 034050.

Harris, N.L., Gibbs, D.A., Baccini, A., Birdsey, R.A., De Bruin, S., Farina, M., Fatoyinbo, L., Hansen, M.C., Herold, M., Houghton, R.A. and Potapov, P.V., 2021. Global maps of twenty-first century forest carbon fluxes. *Nature Climate Change* 11 (3), pp.234-240.

Herritt, M.T., Pauli, D., Mockler, T.C. and Thompson, A.L., 2020. Chlorophyll fluorescence imaging captures photochemical efficiency of grain sorghum (Sorghum bicolor) in a field setting. *Plant Methods*, 16, pp.1-13.

Hoel, B. O., and Solhaug K. A., 1998 Efect of irradiance on chlorophyll estimation with the Minolta SPAD-502 leaf chlorophyll meter. *Annals of Botany* 82, pp.389 - 392.

Hu, X., Tanaka, A. and Tanaka, R., 2013. Simple extraction methods that prevent the artifactual conversion of chlorophyll to chlorophyllide during pigment isolation from leaf samples. *Plant Methods* 9, pp.1 - 13.

Inoue, Y., Guérif, M., Baret, F., Skidmore, A., Gitelson, A., Schlerf, M., Darvishzadeh, R. and Olioso, A., 2016. Simple and robust methods for remote sensing of canopy chlorophyll content: a comparative analysis of hyperspectral data for different types of vegetation. *Plant, Cell and Environment* 39(12), pp.2609 - 2623.

Iqbal, M., Shafiq, M., Zaidi, S. and Athar, M., 2015. Effect of automobile pollution on chlorophyll content of roadside urban trees. *Global Journal of Environmental Science and Management* 1(4), pp.283 - 296.

Janizadeh, S., Avand, M., Jaafari, A., Phong, T.V., Bayat, M., Ahmadisharaf, E., Prakash, I., Pham, B.T and Lee, S., 2019. Prediction Success of Machine Learning Methods for flash flood susceptibility mapping in the Tafresh Watershed, Iran. *Sustainability* 11 (5426). doi:10.3390/su11195426.

Jiang, M., Medlyn, B.E., Drake, J.E., Duursma, R.A., Anderson, I.C., Barton, C.V., Boer, M.M., Carrillo, Y., Castañeda-Gómez, L., Collins, L. and Crous, K.Y., 2020. The fate of carbon in a mature forest under carbon dioxide enrichment. *Nature* 580(7802), pp.227 - 231.

Jin, L., Yi, Y. and Xu, J., 2020. Forest carbon sequestration and China's potential: the rise of a nature-based solution for climate change mitigation. *China Economic Journal* 13(2), pp.200 - 222.

Kitao, M., Yazaki, K., Tobita, H., Agathokleous, E., Kishimoto, J., Takabayashi, A. and Tanaka, R., 2024. Anthocyanins act as a sugar-buffer and an alternative electron sink in response to starch depletion during leaf senescence: a case study on a typical anthocyanic tree species, Acer japonicum. *Journal of. Experimental Botany* erae109. doi:10.1093/jxb/erae109.

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* 28(5), pp.1 - 6.

Kuhn, M. Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer Z., Kenkel, B., and R Core Team (Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C and Hunt, T., 2023. Classification and regression training: Package 'caret" 6.0-94. *The R Journal*, 223(7). *Journal of Statistical Software* pp.1 - 26.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., and Hunt, T., 2020. Package 'caret'. *The R Journal*,*223*(7), p.48.

Kulkarni, M., Schneider, B., Raveh, E. and Tel-Zur, N., 2010. Leaf anatomical characteristics and physiological responses to short-term drought in Ziziphus mauritiana (Lamk.). *Scientia Horticulturae*, *124*(3), pp.316-322.

Kursar, M.B, and Rudnicki, W.R., 2022. Boruta. Wrapper Algorithm for All relevant feature selection version 8.0.0 online https://gitlab.com/mb/Boruta/-/issues.

Kursar, M.B., and Rudnicki, W.R., 2010. Feature Selection with Boruta Package. *Journal of Statistical Software*, 36(11), pp.1-13. doi:10.18637/jss.v036.i11.

Lengerich, B., Xing, E.P. and Caruana, R., 2020. On dropout, overfitting, and interaction effects in deep neural networks. *arXiv preprint arXiv:2007.00823*, *2*.

Li, H., Cui, L., Dou, Z., Wang, J., Zhai, X., Li, J., Zhao, X., Lei, Y., Wang, J., Li, W., 2023. Hyperspectral analysis and regression modeling of SPAD measurements in leaves of three mangrove species. *Forests*, 14; 1566. https://doi.org/10.3390/F14081566.

Li, W., Guo, W.Y., Pasgaard, M., Niu, Z., Wang, L., Chen, F., Qin, Y. and Svenning, J.C., 2023. Human fingerprint on structural density of forests globally. *Nature Sustainability*, 6(4), pp.368-379.

Liu, Y., Liu, R., Qi, L., Chen, J., Dong, J. and Wei, X., 2024. Global mapping of fractional tree cover for forest cover change analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, *211*, pp.67-82.

Li, Y., Sulla-Menashe, D., Motesharrei, S., Song, X.P., Kalnay, E., Ying, Q., Li, S. and Ma, Z., 2017. Inconsistent estimates of forest cover change in China between 2000 and 2013 from multiple datasets: Differences in parameters, spatial resolution, and definitions. *Scientific Reports*, 7(1), 8748.

Li, G., Törnqvist, T.E. and Dangendorf, S., 2024. Real-world time-travel experiment shows ecosystem collapse due to anthropogenic climate change. *Nature Communications*, *15*(1), p.1226.

Li, Y., Wang, H., Liu, C., Sun, J. and Ran, Q., 2024. Optimizing the Valuation and Implementation Path of the Gross Ecosystem Product: A Case Study of Tonglu County, Hangzhou City. *Sustainability*, *16*(4), p.1408.

Li, J., Wijewardane, N.K., Ge, Y. and Shi, Y., 2023. Improved chlorophyll and water content estimations at leaf level with a hybrid radiative transfer and machine learning model. *Computers and Electronics in Agriculture*, 206, 107669.

Li, Y., Xu, X., Wu, W., Zhu, Y., Gao, L., Jiang, X., Meng, Y., Yang, G. and Xue, H., 2025. Hyperspectral estimation of chlorophyll content in grapevine based on feature selection and GA-BP. *Scientific Reports*, *15*(1), p.8029.

Li, C., Zhu, X., Wei, Y., Cao, S., Guo, X., Yu, X. and Chang, C., 2018. Estimating apple tree canopy chlorophyll content based on Sentinel-2A remote sensing imaging. *Scientific Reports* 8(1), 3756.

Liaw, A., and Wiener, M., 2002. Classification and Regression by randomForest. *R News* 2(3), pp.18 - 22.

Liu, L., Gudmundsson, L., Hauser, M., Qin, D., Li, S. and Seneviratne, S.I., 2020. Soil moisture dominates dryness stress on ecosystem production globally. *Nature Communications* 11(1), 4892.

Liu, Y., Liu, R., Qi, L., Chen, J., Dong, J. and Wei, X., 2024. Global mapping of fractional tree cover for forest cover change analysis. *ISPRS Journal of Photogrammetry and Remote Sensing* 211, pp.67 - 82.

Lyu, C., Joehanes, R., Huan, T., Levy, D., Li, Y., Wang, M., Liu, X., Liu, C., and Ma, J., 2024. Enhancing selection of alcohol consumption-associated genes by random forest. *British Journal of Nutrition*, pp.1 - 10.

Ma, J., Li, J., Wu, W. and Liu, J., 2023. Global forest fragmentation change from 2000 to 2020. *Nature Communications*, 14(1), 3752. https://doi.org/10.1038/s41467-023-39221-x.

Makariev, A.M and Gorshkov, V.G., 2007. Biotic pump of the atmospheric moisture as driver of the hydrological cycyle on land. *Hydrol. Earth Systems Science,* 11(2), pp.1013 - 1033.

Malenovský, Z., Homolová, L., Lukeš, P., Buddenbaum, H., Verrelst, J., Alonso, L., Schaepman, M.E., Lauret, N. and Gastellu-Etchegorry, J.P., 2019. Variability and uncertainty challenges in scaling imaging spectroscopy retrievals and validations from leaves up to vegetation canopies. *Surveys in Geophysics 40*, pp.631 - 656.

Manikandan, G., Pragadeesh, B., Manojkumar, V., Karthikeyan, A.L., Manikandan, R. and Gandomi, A.H., 2024. Classification models combined with Boruta feature selection for heart disease prediction. *Informatics in Medicine Unlocked 44*, p.101442.

Mattila, H., Valev, D., Havurinne, V., Khorobrykh, S., Virtanen, O., Antinluoma, M., Mishra, K.B. and Tyystjärvi, E., 2018. Degradation of chlorophyll and synthesis of flavonols during autumn senescence—the story told by individual leaves. *AoB Plants* 10(3), p.ply028. doi:10.1093/aobpla/ply028.

Mavhura, E., 2019. Systems analysis of vulnerability to hydrometeorological threats: An exploratory study of vulnerability drivers in Northern Zimbabwe. *International Journal of Disaster Risk Science* 10, pp.204 - 219.

Mavhura, E., Manyangadze, T., Mudavanhu, C. and Pedzisai, E., 2022. An assessment of riparian communities' preparedness to flood risk: The case of Mbire communities in Zimbabwe. *GeoJournal* 87(3), pp.1671 - 1693.

Mavhura, E. and Mucherera, B., 2020. Flood survivors' perspectives on vulnerability reduction to floods in Mbire district, Zimbabwe. *Jàmbá: Journal of Disaster Risk Studies*, *12*(1), pp.1-12.

Maxwell, K. and Johnson, G.N., 2000. Chlorophyll fluorescence - a practical guide. *Journal of Experimental Botany*, 51(345), pp.659 - 668.

Mo, L., Zohner, C.M., Reich, P.B. *et al.* 2023. Integrated global assessment of the natural forest carbon potential. *Nature* 624, pp.92 - 101.

Moore, K.G., 1965. Senescence in Leaves of *Acer pseudoplatanus* L. and *Parthenocissus tricuspidata* Planch: 1. Changes in some leaf Constituents during Maturity and Senescence. *Annals of Botany* 29(3), pp.433 - 444,

Mori, A.S., Dee, L.E., Gonzalez, A., Ohashi, H., Cowles, J., Wright, A.J., Loreau, M., Hautier, Y., Newbold, T., Reich, P.B. and Matsui, T., 2021. Biodiversity–productivity relationships are key to nature-based climate solutions. *Nature Climate Change* 11(6), pp.543 - 550.

Mudavanhu, C., Manyangadze, T., Mavhura, E., Pedzisai, E. and Manatsa, D., 2020. Rural households' vulnerability and risk of flooding in Mbire District, Zimbabwe. *Natural Hazards* 103(3), pp.3591-3608.

Mugabowindekwe, M., Brandt, M., Chave, J., Reiner, F., Skole, D.L., Kariryaa, A., Igel, C., Hiernaux, P., Ciais, P., Mertz, O. and Tong, X., 2023. Nation-wide mapping of tree-level aboveground carbon stocks in Rwanda. *Nature Climate Change* 13(1), pp.91 - 97.

Muller, P., Li, X.P. and Niyogi, K.K., 2001. Non-photochemical quenching. A response to excess light energy. *Plant Physiology* 125(4), pp.1558 - 1566.

Neugarten, R.A., Chaplin-Kramer, R., Sharp, R.P., Schuster, R., Strimas-Mackey, M., Roehrdanz, P.R., Mulligan, M., van Soesbergen, A., Hole, D., Kennedy, C.M. and Oakleaf, J.R., 2024. Mapping the planet's critical areas for biodiversity and nature's contributions to people. *Nature Communications* 15(1), p.261.

Nevala, N.E., and Baden, T. A., 2019. low-cost hyperspectral scanner for natural imaging and the study of animal colour vision above and under water. *Scientific Reports* 9, 10799. doi:10.1038/s41598-019-47220-6.

Perez-Harguindeguy, N., Diaz, S., Garnier, E., Lavorel, S., Poorter, H., Jaureguiberry, P., Bret-Harte, M.S., et al. 2016. Corrigendum to: New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany* 64(8), pp.715 - 716.

Qin, Y., Xiao, X., Dong, J., Zhang, G., Roy, P.S., Joshi, P.K., Gilani, H., Murthy, M.S.R., Jin, C., Wang, J. and Zhang, Y., 2016. Mapping forests in monsoon Asia with ALOS PALSAR 50-m mosaic images and MODIS imagery in 2010. *Scientific Reports* 6(1), 20880.

Qin, Y., Xiao, X., Dong, J., Zhang, Y., Wu, X., Shimabukuro, Y., Arai, E., Biradar, C., Wang, J., Zou, Z. and Liu, F., 2019. Improved estimates of forest cover and loss in the Brazilian Amazon in 2000–2017. *Nature Sustainability* 2(8), pp.764 - 772.

Qiu, J.J.; Liu, Y.H.; Chen, C.J.; Huang, Q.Y. 2023. The spatial pattern and driving mode of the coupling between ecological product value and human welfare—A case study of Guangzhou. *Journal of Natural Resources* 38, pp.760 - 778.

Ramachandra, T.V., Soman, D., Naik, A.D. and Chandran, M.S., 2017. Appraisal of forest ecosystems goods and services: challenges and opportunities for conservation. *Journal of Biodiversity* 8(1), pp.12 - 33.

Rathore, M.M., Shah, S.A., Shukla, D., Bentafat, E. and Bakiras, S., 2021. The role of ai, machine learning, and big data in digital twinning: A systematic literature review, challenges, and opportunities. *IEEE Access* 9, pp.32030-32052.

Reiner, F., Brandt, M., Tong, X., Skole, D., Kariryaa, A., Ciais, P., Davies, A., Hiernaux, P., Chave, J., Mugabowindekwe, M. and Igel, C., 2023. More than one quarter of Africa's tree cover is found outside areas previously classified as forest. *Nature Communications* 14(1), 2258, doi:10.1038/s41467-023-37880-4.

Ren, Y., Qiu, J., Zeng, Z., Liu, X., Sitch, S., Pilegaard, K., Yang, T., Wang, S., Yuan, W. and Jain, A.K., 2024. Earlier spring greening in Northern Hemisphere terrestrial biomes enhanced net ecosystem productivity in summer. *Communications, Earth and Environment* *5*(1), 122.

Roebroek, C.T., Duveiller, G., Seneviratne, S.I., Davin, E.L. and Cescatti, A., 2023. Releasing global forests from human management: How much more carbon could be stored? *Science* 380(6646), pp.749 - 753.

Rotenberg, E. and Yakir, D., 2010. Contribution of semi-arid forests to the climate system. *Science* 327(5964), pp.451 - 454.

Sandel, B. and Svenning, J.C., 2013. Human impacts drive a global topographic signature in tree cover. *Nature Communications* 4(1), 2474.

Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., Elberg, A., Toomet, O., Crowley, J., Hofmann, H., and Hadley Wickham, H., 14/02/2024. GGally: Extension to 'ggplot2'. DOI: 10.32614/CRAN.package.GGally

Scogings, P.F., 2023. Perspective: Monitoring global forests using only structural metrics–problems and solutions from a savanna viewpoint. *Forest Ecology and Management* 546, 121381.

Sebastián, C. and González-Guillén, C.E., 2024. A feature selection method based on Shapley values robust for concept shift in regression. *Neural Computing and Applications* pp.1 - 23.

Shah, S.H., Angel, Y., Houborg, R., Ali, S. and McCabe, M.F., 2019. A random forest machine learning approach for the retrieval of leaf chlorophyll content in wheat. *Remote Sensing* 11(8), 920.

Sheil, D., 2018. Forests, atmospheric water and an uncertain future: the new biology of the global water cycle. *Forest Ecosystems* 5(1), pp.1 - 22.

Shi, Y., Wang, Z., Zhang, G., Wei, X., Ma, W. and Yu, H., 2024. Evaluating the Research Status of the Remote Sensing-Mediated Monitoring of Forest Biomass: A Bibliometric Analysis of WOS. *Forests* 15(3), 524.

Signorell, A., Aho, K., Alfons, A., Anderegg, N., Aragon, T., Arachchige, C., Arppe, A., Baddeley, A., et al., …and Zeileis, A., 03-02-2024. DescTools: Tools for descriptive statistics. Version 0.99.54.DOI: 10.32614/CRAN.package.DescTools.

Spawn, S.A., Sullivan, C.C., Lark, T.J. and Gibbs, H.K., 2020. Harmonized global maps of above and belowground biomass carbon density in the year 2010. *Scientific Data* **7**(1), 12.

Soysal, D., 2024. A non-destructive leaf area prediction model and some physical leaf properties in apples. *Applied Fruit Science* pp.1 - 6.

Steel, E.A., Hinckley, T.M., Richards, W.H. and D'Amore, D.V., 2024. Forests then and now: Managing for ecosystem benefits, services to humans, and healthy forests across scales. In *Future Forests* pp.49 - 64.

Stephenson, N.L., Das, A.J., Condit, R., Russo, S.E., Baker, P.J., Beckman, N.G., Coomes, D.A., Lines, E.R., Morris, W.K., Rüger, N. and Alvarez, E., 2014. Rate of tree carbon accumulation increases continuously with tree size. *Nature* 507(7490), pp.90 - 93.

Sun, Y., Frankenberg, C., Wood, J.D., Schimel, D.S., Jung, M., Guanter, L., Drewry, D.T., Verma, M., Porcar-Castell, A., Griffis, T.J. and Gu, L., 2017. OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence. *Science* 358(6360), eaam5747.

Takahashi, S., Tamashiro, A., Sakihama, Y., Yamamoto, Y., Kwamitsu, Y, and Yamasaki H., 2002. High-susceptibility of photosynthesis to photoinhibition in the tropical plant *Ficus microcarpa* L. f. cv. Golden Leaves. *BMC Plant Biology* 2**(**2), doi:10.1186/1471-2229-2-2.

Taubert, F., Fischer, R., Groeneveld, J., Lehmann, S., Müller, M.S., Rödig, E., Wiegand, T. and Huth, A., 2018. Global patterns of tropical forest fragmentation. *Nature* 554, pp.519 - 522.

The European Space Agency QuickBird-2 ESA archive earth.esa.int/geogateway/catalog/quickbird-2-esa-archive (retrieved 03/05/2024).

Tucker, C., Brandt, M., Hiernaux, P., Kariryaa, A., Rasmussen, K., Small, J., Igel, C., Reiner, F., Melocik, K., Meyer, J. and Sinno, S., 2023. Sub-continental-scale carbon stocks of individual trees in African drylands. *Nature* 615(7950), pp.80 - 86.

Vainio, K., Korrensalo, A., Takala, T., Räsänen, A., Lummaa, K. and Tuittila, E.S., 2024. Do you have a tree friend? - Human–tree relationships in Finland. *People and Nature* 6(2), pp.646 - 659.

Vieilledent, G., Grinand, C., Rakotomalala, F.A., Ranaivosoa, R., Rakotoarijaona, J.R., Allnutt, T.F. and Achard, F., 2018. Combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in Madagascar. *Biological Conservation* 222, pp.189 - 197.

Verhegghen, A., Kuzelova, K., Syrris, V., Eva, H. and Achard, F., 2022. Mapping canopy cover in African dry forests from the combined use of Sentinel-1 and Sentinel-2 data: Application to Tanzania for the year 2018. *Remote Sensing* 14(6), 1522.

Wagner, F.H., Dalagnol, R., Silva-Junior, C.H., Carter, G., Ritz, A.L., Hirye, M.C., Ometto, J.P. and Saatchi, S., 2023. Mapping tropical forest cover and deforestation with Planet NICFI satellite images and deep learning in Mato Grosso State (Brazil) from 2015 to 2021. *Remote. Sensing* 15(2), 521.

Waldeland, A.U., Trier, Ø.D. and Salberg, A.B., 2022. Forest mapping and monitoring in Africa using Sentinel-2 data and deep learning. *International Journal of Applied Earth Observation and Geoinformation* 111, 102840.

Wang, N., Fu, F., Wang, H., Wang, P., He, S., Shao, H., Ni, Z. and Zhang, X., 2021. Effects of irrigation and nitrogen on chlorophyll content, dry matter and nitrogen accumulation in sugar beet (*Beta vulgaris L.*). *Scientific Reports* 11(1), 16651.

Wang, H., Prentice, I.C., Keenan, T.F., Davis, T.W., Wright, I.J., Cornwell, W.K., Evans, B.J. and Peng, C., 2017. Towards a universal model for carbon dioxide uptake by plants. *Nature Plants* 3(9), pp.734 - 741.

Wang, L., Wei, F., Tagesson, T., Fang, Z. and Svenning, J.C., 2025. Transforming forest management through rewilding: Enhancing biodiversity, resilience, and biosphere sustainability under global change. *One Earth*, *8*(3).

Watson, J.E., Evans, T., Venter, O., Williams, B., Tulloch, A., Stewart, C., Thompson, I., Ray, J.C., Murray, K., Salazar, A. and McAlpine, C., 2018. The exceptional value of intact forest ecosystems. *Nature Ecology and Evolution* 2(4), pp.599 - 610.

Xie, X.S.; Qiu, X.L.; Zhao, R.; Chen, S.Z. 2023. The main crux and promotion path of ecological product value realization mechanism. *Forestry Industries* 60, pp.82 - 86.

Xu, J., Fu, G., Yan, L., Yu, L., Kuang, F., Huang, Q. and Tang, X., 2024. Estimation of the Relative Chlorophyll Content of Carya illinoensis Leaves Using Fractional Order Derivative of Leaf and Canopy Scale Hyperspectral Data. *Journal of Soil Science and Plant Nutrition* pp.1 - 17.

Xu, H., He, B., Guo, L., Yan, X., Dong, J., Yuan, W., Hao, X., Lv, A., He, X. and Li, T., 2024. Changes in the Fine Composition of Global Forests from 2001 to 2020. *Journal of Remote Sensing* 4, 0119.

Yamada, Y., Lindenbaum, O., Negahban, S. and Kluger, Y., 2020, November. Feature selection using stochastic gates. *International Conference on Machine Learning pp.*10648 - 10659.

Yamashita, H., Sonobe, R., Hirono, Y., Morita, A. and Ikka, T., 2020. Dissection of hyperspectral reflectance to estimate nitrogen and chlorophyll contents in tea leaves based on machine learning algorithms. *Scientific Reports* 10(1), 17360.

Yang, H., Ciais, P., Frappart, F., Li, X., Brandt, M., Fensholt, R., Fan, L., Saatchi, S., Besnard, S., Deng, Z. and Bowring, S., 2023. Global increase in biomass carbon stock dominated by growth of northern young forests over past decade. *Nature Geoscience* 16(10), pp.886 - 892.

Yang, F., Jiang, X., Ziegler, A.D., Estes, L.D., Wu, J., Chen, A., Ciais, P., Wu, J. and Zeng, Z., 2023. Improved fine-scale tropical forest cover mapping for Southeast Asia using Planet-NICFI and Sentinel-1 imagery. *Journal of Remote Sensing 3*, 0064.

Yee, T.W. (2022). On the Hauck-Donner effect in Wald tests: Detection, tipping points and parameter space characterization, *Journal of the American Statistical Association* 117, pp.1763 - 1774.

Yee, T., 2024. Vector Generalized Linear and Additive Models: Package 'VGAM' 1.1-10.

Yin, H., Khamzina, A., Pflugmacher, D. and Martius, C., 2017. Forest cover mapping in post-Soviet Central Asia using multi-resolution remote sensing imagery. *Scientific Reports* 7(1), 1375.

Zhang, M.C., Duan, L.S., Tian, X.L., He, Z.P., Li, J.M., Wang, B.M., and Li, Z.H., 2007. Uniconazole-Induced Tolerance of Soybean to Water Deficit Stress in Relation to Changes in Photosynthesis, Hormones and Antioxidant System, *Journal of Plant Physiology* 164, pp.709 - 717.

Zhang, X., Huang, H., Tu, K., Li, R., Zhang, X., Wang, P., Li, Y., Yang, Q., Acerman, A.C., Guo, N. and Liu, Y., 2024. Effects of plant community structural characteristics on carbon sequestration in urban green spaces. *Scientific Reports* 14(1), 7382.

Zhang, H., Lee, C.K., Law, Y.K., Chan, A.H., Zhang, J., Gale, S.W., Hughes, A.C., Ledger, M.J., Wong, M.S., Tai, A.P. and Hau, B., 2024. Integrating both restoration and regeneration potentials into the real-world forest restoration planning. *Integrating both restoration and regeneration potentials into the real-world forest restoration planning*.

Zhang, C.L.; Liu, Q.X.; Zhang, S.J.; Hao, L. 2024. Research on the value realization mechanism of ecological products in the source area of the Yellow River basin. *Ecological Economics* 40, pp.212 - 217.

Zhang, W., Rong, L., Xiong, K., Zhang, Z. and Chang, H., 2024. Spatial Pattern of Ecosystem Services and the Mechanism of Eco-Industry Formation in South China Karst Nature Reserves. *Forests* 15(3), 493.

Zhang, Z., Xiong, K., Zhang, Y. and Ning, Y., 2024. Research Progress on Forest Eco-Product Value Realization and Eco-Industry: The Inspiration for Planted Forests in Karst Desertification Control. *Forests* 15(3), 517.

Zhang, R., Yang, P., Liu, S., Wang, C. and Liu, J., 2022. Evaluation of the methods for estimating leaf chlorophyll content with SPAD chlorophyll meters. *Remote Sensing* 14(20), 5144.

Zeng, J., Liu, F., Zhu, Y., Li, J., Ruan, Y., Quan, X., Wang, C. and Wang, X., 2024. Degree of shade tolerance shapes seasonality of chlorophyll, nitrogen and phosphorus levels of trees and herbs in a temperate deciduous forest. *Journal of Forestry Research* 35(1), 72.

CHAPTER 6: LONG-TERM INFLUENCE OF FLOOD-RECHARGED SOIL MOISTURE ON TREE BIOPHYSICAL CHARACTERISTICS IN SEMI-ARID LANDS

The chapter is based on: **Pedzisai, E.,** Mutanga, O., and Odindi, J. Long-term influence of floods-recharged soil moisture on tree biophysical characteristics in semi-arid lands. *Under review*.

Abstract

In arid and semi-arid landscapes, soil moisture is limited, thus limiting tree growth. However, multipurpose tree species in these landscapes e.g., *Ziziphus mauritiana* have critical ecological, socio-economic, and environmental benefits that include food and habitat for biodiversity, edible fruits and medicinal value, soil protection from erosion, and climate change regulation through carbon sequestration. Therefore, episodic floods in these landscapes are a critical source of soil moisture recharge through deep infiltration that, in the long-term, supports tree clusters that are important biodiversity and socio-economic hotspots. However, the long-term influence of flood-recharged soil moisture (FRSM) on tree biophysical characteristics in semi-arid lands is poorly understood. Therefore, this chapter sought to assess the long-term influence of FRSM on tree growth in a semi-arid landscape. To achieve this, the study sampled 366 *Z. mauritiana* in- and out-side a floodplain. Specifically, the study measured canopy diameter, tree height, diameter at breast height (DBH), and locational co-ordinates and computed sample's distance to the river and floodplain by applying a vector overlay. Firstly, the study compared the differences in biophysical characteristics of trees within and outside the floodplain using Levene's test for homogeneity of variances, nonparametric Mann-Whitney-Wilcoxon independent rank sum test, and post-hoc pairwise Wilcoxon test. Secondly, to explain the observed spatial variability, the study performed linear regression of the three tree biophysical characteristics on key predictors namely distance to river, distance to floodplain, and flood status. Results of Levene's test showed significant differences of variances for canopy diameter and DBH at $p<0.001$, and tree height ($p<0.05$) between flooded and non-flooded trees. Similarly, the Mann-Whitney-Wilcoxon test showed significant differences for canopy diameter and tree height at $p<0.001$, and DBH at $p<0.01$ between flooded and non-flooded trees, respectively. The differences between floodplain and non-floodplain canopy diameter, tree height and DBH were confirmed as significant using post-hoc pairwise Wilcoxon rank sum test. Specifically, trees inside the floodplain had bigger canopy diameter, were taller and had larger stems, hence bigger than those outside. Linear regression showed significant coefficients for both distance to river and distance to floodplain at $p<0.001$, and flood status ($p<0.01$) on tree growth, further explaining the influence of spatial predictors. We conclude that FRSM contributes positively to the long-term tree growth in semi-arid landscapes, hence significant implications on climate change, ecological, economic, and environmental management decisions and policies.

Keywords

Canopy diameter; diameter at breast height; distance to floodplain; distance to river; floodplain; flood-recharged soil moisture; tree height.

6.1 Introduction

Clustered trees in semi-arid regions form important biodiversity hotspots that provide multiple social, ecological, environmental and economic benefits (Wallace et al., 2021; Talbot et al., 2018). Broadly, trees provide diverse ecosystem goods and services. Trees, for instance, have a net global cooling effect through sinking carbon, a greenhouse gas, hence are important in mitigating global warming (IPCC, 2019; Grassi et al., 2017; Rotenberg and Yakir, 2010). Multipurpose tree species in semi-arid floodplains e.g., jujube (*Ziziphus mauritiana*) provides edible fruits (Nyanga et al., 2008), medicine (Goyal et al., 2012), timber (Dlamini and Geldenhuis, 2009), and shade for social gatherings (Mausse et al., 2021). Furthermore, semi-arid trees provide habitat that support biodivesirty (Betts et al., 2017) and help to conserve soil against erosion (Fischer et al., 2021). Ultimately, in view of their importance, especially in semi-arid landscapes, a study of tree size characteristics, relative to their proximity to moisture is imperative.

Semi-arid landscapes are typified by limited soil moisture due to low precipitation, which limits tree growth (Pedzisai et al., 2022). As such, floods facilitate deep infiltration mechanism which significantly recharges soil moisture in drylands (Dahan et al., 2008). However, a knowledge gap exists on understanding how flood-recharged soil moisture (FRSM) influences tree biophysical characteristics in such areas of precipitation scarcity. Soil moisture promotes tree growth (Madani et al., 2018; Iizuka et al., 2017; Blanchard et al., 2016) hence, Pedzisai et al. (2022) concluded that FRSM is important for tree growth in arid and semi-arid landscapes. Within a flooded semi-arid floodplain, FRSM promotes more tree growth than outside (Pedzisai et al., 2022). Since floods are predicted to increase in the future (Alifu et al., 2022), escalated by an upsurge of high rainfall extreme events (Wang et al., 2014), it is anticipated that the increase will trigger more FRSM that promotes tree growth in semi-arid regions. Therefore, assessing the influence of FRSM on tree biophysical characteristics inside versus those outside the floodplain is necessary for understanding their long-term adaptation to FRSM

A shrub is shorter in height tha a tree (Körner, 2012). Recently, tree height, canopy size (e.g., diameter, area, etc.), and trunk diameter at breast height (DBH) have become valuable proxies to estimate aboveground biomass (AGB) (Mulatu et al., 2024; Wagner, et al., 2024; Qin et al.,

2021). Tree crowns can be mapped either manually (Disney et al., 2020) or automatically by applying algorithms developed using artificial intelligence on remotely sensed data (Chen et al., 2021; Onishi and Ise, 2021; 2018). Tree DBH is measured either using calipers, diameter tape, hypsometer or laser rangefinders and remotely sensed LiDAR data (Guenther et al., 2024; Su et al., 2024; Wardius and Hein, 2024; Xie et al., 2022). Gülci et al. (2023) for instance, compared LiDAR with caliper DBH measurements, while Huang et al. (2011) measured DBH and tree heights using automated scanning LiDAR. Iizuka et al. (2017) performed tree height mapping using ortho-photographs, digital surface, and images acquired using drones.

A study by Zhang et al. (2023) used vertex hypsometer and a diameter tape to estimate tree height and DBH, respectively. Such measurements are important in modelling tree variables using allometry. Whereas empirical allometry can relate tree height to DBH (e.g., Chave et al., 2005) using formulae (Chojnacky et al., 2014), modelling is subject to data quality and allometry commonly uses destructive sampling approach (Kizha and Han, 2016). Furthermore, studies often overlook small scale tree stands (Mugabowindekwe et al., 2024) within semi-arid lands. Nevertheless, field measurements remain indispensable for modelling tree characteristics (Ploton et al., 2020).

Unlike fieldwork that is constrained by limited spatial coverage, high cost, and human error, remote sensing (e.g., terrestrial laser scanning and LiDAR sensors) has wide spatial coverage (Vähä-Konka et al., 2024; Ghosh, 2018; Kumar and Mutanga, 2017; Bastin et al., 2015; Mertz et al., 2012). However, airborne LiDAR data useful to estimate tree height is costly (Stone et al., 2016). Therefore, since modern fieldwork instruments such as hypsometers (Guenther et al., 2024; Larjavaara and Muller-Landau, 2013) and terrestrial laser scanners (Calvert, et al., 2024; Wardius and Hein, 2024) have improved accuracy and non-destructive, they have become the golden standard for tree height measurement. As such, to understand the influence of FRSM on tree biophysical characteristics, we used a modern hypsometer that uses laser-pointing technology based on the sine method (Larjavaara and Muller-Landau, 2013) to collect *Z. mauritiana* tree height inside and outside a semi-arid floodplain for comparison.

The Indian jujube (*Ziziphus mauritiana*), known as Musawu in Zimbabwe, is part of the of the approximately 170 tree-crop species grown in many parts of the globe with approximately 20,000 ha cultivated in India (Pareek and Yahia, 2013). Musawu is a valuable tree species providing multiple uses (Goyal et al., 2012). Specifically, it produces a multipurpose nutritious 'super-fruit' (Shahrajabian et al., 2019; Arndt et al., 2001) that has tasty pulp, eaten fresh or dry, and is

an important food additive in Asian diets (Hussain et al., 2021). It has ethnomedicinal (Shahrajabian et al., 2020) and pharmacological value (Jha et al., 2023; Siddiqui et al., 2023; Butt et al., 2021; Mongalo et al., 2020; Lim and Lim, 2013). For instance, its extracts are processed into anticancer drugs (Khurshid et al., 2021). Mokgolodi et al. (2011) notes that the tree is medically underutilized in Africa. Musawu fruit provides more proteins, vitamins, and mineral nutrients than mango and apple, more phosphorus and iron than orange (Pareek and Yahia, 2013) and sugars (Muchuweti et al., 2005). In Zambia, the Musawu fruit dried pulp is baked into a cake (Kalikiti, 1998), while in southern African countries, it is brewed into alcoholic beverage (Rampedi and Olivier, 2013; Nyanga et al., 2008; Siddiqui et al., 2023). In Zimbabwe, it is used to produce fruit juices (Nemapare et al., 2023; Chivero and Mawadza, 1998), jam spread (Maposa and Chisuro 1998) and traditional bread (Kadzere, 1998). Its leaves are an important stockfeed (Maruza et al., 2017; Funkhouser, 1999). These multiple uses make the fruit a viable economic product (Saied et al., 2008). Furthermore, like other trees in the semi-arid landscapes, Musawu provides shade for families and community meetings, cultural gatherings, and religious activities (Saied et al., 2008; Funkhouser, 1999). Its biophysical features can be useful in understanding the ecological importance of FRSM.

Given the importance and projected increase in occurrence of floods (IPCC, 2019), the aim of this study was to assess the influence of FRSM on tree biophysical characteristics using *Z. mauritiana* as a proxy. Comparing the biophysical characteristics between FRSM sustained locations inside the floodplain (flood treatment) as opposed to those outside (control), the floodplain respectively facilitates isolation of the effect of floods on trees structure. Specifically, we measured the height, DBH, and canopy diameter, sampling *Z. mauritiana* trees from locations both inside and outside the floodplain in the middle Zambezi valley, northern Zimbabwe and compared their respective biophysical features. The study hypothesized that there is a significant difference in tree size between the flooded and non-flooded locations. Specifically, we investigated the long-term influence of the FRSM to sustain *Z. mauritiana's* higher biomass and productivity in a floodplain using biophysical variables. Implications from this study are important to our understanding of the potential ecological influence of flooded trees in mitigating global warming. The study informs research seeking to quantify the social, ecological and economic value of trees in arid and semi-arid floodplains, the global carbon budget and climate change policy.

6.2. Materials and methods

6.2.1 Study sites

The study area is a gently sloping, low elevation and highly drained terrain located between 29°E - 31°E and 15°S - 17°S floodplains in northern Zimbabwe (Figure 6.1). Mbire and Muzarabani floodplains are susceptible to both floods and droughts (Kasimba et al., 2023; Pedzisai et al., 2023; Mavhura et al., 2022). The area is a semi-arid tropical savanna characterized by low annual rainfall and moderately cold-dry winter and hot-wet summer seasons (Mavhura et al., 2022). The area is drained by two main rivers; Musengezi and Hunyani on the eastern and western sides, respectively. Both rivers originate from the escarpment with higher precipitation and drain into the Kahora Bassa Dam located on the Zambezi River. Two forms of floods occur in the area. Firstly, during summer season, floods are triggered by intense rainfall activity, such as tropical cyclones, often concentrated in either a few months or in few episodic events. Secondly, during winter season, floods occur due to backflow when the Cahora Bassa Dam is full with limited discharge from the two named rivers (Chingombe et al., 2015). Conversely, extended periods of high temperatures, and windy conditions often trigger droughts leading to high evaporation that depletes soil moisture outside the floodplain. This area has grasslands and forests with trees located inside and outside the floodplain. Consequently, the wet and dry extremes influence trees in the semi-arid area differently, with floodplain trees benefiting from episodic floods, while trees outside the floodzone are often subjected to prolonged soil moisture deficit. The area is a sparsely populated rural setting with crop and livestock rearing activities. Farmers till inside the floodplain with depositional silt and outside the floodplain. The study area was spatially delimited by floodplain extent to compare of flooded versus non-flooded *Z. mauritiana* tree biophysical characteristics using canopy diameter, tree height and DBH as growth proxies.

6.2.2 Field Data

The study sought to assess the effect of FRSM on tree biophysical characteristics. Therefore, field data on tree height, canopy diameter, diameter at breast height (DBH) and flood status for sampled trees (n = 366) were collected by the candidate with assistance of a driver and guided by few local community members during a field campaign between 2017/08/20 and 2017/09/06. The flood status annotation was informed by community members during fieldwork, while GPS coordinates were captured using a hand-held Garmin eTrex 10J device (Garmin Ltd, USA). The flood status for individual samples was coded as either flooded or non-flooded for trees inside and outside regularly flooded (floodplain) locations, respectively. The canopy radius was

estimated using a tape measure by estimating distance from the trucnk to the edge of one side of the canopy. Since radius on a circle is half of the diameter, this measurement was thus doubled to determine canopy diameter (CD), assuming a circular canopy for savanna trees as used by Mugabowindekwe et al. (2022). Equation 6.1 shows how CD was measured after determining canopy radius;

$$CD = 2*\text{canopy radius} \quad (6.1)$$

The study used a hypsometer (Haglöf Vertex Laser Geo, Sweden) to measure tree height. The circumference was measured using a tape measure encircled around the tree trunk at 1.3 m above ground (Guenter, 2024; Bastian et al., 2015) to determine the DBH as shown by Equation (6.2);

$$DBH = \text{Circumference}/\pi \quad (6.2)$$

where: DBH is the diameter at breast height (cm) and $\pi = 22/7$. To accurately estimate tree allometric AGB, fieldwork measurements of DBH and tree height are necessary (Chae et al., 2014, Junker, 2017).

The study extracted distance to river and distance to floodplain for both flooded and non-flooded area tree samples using a vector-GIS field calculator within QGIS (3.6.11) tool 'overlay_nearest' for distance to river polyline executing the code;

```
distance( $geometry, nearest_point(overlay_nearest(layer:='lines', expression:=$geometry)))   (6.3)
```

where: 'lines' is the name of the river layer. The same code was customized to determine shortest distance to floodplain boundary replacing the layer = 'lines' with 'polygon' (Equation 6.3). Both distance to river and distance to floodplain were extracted into spreadsheet and joined into a table with the field data measurements to enable analysis in R calling relevant packages.

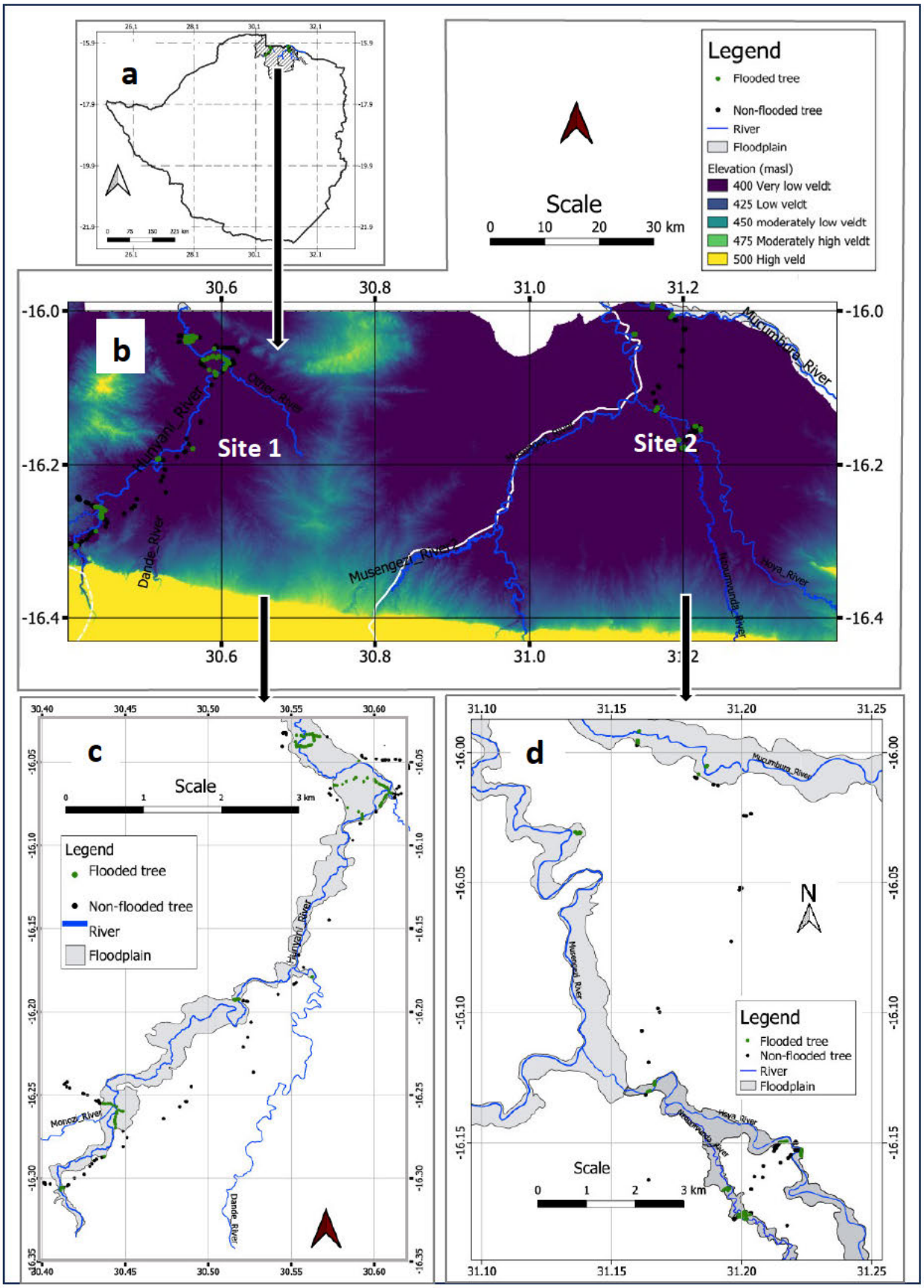


Figure 6.1: Location of study area in Zimbabwe (a), elevation and drainage in study area (b), with zoomed-in Mbire, Site 1 (c) and Muzarabani, Site 2 (d) study sites showing drainage, sample distribution for flooded (green) and non-flooded (black) of *Ziziphus mauritiana* trees.

6.2.3 Comparison of tree canopy diameter, tree height and DBH based on flood status

The study sampled the *Ziziphus mauritiana* trees located inside and outside the floodplains to determine the influence of FRSM on tree growth. The study used a two-stepped modelling method as shown on Figure 6.2. Firstly, we tested normality of the three *Z. mauritiana* data using the Shapiro-Wilk test (Shapiro and Wilk, 1965), and heteroscedasticity with the Breusch-Pagan test (Breusch and Pagan, 1979), and Rainbow linearity test (Utts, 1982). Secondly, upon confirming non-normality, Levene's test, nonparametric test for unpaired two sample test was relevant hence the Mann-Whitney-Wilcoxon rank sum test, biserial rank correlation tests used to compare flooded versus non-flooded trees. Thirdly, as post-hoc tests, the study used a pairwise Wilcoxon to conclude the significance of differences between flooded (hence FRSM) and non-flooded trees. Finally, the study applied multiple linear regression to explain the observed variability based on three spatial predictors namely, distance to river, distance to floodplain and flood status designated as either flooded or non-flooded. Due to local spatial scale, other factors such as precipitation, temperature, wind, and geology were held constant. These explanatory variables were selected since they have a close relation with the flooding, thus help to reinforce the importance of FRSM on tree biophysical characteristics. Secondly, the study applied linear regression with Bootstrapping to compare model coefficients.

6.2.4 Difference of biophysical characteristics between flooded and non-flooded trees

Firstly, the study tested the difference of the three biophysical characteristics based primarily on the flood status of the *Z. mauritiana* trees. The flood status was binary with flooded and non-flooded for inside inundated floodplain and outside locations, respectively. The study also compared the mean variances between flooded and non-flooded trees biophysical characteristics using the Levene's test (Levene, 1960) by running the script in car (3.1-2) library within R (4.3.3) (Equation 6.4);

leveneTest(data = model_data, canopy_diameter ~ flood_status, center = mean) (6.4)

where: data, canopy_diameter, flood_status, and center were the data (csv) file, response variable, predictor variable, and the measure of central tendency statistic (arithmetic mean) respectively. The study applied unpaired non-parametric Mann-Whitney-Wilcoxon rank sum test to assess the differences between flooded and non-flooded *Z. mauritiana* trees. It executed this test in which we computed the p-Value, biserial rank correlation, and confidence interval in one 'ggbetweenstats' function within ggstatsplot (0.12.4) library to produce a graphical visualisation (Appendix 5.2). A biserial correlation coefficient test was also important to compare differences between two independent samples. A higher coefficient shows a stronger relationship, while

zero, positive and negative biserial values show no significant difference, first group is greater than second group, and second group is greater than first group, respectively for a two-group design (Bonett, 2020). To further corroborate the differences between flooded and non-flooded trees, the study performed a post-hoc pairwise Wilcoxon test. This post-hoc Wilcoxon test computed a p-value adjusted using Benjamini-Hochberg method that uses a false discovery rate control (Benjamini and Hochberg, 1995). The same was applied for tree height and DBH.

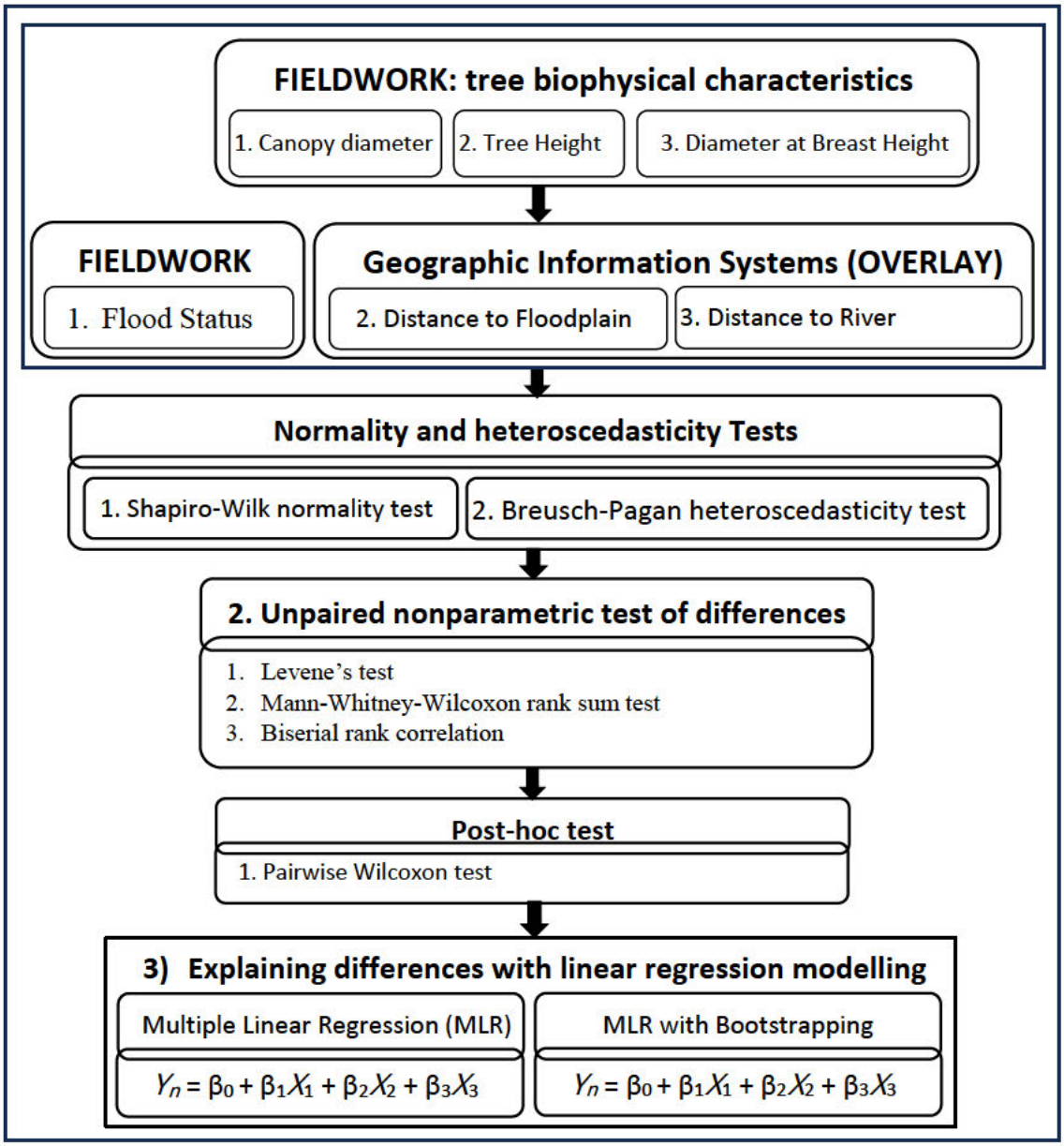


Figure 6.2: Framework to assess the difference influenced of FRSM on *Ziziphus mauritiana* tree biophysical characteristics in semi-arid floodplain, northern Zimbabwe.

6.2.5 Multiple linear regression model of *Z. mauritiana* tree growth on spatial predictors

To explain the variability between flooded and non-flooded area trees, the study performed multiple linear regression. Canopy diameter, tree height, and DBH were the response variables while distance to river, distance to floodplain, and flood status were the predictors. Specifically, we regressed one response tree variable, e.g., canopy diameter on the four predictors. The same was applied for tree height and DBH. The three predictors model coefficients were checked using the Bootstrapping algorithm with 5000 replications. Linear regression of canopy diameter, tree height and DBH on distance to river, distance to floodplain and flood status and testing for significance of predictors followed Equations (6.5);

$$Y_3 = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 \quad (6.5)$$

where: the Y_n is the model dependent tree biophysical variable (e.g. canopy diameter, tree height or DBH) and X_n is the predictor where n = 1, 2, 3 being distance to river, distance to floodplain and flood status, respectively. Each linear model shows a tree characteristic (Y) regressed on the predictor variables X_1 (distance to river) and X_2 (distance to floodplain) and X_3 (flood status: flooded/non-flooded), with β_1, β_2 and β_3 being the respective coefficients with β_0 (intercept). Variable X_3 was exclusively applicable to non-flooded trees. Linear regression models are sensitive to outlier influence, hence, diagnostic test is a golden standard (Stevens, 1984). Section 6.3 presents the results of the observed differences between flooded and non-flooded tree biophysical characteristics and their respective explanation through modelling using multiple linear regression on three selected spatial predictors. The script used to analyse the data is provided in Appendix 6.1.

6.3. Results

Results of exploratory descriptive analysis on canopy diameter, tree height, and DBH for *Ziziphus mauritiana* trees are presented in Figure 6.3. Firstly, the study computed the difference between trees in the two spatial locations; inside (flooded), and outside the floodplain (non-flooded). The repeatedly flooded trees are within FRSM, which influences tree growth as opposed to those in non-flooded areas constrained by droughts. Section 6.3.1 presents preliminary tests and Section 6.3.2 is a comparison between flooded with the non-flooded trees, with nonparametric tests of significance of differences. In the long-term, flooded trees within the FRSM location grow bigger than those in non-flooded areas, hence bigger canopy diameters, taller, and have larger DBH than those outside the floodplain (Figure 6.3). Figure 6.3 shows that

trees outside the floodplain (grey) depict smaller growth characteristics than those within the floodplain (blue). These results presented were analysed as outlined earlier in Section 6.2.4.

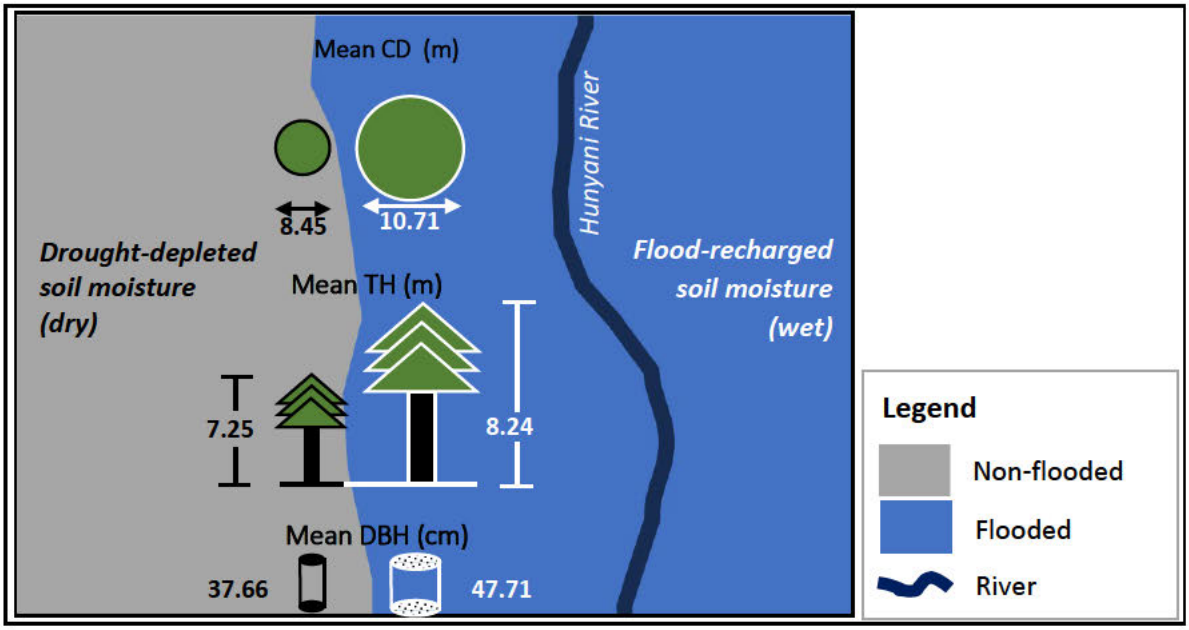


Figure 6.3: Differences of canopy diameter, tree height and DBH *Z. mauritiana* between flooded (inside) and non-flooded (outside) floodplains in semi-arid northern Zimbabwe.

6.3.1 Diagnostic plots for the three biophysical characteristics

Linear models were checked for significance using requisite normality, outlier and linearity diagnostic tests as suggested by Stephens (1984). Thus, all three regression models were evaluated in R using requisite packages with Shapiro-Wilk, Breusch-Pagan, and Rainbow tests for normality, heteroscedasticity, and linearity, respectively. Normality, and outlier diagnostic plots relating to the canopy diameter, tree height and DBH are shown on Figure 6.4.

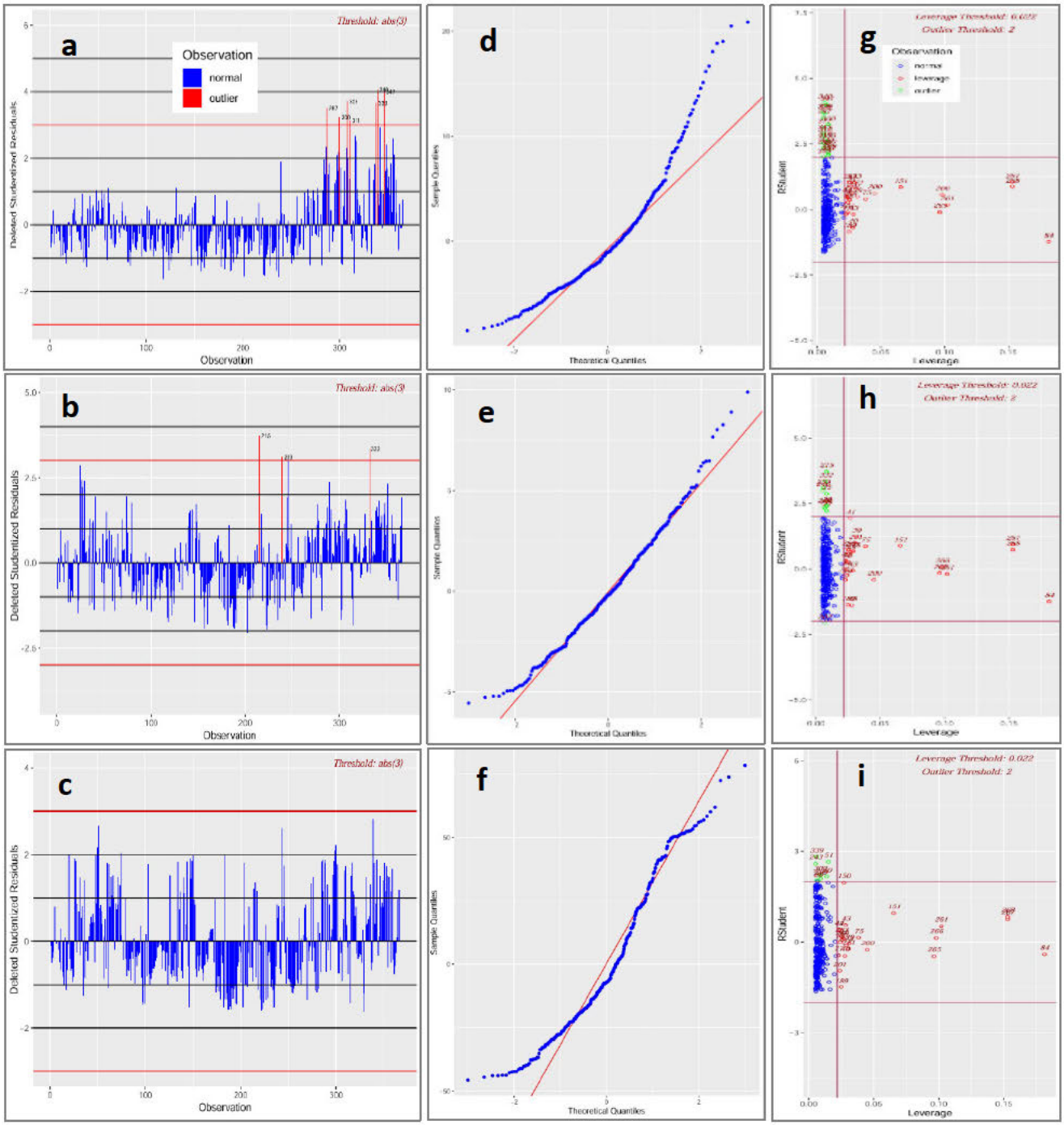


Figure 6.4: Diagnostic Studentised residuals (a-c), QQ (d - f), outlier and leverage (g - i) plot for canopy diameter, tree height and DBH in semi-arid floodplain, Zimbabwe.

The Studentised residual plots indicated a few deleted outliers in canopy diameter and tree height but none for DBH (Figure 6.4 a-c). The QQ plots (Figure 6.4 d-f) show that tree height is closer to normal distribution, unlike canopy diameter and DBH. The three plots (Figure 6.4 g- i) show positive outlier and leverage trees without due influence on the outcome in case linear modelling is undertaken. Therefore, all the above plots show that the normality assumption was not sustained. With that outcome, non-parametric tests for difference become applicable.

6.3.2 Comparison of *Z. mauritiana* tree biophysical characteristics by flood status

The flood status-based differences between flooded and non-flooded *Z. mauritiana* canopy diameter, tree height, and DBH were assessed. A quantitative description showing the median, mean, and maximum of the sampled trees is shown on Table 6.1.

Table 6.1: Comparing flooded (195) with non-flooded (171) *Z. mauritiana* trees.

Tree Biophysical Variable	Flood Status	Descriptive Statistical Metric		
		Median	Mean	Maximum
Canopy Diameter (m)	Flooded	9.00	10.71	32.00
	Non-flooded	*7.90*	*8.45*	*26.00*
Tree Height (m)	Flooded	8.10	8.24	17.80
	Non-flooded	*7.00*	*7.25*	*15.00*
Diameter at Breast Height (cm)	Flooded	40.00	47.52	126.00
	Non-flooded	*28.00*	*34.42*	*110.00*

As depicted on Table 6.1, all the three metrics for the three selected biophysical variables, flooded trees show that the median, mean and maximum canopy diameter, tree height and DBH exceed those of the non-flooded areas. Trees located inside the floodplain depict more growth characteristics as compared to those outside, as confirmed using non-parametric tests of significance of differences.

6.3.2.1 Canopy diameter differences between flooded and non-flooded Z. mauritiana trees

To assess the long-term influence of FRSM on trees, the canopy diameter was compared between flooded and non-flooded locations. Levene's test confirmed significant difference of variances ($p<0.001$) of canopy diameter between flooded and non-flooded trees. This result is confirmed on Figure 6.5, where flooded trees have bigger mean canopy diameter than those in non-flooded areas. Specifically, flooded trees showed a 1.10 m higher median and 6 m bigger maximum canopy diameter than the non-flooded. These differences between flooded and non-flooded trees were statistically significant using the Mann-Whitney-Wilcoxon test ($p<0.001$). Since the biserial rank correlation ($r = 0.22$) is positive, though weak, the first group (flooded) has bigger canopy diameter than the second (non-flooded).

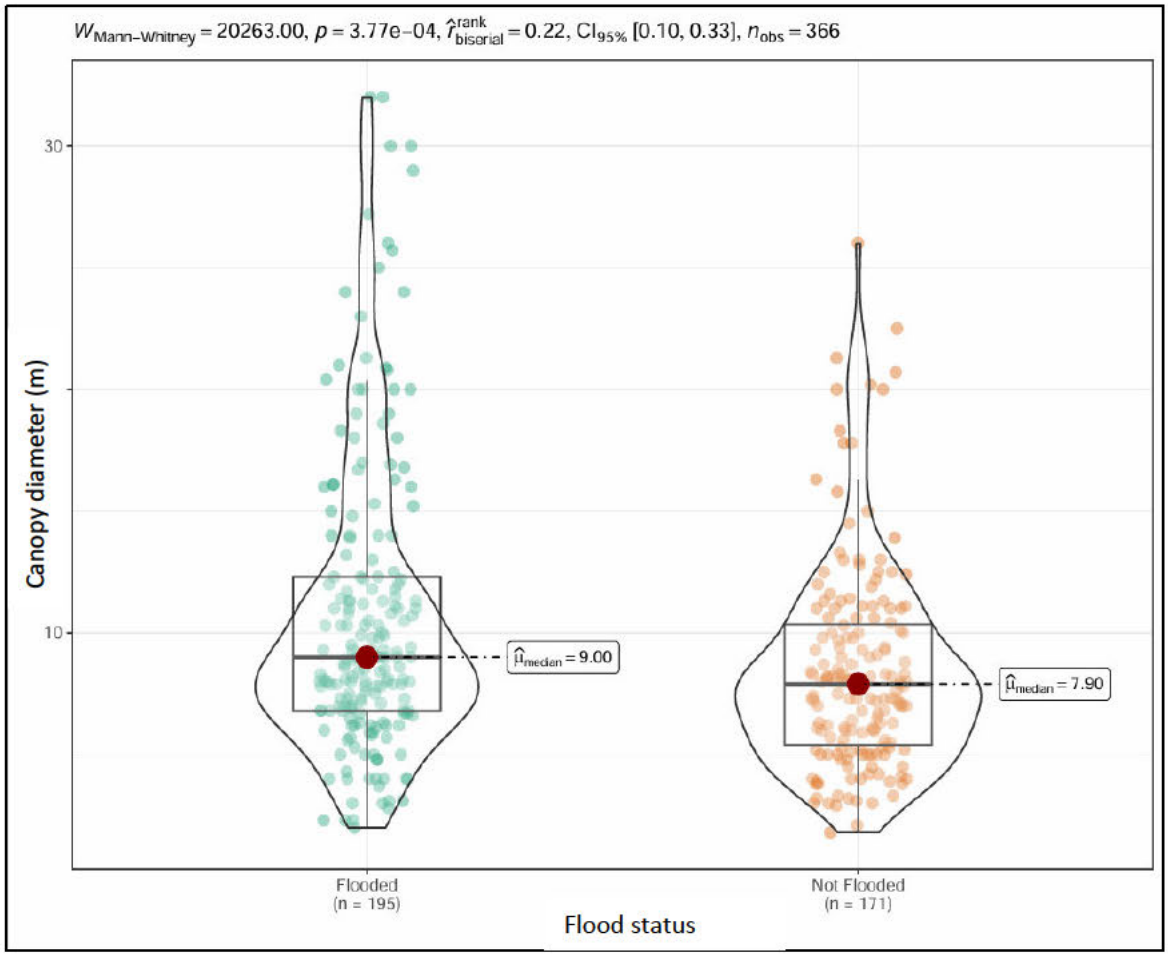


Figure 6.5: The boxviolin plots showing flooded and non-flooded *Z. mauritiana* tree canopy diameter, northern Zimbabwe.

Flooded trees show higher maximum, and median canopy diameter than those not flooded. This difference was statistically significant ($p<0.001$) as shown on Figure 6.4.

6.3.2.2 The tree height differences between flooded and non-flooded *Z. mauritiana* trees

Height was compared between flooded and non-flooded trees to enable assessment of the long-term influence of FRSM. Overall, tree height in flooded locations is higher than outside. This is shown by the differences in the median (1.10 m) and maximum (2.80 m) of tree height between flooded and non-flooded trees, which were confirmed to be statistically significant using the Mann-Whitney-Wilcoxon test ($p<0.001$). Furthermore, the positive biserial correlation coefficient confirmed that trees in flooded locations grow taller than those outside (Figure 6.6). However, the mean tree height inside and outside the floodplain is small. Similarly, the average tree height outside the floodplain is lower than those inside the floodplain. Furthermore, the

Levene's homogeneity of variances test also confirmed significant difference ($p<0.05$) in the tree height between flooded versus non-flooded samples.

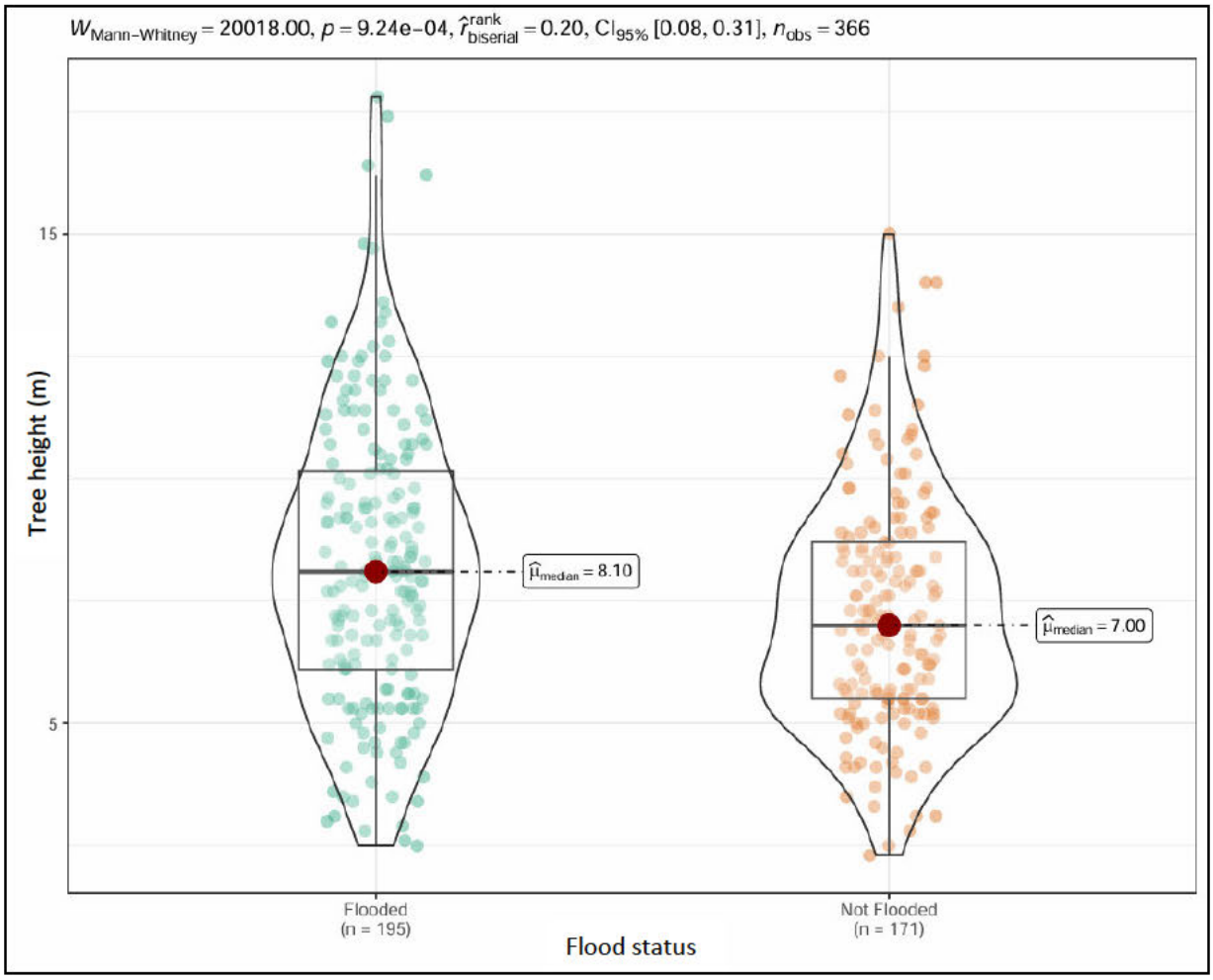


Figure 6.6: The boxviolin plots showing flooded and non-flooded *Z. mauritiana* tree height, northern Zimbabwe.

Figure 6.6 shows the boxviolin plots depicting the height differences for flood zone versus non-flood zone trees. Flood zone trees grow taller than non flood zone trees as confirmed using the Mann-Whitney-Wilcoxon test ($p<0.001$) and weak positive biserial correlation coefficient.

6.3.2.3 Comparing DBH differences between flooded and non-flooded *Z. mauritiana* trees

The tree trunk size, determined using the DBH, was an important determinant of tree biophysical characteristics used to compare trees inside floodplain with those outside. The range of the median (12 cm) and maximum (16 cm) in DBH between flooded and non-flooded trees was statistically significant using the Mann-Whitney-Wilcoxon test ($p<0.01$). Trees inside the floodplain have bigger trunk mean as compared to those outside. The Levene's homogeneity of variances test also confirmed a significant difference between samples in flooded versus those

non-flooded (p<0.001). The positive biserial rank correlation (r = 0.18) shows that the flooded trees have bigger trunks than those non-flooded areas (Figure 6.7).

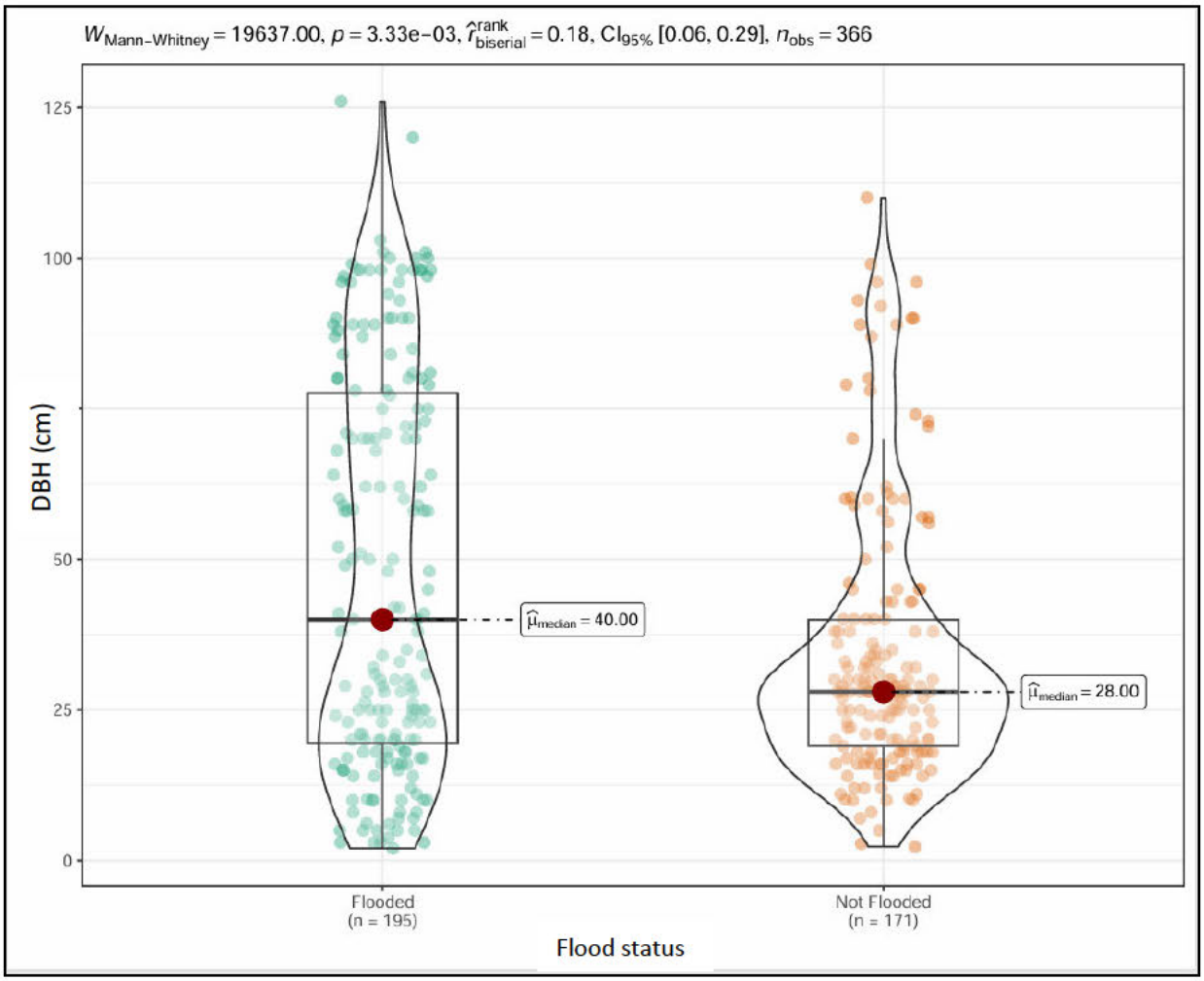


Figure 6.7: The boxviolin plots showing flooded and non-flooded *Ziziphus mauritiana* tree DBH, northern Zimbabwe.

To assess the influence of FRSM on the three biophysical characteristics, a Wilcoxon rank sum test on flood status was done. The result shows an overall significant statistical difference between flooded and non-flooded trees for canopy diameter (p<0.001), tree height (p<0.001) and DBH (p<0.01). Therefore, FRSM positively influences tree growth as illustrated using the Z. *mauritiana* tree. To further evaluate the above-noted differences in the selected tree biophysical characteristics, we present results of the pairwise Wilcoxon rank sum post-hoc test between flooded and non-flooded trees. The post-hoc Wilcoxon test confirmed significant differences between flooded and non-flooded trees for the three selected biophysical characteristics. For both canopy diameter and DBH, there was higher statistical significance (p<0.001) than for tree height (p<0.01) in the differences between the flooded and non-flooded trees.

6.3.3. Linear regression to model *Z. mauritiana* tree biophysical characteristics

To explain the differences noted between flooded and non-flooded area tree characteristics, multiple linear regression modelling was applied on selected spatial variables while holding other factors constant. The modelling of the relationships was between the three dependent variables as influenced by the three independent variables. Linear regression was used to separately model the canopy diameter, tree height and DBH tree biophysical characteristics based on distance to river, distance to floodplain, and flood status predictors. Canopy diameter, tree height and DBH were dependent variables using three independent factors namely distance to river, distance to floodplain for non-flooded trees exclusively and flood status to help deduce the effect of FRSM. Firstly, the canopy diameter biophysical variable was regressed on distance to river + distance to floodplain + flood status, hence one multiple regression linear model. The same was applied for tree height and DBH. Table 6.2 shows model coefficients and respective significance levels.

Table 6.2: Three predictors' model compared with bootstrapping coefficients

DV	Model coefficients (MC)				Bootstrapping (R = 5000) bootMed Coefficients (BC)				Absolute Coefficients Difference |BC – MC|			
	β_0	β_1FS(nf)	β_2DF	β_3DR	β_0	β_1FS(nf)	β_2DF	β_3DR	β_0	β_1FS(nf)	β_2DF	β_3DR
CD	11.505[1]	-1.639[2]	0.003[3]	-0.003[2]	11.516	-1.617	0.003	-0.003	0.011	0.022	0.000	0.000
TH	8.517[1]	-0.734[2]	0.001[5]	-0.001[3]	8.530	-0.723	0.001	-0.001	0.013	0.012	0.000	0.000
DBH	47.870[1]	-9.598[2]	-0.004[5]	-0.001[5]	47.950	-9.634	-0.004	-0.001	0.080	0.136	0.000	0.000

β_0 = coefficient of the intercept, β_1 = coefficient of first independent variable 'flood status (FS) category 'non-flooded' (nf); β_2 = coefficient of second independent variable for distance to floodplain (DF), β_3 = coefficient of third independent variable for distance to river (DR); CD = canopy diameter; DBH = diameter at breast height; DV = dependent variable; TH = tree height; significant codes: [1]0 – 0.001; [2]0.001 - 0.01; [3]0.01 – 0.05; [4]0.05 – 0.1; and [5]0.1 – 1.0.

For all three linear models, there was small (flood status) and no absolute difference (distance to river and distance to floodplain) between coefficients of the independent variables compared with the bootstrapping. Table 6.2, shows results of the linear models for the three dependent variables regressed on three independent variables. Table 6.2 shows smaller coefficients for β_0, and β_1 than β_2 and β_3. Thus, bootstrapping with 5000 replications test slightly improved the linear models. Non-flooded trees showed significant intercept ($p<0.001$) and negative predictor coefficients for distance to river and flood status, though with varying significance levels for the three models.

6.3.3.1 Pairwise linearity test

Linearity was tested using pairwise scatterplots with GGally (2.2.1) library in R for all the three biophysical characteristics in the study. The scatterplots show the distribution of canopy diameter with distance to river, distance to floodplain, and flood status, respectively. The canopy diameter depicts a decrease in size away from the two sources of soil moisture. The canopy diameter, tree height and diameter at breast height was plotted against distance to floodplain, distance to river and flood status (Figure 6.8).

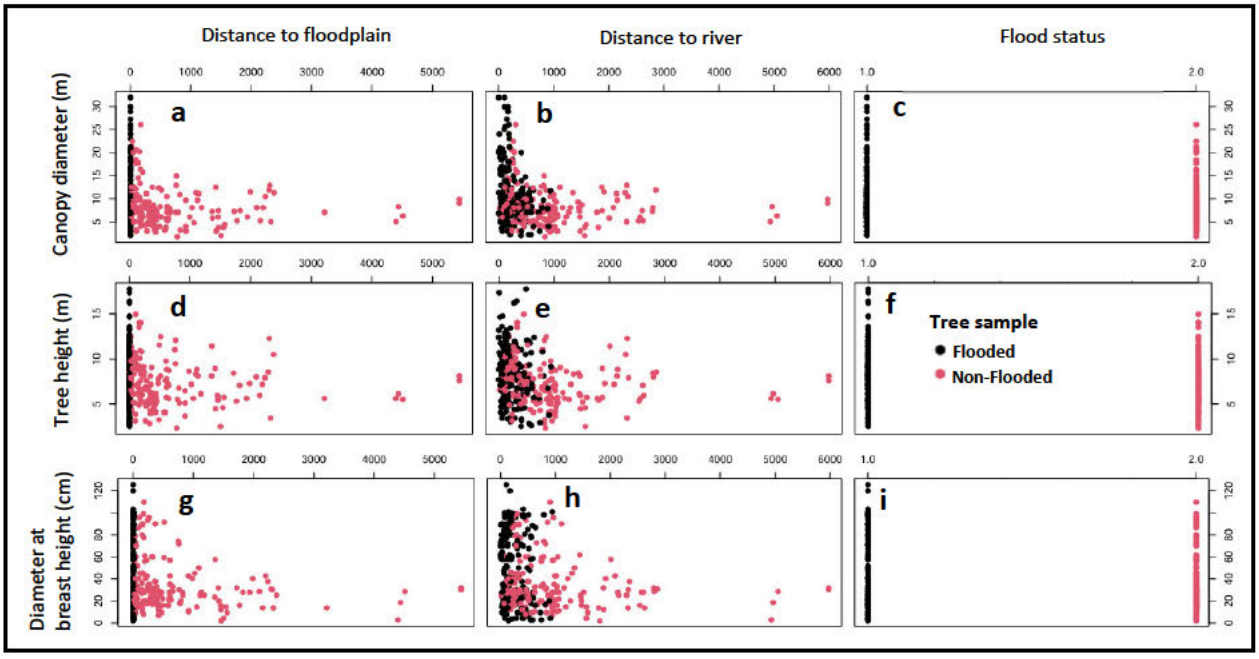


Figure 6.8: Scatterplots showing linearity of canopy diameter (a-c), tree height (d-f) and diameter at breast height (g-i) plotted against distance to floodplain (left), distance to river (middle) and flood status (right), depicted for flooded and non-flooded *Z. mauritiana* trees inside (black) and outside (red) floodplains of northern Zimbabwe.

The canopy diameter (m) (Figure 6.8a-c), tree height (m) (d-f) and DBH (cm) (g-i) in the flooded against and non-flooded was plotted against distance to floodplain, distance to river and flood status. The linearity is not very distinct in the scatterplots as was the case in the canopy diameter. Trees were discriminated by flood status, with those inside the floodplain treated as the experiment, while those on the outside were the control samples. The river and the floodplain are important sources of soil moisture in the study area. As such, linear regression of canopy diameter on distance to river and distance to floodplain were modelled empirically. Results show that the canopy diameter decreases with increasing distance from the river within and outside of the floodplain (Figure 6.8a). With regards to trees located outside the floodplain, increasing distance from floodplain also shows a decrease in canopy diameter (Figure 6.8a), while those

inside the floodplain are depicted as single brown dotted linear samples. The results show that trees within the floodplain have bigger canopy diameter than those outside. Pairwise scatterplots show tree characteristics regressed on distance to river, distance to floodplain and flood status. The maximum canopy diameter, tree height and DBH values are larger for trees inside floodplain, which are flooded than outside. All sample points within the floodplain for distance to floodplain are on the zero distance, while those outside have notable distances. Pairwise scatterplots for canopy diameter, tree height and diameter at breast height with distance to floodplains, distance to river and flood status were also tested using the linearity test. Table 6.3 shows statistically significant linear models though with variable heteroscedasticity, not normal, and not linear with very small predictor coefficients. All flood status-based tests are in bold text in Table 6.3.

Table 6.3: Statistical significance test for heteroscedasticity, normality, linearity using single predictor linear regression models

Model	**Breusch-Pagan test (*Heteroscedasticity)***		**Shapiro-Wilk test (*Normality)***		**Rainbow test *(Linearity)***	
	X^2	**Prob > X^2**	**W**	**p-Value**	**Rain**	**p-Value**
Canopy diameter ~ Flood Status	**21.947**	**2.30e-06[1]**	**0.886**	**6.83e-16[1]**	**5.317**	**<2.2e-16[1]**
Canopy diameter ~ Distance to floodplain	8.546	3.46e-03[2]	0.877	<2.2e-16[1]	5.167	<2.2e-16[1]
Canopy diameter ~ Distance to river	13.200	2.90e-04[1]	0.885	5.92e-16[1]	5.044	<2.2e-16[1]
Tree height~ flood status	**6.998**	**8.16e-03[2]**	**0.983**	**2.60e-04[1]**	**1.441**	**7.07e-03[2]**
Tree height ~ Distance to floodplain	2.830	9.25e-02[4]	0.980	6.64e-05[1]	1.340	2.46e-02[3]
Tree height ~ Distance to river	1.905	1.68e-01	0.980	7.10e-05[1]	1.311	3.45e-02[3]
DBH ~ Flood status	**22.209**	**2.45e-06[1]**	**0.945**	**1.85e-10[1]**	**1.574**	**1.19e-03[2]**
DBH ~ Distance to floodplain	10.231	1.38e-03[2]	0.921	5.24e-13[1]	1.529	2.19e-03[2]
DBH ~ Distance to river	7.929	4.87e-03[2]	0.918	2.80e-13[1]	1.510	2.83e-03[2]

Significant codes: [1]0 – 0.001; [2]0.001 - 0.01; [3]0.01 – 0.05; [4]0.05 – 0.1; and [5]0.1 – 1.0.

6.3.4.2 Linear model of *Z. mauritiana* tree biophysical characteristics by flood status

Table 6.4 shows linear models of tree biophysical characteristics (Y) regressed on distance to river and floodplain with coefficients β_1 and β_2 respectively, while β_0 term is for intercept.

Table 6.4: Linear regression models of biophysical features on distance for flooded, non-flooded and both (global) for *Z. mauritiana* tree measurements (n = 366).

Location	Dependent variable (Y)	Model Coefficients[significance code]		
		β_0	β_1	β_2
Flooded	Canopy diameter (m)	12.486[1]	-0.008[1]	**N/A**
Non-flooded		10.980[1]	-0.004[1]	0.003[3]
Global		9.220[1]	-0.001[5]	0.001[5]
Flooded	Tree height (m)	8.6714[1]	0.002 [4]	**N/A**
Not flooded		8.2815[1]	-0.001[3]	0.001[5]
Global		7.6803[1]	-0.001[5]	0.001[5]
Flooded	Diameter at breast height (cm)	51.1579[1]	-0.015[5]	**N/A**
Not flooded		45.6710[1]	-0.004 [5]	-0.0038[5]
Global		39.4816[1]	0.003 [5]	-0.008[5]

Significant codes: [1]0 -0.001; [2]0.001 - 0.01; [3]0.01 - 0.05; [4]0.05 - 0.1; and [5]0.1 - 1.0

β_0, β_1 and β_2 denote the intercept, coefficients of distance to river (X_1) and distance to floodplain (X_2), respectively. The flooded trees showed higher statistically significant intercept (β_1), ($p<0.001$) than for coefficient (β_1) for X_1 (Table 6.4).

6.4 Discussion

This study sought to assess the long-term influence of FRSM on tree biophysical characteristics. Firstly, the study compared growth characteristics between flooded and non-flooded trees using nonparametric Mann-Whitney-Wilcoxon, Levene and post-hoc Wilcoxon tests to quantify the long-term influence of FRSM on trees. Secondly, the study assessed the strength of the linear relationship between canopy diameter, tree height and DBH with distance to river and floodplain and flood status. Small floods that are highly frequent, and connected have positive important ecological effects. Thus, the difference between flooded (hence FRSM) and non-flooded *Z. mauritiana* tree growth characteristics was statistically significant while linear regression was essential to explain how increasing distance to river and floodplain negatively influenced the tree biophysical characteristics to prove the importance of FRSM.

6.4.1 Overall influence of FRSM on tree biophysical characteristics in semi-arid areas

Comparing results of trees in the flooded with non-flooded areas revealed that the former depicted more growth than the latter (see Figures 6.3; Table 6.1). Flooded locations depict more biomass expressed in the growth, hence larger tree size characterized by canopy diameter (Figure 6.5), tree height (Figure 6.6) and DBH (Figure 6.7). Ultimately, trees inside the floodplain have

bigger canopies (Figure 6.5) are taller (Figure 6.6), and thicker trunks (Figure 6.7) than outside. Semi-arid floodplains are perturbed by drought and floods, hence unique in character (Colloff and Baldwin, 2010). Th FRSM lasts longer inside the floodplain than outside (Pedzisai et al., 2022), hence supporting productivity for longer duration. Ma et al. (2021) confirmed that soil moisture promotes tree primary productivity, while Swemmer et al. (2023) noted that in riparian semi-arid floodplains, soil moisture is an important variable for tree survival. The same is true for trees inside floodplain of semi-arid areas benefiting from FRSM through deep infiltration.

However, FRSM is only one of the variables influencing tree growth. Ma et al. (2021) for instance, noted that high temperature, soil water and nutrients are important variables promoting plant growth. As such, soil moisture and nutrients are often high in floodplains because of silt and leachate deposition (Tsheboeng et al., 2014). Therefore, flooding promotes soil moisture recharge while silt deposits are rich in nutrients (Gordon et al., 2020; Tsheboeng et al., 2014; Rinklebe et a., 2007), promoting enhanced *Z. mauritania* tree growth within than outside the floodplain. Goebe et al. (2012) noted that frequently occuring smaller floods promote growth of vegetation than high magnitude low frequency events.

6.4.1.1 The FRSM influence on canopy diameter

Floodplain trees benefit from FRSM. As such, the trees inside the periodically flooded floodplain depicted more growth in canopy diameter than those outside affected by flash drought (Figure 6.3 and 6.4). In the long-term, trees within the floodplain have bigger canopies than those outside (Table 6.1). Therefore, trees inside the floodplain depict mesic conditions, while those on the outside show aridity, yet they exist in the same microclimate, except soil moisture. Drought-induced vegetation stress is influenced by background climate (Sungmin and Park, 2024) in which FRSM lasts longer in areas outside floodplains than outside (Pedzisai et al., 2022**).** A high soil moisture stress was observed in arid regions (Sungmin and Park, 2024). Pedzisai et al. (2022) also noted that in the short-term, trees inside floodplain depict a longer green-up phase, hence have a longer productivity window than those outside. The annual cycle with a longer wet season and productive phase leads to trees inside floodplains to outgrow those outside. Repetitive and less intense inundation is advantageous to trees inside the floodplain.

While trees such as those outside the soil-moisture-rich floodplains can adapt to overcome low soil moisture content through, among others, drought avoidance and tolerance root traits (Brunner et al., 2015), in this study, we found a distinct difference between flooded versus non-flooded area tree characteristics (Figures 6.3, and Table 6.4). Talbot et al. (2018) noted that frequent floods promote net primary productivity, while Brunner et al. (2015) noted that drought

affected trees tend to develop more roots to tap more soil moisture compared to shoots than those in water sufficient or surplus areas. Some trees outside the floodplain were observed to have shed their leaves as a water conservation measure as compared to those with excess water inside floodplain that had larger leaves. Similarly, Swemmer et al. (2023) argued that floods and droughts in semi-arid floodplains are important determinants for tree persistence, as demonstrated using the *Z. Mauritiana* species in northern Zimbabwean floodplains in Mbire and Muzarabani districts. Seasonal floods are better temporally connected than large ones and have positive ecological benefits (Talbot et al., 2018).

In this study, FRSM within and outside semi-arid floodplains was assessed. Trees interact with soil moisture through roots (Sungmin and Park, 2024), which constitute ~40% of the tree biomass (Jackson et al., 1997). Roots help trees avoid and tolerate different soil moisture levels (Brunner et al., 2015). Within the tree root system, the xylem vessels transport water up the tree (Grote et al., 2016). Hence, Kühnhammer et al. (2023) deduced that deep roots of tropical trees reach up to 3.3m deep to mitigate the effects of droughts by either stress avoidance or tolerance. The tree height, and canopy size are important determinants of tree adaptation to water demand and supply, hence influence growth or mortality (Grote et al., 2016). Droughts are noted to facilitate succession of linked effects (Brunner et al., 2015), while floods influence distinct tree size biophysical structure (Pedzisai et al., 2022). The difference is due to a decrease in soil moisture, which in turn decreases vegetation productivity and growth (Sungmin and Park, 2024). In semi-arid floodplains, soil moisture is additionally recharged by floods, while outside, it is precipitation-only recharged. The latter is rapidly depleted by flash droughts. Thus, trees in semi-arid floodplains depict the influence of higher soil moisture as compared to the outside locations.

6.4.1.2 Influence of FRSM on tree height in semi-arid areas

Tree height is an important indicator of tree growth that is sustained by soil moisture and nutrient supply. Additionally, competition for sunlight influences tree height, hence phototropic growth (Lang et al., 2010). Trees inside floodplain are taller in comparison to those outside the floodplain (Figure 6.7; Table 6.1). Observed tree height results confirm that the FRSM inside floodplain is important for tree growth (Pedzisai et al., 2022). Additionally, availability of mineral nutrients due to silt deposition also promotes taller trees inside the floodplain (Gordon et al., 2020; Rinklebe et al., 2007). Tree growth shows vitality and response to stresses from the environment (Dobbertin, 2005). However, Grote et al. (2016) noted that while roots are important for water uptake, tree height influences stomatal sensitivity, xylem and mesophyll conductance. Therefore, tree height both influences and is influenced by soil moisture. As such, the near annual flood

pattern (Nharo et al., 2019) observed supports the finding that floodplain located trees are taller and have higher structural characteristics distinguishable from those on the inside.

6.4.1.3 The influence of FRSM on DBH in semi-arid areas

The effect of FRSM on tree biophysical characteristics was also demonstrated using DBH. Results have shown that trees inside floodplain depict clearly distinct bigger DBH (Figures 6.7; Tables 6.1). This was probably due to soil moisture differences. Floodplain trees benefit from the repetitive flood episodes, while moisture scarcity is aggravated by excessive heat, scanty precipitation, hence soil moisture deficit in the areas outside the floodplain. As summarized on Figure 6.3, and Table 6.1, the DBH, among other features for floodplain trees exceed that of the trees outside. The floodplain trees benefit from FRSM, which persists for a longer duration (Pedzisai et al., 2022), hence the noted disparity in DBH.

While Sungmin and Park (2024) noted that flash droughts for instance deplete soil moisture faster, Pedzisai et al. (2022) found that FRSM lasted longer than outside flooded locations in semi-arid landscapes. The soil moisture-deficit-excess dichotomy manifests in the biophysical characteristics as shown using *Z. maurutiana* in northern Zimbabwean floodplains. Talbot et al. (2018) noted that frequent, and low magnitude floods have more ecological benefits and less negative impacts than low frequency high magnitude floods. The small events are usually more frequent and better connected to contribute to sustained soil moisture supply, hence floodplain trees become better adapted to higher soil moisture facilitated by the deep infiltration. The study has demonstrated that regular episodes of floods as noted in the study area by Chingombe et al. (2015) are essential to promote tree productivity in floodplains of semi-arid areas while droughts retard growth.

6.4.2 Relationship between tree biophysical characteristics and soil moisture predictors

Linear regression of canopy diameter, tree height and DBH on distance to river, distance to floodplain and flood status showed that all the three biophysical characteristics are negatively related to these predictor variables, especially for non-flooded trees. This was true for all scenarios albeit with variable statistical significance (Table 6.3). Apparently, tree growth features decrease with increasing distance to the river and distance to floodplain. For trees outside the floodplain, distance to floodplain was statistically significant ($p<0.01$). The distance to river was less significant in comparison to distance to floodplain, particularly for trees outside floodplain.

However, Heklau et al. (2019) noted that, contrary to promoting tree growth, flooding can also lead to anoxic conditions which promote toxic ethanol production in roots removed through leaves by transpiration. Furthermore, anoxic condition is species-specific (Kreuzwieser et al., 2004). However, this study only focused on one tree species, hence further exploration of this phenomenon is necessary.

6.5 Conclusion

The study assessed the influence of flood-recharged soil moisture to the *Z. mauritiana* tree species biophysical structure on semi-arid northern Zimbabwe. We used Levene's, non-parametric Mann-Whitney-Wilcoxon and biserial correlation tests to compare tree growth characteristics inside periodically flooded areas with those outside to draw the following conclusions;

1. Canopy diameter, tree height, and DBH for trees within the floodplain were significantly greater than those outside the floodplain.
2. Flooded trees (inside the floodplain) are sustained with FRSM, hence grow more than those outside that are subjected to flash droughts.
3. Distance to river and floodplain were crucial sources of soil moisture linked to floods, hence FRSM, were negatively related to the three *Z. mauritiana* tree characteristics in semi-arid sites.

6.6 Recommendations

Understanding the influence of FRSM on trees is important to various fields of applications. As such, further research is necessary to understand the influence of FRSM on tree biophysical characteristics in semi-arid landscapes. Hence, the following recommendations are suggested;

1. More field data may be necessary to provide more insights to facilitate follow-up studies to check on the consistency of the findings of this study.
2. It is ideal to test the differences using different species within and outside the floodplains of the study area.
3. There is need to separate the respective growth influence of rain season floods from winter backflow floods.

6.7 References

Alifu, H., Hirabayashi, Y., Imada, Y. and Shiogama, H., 2022. Enhancement of river flooding due to global warming. *Scientific Reports 12*(1), p.20687.

Arndt, S.K., Clifford, S.C. and Popp, M., 2001. Ziziphus - A multipurpose fruit tree for arid regions. In *Sustainable land use in deserts* (pp.388 - 399). Springer Berlin Heidelberg.

Bastin, J.F., Barbier, N., Réjou-Méchain, M., Fayolle, A., Gourlet-Fleury, S., Maniatis, D., De Haulleville, T., Baya, F., Beeckman, H., Beina, D. and Couteron, P., 2015. Seeing Central African forests through their largest trees. *Scientific Reports 5*(1), p.13156.

Benjamini, Y., and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*. 57 (1), pp. 289 - 300.

Betts, M.G., Wolf, C., Ripple, W.J., Phalan, B., Millers, K.A., Duarte, A., Butchart, S.H. and Levi, T., 2017. Global forest loss disproportionately erodes biodiversity in intact landscapes. *Nature* 547 (7664), pp.441 - 444.

Blanchard, E., Birnbaum, P., Ibanez, T., Boutreux, T., Antin, C., Ploton, P., Vincent, G., Pouteau, R., Vandrot, H., Hequet, V. and Barbier, N., 2016. Contrasted allometries between stem diameter, crown area, and tree height in five tropical biogeographic areas. *Trees 30*, pp.1953 - 1968.

Bonett, D.G., 2020. Point-biserial correlation: Interval estimation, hypothesis testing, meta-analysis, and sample size determination. *British Journal of Mathematical and Statistical Psychology 73*, pp.113 - 144.

Breusch, T.S. and Pagan, A.R., 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society* pp.1287 - 1294.

Brunner, I., Herzog, C., Dawes, M. A., Arend, M. and Sperisen, C., 2015. How tree roots respond to drought. *Frontiers of Plant Science* 6, p.547.

Butt, S.Z., Hussain, S., Munawar, K.S., Tajammal, A. and Muazzam, M.A., 2021. Phytochemistry of Ziziphus Mauritiana; its nutritional and pharmaceutical potential. *Scientific Inquiry and Review 5*(2), pp.1 - 15.

Calvert, J., Yatsko, A.R., Bresgi, J., Cheesman, A.W., Cook, K., Crowe, J., Gambold, I., Jones, C., O'Connor, L., Peter, T. and Russell-Smith, P., 2024. Modelling internal stem damage in savanna trees: Error in aboveground biomass with terrestrial laser scanning and allometry. *Methods in Ecology and Evolution*.

Chave, J., Andalo, C., Brown, S., Cairns, M.A., Chambers, J.Q., Eamus, D., Fölster, H., Fromard, F., Higuchi, N., Kira, T. and Lescure, J.P., 2005. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia 145(1)*, pp.87 - 99.

Chivero, E., and C. Mawadza. 1998. Traditional processing of Ziziphus mauritiana fruits in Zimbabwe. Proc. Intl. Workshop on Ziziphus mauritiana, Harare, Zimbabwe. pp.75 - 78.

Chojnacky, D. C., Heath, L. S. & Jenkins, J. C., 2014. Updated generalized biomass equations for North American tree species. *Forestry* 87(1), pp.129 - 151.

Colloff, M.J. and Baldwin, D.S., 2010. Resilience of floodplain ecosystems in a semi-arid environment. *The Rangeland Journal* 32(3), pp.305 - 314.

Dahan, O., Tatarsky, B., Enzel, Y., Kulls, C., Seely, M. and Benito, G., 2008. Dynamics of flood water infiltration and ground water recharge in hyperarid desert. *Groundwater* 46(3), pp.450 - 461.

Disney, M., Burt, A., Wilkes, P., Armston, J. and Duncanson, L., 2020. New 3D measurements of large redwood trees for biomass and structure. *Scientific Reports* 10(1), p.16721.

Dlamini, C., and Geldenhuys, C.J., 2009. The socioeconomic status of the non-timber forest product subsector in Swaziland. *Southern Forests* 71(4). DOI:10.2989/SF.2009.71.4.9.1036.

Dobbertin, M., 2005. Tree growth as indicator of tree vitality and of tree reaction to environmental stress: a review. *European Journal of Forest Research* 124, pp.319 - 333.

Fischer, S., Greet, J., Walsh, C.J., and Catford, J.A., 2021. Restored river-floodplain connectivity promotes woody plant establishment. *Forest Ecology and Management* 493, 119264. doi:10.1016/j.foreco.2021.119264.

Funkhouser, S.L., 1999. *Local community tree management of Ziziphus mauritiana: An exploratory study from the mid-Zambezi Valley, Zimbabwe*. Michigan State University.

Ghosh, D., 2018. Risky fieldwork: the problems of ethics in the field. *Energy Research and Social Science* 45, pp.348 - 354.

Gordon, B.A., Dorothy, O. and Lenhart, C.F., 2020. Nutrient retention in ecologically functional floodplains: A review. *Water* 12(10), p.2762.

Goyal, M., Nagori, B.P. and Sasmal, D., 2012. Review on ethnomedicinal uses, pharmacological activity and phytochemical constituents of Ziziphus mauritiana (Z. jujuba Lam., non Mill). *Spatula DD* 2(2), pp.107 - 116.

Grassi, G., House, J., Dentener, F., Federici, S., den Elzen, M. and Penman, J., 2017. The key role of forests in meeting climate targets requires science for credible mitigation. *Nature Climate Change* 7(3), pp.220 - 226.

Green, J.K. and Keenan, T.F., 2022. The limits of forest carbon sequestration. *Science* 376(6594), pp.692 - 693.

Grote, R., Gessler, A., Hommel, R., Poschenrieder, W. and Priesack, E., 2016. Importance of tree height and social position for drought-related stress on tree growth and mortality. *Trees* 30, pp.1467 - 1482.

Guenther, M., Heenkenda, M.K., Leblon, B., Morris, D. and Freeburn, J., 2024. Estimating Tree Diameter at Breast Height (DBH) Using iPad Pro LiDAR Sensor in Boreal Forests. *Canadian Journal of Remote Sensing* 50(1), p.2295470.

Gülci, S., Yurtseven, H., Akay, A.O. and Akgul, M., 2023. Measuring tree diameter using a LiDAR-equipped smartphone: a comparison of smartphone-and caliper-based DBH. *Environmental Monitoring and Assessment* 195(6), p.678.

Huang, H., Li, Z., Gong, P., Cheng, X., Clinton, N., Cao, C., Ni, W. and Wang, L., 2011. Automated methods for measuring DBH and tree heights with a commercial scanning lidar. *Photogrammetric Engineering and Remote Sensing* 77(3), pp.219 - 227.

Hussain, S.Z., Naseer, B., Qadri, T., Fatima, T. and Bhat, T.A., 2021. Ber/Jujube (Ziziphus mauritiana): Morphology, taxonomy, composition and health benefits. In *Fruits grown in highland regions of the Himalayas: Nutritional and health benefits* (pp. 157 - 168). Cham: Springer International Publishing.

Iizuka, K., Yonehara, T., Itoh, M. and Kosugi, Y., 2017. Estimating tree height and diameter at breast height (DBH) from digital surface models and orthophotos obtained with an unmanned aerial system for a Japanese cypress (Chamaecyparis obtusa) forest. *Remote Sensing*, *10*(1), p.13.

IPCC, 2019. 2019 refinement to the 2006 IPCC guidelines for national greenhouse gas inventories. https://www.ipcc.ch/report/2019-refinement-to-the-2006-ipcc-guidelines-for-national-greenhouse-gas-inventories.

Jackson, R. B., Mooney, H. A., and Schulze, E. D., 1997. A global budget for fine root biomass, surface area, and nutrient contents. *Procedings of the National Academy of Sciences, U.S.A.* 94, pp.7362 - 7366. doi: 10.1073/pnas.94.14.7362.

Jha, D., Hangargekar, P., Akbar, M., Parihar, A.S., Kashyap, S., Joshi, A. and Rahman, M.A., 2023. Ziziphus mauritiana: An in-depth review of its medicinal attributes and pharmacological activities. *Intelligent Pharmacy*.

Jiang, M., Medlyn, B.E., Drake, J.E., Duursma, R.A., Anderson, I.C., Barton, C.V., Boer, M.M., Carrillo, Y., Castañeda-Gómez, L., Collins, L., Crous, K.Y., …… and Ellsworth, D.S., 2020. The fate of carbon in a mature forest under carbon dioxide enrichment. *Nature 580*(7802), pp.227 - 231.

Kadzere, I. 1998. Role of Ziziphus mauritiana in the livelihood of some communities in Zimbabwe. Proceedings of the International Workshop on *Ziziphus mauritiana*, Harare. Zimbabwe.

Kalikiti, F. 1998. Ziziphus mauritiana in Siavonga district, Zambia. Proceedings of the International Workshop on *Ziziphus mauritiana*, Harare. Zimbabwe.

Kasimba, R., Muqayi, S. and Chirisa, I., 2023. Towards a social capital resilience model in coping with floods and droughts: The case of Muzarabani, Zimbabwe. In *The Palgrave encyclopedia of urban and regional futures* (1850 - 1869). Cham: Springer International Publishing.

Khurshid, S., Awan, S.J., Naz, A., Hayat Khan, S. and Fiaz, S., 2021. Mechanism of anticancer activity of compounds isolated from two Species of Ziziphus (Z. jujube and Z. mauritiana). *Journal of Pharmaceutical Research International* 33(50A), pp.170 - 183.

Kizha, A. R. and Han, H.-S., 2016. Predicting aboveground biomass in second growth coast redwood: Comparing localized with generic allometric models. *Forests* 7, p. 96.

Körner, C., 2012. Treelines will be understood once the functional difference between a tree and a shrub is. *Ambio*, *41*(Suppl 3), pp.197-206.

Kreuzwieser, J., Papadopoulou, E. and Rennenberg, H., 2004. Interaction of flooding with carbon metabolism of forest trees. *Plant Biology* 6(03), pp.299 - 306.

Kühnhammer, K., van Haren, J., Kübert, A., Bailey, K., Dubbert, M., Hu, J., Ladd, S.N., Meredith, L.K., Werner, C. and Beyer, M., 2023. Deep roots mitigate drought impacts on tropical trees despite limited quantitative contribution to transpiration. *Science of the Total Environment* 893, p.164763.

Larjavaara, M. and Muller-Landau, H.C., 2013. Measuring tree height: a quantitative comparison of two common field methods in a moist tropical forest. *Methods in Ecology and Evolution* 4(9), pp.793 - 801.

Levene, H., 1960. *Robust Tests for Equality of Variances*, in Contributions to Probability and Statistics, ed. I. Olkin, Palo Alto, CA: Stanford University Press.

Lim, T.K. and Lim, T.K., 2013. Ziziphus mauritiana. Edible Medicinal and Non-Medicinal Plants: *Fruits* 5*,* pp.605 - 613.

Liu, G., Wang, J., Dong, P., Chen, Y. and Liu, Z., 2018. Estimating individual tree height and diameter at breast height (DBH) from terrestrial laser scanning (TLS) data at plot level. *Forests* 9(7), p.398.

Ma, H., Mo, L., Crowther, T.W., Maynard, D.S., van den Hoogen, J., Stocker, B.D., Terrer, C. and Zohner, C.M., 2021. The global distribution and environmental drivers of aboveground versus belowground plant biomass. *Nature, Ecology and Evolution* 5(8), pp.1110 - 1122.

Madani, N., Kimball, J.S., Ballantyne, A.P., Affleck, D.L., Van Bodegom, P.M., Reich, P.B., Kattge, J., Sala, A., Nazeri, M., Jones, M.O. and Zhao, M., 2018. Future global productivity will be affected by plant trait response to climate. *Scientific Reports* 8(1), p.2870.

Maposa, M., and D. Chisuro. 1998. Importance of Ziziphus mauritiana (masau) in the Mukumbura area of Zimbabwe: From a farmer's and extensionist's point of view. Proc. Int. Workshop on Ziziphus mauritiana, Harare, Zimbabwe.

Maruza, I.M., Musemwa, L., Mapurazi, S., Matsika, P., Munyati, V.T. and Ndhleve, S., 2017. Future prospects of Ziziphus mauritiana in alleviating household food insecurity and illnesses in arid and semi-arid areas: A review. *World Development Perspectives* 5, pp.1 - 6.

Mertz, O., Müller, D., Sikor, T., Hett, C., Heinimann, A., Castella, J.C., Lestrelin, G., Ryan, C.M., Reay, D.S., Schmidt-Vogt, D. and Danielsen, F., 2012. The forgotten D: challenges of addressing forest degradation in complex mosaic landscapes under REDD+. *Geografisk Tidsskrift-Danish Journal of Geography* 112(1), pp.63 - 76.

Mokgolodi, N.C., Hu, Y., Shi, L.L. and Liu, Y.J., 2011. Ziziphus mucronata: an underutilized traditional medicinal plant in Africa. *Forestry studies in China* 13, pp.163 - 172.

Mongalo, N.I., Mashele, S.S. and Makhafola, T.J., 2020. Ziziphus mucronata Willd. (Rhamnaceae): it's botany, toxicity, phytochemistry and pharmacological activities. *Heliyon 6*(4), pp.

Mugabowindekwe, M., Brandt, M., Mukuralinda, A., Ciais, P., Reiner, F., Kariryaa, A., Igel, C., Chave, J., Mertz, O., Hiernaux, P. and Tong, X., 2024. Trees on smallholder farms and forest restoration are critical for Rwanda to achieve net zero emissions. *Communications Earth and Environment* 5(1), p.113.

Mugabowindekwe, M., Brandt, M., Chave, J., Reiner, F., Skole, D., Kariryaa, A., Igel, C., Hiernaux, P., Ciais, P., Mertz, O. and Tong, X., 2022. Nation-wide mapping of tree level carbon stocks in Rwanda.

Mulatu, A., Negash, M. and Asrat, Z., 2024. Species-specific allometric models for reducing uncertainty in estimating above ground biomass at Moist Evergreen Afromontane Forest of Ethiopia. *Scientific Reports* 14(1), p.1147.

Muchuweti, M., Zenda, G. Ndhlala, A.R. and Kasiyamhuru, A., 2005. Sugars, organic acid and phenolic compounds in Ziziphus mauritiana fruit. *European Food Research and Technology,* 221, pp.570 - 574.

Muhonda,P., Mabiza, C., Makurira, H., Kujinga, K., Nhapi, I., Goldin, J., and Mashauri, D.A., 2014. Analysis of institutional mechanisms that support community response to impacts

of floods in the middle-zambezi river basin, Zimbabwe. *Physics and Chemistry of the Earth, Parts A/B/C*, 76–78, pp.64 - 71, doi:10.1016/j.pce.2014.11.013.

Nemapare, P., Gadaga, T.H. and Mugadza, D.T., 2023. Edible indigenous fruits in Zimbabwe: A review on the post-harvest handling, processing, and commercial value. *Cogent Social Sciences* 9(1), p.2229686.

Nharo, T., Makurira, H. and Gumindoga, W., 2019. Mapping floods in the middle Zambezi Basin using earth observation and hydrological modeling techniques. *Physics and chemistry of the earth, Parts A/B/C* 114, p.102787.

Nyamwanza, A.M., 2018. Local institutional adaptation for sustainable water management under increasing climatic variability and change: A case in the mid-Zambezi Valley, Zimbabwe. *International Journal of Climate Change Strategies and Management*, 10(3), pp.453 - 471.

Nyanga, L.K., Nout, M.J., Gadaga, T.H., Boekhout, T. and Zwietering, M.H., 2008. Traditional processing of masau fruits (Ziziphus mauritiana) in Zimbabwe. *Ecology of Food and Nutrition* 47(1), pp.95 - 107.

Onishi, M. and Ise, T., 2018. Automatic classification of trees using a UAV onboard camera and deep learning. *arXiv preprint arXiv:1804.10390*.

Onishi, M. and Ise, T., 2021. Explainable identification and mapping of trees using UAV RGB image and deep learning. *Scientific Reports* 11(1), p.903.

Pan, Y., Birdsey, R.A., Phillips, O.L., Houghton, R.A., Fang, J., Kauppi, P.E., Keith, H., Kurz, W.A., Ito, A., Lewis, S.L. and Nabuurs, G.J., 2024. The enduring world forest carbon sink. *Nature* 631(8021), pp.563 - 569.

Pareek, S. and Yahia, E.M., 2013. Postharvest biology and technology of ber fruit. *Horticultural Reviews* 41, pp.201 - 240.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N. and Lyapustin, A., 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature communications* 11(1), p.4540.

Qin, L., Meng, S., Zhou, G., Liu, Q. and Xu, Z., 2021. Uncertainties in above ground tree biomass estimation. *Journal of Forestry Research* 32, pp.1989 - 2000.

Rampedi, I.T. and Olivier, J., 2013. Traditional beverages derived from wild food plant species in the Vhembe District, Limpopo Province in South Africa. *Ecology of Food and Nutrition* 52(3), pp.203 - 222.

Rinklebe, J., Franke, C. and Neue, H.U., 2007. Aggregation of floodplain soils based on classification principles to predict concentrations of nutrients and pollutants. *Geoderma* 141(3-4), pp.210 - 223.

Rotenberg, E. and Yakir, D., 2010. Contribution of semi-arid forests to the climate system. *Science* 3 27(5964), pp.451 - 454.

Safriel, U. and Adeel, Z., 2005. Dryland systems. In Hassan, R., Scholes, R., and Ash, N. (eds) Ecosystems and human well-being, current state and trends, 1, pp 625 – 658, Island Press, Washington.

Saied, A.S., Gebauer, J., Hammer, K. and Buerkert, A., 2008. Ziziphus spina-christi (L.) Willd.: a multipurpose fruit tree. *Genetic Resources and Crop Evolution* 55, pp.929 - 937.

Shahrajabian, M.H., Khoshkharam, M., Zandi, P., Sun, W. and Cheng, Q., 2019. Jujube, a super-fruit in traditional Chinese medicine, heading for modern pharmacological science. *Journal of Medicinal Plants Studies* 7(4), pp.173 - 178.

Shahrajabian, M.H., Sun, W. and Cheng, Q., 2020. Chinese jujube (Ziziphus jujuba Mill.)–a promising fruit from Traditional Chinese Medicine. *Annales Universitatis Paedagogicae Cracoviensis Studia Naturae* pp.194 - 219.

Shapiro, S.S., and Wilk, M. B., 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52(3-4), pp.591 - 611.

Siddiqui, S.A., Bhardwaj, T.R., Sharma, B. and Gull, B.S., 2023. A review on Ziziphus mauritana plant. *World Journal of Pharmaceutical Research* 12(22), 133 - 145.

Spawn, S.A., Sullivan, C.C., Lark, T.J. and Gibbs, H.K., 2020. Harmonized global maps of above and belowground biomass carbon density in the year 2010. *Scientific Data* 7(1), p.112.

Stevens, J. P. (1984). Outliers and influential data points in regression analysis. *Psychological Bulletin* 95(2), pp.334 – 344. https://doi.org/10.1037/0033-2909.95.2.334.

Stone, C., Webster, M., Osborn, J. and Iqbal, I., 2016. Alternatives to LiDAR-derived canopy height models for softwood plantations: a review and example using photogrammetry. *Australian Forestry* 79(4), pp.271 - 282.

Su, J., Fan, Y., Mannan, A., Wang, S., Long, L. and Feng, Z., 2024. Real-Time Estimation of Tree Position, Tree Height, and Tree Diameter at Breast Height Point, Using Smartphones Based on Monocular SLAM. *Forests* 15(6), p.939.

Sudhersan, C., Jibi, S. and Ashkanani, J., 2014, August. Ziziphus: a highly potential multipurpose woody perennial for desert environmental rehabilitation. In *XXIX International Horticultural Congress on Horticulture: Sustaining Lives, Livelihoods and Landscapes (IHC2014): III 1116* (pp. 9 - 14).

Sungmin, O. and Park, S. K., 2023. Flash drought drives rapid vegetation stress in arid regions in Europe. *Environmental Research Letters* 18(1), p.014028.

Sungmin, O., and Park, S. K., 2024. Global ecosystem responses to flash droughts are modulated by background climate and vegetation conditions. *Communications Earth & Environment*, *5*(1), pp.1-7.

Swemmer, A. M., Nippert, J.B., and O'Connor, T.G., 2023. The effects of floods, droughts and elephants on riparian tree mortality in a semi-arid savanna. *Forest Ecology and Management* 545, p.121264, https://doi.org/10.1016/j.foreco.2023.121264.

Talbot, C.J., Bennett, E.M., Cassell, K., Hanes, D.M., Minor, E.C., Paerl, H., Raymond, P.A., Vargas, R., Vidon, P.G., Wollheim, W. and Xenopoulos, M.A., 2018. The impact of flooding on aquatic ecosystem services. *Biogeochemistry* 141, pp.439 - 461.

Tanwar, S.P.S., Kumar, P., Verma, A., Bhatt, R.K., Singh, A., Lal, K., Patidar, M. and Mathur, B.K., 2019. Carbon sequestration potential of agroforestry systems in the Indian arid zone. *Current Science* 117(12), pp.2014-2022.

Tsheboeng, G., Bonyongo, M. and Murray-Hudson, M., 2014. Flood variation and soil nutrient content in floodplain vegetation communities in the Okavango Delta. *South African Journal of Science* 110(3-4), pp.01 - 05.

Utts, J.M., 1982. The rainbow test for lack of fit in regression. *Communications in Statistics-Theory and Methods* 11(24), pp.2801-2815.

Vähä-Konka, V., Korhonen, L., Kärhä, K. and Maltamo, M., 2024. Estimating the accuracy of smartphone app-based removal estimates against actual wood-harvesting data from clear cuttings. *iForest-Biogeosciences and Forestry* 17(3), p.140.

Wagner, F.H., Roberts, S., Ritz, A.L., Carter, G., Dalagnol, R., Favrichon, S., Hirye, M., Brandt, M., Ciais, P., and Saatchi, S., 2024. Sub-meter tree height mapping of California using aerial images and LiDAR-informed U-Net model. *Remote Sensing of Environment* 305, p.114099. https://doi.org/10.1016/j.rse.2024.114099.

Wallace, T.A., Gehrig, S. L, Doody, T.M., Davies, M., Walsh, R., Fulton, C. , …, . and Nolan, M., 2021. A multiple-lines-of-evidence approach for prioritising environmental watering of wetland and floodplain trees. *Ecohydrology*. doi:10.1002/eco.2272.

Wang, Y., Lehtomäki, M., Liang, X., Pyörälä, J., Kukko, A., Jaakkola, A., Liu, J., Feng, Z., Chen, R. and Hyyppä, J., 2019. Is field-measured tree height as reliable as believed–A comparison study of tree height estimates from field measurement, airborne laser scanning and terrestrial laser scanning in a boreal forest. *ISPRS Journal of Photogrammetry and Remote Sensing 147*, pp.132 - 145.

Wang, C., Wang, X., Liu, D., Wu, H., Lü, X., Fang, Y., Cheng, W., Luo, W., Jiang, P., Shi, J. and Yin, H., 2014. Aridity threshold in controlling ecosystem nitrogen cycling in arid and semi-arid grasslands. *Nature Communications* 5(1), p.4799.

Wardius, Y. and Hein, S., 2024. Terrestrial laser scanning vs. manual methods for assessing complex forest stand structure: a comparative analysis on plenter forests. *European Journal of Forest Research* 143(2), pp.635 - 649.

Wu, J., Yao, W., Choi, S., Park, T. and Myneni, R.B., 2015. A comparative study of predicting DBH and stem volume of individual trees in a temperate forest using airborne waveform LiDAR. *IEEE geoscience and Remote Sensing letters* 12(11), pp.2267 - 2271.

Xie, Y., Yang, T., Wang, X., Chen, X., Pang, S., Hu, J., Wang, A., Chen, L. and Shen, Z., 2022. Applying a portable backpack LiDAR to measure and locate trees in a nature forest plot: Accuracy and error analyses. *Remote Sensing* 14(8), p.1806.

Xu, Z., Qin, L., Zhou, G., SiQing, B., Du, W., Meng, S., Yu, J., Sun, Z. and Liu, Q., 2024. Exploring carbon sequestration in broad-leaved Korean pine forests: Insights into photosynthetic and respiratory processes. *Science of the Total Environment 906*, p.167421.

Yang, H., Ciais, P., Frappart, F., Li, X., Brandt, M., Fensholt, R., Fan, L., Saatchi, S., Besnard, S., Deng, Z. and Bowring, S., 2023. Global increase in biomass carbon stock dominated by growth of northern young forests over past decade. *Nature Geoscience* 16(10), pp.886 - 892.

Zeng, N. and Hausmann, H., 2022. Wood Vault: remove atmospheric CO_2 with trees, store wood for carbon sequestration for now and as biomass, bioenergy and carbon reserve for the future. *Carbon Balance and Management* 17(1), p.2.

Zhang, Z., Wang, T., Skidmore, A.K., Cao, F., She, G. and Cao, L., 2023. An improved area-based approach for estimating plot-level tree DBH from airborne LiDAR data. *Forest Ecosystems* 10, p.100089.

CHAPTER 7: SYNTHESIS, CONCLUSION AND RECOMMENDATIONS

7.1 Synthesis

The influence of floods-recharged soil moisture (FRSM) on trees that provide multiple societal benefits is poorly understood. Hence, we assessed the ecological influence of FRSM on jujube *(Ziziphus mauritiana)* species as a proxy for trees in semi-arid floodplains of northern Zimbabwe to address this knowledge gap. We undertook a cross-cutting research with modelling; 1) to improve flood extent mapping using a novel change detection and thresholding Ensemble of Scenarios Pyramid (ESP) technique using Sentinel-1 data, 2) three FRSM features namely, 'anomaly', 'lag' and 'memory' using a novel hybrid deep learning Long Short-Term Memory embedded within an AutoEncoder architecture, (LSTM-AE), within Python, 3) short-term influence of FRSM on leaf chlorophyll content (LCC), and 4) long-term influence of FRSM on tree biophysical characteristics. The specific objectives of the study were to;

1 assess the utility of remote sensing indices in understanding the nexus between FRSM with tree biochemical and biophysical features in tropical semi-arid floodplains.
2 model flood extent mapping using Ensemble of Scenarios Pyramid in semi-arid floodplains in northern Zimbabwe using Sentinel-1 SAR data.
3 assess the utility of deep learning long short-term memory autoencoder algorithm applied on Sentinel-1 SAR data to model FRSM features in semi-arid floodplains in northern Zimbabwe.
4 assess the influence of FRSM on short-term biochemical properties of *Z. mauritiana* tree in semi-arid floodplain in northern Zimbabwe.
5 evaluate the long-term influence of FRSM on *Z. mauritiana* tree biophysical characteristsics in semi-arid lands in northern ZImbabwe.

The results showed that the novel ESP as method achieved higher accuracy in flood extent mapping as compared to using one scenario as in the Normalized Difference Flood Index (NDFI). While the Sentinel-1 data pre-processing costs were high, the ESP superior accuracy was confirmed using six model evaluation metrics. The utility of the ESP method, which is based on radar data, helped to solve the challenge of cloudy weather and night-time constrained flood extent mapping using passive remote sensing data. To optimize accurate flood extent mapping

using the ESP, at least two multi-temporal images; before and after flood; are necessary. At the base of the ESP, the vertical-vertical polarized proved superiority to vertical-horizonal polarized image in flood extent mapping. However, while the ESP demonstrated efficacy in operational flood disaster management, more use cases are necessary to check its applicability elsewhere. This spatial flood extent delimitation facilitated sampling of points inside and outside the flooded zone for subsequent FRSM modelling.

The LSTM-AE deep learning algorithm detected a FRSM anomaly that had a one-week lag and profound 2 < months longer memory not evident outside the flooded zone. The lag feature introduced in this study, is the duration of the anomalous condition implemented in Spyder (Python). Two-way ANOVA confirmed that the LCC inside the floodplain was significantly higher than outside, using spatial plus other variables, exposure to sunlight, and height on canopy. Spatial variables alone were statistically significant, further confirming that FRSM influenced higher LCC than outside the floodplain. Ultimately, linear modelling of the relationship of three selected tree biophysical characteristics with distance to source of moisture confirmed that on the long-term, FRSM showed significant influence. The trees inside flooded locations showed higher tree size values than those outside. This finding was statistically significant using the Mann-Whitney-Wilcoxon test, Levene's test and pairwise Wilcoxon post-hoc tests. Logically, the post-flood longer FRSM memory gives trees inside flooded locations a longer productivity window as shown using the higher floodplain LCC than outside, hence, produce more growth with each flood occurence. Consequently, trees inside flooded locations grow taller, have bigger canopy diameter, and DBH than those outside, thus showing the influence of the FRSM on trees. Therefore, this study addressed the knowledge gap on the influence of FRSM on trees in semi-arid areas.

7.2 Conclusion

This study aimed to assess the influence of FRSM on trees in semi-arid areas in northern Zimbabwe using *Z. mauritiana* tree as a l proxy since it exists abundantly inside and outside the floodplain. The aim was achieved through sequenced steps involving accurate spatial delimitation of the floodplain, temporal modelling of three FRSM features (anomaly, lag and memory), modelling seasonal LCC that compared flood- with non-floodplain trees, and finally the linear modelling of canopy diameter, tree height and DBH on distance to water sources and flood status. The study also quantified the difference between the three selected tree characteristics based on flood status. Results from the study showed that FRSM promotes higher

chlorophyll content and subsequently more tree growth for trees inside flooded locations than outside. The review of literature showed that floods recharge soil moisture through deep infiltration. This phenomenon was shown in FRSM LSTMAE deep learning modelling that showed distinct positive anomaly, lag and memory inside flooded locations while not evident outside. The prolonged FRSM memory also corroborated the higher LCC and longer green-up phenology window inside flooded locations than outside. The main conclusions of the study are as follows:

i) Using the northern Zimbabwe floodplain use-case, the ESP as a new alternative to flood extent mapping based on change detection and thresholding technique proved higher performance in accuracy in comparison to single scenario of NDFI.

ii) Three FRSM features (anomaly, lag and memory) were evident on a 2017 flood event within flooded locations as opposed to outside ones. Consequently, within same locality, the deep infiltration effect contributes to more soil moisture recharge in flooded locations unlike outside that rely on precipitation within flash droughts due to high temperatures and windy conditions associated with high evaporative demand.

iii) The LCC of *Z. mauritiana* trees inside the floodplain were compared to those outside the floodplain to deduce the effect of FRSM within an area affected by antagonistic hydro-meteorological extremes (floods and flash droughts). Results showed that trees outside the floodplain were exposed to numerous in-season flash droughts while inside, the deep infiltration triggered FRSM anomaly and memory persists for 2< months, hence longer green-up window. In-shadow LCC was significantly higher than that in-sunlight LCC because of photoinhibition and photodamage. LCC inside floodplain was higher than outside the floodplain.

iv) Trees inside floodplain benefit from repeated episodes of deep infiltration, hence longer growth window with taller, and larger canopy diameter and DBH than those outside.

Overall, the flood extent delimitation was useful in discriminating soil moisture by flood status. Locations inside flooded area depicted higher FRSM which last longer than outside. Thus, *Z. mauritiana* tree inside flooded locations showed higher LCC than outside. Ultimately, trees inside flooded location grow more than those outside, probably due to the contribution of the FRSM. This thesis showed the ecological importance of FRSM in semi-arid environments, hence applicable to carbon accounting, ecosystem goods and services as important areas of application.

7.3 Recommendations

Quantifying the influence of FRSM on trees lies in understanding the flood extent, spatio-temporal FRSM characteristics, the length of the green-up phenology window assessed using LCC, and ultimately the tree size characteristics. Modelling the relationship between tree features with distance to source of soil moisture and comparing same features between flood- and nonfloodplain trees was essential to evaluate the contribution of FRSM to trees using *Z. mauritiana* species in semi-arid Zimbabwe. Therefore, the study recommends the following for future studies:

i) To accurately model flood extent, application of machine and deep learning towards developing of new indices that may combine both optical and radar images to map flood extent for various applications is necessary. This is potentially possible in view of the ongoing machine learning algorithm developments and the data deluge.

iii) Whereas this study statiscally derived numerous important variables, monitoring vegetation changes consistently and over longer periods through a distributed network of instrumentation could be valuable in determining the precise influences of flooding on LCC. The daily fluctuations and multiple species differentiation should also be investigated.

iv) While long term influence of floods on tree structural characteristics is visible, across species research may be necessary to extract the actual tree height, DBH, canopy size, among others. Furthermore, the study was based on limited variables, hence more variables such as edaphic, climatic, topographic and human factors may be incorporated to effectively discriminate the most versus the least influential ones.

Overall, more data, improvements and innovations in machine learning and deep learning algorithm development is promising to help expedite similar efforts. Figure 3.2

8. APPENDICES

Appendix 3.1: SNAP Sentinel-1 Data preprocessing

SNAP Graph Processing Tool (GPT) customised script for all steps image preprocessing.

```
<graph id="Graph">
  <version>1.0</version>
<!--
        <root xmlns:copyright="http://www.w3.org/1999/xhtml"
                <metadata>
                        <author name="Ezra Pedzisai" />
                        <author contact="218085751@stu.ukzn.ac.za" />
                        <author details="PhD Student 218085751 University of KwaZulu Natal,
School of Agriculture, Earth and Environmental Science, Discipline of Geography,
Pietermaritzburg Campus, South Africa/>
                        <copyright name="CC BY-SA" />
                        <license type="GPL" />
                        <license version="3" />
                </metadata>
        </root>
  -->
<!-- usage
  <node id="Read">
        <operator>Read</operator>
        <sources/>
        <parameters class="com.bc.ceres.binding.dom.XppDomElement">
        <file>C:/Users/Ezra/Desktop/S1A_IW_GRDH_1SDV_20150420T162425_20150420T
162450_005569_007204_8CB0.zip</file>
    </parameters>
  </node>
<node id="Apply-Orbit-File">
    <operator>Apply-Orbit-File</operator>
```

```
        <sources>
          <sourceProduct refid="Read"/>
    </sources>
    <parameters class="com.bc.ceres.binding.dom.XppDomElement">
      <orbitType>Sentinel Precise (Auto Download)</orbitType>
      <polyDegree>3</polyDegree>
      <continueOnFail>true</continueOnFail>
    </parameters>
  </node>
<node id="ThermalNoiseRemoval">
    <operator>ThermalNoiseRemoval</operator>
    <sources>
      <sourceProduct refid="Apply-Orbit-File"/>
    </sources>
    <parameters class="com.bc.ceres.binding.dom.XppDomElement">
      <selectedPolarisations/>
      <removeThermalNoise>true</removeThermalNoise>
      <reIntroduceThermalNoise>false</reIntroduceThermalNoise>
    </parameters>
</node>
<node id="Remove-GRD-Border-Noise">
        <operator>Remove-GRD-Border-Noise</operator>
        <sources>
                <sourceProduct refid="ThermalNoiseRemoval"/>
        </sources>
  <parameters class="com.bc.ceres.binding.dom.XppDomElement">
        <selectedPolarisations/>
        <borderLimit>500</borderLimit>
        <trimThreshold>0.50</trimThreshold>
        </parameters>
```

```
</node>
<node id="Calibration-Sigma">
        <operator>Calibration</operator>
        <sources>
                <sourceProduct refid="Remove-GRD-Border-Noise"/>
        </sources>
        <parameters class="com.bc.ceres.binding.dom.XppDomElement">
                <sourceBands/>
                <auxFile>Product Auxiliary File</auxFile>
                <externalAuxFile/>
                <outputImageInComplex>false</outputImageInComplex>
                <outputImageScaleInDb>false</outputImageScaleInDb>
                <createGammaBand>false</createGammaBand>
                <createBetaBand>false</createBetaBand>
                <selectedPolarisations/>
                <outputSigmaBand>true</outputSigmaBand>
                <outputGammaBand>false</outputGammaBand>
                <outputBetaBand>false</outputBetaBand>
        </parameters>
</node>
<node id="Speckle-Filter">
   <operator>Speckle-Filter</operator>
   </sources>
     <sourceProduct refid="Calibration-Sigma"/>
   </sources>
   <parameters class="com.bc.ceres.binding.dom.XppDomElement">
     <sourceBands/>
     <filter>Lee Sigma</filter>
     <filterSizeX>3</filterSizeX>
          <filterSizeY>3</filterSizeY>
```

```
            <dampingFactor>2</dampingFactor>
            <estimateENL>true</estimateENL>
            <enl>1.0</enl>
            <numLooksStr>1</numLooksStr>
            <windowSize>7x7</windowSize>
            <targetWindowSizeStr>3x3</targetWindowSizeStr>
            <sigmaStr>0.9</sigmaStr>
            <anSize>50</anSize>
    </parameters>
</node>
<node id="Terrain-Correction">
    <operator>Terrain-Correction</operator>
    <sources>
      <sourceProduct refid="Speckle-Filter"/>
    </sources>
    <parameters class="com.bc.ceres.binding.dom.XppDomElement">
      <sourceBands/>
      <demName>SRTM 3Sec</demName>
      <externalDEMNoDataValue>0.0</externalDEMNoDataValue>
      <externalDEMApplyEGM>true</externalDEMApplyEGM>
      <demResamplingMethod>BILINEAR_INTERPOLATION</demResamplingMethod>
      <imgResamplingMethod>BILINEAR_INTERPOLATION</imgResamplingMethod>
      <pixelSpacingInMeter>10.0</pixelSpacingInMeter>
      <pixelSpacingInDegree>8.983152841195215E-5</pixelSpacingInDegree>
      <mapProjection>PROJCS[&amp;quot;UTM Zone 36, South / World Geodetic System
1984&amp;quot;, &amp;#xd;
  GEOGCS[&amp;quot;World Geodetic System 1984&amp;quot;, &amp;#xd;
    DATUM[&amp;quot;World Geodetic System 1984&amp;quot;, &amp;#xd;
      SPHEROID[&amp;quot;WGS 84&amp;quot;, 6378137.0, 298.257223563,
AUTHORITY[&amp;quot;EPSG&amp;quot;,&amp;quot;7030&amp;quot;]], &amp;#xd;
```

AUTHORITY["EPSG","6326"]],
PRIMEM["Greenwich", 0.0,
AUTHORITY["EPSG","8901"]],
UNIT["degree", 0.017453292519943295],
AXIS["Geodetic longitude", EAST],
AXIS["Geodetic latitude", NORTH]],
PROJECTION["Transverse_Mercator"],
PARAMETER["central_meridian", 33.0],
PARAMETER["latitude_of_origin", 0.0],
PARAMETER["scale_factor", 0.9996],
PARAMETER["false_easting", 500000.0],
PARAMETER["false_northing", 10000000.0],
UNIT["m", 1.0],
AXIS["Easting", EAST],
AXIS["Northing", NORTH]]</mapProjection>
<alignToStandardGrid>true</alignToStandardGrid>
<standardGridOriginX>0.0</standardGridOriginX>
<standardGridOriginY>0.0</standardGridOriginY>
<nodataValueAtSea>false</nodataValueAtSea>
<saveDEM>false</saveDEM>
<saveLatLon>false</saveLatLon>
<saveIncidenceAngleFromEllipsoid>false</saveIncidenceAngleFromEllipsoid>
<saveLocalIncidenceAngle>false</saveLocalIncidenceAngle>
<saveProjectedLocalIncidenceAngle>false</saveProjectedLocalIncidenceAngle>
<saveSelectedSourceBand>true</saveSelectedSourceBand>
<outputComplex>false</outputComplex>
<applyRadiometricNormalization>false</applyRadiometricNormalization>
<saveSigmaNought>false</saveSigmaNought>
<saveGammaNought>false</saveGammaNought>
<saveBetaNought>false</saveBetaNought>

```
    <incidenceAngleForSigma0>Use projected local incidence angle from
DEM</incidenceAngleForSigma0>
    <incidenceAngleForGamma0>Use projected local incidence angle from
DEM</incidenceAngleForGamma0>
    <auxFile>Latest Auxiliary File</auxFile>
   </parameters>
  </node>
<node id="LinearToFromdB">
        <operator>LinearToFromdB</operator>
   <sources>
            <sourceProduct refid="Terrain-Correction"/>
   </sources>
        <parameters class="com.bc.ceres.binding.dom.XppDomElement">
                <sourceBands/>
            </parameters>
</node>
<node id="Subset">
   <operator>Subset</operator>
   <sources>
    <sourceProduct refid="LinearToFromdB"/>
   </sources>
   <parameters class="com.bc.ceres.binding.dom.XppDomElement">
    <sourceBands>Sigma0_VH</sourceBands>
    <region>0,0,0,0</region>
    <referenceBand/>
    <geoRegion>POLYGON ((30.65216636657715 -16.018808364868164,
31.434921264648438 -16.018808364868164, 31.434921264648438 -16.34048843383789,
30.65216636657715 -16.34048843383789, 30.65216636657715 -16.018808364868164,
30.65216636657715 -16.018808364868164))</geoRegion>
    <subSamplingX>1</subSamplingX>
```

```
      <subSamplingY>1</subSamplingY>
      <fullSwath>false</fullSwath>
      <tiePointGridNames/>
      <copyMetadata>true</copyMetadata>
    </parameters>
</node>
<node id="Subset(2)">
    <operator>Subset</operator>
    <sources>
      <sourceProduct refid="Terrain-Correction"/>
    </sources>
    <parameters class="com.bc.ceres.binding.dom.XppDomElement">
      <sourceBands>Sigma0_VV</sourceBands>
      <region>0,0,0,0</region>
      <referenceBand/>
      <geoRegion>POLYGON ((30.65216636657715 -16.018808364868164,
31.434921264648438 -16.018808364868164, 31.434921264648438 -16.34048843383789,
30.65216636657715 -16.34048843383789, 30.65216636657715 -16.018808364868164,
30.65216636657715 -16.018808364868164))</geoRegion>
      <subSamplingX>1</subSamplingX>
      <subSamplingY>1</subSamplingY>
      <fullSwath>false</fullSwath>
      <tiePointGridNames/>
      <copyMetadata>true</copyMetadata>
    </parameters>
</node>
<node id="Write">
    <operator>Write</operator>
    <sources>
      <sourceProduct refid="Subset"/>
```

```
    </sources>
    <parameters class="com.bc.ceres.binding.dom.XppDomElement">
<file>C:/Users/Ezra/Subset_S1A_IW_GRDH_1SDV_20150420T162425_20150420T162450_005569_007204_8CB0_Orb_NR_Cal_TC.dim</file>
      <formatName>BEAM-DIMAP</formatName>
    </parameters>
</node>
<node id="Write(2)">
    <operator>Write</operator>
    <sources>
      <sourceProduct refid="Subset(2)"/>
    </sources>
    <parameters class="com.bc.ceres.binding.dom.XppDomElement">
      <file>
C:/Users/Ezra/Subset_S1A_IW_GRDH_1SDV_20150420T162425_20150420T162450_005569_007204_8CB0_Orb_NR_Cal_TC.dim</file>
      <formatName>BEAM-DIMAP</formatName>
    </parameters>
</node>
<applicationData id="Presentation">
  <Description>Sentinel-1 GRD Sigma0 Standard Preprocessing Graph</Description>
  <node id="Read">
          <displayPosition x="37.0" y="134.0"/>
  </node>
  <node id="Apply-Orbit-File">
    <displayPosition x="131.0" y="135.0"/>
  </node>
  <node id="ThermalNoiseRemoval">
    <displayPosition x="258.0" y="136.0"/>
  </node>
```

```
  <node> id="Remove-GRD-Border-Noise">
           <displayPosition x="00.0" y="00.0"/>
         </node>
   <node id="Calibration-Sigma">
     <displayPosition x="427.0" y="135.0"/>
   </node>
   <node id="Terrain-Correction">
     <displayPosition x="531.0" y="134.0"/>
   </node>
         <node id="LinearToFromdB">
           <displayPosition x="00.0" y="00.0"/>
         </node>
   <node id="Subset">
     <displayPosition x="679.0" y="136.0"/>
   </node>
   <node id="Subset(2)">
     <displayPosition x="687.0" y="212.0"/>
   </node>
   <node id="Write(2)">
     <displayPosition x="869.0" y="213.0"/>
   </node>
   <node id="Write">
         <displayPosition x="867.0" y="135.0"/>
   </node>
  </applicationData>
</graph>
"""
@author: Ezra Pedzisai
University of KwaZulu Natal
College of Agriculture, Engineering and Science
```

School of Agricultural, Earth and Environmental Sciences

Discipline of Geography

Pietermaritzburg Campus

Pietermaritzburg

South Africa

218285751@stu.ukzn.ac.za; ezpedzisai@gmail.com

Appendix 4.1: Hybrid Deep Learning (LSTMAE) FRSM modelling in Spyder Python

```
################################################################################
##########                Start of Appendix 4.1 – Script                ################
################################################################################
```

 # STEP 4.1A1: Import/Call relevant Python Libraries in Spyder (4) IDE (Anaconda 3.8.8) Python Programming Language on TensorFlow Backend

```
import numpy as np
import matplotlib.ticker as tkr # to check
from scipy import stats # to check
import math # to check
import tensorflow as tf
from tensorflow import keras
from keras.models import Model
from keras.models import Sequential
from keras.layers import LSTM
from keras.layers import Input
from keras.layers import Dropout
from keras.layers import Dense
from keras.layers import RepeatVector
from keras.layers import TimeDistributed
from keras.callbacks import EarlyStopping # to check
import pandas as pd
from matplotlib import pyplot as plt
from matplotlib import rc
from sklearn.preprocessing import MinMaxScaler
#from sklearn.linear_model import LinearRegression
#from sklearn.preprocessing import StandardScaler
from sklearn.metrics import mean_squared_error # to check
from sklearn.metrics import mean_absolute_error # to check
# from sklearn.metrics import root_mean_squared_error # to check
from sklearn.model_selection import train_test_split
import seaborn as sns
from pylab import rcParams
sns.set(style ='whitegrid', palette ='muted', font_scale = 2.0)
rcParams['figure.figsize'] = 20, 8
```

```
################################################################################
# Step 4.1A2: Load, Inspect Data and Test for Normality
################################################################################
dataframe = pd.read_csv('C://Users//user//Desktop//LSTMAE_Data1.csv')

# Point to the dataframe Select one each time a scenario based on polarisation is executed,
# NB: Notes on the codes are shown on the right of the code
#df = dataframe[['Date', '136FloVV']]  #  FLo - is sampled INSIDE  a Flood location and VV - is for Co-polarised band
df = dataframe[['Date', '136FloVH']] #  FLo - is sampled INSIDE  a Flood location and VH - is for Cross-polarised band
#df = dataframe[['Date', '171NFloVV']] # NFLo - is sampled OUTSIDE a Flood location and VV - is for Co-polarised band
#df = dataframe[['Date', '171NFloVH']] # NFLo - is sampled OUTSIDE a Flood location and VH - is for Cross-polarised band

df['Date'] = pd.to_datetime(df['Date']) # OR
# OR Alternatively
#df.Date = pd.to_datetime(df.Date)
#df.set_index('Date', inplace = True)
df['136FloVH'] = df['136FloVH']
#Inspect to verify the start and end of time series
print("Start of Sequence: ", df['Date'].min())
print("End of Sequence: ", df['Date'].max())
# Test for normality
# Statistical Normality Test
# There are several statistical tests that we can use to quantify whether our data looks as though it was drawn from a Gaussian distribution.
# And we will use D'Agostino's K² Test.
# implementation the test by interpretting the p value as follows.

#p<=alpha: reject H0, not normal.
#p>alpha: fail to reject H0, normal.
#p <= alpha: reject H0, not normal
#p > alpha: fail to reject H0, normal
stat, p = stats.normaltest(df['136FloVH'])
print('Statistics=%.3f, p=%.3f' % (stat, p))
```

```
alpha = 0.05
if p > alpha:
    print('Data resembles Gaussian (fail to reject H0)')
else:
    print('Data does not resemble Gaussian (reject H0)')
#============================================================
```

 # Step 4.1B Preprocess Sequential Data (Train/Test Data Splitting)

```
# ============================================================#
# Train data is for the two years with no flood occurrence, i.e. since January 2018 through December 2019 (Ntrain = 61)
# Test data  is for the one year with a flood occurrence, i.e. from January 2017 through December 2017 (Ntest = 30)
# (SINGLE BAND - LSTMAE MODELLING) (I.E. SEQUENCE SINGLE POLARIZATION SEQUENCE) LSTM-AE MODELLING OF ANOMALY AND TEMPORAL MEMORY
# (INSIDE AND OUTSINE) FLOODED ZONE IN CHITSUNGO OF MBIRE DISTRICT, ZIMBABWE
# LSTM uses sigmoid and tanh that are sensitive to magnitude so values need to be normalized

# Data Partitioning into Train (67%) and Test (33%) ARRAYS.
train, test = df.loc[df['Date'] > '2017-12-31'], df.loc[df['Date'] < '2017-12-31']

print(train)
print(test)
# There was no need to transform data, after testing with both MinMax scaler,
# the transformation did not give any differences since the variances in the data were low.

#scaler = MinMaxScaler()
#scaler = MinMaxScaler(feature_range=(0, 1))
#df = scaler.fit_transform(df[['136FloVH']])
# scaler = scaler.fit(df[['136FloVV']])
# scaler = scaler.fit(test[['136FloVV']])
# train['136FloVV'] = scaler.fit_transform(train[['136FloVV']])
# test['136FloVV'] = scaler.fit_transform(test[['136FloVV']])

#############################################################################
```

 # Step 4.1B1: Define the Batch_size =2

```
#############################################################################
```

```
seq_size = 2  # aka lookback or Number of time steps to look back (i.e. moving average window used)
#Seq_size is also known as lookback hence
# While larger sequences (look further back) may improve fecasting, however, in this case, since the Flood occurred on the
# third date, using a timestep larger than 3 means we fail to predict starting from the date of the flood event in the study area.

################################################################################
```
 # STEP 4.1: Create arrays of y given the x arrays using the following calculation;1C
```
################################################################################
def to_sequences(x, y, seq_size = 1):
    x_values, y_values = [], []
    for i in range(len(x)-seq_size):
        #print(i)
        x_values.append(x.iloc[i:(i + seq_size)].values)
        y_values.append(y.iloc[i + seq_size])

    return np.array(x_values), np.array(y_values)

trainX, trainY = to_sequences(train[['136FloVH']], train['136FloVH'], seq_size)
testX, testY = to_sequences(test[['136FloVH']], test['136FloVH'], seq_size)

print(trainX.shape, trainY.shape)
print(testX.shape, testY.shape)

trainX = trainX.reshape(trainX.shape[0], trainX.shape[1], 1)
testX= testX.reshape(testX.shape[0], testX.shape[1], 1)

print(trainX.shape)
print(trainY.shape)
print(testX.shape)
print(testY.shape)
# ################################################################################
```
 # Step 4.2: Define LSTM Autoencoder architecture and parameterise
```
################################################################################

#Build the model: Autoencoder LSTM model1
```

```
#model1- LSTMAE No Dropout
model = Sequential()
model.add(LSTM(128, activation='relu', input_shape=(trainX.shape[1], trainX.shape[2]),
return_sequences=True))
model.add(LSTM(64, activation='relu', return_sequences=False))
model.add(RepeatVector(trainX.shape[1]))
model.add(LSTM(64, activation='relu', return_sequences=True))
model.add(LSTM(128, activation='relu', return_sequences=True))
model.add(TimeDistributed(Dense(trainX.shape[2])))
model.compile(optimizer='adam', loss = 'mae', metrics = 'mse')
model.summary()

# LSTMAE Model Architecture With Dropout of 0.2
# Define the LSTM with 128 neurons in the first hidden layer and 64 neurons in the second output layer
for predicting Global_active_power. The input shape will be 1 time step with 30 features.
# Dropout 20%.
# Use the MSE loss function and the efficient Adam version of stochastic gradient descent.
# The model will be fit for 61 training epochs with a batch_size = 2, with 100 epochs
#model = Sequential()
#model.add(LSTM(128, input_shape=(trainX.shape[1], trainX.shape[2])))
#model.add(Dropout(rate=0.2))
#model.add(RepeatVector(trainX.shape[1]))
#model.add(LSTM(128, return_sequences=True))
#model.add(Dropout(rate=0.2))
#model.add(TimeDistributed(Dense(trainX.shape[2])))
#model.compile(optimizer='adam', loss='mae', metrics ='mse')
#model.summary()

#model1- LSTMAE No Dropout
model = Sequential()
model.add(LSTM(128, activation='relu', input_shape=(trainX.shape[1], trainX.shape[2]),
return_sequences=True))
model.add(LSTM(64, activation='relu', return_sequences=True))
model.add(LSTM(32, activation='relu', return_sequences=True))
model.add(LSTM(16, activation='relu', return_sequences=True))
model.add(LSTM(8, activation='relu', return_sequences=True))
model.add(LSTM(4, activation='relu', return_sequences=False))
```

```
model.add(RepeatVector(trainX.shape[1]))
model.add(LSTM(4, activation='relu', return_sequences=True))
model.add(LSTM(8, activation='relu', return_sequences=True))
model.add(LSTM(16, activation='relu', return_sequences=True))
model.add(LSTM(32, activation='relu', return_sequences=True))
model.add(LSTM(64, activation='relu', return_sequences=True))
model.add(LSTM(128, activation='relu', return_sequences=True))
model.add(TimeDistributed(Dense(trainX.shape[2])))
model.compile(optimizer='adam', loss = 'mae', metrics = 'mse')
model.summary()

##############################################################################
```

 # # Section 4.3: 1) Fit the Autoencoder LSTMAE model as a Figure 10 (line plot for train (Calibration) and test (validation))))

```
##############################################################################
# Figure 10 model evaluation, a) VV Flooded, b) VV Not Flooded, c) VH Flooded, d) VH Not Flooded
history = model.fit(trainX, trainY, epochs = 100, batch_size = 2, validation_split = 0.33, verbose = 1)

#trainX = trainX.reshape(1, -1)
#trainX.shape
#testX = testX.reshape(1, -1)
#testX.shape

train_predict = model.predict(trainX)
test_predict = model.predict(testX)
#train_predict = scaler.inverse_transform(train_predict)
#test_predict = scaler.inverse_transform(test_predict)

#train_predict = train_predict.reshape(59, 1, -1)
#test_predict = test_predict.reshape(28, 1, -1)

print('Train Mean Absolute Error:', mean_absolute_error(trainY[0:,], train_predict[:,0]))
print('Train Mean Squared Error:', np.sqrt(mean_squared_error(trainY[0:,],
train_predict[:,0]))*np.sqrt(mean_squared_error(trainY[0:,], train_predict[:,0])))
print('Train Root Mean Squared Error:', np.sqrt(mean_squared_error(trainY[0:,], train_predict[:,0])))
print('Test Mean Absolute Error:', mean_absolute_error(testY[0:,], test_predict[:,0]))
```

```
print('Test Mean Squared Error:', np.sqrt(mean_squared_error(testY[0:,],
test_predict[:,0]))*np.sqrt(mean_squared_error(testY[0:,], test_predict[:,0])))
print('Test Root Mean Squared Error:', np.sqrt(mean_squared_error(testY[0:,], test_predict[:,0])))

#: Figure 11a: Evaluate the model performance using observed versus predicted VV in Flood sigma
nought values
trainval = [x for x in range(59)]
plt.figure(figsize = (20, 8))
plt.plot(trainval, trainY[0:][:59], color = 'gray', linewidth = 8, label = "Observed")
plt.plot(trainval, train_predict[:,0][:59], color = 'red', linewidth = 4, label = "Predicted")
# plt.tick_params(left=False, labelleft=True) #remove ticks
plt.tight_layout()
sns.despine(top = True)
plt.title('b) VH in Flood: Train', size = 30, color = 'black')
plt.subplots_adjust(left = 0.07)
plt.ylabel('Sigma Nought', size = 26)
plt.xlabel('Timestep', size = 26)
plt.legend(fontsize = 26)
plt.show()
#: Figure 11e: Evaluate the model performance using observed versus predicted VV in Flood sigma
nought values
testval=[x for x in range(28)]
plt.figure(figsize = (20, 8))
plt.plot(testval, testY[0:28], marker = '.', color = 'gray', linewidth = 8, label = "Observed")
plt.plot(testval, test_predict[:, 0][:28], marker = '+', color = 'red', linewidth = 4, label = "Predicted") #
alpha = 0.05,
plt.tight_layout()
plt.title('g) VH in Flood: Test', size = 30, color = 'black')
sns.despine(top = True)
plt.subplots_adjust(left = 0.07)
plt.ylabel('Sigma nought', size = 26)
plt.xlabel('Timestep', size = 26)
plt.legend(fontsize = 26)
plt.show()

#:test_predict = test_predict.flatten()
#:train_predict = train_predict.flatten()
```

```
#############################################################################
# OR in sns
#sns.regplot(data = trainval, x = trainY, y = train_predict, scatter_kws = dict(color = 'gray', s = 10, alpha
= 0.5), line_kws = dict(color = 'orange'))
#############################################################################
```
 # Evaluate model using MAE and MSE on Train and Validation data using loss functions Figure 9
```
#############################################################################
plt.title('b) VH in Flood: Loss Metrics ', size=30, color='black')
plt.plot(history.history['loss'], label='Training MAE', linewidth = 5, color = 'black')
plt.plot(history.history['val_loss'], label = 'Testing MAE', linewidth = 5, color = 'green')
plt.plot(history.history['mse'], label = 'Training MSE', linewidth = 5, color = 'blue')
plt.plot(history.history['val_mse'], label = 'Testing MSE', linewidth = 5, color = 'orange')
plt.xlabel('Epochs', size=26, color='black')
plt.ylabel('Loss', size=26, color='black')
plt.legend()

# Make predictions based on the LSTMAE model built above
train_pedict = model.predict(trainX)
test_predict = model.predict(testX)

print(train_predict) # extract for graphics
print(test_predict) # extract for graphics

#################################################################
```
#Section 4.3: 2) Evaluate the model based on testX and testY
```
#################################################################
# To extract by scenrio and Create Table 2:
#The quantitative LSTMAE MAE by polarisation and location.
model.evaluate(testX, testY)
model.evaluate(trainX, trainY)
#################################################################
# Train MAE and MSE Loss
_, loss = model.evaluate(trainX, trainY)
print('Train MAE Loss: %.2f', (loss))
_, val_loss = model.evaluate(trainX, trainY)
```

```
print('Train MSE Loss: %.2f', (val_loss))

# Test MAE and MSE Loss
_, loss = model.evaluate(testX, testY)
print('Test MAE Loss: %.2f', (loss))
_, val_loss = model.evaluate(testX, testY)
print('Test MSE Loss: %.2f', (val_loss))
#Theory for Auto-encoder is that Anomaly is where reconstruction error is large.
#We can therefore empirically define this value beyond which we consider ANOMALY.
#This can make use of the mean error (ME) in training prediction and on testing data hence trainME and testME
#These two error estimates values are plotted together in histogram for visual comparison of the distribution

##############################################################################
```

#Section 4.4: Plot the error distribution based on both train and test data as a histogram to help determine the thresholds;

```
##############################################################################

#Figure 7 a) stacked train and test VV Flooded, b) VV Not Flooded, c) VH Flooded, d) VH Not Flooded: Stacked trainME (blue) and testME (orange) for co-polarisation)
plt.title(' a)  VH in Flood: Mean Error', size = 30, color = 'black')
train_predict = model.predict(trainX)
test_predict = model.predict(testX)
#plt.plot(trainX.reshape(-1, 1), trainY.reshape(-1, 1))
trainME = np.mean(train_predict - trainX, axis = 1) # absolute value was not considered since sign of the value distinguishes the wet (-) from The dry (+) anomalies
testME = np.mean(test_predict - testX, axis = 1) # absolute value was not considered since sign of the value distinguishes the wet (-) from The dry (+) anomalies
plt.hist(trainME, bins = 40, label = 'TrainME')
plt.hist(testME, bins = 40, label = 'TestME')
plt.xlabel('Mean Error (ME)', color='black', size=26)
plt.ylabel('Frequency', color='black', size=26)
plt.legend()

print(train_predict) # predicted timeseries based on trainX data
print(test_predict) # predicted timeseries based on trainX data
```

```
print(trainME) # train mean error time series
print(testME) # test mean error time series

#calculate the correlation metric between trainX and trainX_Predict
#x=trainX
#y=trainY

#print(np.corrcoef(x, x))
#plt.scatter(x, x, color = 'black', s = 200)
#plt.title('Observed and Predicted VV polarised data in the flood zone')
#plt.xlabel('Observed')
#plt.ylabel('Predicted')
#plt.plot(np.unique(x),np.poly1d(np.polyfit(x, x, 1))(np.unique(x)), color='red')
#plt.show()

##############################################################################
```

#Section 4.5A: Anomaly Detection: Calculate the model optimum/normal/mean on the time series mean (model_traintestME

```
##############################################################################
## Basic Statistics background
#mean = df['136FloVV'].mean()
#deviation = df['value'].mean
#std_deviation = deviation.std(); # OR
#std_deviation = df['value'].std
#z_score = deviation/std_deviation
#threshold = 3*std_deviation
#anomalies = df[deviation.abs() > threshold]
# Here the threshold was based on the training data extremes
# train data extrems were assumed to be normally distributed
# Therefore the two extremes were tested on test dataset to check for any random anomalies in the timeseries

# Determine the following metrics
mean_trainX = np.mean(trainX)
mean_train = np.mean(train)
mean_train_predict = np.mean(train_predict)
Mean_Train_PredictX = (mean_train_predict-mean_trainX)
```

```
print(Mean_Train_PredictX); # which is also the mean_trainME

#Extract boundaries to help determine empirical thresholds based on minimum, maximum and mean of the MEAN ERROR (NB: Not on the paper)
plt.title('VH in Flood', fontsize = 30)
test_predict = model.predict(testX)
testME = np.mean((test_predict - testX), axis=1)
plt.hist(testME, bins=40)

print(test_predict)
print(testME)
(testME).mean()

# to get min and max of first sub-array:
print(np.min(trainME))
print(Mean_Train_PredictX)
print(np.max(trainME))

print(np.min(testME))
print((testME).mean())
print(np.max(testME))

#std_deviation_testME = df[['testME']].std()

#Determine the EMPIRICAL (Wet and Dry) THRESHOLDS from the trainME and testME based on the plot in C: 4) above (Figures 1.3 and 1.4 respectively) to determine the dry (peak) and wet (trough)
negative_threshold = np.max(trainME)*0.75 # [could multiply by 0.997 if the data proved Gaussian in Section ]. Defined as the empirical Dry threshold (above which the testME is defined as a Dry Anomaly) determined using the trainME and testME histogram plot above.
normal_threshold = Mean_Train_PredictX # Defined as the empirical Normal threshold: Mean_TrainX_Predict = (mean_trainX + mean_trainX_Predict)/2 .
positive_threshold = np.min(trainME)*0.75 # [could multiply by 0.997 if the data proved Gaussian in Section ] #efined as the empirical Wet threshold (below which the testME is defined as a Wet Anomaly) also determined using the trainME and testME histogram plot above.

# OR
```

#negative_threshold = Mean_TrainX_PredictX + 3*std_deviation_testME #0.06 # Defined as the empirical Dry threshold (above which the testME is defined as a Dry Anomaly) determined using the trainME and testME histogram plot above.
#normal_threshold = Mean_TrainX_PredictX # -0.014 # Defined as the empirical Normal threshold: Mean_TrainX_Predict = (mean_trainX + mean_trainX_Predict)/2 .
#positive_threshold = Mean_TrainX_PredictX - 3*std_deviation_testME #-0.017 #efined as the empirical Wet threshold (below which the testME is defined as a Wet Anomaly) also determined using the trainME and testME histogram plot above.

```
print(negative_threshold)
print(normal_threshold)
print(positive_threshold)

#a] dry and wet anomalies definition and modelling
negative_anomaly_df = pd.DataFrame(test[seq_size:])
mean_normal_df = pd.DataFrame(test[seq_size:])
positive_anomaly_df = pd.DataFrame(test[seq_size:])
negative_anomaly_df['testME'] = testME
mean_normal_df['testME'] = testME
positive_anomaly_df['testME'] = testME

#dry_anomaly_df['meantrainX'] = meantrainX
#dry_anomaly_df['meantrainXPredict'] = meantrainXPredict
negative_anomaly_df['negative_threshold'] = negative_threshold
mean_normal_df['normal_threshold'] = normal_threshold
positive_anomaly_df['positive_threshold'] = positive_threshold

#Capture data attributes in one dataframe for easy plotting
#dry_anomaly_df['MeanTrainXPredict'] = MeanTrainXPredict
negative_anomaly_df['negative_anomaly'] = negative_anomaly_df['testME'] >
negative_anomaly_df['negative_threshold']
mean_normal_df['mean_normal'] = mean_normal_df['testME'] == mean_normal_df['normal_threshold']
positive_anomaly_df['positive_anomaly'] = positive_anomaly_df['testME'] <
positive_anomaly_df['positive_threshold']
negative_anomaly_df['136FloVH'] = test[seq_size:]['136FloVH']

#print(mean_normal)
```

##Figure 9: The VV in (a) anomaly detection, lag and memory in Mbire Floodplain, Zimbabwe on Loss ME on testing Plot testME vs max_trainMAE curve with threshold value shown as a red line

```
plt.title('a) VH in Flood: Thresholds', fontsize = 30)
sns.lineplot(x = negative_anomaly_df['Date'], y = negative_anomaly_df['negative_threshold'], color = 'red', label = 'negative threshold')
sns.lineplot(x = negative_anomaly_df['Date'], y = negative_anomaly_df['testME'], color = 'black', label='test mean error')
sns.lineplot(x = mean_normal_df['Date'], y = mean_normal_df['normal_threshold'], color = 'green', label = 'normal threshold')
sns.lineplot(x = positive_anomaly_df['Date'], y = positive_anomaly_df['positive_threshold'], color = 'blue', label = 'positive threshold')
plt.xlabel('Date (year-month)')
plt.ylabel('Mean Error')
plt.legend()

##############################################################################
#SECTION 5B - STEP 1A: Define both Dry and Wet anomalies using boolean function True/False,
##############################################################################

negative_anomalies = negative_anomaly_df.loc[negative_anomaly_df['negative_anomaly'] == True]
normal = mean_normal_df.loc[mean_normal_df['mean_normal'] == True]
positive_anomalies = positive_anomaly_df.loc[positive_anomaly_df['positive_anomaly'] == True]

#1B: Extract dates of both positive and negative Anomalies in Tabular Format
print(negative_anomalies)
print(normal)
print(positive_anomalies)

plt.title('c) VH in Flood: Anomaly Determination', fontsize = 30)
sns.lineplot(x = negative_anomaly_df['Date'], y = negative_anomaly_df['136FloVH'], label ='Test Mean Error', color = 'black')
sns.scatterplot(x = negative_anomalies['Date'], y = negative_anomalies['136FloVH'], color='red', label ='Negative Anomaly', s = 300)
sns.scatterplot(x = normal['Date'], y = normal['136FloVH'], color='green', label = 'Normal Level')
sns.scatterplot(x = positive_anomalies['Date'], y = positive_anomalies['136FloVH'], color='blue', label ='Positive Anomaly', s = 300)
plt.xlabel('Date (year-month)')
```

```
plt.ylabel('Maximum trainME(Sigma Nought)')
plt.legend()

################################################################################
#Step 5: LAG And MEMORY Calculation - Inference to soil moisture
################################################################################

# Step 5.1A: Determine the Wet Anomaly Lag inside the floodplain

positiveMaxDate = positive_anomalies[['Date']].max()
positiveMinDate = positive_anomalies[['Date']].min()
print('End of wet regime', positiveMaxDate)
print('Start of wet regime', positiveMinDate)

#Calculate FRSM Positive Lag
Positive_FRSM_Lag = positiveMaxDate - positiveMinDate
print('Positive FRSM Lag', Positive_FRSM_Lag)

#Step 5.1B: Determine the DRY ANOMALY LAG inside the Floodplain

negativeMaxDate = negative_anomalies[['Date']].max()
negativeMinDate = negative_anomalies[['Date']].min()
print('End of dry regime', negativeMaxDate)
print('Start of dry regime', negativeMinDate)

#Calculate FRSM Positive Lag
Negative_FRSM_Lag = negativeMaxDate - negativeMinDate
print('Negative FRSM Lag', Negative_FRSM_Lag)

################################################################################
# Call Train evaluation dataset for plotting the following Figures 12a - 12d
dataframe = pd.read_csv('C://Users//user//Desktop//TrainEval.csv')
df_train_eval_NFloVV = dataframe[['Date', '171NFloVH_Obs', '171NFloVH_Pred']] # Repeat these
four scenarios for TEST DATASET to plot Figure 12e) Test: VV Flood; 12f) Test: VV Not Flood; 12g)
Test: VH Flood  and 12h) Test: VH Not Flood
df_train_eval_NFloVV['Date'] = pd.to_datetime(df_train_eval_NFloVV['Date'])
```

#Figure 12: Plot of Observed versus Predicted (for TRAIN DATASET starting with Figure 12a) Train: VV Flood).

```
plt.title('b) VH in  Flood: Train and Test Prediction', size = 30)
#sns.lineplot(x=df['Date'], y=df['136FloVV'], label ='Flooded VV', color='darkblue')
sns.lineplot(x=df['Date'], y=df['trainPredict'], label='Flooded VH', color='blue', linewidth = 8)
sns.lineplot(x=df['Date'], y=df['testPredict'], label='Flooded VH', color='red', linewidth = 3)
#sns.lineplot(x=df['Date'], y=df['136FloVH'], label='136 in Flooded VH', color='blue')
#sns.lineplot(x=df['Date'], y=df['171NFloVV'], label='Not Flooded VV', color='green')
#sns.lineplot(x=df['Date'], y=df['171NFloVH'], label='Not Flooded VH', color='red')
#plt.legend()
```

Repeat for Figure 12b) Train: VV Not Flood, 12c) Train: VH Flood and 12d) Train: VH NotFlood)

```
# Call Test evaluation dataset for flotting the following Figures 12e- 12h)
dataframe = pd.read_csv('C://Users//user//Desktop//TestEval.csv')
df_test_eval_NFloVH = dataframe[['Date', '171NFloVH_Obs', '171NFloVH_Pred']] # Repeat these four scenarios for TEST DATASET to plot Figure 12e) Test: VV Flood; 12f) Test: VV Not Flood; 12g) Test: VH Flood  and 12h) Test: VH Not Flood
df_test_eval_NFloVH['Date'] = pd.to_datetime(df_test_eval_NFloVH['Date'])
```

Plot the Observed versus Predicted Test for Flooded VV polarisation

#Figure 8: LSTMAE evaluation of the test dataset (Figure 8a to 88d) and testing (Figure 8e to 8h) by location, (left = flood; right = non-flood) and polarisation; training VV (a, b) and VH(c, d) and testing VV(e, f) and VH (g, h) with the observed (thick light blue) while the rpedicted (thin black) lines respectively.

Singly run the following to generate Figures 8b to 8d individually

#Later after building model and running the predictions, the predicted vrsus observed, the pairwise datasets are as shown below

```
#df = dataframe[['Date', '136FloVV_Obs', '136FloVV_Pred']] # for Flooded Co-polarized location
df = dataframe[['Date', '136FloVH_Obs', '136FloVH_Pred']] # for Flooded Cross-polarized location
#df = dataframe[['Date', '171NFloVV_Obs', '171NFloVV_Pred']] # for Non-Flooded Co-polarized location
#df = dataframe[['Date', '171NFloVH_Obs', '171NFloVH_Pred']] # for Non-Flooded Cross-polarized location
```

repeat to run each line of the above for each graph (Figure 8e to 8h)

```
################################################################################
# Step 6.:Exploratory Data Analyses
################################################################################
# Point to the dataframe
dataframe = pd.read_csv('C://Users//user//Desktop//TestEval.csv')
# For ALL df time series are to be computed, then we command as follows;
df_all = dataframe[['Date', '136FloVV_Obs', '136FloVV_Pred', '136FloVH_Obs', '136FloVV_Pred',
'171NFloVV_Obs', '171NFloVV_Pred', '171NFloVH_Obs', '171NFloVH_Pred']]

df_all['Date'] = pd.to_datetime(df_all['Date'])

# Explore the dataset by plotting some time series using the following code;
# #Plot the fourscore time series based on polarisation (FIGURE 13a)
#Figure 13a: Time Series Plot of the DATA (e.g.SP136VV) by polarization and Flood Zone Location
(inside/outside)
plt.title('a) Time series by polarisation inside and outside flood', fontsize = 30, color ='black')
sns.lineplot(x=df_all['Date'], y=df_all['136FloVV_Obs'], label='VV inside flood', linewidth=5,
color='blue')
sns.lineplot(x=df_all['Date'], y=df_all['136FloVH_Obs'], label='VH inside flood', linewidth=5,
color='black')
sns.lineplot(x=df_all['Date'], y=df_all['171NFloVV_Obs'], label='VV outside flood', linewidth=5,
color='red')
sns.lineplot(x=df_all['Date'], y=df_all['171NFloVH_Obs'], label='VH outside flood', linewidth=5,
color='green')
plt.xlabel('Date (year-month)', fontsize = 26, color = 'black')
plt.ylabel('Sigma Nought', fontsize = 26, color = 'black')
plt.legend()

# Plot the fourscore time series based on polarisation (FIGURE 13b)
plt.title('b) Moving average by polarisation inside and outside flood', fontsize = 30, color ='black')
minFlo_roll = df_all[['136FloVV_Obs','136FloVH_Obs','171NFloVV_Obs', '171NFloVH_Obs']]
minFlo_roll.rolling(12).mean().plot(figsize=(20,8), linewidth=5, fontsize=26, color = ['blue', 'black',
'red', 'green'])
#plt.xlabel('Date', fontsize=26)

# Plot the pairwise time series based on polarisation (FIGURE 13c)
```

```
# Calculting the rolling difference ;
plt.title('c) Rolling difference by polarisation inside and outside flood', fontsize = 30, color ='black')
minFlo_roll.diff().plot(figsize=(20,8), linewidth=5, fontsize=26, color = ['blue', 'black', 'red', 'green']);
plt.xlabel('Date', fontsize=26, color = 'black');

################################################################################
#########################      End of Appendix 4.1      ##########################
```

Appendix 5.1: List of acronymns used in the Boruta variable importance projection

Attributes ingested into Boruta Variable Importance Projection for leaf chlorophyll content (LCC) (n; fieldwork = 18, statistically derived = 40, spatial = 3, shadow = 3)

Acronymn	Description of leaf by Light Exposure, Canopy Height, and Aspect
Б shadowMin	Minimum of Chlorophyll for Boruta 'shadow' sample leaves
† MWS	Chlorophyll for Middle of Canopy-Height, West-facing in-Shadow leaves
‡ CHccMmin	Minimum of Chlorophyll for Middle of Canopy-Height leaves
‡ EChlMin	Minimum of Chlorophyll for East-facing leaves
Б shadowMean	Mean of Chlorophyll for Boruta 'shadow' sample leaves
‡ CHccMmean	Mean of Chlorophyll for Middle of Canopy-Height leaves
‡ CHccBstdev	Standard deviation for LCC for Bottom Canopy-Height leaves
† MWL	Chlorophyll for Middle of Canopy-Height, West-facing, in-Light leaves
† MSL	Chlorophyll for Middle of Canopy-Height, South-facing, in-Light leaves
† MNL	Chlorophyll for Middle of Canopy-Height, North-facing, in-Light leaves
† MEL	Chlorophyll for Middle of Canopy-Height, East-facing, in-Light leaves
‡ WChlMean	Mean of LCC for West-facing leaves
‡ WChlStdev	Standard deviation of Chlorophyll of West-facing leaves
†MNS	Chlorophyll for Middle of Canopy-Height, North-facing, in-Shadow leaves
‡ LEccSmin	Minimum of Chlorophyll for leaves exposed to Shadow
‡ WChlMin	Minimum of Chlorophyll of West-facing leaves
‡ CHccMmax	Maximum of Chlorophyll for Middle of Canopy-Height leaves
† MSS	Chlorophyll for Middle of Canopy-Height, South-facing in-Shadow leaves
‡ GChlMin	Minimum of Chlorophyll for all leaves
‡ NChlMin	Mimimum of Chlorophyll for North-facing leaves
‡ LEccLmean	Mean of Chlorophyll for leaves exposed to Light
CHccTsdev	Standard deviation of Chlorophyll for Top of Canopy-Height leaves
EChlStdev	Stabdard deviation of Chlorophyll for East-facing leaves
SChlMean	Mean of Chlorophyll for South-facing leaves
† MES	Chlorophyll for Middle of Canopy-Height, East-facing, in-Shadow leaves
‡ CHccMstdev	Standard deviation of Chlorophyll for Middle-height
‡ GChlStdev	Standard deviation of Chlorophyll for all leaves
‡ SChMin	Minimum of Chlorophyll for all South-facing leaves

‡ LEccLmin	Minimum of Chlorophyll for leaves exposed to Light
‡ LEccLstdev	Standard deviation of Chlorophyll for leaves exposed to Light
‡ GChlMean	Mean of Chlorophyll for all leaves
† BSL	Chlorophyll for Bottom of Canopy-Height South-facing, in-Light leaves
† BNL	Chlorophyll for Bottom of Canopy-Height North-facing, in-Light leaves
‡ LEccSstdev	Standard deviation of Chlorophyll for leaves exposed to Shadow
‡ LEccSmean	Mean of Chlorophyll for leaves exposed to Shadow
† BSS	Chlorophyll for Bottom of Canopy-Height South-facing in-Shadow leaves
‡ EChlMax	Maximum Chlorophyll for East-facing leaves
‡ NChlMean	Mean of Chlorophyll for all North-facing leaves
† BWL	Chlorophyll for Bottom of Canopy-Height West-facing in-Light leaves
† BNS	Chlorophyll for Bottom of Canopy-Height, North-facing in-Shadow leaves
‡ EChlMax	Maximum of Chlorophyll for East-facing leaves
‡ NChlMean	Mean of Chlorophyll for North-facing leaves
‡ CHccBmin	Minimum Chlorophyll for Bottom of Canopy-Height

Acronymn	**Description of leaf by Light Exposure, Canopy Height, and Aspect**
† BWS	Chlorophyll for Bottom of Canopy-Height West-facing in-Shadow leaves
‡ SChlStdev	Standard deviation of Chlorophyll for South-facing leaves
‡ WChlMax	Maximum of LCC for all West-facing leaves
† BES	LCC for Bottom of Canopy-Height, East-facing in-Shadow leaves
Б shadowMax	Maximum of LCC for Boruta ‘shadow’ sample leaves
‡ NChlStdev	Standard deviation of LCC for all North-facing leaves
‡ NChlMax	Maximum of LCC for all north-facing leaves
† TL	LCC for all Top of Canopy-Height in-Light leaves
† BEL	LCC for Bottom of Canopy-Height East-side in-Light leaves
‡ CHccBmean	Mean chlorophyll content for bottom Canopy-Height leaves
‡ SChlMax	maximum of LCC for South-facing leaves
‡ CHccTmin	Minimum LCC for Top of Canopy-Height leaves
‡ LEccSmax	Maximum of LCC for Leaves Exposed to Shadow leaves
‡ GChlMax	Maximum of LCC for all leaves
‡ LeccLmax	Maximum of LCC for Leaves Exposed to Light
‡ CHccTmax	Maximum LCC for Top of Canopy-Height leaves
‡ CHccTmean	Mean LCC for Top of Canopy-Height leaves

‡ CHccBmax	Maximum LCC for Bottom of Canopy-Height leaves
† TS	LCC for Top of Canopy-Height in-Shadow leaves
§ Alt	Altitude
§ D2R	Distance to River
§ D2FP	Distance to Floodplain

---

† - Field sample leaves per tree (n = 18)

‡ - Statistically-derived attributes per tree (n = 40)

§ - Spatially-derivative attributes per tree (n = 3)

Б - Boruta shadow attributes per tree (n = 3)

Appendix 5.2: Script used to model *Z. mautitiana* Leaf Chlorophyll Content (LCC) in R

Machine learning code for variable importance projection for leaf chlorophyll content

 # EZRA PEDZISAI

 # PhD Candidate

 # University of KwaZulu Natal, Pietermaritzburg, SOUTH AFRICA

#

 # Leaf Chlorophyll Content Analyses

 # Step 5.2.1: Testing the Regression Problem

Assumption 1*:*

There exists a linear relationship (hence general linear model (glm) function) explaining floodplain class based on 1. **altitude, 2. chlorophyll**, and 3. **distance** (from river and floodplain) variables.

Call relevant **libraries** for use in the **glm calculation**;

library(DescTools)

library(manipulate)

Call and name the data (as 'data') in the correct directory;

data <- read.csv("C:/Users/user/Documents/Untitled Folder 1/Chlorophyll.csv", header = TRUE, sep = ",")

visualise the data

data

View the data structure by variable domain type;

str(data)

Attach the data for easy definition of data variables;

attach(data)

 # General Linear Modelling (glm)

Name the data independent and dependent variables to enable specified modelling

x = predictor variables; y = response variable (with a factor domain);

Model or fit the Floodplan binary locations (**Class**) using **Altitude**, **Chlorophyll**, and **Distance** variables; and **summarise** the model.

```
fit <- glm(Class ~ ., data = data, family = "binomial")
summary(fit)
```

Check for **significance** of results
If **NO** then assess if the causes are not among the following
Warning messages:
1: glm.fit: algorithm did not converge
2: glm.fit: fitted probabilities numerically 0 or 1 occurred
3. Others, etc.

##

PROBLEM:

 # 3. Wald's Test is used to test goodness of fit, which suffers from; (1) lack of invariance under reparameterization, (2) inaccuracy in small samples (which one is this as well) **YET** this is the most widely used test.

 # The HDE is less known (Yee 2022) yet may influence scientific conclusions

THERE4 there is a NEED to perform a HDE-free (hypotheses testing in scientific analysis. There are several options including Boruta variable selection for multiple variables, logistic regression, HDI, etc.

 # MOTIVATION for BORUTA use in feature selection (R-package)

Boruta is a wrapper, built around **randomForest Classification algorithms**, capable of capturing IMPORTANT features/variables in a dataset, using a shuffling concept/process that ensures randomness.
It duplicates the dataset into two, one measured and shadow variable.
It trains the classifier, uses **Mean Decrease Accuracy** or **Mean Decrease Impurity**.
(The Mean Decrease Accuracy plot expresses how much accuracy the model losses by excluding each variable. The more the accuracy suffers, the more important the variable is for the successful classification)
Checks for higher importance; higher z-score against maximum z-scores of its shadow feature to record **hits**

The method is iterative, with each iteration comparing z-score of shuffled features against original features.
Uses Variable importance measure (VIM)hence removes bad predictors
Method has 3 evaluation outcomes: Confirmed **important**, confirmed **unimportant**, and **Tentative**

#Therefore, Boruta is;
1. useful when there are **too many features, s**ome may be less useful/important ones in classification or prediction.
2. It is **time** consuming in computation,
3. When the accuracy is reduced due to weak simulation ability
4. computation may constraint **computational resources** (storage, calculation, …) challenges are expected

Application of the Boruta variable selection algorithm
The Boruta variable selection approach creates a **shadow attribute** # For each **attribute**, it by randomly shuffling all values.
It assesses **variable importance**; hence calculates/computes the variable importance measure (VIM)
As such Boruta variable selection becomes relevant to identify predictor to the locational attributes based on Floodplan binary characteristics.

Too many features/variables; time?, accuracy?, resources?
NB: Boruta maybe useful under these /such circumstances. **HDI** can also handle many variables

##

 # STEP 5.2.2A VARIABLE SELECTION USING BORUTA

 # Script for Boruta Variable Selection pipeline

LOAD requisite libraries

```
library(Boruta) # for variable selection

library(mlbench)
```

```
library(caret) # Call caret for model prediction

library(randomForest)

library(DescTools)

library(caTools) # Call CaTools package for Logistic Regression

library(e1071)

library(dplyr)


# Build a Linear MODEL of Floodplain characteristics using selected aspect (direction), height (form base to apex of canopy), and exposure to sunlight of leaf chlorophyll as predictor variable e.g., Leaf Chlorophyll Content with the data

GLM_LCC <- glm(Class ~ BEL+BES+MEL+MES+BSL+BSS+MSL+MSS+BWL+BWS+MWL+MWS+BNL+BNS+MNL+MNS+TL+TS, data = data, family = "binomial")

GLM_LCC

> summary(GLM_LCC)


borutaLCC <- Boruta(Class ~ BEL+BES+MEL+MES+BSL+BSS+MSL+MSS+BWL+BWS+MWL+MWS+BNL+BNS+MNL+MNS+TL+TS, data = data, doTrace = 2, maxRuns = 1000)

print(borutaLCC)

attStats(borutaLCC)

plotImpHistory(borutaLCC, cex.axis = 0.8, xlab = "Classifier Runs", ylab = "Importance")

plot(borutaLCC, las = 2, cex.axis = 0.8, xlab = substitute(paste(bold('Attributes'))), ylab = substitute(paste(bold('Importance'))), legend = “top left”)

borutaGlobal <- Boruta(Class ~ Alt+BEL+BES+MEL+MES+BSL+BSS+MSL+MSS+BWL+BWS+MWL+MWS+BNL+BNS+MNL+MNS+TL+TS+D2R+D2FP+GChlMean+EChlMean+SChlMean+WChlMean+NChlM
```

ean+GChlStdev+EChlStdev+SChlStdev+WChlStdev+NChlStdev+GChlMax+EChlMax+SChl
Max+WChlMax+NChlMax+GChlMin+EChlMin+SChlMin+WChlMin+NChlMin+CHccBma
x+CHccBmean+CHccBmin+CHccBstdev+CHccMmax+CHccMmean+CHccMmin+CHccMst
dev+CHccTmax+CHccTmean+CHccTmin+CHccTstdev+LEccLmax+LEccLmean+LEccLmin
+LEccLstdev+LEccSmax+LEccSmean+LEccSmin+LEccSstdev, las = 2, cex.axis = 0.8,
doTrace = 2, maxRuns = 1000

libraries:

#STEP 5.2.2B: DATA PARTITION FOR RF TRAIN[70]/TEST[30] PARAMETERIZATION

set.seed(112)

OPTION 1

```
library(dplyr)
library(caTools)
library(e1071)
# split data into training (70%) and testing (30%)
split <- sample.split(data, SplitRatio = 0.7)
split
# Calculate train dataset;
data_train <- subset(data, split == "TRUE")
# Calculate test dataset;
data_test <- subset(data, split == "FALSE")
# Check the dimensions of the train and test datasets respectively;
dim(data_train)
dim(data_test)
```

 # OPTION 2

```
#data$Class <- as.factor(data$Class)
#data
 #dim(data)
#split <- sample.split(data, SplitRatio = 0.7)
```

```
#split
#split <- sample.split(data, split=="TRUE")
#data_train$Class <- as.factor(data_train$Class)
#data_test$Class <- as.factor(data_test$Class)
```

 # STEP 5.2.2C: RANDOM FOREST MODEL TRAINING

```
data$Class <- as.factor(data$Class)
str(data)
data$Class <- as.factor(data$Class)
print(data$Class)
```

 # STEP 5.2.2D: Optimize the mtry

```
data$Class <- as.factor(data$Class)
set.seed(112)

#Alt+BEL+BES+MEL+MES+BSL+BSS+MSL+MSS+BWL+BWS+MWL+MWS+BNL+BNS+MNL+MNS+TL+TS+D2R+D2FP+Class+GChlMean+EChlMean+SChlMean+WChlMean+NChlMean+GChlStdev+EChlStdev+SChlStdev+WChlStdev+NChlStdev+GChlMax+EChlMax+SChlMax+WChlMax+NChlMax+GChlMin+EChlMin+SChlMin+WChlMin+NChlMin
#Alt+BEL+BES+ BSS+TL+TS+D2R+D2FP+EChlMax+SChlMax+ GChlMax+NChlMax

bestmtry <- tuneRF(data_train, data_train$Class, stepFactor = 1.2, improve = 0.01, trace = T, plot = T)

RF_train <- randomForest(Class ~ Alt+BEL+TL+TS+D2R+D2FP+GChlMax+NChlMax+NChlStdev+SChlMax+CHccTmin+LEccSmax+LEccLmax+CHccTmax+CHccTmean+CHccBmax, data = data_train, ntree = 500, keep.forest = FALSE, importance = TRUE)

RF_train

### Visualize variable importance

# Get variable importance from the model fit
```

```
ImpData <- as.data.frame(importance(RF_train))

ImpData$Var.Names <- row.names(ImpData)


var.imp1 <- data.frame(importance(RF_train, type = 2))
var.imp1$Variables <- row.names(var.imp1)
varimp1 <- var.imp1[order(var.imp1$MeanDecreaseGini, decreasing = T),]
par(mar = c(10, 5, 1, 1))
giniplot <- barplot(t(varimp1[-2]/sum(varimp1[-2])), las = 2,
                    cex.names = 1.3,
                   main = "Gini Impurity Index")
```

 # Step 5.2.2E: Make Predictions in the Test Data Using the Model

```
data_test$Class <- as.factor(data_test$Class)

pred_test <- predict (modelLCC, newdata = data_test, type = "Class")

pred_test

confusionMatrix(table(pred_test, data_test$Class)) # to extract Error Matrix,

confusionMatrix(table(pred_test, data_test$Class))


RF_test<-randomForest(formula = Class ~
Alt+BEL+TL+TS+D2R+D2FP+GChlMax+NChlMax+NChlStdev+SChlMax+CHccTmin+LE
ccSmax+LEccLmax+CHccTmax+CHccTmean+CHccBmax, data = data_test, ntree = 500,
keep.forest = FALSE, importance = TRUE)

RF_test

Call:
 randomForest(formula = Class ~ Alt + BEL + BES + BSS + TL + TS +D2R + D2FP +
EChlMax + SChlMax + NChlStdev + GChlMax + NChlMax,      data = data_train)
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 4
```

Mean of squared residuals: 0.02414623

% Var explained: 87.12

Type: Classification

Number of trees: 500

Number of variables at each split: 7

OOB est of error rate: 23.08%

Confusion matrix:

	Floodplain	**Not Floodplain**	**Class Error**
Floodplain			
Not Floodplain			

Prediction and Confusion Matrix Using the Test Data (30%) For RF_train

p1 <- predict(RF_train, test)

confusionMatrix(p1, test$Class)

Cofusion Matrix and Statistics

	REFERENCE	
PREDICTION	**Floodplain**	**Not Floodplain**
Floodplain		
Not Floodplain		

Accuracy :

95% CI :

No Information Rate :

P-Value [Acc > NIR] :

Kappa :

Mcnemar's Test P-Value:

Sensitivity :

Specificity :

Pos Pred Value :

Neg Pred Value :

Prevalence :

Detection Rate :

Detection Prevalence:

Balanced Accuracy :

##

Step 5.2.2F: Regression Using the Selected (Confirmed only) Variables in Boruta

where y is response variable, x is predictor variable(s); fit the model

fit <- glm(y ~ x, data = data, family = "binomial")

summary(fit)

Step 5.2.2G: Model Evaluation (Using ANOVA)

1. Exposure (LIT, SHD), level 1

2. Distance (D2R, D2FP), level 1

3. Aspect (E, S, W, N), level 2

4. Height (Bot, Mid, Top), level 2

to test while east-facing, bottom, sun exposed leaf to shade leaf,

anova(fitxBEL, fitxBES, test = "LRT")

to test for variability on aspect using ANOVA (E, S, W, N) facing;

anova(fitxE, fitxS, fitxW, fitxN, test = "LRT")

to test for variability using ANOVA for Bottom, Middle and Top,

anova(fitBot, fit Mid, fitTop, test = "LRT")

to test while south -facing, middle, sun exposed leaf to shade leaf,

anova(fitxMEL, fitxMES, test = "LRT")

to test while **Altitude** and **Chlorophyll** are removed

anova(fitxAll, fitxDist, test = "LRT")

LEVEL 1

LCC (BEL vs BES; MEL vs MES; BSL vs BSS; MSL vs MSS; BWL vs BWS; MWL vs MWS; BNL vs BNS; MNL vs MNS; TL vs TS

anova(fitNull, fit, test = "LRT")

##

 # Step 5.2.4: Likelihood Ratio Test (LRT) Method [ANOVA-based]

Feature Selection By Penalyzing Functions

#1. Likelihood Ratio Test (LRT) (STEP 4)

Likelihood Ratio Test (Lrt); of the Predictors (not subject to HDE affecting Wald's test, small samples, outlier affected samples, several insignificant variables,

the LRT does not suffer the HDE

anova(model1(e.g.,fitxAll), model2(e.g., fitxChl), test = "LRT")

to test while **Altitude** and **Chlorophyll** are removed

anova(fitxAll, fitxDist, test = "LRT")

Basic scatterplot Matrix

library(GGally)

#Scatterplot matrix of the first four variables of the dataframe

Import libraries

library(randomForest)

library(ggplot2)

set.seed(112)

data(data)

rf.fit <- randomForest(Status ~ ., data=data, ntree=1000, keep.forest=FALSE, importance=TRUE)

```
rf.fit

> rf.fit

var.imp1 <- data.frame(importance(fit, type=2))
var.imp1$Variables <- row.names(var.imp1)
varimp1 <- var.imp1[order(var.imp1$MeanDecreaseGini,decreasing = T),]

par(mar=c(10,5,1,1))
giniplot <- barplot(t(varimp1[-2]/sum(varimp1[-2])),las=2,
                    cex.names=1,
                   main="Gini Impurity Index Plot")

fit <- randomForest(TargetImpData ~.,importance = T,ntree = 500, data=data_train)
fit <- randomForest(ImpData ~.,importance = T,ntree = 500, data=data_train)

var.imp1 <- data.frame(importance(fit, type=2))
var.imp1 <- data.frame(importance(RF, type=2))
var.imp1$Variables <- row.names(var.imp1)

varimp1 <- var.imp1[order(var.imp1$MeanDecreaseGini,decreasing = T),]
par(mar=c(10,5,1,1))
giniplot <- barplot(t(varimp1[-2]/sum(varimp1[-2])),las=2,
+                    cex.names=1,
+                   main="Gini Impurity Index Plot")

giniplot <- barplot(t(varimp1[-2]/sum(varimp1[-2])), las=2,
cex.names = 1,
main = "Gini Impurity Index Plot"
)
giniplot <- barplot(t(varimp1[-2]/sum(varimp1[-2])), las=2,
cex.names = 1,
giniplot <- plot(t(varimp1[-2]/sum(varimp1[-2])), las = 2, xlab = "Variables", ylab = "Gini
    Impurity Index")
```

```
RF <- randomForest(varimp1 ~., importance = T, ntree = 500, data = data_train)
var.imp1 <- data.frame(importance(RF, type=1))
RF <- randomForest(varimpl ~., importance = T, ntree = 500, data = data_train)
varimp1 <- var.imp1[order(var.imp1$MeanDecreaseGini, decreasing = T),]
plot(varimp1, pch = 16, color = "blue", xlab = "Variables", ylab = "MeanDecreaseAccuracy")
var.imp1$Variables <- row.names(var.imp1)
varimp1 <- var.imp1[order(var.imp1$MeanDecreaseAccuracy, decreasing = T),]

plot(varimp1, las = 2)

ggplot2(ImpData, aes(x=Var.Names, y=`%IncMSE`)) +
+ geom_segment( aes(x=Var.Names, xend=Var.Names, y=0, yend=`%IncMSE`),
+ color="skyblue") +
+ geom_point(aes(size = IncNodePurity), color="blue", alpha=0.6) +
+ theme_light() +
+ coord_flip() +
+ theme(
+ legend.position="bottom",
+ panel.grid.major.y = element_blank(),
+ panel.border = element_blank(),
+ axis.ticks.y = element_blank()
+ )

fit <- randomForest(ImpData ~., importance = T, ntree = 500, data=data_train)

ImpData
var.imp1 <- data.frame(importance(fit, type=1))
var.imp1$Variables <- row.names(var.imp1)
var.imp1$Variables
varimp1 <- var.imp1[order(var.imp1$MeanDecreaseGini,decreasing = T),]
par(mar=c(10,5,1,1))
giniplot <- barplot(t(varimp1[-2]/sum(varimp1[-2])), las = 2, cex.names = 1.3, main =
    "Permutation Importance")
```

Appendix 6.1: Script for analysis of long-term tree biophysical characteristics in R programming language

```
import relevant libraries
library(DescTools)
library(manipulate)

flooded <- read.csv('C://Users/user/Downloads/Flood_Final.csv', header = TRUE)
attach(flooded)
str(flooded)
summary(flooded)
names(flooded)

model_TH <- glm(TH~Dist2Riv, data = flooded, family = gaussian)
summary(model_TH)
model_CD <- glm(CD~Dist2Riv, data = flooded, family = gaussian)
summary(model_CD)
model_DBH <- glm(DBH~Dist2Riv, data = flooded, family = gaussian)
summary(model_DBH)

notflooded <- read.csv('C://Users/user/Downloads/Not_Flood_Final.csv', header = TRUE)
attach(notflooded)
str(notflooded)
summary(notflooded)

flooded <- read.csv('C://Users/user/Downloads/Tree_Biophysical.csv', header = TRUE)
attach(flooded)
str(flooded)
summary(flooded)

ggplot(biophysical2, aes(x=Status, y=DBH)) +
geom_boxplot(fill="gray") +
labs(title="Plot of DBH (cm)",x="Status (mg)", y = "DBH")+
```

```
theme_classic()

data3 <- read.csv('C://Users/user/Downloads/Tree_Biophysical.csv', header = TRUE)
attach(data3)
str(data3)
summary(data3)
data3$Status <- as.factor(data3$Status)
summary(data3)

ggplot(data3, aes(x = Status, y = DBH, color = Status)) +
 geom_boxplot() +
 stat_violin()

names(data3)
boxplot(CD~Status, color = "tomato", cex = 1, horiz = T)
boxplot(TH~Status, color = "tomato", cex = 1, horiz = T)
boxplot(DBH~Status, color = "tomato", cex = 1, horiz = T)
boxplot(Dist2Riv~Status, color = "tomato", cex = 1, horiz = T)
boxplot(Dist2FP~Status, color = "tomato", cex = 1, horiz = T)

ggplot(data3, aes(x = Status, y = TH, fill=Status)) +
geom_boxplot()
ggplot(data3, aes(x = Status, y = DBH, fill=Status)) +
geom_boxplot()
ggplot(data3, aes(x = Status, y = CD, fill=Status)) +
geom_boxplot()

data3
mean(data3$Status="Flooded")
mean(data3$DBH~Status="Flooded")
mean(data3$DBH~Status(Flooded))
summary(Status)

data3$Status<-as.factor(data3$Status)
```

```
attatch(data3)
str(data3)
summary(data3)

shapiro.test(data3$DBH)
shapiro.test(data3$TH)
shapiro.test(data3$CD)

install.packages("olsrr")
library(olsrr)
reg.fit.DBH<-lm(data=data3, data3$DBH~Status)
summary(reg.fit.DBH)

reg.fit.CD<-lm(data=data3, data3$CD~Status)
summary(reg.fit.CD)

reg.fit.TH <- lm(data=data3, data3$TH~Status)
summary(reg.fit.TH)

reg.fit.DBH.Dist2Riv.<-lm(data=data3, data3$DBH~Dist2Riv)
summary(reg.fit.DBH.Dist2Riv.)
reg.fit.CD.Dist2Riv.<-lm(data=data3, data3$CD~Dist2Riv)
summary(reg.fit.CD.Dist2Riv.)
reg.fit.TH.Dist2Riv.<-lm(data=data3, data3$TH~Dist2Riv)
summary(reg.fit.TH.Dist2Riv.)

reg.fit.TH.Dist2FP<-lm(data=data3, data3$TH~Dist2FP)
summary(reg.fit.TH.Dist2FP)
reg.fit.CD.Dist2FP<-lm(data=data3, data3$CD~Dist2FP)
summary(reg.fit.CD.Dist2FP)
reg.fit.DBH.Dist2FP<-lm(data=data3, data3$DBH~Dist2FP)
summary(reg.fit.DBH.Dist2FP)

reg.fit.DBH.Dist2FP.Dist2Riv<-lm(data=data3, data3$DBH~Dist2FP+Dist2Riv)
```

```
summary(reg.fit.DBH.Dist2FP.Dist2Riv)
reg.fit.CD.Dist2FP.Dist2Riv<-lm(data=data3, data3$CD~Dist2FP+Dist2Riv)
summary(reg.fit.CD.Dist2FP.Dist2Riv)
reg.fit.TH.Dist2FP.Dist2Riv<-lm(data=data3, data3$TH~Dist2FP+Dist2Riv)
summary(reg.fit.TH.Dist2FP.Dist2Riv)

reg.fit.TH.Dist2FP.Dist2Riv.Status <- lm(data=data3, data3$TH~Dist2FP+Dist2Riv+Status)
summary(reg.fit.TH.Dist2FP.Dist2Riv.Status)
reg.fit.CD.Dist2FP.Dist2Riv.Status <- lm(data=data3, data3$CD~Dist2FP+Dist2Riv+Status)
summary(reg.fit.CD.Dist2FP.Dist2Riv.Status)
reg.fit.DBH.Dist2FP.Dist2Riv.Status <- lm(data=data3,
      data3$DBH~Dist2FP+Dist2Riv+Status)
summary(reg.fit.DBH.Dist2FP.Dist2Riv.Status)

summary(shapiro.test(data3$DBH))
summary(shapiro.test(data3$CD))
summary(shapiro.test(data3$TH))

ols_plot_resid_lev(reg.fit.TH)
ols_plot_resid_lev(reg.fit.DBH)
ols_plot_resid_lev(reg.fit.CD)

ols_plot_resid_lev(reg.fit.CD.Dist2Riv.)
ols_plot_resid_lev(reg.fit.DBH.Dist2Riv.)
ols_plot_resid_lev(reg.fit.TH.Dist2Riv.)

ols_plot_resid_lev(reg.fit.TH.Dist2FP)
ols_plot_resid_lev(reg.fit.CD.Dist2FP)
ols_plot_resid_lev(reg.fit.DBH.Dist2FP)

install.packages("jtools", "moments", "lmtest")
install.packages("jtools")
install.packages("lmtest")
install.packages("moments")
```

```
library(jtools)
library(moments0
library(lmtest)
library(lmtest)
library(moments)

ols_plot_resid_lev(reg.fit.DBH.Dist2FP.Dist2Riv)
ols_plot_resid_lev(reg.fit.CD.Dist2FP.Dist2Riv)
ols_plot_resid_lev(reg.fit.TH.Dist2FP.Dist2Riv)

ols_plot_resid_lev(reg.fit.CH.Dist2FP.Dist2Riv.Status)
ols_plot_resid_lev(reg.fit.CD.Dist2FP.Dist2Riv.Status)
ols_plot_resid_lev(reg.fit.DBH.Dist2FP.Dist2Riv.Status)

ols_plot_cooksd_chart(reg.fit.TH)
ols_plot_cooksd_chart(reg.fit.CD)
ols_plot_cooksd_chart(reg.fit.DBH)

ols_plot_cooksd_chart(reg.fit.DBH.Dist2Riv.)
ols_plot_cooksd_chart(reg.fit.CD.Dist2Riv.)
ols_plot_cooksd_chart(reg.fit.TH.Dist2Riv.)

ols_plot_cooksd_chart(reg.fit.TH.Dist2FP)
ols_plot_cooksd_chart(reg.fit.CD.Dist2FP)
ols_plot_cooksd_chart(reg.fit.DBH.Dist2FP)

ols_plot_cooksd_chart(reg.fit.DBH.Dist2FP.Dist2Riv)
ols_plot_cooksd_chart(reg.fit.CD.Dist2FP.Dist2Riv)
ols_plot_cooksd_chart(reg.fit.TH.Dist2FP.Dist2Riv)
```

Import libraries used for analyses

```
library(DescTools)

library(manipulate)

biophysical <- read.csv('C://Users/user/Downloads/Tree_Biophysical.csv', header = TRUE)

attach(biophysical)

str(biophysical)

summary(biophysical)

hist(CH, col ="blue", xlab= "tree height (m)", ylab = "count", main = "a)")

library(GGally)

ggpairs(biophysical[,2:8])


library(ggplot2)


ggplot(biophysical2, aes(x=Status, y=DBH)) +

geom_boxplot() +

labs(title="Plot of DBH (cm)",x="Status (mg)", y = "DBH")+

theme_classic()


ggplot(biophysical, aes(x = Dist2Riv, y = DBH, color = Status)) +

 geom_point() +

 stat_ellipse()


ggplot2(data3, aes(x = Status, y = DBH)) +

 geom_boxplot(col = Status) +
```

```
geom_point(data3 = DBH,
           aes(x = Satus, y = mean),
           shape = 3,
           size = 2)

# T-Test
comp %>% t_test(freedom_of_choice ~ Status,
              var.equal = TRUE,
              detailed = TRUE) %>%
  glimpse()

boxplot(DBH ~ Status, data = data3, aes(color = “blue”), main = "a) DBH by distance to river",
xlab = "Tree Status", ylab = "DBH (cm)")

biophysical2 <- read.csv('C://Users/user/Downloads/Tree_Biophysical.csv', header = TRUE)
group_by(biophysical3, CH) %>%
  summarise(
  count = n(),
  mean = mean(CH, na.rm = TRUE),
  sd = sd(CH, na.rm = TRUE)
  )

plot(boruta_16, las = 2, cex.axis = 0.8, xlab = substitute(paste(bold('Attributes'))), ylab =
substitute(paste(bold('Importance'))), legend = "top left")
```

```
> library(dplyr)
> library(caTools)
> library(e1071)
> set.seed(123)
> biophysical2$Code <- as.factor(biophysical2$Code)
> dim(biophysical2)
[1] 366  11
split <- sample.split(biophysical2, SplitRatio = 0.7)
split
biophysical2$Floodplain <- as.factor(biophysical2$Floodplain)
biophysical2$Status <- as.factor(biophysical2$Status)
data_train <- subset(biophysical2, split == "TRUE")
data_train
> data_train <- subset(biophysical2, split == "TRUE")
> data_test <- subset(biophysical2, split == "FALSE")
> data_test

> RF_train <- randomForest(Code ~ BSA+CH+CD+DBH+STEM+Dist2FP+Dist2Riv, data =
data_train)
> RF_train

Call:
 randomForest(formula = Code ~ BSA + CH + CD + DBH + STEM + Dist2FP +
Dist2Riv, data = data_train)
```

Type of random forest: classification

Number of trees: **500**

No. of variables tried at each split: **2**

OOB estimate of error rate: **0.43%**

Confusion matrix:

	0	1	class.error
0	109	1	0.009090909
1	0	124	0.000000000

>

> RF_test <- randomForest(Code ~ BSA + CH + CD + DBH + STEM + Dist2FP + Dist2Riv, data = data_test)

> RF_test

Call:

randomForest(formula = Code ~ BSA + CH + CD + DBH + STEM + Dist2FP + Dist2Riv, data = data_test)

Type of random forest: classification

Number of trees: 500

No. of variables tried at each split: 2

OOB estimate of error rate: 0.76%

Confusion matrix:

0	1	class.error

0 60 1 0.01639344

1 0 71 0.00000000

> ImpData <- as.data.frame(importance(RF_train))

> ImpData

	MeanDecreaseGini
TH	2.546450
CD	3.125230
DBH	5.371343
Dist2FP	78.565928
Dist2Riv	19.833061

ImpData <- as.data.frame(importance(RF_test))

ImpData

	MeanDecreaseGini
TH	2.284895
CD	2.216108
DBH	2.975566
Dist2FP	41.252423
Dist2Riv	12.774547

pred_test <- predict (RF_test, newdata = data_test, type = "Class")

pred_test

confusionMatrix(table(pred_test, data_test$Code))

Confusion Matrix and Statistics

pred_test	0	1
0	61	0
1		0 71

Accuracy : 1
95% CI: (0.9724, 1)
No Information Rate: 0.5379
P-Value [Acc > NIR] : < 2.2e-16
Kappa : 1
Mcnemar's Test P-Value : NA
Sensitivity : 1.0000
Specificity : 1.0000
Pos Pred Value : 1.0000
Neg Pred Value : 1.0000
Prevalence : 0.4621
Detection Rate : 0.4621
Detection Prevalence : 0.4621
Balanced Accuracy : 1.0000
'Positive' Class : 0

confusionMatrix(table(pred_train, data_train$Code))
Confusion Matrix and Statistics

pred_train	0	1
0	110	0
1	0	124

Accuracy : 1
95% CI : (0.9844, 1)
No Information Rate : 0.5299
P-Value [Acc > NIR] : < 2.2e-16

```
               Kappa : 1

 Mcnemar's Test P-Value : NA

            Sensitivity : 1.0000
            Specificity : 1.0000
         Pos Pred Value : 1.0000
         Neg Pred Value : 1.0000
             Prevalence : 0.4701
         Detection Rate : 0.4701
   Detection Prevalence : 0.4701
      Balanced Accuracy : 1.0000

       'Positive' Class : 0

> library(ggplot2)
> library(car)

ggplot(biophysical2, aes(x=Status, y=CD, fill = Status)) +
 geom_violin() +
 xlab("Tree status") + ylab("Canopy diameter (m)") +
 theme_classic()+scale_fill_manual(values=c("brown2"," cyan4"))+
 stat_summary(fun.y=mean,  geom="point", color="black")+
 theme(legend.position="none")+
 theme(aspect.ratio=1)


varImpPlot(RF_test, col = "black", pch = 19)

Call:

 randomForest(formula = Code ~ BSA + CH + CD + DBH + STEM + Dist2FP +      Dist2Riv,
data = data_train)

               Type of random forest: classification
```

Number of trees: 500

No. of variables tried at each split: 2

OOB estimate of error rate: 0.43%

Confusion matrix:

```
   0   1 class.error
0 109   1 0.009090909
1   0 124 0.000000000
```

> print(RF_train)

Call:

randomForest(formula = Code ~ BSA + CH + CD + DBH + STEM + Dist2FP + Dist2Riv, data = data_train)

Type of random forest: classification

Number of trees: 500

No. of variables tried at each split: 2

OOB estimate of error rate: 0.43%

Confusion matrix:

```
   0   1 class.error
0 109   1 0.009090909
1   0 124 0.000000000
```

> ImpDataTrain <- as.data.frame(importance(RF_train))

> ImpDataTrain

MeanDecreaseGini

BSA	3.315582
CH	2.546450
CD	3.125230
DBH	5.371343
STEM	3.270986
Dist2FP	78.565928
Dist2Riv	19.833061

> varImpPlot(RF_train, col = "blue", pch = 19)

> RF_test

Call:

randomForest(formula = Code ~ BSA + CH + CD + DBH + STEM + Dist2FP + Dist2Riv, data = data_test)

Type of random forest: classification

Number of trees: 500

No. of variables tried at each split: 2

OOB estimate of error rate: 0.76%

Confusion matrix:

	0	1	class.error
0	60	1	0.01639344
1	0	71	0.00000000

> fit10 <- glm(Code ~ DBH + Dist2Riv + Dist2FP, data = biophysical2, family = binomial)

```
Warning messages:
1: glm.fit: algorithm did not converge
2: glm.fit: fitted probabilities numerically 0 or 1 occurred
> summary(fit10)

Call:
glm(formula = Code ~ DBH + Dist2Riv + Dist2FP, family = binomial,
    data = biophysical2)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  15.11802  339.82331   0.044    0.965
DBH          -0.01417    4.41965  -0.003    0.997
Dist2Riv      0.01487    2.34238   0.006    0.995
Dist2FP     -13.82654  139.81084  -0.099    0.921

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5.0581e+02  on 365  degrees of freedom
Residual deviance: 5.7914e-05  on 362  degrees of freedom
AIC: 8.0001

Number of Fisher Scoring iterations: 25
```

```
>  fit11 <- glm(Code ~ Dist2Riv + Dist2FP, data = biophysical2, family = binomial)
Warning messages:
1: glm.fit: algorithm did not converge
2: glm.fit: fitted probabilities numerically 0 or 1 occurred
> summary(fit11)

Call:
glm(formula = Code ~ Dist2Riv + Dist2FP, family = binomial, data = biophysical2)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  14.2452   199.9856   0.071    0.943
Dist2Riv      0.0154     2.4378   0.006    0.995
Dist2FP     -13.8689   142.4826  -0.097    0.922

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5.0581e+02  on 365  degrees of freedom
Residual deviance: 5.8564e-05  on 363  degrees of freedom
AIC: 6.0001

Number of Fisher Scoring iterations: 25

>  fit12 <- glm(Code ~ Dist2Riv, data = biophysical2, family = binomial)
```

```
> summary(fit12)

Call:
glm(formula = Code ~ Dist2Riv, family = binomial, data = biophysical2)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.0660854  0.2301976   8.975   <2e-16 ***
Dist2Riv    -0.0044773  0.0005166  -8.666   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 505.81  on 365  degrees of freedom
Residual deviance: 339.61  on 364  degrees of freedom
AIC: 343.61

Number of Fisher Scoring iterations: 6

>  fit13 <- glm(Code ~ Dist2FP, data = biophysical2, family = binomial)
Warning messages:
1: glm.fit: algorithm did not converge
2: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
> summary(fit13)

Call:
glm(formula = Code ~ Dist2FP, family = binomial, data = biophysical2)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    16.16     140.43   0.115    0.908
Dist2FP       -14.39     181.10  -0.079    0.937

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5.0581e+02  on 365  degrees of freedom
Residual deviance: 4.3924e-05  on 364  degrees of freedom
AIC: 4

Number of Fisher Scoring iterations: 25

>  fit14 <- glm(Code ~ DBH, data = biophysical2, family = binomial)
>  fit14

Call:  glm(formula = Code ~ DBH, family = binomial, data = biophysical2)

Coefficients:
```

(Intercept) DBH

-0.40895 0.01294

Degrees of Freedom: 365 Total (i.e. Null); 364 Residual

Null Deviance: 505.8

Residual Deviance: 492.8 AIC: 496.8

> summary(fit14)

Call:

glm(formula = Code ~ DBH, family = binomial, data = biophysical2)

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -0.408945 0.184312 -2.219 0.026503 *

DBH 0.012941 0.003684 3.512 0.000444 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 505.81 on 365 degrees of freedom

Residual deviance: 492.76 on 364 degrees of freedom

AIC: 496.76

Number of Fisher Scoring iterations: 4

> fit15 <- glm(Code ~ CD, data = biophysical2, family = binomial)

> summary(fit15)

Call:

glm(formula = Code ~ CD, family = binomial, data = biophysical2)

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.67667	0.23193	-2.918	0.003527	**
CD	0.08549	0.02242	3.814	0.000137	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 505.81 on 365 degrees of freedom

Residual deviance: 489.13 on 364 degrees of freedom

AIC: 493.13

Number of Fisher Scoring iterations: 4

> fit16 <- glm(Code ~ BSA, data = biophysical2, family = binomial)

```
> summary(fit16)

Call:
glm(formula = Code ~ BSA, family = binomial, data = biophysical2)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.235728   0.142654  -1.652 0.098443 .
BSA          0.004109   0.001171   3.510 0.000448 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 505.81  on 365  degrees of freedom
Residual deviance: 489.40  on 364  degrees of freedom
AIC: 493.4

Number of Fisher Scoring iterations: 4

>  fit17 <- glm(Code ~ STEM, data = biophysical2, family = binomial)
> summary(fit17)

Call:
```

glm(formula = Code ~ STEM, family = binomial, data = biophysical2)

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.37305	0.17058	-2.187	0.028749	*
STEM	0.19792	0.05567	3.555	0.000378	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 505.81 on 365 degrees of freedom

Residual deviance: 489.38 on 364 degrees of freedom

AIC: 493.38

Number of Fisher Scoring iterations: 4

> fit18 <- glm(Code ~ CH, data = biophysical2, family = binomial)

> summary(fit18)

Call:

glm(formula = Code ~ CH, family = binomial, data = biophysical2)

Coefficients:

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.91842    0.32741  -2.805 0.005030 **
CH           0.13582    0.04038   3.364 0.000769 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 505.81  on 365  degrees of freedom
Residual deviance: 493.83  on 364  degrees of freedom
AIC: 497.83

Number of Fisher Scoring iterations: 4

>  fit19 <- glm(Code ~ Dist2Riv +Dist2FP + Dist2Riv:Dist2FP, data = biophysical2, family =
binomial)
Warning messages:
1: glm.fit: algorithm did not converge
2: glm.fit: fitted probabilities numerically 0 or 1 occurred
> summary(fit19)

Call:
glm(formula = Code ~ Dist2Riv + Dist2FP + Dist2Riv:Dist2FP, family = binomial,
    data = biophysical2)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	15.475089	346.727743	0.045	0.964
Dist2Riv	0.013487	3.876896	0.003	0.997
Dist2FP	-15.199955	247.636634	-0.061	0.951
Dist2Riv:Dist2FP	0.002541	0.041295	0.062	0.951

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 5.0581e+02 on 365 degrees of freedom

Residual deviance: 1.9377e-05 on 362 degrees of freedom

AIC: 8

Number of Fisher Scoring iterations: 25

> fit20 <- glm(Code ~ Dist2Riv +DBH + Dist2Riv:DBH, data = biophysical2, family = binomial)

> summary(fit20)

Call:

glm(formula = Code ~ Dist2Riv + DBH + Dist2Riv:DBH, family = binomial,

data = biophysical2)

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)

(Intercept)   1.573e+00 4.023e-01   3.912 9.17e-05 ***

Dist2Riv   -3.850e-03 8.675e-04 -4.438 9.06e-06 ***

DBH   1.118e-02 8.169e-03   1.369   0.171

Dist2Riv:DBH -1.457e-05 1.897e-05 -0.768   0.442

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 505.81 on 365 degrees of freedom

Residual deviance: 337.18 on 362 degrees of freedom

AIC: 345.18

Number of Fisher Scoring iterations: 6

> fit21 <- glm(Code ~ Dist2FP +DBH + Dist2FP:DBH, data = biophysical2, family = binomial)

Warning messages:

1: glm.fit: algorithm did not converge

2: glm.fit: fitted probabilities numerically 0 or 1 occurred

> summary(fit21)

Call:

glm(formula = Code ~ Dist2FP + DBH + Dist2FP:DBH, family = binomial,

    data = biophysical2)

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	2.858e+15	6.693e+06	427095110	<2e-16	***
Dist2FP	-4.040e+12	8.123e+03	-497417297	<2e-16	***
DBH	-3.349e+13	1.242e+05	-269664059	<2e-16	***
Dist2FP:DBH	-5.963e+09	2.322e+02	-25682894	<2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 505.81 on 365 degrees of freedom

Residual deviance: 9299.26 on 362 degrees of freedom

AIC: 9307.3

Number of Fisher Scoring iterations: 25

> fit22 <- glm(Code ~ Dist2FP +CH + Dist2FP:CH, data = biophysical2, family = binomial)

Warning messages:

1: glm.fit: algorithm did not converge

2: glm.fit: fitted probabilities numerically 0 or 1 occurred

> summary(fit22)

```
Call:
glm(formula = Code ~ Dist2FP + CH + Dist2FP:CH, family = binomial,
    data = biophysical2)

Coefficients:
              Estimate Std. Error    z value Pr(>|z|)
(Intercept)  1.740e+15  1.162e+07  149734714   <2e-16 ***
Dist2FP     -5.776e+12  1.899e+04 -304197076   <2e-16 ***
CH          -1.866e+14  1.381e+06 -135164097   <2e-16 ***
Dist2FP:CH   2.082e+11  2.619e+03   79495953   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance:  505.81  on 365  degrees of freedom
Residual deviance: 6487.86  on 362  degrees of freedom
AIC: 6495.9

Number of Fisher Scoring iterations: 25

>  fit23 <- glm(Code ~ Dist2FP +CH + Dist2FP:CH, data = biophysical2, family = binomial)
Warning messages:
1: glm.fit: algorithm did not converge
```

```
2: glm.fit: fitted probabilities numerically 0 or 1 occurred
> summary(fit23)

Call:
glm(formula = Code ~ Dist2FP + CH + Dist2FP:CH, family = binomial,
    data = biophysical2)

Coefficients:
              Estimate Std. Error    z value Pr(>|z|)
(Intercept)  1.740e+15  1.162e+07  149734714   <2e-16 ***
Dist2FP     -5.776e+12  1.899e+04 -304197076   <2e-16 ***
CH          -1.866e+14  1.381e+06 -135164097   <2e-16 ***
Dist2FP:CH   2.082e+11  2.619e+03   79495953   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance:  505.81  on 365  degrees of freedom
Residual deviance: 6487.86  on 362  degrees of freedom
AIC: 6495.9

Number of Fisher Scoring iterations: 25
```

```
>  fit24 <- glm(Code ~ Dist2Riv +CH + Dist2Riv:CH, data = biophysical2, family = binomial)
> summary(fit24)

Call:
glm(formula = Code ~ Dist2Riv + CH + Dist2Riv:CH, family = binomial,
    data = biophysical2)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.2774632  0.7265541   3.135 0.001721 **
Dist2Riv    -0.0052356  0.0015630  -3.350 0.000809 ***
CH          -0.0285412  0.0843977  -0.338 0.735232
Dist2Riv:CH  0.0001065  0.0001982   0.537 0.591063
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 505.81  on 365  degrees of freedom
Residual deviance: 339.29  on 362  degrees of freedom
AIC: 347.29

Number of Fisher Scoring iterations: 6
```

```
>  fit25 <- glm(Code ~ Dist2Riv +CD + Dist2Riv:CD, data = biophysical2, family = binomial)
> summary(fit25)

Call:
glm(formula = Code ~ Dist2Riv + CD + Dist2Riv:CD, family = binomial,
    data = biophysical2)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.113e+00  4.672e-01   4.522 6.12e-06 ***
Dist2Riv    -4.478e-03  1.150e-03  -3.893 9.92e-05 ***
CD          -3.599e-03  3.895e-02  -0.092    0.926
Dist2Riv:CD -3.681e-06  1.284e-04  -0.029    0.977
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 505.81  on 365  degrees of freedom
Residual deviance: 339.58  on 362  degrees of freedom
AIC: 347.58

Number of Fisher Scoring iterations: 6
```

```
>  fit26 <- glm(Code ~ Dist2Riv +CD + Dist2Riv:CD, data = biophysical2, family = binomial)
> summary(fit26)

Call:
glm(formula = Code ~ Dist2Riv + CD + Dist2Riv:CD, family = binomial,
    data = biophysical2)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.113e+00  4.672e-01   4.522 6.12e-06 ***
Dist2Riv    -4.478e-03  1.150e-03  -3.893 9.92e-05 ***
CD          -3.599e-03  3.895e-02  -0.092    0.926
Dist2Riv:CD -3.681e-06  1.284e-04  -0.029    0.977
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 505.81  on 365  degrees of freedom
Residual deviance: 339.58  on 362  degrees of freedom
AIC: 347.58

Number of Fisher Scoring iterations: 6
```

```
> fit30 <- glm(CD ~ Dist2Riv + Dist2FP, data = biophysical2, family = gaussian)
> summary(fit30)

Call:
glm(formula = CD ~ Dist2Riv + Dist2FP, family = gaussian, data = biophysical2)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.980205   0.413783  26.536  < 2e-16 ***
Dist2Riv    -0.003710   0.001103  -3.365 0.000848 ***
Dist2FP      0.002626   0.001193   2.201 0.028354 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 28.06256)

    Null deviance: 10782  on 365  degrees of freedom
Residual deviance: 10187  on 363  degrees of freedom
AIC: 2264.1

Number of Fisher Scoring iterations: 2

> fit31 <- glm(BSA ~ Dist2Riv + Dist2FP, data = biophysical2, family = gaussian)
> summary(fit31)
```

Call:

glm(formula = BSA ~ Dist2Riv + Dist2FP, family = gaussian, data = biophysical2)

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	126.97011	9.44560	13.442	< 2e-16	***
Dist2Riv	-0.08803	0.02517	-3.497	0.000528	***
Dist2FP	0.06500	0.02723	2.387	0.017506	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 14623.12)

Null deviance: 5614609 on 365 degrees of freedom

Residual deviance: 5308193 on 363 degrees of freedom

AIC: 4553.7

Number of Fisher Scoring iterations: 2

> fit32 <- glm(CH ~ Dist2Riv + Dist2FP, data = biophysical2, family = gaussian)

> summary(fit32)

Call:

glm(formula = CH ~ Dist2Riv + Dist2FP, family = gaussian, data = biophysical2)

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	8.2815153	0.2125908	38.955	<2e-16	***
Dist2Riv	-0.0013123	0.0005665	-2.317	0.0211	*
Dist2FP	0.0008257	0.0006130	1.347	0.1788	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 7.407474)

Null deviance: 2782.6 on 365 degrees of freedom

Residual deviance: 2688.9 on 363 degrees of freedom

AIC: 1776.6

Number of Fisher Scoring iterations: 2

> fit33 <- glm(DBH ~ Dist2Riv + Dist2FP, data = biophysical2, family = gaussian)

> summary(fit33)

Call:

glm(formula = DBH ~ Dist2Riv + Dist2FP, family = gaussian, data = biophysical2)

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	45.671005	2.328496	19.614	<2e-16	***
Dist2Riv	-0.003504	0.006205	-0.565	0.573	

Dist2FP -0.003751 0.006714 -0.559 0.577

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 888.6525)

Null deviance: 333558 on 365 degrees of freedom

Residual deviance: 322581 on 363 degrees of freedom

AIC: 3528.7

Number of Fisher Scoring iterations: 2

> global <- glm(Code~ BSA + CH + CD + DBH + STEM + Dist2FP + Dist2Riv, data = biophysical2, family = binomial)

Warning messages:

1: glm.fit: algorithm did not converge

2: glm.fit: fitted probabilities numerically 0 or 1 occurred

> summary(global)

Call:

glm(formula = Code ~ BSA + CH + CD + DBH + STEM + Dist2FP + Dist2Riv,

family = binomial, data = biophysical2)

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 3.274e+02 2.536e+04 0.013 0.990

BSA 2.861e+00 2.493e+02 0.011 0.991

CH 8.465e-01 1.897e+02 0.004 0.996

```
CD          -5.738e+01  4.383e+03  -0.013    0.990
DBH          2.309e-01  1.585e+01   0.015    0.988
STEM         2.028e+01  1.320e+03   0.015    0.988
Dist2FP     -1.053e+01  1.264e+02  -0.083    0.934
Dist2Riv     2.056e-02  2.031e+00   0.010    0.992

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5.0581e+02  on 365  degrees of freedom
Residual deviance: 4.0937e-05  on 358  degrees of freedom
AIC: 16

Number of Fisher Scoring iterations: 25

print(n=366,
+ group_by(biophysical3, CH) %>%
+   summarise(
+   count = n(),
+   mean = mean(CH, na.rm = TRUE),
+   sd = sd(CH, na.rm = TRUE)
+   )
+ )
```

```
ggplot(scatterplot_data.plot, aes(x=Status, y=CD, fill = Status)) +
 geom_boxplot() +
 xlab("Tree flood status") + ylab("Canopy diameter (m)") +
 theme_classic()+scale_fill_manual(values=c("brown2"," cyan4"))+
 stat_summary(fun.y=mean,  geom="point", color="black")+
 theme(legend.position="none")+
 theme(aspect.ratio=1)


ggplot(scatterplot_data.plot, aes(x=Status, y=CH, fill = Status)) +
 geom_boxplot() +
 xlab("Tree flood status") + ylab("Tree height (m)") +
 theme_classic()+scale_fill_manual(values=c("brown2"," cyan4"))+
 stat_summary(fun.y=mean,  geom="point", color="black")+
 theme(legend.position="none")+
 theme(aspect.ratio=1)
ggplot(scatterplot_data.plot, aes(x=Status, y=DBH, fill = Status)) +
 geom_boxplot() +
 xlab("Tree flood status") + ylab("Diameter at breast height (cm)") +
 theme_classic()+scale_fill_manual(values=c("brown2"," cyan4"))+
 stat_summary(fun.y=mean,  geom="point", color="black")+
 theme(legend.position="none")+
 theme(aspect.ratio=1)


ggplot(data3, aes(x = Dist2Riv, y = CD, color = Status)) +

+ geom_point()


ggbetweenstats(
  data,
  x = Status,
  y = DBH,
  type = "parametric",
  pairwise.display = "significant",
```

```
  p.adjust.method = "holm",
  effsize.type = "unbiased",
  bf.prior = 0.707,
  bf.message = TRUE,
  results.subtitle = TRUE,
  xlab = NULL,
  ylab = NULL,
  caption = NULL,
  title = NULL,
  subtitle = NULL,
  digits = 2L,
  var.equal = FALSE,
  conf.level = 0.95,
  nboot = 100L,
  tr = 0.2,
  centrality.plotting = TRUE,
  centrality.type = type,
  centrality.point.args = list(size = 5, color = "darkred"),
  centrality.label.args = list(size = 3, nudge_x = 0.4, segment.linetype = 4,
    min.segment.length = 0),
  point.args = list(position = ggplot2::position_jitterdodge(dodge.width = 0.6), alpha =
    0.4, size = 3, stroke = 0, na.rm = TRUE),
  boxplot.args = list(width = 0.3, alpha = 0.2, na.rm = TRUE),
  violin.args = list(width = 0.5, alpha = 0.2, na.rm = TRUE),
  ggsignif.args = list(textsize = 3, tip_length = 0.01, na.rm = TRUE),
  ggtheme = ggstatsplot::theme_ggstatsplot(),
  package = "RColorBrewer",
  palette = "Dark2",
  ggplot.component = NULL,
  ...
)
```

```
# Tree long term model

floodedtreemodel_DBH
notfloodedtreemodel_DBH <- lm(data = data3, DBH~data3$Status!="Flooded")
notfloodedtreemodel_DBH
flooded_distance2river_DBH <- lm(data = data3,
      DBH~data3$Dist2Riv+data3$Status=="Flooded")
data3$Status <- as.factor(data3$Status)
flooded_distance2river_DBH <- lm(data = data3,
      DBH~data3$Dist2Riv+data3$Status=="Flooded")
str(data3)
summary(floodedtreemodel_DBH)
ols_plot_resid_hist(reg.fit.TH)
ols_plot_resid_hist(reg.fit.CD)
ols_plot_resid_hist(reg.fit.DBH)
ols_plot_resid_hist(reg.fit.DBH.Dist2Riv.)
ols_plot_resid_hist(reg.fit.CD.Dist2Riv.)
ols_plot_resid_hist(reg.fit.TH.Dist2Riv.)
ols_plot_resid_hist(reg.fit.TH.Dist2FP)
ols_plot_resid_hist(reg.fit.CD.Dist2FP)
ols_plot_resid_hist(reg.fit.DBH.Dist2FP)
ols_plot_resid_hist(reg.fit.DBH.Dist2FP.Status)
ols_plot_resid_hist(reg.fit.DBH.Dist2FP.Dist2Riv)
ols_plot_resid_hist(reg.fit.CD.Dist2FP.Dist2Riv)
ols_plot_resid_hist(reg.fit.TH.Dist2FP.Dist2Riv)
ols_plot_resid_hist(reg.fit.TH.Dist2FP.Dist2Riv.Status)
ols_plot_resid_hist(reg.fit.CD.Dist2FP.Dist2Riv.Status)
ols_plot_resid_hist(reg.fit.DBH.Dist2FP.Dist2Riv.Status)

#QQ-PLOT
ols_plot_resid_qq(reg.fit.DBH)
ols_plot_resid_qq(reg.fit.CD)
ols_plot_resid_qq(reg.fit.TH)
```

```
ols_plot_resid_qq(reg.fit.TH.Dist2Riv.)
ols_plot_resid_qq(reg.fit.CD.Dist2Riv.)
ols_plot_resid_qq(reg.fit.DBH.Dist2Riv.)

ols_plot_resid_qq(reg.fit.DBH.Dist2FP)
ols_plot_resid_qq(reg.fit.CD.Dist2FP)
ols_plot_resid_qq(reg.fit.TH.Dist2FP)

ols_plot_resid_qq(reg.fit.CH.Dist2FP.Dist2Riv)
ols_plot_resid_qq(reg.fit.CD.Dist2FP.Dist2Riv)
ols_plot_resid_qq(reg.fit.DBH.Dist2FP.Dist2Riv)

ols_plot_resid_qq(reg.fit.DBH.Dist2FP.Dist2Riv.Status)
ols_plot_resid_qq(reg.fit.CD.Dist2FP.Dist2Riv.Status)
ols_plot_resid_qq(reg.fit.TH.Dist2FP.Dist2Riv.Status)

shapiro.test(reg.fit.BSA)
shapiro.test(reg.fit.data3$BSA)
shapiro.test(data3$BSA)

shapiro.test(reg.fit.TH$residuals)
shapiro.test(reg.fit.CD$residuals)
shapiro.test(reg.fit.DBH$residuals)

shapiro.test(reg.fit.DBH.Dist2Riv.$residuals)
shapiro.test(reg.fit.CD.Dist2Riv.$residuals)
shapiro.test(reg.fit.CH.Dist2Riv.$residuals)

shapiro.test(reg.fit.TH.Dist2FP$residuals)
summary(shapiro.test(reg.fit.TH.Dist2FP$residuals))
shapiro.test(reg.fit.CD.Dist2FP$residuals)
shapiro.test(reg.fit.DBH.Dist2FP$residuals)

shapiro.test(reg.fit.DBH.Dist2FP.Dist2Riv$residuals)
```

```
shapiro.test(reg.fit.CD.Dist2FP.Dist2Riv$residuals)
shapiro.test(reg.fit.TH.Dist2FP.Dist2Riv$residuals)

shapiro.test(reg.fit.TH.Dist2FP.Dist2Riv.Status$residuals)
shapiro.test(reg.fit.CD.Dist2FP.Dist2Riv.Status$residuals)
shapiro.test(reg.fit.DBH.Dist2FP.Dist2Riv.Status$residuals)

skewness(reg.fit.DBH$residuals)
skewness(reg.fit.CD$residuals)
skewness(reg.fit.CH$residuals)

skewness(reg.fit.CH.Dist2Riv.$residuals)
skewness(reg.fit.CD.Dist2Riv.$residuals)
skewness(reg.fit.DBH.Dist2Riv.$residuals)

skewness(reg.fit.DBH.Dist2FP$residuals)
skewness(reg.fit.CD.Dist2FP$residuals)
skewness(reg.fit.TH.Dist2FP$residuals)

skewness(reg.fit.TH.Dist2FP.Dist2Riv$residuals)
skewness(reg.fit.CD.Dist2FP.Dist2Riv$residuals)
skewness(reg.fit.DBH.Dist2FP.Dist2Riv$residuals)

skewness(reg.fit.DBH.Dist2FP.Dist2Riv.Status$residuals)
skewness(reg.fit.CD.Dist2FP.Dist2Riv.Status$residuals)
skewness(reg.fit.TH.Dist2FP.Dist2Riv.Status$residuals)

kurtosis(reg.fit.DBH$residuals)
kurtosis(reg.fit.CD$residuals)
kurtosis(reg.fit.TH$residuals)

kurtosis(reg.fit.TH.Dist2Riv.$residuals)
kurtosis(reg.fit.CD.Dist2Riv.$residuals)
kurtosis(reg.fit.DBH.Dist2Riv.$residuals)
```

```
kurtosis(reg.fit.DBH.Dist2FP$residuals)
kurtosis(reg.fit.CD.Dist2FP$residuals)
kurtosis(reg.fit.CH.Dist2FP$residuals)

kurtosis(reg.fit.TH.Dist2FP.Dist2Riv$residuals)
kurtosis(reg.fit.CD.Dist2FP.Dist2Riv$residuals)
kurtosis(reg.fit.DBH.Dist2FP.Dist2Riv$residuals)

kurtosis(reg.fit.DBH.Dist2FP.Dist2Riv.Status$residuals)
kurtosis(reg.fit.CD.Dist2FP.Dist2Riv.Status$residuals)
kurtosis(reg.fit.TH.Dist2FP.Dist2Riv.Status$residuals)

# SIGNIFICANCE TEST FOR SKEWNESS AND KURTOSIS
# SIGNIFICANCE TEST FOR SKEWNESS => AGOSTINO TEST
agostino.test(reg.fit.DBH$residuals)
agostino.test(reg.fit.CD$residuals)
agostino.test(reg.fit.TH$residuals)

agostino.test(reg.fit.CH.Dist2Riv.$residuals)
agostino.test(reg.fit.CD.Dist2Riv.$residuals)
agostino.test(reg.fit.DBH.Dist2Riv.$residuals)

agostino.test(reg.fit.DBH.Dist2FP$residuals)
agostino.test(reg.fit.CD.Dist2FP$residuals)
agostino.test(reg.fit.CH.Dist2FP$residuals)

agostino.test(reg.fit.TH.Dist2FP.Dist2Riv$residuals)
agostino.test(reg.fit.CD.Dist2FP.Dist2Riv$residuals)
agostino.test(reg.fit.DBH.Dist2FP.Dist2Riv$residuals)

agostino.test(reg.fit.DBH.Dist2FP.Dist2Riv.Status$residuals)
agostino.test(reg.fit.CD.Dist2FP.Dist2Riv.Status$residuals)
agostino.test(reg.fit.TH.Dist2FP.Dist2Riv.Status$residuals)
```

SIGNIFICANCE TEST FOR KURTOSIS (NEAR 3)

```
anscombe.test(reg.fit.DBH$residuals)
anscombe.test(reg.fit.CD$residuals)
anscombe.test(reg.fit.TH$residuals)

anscombe.test(reg.fit.TH.Dist2Riv.$residuals)
anscombe.test(reg.fit.CD.Dist2Riv.$residuals)
anscombe.test(reg.fit.DBH.Dist2Riv.$residuals)

anscombe.test(reg.fit.DBH.Dist2FP$residuals)
anscombe.test(reg.fit.CD.Dist2FP$residuals)
anscombe.test(reg.fit.TH.Dist2FP$residuals)

anscombe.test(reg.fit.TH.Dist2FP.Dist2Riv$residuals)
anscombe.test(reg.fit.CD.Dist2FP.Dist2Riv$residuals)
anscombe.test(reg.fit.DBH.Dist2FP.Dist2Riv$residuals)

anscombe.test(reg.fit.DBH.Dist2FP.Dist2Riv.Status$residuals)
anscombe.test(reg.fit.CD.Dist2FP.Dist2Riv.Status$residuals)
anscombe.test(reg.fit.TH.Dist2FP.Dist2Riv.Status$residuals)
```

RAINBOW TEST FOR LINEARITY

```
raintest(reg.fit.DBH)
raintest(reg.fit.TH)
raintest(reg.fit.CD)

raintest(reg.fit.CD.Dist2Riv.)
raintest(reg.fit.TH.Dist2Riv.)
raintest(reg.fit.DBH.Dist2Riv.)

raintest(reg.fit.DBH.Dist2FP)
raintest(reg.fit.TH.Dist2FP)
```

raintest(reg.fit.CD.Dist2FP)

raintest(reg.fit.CD.Dist2FP.Dist2Riv)
raintest(reg.fit.CH.Dist2FP.Dist2Riv)
raintest(reg.fit.DBH.Dist2FP.Dist2Riv)

raintest(reg.fit.DBH.Dist2FP.Dist2Riv.Status)
raintest(reg.fit.CH.Dist2FP.Dist2Riv.Status)
raintest(reg.fit.CD.Dist2FP.Dist2Riv.Status)

ABSENCE OF STRONG MULTICOLLINEARITY
ols_vif-tol(reg.fit.DBH)
ols_vif_tol(reg.fit.DBH)
ols_vif_tol(reg.fit.DBH.Dist2Riv.)
library(olsrr)
ols_plot_dfbetas(reg.fit.DBH)
ols_plot_dfbetas(reg.fit.TH)
ols_plot_dfbetas(reg.fit.CD)

ols_plot_dfbetas(reg.fit.CD.Dist2Riv.)
ols_plot_dfbetas(reg.fit.TH.Dist2Riv.)
ols_plot_dfbetas(reg.fit.DBH.Dist2Riv.)

ols_plot_dfbetas(reg.fit.DBH.Dist2FP)
ols_plot_dfbetas(reg.fit.TH.Dist2FP)
ols_plot_dfbetas(reg.fit.CD.Dist2FP)

ols_plot_dfbetas(reg.fit.CD.Dist2FP.Dist2Riv)
ols_plot_dfbetas(reg.fit.TH.Dist2FP.Dist2Riv)
ols_plot_dfbetas(reg.fit.DBH.Dist2FP.Dist2Riv)

ols_plot_dfbetas(reg.fit.DBH.Dist2FP.Dist2Riv.Status)
ols_plot_dfbetas(reg.fit.CD.Dist2FP.Dist2Riv.Status)
ols_plot_dfbetas(reg.fit.CH.Dist2FP.Dist2Riv.Status)

```
ols_plot_resid_stud(reg.fit.TH)
ols_plot_resid_stud(reg.fit.CD)
ols_plot_resid_stud(reg.fit.DBH)

ols_plot_resid_stud(reg.fit.CD.Dist2Riv.)
ols_plot_resid_stud(reg.fit.TH.Dist2Riv.)
ols_plot_resid_stud(reg.fit.DBH.Dist2Riv.)

ols_plot_resid_stud(reg.fit.DBH.Dist2FP)
ols_plot_resid_stud(reg.fit.CD.Dist2FP)
ols_plot_resid_stud(reg.fit.TH.Dist2FP)

ols_plot_resid_stud(reg.fit.TH.Dist2FP.Dist2Riv)
ols_plot_resid_stud(reg.fit.CD.Dist2FP.Dist2Riv)
ols_plot_resid_stud(reg.fit.DBH.Dist2FP.Dist2Riv)

ols_plot_resid_stud(reg.fit.DBH.Dist2FP.Dist2Riv.Status)
ols_plot_resid_stud(reg.fit.CD.Dist2FP.Dist2Riv.Status)
ols_plot_resid_stud(reg.fit.TH.Dist2FP.Dist2Riv.Status)

reg.fit <-lm(data = data3, treemodel ~ TH + CD + DBH)
reg.fit.DBH <-lm(data = data3, DBH ~ Dist2FP + Dist2Riv + Status)
reg.fit.TH <-lm(data = data3, TH ~ Dist2FP + Dist2Riv + Status)
reg.fit.CD <-lm(data = data3, CD ~ Dist2FP + Dist2Riv + Status)

data.plot.DBH <- data.frame(reg.fit.DBH, Dist2FP, Dist2Riv, Status)
reg.fit.DBH <-lm(data = data3, DBH ~ Dist2FP + Dist2Riv + Status)

summary(reg.fit.DBH)
ols_vif_tol(reg.fit.DBH)
ols_vif_tol(reg.fit.TH)
ols_vif_tol(reg.fit.CD)

ols_vif_tol(reg.fit.DBH.Dist2FP)
```

```
ols_vif_tol(reg.fit.TH.Dist2FP)
ols_vif_tol(reg.fit.CD.Dist2FP)

ols_vif_tol(reg.fit.CD.Dist2Riv.)
ols_vif_tol(reg.fit.TH.Dist2Riv.)
ols_vif_tol(reg.fit.DBH.Dist2Riv.)

ols_vif_tol(reg.fit.DBH.Dist2FP.Dist2Riv)
ols_vif_tol(reg.fit.TH.Dist2FP.Dist2Riv)
ols_vif_tol(reg.fit.CD.Dist2FP.Dist2Riv)

ols_vif_tol(reg.fit.CD.Dist2FP.Dist2Riv.Status)
ols_vif_tol(reg.fit.TH.Dist2FP.Dist2Riv.Status)
ols_vif_tol(reg.fit.DBH.Dist2FP.Dist2Riv.Status)

summary(reg.fit.DBH)
model_data_DBH <- data.frame(reg.fit.DBH ~ Dist2FP + Dist2Riv + Status)
model_data_DBH.plot <- data.frame(reg.fit.DBH ~ Dist2FP + Dist2Riv + Status)
model_data_DBH.plot <- data.frame(DBH ~ Dist2FP + Dist2Riv + Status)

attach(data3)

model_data_DBH.plot <- data.frame(DBH, Dist2FP, Dist2Riv, Status)

pairs(model_data_DBH.plot, pch = 19, lower.panel = NULL, col = Status)

ggplot(data3, aes(x = Dist2Riv, y = DBH, color = Status)) +
geom_point()+
xlab("Distance to River (m)"+
ylab("DBH (cm)")
)
ggplot(data3, aes(x = Dist2Riv, y = DBH, color = Status)) +
geom_point()
```

```
data3

scatterplot_data.plot <- data.frame(TH, CD, DBH, Dist2Riv, Dist2FP,Status)
scatterplot_data.plot

library(GGally)

ggpairs(scatterplot_data.plot[,2:6])
ggpairs(scatterplot_data.plot[,2:16])
ggpairs(scatterplot_data.plot[,1:6])

ggplot(scatterplot_data.plot, aes(x=Status, y=CD, fill = Status)) +
 geom_boxplot() +
 xlab("Tree flood status") + ylab("Canopy diameter (m)") +
 theme_classic()+scale_fill_manual(values=c("brown2"," cyan4"))+
 stat_summary(fun.y=mean,  geom="point", color="black")+
 theme(legend.position="none")+
 theme(aspect.ratio=1)

ggplot(scatterplot_data.plot, aes(x=Status, y=TH, fill = Status)) +
 geom_boxplot() +
 xlab("Tree flood status") + ylab("Tree height (m)") +
 theme_classic()+scale_fill_manual(values=c("brown2"," cyan4"))+
 stat_summary(fun.y=mean,  geom="point", color="black")+
 theme(legend.position="none")+
 theme(aspect.ratio=1)

ggplot(scatterplot_data.plot, aes(x=Status, y=DBH, fill = Status)) +
 geom_boxplot() +
 xlab("Tree flood status") + ylab("Diameter at breast height (cm)") +
 theme_classic()+scale_fill_manual(values=c("brown2"," cyan4"))+
 stat_summary(fun.y=mean,  geom="point", color="black")+
 theme(legend.position="none")+
 theme(aspect.ratio=1)
```

```
ggplot(data3, aes(x = Dist2Riv, y = CD, color = Status)) +
+ geom_point()
ggplot(data3, aes(x = Dist2Riv, y = DBH, color = Status))+
geom_point()
ggplot(data3, aes(x = Dist2Riv, y = TH, color = Status))+
geom_point()

raintest(reg.fit.DBH)
raintest(reg.fit.DBH.Dist2FP)
raintest(reg.fit.DBH)

raintest(reg.fit.DBH)
raintest(reg.fit.CD)
raintest(reg.fit.TH)
raintest(reg.fit.Dist2FP)
raintest(reg.fit.Dist2Riv.)
reg.fit.DBH
```

#Redefine the Linear Models as Follows;

LEVEL 1: dbh_status; cd_status; ch_status

```
# LEVEL 1: dbh_status; cd_status; ch_status
```

LEVEL 1a: dbh_Status; dbh_Dist2FP; dbh_Dist2Riv

```
# LEVEL 1a: dbh_Status; dbh_Dist2FP; dbh_Dist2Riv
# LEVEL 1b: cd_Status; cd_Dist2FP; cd_Dist2Riv
# LEVEL 1c: ch_Status; ch_Dist2FP; ch_Dist2Riv
# LEVEL 2a: dbh_Status_Dist2FP; dbh_Status_Dist2Riv
# LEVEL 2c: cd_Status_Dist2FP; cd_Status_Dist2Riv; cd_Dist2FP_Dist2Riv
# LEVEL 2b ch_Status_Dist2FP; ch_Status_Dist2Riv; ch_Dist2FP_Dist2Riv
# LEVEL 3c: dbh_Status_Dist2FP_Dist2Riv
# LEVEL 3b: cd_Status_Dist2FP_Dist2Riv
```

```
# LEVEL 3a: ch_Status_Dist2FP_Dist2Riv
mod_dbh_Status <- lm(data = data3, DBH ~ Status)
summary(mod_dbh_Status)
mod_cd_Status <- lm(data = data3, CD ~ Status)
summary(mod_cd_Status)
mod_ch_Status <- lm(data = data3, TH ~ Status)
summary(mod_ch_Status)

mod_dbh_Dist2FP <- lm(data = data3, DBH ~ Dist2FP)
summary(mod_dbh_Dist2FP)
mod_cd_Dist2FP <- lm(data = data3, CD ~ Dist2FP)
summary(mod_cd_Dist2FP)
mod_ch_Dist2FP <- lm(data = data3, TH ~ Dist2FP)
summary(mod_ch_Dist2FP)

mod_dbh_Dist2Riv <- lm(data = data3, DBH ~ Dist2Riv)
summary(mod_dbh_Dist2Riv)
mod_cd_Dist2Riv <- lm(data = data3, CD ~ Dist2Riv)
summary(mod_cd_Dist2Riv)
mod_ch_Dist2Riv <- lm(data = data3, TH ~ Dist2Riv)

summary(mod_ch_Dist2Riv)
mod_dbh_Status_Dist2FP <- lm(data = data3, DBH ~ Status + Dist2FP)
summary(mod_dbh_Status_Dist2FP)
mod_cd_Status_Dist2FP <- lm(data = data3, CD ~ Status + Dist2FP)
summary(mod_cd_Status_Dist2FP)
mod_ch_Status_Dist2FP <- lm(data = data3, CH ~ Status + Dist2FP)
summary(mod_ch_Status_Dist2FP)

mod_dbh_Status_Dist2Riv <- lm(data = data3, DBH ~ Status + Dist2Riv)
summary(mod_dbh_Status_Dist2Riv)
mod_cd_Status_Dist2Riv <- lm(data = data3, CD ~ Status + Dist2Riv)
summary(mod_cd_Status_Dist2Riv)
mod_ch_Status_Dist2Riv <- lm(data = data3, TH ~ Status + Dist2Riv)
```

```
summary(mod_ch_Status_Dist2Riv)

mod_dbh_Dist2FP_Dist2Riv <- lm(data = data3, DBH ~ Dist2FP + Dist2Riv)
summary(mod_dbh_Dist2FP_Dist2Riv)
mod_cd_Dist2FP_Dist2Riv <- lm(data = data3, CD ~ Dist2FP + Dist2Riv)
summary(mod_cd_Dist2FP_Dist2Riv)
mod_ch_Dist2FP_Dist2Riv <- lm(data = data3, TH ~ Dist2FP + Dist2Riv)
summary(mod_ch_Dist2FP_Dist2Riv)

mod_dbh_Status_Dist2FP_Dist2Riv <- lm(data = data3, DBH ~ Status + Dist2FP + Dist2Riv)
summary(mod_dbh_Status_Dist2FP_Dist2Riv)
mod_cd_Status_Dist2FP_Dist2Riv <- lm(data = data3, CD ~ Status + Dist2FP + Dist2Riv)
summary(mod_cd_Status_Dist2FP_Dist2Riv)
mod_th_Status_Dist2FP_Dist2Riv <- lm(data = data3, CH ~ Status + Dist2FP + Dist2Riv)
summary(mod_ch_Status_Dist2FP_Dist2Riv)

raintest(mod_dbh_Status)
raintest(mod_cd_Status)
raintest(mod_th_Status)

raintest(mod_dbh_Dist2FP)
raintest(mod_cd_Dist2FP)
raintest(mod_th_Dist2FP)

raintest(mod_dbh_Dist2Riv)
raintest(mod_cd_Dist2Riv)
raintest(mod_ch_Dist2Riv)

raintest(mod_dbh_Dist2FP_Dist2Riv)
raintest(mod_cd_Dist2FP_Dist2Riv)
raintest(mod_th_Dist2FP_Dist2Riv)

raintest(mod_dbh_Status_Dist2FP)
raintest(mod_cd_Status_Dist2FP)
```

```
raintest(mod_th_Status_Dist2FP)

raintest(mod_dbh_Status_Dist2Riv)
raintest(mod_cd_Status_Dist2Riv)
raintest(mod_ch_Status_Dist2Riv)

raintest(mod_dbh_Status_Dist2FP_Dist2Riv)
raintest(mod_cd_Status_Dist2FP_Dist2Riv)
raintest(mod_ch_Status_Dist2FP_Dist2Riv)

ols_vif_tol(mod_dbh_Status)
 ols_vif_tol(mod_cd_Status)
 ols_vif_tol(mod_th_Status)

ols_test_breusch_pagan(mod_dbh_Status)
summary(ols_test_breusch_pagan(mod_dbh_Status))

ols_test_breusch_pagan(mod_dbh_Dist2FP)
ols_test_breusch_pagan(mod_dbh_Dist2Riv)
ols_test_breusch_pagan(mod_dbh_Status)

ols_plot_resid_lev(mod_dbh_Status)
ols_plot_cooksd_chart(mod_dbh_Status)

coeftest(mod_dbh_Status, vcov = vcovHC(mod_dbh_Status, type="HC4"))
ols_plot_resid_hist(mod_dbh_Status)
ols_plot_resid_qq(mod_dbh_Status)

shapiro.test(mod_dbh_Status$residuals)

skewness(mod_dbh_Status$residuals)
skewness(mod_dbh_Status_Dist2FP$residuals)
skewness(mod_cd_Status_Dist2FP$residuals)
skewness(mod_th_Status_Dist2FP$residuals)
```

```
skewness(mod_th_Status_Dist2Riv$residuals)
skewness(mod_cd_Status_Dist2Riv$residuals)
skewness(mod_dbh_Status_Dist2Riv$residuals)
1+1
ols_plot_resid_hist(mod_dbh_Status)

summary(data3)

attach(data3)
boxplot(data3$TH)
boxplot(data3$CD)
boxplot(data3$DBH)
boxplot(data3$CD, data3$DBH)
boxplot(data3$CD, data3$DBH, data3$TH)
boxplot(data3$CD, data3$DBH, data3$TH, col = "red")

model_DBH<- lm(data=data3, DBH~Dist2Riv)
summary(model_DBH)

model_CD <- lm(data=data3, CD~Dist2Riv)
summary(model_CD)

names(data3)
boxplot(data3$CD, data3$DBH, data3$CH, col = "blue")

library(ggplot2)
plot(data3$DBH~data3$Dist2Riv)

plot(data3$DBH~data3$Dist2Riv, col = Status, pch = 10)
model_DBH_Dist2Riv <- lm(data3$DBH~data3$Dist2Riv, col = Status, pch = 10)
model_DBH_Dist2Riv <- lm(data3$DBH~data3$Dist2Riv)
smmary(model_DBH_Dist2Riv
)
```

```
summary(model_DBH_Dist2Riv)
model_DBH_flooded_Dist2Riv <- lm(DBH--"Fooded"~Dist2Riv)
model_DBH_flooded_Dist2Riv <- lm(DBH--"Flooded"~Dist2Riv)
model_DBH_flooded_Dist2Riv <- lm(DBH=="Flooded"~Dist2Riv)
summary(model_DBH_flooded_Dist2Riv)
model_DBH_flooded_Dist2Riv <- lm(DBH==Status$Flooded~Dist2Riv)
model_DBH_flooded_Dist2Riv <- lm(DBH~Dist2Riv + Status == "Flooded")
summary(model_DBH_flooded_Dist2Riv)
model_DBH_flooded_Dist2Riv <- glm(DBH ~ Dist2Riv + Status=="Flooded")
summary(model_DBH_flooded_Dist2Riv)
model_DBH_notflooded_Dist2Riv <- glm(DBH ~ Dist2Riv + Status=="Not Flooded")
summary(model_DBH_notflooded_Dist2Riv)
summary(data3)
summary(data3$Status=="Flooded")
summary(Status=="Flooded")

model_DBH_Dist2Riv <- lm(data3$DBH~data3$Dist2Riv, col = Status, xlab = "Distance to
      River (m)", ylab = "Diameter at Breast Height (m)")
model_DBH_Dist2Riv <- lm(data3$DBH~data3$Dist2Riv, col = Status)
model_DBH_Dist2Riv <- lm(data3$DBH~data3$Dist2Riv)
plot(model_DBH_Dist2Riv)
```

Pairwise scatteplot for D2R, D2FP and and statistical derivatives (Min, Mean, Max and stdev)

```
scatterplotMatrix(data=[c(20, 23, 21)], smooth = TRUE, id = FALSE, legend = TRUE, regLine
= TRUE, ellipse = FALSE, var.labels = colnames(data=[c(20, 23, 21)]), diagonal = TRUE,
plot.points = TRUE, groups = Class, by.groups = TRUE, use = c("complete.obs",
"pairwise.complete.obs"), col = carPalette()[-1], pch = 1:n.groups, cex = par("cex"), cex.axis =
par("cex.axis"), cex.labels = NULL, cex.main = par("cex.main"), row1attop = TRUE)
```

Pairwise goup comparisons between flooded and non-flooded tree characteristics

```
ggbetweenstats(
  data,
  x = Status,
  y = DBH,
```

```
  type = "nonparametric",
  pairwise.display = "significant",
  p.adjust.method = "holm",
  effsize.type = "unbiased",
  bf.prior = 0.707,
  bf.message = TRUE,
  results.subtitle = TRUE,
  xlab = NULL,
  ylab = NULL,
  caption = NULL,
  title = NULL,
  subtitle = NULL,
  digits = 2L,
  var.equal = FALSE,
  conf.level = 0.95,
  nboot = 100L,
  tr = 0.2,
  centrality.plotting = TRUE,
  centrality.type = type,
  centrality.point.args = list(size = 5, color = "darkred"),
  centrality.label.args = list(size = 3, nudge_x = 0.4, segment.linetype = 4,
    min.segment.length = 0),
  point.args = list(position = ggplot2::position_jitterdodge(dodge.width = 0.6), alpha =
    0.4, size = 3, stroke = 0, na.rm = TRUE),
  boxplot.args = list(width = 0.3, alpha = 0.2, na.rm = TRUE),
  violin.args = list(width = 0.5, alpha = 0.2, na.rm = TRUE),
  ggsignif.args = list(textsize = 3, tip_length = 0.01, na.rm = TRUE),
  ggtheme = ggstatsplot::theme_ggstatsplot(),
  package = "RColorBrewer",
  palette = "Dark2",
  ggplot.component = NULL,
  ...
)
```