AN INTEGRATED CONCEPTUAL MODEL TOWARDS SUSTAINABLE RURAL WATER MANAGEMENT BASED REMOTE SENSING AND MACHINE LEARNING

by Thu Thuy Nguyen

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Prof. Huu Hao Ngo, Prof. Wenshan Guo, and Dr. Yiwen Liu

University of Technology Sydney

Faculty of Engineering and Information Technology

May 2022

CERTIFICATION OF ORIGINAL AUTHORSHIP

I, Thu Thuy Nguyen declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Civil and Environmental Engineering/Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature: Signature removed prior to publication.

Date: 10 May 2022

ACKNOWLEDGEMENTS

First of all, it is a pleasure to express my deepest gratitude to my supervisors, Prof. Huu Hao Ngo, Prof. Wenshan Guo, and Dr. Yiwen Liu. This Ph.D. thesis would not have been achievable without their continuous support and guidance. Especially it is my privilege to express a deep sense of gratitude to my principal supervisor, Prof. Huu Hao Ngo, for his unending patience, endless support, and brilliant ideas during my Ph.D. study and research. His DIPP philosophy: the spirit of Dedication, Inspiration, Passion, and Persistence has made me inspired. I humbly appreciate him for what he has provided me and for the time he has spent to help me revise my manuscripts and thesis to complete this study. I am highly grateful to him for his steadfast encouragement, unwavering guidance, and great sense of humor so that I can overcome challenges during my Ph.D. life. I would like to extend my heartfelt thanks to my co-supervisor, Prof. Wenshan Guo, and Dr. Yiwen Liu for their valuable comments and suggestions for completing my Ph.D. study.

I am deeply indebted to the Vietnam International Education Development (VIED) and the University of Technology Sydney (UTS) for providing me VIED-UTS Scholarship for my Ph.D. program. I am thankful to the Australian Postgraduate Research Intern (APR. Intern), Astron Environmental Services Company, and Food Agility CRC for data-sharing and supporting my Ph.D. research.

My special thanks to Dr. Tien Dat Pham, Dr. Kinh Bac Dang, and Dr. Chinh Luu for their data sharing and valuable suggestions. I am thankful to administrative staff and academic officers from the Graduate Research School of UTS, and the Institute for Agricultural Environment, especially Ms. Van Le from the School of Civil and Environmental Engineering, for their unparalleled support. Thanks should also go to my excellent schoolmates, including Loan, Hang, Dora, Khan, Jerry, and Phong, who shared their experience and knowledge. I appreciate my friends, An Le, Chelsey Vu, Allie Nguyen, Quyen Nguyen, Thuy Nguyen, and Minh Vu, who have willingly provided me with assistance during my Ph.D. life.

Finally, the most tremendous and sincere gratitude must be sent to my parents and brother, who have been my constant source of love and strength all my Ph.D. years.

Especially my most profound appreciation and respect should go to my grandma for her encouragement, kindness, and unconditional love. I am also grateful to my uncles, ant, and relatives for their support and encouragement during my study. Last but not least, I would love to thank my husband and my son, who is always with me, support me, and give me love and encouragement. They are my love, strength, and motivation that make me overcome any difficulties in my life. Without their unparalleled love, constant support, and care, I could not pursue and completed my Ph.D. study successfully. This dissertation is dedicated to my extended family, husband, and little boy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AC	Adaptive capacity
AHP	Analytical hierarchy process
AI	Aridity index
ALOS	Advanced land observing satellite
ANN	Artificial neural networks
ASM	Advanced scatter meter
BI	Brightness index
BI2	Brightness index 2
BMPs	Best management practices
BMPs	Best management practices
BRT	Boosted regression trees
CBR	CatBoost gradient boosting regression
CCFSC	Central committee for flood and storm control
CH	Central Highland
CHIRPS	Climate hazards group infrared precipitation with station data
CI	Colour index
CV	Cross validation
DEM	Digital elevation model
DGPS	Differential global positioning system
DPSIR	Driver pressure state impact response
DSM	Digital surface model
E	Exposure
ELM	Extreme learning machine
EO	Earth observation
EPA	Environmental protection agency

ESA	Ecosystem services assessment
FullCAM	Full Carbon Accounting Model
GA	Genetic algorithm
GA	Grass pavement
GBI	Global international geosphere-biosphere programme
GIS	Geographic information system
GLMC	Grey level co-occurrence
GNDVI	Green normalized difference vegetation index
GSD	Ground sampling distance
HDI	Human development index
ICBMs	Integrated component-based models
IPCC	Inter-government panel on climate change
IRECl	Inverted red-edge chlorophyll index
IUDMs	Integrated urban drainage models
IUWCMs	Integrated urban water cycle models
IUWM	Integrated urban water management system
IUWSMs	Integrated urban water system models
KC	Kappa coefficient
LAI	Leaf area index
LCA	Life cycle assessment
LID	Low impact development
MCA	Multi-criteria analysis

MCARI	Modified chlorophyll absorption in reflectance index
MD	Mekong Delta
ML	Machine learning
MODIS	Moderate resolution imaging spectroradiometer
NCC	North Central Coast
NDI45	Normalized difference index using bands 4 & 5 of S-2
NDVI	Normalized difference vegetation index
NDWI	Normalized difference water index
NE	Northeast
NPP	Net primary production
NW	Northwest
OA	Overall accuracy
OA	Overall accuracy
P	Precision
PA	Permeable asphalts
PAC	Priestley–taylor alpha coefficient
PC	Permeable concretes
PICP	Permeable interlocking concrete pavers
PLSR	Partial least squares regression
PPPs	Private public partnerships
R	Recall
R^2	Coefficient of determination
RF	Random forest
RFR	Random forest regression
RI	Redness index

RMSE	Root mean square error
RRD	Red River Delta
RS	Remote sensing
RUE	Rain use efficiency
RVI	Ratio vegetation index
RWH	Rainwater harvesting
RWH	Rainwater harvesting
S	Sensitivity
SAR	Synthetic aperture radar
SAVI	Soil adjusted vegetation index
SCC	South Central Coast
Sentinel 1	S-1
Sentinel 2	S-2
SIs	Soil indices
SM	Soil moisture
SMAP	Soil moisture active passive
SMOS	Soil moisture and ocean salinity
SOC	Soil organic carbon
SuDS	Sustainable drainage system
SVM	Support vector machine
SW	South West
SWMM	Storm water management model
SWS	Soil water stress
TWI	Topographic wetness index
UASs	Unmanned aerial systems
UTM	Universal transverse Mercator
VIs	Vegetation indices
WA	Western Australia
WSC	Water sensitive city

WSE	Wrapper subset evaluator
WSUD	Water sensitive urban design
WUE	Water use efficiency
WVI	Water vulnerability index
XGBoost	Extreme gradient boosting
XGBR	Extreme gradient boosting regression

RESEARCH OUTCOMES

A. Peer-Reviewed Journal Articles

1. **Nguyen, T.T**., Ngo, H.H., Guo, W., Nguyen, H.Q., Luu, C., Dang, K.B., Liu, Y., Zhang, X. 2020a. New approach of water quantity vulnerability assessment using satellite images and GIS-based model: An application to a case study in Vietnam. *Science of The Total Environment*, **737**, 139784 (IF: **7.963**; SJR: **Q1**).
2. **Nguyen, T.T**., Ngo, H.H., Guo, W., Wang, X.C. 2020b. A new model framework for sponge city implementation: Emerging challenges and future developments. *Journal of Environmental Management*, **253**, 109689 (IF: **6.789**; SJR: **Q1**).
3. **Nguyen, T.T**., Ngo, H.H., Guo, W., Wang, X.C., Ren, N., Li, G., Ding, J., Liang, H. 2019. Implementation of a specific urban water management - Sponge City. *Science of The Total Environment*, **652**, 147-162 (IF: **7.963**; SJR: **Q1**).
4. **Nguyen, T.T**., Pham, T.D., Nguyen, C.T., Delfos, J., Archibald, R., Dang, K.B., Hoang, N.B., Guo, W., Ngo, H.H. 2022. A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion. *Science of The Total Environment*, **804**, 150187 (IF: **7.963**; SJR: **Q1**).
5. Dang, K.B., **Nguyen, T.T.**, Ngo, H.H., Burkhard, B., Müller, F., Dang, V.B., Nguyen, H., Ngo, V.L., Pham, T.P.N. 2021. Integrated methods and scenarios for assessment of sand dunes ecosystem services. Journal of Environmental Management, 289, 112485 (IF: **6.789**; SJR: **Q1**).
6. **Nguyen, T.T.**, Ngo, H.H., Guo, W., Chang, S.W., Nguyen, D.D., Nguyen, C.T., Zhang, J., Liang, S., Bui, X.T., Hoang, N.B. 2022. A low-cost approach for soil moisture prediction using multi-sensor data and machine learning algorithm. *Science of The Total Environment*, 833, 155066 (IF: **7.963**; SJR: **Q1**).
7. **Nguyen, T.T**. 2021. Predicting agricultural soil carbon using machine learning. *Nature Reviews Earth & Environment*, **2**(12), 825-825.
8. **Nguyen, T.T**., Ngo, H.H., Guo, W. Fusion of feature selection optimizer and advance machine learning algorithm for improvement of soil carbon prediction. (in preparation).

B. Book chapters

1. Huu Hao Ngo, **Thu Thuy Nguyen**, Wenshan Guo, Dinh Duc Nguyen, Ashok Pandey, Xuan Thanh Bui, Sunita Varjani, Phuoc Dan Nguyen, Chapter 11: Circular bioeconomy for resource recovery from wastewaters using algae-based technologies, In the book series on Current Developments in Biotechnology and Bioengineering: Algae-based biomaterial for sustainable development: biomedical, environmental remediation and sustainability assessment, Huu Hao Ngo, Wenshan Guo, Ashok Pandey, Jo-Shu Chang, Duu-Jong Lee (Eds), Elservier, (In press).
2. Huu Hao Ngo, **Thu Thuy Nguyen**, Wenshan Guo, Lijuan Deng, Sunita Varjani, Yi Liu, Chapter 16: Sustainability assessment of biochar for climate change mitigation, In the book series on Current Developments in Biotechnology and Bioengineering: Biochar towards sustainable environment, Huu Hao Ngo, Wenshan Guo, Ashok Pandey, Sunita Varjani, Daniel CW Tsang (Eds), Elservier, (In press).

C. Conference paper

1. Thu Thuy Nguyen, Huu Hao Ngo, Wenshan Guo, Xiaochang C. Wang. 2019. Developing a Conceptual Water Model for Sponge City. Green Technologies for Sustainable Water (GTSW). Ho Chi Minh City, Vietnam, 1 - 5, December.

D. Research awards

1. Higher Degree Top-Up Scholarship from Food Agility CRC, 2021 for the project: An integrated model for sustainable rural water management based on machine learning and remote sensing, 2021-2022, $10,000 per year.
2. Australian Postgraduate Research Intern offer by Australia Government and Astron Environmental Consulting Company, 2021 for carrying out an internship project: Investigating the use of satellite data for soil carbon monitoring within agricultural areas of Australia, 2021, $9000/ 3months.
3. UTS Thesis Equity Grant, 2021
4. Higher Degree Research Female Top-up Scholarship from the University of Technology, Sydney, 2019 for the outstanding achievement of female students, 2019,
5. PROM Program- International scholarship exchange of PhD students and

academics, 2019" from Poland Government, 2019.

6. FEIT-UTS International Conference Grant Approval, 2019

Ph.D. DISSERTATION ABSTRACT

Author: Thu Thuy Nguyen

Date: 23 February 2022

Thesis title: An integrated conceptual model towards sustainable rural water management based remote sensing and machine learning

Faculty: Faculty of Environmental and Information Technology

School: Civil and Environmental Engineering

Supervisors: Prof. Huu Hao Ngo (Principal supervisor)
Prof. Wenshan Guo (Co-supervisor)

Dr. Yiwen Liu (Co-supervisor)

Abstract

Recently, there have been some improvements in agricultural water supply systems. However, rural areas still face serious water deficiencies including droughts, poor water quality and floods due to inappropriate of water management systems and climate change. This critical issue points out the urgent need for developing an effective integrated rural water model, which can improve water monitoring in rural regions. This study therefore aims to develop the integrated conceptual model for rural sustainable water monitoring to help rural communities overcome the issues of water run-off, water pollution and lack of water for agricultural production.

This thesis presents a novel conceptual model framework including three sub-models (water vulnerability quantity assessment model, soil moisture prediction model, and agricultural soil organic carbon model for supporting rural water modelling using the integration of free-of-charge satellite images including MODIS, Sentinel 1, Sentinel 2, and ALOS DSM imagery and different advanced machine learning algorithms. The framework firstly demonstrates a new approach of water quantity vulnerability assessment based on reliable and updated spatial-temporal datasets (soil water stress, aridity index, rain use efficiency and leaf area index), and the incorporation of the GIS-based model. Notably, this research devises a state-of-the-art machine-learning model for monitoring agricultural drought via predicting soil moisture (SM) using active and ensemble-based decision tree learning combined with multi-sensor data fusion at a national and world scale. This work explores the use of Sentinel-1, Sentinel-2, and an innovative machine learning (ML) approach using an integration of active learning for land-use mapping and advanced Extreme Gradient Boosting Regression (XGBR) for robustness of the SM estimates. The collected soil samples from a field survey in Western Australia were also used for the model validation and indicators including the coefficient of determination (R^2) and root - mean – square - error (RMSE) were applied to evaluate the model's performance. The proposed model XGBR with 21 optimal features obtained from GA was yielded the highest performance (R^2 = 0.891, RMSE = 0.875%). A combination of S1 and S2 sensors could also effectively estimate SOC in farming areas by using ML techniques. Satisfactory accuracy of the proposed XGBoost with optimal features was achieved the highest performance (R^2 = 0.870; RMSE = 1.818 tonC/ha) which outperformed random forest and support vector machine

Conclusively, the described conceptual model can further support precision agriculture, water management and drought resilience programs via water use efficiency, green infrastructure and smart irrigation management for agricultural production.

Keywords: Water vulnerability, spatial datasets, correlation analysis, GIS-based model, machine learning, soil moisture, soil organic carbon, feature selection, genetic algorithm

Chapter 1

Introduction

1.1 Research background

Concerns about water resources sustainability have increased worldwide due to unplanned and poorly managed population growth and environmental problems (Carle et al., 2005; Lee Joong and Heaney James, 2003). Studies on the complexity of water systems and new kinds of sustainable water management concepts are becoming prolific in hydrological scientific research (Salvadore et al., 2015). Today's conventional water management systems, where all components are constructed independently, do not possess the abilities for functioning effectively especially in terms of population growth and climate change requirements (Butler and Schutze, 2005; Rauch et al., 2005). Examples of a diversified approach to achieve an integrated water management system (IUWM) include Best Management Practices (BMPs) in the United States, Water Sensitive Design in Australia, Sustainable Drainage System (SuDS) in the United Kingdom, and Sponge City in China. The objectives of these systems are to: (1) pay good attention to all components of the system so that they work well; (2) implement water systems in both centralized and decentralized contexts; and (3) create multiple ecologically friendly services in urban zones including: water resources conservation, flooding disaster mitigation, relevant amenities, and micro-climate improvements (Bach et al., 2014; Brown et al., 2009; Nguyen et al., 2018).

Integrated water models have been devised and their focus is on interactions amongst all components of water systems management. These transitions to an integrated water model specifically concentrate on the interactions between water systems, which should be the priority of development and societal factors (Rauch et al., 2017). As early as the 1970s, research in integrated water systems was undertaken in Glatt Valley, Switzerland (Gujer et al., 1982), however this research did not document any modelling results. At the first INTERURBA conference in 1993, emerging research on integrated water models was initially reported that marked a milestone in the progress of such integrated models (Lijklema et al., 1993).

Integrated water models are essential tools for planning and management of drainage systems. In 1971, the US Environmental Protection Agency (EPA) developed the Storm Water Management model (SWMM) which is the one of most popular tools for the evaluation of stormwater management systems (Deng et al., 2018). A range of commercial stormwater models such as Mike Urban, InfoWorks and DAnCE4Water, which were built based on SWMM, is commonly used worldwide. Although the development of models has brought benefits for planners and policy-makers, these models encounter many challenges because urban water systems are, in fact, very complex. Moreover, the lack of understanding of interactions between all components, that is, understanding the whole system, and the expense of data requirements, and limitations in computational hardware have affected the model's performance (Candela et al., 2011; Rauch et al., 2005; Vanrolleghem et al., 2005). Having an insufficient understanding of model uncertainties also contributes to the model being at risk of failure (Dotto et al., 2011). However, with the recent advances being made in software package capabilities and technologies, these models have performed better in recent years. Integrated models gained momentum by combining and improving conventional single model packages in the past few decades (Bach et al., 2014).

Water infrastructure implementation promises many benefits for our society in general and urban and rural areas across the world (Jia et al., 2017; Mei et al., 2018; Zhang and Chui, 2019; Zhang et al., 2018). The water infrastructure implementation process consists of four phases (Figure 1.1). Phase 1 is analyzing regional context including water issues and existing water management to identify the demand for the implementation. Next phase is developing scenarios based on climate change scenarios, population growth scenarios, and water demand scenarios. Phase 3 indicates the selection and development of water modelling to simulate water measurements performance. The final phase is the planning and implementation of water practices.

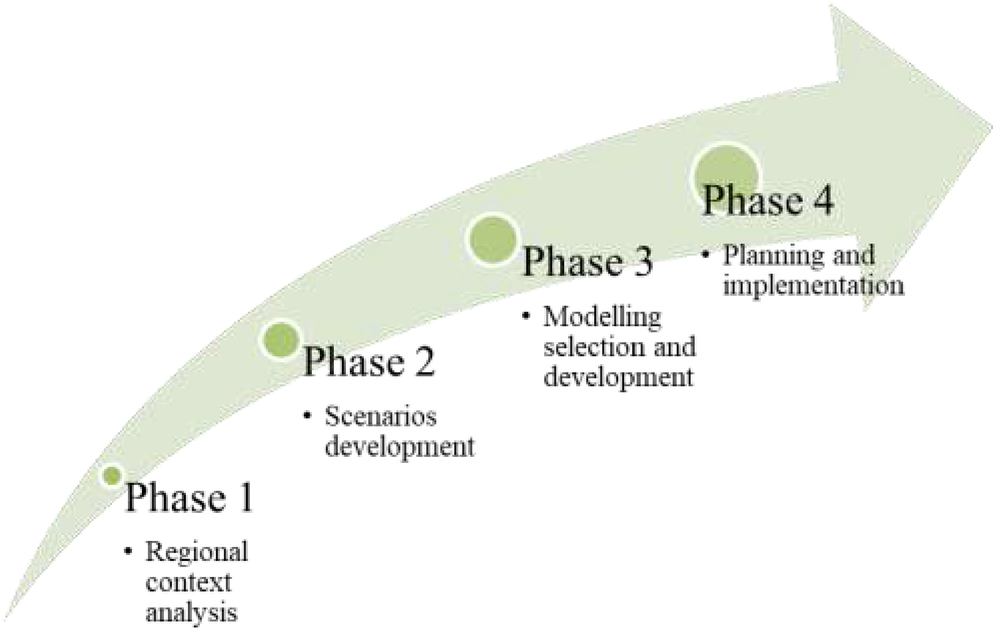


Figure 1.1 Summary of water infrastructure implementation process

To obtain the promising benefits of water infrastructure, planning and development of water measurements is important. However, it is difficult work as water systems are indeed highly complex combined with uncertain futures, as having a variety of aspects to be scrutinized, including water infrastructure planning and measurements feasibility evaluations to be conducted. An interdisciplinary approach to be developed for integrated models should be based on water infrastructure concepts in order to deal with interdisciplinary planning problems is necessary. In addition to this, some models were applied for assessing the water infrastructure performance. Storm Water Management Model (SWMM) and the Analytical Hierarchy Process (AHP) method, for example, served to quantify the benefits of LID practices (Li et al., 2019). An energy analysis and GIS model were combined for application to selected pivotal areas for water infrastructure construction (Zhao et al., 2018). In addition, spatial data like Landsat-8TIRS was used to evaluate the effects of LID practices on thermal landscapes (Hou et al., 2019). An integrated model named Uwater was innovated to support Sponge City development recently (Deng et al., 2018). The development of Uwater is based on the integration of SWMM and spatial data management tools such as in GIS. This model can evaluate drainage capacity of the stormwater system and design water infrastructure.

Sponge City's construction in China is unique in many respects when compared to other concepts (e.g., SuDS, BPMs) since the city not only addresses storm water, but also tackles flooding disasters, water restoration and water purification. Nevertheless, the simulation and evaluation tools to predict the comprehensive Sponge City's performance are still limited. This study focuses on establishing a new integrated conceptual model towards sustainable rural water management based on the Sponge City concept for not only water improvement factors, but also other environmental associated elements are taken into account. A new integrated conceptual model should be able to integrate the sub-models and include the following: (1) water vulnerability assessment model to identify suitable areas for water infrastructures construction; (2) soil moisture (SM) content prediction model to evaluate the performance water infrastructures on flood control; drought resistance; and precision irrigation decisions; (3) and agricultural soil organic carbon (SOC) monitoring model which serves to assess the effectiveness of different water strategies on soil carbon improvement to generate net zero emissions for the agriculture sector.

1.2 Research scope and objectives

Although a variety of rural water models have been proposed and applied in many studies, there are several challenges associated with these models: (1) model complexity, (2) the unavailability and inaccurately prediction variables, (3) limitations in applying free-of-charge spatial and temporal data from satellite images, (4) the unfeasibility in practical applications of the models, and (5) model cost-ineffectiveness. Thus, the general aim of this study is to implement an integrated model for rural sustainable water management using remote sensing data and machine learning algorithms. This research mainly focuses on applying free-of-charge and reliable remote sensing datasets and advanced machine learning to support rural water assessment and monitoring with high spatial resolution.

The specific objectives of this study comprise as follows:

1) To identify the promising conceptual model for supporting rural sustainable water management;
2) To develop a water vulnerability framework for rural water management;
3) To explore next-generation spatial modelling of soil moisture for flood, drought monitoring, and water management strategies; and
4) To determine the approach for agricultural soil organic carbon prediction for different water regimes and farming practices

1.3 Research significance

This study is the first attempt to implement a new assessment framework that considers the contribution of satellite datasets like Terra MODIS and the utilization of a GIS-based model for the water vulnerability assessment. Moreover, the present study pioneers the use of predictor features (dual polarization and transformed bands) from SAR remote sensing imagery (S-1), predictor variables derived from optical remote sensing imagery (S-2), and ALOS DSM derived indications with active and ensemble machine learning techniques for estimating soil moisture and soil carbon in rural areas). This thesis presents novel approaches that contribute significantly to various water vulnerability assessment, agricultural SM, and SOC retrieval studies globally. The innovative methods described in the research allow rapid and reliable estimation of the spatial variability of SM and SOC. More importantly, this described model framework can further support decision making on precision agriculture and drought resilience programs via water use efficiency and smart irrigation management for crop production which makes possible carbon neutrality for agriculture towards additional revenue via carbon credits.

1.4 Thesis structure

The thesis is structured with eight chapters, which are illustrated in Figure 1.2.

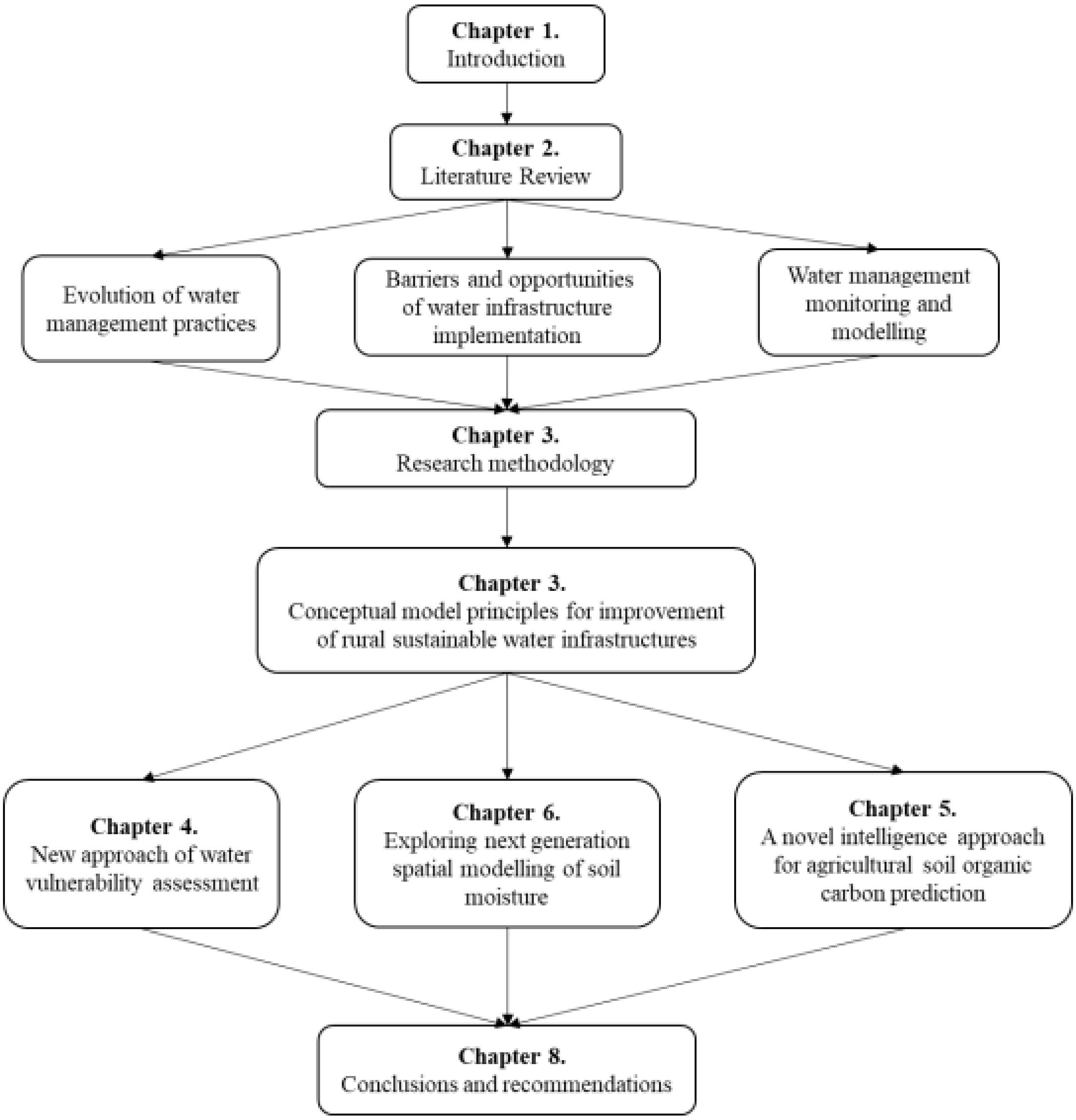


Figure 1.2 Thesis structure

Chapter 1 outlines a brief overview of existing water issues, water management practices, and their challenges. The research motivations, scope and significance are also mentioned afterwards.

Chapter 2 presents the evolution of water management strategies, the principle of the latest water management framework and their future perspectives. It also demonstrates a critical review of conventional water models for water monitoring. The advantages and disadvantages of each model to support water management strategies are illustrated in detail.

Chapter 3 introduces the adaptive research methodologies in the research including materials and methods. The used materials for the present thesis include various satellite imagery and field survey data. The methods include field data collection, spatial data processing, vulnerability assessment, machine learning analysis, feature selection algorithm and model performance evaluation.

Chapter 4 demonstrates of conceptual model principles for improvement of rural sustainable water infrastructures.

Chapter 5 illustrates the new approach for water quantity vulnerability assessment using various remote sensing datasets and GIS-based model.

Chapter 6 explores next generation spatial modelling of soil moisture based multi-sensor data fusion and machine learning approach, which provides valuable data for different stakeholders like water managers, local authorities, and landholders to practice precision agriculture.

Chapter 7 indicates a novel framework using free-of-charge multi-sensor Sentinel 2 and Sentinel 1 with state-of-the-art extreme gradient boosting to predict agricultural SOC stocks for different water and farming practices.

Chapter 8 summarizes the key findings, statements and conclusions from this study and provides recommendations for future research.

Chapter 2

Literature Review

2.1 Introduction

Water management is a vital factor in the sustainable development of any areas (Schaffer and Vollmer, 2010). Water-related problems have raised concerns worldwide among the scientific community (Marlow et al., 2013). Due to rapid population growth and the extreme weather phenomenon, water issues now include more floods, over-exploitation of groundwater, water shortages, the wasting of rainwater resources, and water pollution (Jia et al., 2015; Marlow et al., 2013). For instance, the over-use of grey construction such as concrete and asphalt in regional development has created impermeable surfaces that are not able to absorb water. This leads to floods. The construction of buildings has been about accommodating the rapidly increasing populations. This has resulted in the removal of natural rainwater- detaining infrastructure including woodlands, green spaces, natural lakes and wetlands for rainwater recycling processes. For instance, storm water has been discharged as wastewater rather than being absorbed into the soil that should has been added to groundwater reserves for water conservation, or reused as water resources for sustaining people’s lives and agricultural production. Another main reason for water flooding hazards is maladaptive drainage systems in rural regions. The inappropriate management of stormwater is not only detrimental to human health, but the aquatic ecosystems as well. The marked change in the last few decades has focused mainly on reducing floods, but now a number of targets need to be meet for water quality improvement and rainwater recycling. From an rural water perspective, many scientists illustrated that the current model of centralised drainages are inappropriate to constraints associated with climate change, the remarkable increase of population and social circumstances (Brown et al., 2009; Pahl-Wostl, 2007). Water flooding and pollution has received much attention from developed countries since the 1970s (Fletcher et al., 2014). A number of solutions are suggested to address water issues for rural areas including green infrastructures (Liu & Jensen, 2018). Scientists and policy-makers have proposed several concepts and theories for water planning. These include best management practices (BMPs), which were introduced in the United State in the 1970s (Fletcher et al., 2014; Scholz, 2006). At the same time, in the United Kingdom, sustainable urban drainage systems (SUDS) were issued with the purpose of addressing water pollution and flood hazards (Fletcher et al., 2014). Since the 1990s, the low impact development (LID) strategy was accepted not only in the United States, but in New Zealand as well (Chui et al., 2016; Fletcher et al., 2014; Mao et al., 2017). In Australia, experiencing six development stages of water management, water sensitive city (WSC) was initiated in 21st century to bring a range of benefits which not only protect the degradation of water resources, but also manage and recycle stormwater to make cities become sustainable, liveable and resilient (Brown et al., 2009; Ashley et al., 2013). However, these concepts and strategies are still being developed for industrial countries (Chan et al., 2018) and they have applied in small-scales like experimental pilots and localized areas. The developing countries like China, most areas have witnessed a high density of population growth, intensive expansion of impermeable roads and rooftops and pressures of water flood disasters in terms of climate change. As such, the new approach of water management is essential for developing countries. The Sponge City program was launched in 2013-14 to address and overcome the above-mentioned issues (Li et al., 2017) by delivering multiple advantages for communities associated with water run-off reduction, water quality enahancement, water storage increase and greenhouse gases (GHGs) emission mitigation (Wang et al., 2018). Many pilot Sponge City programs commenced with the Chinese government stating that approximately 70% of stormwater would be recyled from implementing measures to improve permeation, detention, storage, purification and drainage systems (Li et al., 2017). This literature review chapter aims to illustrate clearly some aspects: (1) understanding the evolution of water management strategies in the world in general and the drivers for water infrastructures implementation, (2) clarifying the barriers and uncertainties of water infrastructure construction, (3) highlighting water management modelling and the roles of remote sensing and machine learning on water ecosystem services' monitoring.

Major parts of this chapter were published in a peer-reviewed journal (A-rated journal):

Nguyen, T.T., Ngo, H.H., Guo, W., Wang, X.C., Ren, N., Li, G., Ding, J., Liang, H. 2019. Implementation of a specific urban water management - Sponge City. Science of the Total Environment, 652, 147-162 (IF: 7.963; SJR: Q1).

Nguyen, T.T., Ngo, H.H., Guo, W., Wang, X.C. 2020b. A new model framework for sponge city implementation: Emerging challenges and future developments. Journal of Environmental Management, 253, 109689 (IF: 6.789; SJR: Q1).

Water management practices predominantly aim to prevent flood disasters. There are three main factors causing flooding - climate change, population growth and inappropriate water planning cooperation with the region's management system (see in Figure 2.1) (Jiang et al., 2018). According to Jiang et al. (2018), one of the leading factors causing floods is climate change, which leads to heavy rainfall events with high intensity within a short time resulting in flooding situations. In addition to this, climate change means frequent precipitation extremes that increase flood hazards. Another reason why flood disasters are occurring is due to population growth (Wang et al., 2018; Wang et al., 2016; Wang et al., 2017b; Xia et al., 2017b). Consequently, during the increase of population, natural lands were converted to residential or commercial purposes, resulting in more impervious surfaces in rural areas that creates water runoff, thus leading to increased flooding occurrences. As an example, the priority for economic development is to erect skyscrapers and buildings, which have actually replaced and/or threatened aquatic ecosystems such as lakes and wetland areas, and in turn affecting the water resources balance. Therefore, large city areas do not have the capacity of absorption, purification and filtration of rainwater leading to flooding disasters or their re-occurrence (Dong et al., 2018; Li et al., 2018; Wang et al., 2018). Unsuitable planning strategies including poor and insufficient drainage systems and rural development on the floodplains have somewhat created serious problems for the environment and simply increase the threat of flooding (Li et al., 2017; Li et al., 2016; Liu, 2016; Liu et al., 2017). For instance, traditional drainage systems are inconsistent with rapid population growth and climate change, their facilities are outdated and their design standards for stormwater management have failed to adapt with the growth of population (Jiang et al., 2018; Nkwunonwo et al., 2016). The construction and development of buildings and industrial areas on the floodplains on one hand exacerbates the risk of flooding (Nkwunonwo et al., 2016). For these reasons, it is necessary to construct an effective and conventional rainwater management system to be consistent with development plans and climate extreme events as well.

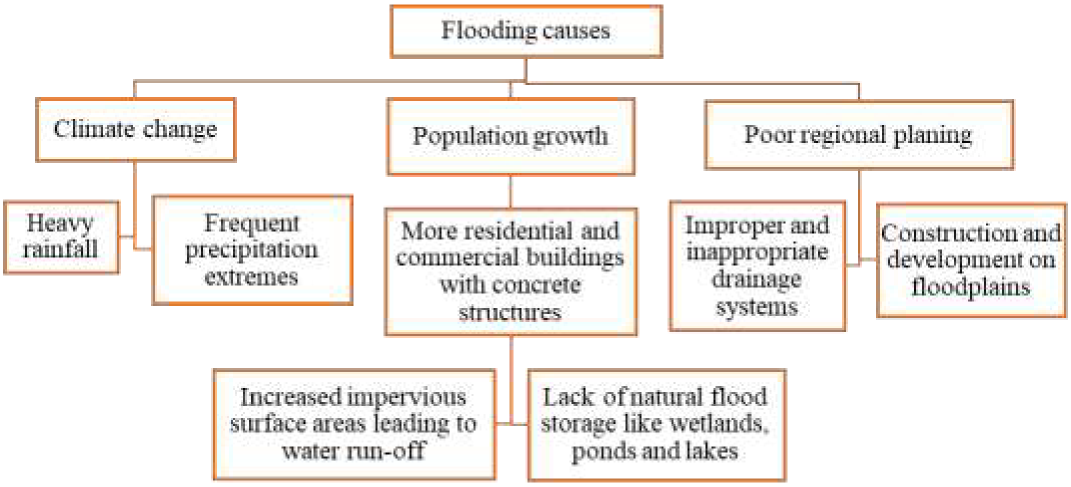


Figure 2.1. The main causes of flooding

(Modified from Nkwunonwo et al., 2016.)

2.2.2 Evolution of water management practices

The traditional approach of stormwater management appeared as early as 3000 BC (Burian and Edwards, 2002) with the primary purposes of avoiding flooding and collecting rainwater. Drainage development system has evolved through four main periods of human history, these being: the ancient world; the Roman Empire; the post-Roman era to the 19th century; and lastly, from the 19th century to now which are synthesised in Table 2.1. Moreover, according David (2014), water technologies were divided into 4 development periods (designated Water **1.0** to **4.0**) from 2500BC to the present day. In Water **1.0,** the growth of Rome's population during the Roman Empire period led to increased demand for water; therefore, Roman engineers constructed an initial water system for importing and its distribution to households and public spaces through pipe networks and then the used water released back into the environment. Water **2.0** began when the United States' economy grew but industries contaminated river through polluted wastewater. To deal with this situation, American bacteriologists created an innovative wastewater system known as biofilm to purify the contaminated water for drinking purpose. During Water **3.0,** the development of sewage treatment was the focus by building holding ponds to consolidate water and regularize the speed of sewage flow through filters, which controlled microbes to treat toxic waste before discharging into rivers. The final stage, Water **4.0**, is one where the next generation of drainage system solves all problems of the previous three water systems by replacing outdated water infrastructure and making the community more aware about how to manage water resources (Sedlak, 2014).

Table 2.1 A comparison of different drainage systems from 3000BC to the present

No**.**	Period	Objective	Achievement	Limitations
1	Ancient civilization	Rainwater collection; flooding mitigation; and conveying wastes	Numerous successful and uneconomical sewer systems parallel with social planning	Lack of optimization and numerical standards before construction (Herbert, 1961)
2	Roman Empire	Rooftop rainwater collection;	Uniform roadway drainage practices and	Lack of calculation in balancing

No.	Period	Objective	Achievement	Limitations
		flooding mitigation; and water storage in underground structures	large underground sewers construction	water storage volume and water supply, leading to water overflowing into streets and public areas (Hodge, 1992).
3	Post-Roman era to the 1800s	Flooding mitigation and wastewater removal	Stone roadways with surface and subsurface drainage systems constructed with a crown in the centre and gutters along the sides; and sewers made of wood	Exposing many drainage problems due to the insufficiency maintenance of the sewers and then the spread of diseases.
4	From the 19^{th} century until now	Integrated water -related problems	Construction method and maintenance practices improving; wastewater treatment	Methods to design and planning a sustainable

No.	Period	Objective	Achievement	Limitations
			construction; application of computer modelling for design and construction; and enhancing environmental awareness	water management are still in the research and testing phases

The drainage systems have evolved over many centuries during trial-and-error modifications and after their implementation with increasing sophistication and multi-purpose concepts. Whilst initially the primary objective of rainwater management was flood mitigation in combination with rainwater collection for private purposes, from the 19^{th} century until now, this goal has expanded to integrate other aspects. They include water resources management, biodiversity and recreational and community purposes. In the 20^{th} century, industrial countries developed policies and strategies to address water-related issues due to their industrial expansion. Strategies included: best management practices (BMPs); low impact development (LID) in the United State in the 1940s (Ice, 2004); water sensitive urban design (WSUD) in 1990s Australia (Wang et al., 2018); and 2000s with the sustainable urban drainage systems (SUDS) in the United Kingdom. The above-mentioned Water **4.0** focuses on water protection and improving local people's awareness of water usage. There are various Sponge City models under BMPs-LID in the U.S. for rainwater collection and water quality improvement. BMPs-LID were divided into two broad groups: structural and non-structural. Structural BMPs-LID included ponds, wetlands and green rooftops (Scholz, 2006) that were built as multi-functional concepts for flooding mitigation, water quality enhancement, creating green spaces for recreation, and supporting ecosystems and wildlife. Water management in Australia experienced five different stages before implementing water sensitive cities under water sensitive urban design (WSUD) with the objective of climate change resilience and ecological integrity (Brown et al., 2009). The purpose of these concepts is to introduce integrated water management strategies toward more sustainable development that ensures the suitable management of water supply, water treatment and limiting flooding from rainwater.

These initiatives involved the design of drainage systems that would be constructed according to a sustainable development trend by considering and balancing all issues including quantity, quality and amenity in stormwater management. Hence, these strategic designs have resulted in improving community values, biodiversity, educational and recreational functions and multi-purposes of space (Developer, 2007). These strategies are different to the traditional approaches of rainwater management that did not balance amenity aspects in comparison with other aspects in terms of quality or quantity (Figure 2.2).

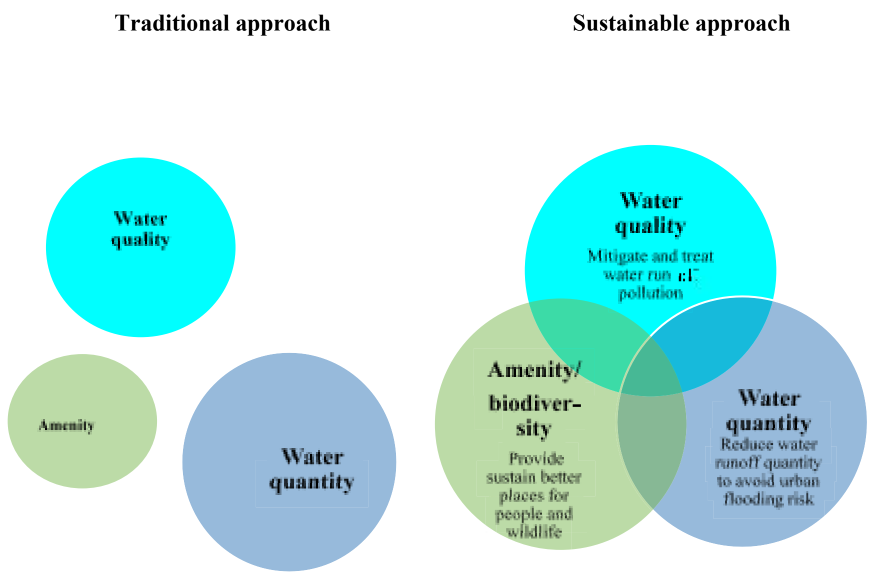


Figure 2.2 Differences between conventional and sustainable approaches of water systems

Although the structural BMPs-LID, sustainable urban drainage systems (SUDS), water sensitive urban design (WSUD), and Water 4.0 have been applied to different water problems in developed countries, it is still very much a new concept in developing countries. It is vital to utilize the most appropriate water management practices that are suitable for developing countries and their regions. SUDS and WSUD focus on natural hydrological protection and management of stormwater. In addition, water 4.0, which was introduced by David (2014), mainly considers water supply distribution and its value to society in the U.S. This type of strategies is difficult to address the complicated water-related problems that occur in developing countries like China (Ren et al., 2017). Water management problems first raised the concerns of scientists and authorities in China in the 1990s (Wang et al., 2018). Since then they have concentrated on how to design drainage systems as a sustainable water supply for the cities. From 2000 to 2007, the objective of water management in China was to recycle stormwater and treat wastewater. In the following 5 years, water strategies in China have increasingly focused on optimizing water drainage system for city water distribution and sewage purification. However, some water-related issues associated with flooding disasters and water shortages still exist in large areas of China. This started to occur when the Chinese government launched the “Open Door Policy” in the late 1970s and the country witnessed rapid population growth and socio-economic growth. For example, the population quadrupled from 1978 to 2015 (Chan et al., 2018). Due to this situation, the land use for green spaces, wetlands, forestry and agricultural land changed to urban areas for the purpose of commercial, residential and industrial development. This drastic loss in natural environment capital resulted in rainfall infiltration and absorption reduction, and less recharge of groundwater due to a lower retention capacity of rainwater in many regions in China, which results in water shortage in these regions (Arshad et al., 2014; Qin et al., 2013). Consequently, many cities in China have experienced flood disasters because of the inappropriate drainage system and the phenomenon of unpredictable weather. For this reason in 2014, the Sponge City concept became a reality in China to help develop sustainable cities (Jia et al., 2017; Wang et al., 2018). Sponge City is considered as an integrated water management solution.

A Sponge City is designed and implemented according to LID strategies that require the designer and builder having to mitigate the impact of construction on the environment including water, soil, vegetation, and biodiversity. Thus, respecting nature is the core value of LID strategies. Low impact development strategies aim to build sustainable drainage systems, better water cycle design and rainwater source controls. Technical methods in LID assist water infiltrating, filtering, evaporating, harvesting and surface runoff reduction while mitigating pollution. Therefore, the purpose of Sponge City is to absorb, store, treat stormwater and provide stored water for the public's use when needed through green infrastructure applications, for example, green roof, raingarden or bio-detention (Wu, 2015). These measures help to balance the water circulation system and create a high-quality living environment for both people and wildlife (Figure 2.3).

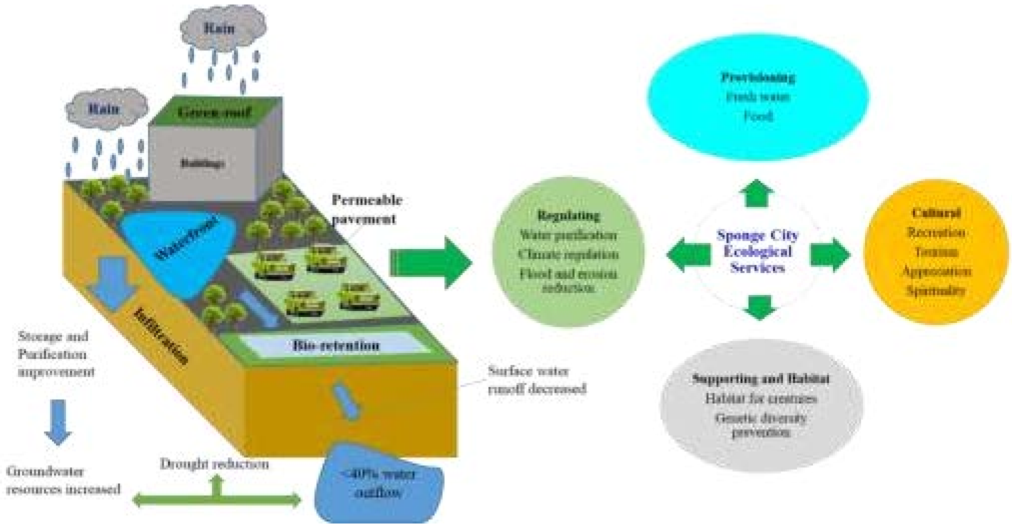


Figure 2.3 Schematic design of the Sponge City and Sponge City Ecological Services

There are many Sponge City objectives (see Figure 2.3). Firstly, it aims to control flooding disasters (Jia et al., 2017; Wang et al., 2018). In terms of climate change and urbanization, many cities in China face extreme flooding hazards. To overcome this, Sponge City has developed green infrastructure, such as, green roofs, bio-retention and permeable pavements in order to increase water absorption and water runoff reduction. As a result, flooding can be mitigated, however; it is recognized that an increase of wastewater due to population growth and industrialization results in serious water pollution in Chinese cities, affecting people's health and well-being. Secondly, the purpose of the Sponge City is to enhance water quality in urban areas by self-purification systems and ecological waterfronts. Therefore, a Sponge

City can help the water's ecological function restoration. The next goal is to recycle stormwater for water supply. The purpose of Sponge City is to alter rainwater as a resource, with the purpose of tackling the water shortage in the cities especially providing water back for city in terms of drought. In the period of 2000 to 2007, the Chinese government implemented of green infrastructure for stormwater utilization, however, the efficiency of rainwater utilization was initially not high in this period due to the lack of optimal elements including green technologies and materials (Liu et al., 2013; Shi et al., 2015). Sponge City aims at improving green infrastructures in order to consume or utilize until 70% of stormwater regionally and mitigate the effect or urbanization on the ecological system through green infrastructures. The final aim of Sponge City is to create a pleasant regional microclimate. Reducing the city's heat through increasing the green spaces with green rooftops, lakes and wetland areas is a one of major aspect of the Sponge City. Therefore, Sponge City concept can be applied for water management in both urban areas, sub-urban areas, and rural areas.

2.2.3 Recent water management initiative –Sponge City

2.2.3.1 Principle concepts

There are four main principle concepts of a Sponge City (Figure 2.3). The first aim is to make the surface of the city more capable of absorbing and storing rainwater in order to supply water and to mitigate the water runoff, which leads to floods. The second principle is about water ecology management via water self-purification systems and the provision for ecological waterfront design. The third is concerned with the application of green infrastructure to purify, restore, adjust and reuse stormwater, which helps the cities avoid water and soil pollution. This reduces the heat island effect and supports sustainable urbanization. Fourthly, using permeable pavements in road construction will benefit a Sponge City.

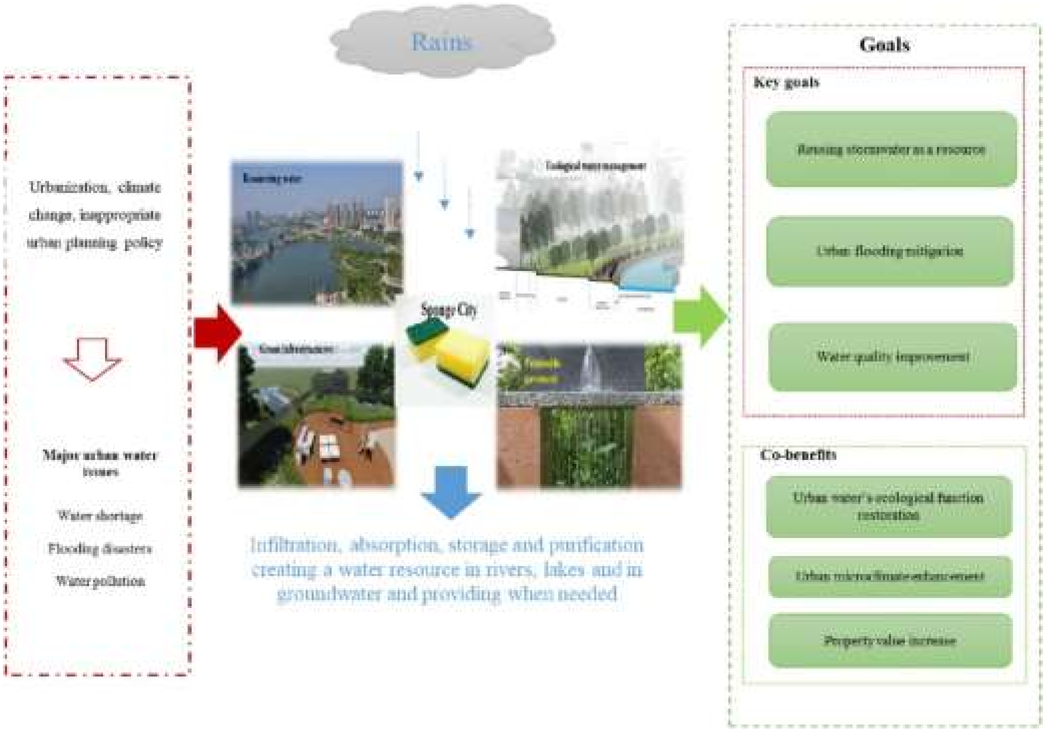


Figure 2.4 Sponge City's principle concepts and objectives

Sponge City implementation can be divided into macro-scale and micro-scale scenarios. In the micro-scale context, Sponge City mainly focuses on implementing designs for site level including rain gardens, stormwater-bio-retention and constructed wetlands. From maximizing the effective of micro-scale Sponge City in site level and localized level, Sponge City is scaled up into catchment level in order to enhance hydrological and bio-ecological benefits (Zhang and Chui, 2019). With macro-scale, stormwater infrastructure systems are integrated with natural hydrology systems to protect riparian corridors including grass, trees, shrubs and the buffer areas of these corridors. A novel model and the availability of spatial data (social-economic, land use, climate, green infrastructure practices and hydrological condition information) play a vital role in upscaling of Sponge City technologies from plots and localized areas to catchments (Golden and Hoghooghi, 2018; Zhang and Chui, 2019). The successful implementation of a catchment-scale Sponge City by maximizing green infrastructure practice in large-scale contributes to the thorough Sponge City program.

2.2.3.2 Resourcing rainwater

The rapid increase in population, growth of population and industrialization and excessive use water for agricultural practices have posed a threat to existing water resources, which damages the city's development and the basic lives of the inhabitants. Therefore, an alternative water resource is required to combat the lack of water resources, given that sustainable water resources management plays a vital role in socio-economic development. Rainwater harvesting (RWH) is the most ancient method to address water shortage (Campisano et al., 2017). The existing RWH initially focused on the restoring and recycling of stormwater without paying attention on the multi- benefits of RWH. RWH is the approach of LID, SuDS and Sponge City concept, which aims at decreasing peaks, frequency and volumes of water runoff. The implementation of RWH systems improve water self-sufficiency and mitigate the impact of urbanization on water bodies (Christian Amos et al., 2016). In order to implement a successful Sponge City for resourcing rainwater, it is necessary to understand the area's hydrological characteristics including water surface runoff, flow time, discharge, speed, size and peak time to better connect between natural water networks and drainage systems to control flooding and enhance the water storage capacity of infrastructure systems. In addition, wetland ecosystem design including natural and artificial designs for RWH is considered as an important aspect in terms of Sponge City rainwater resourcing construction, which improves climate regulation, flood prevention and water purification. It also provides the landscape for entertainment and leisure activities that the community can enjoy.

2.2.3.3 Ecological water management

Rapid increasing of population and industrialization have threatened the water quality in many areas. Therefore, water pollution reduction and water quality enhancement is one of the important roles of Sponge City implementation. The Sponge City ensures the water environment being restored ecologically through a self-purification system and waterfront design and creates healthy water landscapes for people and wildlife. The water self-purification process is very complicated in that it includes physical, chemical and biological processes. Normally, the biological purification is considered an environmentally friendly method. There are four factors, which significantly influence water self-purification. The first factor is the hydrodynamic force, which adds to a mixture of pollutants, their movement and the water dissolved oxygen content. Soil is the second factor and it serves to remove pollutants through absorption, sedimentation and filtration. The third one is plants that can remove heavy metal pollutants, nitrogen and phosphorus. Finally, microorganisms can help contaminant degradation. Another form of ecological water management is ecological waterfront design. Waterfront design aims to integrate cities and a water system in order to develop macro and micro environments. In the construction of a Sponge City, both natural and artificial ecological waterfronts are considered to protect against riverbanks' erosion and consolidate the water self-purification system (Wu, 2015).

2.2.3.4 Green infrastructure

Green infrastructure has emerged as the solution to protect the environment and make environments sustainable. There are two main types of green infrastructure in Sponge City implementation including green roofs and bio-retention. These infrastructures are described in more detail below:

➢ Green roofs

Green roofs are also known as living roof or rooftop garden (Mentens et al., 2006; Sailor, 2008; Shafique et al., 2018; Stovin, 2010). Green roofs are constructed with roofs covered by vegetation and/or plants. Green roofs concept also has been mentioned in LID strategies and SUDS techniques. Green-roofs promote vegetation planting on the top of buildings. They are to reduce stormwater run-off, mitigate the heat island effect; reduce energy consumption; enhance air and water quality, improve wildlife habitat and plant life, and boost recreational activities through green areas (Besir and Cuce, 2018; Brudermann and Sangkakool, 2017; Chen et al., 2018; Coma et al., 2018; Mentens et al., 2006; Sailor, 2008; Shafique et al., 2018). The application of green roofs started long years ago but the modern green roof system started initially in Germany during the early 1960s with the initial aim of energy consumption reduction for buildings. Germany is considered as a leader in green roof technology because its system is well developed, designed and implemented on a large-scale (Zhang et al., 2011).

➢ Bio-retention

Bio-retention systems are also known as bio-filters or raingardens (Davis et al., 2009; Fujita, 1997; Laurenson et al., 2013; Mangangka et al., 2015; Trowsdale and Simcock, 2011). Normally, a system of bio-retention consist of five main components, which are drainage layer, transition layer, submerged zone, filter media and detention layer (see figure 2.5). Each layer of the bio-retention system is constructed according a specific area's condition. These help filter polluted stormwater and remove contaminants via biological processes using active plants and sandy loam layers. Bio-retention systems are considered as feasible solutions in sustainable rainwater management practices (Muthanna et al., 2007). Furthermore, these systems are designed to manage stormwater peak flow, runoff volume reduction, groundwater recharge enhancement and stream base flow maintenance (Davis et al., 2009; Randelovic et al., 2016; Rycewicz-Borecki et al., 2017; Wang et al., 2017a).

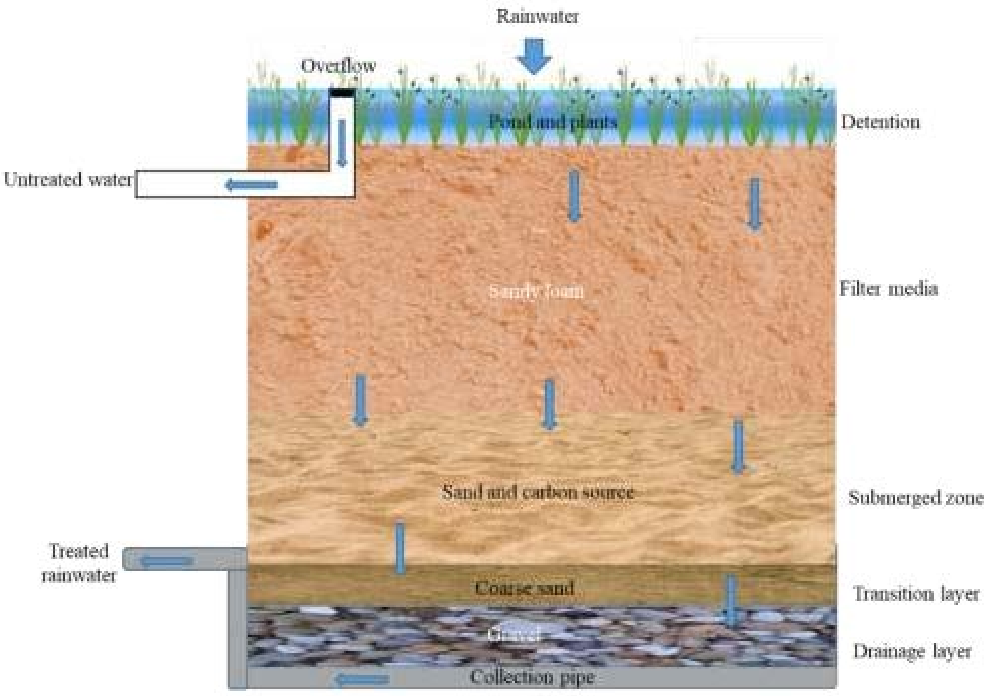


Figure 2.5 Bio-retention design

2.2.3.5 Permeable pavement

Permeable pavement is a Sponge City technology that utilizes permeable materials to build ground pavement with the purpose of improving rainwater infiltration, and purification of groundwater for water supply, reducing water runoff, cooling, humidification, noise reduction, and environmental and ecological soil restoration (Hu et al., 2018; Liu et al., 2018; Scholz and Uzomah, 2013; Yu et al., 2017) (Figure 2.6). Pavement infrastructures includes roads, squares, parking, and rural site walkways (Kamali et al., 2017). Each type of pavement is designed differently depending on traffic quantity and road loads. There are four main types of permeable pavements: grass pavement (GP) permeable asphalts (PA), permeable concretes (PC), and permeable interlocking concrete pavers (PICP) (Woods Ballard et al., 1015). The performance of PC is better than PA and PICP if without clogging influence (Hu et al., 2018). Such permeable pavements have been adopted in many regions the world over to mitigate flood disasters using urban water management practices, such as low impact development and best management practices (Brunetti et al., 2016; Hu et al., 2018). These permeable pavement parking lots may bring significant benefits for eco-systems because of their potential for the rapid infiltration of storm water to reduce a high level of water runoff (Kumar et al., 2016). Some studies illustrated that the performance of permeabele pavements in terms of flooding reduction are even better and more effective than other Sponge City technologies, including greenroofs or rainwater resourcing (Chandana et al., 2010; Hu et al., 2018). For example, the construction of permeable pavements decreased about 35.6% of total water run-off in Tianjin University campus, China (Huang et al., 2014) However, the construction of permeable pavement should be suit the local condition. In high polluted regions and unfavorable soil permeabiltiy, this technique is not suitable (Yu et al., 2017).

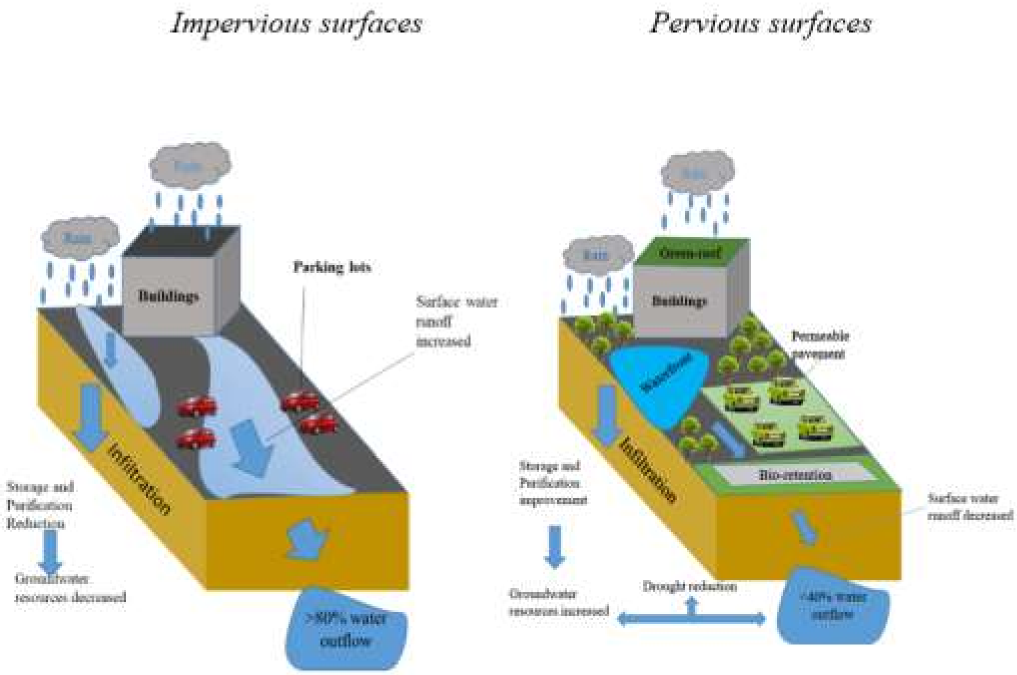


Figure 2.6 Difference between impervious surfaces and pervious surfaces

2.3 Challenges and opportunities of water management strategies adoption

2.3.1 Technical and physical challenges

There are some major technical challenges in implementing the water management practices. There is a technical gap between developed and developing countries, where the latter have limited expertise or skills regarding green materials for green roofs or bio-detention, lack of technical guidance and training, outdated supportive simulation models, and insufficient performance data for planning and designing. In addition, operation and maintenance difficulties and some physical challenges including climate, soil and geographical conditions are also barriers to implementing a successful water management infrastructure.

2.3.1.1 Technical gaps and limitations

The first overall problem is the technical gap between developed and developing country, so developing countries find it difficult to apply or accept technologies from developed countries (Jia et al., 2017; Li et al., 2017; Li et al., 2016; Li et al., 2018; Ma et al., 2017; Xia et al., 2017a). All countries have their own distinctive geographical, climatic and soil conditions, so the design strategies implemented and the measures for construction need to be compatible with each location and or region, and should not be modelled incorrectly. Although the LID practices were introduced across the world some decade ago, large - scale developments are still limited due to the lack of domestic and local research. In addition, water management technologies and the green infrastructure industry are much more advanced in wealthier countries, such as, Germany, the U.K and the U.S than developing countries. These developed countries have industries that can provide a range of green materials for green infrastructure building. The unavailability and uncertainty of rain garden system, green roof system, tree planter, infiltration planter system, underground infiltration and monitoring systems can greatly affect the effectiveness of the water infrastructure programs (Li et al., 2017; Xia et al., 2017).

2.3.1.2 Lack of technical guidance and training

The major factor that contributes to the unsuccessful adoption of the best water management practices is the limited design strategies that encompass relevant standards and codes for diverse regions. The lack of the presence of experts (architects and regional planners, hydraulic and environmental engineers, hydrogeologists and agronomists) to support the implementation of water management practices and the limitation of educational and training courses for the up-skilling of staff leads to inappropriate approaches and planning that will not produce successful outcomes.

2.3.1.3 Current and relevant simulation models for rural water management design

To assist rural water planning and assessment the correct simulation model is required. Computer models apply a simulated design, policy, and strategy including widely used or accepted measures. Computer software modelling has the ability to consider various factors and environmental scenarios, for example; stormwater management in developed countries as part of adopting the sustainable water management model (SWMM); stormwater quality model (SQM); and water modelling software (MIKE-URBAN) (Bach et al., 2014). Enhanced models that can be adapted to the complex diverse water system variables in developing countries must produce a good fit and align well with the objectives of rural water management purpose and mission.

2.3.1.4 Unavailable rural performance data

Creating a water management strategies is very complicated and many variable factors have to be accounted for: hydrology, land-use systems, regional development, and biodiversity (see Figure 2.7). All these aspects require evidence-based research. Performance of data for specific locations helps us to understand the range of climatic conditions and/or natural precipitation data, soil data, surface and groundwater characteristics, existing drainage systems and information on how buildings are constructed (Li et al., 2017; Shao et al., 2016).

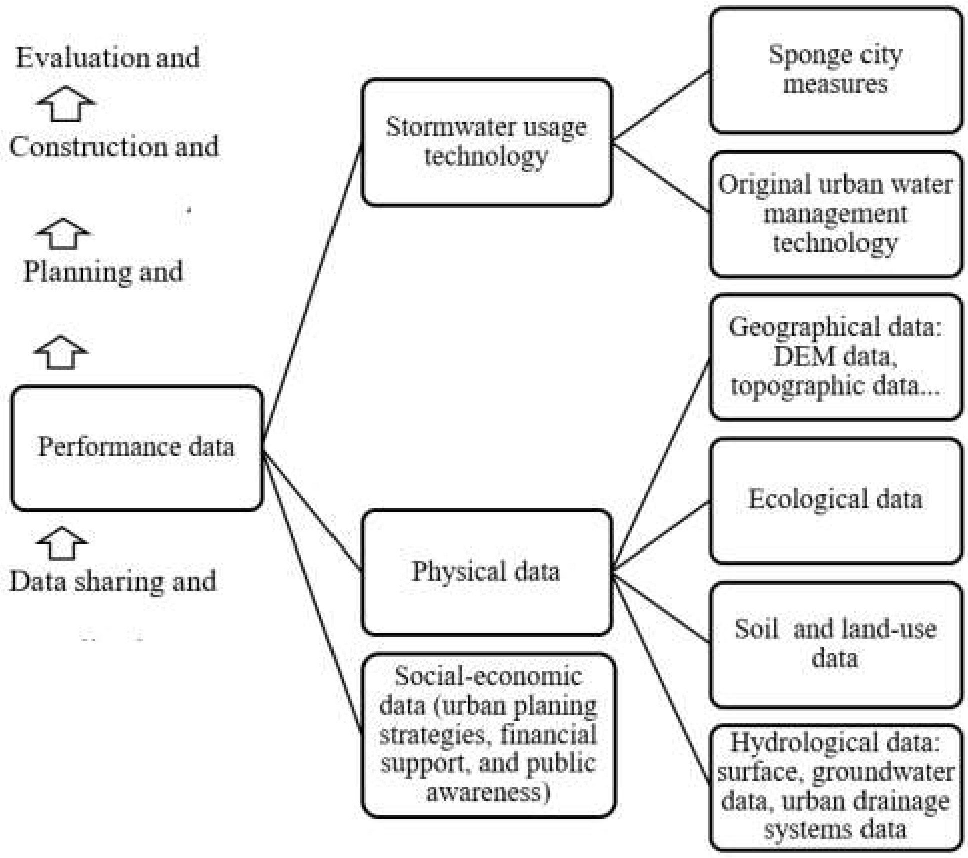


Figure 2.7 Required data for water management design

2.3.2 Financial challenges

Water infrastructure construction requires substantial investment. Research on cost-benefit analysis of such projects is still limited, so the cost and benefit of water infrastructure projects is difficult to compare. Liang (2018) conducted a project to assess the socio-economic and financial or investment aspects of Sponge City projects in the city of Chang-De in China. The project found that the environmental and social benefits of the Sponge City program should be encouraged by the government and the wider society. However, it determined that the private sector would not invest in a water infrastructure project due to higher costs not being matched by higher revenues (Liang, 2018). As a result, the objective appears to be somewhat challenging in terms of the Chinese government attracting the investment of public-private partnerships, as this would require approximately 50% of total costs and may not be feasible. Nonetheless, this is only one scenario where the life cycle costs have to consider the design, construction costs, operation, maintenance costs, and compared them to the uncertain environmental, ecological and social benefits. Consequently, more research is required on the financial viability of water infrastructure projects, in which life-cycle benefits should be clearly articulated in terms of social wellbeing, return on investment, the value of private-public partnerships (PPP) and the role of local or regional organizations.

2.3.3 Administrative fragmentation

Local government administration system lacks cooperation between related functions or agencies (Jiang et al., 2017b; Li et al., 2017). As inter-connectedness is required to promote the water infrastructure's aims for positive societal outcomes with rural planning, water management, land use and supporting eco-systems, so the value of community cooperation and the support from all levels of administration and agencies are essential for water infrastructure's construction. The complexity of water infrastructure implementation requires not only appropriate acceptance of technologies but also strong management systems and governance capabilities. The complex and fragmented structure of governance offers less opportunity for participation and collaboration between ministries, public/private sectors and local government bodies. Furthermore, the objectives of water infrastructures may prove too difficult to achieve without a sense of collaboration and the strong coordination between multiple stakeholders in sharing data, financial resources, etc. Having an appropriate legal framework or agency body that can ensure sharing of benefits between sectors will help cooperation and involvement of inter-agencies.

2.3.4 Public awareness and acceptance challenges

Community acceptance challenges is one of the strongest barriers against the adoption of water infrastructure. The subsidy from the central government is inadequate due to the high funding requirement for water infrastructure construction. Therefore, the mobilization of non-government sources of finance is very important in terms of water infrastructure planning and building especially when insufficient financial resources from the government make these sorts of projects difficult to achieve. Achieving public participation, their willingness to invest, and having the education, training and information dissemination methods to support regions should be discussed with a broad array of public groups ranging from political leaders to everyday citizens (Li et al., 2017; Wang et al., 2017b).

2.3.5 Opportunities and future perspective for water infrastructure implementation

The significant technical difficulties are the lack of an appropriate simulation model that includes the relevant factors based on evidence for implementing water infrastructures designs. The second serious barrier is a financial aspect in that building and maintaining rural water infrastructures requires a huge monetary input, which the central government is reluctant to provide. It is important to attract funding from the public-private groups and international organizations. The successful implementation of water management programs bases on four key aspects, which are illustrated in Figure 2.8.

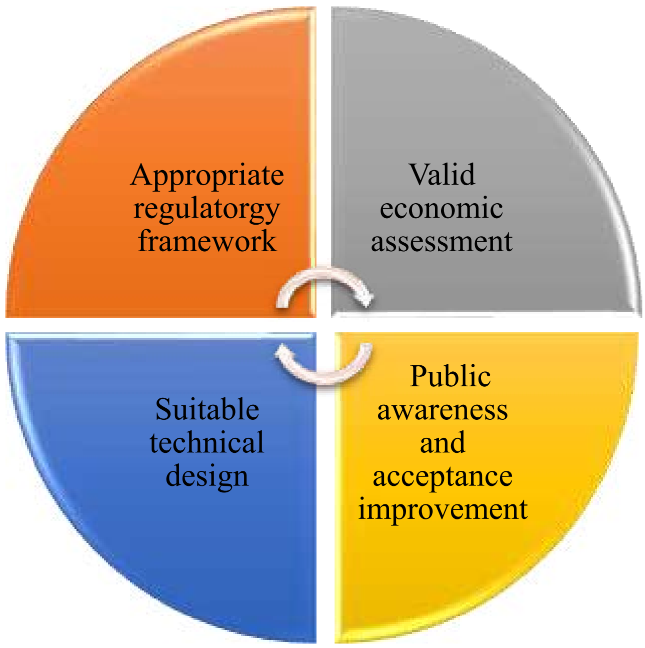


Figure 2.8 Key aspects to successful implementation of water management practices

The various technical, physical financial, regulatory and community challenges of water management are summarized in the table 2.2. The presence of these challenges can hinder the overall uptake of the water infrastructures implementation in rural areas.

Table 2.2 Challenges and outlook for future development of water management strategies

No.	Type	Challenges	Efforts for future development	References
1	Technical and physical challenges	• Deficiency in technical standardization and guidance for each region and each city • Lack of performance data (e.g. hydrological data, soil and climate data for technical feasibility assessment before project implementation). • Lack of simulation model for water measurements development • Unavailable materials and green products like pavement materials for water absorption and green rooftop system.	• Improving local guidance and standard for each city • Providing sufficient performance data; Establishing an effective national database to determine the suitability of Sponge City • Building new simulation model for water infrastructures design before implementing it; No	Chan et al., 2018; Che et al., 2015; Jia et al., 2017; Li et al., 2017; Ren et al., 2017; Roy et al., 2008; Šakić Trogrlić et al., 2018; van de Meene et al., 2011

No.	Type	Challenges	Efforts for future development	References
		• Lack of land for designing wetlands and ponds due to urbanization and industrialization • Litter or no maintenance and monitoring water infrastructures construction • Lack of the presence of experts (architects and planners, hydraulic and environmental engineers, hydrogeologists and agronomists) to support the implementation of Sponge City	outdated modelling software for water strategies implementation • Developing suitable material and green products for each region • Issuing appropriate land use policy to limit land use change from natural land to residential and commercial areas • Establishing a suitable maintenance and monitoring system	

No.	Type	Challenges	Efforts for future development	References
			• Raising the involvement and support of related experts; training for localized staff to implement Sponge City technologies	
2	Financial challenges	• Requirement of a substantial funding for implementation and the unknown cost of maintenance and operation. • Insufficient data for life cycle economic feasibility assessment of Sponge City • Shortage of funding sources and effective market incentives	• Establishing an economic feasibility assessment tool for Sponge City planning. • Raising financial resources from the private and community sectors through providing environmental and economic benefits of	Li et al., 2017; Li et al., 2018; Liang, 2018; Roy et al., 2008; Zhang et al., 2018a

No.	Type	Challenges	Efforts for future development	References
		• The high value of land for Sponge City construction • The low amount of willing to pay from community	Sponge City projects for those sectors	
3	Legal and Regulatory Challenges	• Lack of understanding, close cooperation between agencies involved in Sponge City implementation • Lack of legislative mandate, fragmented responsibilities, institutional capacity • Setting ambitious target without background knowledge	• Strengthening coordination between agencies • The secret to success is producing clear and right objectives which are suitable for each location • Enhancing local legislation framework	Chan et al., 2018; Jiang et al., 2017b; Li et al., 2017; Li et al., 2018; van de Meene et al., 2011; Wu et al., 2017

No.	Type	Challenges	Efforts for future development	References
4	Public's awareness and acceptance	• Lack of knowledge about overall significance of the water strategies and ineffective communication • Resistance to change, limited community engagement and missing support from public-private sector	• Conducting both informal and formal education and training courses to enhance public perception and engagement	Li et al., 2017; Li et al., 2016; Li et al., 2018; van de Meene et al., 2011; Wang et al., 2018

2.4 Water management monitoring and modelling

2.4.1. Overview of current water modelling

A number of water modelling and analysis approach have been developed to support for sustainable water management in both rural and urban areas. Integrated water management models might include water treatment models, wastewater collection models, wastewater treatment models, rainfall and surface runoff models, river models, water distribution models, environmental assessment models, economic assessment models, and social assessment models. Integrated water models can be divided into various groups according to their functions and their integration levels. According to their function, they could be classified integrated water management models into integrated water treatment models or integrated water quality models, integrated water supply models, and integrated water multi-criteria analysis models, which describe in Figure 2.9.

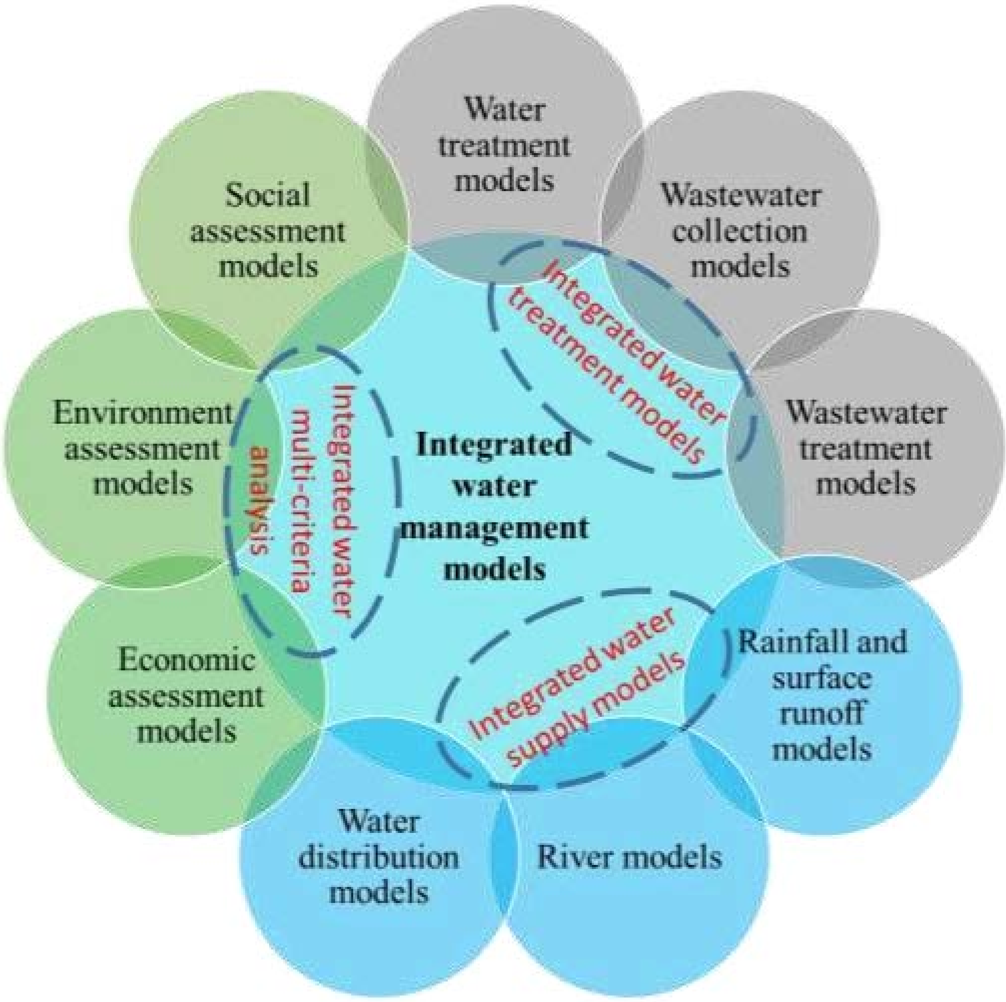


Figure 2.9 Sub-models of integrated water modelling.

According to Bach et al. (2014), integrated urban water models were divided into four groups based on different integration levels. These are: (i) integrated component-based models (ICBMs); (ii) integrated urban drainage models (IUDMs) of integrated water supply models (IWSMs); (iii) integrated urban water cycle models (IUWCMs); and (iv) integrated urban water system models (IUWSMs) (Figure 2.110). While ICBMs represent the lowest level of integration, IUWSMs are the highest level and this emphasizes the importance of water flows in the urban environment. For the four groups noted above, the scope of these models is broader.

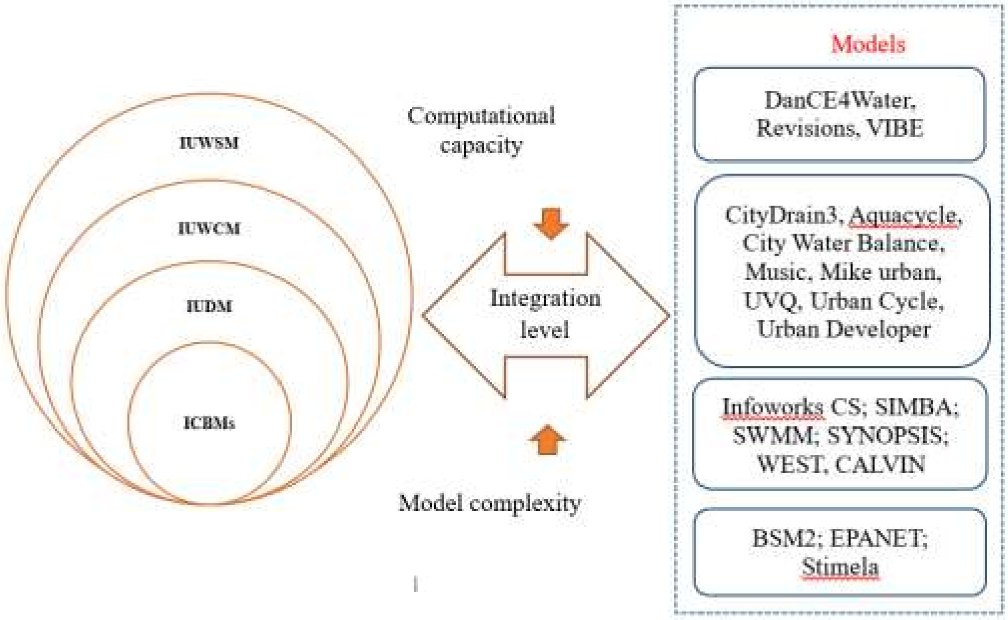


Figure 2.10 Classification of integrated water models (adapted from Bach et al., 2014).

One limitation of integrated water models is that they are not able to evaluate comprehensively economic, social, environmental benefits and ecosystem services of water strategies, while stakeholders tend to depend on this evaluation to make their decisions (Castonguay et al., 2018) (Figure 2.10). For example, ICBM models including BSM2, EPANET and Stimela are considered as a form of plant-wide integration that does not pay attention to flooding problems of urban water. IUDMs and IUWCMs such as InforWorks CS, SWMM, and MIKE-URBAN only link between urban development and urban water infrastructure and do not consider the environmental, economic and social of urban water management infrastructures (Schellart et al., 2010; Burger et al., 2014; DHI, 2009). Although the DAnCE4Water model was defined as the highest integration level model, it only considers partially the interactions between urban water infrastructure and environmental, social and economic aspects.

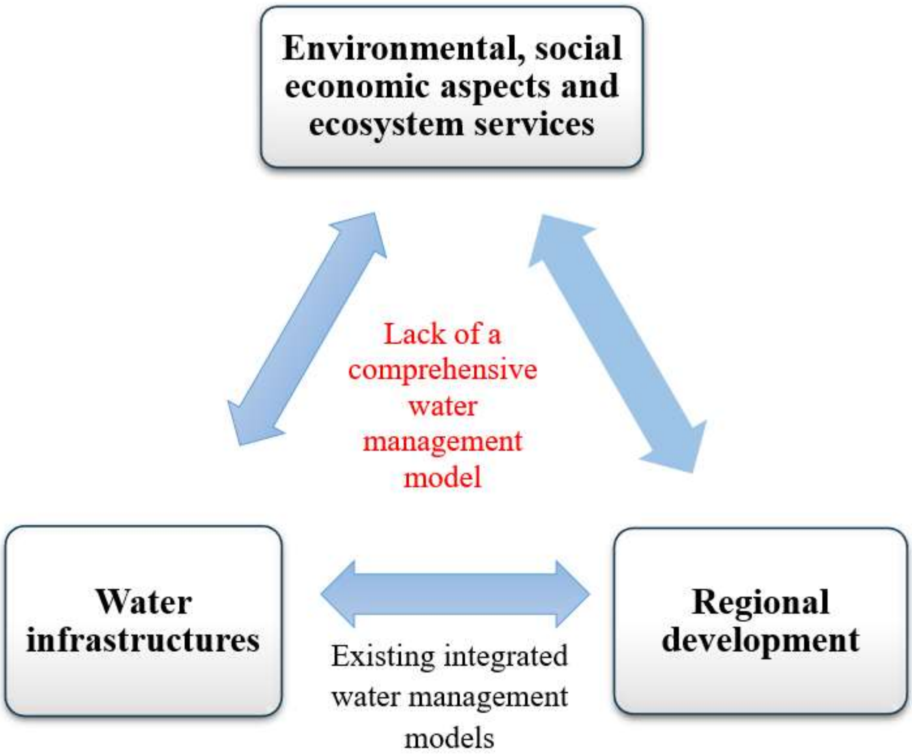


Figure 2.11 The limitation of existing integrated water management models.

Moreover, there are several key barriers from case studies that applied the existing integrated water management models such as SWMM, LCA and SUSTAIN for the assessment of rural water infrastructure implementation's performance (see figure 2.11). They are (1) the uncertainty of spatial and temporal data, (2) the limitations in the comprehensive assessment of ecosystem services of water infrastructures, (3) the lack of the assessment of long-term benefits of water measurements, and (4) limitation in simulation with long-time series data such as rainfall data. Therefore, the development of a comprehensive integrated water model will be significantly reduced these limitations, which is summarised in Table 2.3.

Table 2.3 Applying different integrated water modelling for water strategies assessment

Case study	Model (integration)	Study site	Description	Results	Barriers
Li et al., 2019	SWMM, AHP	Guangxi, China	Simulation the benefits of Low Impact Development (LID) practices (bio-retention, grassed swale, sunken green space, permeable, storage tank) in Sponge City program	Quantified environmental, economic, social benefits of these practices	✓ Lack of assessment of long-terms benefits and performance of LID practices ✓ Lack of assessment of the effect of climate on LID measurements. ✓ Lack of comprehensive evaluation of

Case study	Model (integration)	Study site	Description	Results	Barriers
					ecosystem services of these practices.
Zhao et al., 2018	The emergy-GIS framework based on SCS-CN model, L-THIA model and energy balance model	Shenzhen, China	Identification of appropriate areas for Sponge City construction	Selected Sponge City implementation areas based on the degree of water runoff, water pollution, heat discharge	✓ Limitations in collection of precision satellite imagery data ✓ The probability of deviations and errors of sub-models

Case study	Model (integration)	Study site	Description	Results	Barriers
Hou et al., 2019	SWMM, GIS	Yinchuan, China	Simulation ecological stormwater processes of different LID facilities in a Sponge City	Simulated water runoff, thermal landscape, purification process of Sponge City measurements	✓ Limitation in simulation with long-time series data such as rain-fall data ✓ The precision of input data such as DEM data, pipe network needs to be higher

Case study	Model (integration)	Study site	Description	Results	Barriers
Deng et al., 2018	SWMM, GIS, CAD	Yuelai, Chian	An integrated stormwater management system model to evaluate the whole life cycle of LID facility in Sponge City program	Stormwater network system construction LID facilities design and optimization	✓ Need a huge amount of data for calibration such as long-term climate data, soil infiltration coefficient… ✓ Lack of the evaluation of economic, social feasibility and ecological services of LID facilities in Sponge City

Case study	Model (integration)	Study site	Description	Results	Barriers
Mei et al., 2018	SWMM, Life Cycle Cost Analysis	Liangshuihe watershed	Integrated evaluation of green infrastructure for flood mitigation to support Sponge City implementatio n	Assessed hydrological performance assessment of green infrastructure (GI) practices Evaluated cost-effectiveness of GI strategies	✓ Lack of experimental data for calibration of the integrated assessment system causing model uncertainties ✓ Long-term benefits of GI practices are not evaluated ✓ Lack of GI practices planning and limitation in

Case study	Model (integration)	Study site	Description	Results	Barriers
					ecological services of GI practices under Sponge City program
Mao et al., 2017	SUSTAIN	Foshan New City, China	Application of SUSTAIN model to assess the ecological benefits of aggregate LID-BMPs in Sponge City program	Planned LID-BMPs facilities for the city Evaluated the ecological benefits (e.g., water runoff control performance) of LID-BMPs	✓ The cost-effectiveness of LID-BMPs practices is not calculated ✓ Limitation in assessment of comprehensive ecological services of LID-

Case study	Model (integration)	Study site	Description	Results	Barriers
				and the costs of these practice	BMPs practice including environmental and social benefits.

2.4.2 Remote sensing and machine learning application for water monitoring

Recent advances in geospatial methods using earth observation (EO) datasets and advanced machine learning (ML) techniques can be effective in water modelling as well as the monitoring of water infrastructure. Included here are soil moisture and carbon prediction originating from different water management practices (Vaudour et al., 2019). The use of multispectral, hyperspectral, or synthetic aperture radar (SAR) data from space-borne, air-borne remote sensing platforms, or unmanned aerial systems (UASs) has emerged as an innovative solution to address the issues of soil moisture and soil carbon prediction on farming lands. Although the performance of airborne RS and UAS with high spatial resolutions of hyperspectral images and extensive spectral information in the prediction outperforms the space-borne sensors with multispectral bands, the scarcity and high costs of hyperspectral data hinder their application in large-scale soil moisture and soil carbon estimation (Angelopoulou et al., 2019; Guo et al., 2021; Table 2.4).

Table 2.4 Prediction performance of agricultural SOC in the recent literature.

Type of sensor	Sensor	ML Algorithm	R^2	Reference
Space-borne	Hyperion	PLSR	0.49	(Gomez et al., 2008)
	PRISMA	PLSR	0.51	(Castaldi et al., 2016)
	Landsat ETM+	ANN	0.63	(Mirzaee et al., 2016)
	S2	PLSR	0.56	(Vaudour et al., 2019)
	Gaofen 1	ELM	0.84	(Guo et al., 2020)
	S-1+S2 +DEM	BRT	0.44	(Zhou et al., 2020b)
Air-borne	AHS-160	SVM	0.89	(Stevens et al., 2010)

Type of sensor	Sensor	ML Algorithm	R^2	Reference
	HyMap	PLSR	0.85	(Vohland et al., 2017)
Unmanned Aerial Systems	Mini-MCA6	SVM	0.95	(Aldana-Jague et al., 2016)

PLSR: Partial Least Squares Regression; SVM: Support Vector Machines; ANN: Artificial Neural Networks; ELM: Extreme Learning Machine; BRT: Boosted Regression Trees;

Multispectral remote sensing sensors such as Hyperion, S-2, S-1, Gaofen 1, Landsat ETM+, and PRISMA have demonstrated their usefulness in agricultural SOC estimation. The free-of-charge multispectral images constitute an effective solution to address the problems concerning hyperspectral images in agricultural SOC monitoring. Gaofen 1 - launched by China National Space Administration – has great potential in estimating agricultural SOC with 0.84 R^2 compared to other multispectral images (Guo et al., 2020). However, its spectral bands are not widely supported by various agencies of the Chinese government. Combining multi-sensors in predicting agricultural SOC has been done in recent research such as: the integration of Sentinel 1 and Sentinel 2; and joining Sentinel 2 and Sentinel 3 (Zhou et al., 2020b; Zhou et al., 2021). Multi-sensor data fusion technology is a promising way to improve prediction performance compared to single sensor technology (Khaleghi et al., 2013; Le et al., 2021).

A few studies have combined optical data (S-2) and SAR data (S-1) to estimate agricultural SOC and SM content (Zhou et al., 2020). Recently, Zhou et al (2020) explored the potential of using S-1, S2, and digital elevation model (DEM) data in predicting agricultural SOC by Boosted Regression Tree (BRT) machine learning technique. This had a prediction accuracy of 0.44 R^2, which is quite low compared to other research (Table 2.4). This is likely due to the optimization of hyper-parameters tuning and the selection of predictor variables during the construction phase of the ML techniques. A range of ML algorithms were used for agricultural SOC and SM monitoring which are presented in Table 2.4. ML techniques using satellite datasets are considered to be effective and low-cost methods for rural water monitoring. Therefore, with these considerations in mind this should be integrated into the sustainable rural water management model.

2.4.3 Barriers of existing integrated water planning models

There are seven challenges associated with integrated water planning models (Lee, 1973) that are also considered as the barriers of integrated water models according to the vision of integrated modellers (Tscheikner-Gratl et al., 2019). These challenges include (1) hyper-comprehensiveness, (2) complicatedness, (2) grossness, (4) hungriness, (5) mechanicalness, (6) wrongheadedness, and (7) expensiveness. The main challenges of a Sponge City model that the paper want to highlight here are:

2.4.2.1 Model complexity

The integrated model should be able to simulate water strategies, efficient hydrological performance and the assessment of catchment water and environmental indicators. The multiple objectives of the water model originate in model's complexity due to high level of integration or too much linkage in the model and the remarkable amount of simulations creating the errors of model's computation and coding processes so that modellers need to take into account these errors. The complexity of model creates a huge amount of data availability requirements and uncertainty problems in the model that should be addressed by sensitivity and uncertainty analysis to improve the effectiveness of integrated models (Schellart et al., 2010). For this reason, it is crucial to determine the adequate degree of integration and simplified solutions for each model being implemented to ensure issues regarding uncertainty are minimized.

2.4.2.2 Limited knowledge about water systems

Water networks are complicated systems, which are incorporated by various processes including storage, infiltration, transport and distribution of water, and their interactions (García et al., 2015). The process of each water management system is different. These factors lead to the limitations in understanding the interactions between water cycle components and their performance, which threatens the success of any hydrology simulation model. A comprehensive understanding of water systems is extremely important to prevent major problems such as flooding, water pollution and water shortages (Marlow et al., 2013). Especially, integrated water management practices have emerged intending to build in future sustainable water systems. Therefore, the implementation of any water infrastructures should consider climate change, population growth, and regional development scenarios. At present, hydrological advances such as new technologies for recording and predicting rainfall in areas can support and manage water resources.

2.4.2.3 The lack of stakeholders' involvement

Single tools or models for water management like ICBMs might be easy for practitioners to use if the interfaces are properly set up. Integrated modelling aims to simulate a range of processes and components in the system with a spectrum of temporal and spatial data (Tscheikner-Gratl et al., 2019). This generates complex links and interactions in integrated models which affects the stakeholders' adoption due to a poorly established interface (Marsalek et al., 1993). Another significant problem causing integrated models to not be entirely user-friendly is the lack of training to transfer the models due to limited time and rising costs. Therefore, building a user-friendly interface model that is suitable for both users either with an immediate level of model application skill and/or an advanced modelling professional is essential that helps improve stakeholders' involvement and participation in the modelling process (Bach et al., 2014; Heusch et al., 2010).

2.4.2.4 Model's cost-effectiveness

The monetary investment is a fundamental factor in any project development, and the water model requires substantial investment costs. Modellers are still required to cover all the costs associated with model development if they cannot obtain government funding or the backing of policy-makers. The required costs and efforts need to be able to create and manage huge data requirements; model calibration and building an integrated rural water model but the costs involved might exceed the value of the output (Ahyerre et al., 1998). For this reason, the integrated models work better as research models but not as practical ones (Bach et al., 2014). To increase the cost- effectiveness of the rural model and retain the interest of practitioners, all uncertainty issues in model development have to be solved or minimized (Diaz-Granados et al., 2009; Sriwastava Ambuj et al., 2018). Despite the integrated model's development costs, it does play an important role in assisting decision-makers in making investments and the development of strategies that support various regions.

2.5 Conclusions

Climate change, rapid population growth and inappropriate development policies in many countries have resulted in water-related problems, such as flooding disasters, water pollution and water shortages. To tackle these issues, water management strategies such as green infrastructures, low impact development strategies, sustainable water management, and Sponge City have been implemented. There are complex systems with many challenges. Uncertainties in water infrastructure design and planning, and financial insufficiencies are the most serious problems that can risk the failure of the water management concept. While significant barriers exist, the opportunities for implementing sustainable rural water systems are evident. To obtain multi-ecosystem services of water measurements, they should be implemented at the watershed scales and be flexible, depending on different decision levels or catchment characteristics. It is essential to apply an intelligent decision-making mechanism and consider the need for close cooperation between various agencies with which the central government can work. A suitable sized and harmonious rural water infrastructure, ensuring a good balance between socio-economic development and environmental conservation, is the ideal.

More research is required to build a comprehensive computerized model using free-of-charge satellite images and advanced machine learning model for sustainable rural water infrastructure design, identify appropriate technical measures including bio-detention systems, and green roofs. These are possibilities that can be applied in each locality. It is also important to improve co-ordination across government bureaucracies through the establishment of a rural water database system and experiences-sharing networks to deliver a successful, large-scale rural model. Finally, an economic valuation of rural water practices which highlights the whole life cycle benefits and risk of failure should be conducted to enhance public-private knowledge and perceptions. Doing so could enhance their willingness to support the implementation of the rural water management practices and its sustainability features. Rural strategies and policies focused on promoting this concept play an important role in developing healthier, resilient and sustainable cities in an era of climate change and massive urbanization.

Chapter 3

Research Methodology

3.1 Materials

3.1.1 Spatial datasets

Six global gridded geographic datasets, specifically GADM, ASTER GDEM, CHIRPS, Terra MODIS, Global Aridity datasets and Global Soil Water Balance datasets were used to conduct the water vulnerability assessment (Table 3.1). The property characteristics of these data differ widely, so it is vital to harmonize data layers so that detailed spatial variables that appropriately constructing the indices of water vulnerability can be properly processed.

Table 3.1 Spatial variables and datasets used in water quantity vulnerability assessment

Variable	Scene ID	Dataset	Spatial Resolution	Temporal Resolution
Administrative boundaries (all levels of sub-division)	GADM	Gadm36_VNM version 3.6		2018
Elevation	ASTERGDEMV003	ASTER GDEM v3	30m	2000
Precipitation	CHIRPS-v2.0	Climate Hazards Group InfraRed Precipitation	4.8-km grid (1/20 degree)	Yearly, 1981-2018
Net Evapo-transpiration	MOD16A3GF.006	Terra MODIS	500m	Yearly, 2000-2019

Soil Water Stress	SWC_fr	Global Soil-Water Balance datasets	1000m	Montly, 1970-2000
Priestley–Taylor alpha coefficient	alpha	Global Soil-Water Balance datasets	1000m	Yearly, 1970-2000
Aridity Index	Ai_et0	Global-Aridity datasets	1000m,	Yearly, 1970-2000
Leaf Area Index	MOD15A2H.006;	Terra MODIS	500m	8 day, 2000-2019
Net Primary Production	MOD17A3HGF.006	Terra MODIS	500m	yearly, 2000-2019

Satellite images were generated from S-2 multispectral satellite data, S-1 C-band dual polarimetric SAR imagery, and ALOS DSM (table 3.2) for soil carbon and soil moisture estimation. While Sentinel 1 and Sentinel 2 images were generated from the Copernicus Open Access Hub from European Space Agency (ESA), the ALOS DSM 30m imagery was acquired from JAXA Earth Observation Research Centre. The SNAP Sentinel Application Platform toolbox were used for both Sentinel datasets processing, whereas ArcGIS 10.4 was employed to process ALOS imagery and compute the ALOS-DSM derived features. All images were resampled to a ground sampling distance (GSD) of 10 m and geocoded in the same projection of World Geodetic System (WGS84) - Universal Transverse Mercator (UTM) zone 50 South (50S).

Table 3.2 Remote sensing data acquisition for the study areas

Sensor	Scene / Tile ID	Acquisition date (month/day/year)	Processing level	Spatial resolution (m)	Spectral band/ polarization
S-2	50JML	04/17/2021	1C	10 – 20	13 multispectral bands
S-1	S-1B_IW_GRDH1SDV	04/27/2021	GRD	10	Dual-polarization (VV and VH)
ALOS-DSM	AW3D30	04/01/2021		30m	

Source: European Space Agency ESA, 2021 and JAXA Earth Observation Research Centre

3.1.2 Statistical data

The study also used data from statistical sources and other documents. Administration unit, land use types, education, socioeconomic conditions and water supply systems were collected from Vietnam's national statistical yearbook for 2018. Water damage data, economic loss and flood risk index were recently provided by Luu et al. (2019) and the Central Committee for Flood and Storm Control (CCFSC) of Vietnam.

3.1.3 Field survey data

Soil samples were collected the Wests, Goomalling shire (latitude coordinate: -31°18'S and longitude coordinate: 116° 49' E), and Cookies area - Northam shire (latitude: -31° 39' S, and longitude: 116° 39' E) in the agricultural region of Western

Australia (WA).

3.2 Methods

3.2.1 Spatial data processing

ESRI ArcMap v10.4 was employed to process spatial data. Administrative boundaries from GADM served to clip or extract elevation, precipitation and other remote sensing data to the GADM country codes. ModelBuilder in ArcGIS 10.4 helped to convert spatial datasets to Excel datasets for further calculations. All spatial and statistical datasets were compiled according to the study site's provincial and ecological areas (Figure 3.1).

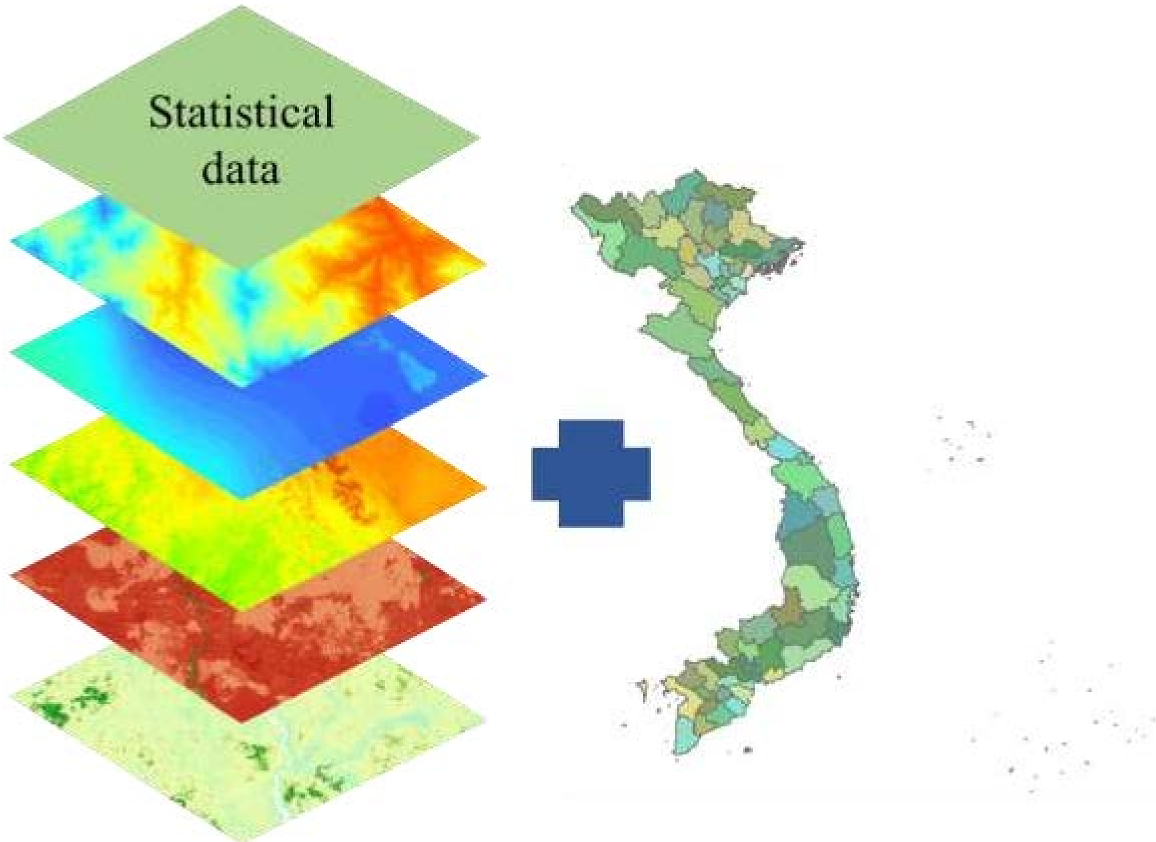


Figure 3.1 Incorporation of spatial datasets, statistical data set and the study site

3.2.1.1 Sentinel images processing

The S-2 image was processed via four main steps which presented in the Figure 3.3. Ten multispectral bands were extracted for the study including B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12. Vegetation indices, soil indices, and water index were computed by thematic land processing function in the SNAP toolbox (Pasqualotto et al., 2019). Vegetation, soil and water indicators are presented as being sensitive to soil moisture content which recently were used for soil moisture properties estimation (Jin et al., 2017). Predictor variables derived from S-2 were illustrated in table 2 below. A total of 22 indicators were computed from S-2 for the SM prediction.

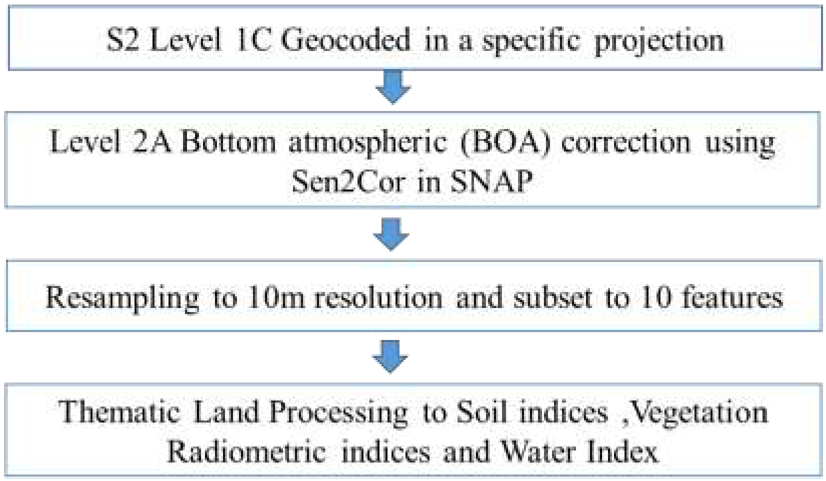


Figure 3.2 The steps of Sentinel images processing using SNAP Toolbox

Table 3.3 Vegetation, soil, and water predictor variables derived from Sentinel 2 (modified from (Pham et al., 2020))

Vegetation and Soil Index	Acronyms	S-2 band wavelengths	References
Ratio Vegetation Index	RVI	$\frac{\text{NIR}}{\text{Red}}$	(Tucker, 1979)
Normalized Difference Vegetation Index	NDVI	$\frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$	(Rouse et al., 1973)
Green Normalized Difference Vegetation Index	GNDVI	$\frac{\text{NIR} - \text{Green}}{\text{NIR} + \text{Green}}$	(Gitelson et al., 1996)
Normalized Difference Index using Bands 4 & 5 of S-2	NDI45	$\frac{\text{RE1} - \text{Red}}{\text{RE1} + \text{Red}}$	(Delegido et al., 2011)
Soil Adjusted Vegetation Index	SAVI	$(1 + L)(\frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red} + \text{L}})$ L = 0.5 in most conditions	(Huete, 1988)
Inverted Red-Edge Chlorophyll Index	IRECl	$\frac{\text{RE3} - \text{Red}}{\text{RE1}/\text{RE2}}$	(Frampton et al., 2013)
Modified Chlorophyll Absorption in Reflectance	MCARI	[(RE1 − Red) − 0.2 × (RE1 − Green)] × (RE1 − NIR)	(Daughtry et al., 2000)

Vegetation and Soil Index	Acronyms	S-2 band wavelengths	References
Index			
Brightness index	BI	$\sqrt{\frac{(\text{Red} \times \text{Red}) + (\text{Green} \times \text{Green})}{2}}$	(Escadafal, 1989)
Brightness index 2	BI2	$\sqrt{\frac{(\text{Red} \times \text{Red}) + (\text{Green} \times \text{Green}) + (\text{NIR x NIR})}{2}}$	(Escadafal, 1989)
Redness index	RI	$\frac{\text{Red} \times \text{Red}}{\text{Green} \times \text{Green} \times \text{Green}}$	(Mathieu et al., 1998)
Colour index	CI	$\frac{\text{Red} - \text{Green}}{\text{Red} + \text{Green}}$	(Mathieu et al., 1998)
Normalized difference water index	NDWI	$(\text{NIR} - \text{SWIR})/(\text{NIR} + \text{SWIR})$	(Gao, 1996)

Note: Band wavelengths of S-2: B2: Blue (492 nm), B3: Green (560 nm), B4: Red (665 nm), B5: Red-edge 1 (RE1) (704 nm), B6: Red-edge 2 (RE2) (740 nm), B7: Red-edge 3 (RE3) (783nm), B8: near-infrared (NIR) (833 nm), B8A: Narrow-NIR (865 nm), B11: short-wavelength infrared (SWIR1) (1614 nm), and B12: SWIR2 (2202 nm).

The extraction of Sentinel 1 data included eight steps which were conducted in the SNAP application using the Radar toolset to convert the S-1 C-band SAR raw intensity signal data to scale backscatter coefficient (σ0) in decibel (dB) as suggested by Pham et al.,(2020). The steps includes: (1) Correct the orbit file; (2) Thermal and border noise removal; (3) Radiometric calibration; (4) Speckle filtering; (5) Range Doppler terrain correction; (7) Normalized radar backscattering coefficient by the equation 1 below; (8) S-1 SAR band transformation to create five predictor features including VV/VH; VH/VV; VV-VH; VH-VV; (VV+VH)/2; and (9) computation of 20 features using grey level co-occurrence matrix (GLMC) from S-1 VV and VH Polarizations (Fig 3.3).

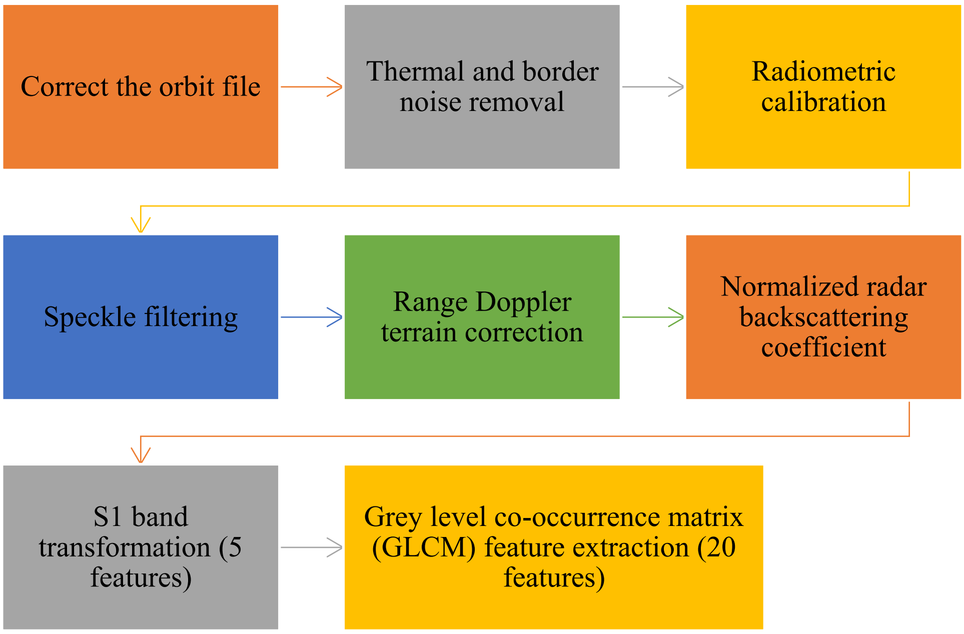


Figure 3.3 Steps of Sentinel 1 pre-processing and processing

A total of 27 features were extracted and computed from Sentinel 1. These features contained: the two bands from dual polarization (VH and VV); the five SAR transformed bands (VV/VH; VH/VV; VV-VH; VH-VV; (VV+VH)/2); and the 20 new features extracted from VV and VH using the GLMC algorithm (VV_Contrast, VV_Dissimilarity, VV_Homogeneity, VV_Angular Second Moment, VV_Energy, VV_Maximum Probability, VV_Entropy, VV_GLCM Mean, VV_GLCM Variance, VV_GLCM Correlation, VH_Contrast, VH_Dissimilarity, VH_Homogeneity, VH_Angular Second Moment, VH_Energy, VH_Maximum

Probability, VH_Entropy, VH_GLCM Mean, VH_GLCM Variance, and VH_GLCM Correlation).

3.2.1.2 ALOS image processing

The Advanced Land Observing Satellite (ALOS) was introduced by the Japan Aerospace Exploration Agency (JAXA) in 2006. JAXA recently provided the product of ALOS-DSM which is one of the newest remote sensing-based DEM. The ALOS-DSM has two kinds of resolution. ALOS-DSM with the resolution of 30m is a free-of-charge dataset and higher prediction performance compared to Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM) ASTER GDEM and Shuttle Radar Topography Mission Digital Elevation Model (SRTM-DEM) (Nikolakopoulos, 2020). DEM and SLOPE derived indicators were generated by raster processing and calculation in ArcGIS 10.4. Figure 3.4 shows the elevation of the study sites which ranges from 139 m to 480 m and slope is between 0 and 87 degree.

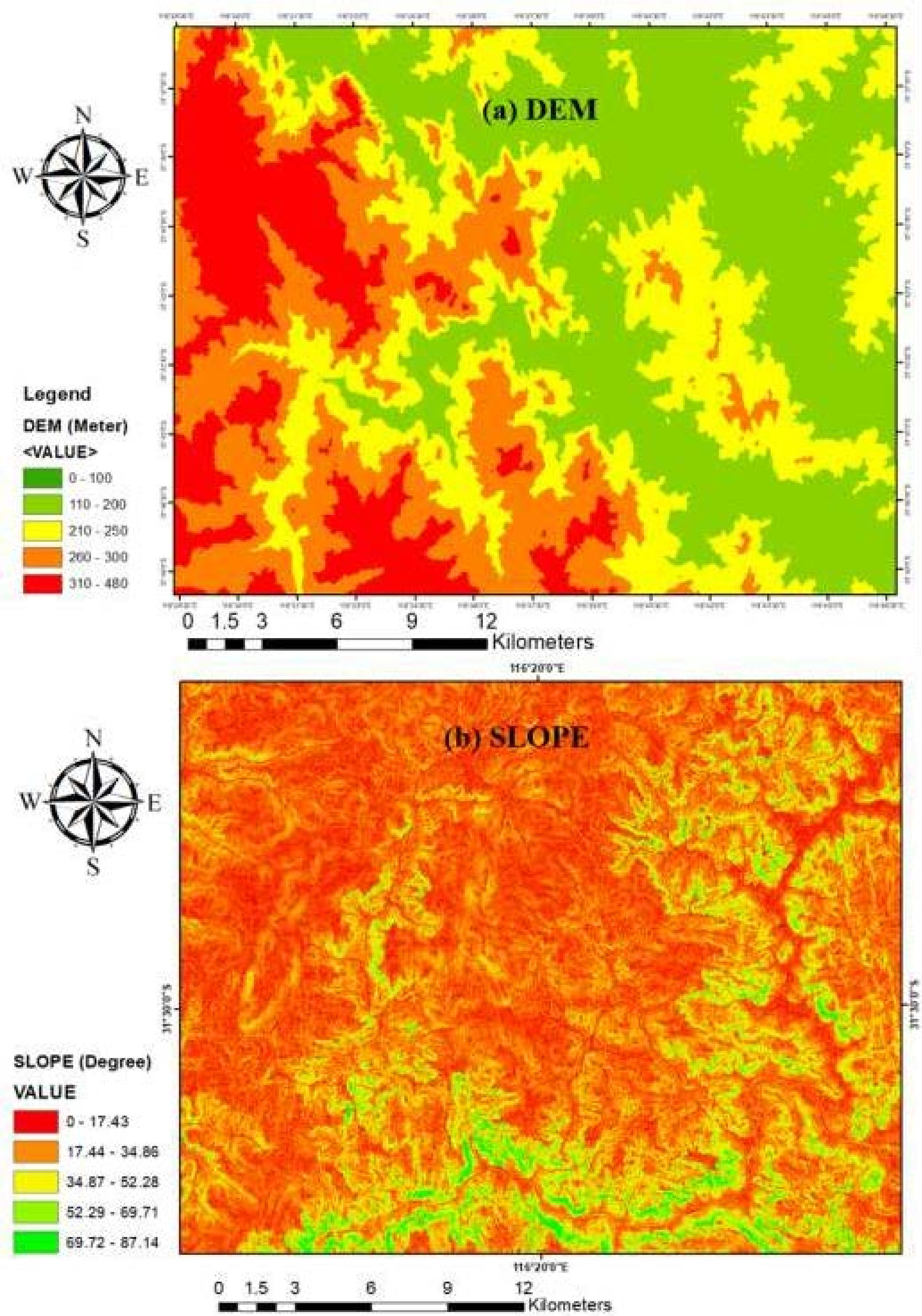


Figure 3.4 Indices generated from ALOS DSM: (a) DEM and (b) SLOPE

Topographic Wetness Index (TWI) generated from digital elevation model (DEM) have been used for soil moisture estimation because TWI is helpful to identify the place where water is accumulated in the specific area with the differences of elevation (Figure 3.5). TWI highlights the terrain-driven balance of the catchment water supply and the water drainage of specific local areas. However, there are various algorithms such as a flow accumulation, a flow width, or a slope algorithm can be employed to compute TWI. It should be select the best one that the TWI obtain the high correlation with soil moisture content. The best TWI for soil moisture prediction is Freeman flow algorithm, local slope, and the equal cell size of flow width which was generated by the following equation (Kopecký et al., 2021).

$$\text{TWI} = \ln \frac{Total\ catchment / _{Flow\ Width}}{\tan(slope)} \quad (1)$$

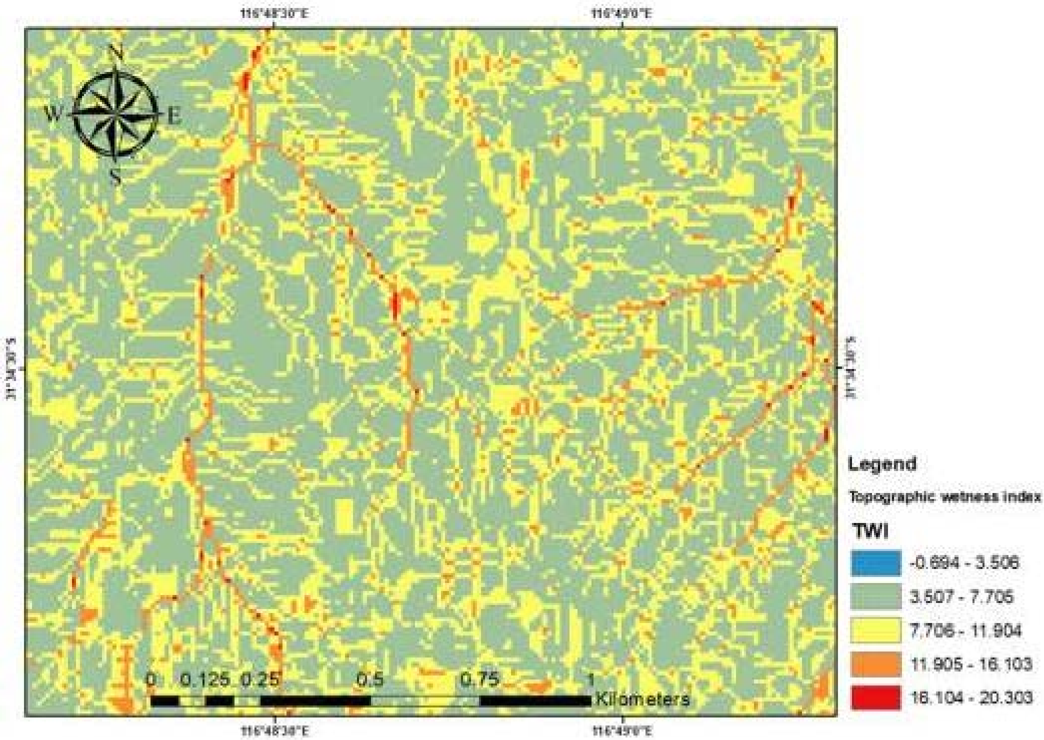


Figure 3.5 TWI mapping in the study site

3.2.2 Vulnerability assessment

One of the most popular concepts of vulnerability was presented in the 2001 IPCC report (Thornes, 2002). There are six steps in the vulnerability assessment process. The most important step is to determine appropriate indicators or variables through consultations with experts, conducting a literature review or a field survey. Next is the collection of input data for those indicators by a variety of sources such as field survey, statistical data and remote sensing datasets. The third step is normalization of input datasets via the formula involving the UNDP's Human Development Index (HDI) (Duong et al., 2017; McNicoll, 2007).

$$X_{ij} = \frac{X_{ij} - MinX_{ij}}{MaxX_{ij} - MinX_{ij}} \qquad (2)$$

Where X_{ij} represents for normalized score of the j indicator for the ith area.

The next one is to calculate the indicators' weight according to the Iyengar and Sudarshan method (Iyengar and Sudarshan, 1982) method as stated here:

$$W_j = \frac{C}{\sqrt{Var(X_{ij})}} \qquad (3)$$

$$C = \sum_{j=1}^{K} \frac{1}{\sqrt{Var(X_{ij})}} \qquad (4)$$

Where K is the number of indicators, C is a normalizing constant, and W_j is the weight of indicator j (0<1<W_j)

The index for each indicator for in each area was estimated with the calculated weight in the previous step via the following formula:

$$M_{ij} = W_J \times X_{ij} \qquad (4)$$

Where M_{ij} is the index of j factor in ith area

After establishing the indices of indicators, the index of exposure, sensitivity, adaptive capacity and water vulnerability are estimated using the equations below:

$$V_h = \frac{\sum_1^n M_{ij}}{n} \qquad (5)$$

$$WVI = \left(V_E + V_S + (1 - V_{AC})\right) / 3 \qquad (6)$$

Where V_h is the index of E, S and AC component (h=E, S, AC); n is the total number of indicators for each E, S and AC; and WVI represents the water vulnerability index.

All vulnerability indicators are integrated and visualized by ArcGIS 10.4 software. Vulnerability maps are constructed according to ecological and provincial scale.

There are five level of vulnerability including: very low, low, medium, high and very high.

3.2.3 Field data collection

From very high spatial resolution Google Earth imagery and Sentinel 2 imagery, a total of 266 digitizing points for both vegetation and bare soil locations were selected to generate land-use binary maps (Figure 3.6). An Advanced ML technique with five-fold cross validation (CV) method were applied for binary land-use classification mapping. The classification accuracy of the XGBoost model were compared with the two well-known ML algorithm such as the RF and SVM technique. The overall accuracy, kappa coefficient, precision, recall and F1_score served as evaluation metrics. The best model with the highest value of overall accuracy, F1 score and Kappa coefficient was chosen to produce the binary land-use map. The binary land use classification map devised in the study areas served to identify bare-soil points for soil sampling. The active learning technique in remote sensing classification was employed to assist in designing and sampling soil carbon, which helps minimise effects of vegetation on SOC contents (Fu et al., 2010; Tuia et al., 2011).

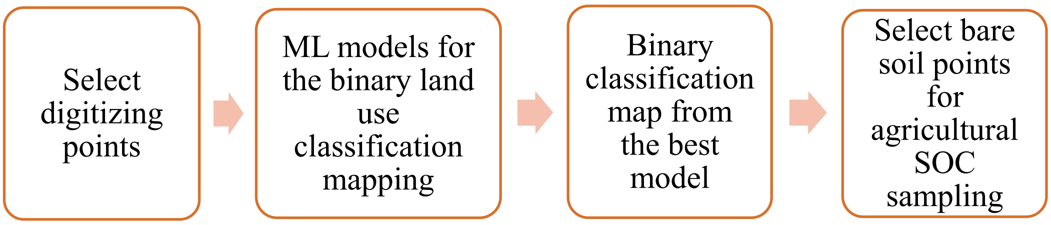
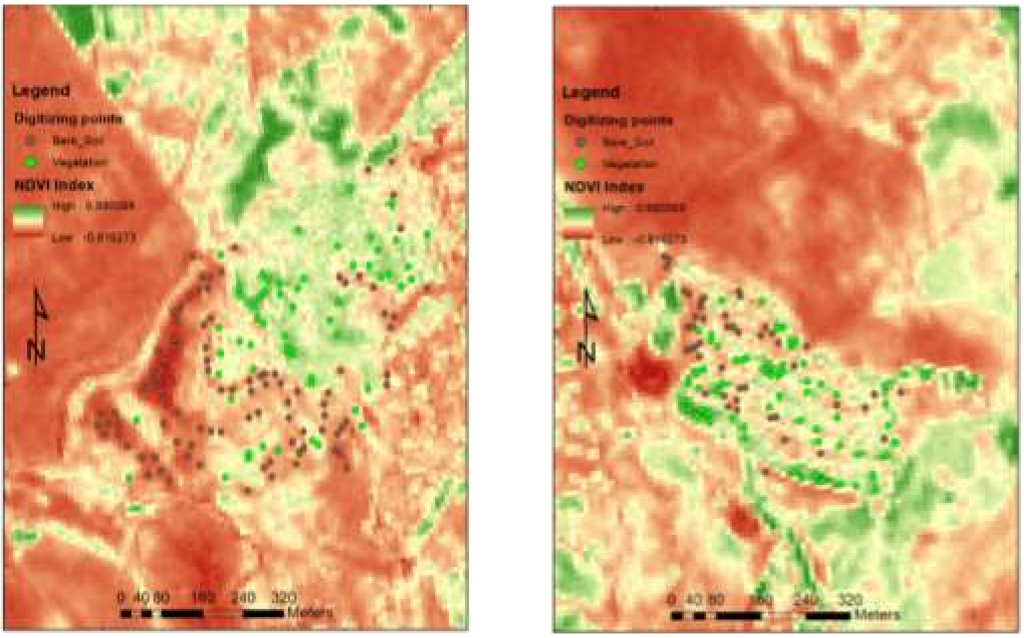


Figure 3.6 Flow chart of land-use binary mapping and SOC samples selection using an active learning method.

The agricultural SOC field survey was carried out in April 2021. Forty bare-soil sampling locations with a pixel (size of 10m x 10m) across the study areas (20 points for each area) were selected based on the binary map (Figure 3.7). A Differential Global Positioning System (DGPS) - a refined version of the Global Positioning System (GPS) - was used to identify precisely the samples' location with an accuracy of 1-3 cm (Michalski and Czajewski; 2004). Four soil cores were taken in each sampling plot. The dimensions of the core was 7 cm in depth and 7.3 cm in diameter. The total agricultural SOC of soil samples was analysed in the laboratory by Rayment and Lyons Method 6B1 (Heanes, 1984).



(a)Wests (b) Cookies

Figure 3.7 Study areas and digitizing point selection: (a) Wests, and (b) Cookies

3.2.4 Machine learning algorithms

3.2.4.1 Extreme gradient boosting (XGBoost)

The XGBoost technique was introduced by Chen and Guestrin (2016). It shares the same theory with other gradient tree boosting algorithms. The XGBoost algorithm is described as a scalable end-to-end tree boosting which is a highly accurate machine learning technique and has widely applied to solve data mining problems (Chen and Guestrin, 2016). The novelty of XGBoost is its scalability in all scenarios so it can handle sparse data challenges. This advanced ML techniques is able to handle both classification and regression tasks (Ha et al., 2021b). The further merits of the XGBoost are parallelization, out-of-core computation, and cache optimization, which help the training process of the system more quickly than existing gradient boosted regression tree methods. This technique can easily deal with the problem of a model’s complexity especially if it has a large dataset. Moreover, the XGBoost method can use integrated optimization algorithms to tune important hyper-parameters such as the number of trees and the rate of learning to suit a specific dataset. In this study, the best structure with 100 trees, and a learning rate set at 0.5 and gamma value of 5 was found the highest performance in the

XGBoost model.

3.2.4.2 Random forest (RF)

The RF algorithm is one of the most popular machine learning algorithms, and it can be used effectively for a wide range of applications (Breiman, 2001; Pham et al., 2020). This technique includes a large number of regression trees. Each regression tree is built by the unique bootstrap sample from the original dataset, which decreases the sensitivity of the RF method to overfitting problems. Normally, the dataset will be divided with about two-thirds of the samples (in-bag data) for the training sets and the remaining samples for the test sets (Out-Of-Bag (OBB data). Two essential parameters including the number of regression trees and number of predictor variables must be defined in the RF model. In the current work, the RF model with 100 trees and the maximum number of 11 features had the highest performance for this study area.

3.2.4.3 Support vector machine (SVM)

Developed by Cortes and Vapnik (1995), the SVM algorithm is a well-known supervised learning technique based on the kernel approach and statistical theory, which can applied for classification, regression and outliers' detection (Cortes & Vapnik, 1995; Cristianini & Ricci, 2008). While the SVM can help solve non-linear dataset, this method is not effective with a noisy and overlapped dataset. One of the advantages of SVM is that it can work accurately with a small number of training datasets. The SVM algorithm's performance is based on the selection of kernel functions and their parameters. There are three hyper-parameters in the SVM method including regularization parameter, the kernel function, and gamma controlling the overfitting. The hyper-parameters of the SVM method are fewer than other machine learning algorithms. Four kernel function types include polynomial, sigmoid, linear and radial basis function. In this study, the grid search with a five-fold CV was used to determine the optimal hyper-parameters of each ML algorithm in the Python environment. In this work, the SVM algorithm with the radial basis function (RBF) kernel and the C value of 10000 was used, and the epsilon value of 0.01 as the best values for tuning hyper-parameters of the SVM model.

3.2.4.4 CatBoost gradient boosting regression (CBR)

CBR is known as a family member of gradient boosted decision trees (GBDT's). It is an interdisciplinary approach for classification and regression tasks in time-series and big data (Hancock & Khoshgoftaar, 2020). It can also solve and minimize the issue of over-fitting by identifying the best tree structure for the calculation of the leaf values (Dorogush et al., 2018). CBR have recently been employed for soil parameters and soil carbon estimation (Xu et al.). Max depth, learning rate, and the number of iterations is the key hyper-parameters of the CBR model. It is similar to XGBR, important hyper-parameters were tuned by hyper-parameter tuning using grid search with five-fold CV to select optimal ones which helps improve the CBR model performance.

3.2.5 Genetic Algorithm (GA) optimizer for optimal feature selection

Features selection is vital for the ML model's performance. It also helps simplify the models, reduce the time for training and testing model, and address overfitting issues. A genetic algorithm method was employed to determine automatically optimal indicators for the SM content retrieval in the study from the total of 52 variables derived from selected RS missions. GA implementation includes the following stages: (1) population formation from soil samples; (2) generation of a mating pool based on the highest fitness individual values; (3) the selection of parents from the mating pool by random selection methods; and (4) the generation of parents' offspring using crossover and mutation operators. The prediction accuracy of ML models for soil properties can be improved with the use of the GA for the selection of predictor variables (Xie et al., 2015).

3.2.6 Model evaluation

To assess the model performance of binary land-use classification, five evaluation criteria have been used including overall accuracy (*OA*), kappa coefficient (*KC*), precision (*P*), Recall (*R*), and F1 score (*F1*) (Chicco & Jurman, 2020; Ha et al., 2021).

For agricultural SOC, and SM retrieval, two common validation criteria were employed to assess the performance of machine learning techniques with different scenarios including: the root mean square error (RMSE), and the coefficient of determination (R^2). Superior model performance illustrates the higher R^2 and lower RMSE. These criteria are evaluated using the equations below:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \quad (7)$$

$$R^2 = \frac{\sum_{i=1}^{n}(P_i - \overline{O_i})}{\sum_{i=1}^{n}(O_i - \overline{O_i})} \quad (8)$$

Where: n indicates the number of soil samples; P_i and O_i illustrate the predicted value and measured value of the i sample, respectively.

Chapter 4

A new integrated conceptual model for sustainable rural water infrastructures implementation

4.1 Introduction

The majority of integrated water models have focused on evaluating potential hydrological performance, and water quality assessment of water management solutions. However, in the meantime they have neglected their full assessment of environmental, economic, and social benefits such as better human health and economic growth (Zomorodian et al., 2018). Numerous countries have developed their own models to support their water management programs. The United States implemented many models including SWMM, MIKE, SUSTAIN, MapShed, SWAT, PondNet, etc., to support the Best Management Practices program. Some models were also devised in the United Kingdom such as UWOT, Sobek-Urban, and WaterMet2 to assist that nation's Sustainable Urban Drainage Systems. To improve the efficiency of its Water Sensitive Urban Design program, Australia has devised some effective integrated models including DAnCE4Water, Urban Bests, UrbanCycle, UrbanDeveloper, and Aquacycle. DAnCE4Water is able to assess a range of integrated planning, various urban water infrastructure systems, and in conjunction with the dynamics of various social systems (Löwe et al., 2017; Rauch et al., 2017; Zischg et al., 2019; Urich et al., 2013).

There are three general methods in which to develop integrated models: (1) modifying conventional integrated models; (2) combining existing sub-integrated models into more comprehensive as well as integrated ones; and (3) innovating new integrated models. Integrated water models are normally constructed by computationally linking a sequence of two or more sub-models that illustrate the different components of water bodies (Rauch et al., 2002). Each integrated model has developed its own distinct approaches and methods according to Bach et al. (2014). A variety of processes are involved in water management and these include: hydrology; hydraulics; pollution; treatment; downstream impact; storage-behaviour; water consumption; groundwater interaction and flooding; different water components, i.e., water transportation network, treatment plants, decentralized technologies, receiving water bodies and built environments. All these processes need to be taken into account when building an integrated model (Bach et al., 2014). Integrated modelling is based on several types of model applications including: life cycle assessment; operations and control; risk and impact assessment; social implications; economic issues; ecological implications; conceptual design; and strategic planning.

A range of papers illustrates the development, barriers, and opportunities of integrated water models. Bach et al. (2014) reviewed 30 years of research on the adoption of integrated water models and classified these models into four groups according to their level of integration. Their review paper did note that user-friendliness, administrative fragmentation, model complexity, and communication are crucial factors, which guide the uptake of integrated water models (Bach et al., 2014). Zomorodian et al. (2018) analyzed the feasibility of System Dynamics (SD) application on addressing the complexity of integrated water management modelling. Salvadore et al. (2015) compared 43 hydrological modelling approaches and identified a blueprint for future hydrological modelling development. The study ascertained that a high degree of uncertainty will be reduced if remote sensing data, measurement model parameters, and spatial calibration methods are applied.

This study aims to build a new integrated conceptual model framework to better address the multiple sustainability objectives of water infrastructure concepts including water vulnerability improvement, soil moisture enhancement, soil carbon sequestration of different water strategies, and agricultural practices.

4.2. Conceptual model principles

4.2.1 Selection of model's features

The features of integrated water models do vary as a consequence of the diverse requirements at each level of integration. There are six components that need to be considered as being essential to a specific model's features: data requirement and availability; computational power and software development; process methods; spatial and temporal detailing; simulation configuration; and model structure (Bach et al., 2014). The key model features of the rural water management model are summarized in Table 4.1.

Table 4.1 Key model features of the rural water model

Category	Model features selection	Challenges
Data requirement and availability	Both qualitative and quantitative data, both spatial and temporal data	Difficult to collect all types of data; data uncertainty problems
Model structure	Conceptual	Model computational burden due to high level of integration
Simulation configuration	Both parallel and sequential	The inaccuracy of each sub-models affect the overall result of model
Spatial detailing	Both branched and looped	Basing on considered interactions in models
Temporal detailing	Continuous simulation and uniform time step	Huge data requirement and challenges in collection the historical data.
Process nature	Quantity: Hydrology performances Quality: Both biological and physical	Hydrology/hydraulic systems is very complex and different between watersheds.
Computational power	Multi-core processing, optimisation and scenario analysis	Uncertainties in model parameters, the big size of integration model and doubtful mathematical
Software development	Supermodel, interface and hybrid	

Category	Model features selection	Challenges
		formulation of processes Model complexity

Problems concerning the features of this model can range from data uncertainty issues to doubtful mathematical formulation of processes given that these complex integrated models are so complex and could be prone to error. Challenges in model structure and the nature of processes (hydrodynamic, biological and physical) should be given priority in any rural water model's development. Data collection and data reliability are fundamental variables that encompass the requirements for building, testing and calibration (Bach et al., 2014; Deletic et al., 2012; Elliott and Trowsdale, 2007; Nguyen et al., 2007).

4.2.2 Selection of relevant sub-models

A developed integrated rural water model should consider incorporating three models: water vulnerability assessment; soil moisture prediction; and carbon sequestration estimation. The excessive, wasteful and inappropriate use of water resources has increased dramatically throughout the world. According to the Global International Geosphere-Biosphere Programme (IGB), total water global freshwater withdrawals amounted to 4 trillion m^3 in 2014, representing a six-fold increase between 1900 to 2014 (Alcamo et al., 2003; aus der Beek et al., 2010; Flörke et al., 2013). Water use per capita throughout the world varies greatly depending on the latitude, climate, and level of countries' or regions' economic development. Appropriate evaluation of water vulnerability is vital if we are to understand the impacts of climate change and human activities on water resources. As well, the information from the water vulnerability assessment projects can support managers and decision-makers to select suitable areas for water infrastructure implementation. Therefore, the important sub-component of a proposed conceptual model for sustainable rural water management should be the water vulnerability assessment model.

Linking these sub-models should help support water management systems to interact effectively. The in-built key innovation is created by the integration and simplification of these models and their features. The proposed sustainable rural water model below is better able to predict actual effectiveness of water management practices in reducing water vulnerability, improving soil organic carbon, and enhancing soil moisture content (Figure 4.1). Moreover, it provides valuable data for different stakeholders including the following: water systems managers, local authorities, and landholders. They need to be able to implement appropriate water management strategies and practice precision agriculture, which would increase revenue for landholders via carbon credits and crop production. The last point could see a marked improvement in climate, soil and air quality, as well as better hydrological performance through water runoff and flooding disaster reduction. Appropriate rural water strategies are selected through the harmonization of economic, and environmental feasibility of rural water infrastructure. These desirable attributes of the proposed rural water management model will require an effective support system for managing water-related problems, where decisions are based on continuous cooperation (consensus) between modellers and policy-makers throughout water infrastructure development processes (Liu et al., 2008; Makropoulos et al., 2008). Nonetheless, ensuring strong communication and collaboration with stakeholders will help support and build an integrated rural water simulation model.

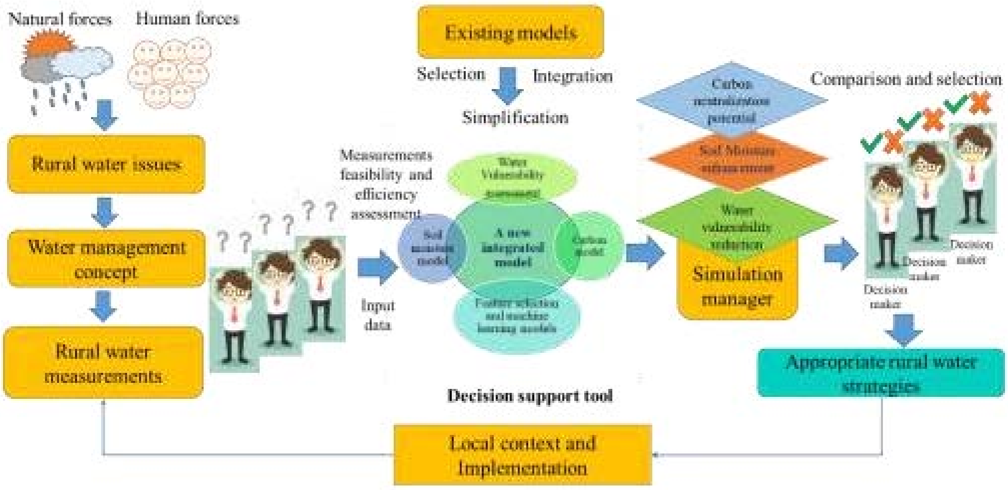


Figure 4.1 Overall development proposal of the sustainable rural water management model

The most difficult challenge for devising a model is to answer the question of how to incorporate sub-models into a complete one for testing and application. Therefore, the main methods used in the newly integrated models could be: (i) the construction of an integrated database system; (ii) identifying the integration of multiple data sources, geo-spatial data computing; and (iii) design and visualization methods (Zhang et al., 201; Xu and Yu, 2017; Deng et al., 2018). Real-time monitoring and warning systems can be built based on modern technologies, for instance the Internet of Things, Big Data, and Cloud Computing (Deng et al., 2018).

4.2.3 Selection of model variables

Each sub-model requires a range of variables. For example, the water vulnerability model database system includes data on population and technologies scenarios, social characterization, surface characteristics, and climate information data. The inaccuracy of input data results in model uncertainties and subsequently serious errors (Deletic et al., 2012; Bach et al., 2018, Wijesiri et al., 2016). An adequate and reliable database system is hugely important for any kind of integrated water management model including the proposed integrated model. The capabilities of the variables regarding the integrated model according the framework in Fig. 1 can be divided into five components: (1) climate indicators, for example the history of daily precipitation, temperature, and solar radiation; (2) social indicators like population, living standards, poverty rate, economic loss; (3) environmental indicators such as soil moisture, soil properties, and carbon sequestration; (4) water indicators such as water demand, flood index, normalized difference water index, and aridity index; and (5) geo-database including soil and elevation map, surface coverage, land use map, administration map, vegetation indices, and soil indices.

4.3 Validation and calibration process of integrated model

The first step in generating the integrated model is to develop the conceptual basis based on scenarios about rural water management practices, climate, and development plans where data are available. The conceptual model will apply mathematical or computational methods to solve the problems noted above. The second step is to implement the integrated simulation model as a computer-based project. This integrated simulation model will predict the feasibility of relevant water practices in terms of addressing rural water issues, and maximizing water ecosystem services according to different scenarios. Here the validation and calibration steps are emphasized and they are the most complex tasks to perfect (Bach et al., 2014; Deletic et al., 2012; Rauch et al., 2002). The model validation comprises three dimensions, these being: conceptual validation; operational validation; and data validation (Sargent, 1991) (Fig. 4.2). Conceptual validity seeks to verify the accuracy of theories and assumptions in the conceptual model. Operational validity aims to ascertain the reliability of the model's output (Sargent, 1991). Data validity aims to ensure quantitative data is available for model testing, model calibration, and model simulation. Model calibration is defined as ensuring the accuracy of the model's output by comparing its output with actual data from measurement, observation, and collection.

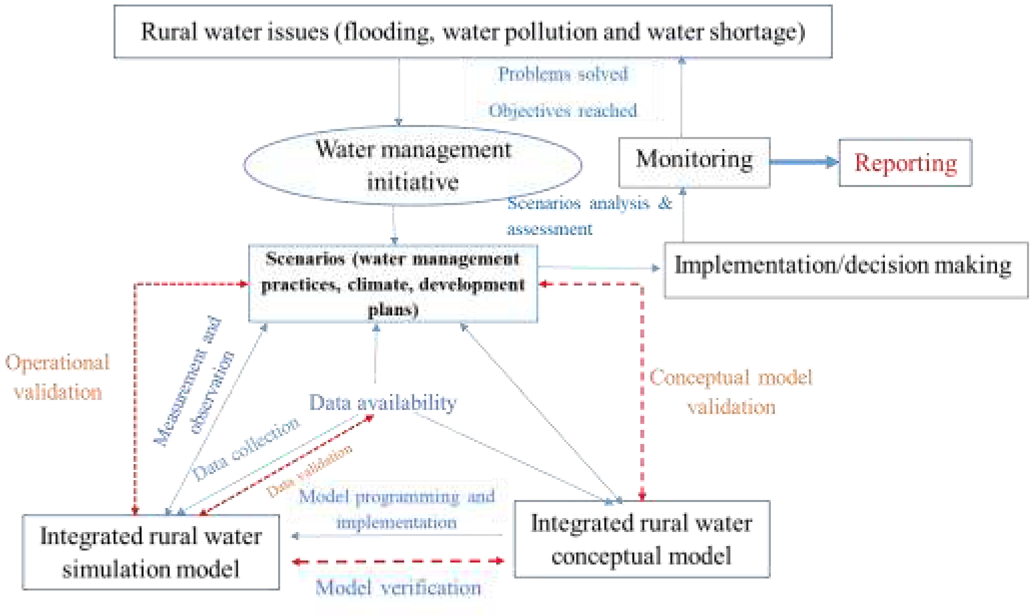


Figure 4.2 A generic framework of the integrated rural water model development (modified from Sargent, 1991).

4.4 Mapping model uncertainties

Some models or software packages are still used although uncertainties will exist amongst these (Schellart et al., 2010). For example, regarding the output of such models, it may not always be reliable. Although it is vital to solve the sources and outcomes of uncertainty in drainage models, not much research addresses these sorts of problems in rural water management modelling (Harremoës & Madsen, 1999; Schellart et al., 2010). Advanced water management models need to identify uncertainties and challenges from different perspectives. Uncertainties in hydrological models are caused by input data, model parameters, and model structure uncertainties (Guzman et al., 2015). The models without sufficient uncertainties being properly accounted for have led to incorrect simulation and this then has serious implications for strategic management and planning where errors can occur (Mannina and Viviani, 2010; Voinov and Shugart, 2013; Ahyerre et al., 1998). Accurate model predictions might be achieved if the measured and observed data are enhanced (Dotto et al., 2011). There are three main components causing model uncertainties and these are: model input uncertainties; model structure uncertainties; and model calibration uncertainties (Deletic et al., 2012) (Fig. 4.3).

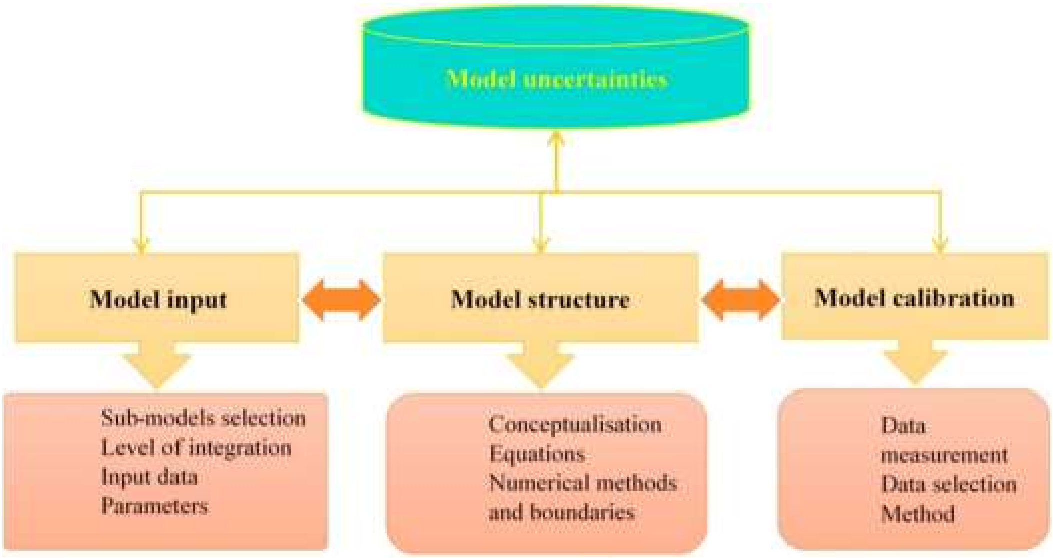


Figure 4.3 The main components of uncertainties in the integrated rural water model (modified from Deletic et al., 2012).

Firstly, the highest level of integration in the rural water model requires a vast temporal and spatial scale of data, and where components of sub-models are intricately linked, as these may result in model input uncertainties. Secondly, the uncertainty comes from the model structure itself where any inaccuracies in the scale and selection of key processes selected could lead to conceptualization errors being inherent within these models. The application of wrong equations or numerical methods and boundary conditions creates inaccurate outputs and solutions. Finally, calibration uncertainty represents another source of uncertainty. It is caused by the errors in measurement or monitoring of both input and output data, and this affects the selection of variables for the calibration process and related methods (Deletic et al., 2012; Dotto et al., 2014; Seppelt et al., 2009). To evaluate overall uncertainty, the Generalised Likelihood Uncertainty Estimation (GLUE) method developed by Beven and Binley (1992) has been applied in many studies (Freni et al., 2009; Mannina et al., 2006; Thorndahl et al., 2008). Attempts to mitigate a model's uncertainties include: model simplification; detail reduction; maximum application of computational or equation resources; and selection of appropriate calibration and validation methods (Jamali et al., 2018).

4.5 The proposed model's structure

The new model is built for multiple stakeholders who include: researchers; engineers; policy-makers; and other practitioners. It will assist them in the provision of appropriate decisions for rural water infrastructure implementation. The model is able to assess the efficiency of water infrastructure including LID practices, and green infrastructure practices. Environmental aspects of these practices will be assessed. The model is developed based on the structure language of Python, GIS and R with the idea of simplification, and integration. There are five layers in the integrated model and they are: sub-model layer; input layer; module layer; output layer; and programming language layer which is illustrated in Figure 4.4. Each component of the system is designed throughout the identification of each component's function. Then, the corresponding modules are formed based on their functions. This establishes the simulation process for the whole system.

Key aspects that need to be considered for sustainable model development include: database construction; data integration; temporal and spatial data computing; and the visualization process (Deng et al., 2018). The post-implementation monitoring system is made possible by cooperation between the Internet of Things, Big Data, Cloud Computing technologies, and the related standards associated with water infrastructure (Deng et al., 2018).

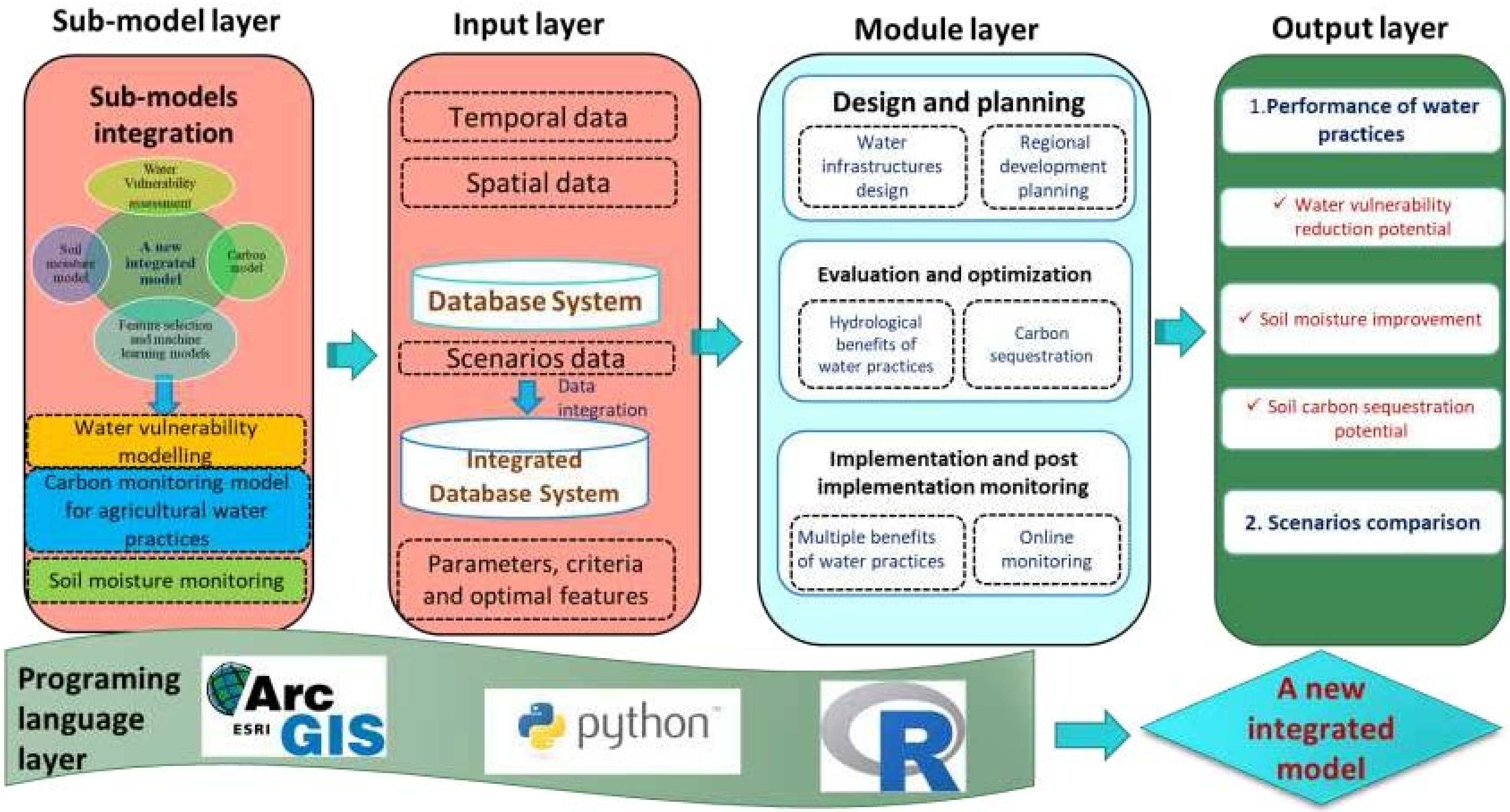


Figure 4.4 The possible structure of integrated rural water model

4.5 Conclusions

The uptake of integrated models to deal with environmental problems has increased in recent times. Although, many models have been developed recently, and are now incorporated into wider ranging water management practices to address such water issues, at present there is a lack of an effective overarching model that supports rural water infrastructure implementation. This is due to its multiple objectives and complexity. Therefore, it will be inevitable to essentially build the sustainable rural model based on the current conventional integrated sub low-cost models including ML models and feature selection models. This chapter critically highlights the importance of the comprehensive rural water model where the combined format can properly evaluate the challenges that include the model's cost-effectiveness, ambiguous data, etc.

In this chapter, a novel framework for the rural sustainable water management model was identified to simulate the efficiency of water management practices. The framework of the integrated model developed here was based on integrating four important sub-models, i.e. soil moisture model, soil carbon model, water vulnerability assessment model, machine learning model, and feature selection model. This model not only predicts the multi-benefits of rural water management measures in terms of water vulnerability improvement, but includes soil moisture enhancement and the increase of soil organic carbon for agricultural prediction. With these in place it can identify the most appropriate rural water management strategy.

Applying the integrated model requires different scenarios being taken into account and a large amount of required data. In this way, it will demonstrate that water management practices, development patterns and the dynamics of natural processes including hydrological systems are working properly or where changes need to be made. Uncertainties associated with this model can be overcome through the improvement of spatial data-sharing systems, the implementation of efficient computation and software design, and inter-disciplinary work. Finally, the novel model framework described in this research will assist modellers to develop a comprehensive sustainable rural water model for future applications. The model development specifically focuses on assisting the sustainable rural water program, but in the future, it should be broadened in scope for catchment areas.

Chapter 5

New approach of water quantity vulnerability assessment using satellite images and GIS-based model

5.1 Introduction

Water is a vital resource for people and many industries including agricultural, industrial and domestic applications (Anandhi & Kannan, 2018; Vorosmarty et al., 2010). It helps to sustain ecosystems but it causes disasters like floods or droughts for human communities (Brown et al., 2015). Water stress levels vary greatly in the world's regions and countries. The Middle East and North Africa regions have experienced extremely high rates of water stress when their freshwater withdrawals are greater than 80% (Ritchie and Roser, 2020). Several countries throughout South Asia and East Asia are experiencing medium to high levels of water stress (Ritchie and Roser, 2020). Nearly 80% of people on our planet have suffered high threats regarding water security (Vörösmarty et al., 2010). Water security has been influenced by abiotic factors like climate, and anthropogenic factors such as population and changes in land cover. Their relationships are explored when a water vulnerability assessment is conducted (Plummer et al., 2012). Identifying appropriately the list of indicators with sufficient input data is crucial and these contribute to a proper vulnerability assessment. Figure 5.1 illustrates a summary of water vulnerability components and sub-components. There are five main components, namely water resources, physical environment, economic, social and institution with several sub-components which are also depicted in this figure. Those sub-components or indicators are then categorized into three vulnerability assessment components (Exposure, Sensitivity and Adaptive Capacity). In general, the assessment of vulnerability is a very complicated process due to the multi-disciplinary nature of the problem, lack of knowledge and understanding of vulnerability theoretical frameworks and input data for required indicators related problems (Anandhi and Kannan, 2018; Gain et al., 2012).

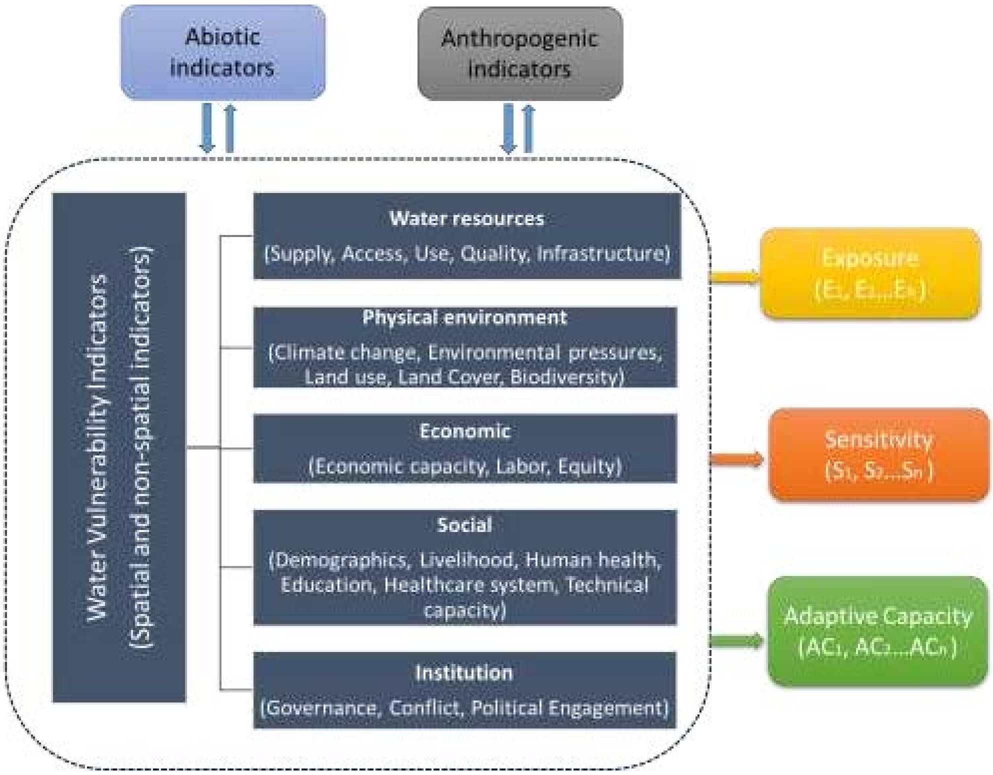


Figure 5.1 List of potential water vulnerability indicators (modified from (Plummer et al., 2012))

Previous studies have applied econometric methods by collecting information from surveys and questionnaires or index-based methods. These are derived from indicators and quantitative analyses of water vulnerability assessments (Bär et al., 2015). Indicators are identified by systems thinking approaches developed by experts in working in the water sector. The variables of vulnerability assessment can be selected through the Driver – Pressure – State - Impact – Response (DPSIR) framework (Jun et al., 2011). Satellite remote sensing datasets like MODIS, Sentinel or Landsat have been utilized to monitor water resources and to acquire input data for water assessment over the world (Khosravi et al., 2018; Sheffield et al., 2018).

A variety of methods and frameworks for vulnerability assessment have been proposed and applied in many studies. For example, the DRASTIC model and Catastrophe Theory have been used to assess groundwater vulnerability (Khosravi et al., 2018; Sadeghfam et al., 2016). However, there are several challenges associated with these methods: (1) not enough variables for water vulnerability assessment, 2) the unavailability and inaccurately input data for indicators, (3) limitations in applying spatial and temporal data from satellite images, (4) the practical applications of the framework, and (5) tools being limited in supporting water managers with their required water planning and management needs. This study aims to address these issues. Specifically, the research attempts to build a new spatial approach framework for water quantity vulnerability assessment based on mainly time series remote sensing data. Those spatial data were integrated with statistical data for water quantity vulnerability assessment. The novelties of the study are: (1) applying updated spatial-temporal satellite images from reliable datasets as important vulnerability indicators such as elevation from ASTER GDEM version 3, leaf area index and net primary production from MODIS datasets, and soil water stress from Consortium for Spatial Information; (2) utilizing GIS-based model for assessment; (3) incorporating different satellite datasets and statistical datasets in the ArcGIS 10.4 platform to construct spatial distribution of water vulnerability across ecological and provincial contexts. Overall, the evaluation of water quantity vulnerability is of vital importance for water managers in making the best decisions that improve the sustainability of water resource withdrawals.

Major parts of this chapter were published in a peer-reviewed journal (A-rated journal):

Nguyen, T.T., Ngo, H.H., Guo, W., Nguyen, H.Q., Luu, C., Dang, K.B., Liu, Y., Zhang, X. 2020a. New approach of water quantity vulnerability assessment using satellite images and GIS-based model: An application to a case study in Vietnam. Science of The Total Environment, 737, 139784 (IF: 7.963; SJR:

5.2 Materials and methods

5.2.1 Study area

The study area is the country of Vietnam. Vietnam's climate is strongly influenced by a monsoon-influenced tropical system with average temperature, precipitation and humidity ranging from 22-27^{o}C, 1500-3300 mm, and 70% - 85%, respectively. Based on the similarity of geographical and climatic conditions, Vietnam comprises eight ecological zones and these are the Northeast (NE), Northwest (NW), Red River Delta (RRD), South Central Coast (SCC), North Central Coast (NCC), Central Highland (CH), South West (SW) and Mekong Delta (MD) (Figure 5.2). Water sources in Vietnam mainly originate from its river basin system. There are 2,360 rivers in Vietnam with a length greater than 10km. Red River, Mekong River and Dong Nai River are the three main river watersheds where about 65% of the country's population living along these rivers (Le Luu, 2019). Water resources management in Vietnam has historically focused on freshwater conservation for agricultural production for hundreds of years. According to the data of FAOSTAT, Vietnam is one of the world's main agricultural water users with around 77.75 billion m^3 per year and, importantly, water stress reached a medium-to-high level in 2007 (Ritchie and Roser, 2020). In recent years there has been no data for water stress in Vietnam yet.

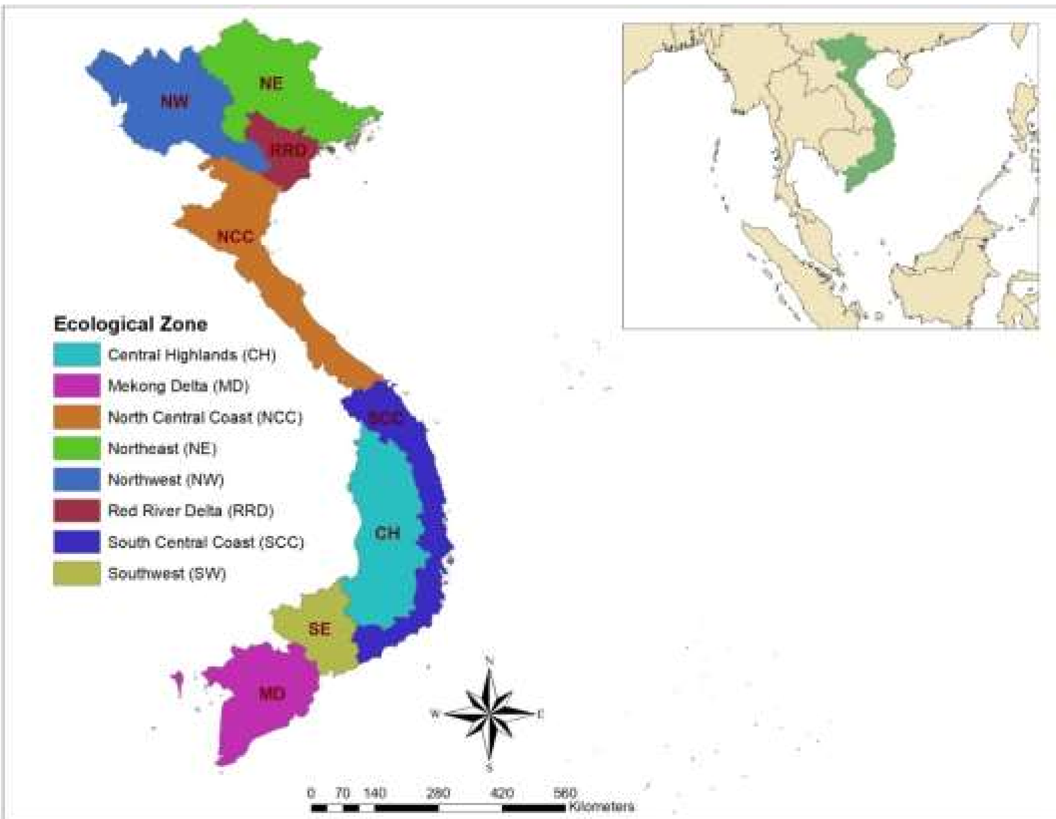


Figure 5.2 Location map of the study area

Since the late 1980s, Vietnam has experienced many water-related problems like water pollution, flood disasters and water shortages which have compromised the economic transition process and highlighted the dangers posed by climate change (Dang et al., 2019; Ngo et al., 2018; Norrman et al., 2008). A comprehensive understanding of water vulnerability not only minimizes the future vulnerabilities but also reduces the damage caused by water disasters on fragile ecosystems. Specially, it contributes to implementing an effective integrated water management system for Vietnam which not only addresses water problems but also introduces other environmental benefits like recognition of climate change and being prepared for natural disasters.

5.2.2 Data acquisition

The spatial variables and statistical data used in water quantity vulnerability assessment are presented in Chapter 3, Section 3.1.1.

5.2.3 Data analysis

5.2.3.1 Images processing

Methods for images processing in this chapter is illustrated in Chapter 3, Section 3.2.1.

5.2.3.2 Vulnerability assessment

Vulnerability assessment method is described in Chapter 3, Section 3.2.2.

5.3 Results and Discussion

5.3.1 Water quantity vulnerability assessment framework

The new water vulnerability assessment framework is built based on the general vulnerability assessment of IPCC, while the quantitative vulnerability assessment is devised by applying the index calculation method, normalization method, and weight evaluation method devised by Iyengar and Sudanrshan (Duong et al., 2017). The availability and accurately of input data for identified indicators is an integral part of this framework. The calculated results will help researchers in water resource management strategies to develop models for mitigating or adapting to climate change in the context of population growth and the drivers of economic development.

The water quantity vulnerability assessment framework is divided into three main layers (Figure 5.3): a data collection layer, a calculation layer and an output layer. The data collection layer will provide input data for the calculation layer. The major sources including satellite data, national or regional statistical data and information from experts, journals and regional documents will be used to collect data for their variables. Collected input data are stored in a database system, while statistical data can be displayed in Microsoft Excel and spatial data is processed by ArcGIS 10.4.

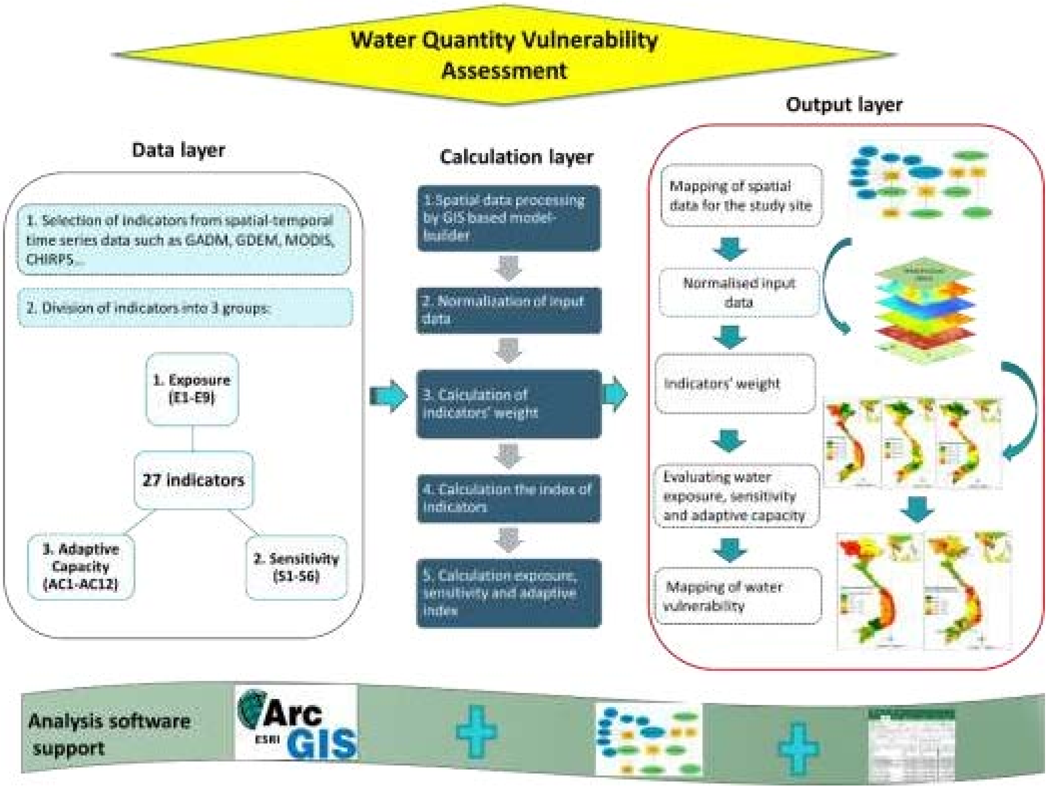


Figure 5.3 The framework of water vulnerability assessment

The calculation layer comprises five steps. The first step is to process satellite images by a GIS-based ModelBuilder function. The other steps including normalization of input data, calculating the weight of these data and the calculation of components and indications follow the vulnerability assessment method. The output layer will display results of the water vulnerability assessment in the form of tables, maps and graphs; this entails integration with Excel program and ArcGIS

10.4. This module is able to demonstrate the calculated outputs of exposure, sensitivity and adaptive indices for each province or each ecological zone. It also makes it possible to present calculated water quantity vulnerability index (WVI) of the study site for specific time periods. The results can be illustrated by maps so that policy-makers or local authorities can easily identify the degree of water vulnerability for their regions in order to implement the best strategies for water management.

5.3.2 Selection of water vulnerability indicators

Indicators collected are divided into three groups, i.e. Exposure (E), Sensitivity (S) and Adaptive Capacity (AC) (Table 5.1). A greater number of indicators which can be acquired will create more appropriate results for understanding vulnerability assessment. Through a literature review and the limitation of data at the study site, the total number of variables is 27 indicators including 9 for exposure components, 6 for sensitivity components and 12 for adaptive components. These indicators originate from many sources. Spatial data are collected from satellite datasets such as MODIS images, ASTER GDEM and Consortium for Spatial Information. Other data are derived from national and provincial statistical yearbooks and relevant journal papers.

Table 5.1 The selected indicators for water vulnerability assessment

Component	Indicator	Code	Unit	Period	Data Source
Exposure	Evapotranspiration	E1	mm	1981-2018	Trabucco and J.Zomer, 2018
	Annual rainfall	E2	mm	1981-2018	Climate Hazards Group

Component	Indicator	Code	Unit	Period	Data Source
					InfraRed Precipitation (CHIRPS)
	Aridity index	E3		1970-2000	Consortium for Spatial Information
	Flood index	E4		1989-2015	(Luu, von Meding, & Mojtahedi, 2019)
	Elevation	E5	m	2016	ASTER GDEM version 3
	Priestley–Taylor alpha coefficient	E6		1970-2000	Consortium for Spatial Information
	Impervious surface ratio	E7	%	2000-2018	Vietnam statistical yearbook
	Population density	E8	person/km^2	2000-2019	Vietnam statistical yearbook

Component	Indicator	Code	Unit	Period	Data Source
	Population growth rate	E9	%	2000-2020	Vietnam statistical yearbook
Sensitivity	Irrigation -Eroded earth, rock	S-1	m^3	1989-2015	(Luu et al., 2019)
	Economic loss	S2	Million VND	1989-2015	(Luu et al., 2019)
	Soil Water Stress	S3		1970-2000	Consortium for Spatial Information
	Agricultural production land	S4	km^2	2000-2018	Vietnam statistical yearbook
	Female	S5	Thous-and people	2000-2018	Vietnam statistical yearbook
	Poverty rate	S6	%	2000-2018	Vietnam statistical yearbook

Component	Indicator	Code	Unit	Period	Data Source
Adaptive Capacity	Water Use Efficiency	AC1	g/Cm^2 mm	2000-2019	Calculated from NPP and evapotranspiration
	Rain Use Efficiency	AC2	g/Cm^2 mm	2000-2019	Calculated from NPP and precipitation
	Leaf area index (LAI)	AC3	m^2/m^2	2000-2019	MODIS data (MOD15A2)
	River density	AC4	m/km^2	2010	Hanoi University of Science
	Road density	AC5	m/km^2	2010	Hanoi University of Science
	Pervious surface	AC6	%	2000-2018	Vietnam statistical yearbook

Component	Indicator	Code	Unit	Period	Data Source
	Percentage of trained employed workers at 15 years of age and above	AC7	%	2000-2018	Vietnam statistical yearbook
	Total foreign direct investment until 2018	AC8	Mil.USD	2000-2018	Vietnam statistical yearbook
	Number of health establishments	AC9	Establish-ments	2000-2018	Vietnam statistical yearbook
	Percentage of urban population provided with clean water by centralized water supply system	AC10	%	2000-2018	Vietnam statistical yearbook
	Percentage of household having hygienic water	AC11	%	2000-2018	Vietnam statistical yearbook
	Annual income	AC12	Thousand VND	2000-2018	Vietnam statistical yearbook

The water vulnerability assessment was determined by combining important spatial indicators that include the evapotranspiration, annual precipitation, aridity index, soil water stress, Priestley–Taylor alpha coefficient, leaf area index, water use efficiency and rain use efficiency. Annual precipitation data was collected from Climate Hazards Group Infrared Precipitation with Station data (CHIRPS) for the years 1981-2018 when floods and droughts were studied (Isundwa and Mourad, 2019). Aridity index (AI) was evaluated from mean annual precipitation and mean annual reference evapo-transpiration (Trabucco and J. Zomer, 2018). It can be acquired from the Consortium for Spatial Information. Higher aridity index indicates less aridity and this data can be applied for research on environmental conservation, sustainable water development and climate change projects.

The annual soil water stress was estimated by the average of monthly soil water stress. Soil stress coefficient (*Ksoil*) represents soil water stress (SWS) which was calculated by the ratio of monthly soil water content (*SWCm*) and the maximum amount of soil water content for evapotranspiration process (*SWCmax*) according to this formulation (Trabucco, 2010):

$$Ksoil = SWCm / SWCmax$$

$$SWS = Ksoil * 100$$

Leaf Area Index (LAI) is defined as the leaf occupied area in a unit of land (Fang and Liang, 2014). The research used the annual LAI values in standard deviation of the leaf area index (LAI) for the 19 years from February 2000 to December 2019, which processed the MODIS images – MOD15A2 version 6 product. This dataset is the acquisition of the Terra sensor in an 8-day composite dataset with a 500m resolution. The version 6 product is of superior quality compared to other versions of 1000m resolution. MODIS land products were validated by the MODIS Land Team and Earth Observing System Validation Program Office (Justice et al., 2002; Morisette et al., 2002).

Water Use Efficiency (WUE) is presented as the amount of biomass produced (gram of carbon mass per m^2) per mm of water used by crops (Hatfield and Dold, 2019). Annual water use efficiency is calculated as the ratio of net primary production (g $C/m^2/mm$) per amount of water loss which were defined by units of annual evapotranspiration (mm). Rain Use Efficiency (RUE) is identified by the ratio of the net primary production and amount of annual precipitation (Dardel et al., 2014). Net primary production (NPP) is identified as total amount the carbon which ecosystems receive through the photosynthetic reduction of CO_2 discounted for plant autotrophic respiration (Chapin and Eviner, 2007; Running et al., 2000). Photosynthesis is affected by droughts, floods and other types of extreme climate patterns (Zhang et al., 2017). NPP has the negative correlation with water disasters. So, the estimation of NPP plays an important role in predicting climate change and its impact on water issues that increasingly threaten the ecosystem. This paper used NPP as an indicator to assess water vulnerability index. NPP datasets were acquired from Terra MODIS 17A3H with a resolution of 500m for the study site covering the period 2000-2019.

5.3.3 Mapping of satellite data

Spatial data after collecting were processed in ArcGIS 10.4 using ModelBuilder function. Leaf Area Index, Precipitation or Net Primary Production are displayed as time series spatial raster image. ModelBuilder makes it possible to analyse these data through analytical procedures. The model created from this function can be transferred into the tool and then allowed to share and be applied in other studies that examined different regions. ModelBuilder is a visual programming tool that serves to build workflows with a sequence. Figure 5.4 presents the explicit modelling process of extracting precipitation from remote sensing datasets. Other spatial data are analysed with a similar process. The blue blocks are primary input data, the yellows ones are geo-processing tools and the green blocks are the results of one geo-processing tool and subjected to the input of another tool in the model. The model in Figure 5.4 consists of five steps each one using a processing tool: (1) evaluating the world's average precipitation from 1970 to 2018 by raster calculation tool or cell calculation tool; (2) extracting data for study area by extract by mask tool; (3) determining statistic values for each region or area by zonal statistic tool; (4) reporting the results on a table by zonal statistic; and (5) converting these data into an Excel spreadsheet. Extracted data from satellite images are analysed in the following steps using Excel and R programming.

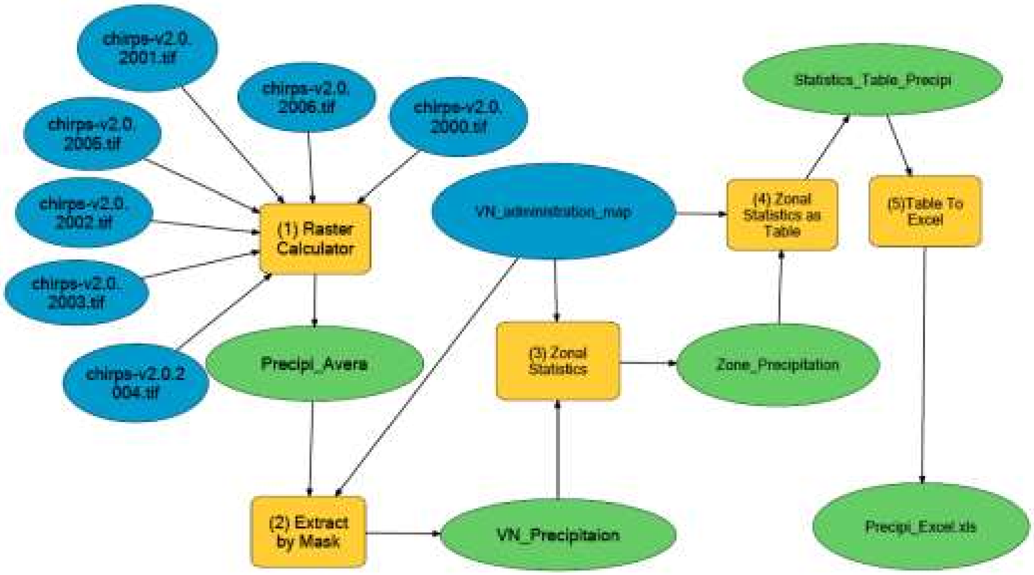


Figure 5.4 The example of a GIS-based model for precipitation data processing

Yearly actual evapotranspiration data was estimated by the average of evapotranspiration from 1950 to 2000 (Trabucco, 2010). The annual precipitation in Vietnam ranges from about 1480 mm to 3270 mm and the mean annual evapotranspiration is about 810 mm to 1220 mm (Figure 4.5). While annual precipitation is higher in the central region and lower in the north area of Vietnam, the provinces in the country's south west region represent have more annual evapotranspiration. Figure 5.5 illustrates the Vietnam aridity index which is extracted from the Global Aridity Index dataset with high-resolution (30 arc-seconds) global raster climate data from 1970 to 2000. The aridity index for all provinces in Vietnam is higher than 0.65 which means Vietnam belongs to the humid climate class according to Table 5.2. The most humid area in Vietnam is the central coast region with an aridity index of nearly 1.5.

Table 5.2 The Classification of climate types based on aridity index (adapted from (Trabucco and J. Zomer, 2018))

Aridity Index	Climate type
< 0.03	Hyper Arid
0.03-0.2	Arid
0.2-0.5	Semi-Arid
0.5-0.65	Dry sub-humid
> 0.65	Humid

Provinces in Vietnam's south central coast region have the highest value of soil water stress with more than 80% of water available for evapotranspiration, indicating high water content in soil so there is less water vulnerability in this area. In contrast, Khanh Hoa and Ninh Thuan provinces experience less soil water stress and higher water vulnerability with SWS value ranging from 61% to 67%. The Priestley–Taylor alpha coefficient (PAC) was calculated by the fraction of annual actual evapotranspiration and the annual potential evapotranspiration (Trabucco, 2010). The value of PAC ranges from 0-100%. The higher value of PAC illustrates lower water vulnerability so consequently, PAC was selected as one of the indicators for assessing water vulnerability. It can be seen in the figures below that while the north of Vietnam has higher PAC values than other regions, the central highland and south central coast experience lower PAC values.

Low values of LAI can be seen in the southeast area and Mekong Delta, followed by the Central Highlands zone. Hau Giang and Long An provinces had a LAI index below 30 m^2/m^2, while in comparison, Ca Mau and Bac Kan experienced higher LAI values with nearly 115 m^2/m^2 (Figure 4.5). LAI are influenced by natural factors, for example climate but there are also human factors involved such as farming and deforestation activities. LAI is deemed to be an indicator which impacts on water resources. This is due to the processes of evapotranspiration, water flow and infiltration, and aquifer recharge (Taugourdeau et al., 2014). Higher level of LAI value results in a much reduced water risk (Isundwa and Mourad, 2019).

Higher WUE and RUE values can be observed in the Mekong Delta. In contrast, the Central Coast area and Central Highlands have experienced lower levels of WUE and RUE. The country's northwest region also suffered from poor water use efficiency. The provinces Lam Dong, Kon Tum and Dak Nong in the Central Highlands area have the lowest value of NPP; about 50 g C/m^2 resulting in lower WUE and RUE. In contrast, Mekong Delta and Red River Delta experienced higher levels of NPP. Ca Mau province witnessed the highest value of NPP with 230 g C/m^2 and it had the highest level of WUE and RUE, followed by Bac Lieu province.

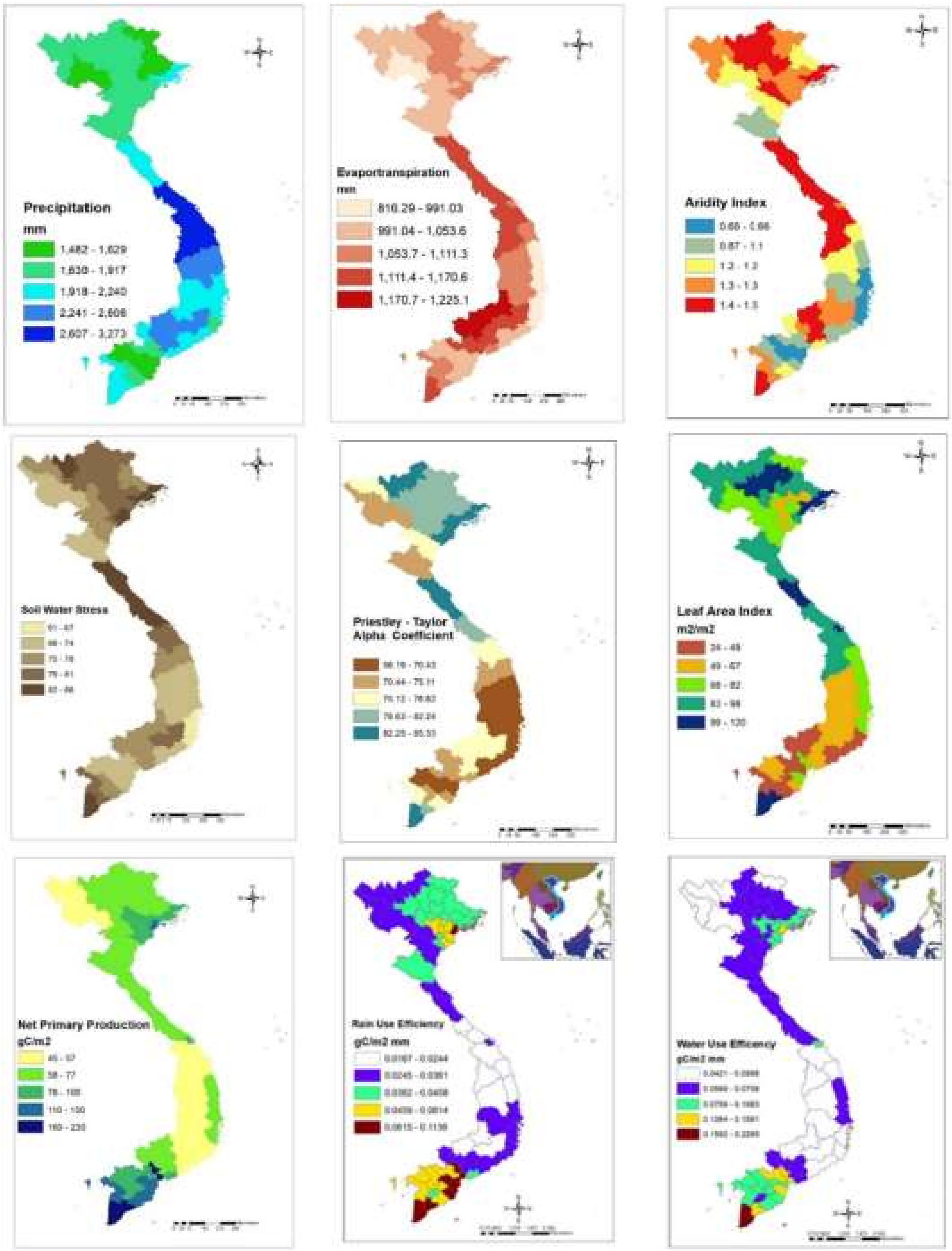


Figure 5.5 Mapping of spatial distribution for Vietnam's provincial remote sensing indicators

5.3.4 The weights of indicators

The weights of selected indicators are identified in Table 5.3 that represent their contribution to the issue of water quantity vulnerability. There are four main methods to evaluate indicators' weights: (1) identifying the weights by expert consultants and interviews; (2) assumption of the equal weight for all variables; (3) applying multivariate statistical techniques; and (4) using the Iyengar and Sudarshan method. The fourth method is the easiest to apply and the most feasible for this study. The indicators' weights were determined by the Iyengar and Sudarshan method via the Excel function. These weights were calculated independently for each component of exposure, sensitivity and adaptive capacity that are illustrated in Table 5.3. Impervious surface ratio, female ratio, population density and river density were identified as having higher relative importance compared to other indicators in this vulnerability assessment.

Table 5.3 Results of the weight calculation

Component	Indicator	Code	Weight
Exposure	Evaporation	E1	0.135
	Annual rainfall	E2	0.0951
	Aridity index	E3	0.0908
	Flood index	E4	0.0877
	Elevation	E5	0.079
	Priestley–Taylor alpha coefficient	E6	0.1013
	Impervious surface ratio	E7	0.1709
	Population density	E8	0.1433
	Population growth rate	E9	0.097
Sensitivity	Irrigation -Eroded earth, rock	S-1	0.1666
	Economic loss	S2	0.1482
	Soil Water Stress	S3	0.1851
	Agricultural production land	S4	0.1276

Component	Indicator	Code	Weight
	Female ratio	S5	0.215
	Poverty rate	S6	0.1575
Adaptive Capacity	Water Use Efficiency	AC1	0.0833
	Rain Use Efficiency	AC2	0.0939
	Leaf area index (LAI)	AC3	0.0703
	River density	AC4	0.1325
	Road density	AC5	0.1062
	Pervious surface	AC6	0.0665
	Percentage of trained employed workers at 15 years of age and above	AC7	0.079
	Total foreign direct investment until 2018	AC8	0.0816
	Number of health establishments	AC9	0.0897
	Percentage of population provided	AC10	0.0729

Component	Indicator	Code	Weight
	with clean water by centralized water supply system		
	Percentage of household having hygienic water	AC11	0.0468
	Annual income	AC12	0.0773

5.3.5 Spatial distribution of water vulnerability

The indices of E, S, AC and WVI were then calculated and ranked for ecological zones and provincial areas as well. Of the eight ecological zones, the northeast area experienced both the lowest level of exposure, sensitivity and adaptive capacity which are indicated in Table 5.4. Very low vulnerability occurs in Southeast area. The Mekong Delta although has the highest level of sensitivity, it exposes low vulnerability due to exhibiting very high level of adaptive capacity compared to the other ecological zones. The South Central Coast area is greatly influenced by flood disasters and extreme climate events (Luu et al., 2019). It is one reason causing the highest level of water vulnerability sensitivity in the South Central Coast (Table 5.4). The study results suggest that the very high vulnerable area are the South Central Coast and Northeast, followed by Red River Delta and Central Highlands.

Table 5.4 Results of water quantity vulnerability assessment for the ecological zones

EZ	Exposure	Sensitivity	Adaptive Capacity	WVI	Vulnerability level
Northeast	0.3193	0.1671	0.1671	0.4398	Medium

Northwest	0.3323	0.1698	0.1764	0.4419	Very high
Red River Delta	0.3229	0.1681	0.1701	0.4403	High
North Central Coast	0.3270	0.1795	0.1885	0.4393	Low
South Central Coast	0.3401	0.1763	0.1861	0.4435	Very high
Central Highlands	0.3280	0.1695	0.1757	0.4406	High
Southeast	0.3202	0.1713	0.1816	0.4366	Very low
Mekong Delta	0.3269	0.1801	0.1891	0.4393	Low

The indices for exposure, sensitivity and adaptive capacity at the provincial level were also established (see Figure 5.6). The most vulnerable provinces are Yen Bai, Binh Dinh, Khanh Hoa, Ho Chi Minh, Ninh Thuan, Da Nang and Quang Nam. In contrast, Binh Duong, Hoa Binh, Tay Ninh, Nghe An, An Giang, Ha Tinh experienced a low level of water vulnerability (Figure 5.7a). Results concerning provincial water vulnerability indices can help local authorities with their water systems planning and deciding on what is the best strategy to implement.

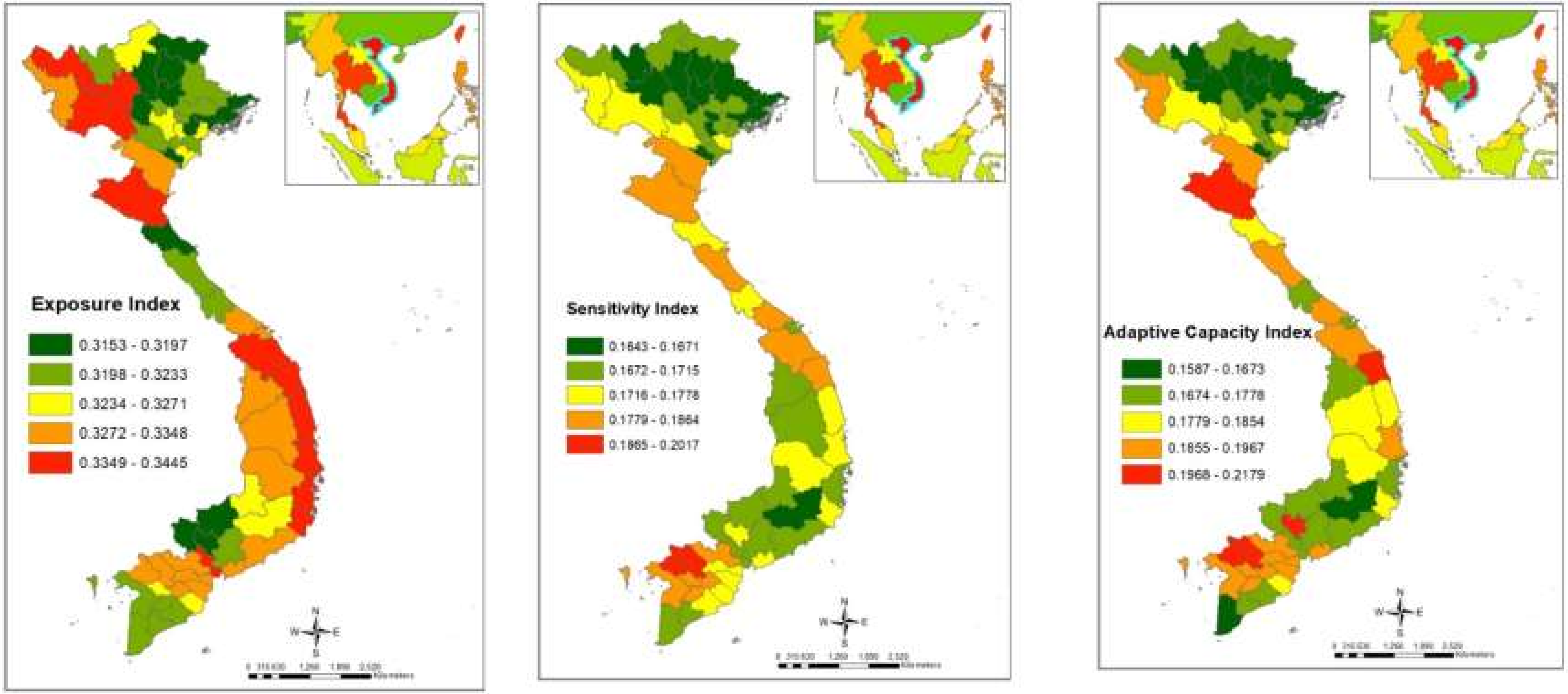


Figure 5.6 Spatial distribution for provincial exposure index (a), sensitivity index (b) and adaptive capacity index (c)

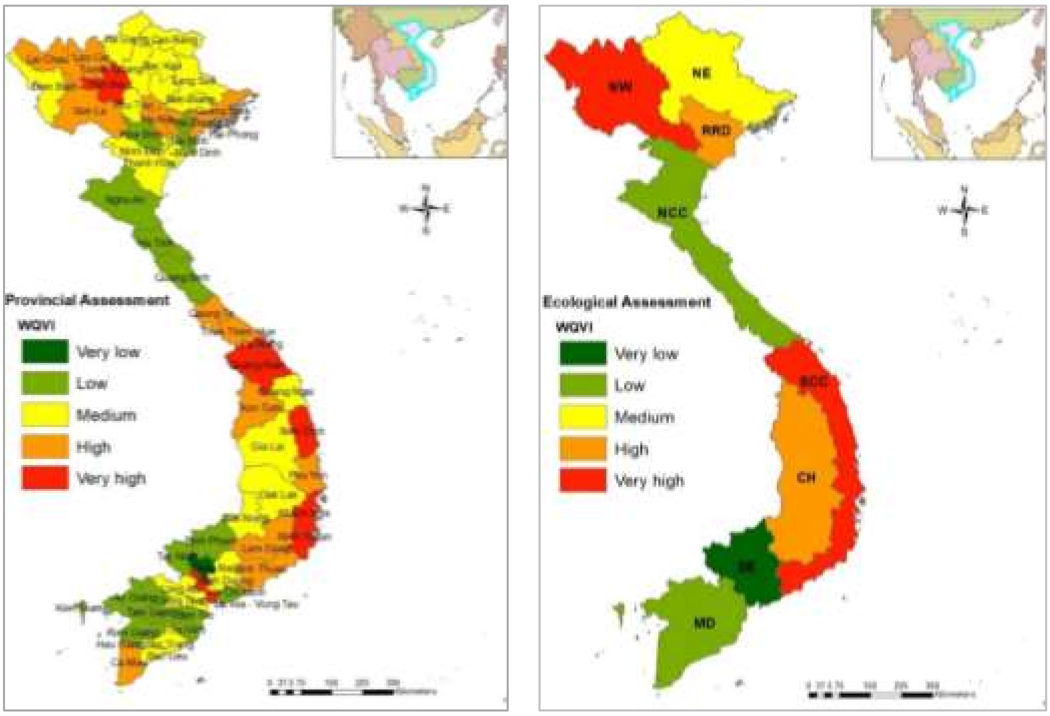


Figure 5.7 Spatial distribution of provincial water quantity vulnerability index (a), and ecological water quantity vulnerability index (b)

The vulnerability index is influenced by all three components of exposure, sensitivity and adaptive capacity. The spatial distribution of water quantity vulnerability for Vietnam's eight ecological zone is provided in Figure 5.7b. This picture clearly indicates the most vulnerable ecological zones which are the South Central Coast and the North West. Those areas also experienced poor water use efficiency levels. The South Central Coast area has a higher exposure index with lower adaptive capacity and it suffers the highest vulnerability index. It is also evident that the South East is the least vulnerable due to the lower level of exposure and the fact that the Mekong Delta is highly resilient because adaptive capacity there is very high.

5.4 Overall discussion

The approach can be applied easily for other studies and regions with freely accessible spatial data sources. Assessing water quantity vulnerability is to identify regions and communities that need to prioritize planning and implementing water strategies in terms of water stress across the world. The finding indicates the South Central Coast has very high level of water quantity vulnerability. The South Central Coast area also was identified as the highest vulnerable area to typhoons and floods (Nguyen et al., 2019; Luu et al., 2019). The Northwest region was illustrated as very high vulnerable area in the study due to the lower level of community's resilience which was presented in the other studies of vulnerability assessment for Vietnam (Few & Tran, 2010; Thanh Thi Pham et al., 2020). The framework employed 27 important spatial-temporal indicators, however; it did not explore other influential factors such as freshwater availability, water stress, water withdrawal for different reasons or other institutional indicators due to the unavailability of data in the study area. In addition, water related hazards like water pollution and contamination should be undertaken in the research. These aspects should be incorporated in further water vulnerability assessment studies.

Uncertainties in the study should be addressed through verification processes by survey, questionaries and a communication between experts from different fields such as social, technical and political fields. For example, indicator's selection and indicator's weight should be cross-checked through consultations with scientists and local officials so that it will help the current framework become more refined. In the scope of the study, the verification of the results were carried out on the total number of human affected (deaths, injured, missing) by flood, flash-flood, rain, storm and typhoon and water quality vulnerability index using a multiple linear regression analysis in R. Human affected data were collected from Sendai Framework for Disaster Risk Reduction the Central Committee for Flood and Storm Control (CCFSC) of Vietnam from 1989 to 2015. The coefficient of determination (R^2) is one of statistical measures which is applied to assess the model performance (Anandhi and Kannan, 2018). If the R-squared is greater than 60%, it is considered acceptable for hydrological simulation and prediction (Anandhi and Kannan, 2018; Moriasi et al., 2007; Santhi et al., 2001). There was a linear relationship between the water quantity vulnerability and the observations of human affected by water related hazards with the multiple coefficient of determination (R-squared) of 0.6463 and residual standard error of 0.6406. It can be seen from figure 5.8 that the results of new approach are reasonable as residuals are close to straight dashed lines.

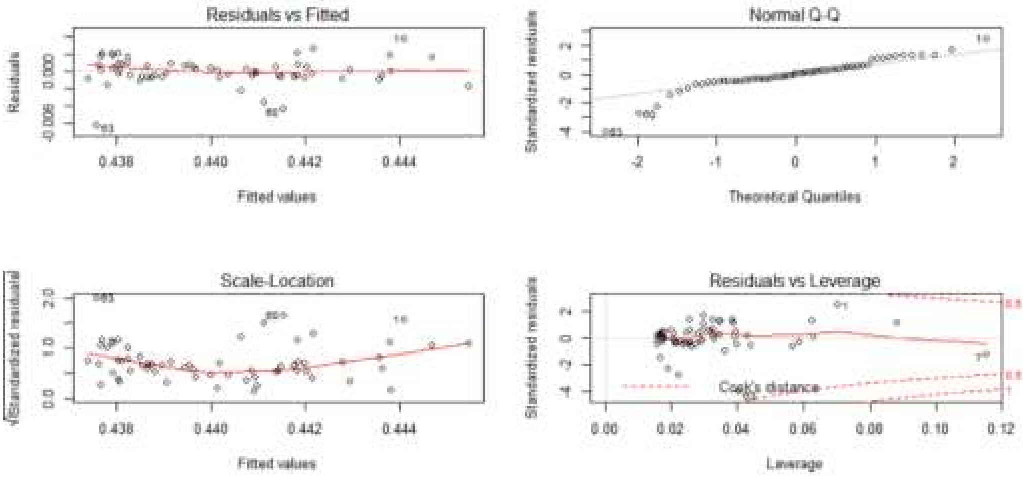


Figure 5.8 Linear regression plots

Furthermore, the further water vulnerability assessment should consider the concept of sustainability in the DPSIR (Driver-Pressure-State-Impact-Response) framework for the selection of vulnerability's indicators. Moreover, it should be applied machine learning methods like Analytical Hierarchy Process (AHP), the Fuzzy Logic, Weights of Evidence (WOE) and Logistic Model Tree (LMT) to improve the results of the framework (Khosravi et al., 2018). Finally, the accepted framework should integrate climate change and population growth scenarios so that a future water vulnerability index can be predicted.

The results of the study are still beneficial although its limitations and uncertainties. Firstly, it can be used for water quantity vulnerability adaptation and mitigation research by providing standardized input data for site selection of alternative water practices. Secondly, it helps practitioners and administrators identify influencing factors to water quantity vulnerability in order to support them more understanding the water system. In addition, it allows policy makers appreciate baseline data and wide range of information for water implementation practices. Finally, the water quantity vulnerability framework of the study employing updated spatial datasets is an indispensable approach for countries where lack of efficient data for conducting vulnerability and impact analyses.

5.5 Conclusions

Appropriate evaluation of water vulnerability is vital if we are to understand the impact of climate change and human activities on water resources. Water managers must have the most effective water management strategies in the future where climate change will influence much of what human societies do. Several methods and frameworks of vulnerability assessment are available, but their performance can be compromised by several obstacles due to the unfeasibility and unavailability of input data. This study developed a new assessment framework that considered the contribution of satellite datasets like Terra MODIS and the utilization of GIS-based model. The integration is very useful and flexible for the complexity of vulnerability assessment. Moreover, this study calculated the indices of water quantity vulnerability components and constructed spatial distribution maps of water exposure, sensitivity and adaptive capacity in different scales for a case study. This part of the study was based on 27 chosen time series indicators. The findings indicated that the South Central Coast experiences extremely vulnerable, while the South East region is the least vulnerable region. Yen Bai, Binh Dinh, Khanh Hoa, Ho Chi Minh, Ninh Thuan, Da Nang and Quang Nam provinces are classified as very high vulnerability. Through the calculation of indicators' weight, it can be concluded that more responsible variables for high level of water vulnerability are impervious surface ratio, population density, healthcare establishment, water use efficiency and river density. Vietnam's growing population is accompanied by an increase in the impervious surface infrastructures like high-rise building and deforestation. These activities are triggering a high rate of water quantity vulnerability. A future work should be considered scenarios namely population, water availability, climate change to enhance water vulnerability evaluation and increase the resolution of the framework in identifying vulnerable hotspots. In addition, the machine learning methods and field observations should be integrated to increase the accuracy of such framework.

Chapter 6

Exploring next generation spatial modelling of soil moisture

6.1 Introduction

Soil moisture (SM) has played vital roles in hydrological state and ecological processes which affects energy, water, and carbon cycles such as evaporation, transpiration, diversity and rainfall-runoff of various ecosystems (Ågren et al., 2021; efBabaeian et al., 2021; Robinson et al., 2008). Soil moisture is also a crucial predictor indicator for identify crop water stress, which helps agricultural drought monitoring. Thorough knowledge about the spatiotemporal patterns of SM is of essential importance for understanding water budgets in hydrological systems which helps prevent agricultural drought problems, water vulnerability, the issues of water shortage, and improve properly crop production across the world (Chaudhary et al., 2021; Tuller et al., 2019). Traditional ground techniques of soil moisture based on field experiments, in-situ soil sensing instrumentation, and geophysical and mobile sensing (Cheng et al., 2022; Robinson et al., 2008). The disadvantages of these method are high cost with small-scale monitoring. Remotely sensed measurements including active remote sensing and passive remote sensing recently have employed effectively for SM monitoring globally (Chaudhary et al., 2021; Cheng et al., 2022; Dubois et al., 2021; Prasad et al., 2018; Warner et al., 2021). At present, various satellite systems via microwave remote sensing like Soil Moisture Active Passive (SMAP) (Entekhabi et al., 2010), Advanced Scatter Meter (ASCAT) (Wagner et al., 2013), and Soil Moisture and Ocean Salinity (SMOS) (Kerr et al., 2001) have been explored for global SM monitoring with spatial resolutions of 10km, 50km, and 35km, respectively. With the low spatial resolution, SM data obtained from these aforementioned missions have not been used widely in farm scales for agricultural management.

Recent advances in earth observation technology such as using active and passive remote sensing (RS) imagery have been dedicated to solving the problems of SM dynamics retrieval on farming lands. Active remote sensing like Unmanned Aerial System (UAS) with highly flexible flight schedules and high spatial resolutions of images offer a great opportunity to estimate the SM for farm-scales (efBabaeian et al., 2021). The application of high-resolution about 2m images from airborne LIDAR can accurately estimate the SM dynamics to support precision agriculture production (Ågren et al., 2021). However, the deployment of UAS and LIDAR have struggled with some obstacles such as limited fight time, high operation cost, and challenges with hyperspectral images processing which limits the application of active RS for SM monitoring (Gago et al., 2015).

Multispectral remote sensing sensors such as Sentinel 1 and Sentinel 2 datasets from European earth observation program Copernicus have employed recently to capture effectively the SM content in several agricultural areas across the world with the spatial resolution of 10-100m (El Hajj et al., 2017; Georganos et al., 2018). The free-of-charge imagery from Sentinel date at high spatial and temporal resolutions are a proper solution to address the challenges of hyperspectral images in agricultural SM prediction. The C-band of the Sentinel-1A and-1B Synthetic Aperture Radar (SAR), and vegetation and soil indices from Sentinel -2A and -2B have been generated to estimate SM properties at high spatial resolution in the pilot scale (Aksoy et al., 2021; El Hajj et al., 2017; Karthikeyan & Mishra, 2021; Ma et al., 2021; Prasad et al., 2018; Schönauer et al., 2021; Senanayake et al., 2021). In addition, terrain indices from digital elevation (DEM) models such as slope, topographic wetness index (TWI), and death-to-water (DTW) index have also been used to predict the agricultural SM (Ågren et al., 2021; Murphy et al., 2008). Topo-hydrological indicators generated from high-resolution DEM data illustrated high correlations with soil properties and soil moistures by capturing the hydrological processes' characteristics of specific sites (Zhao et al., 2021; Zhou et al., 2020). According to Florinsky et al., (2002), soil properties including soil moisture have a significant relationship with topographic attributes, especially in agricultural landscapes.

Machine learning techniques are already commonly applied to handle diverse and large volumes of remote-sensing datasets, with very high performances (Carranza et al., 2021; Gómez et al., 2020; Gómez et al., 2021; Hosoda et al., 2020; Karthikeyan & Mishra, 2021; Ma et al., 2021; Prasad et al., 2018; Schmidt et al., 2020). Artificial intelligence techniques such as random forest regression (RFR), support vector machine (SVM), extreme gradient boosting regression (XGBR), CatBoost gradient boosting regression (CBR) have been employed widely to estimate soil moisture products with high prediction accuracy (Ågren et al., 2021; Carranza et al., 2021; Senanayake et al., 2021). The RFR algorithm performed well to predict the field-scale of soil moisture in China using unmanned aerial vehicle (UAV) imagery with coefficient determination (R2) of 0.91 (Ge et al., 2019). The XGBR technique was used to estimate the SM dynamics in Swedish forest landscape using multiple LIDAR derived digital terrain indices with high performance values compared to RFR and SVM (Ågren et al., 2021). In general, ML algorithms provide a substantial potential for the SM estimation accurately. In this study, a new approach for soil moisture monitoring using the combination of three free-of-charge and high-resolution remote sensing datasets including Sentinel 1, Sentinel 2, and ALOS DSM was presented to estimate the soil moisture in field-scale. Four well-known ML algorithms including RFR, SVM, XGBR, and CBR were employed to test the performance of predictor variables from these datasets. The optimisation of hyper-parameters tuning and the selection of predictor variables during the construction phase of the ML techniques was applied to improve the performance of ML models. The study aims to: (1) assess the correlation of prediction indicators derived from multi-spectral images, SAR datasets, and ALOS DSM in the SM retrieval; (2) select and optimize features from these indicators using genetic algorithm (GA); (3) evaluate the prediction performance of the selected ML model (XGBR) with various scenarios of data-fusion level in the SM prediction while exploring the effectiveness of GA feature optimization on the ML model in mapping the SM content at 10 m spatial resolution; and (4) compare the estimation accuracy of XGBR model with other three well-know ML models using optimal features. The novel framework will be expanded to other field-scales or regional scales to build the SM map, which provides valuable data for different stakeholders like water managers, local authorities, and landholders to practice precision agriculture.

Major parts of this chapter were under revision in a peer-reviewed journal (A-rated journal):

Nguyen, T.T., Ngo, H.H., Guo, W., Chang, S.W., Nguyen, D.D., Nguyen, C.T., Zhang, J., Liang, S., Bui, X.T., Hoang, N.B. 2022. A low-cost approach for soil moisture prediction using multi-sensor data and machine learning algorithm. *Science of The Total Environment*, **833**, 155066.

6.2 Materials and methods

6.2.1 Study area and soil sample collection

The study sites are located in the Wests, Goomalling shire (latitude coordinate: -31°18'S and longitude coordinate: 116° 49' E), and Cookies area - Northam shire (latitude: -31° 39' S, and longitude: 116° 39' E) in the agricultural region of Western Australia (WA) (Figure 6.1). The WA has a diverse type of agricultural production including vegetable industries which contributes a majority of total value of agricultural production in the region. Pastoral and cropping are two key agricultural practices in the WA (Kingwell et al., 2020). According to Australian Bureau of Agricultural and Resource Economics, high-rainfall, wheat-sheep, and pastoral zones are the main agricultural climatic zones in Australian (Salim & Islam, 2010). The type of climate in the WA is a Mediterranean climate where is hot and dry in summer, and cool and wet in winter seasons. The rainfall season is from April and October which ranges between 300 and 600 mm (Kingwell et al., 2020).

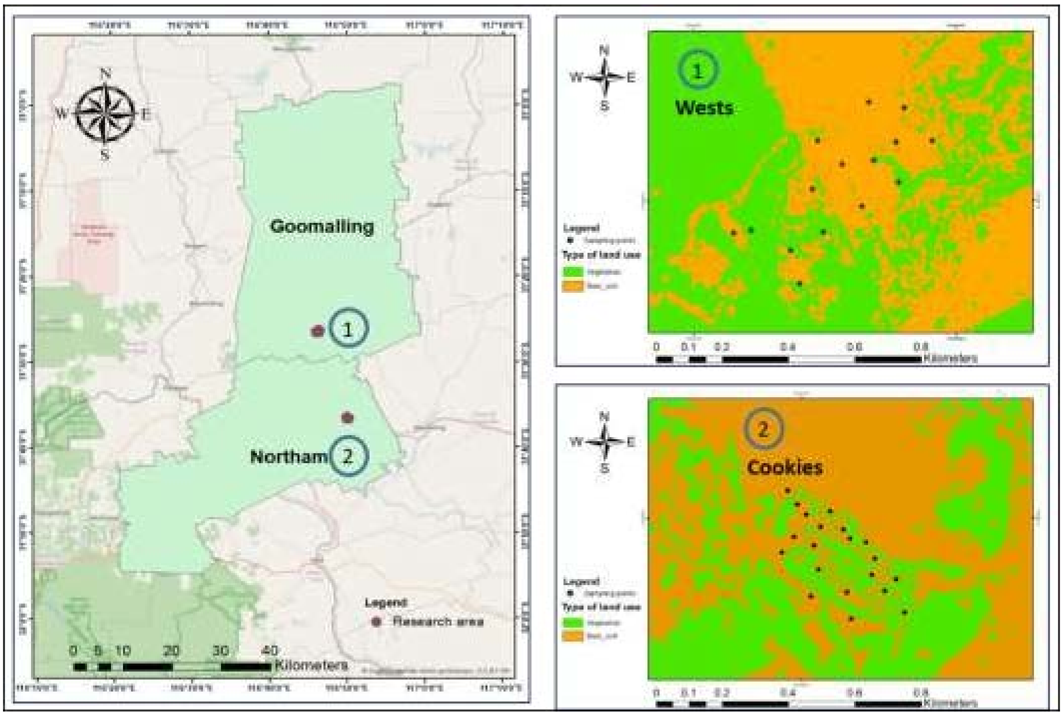


Figure 6.1 Location of the study sites and sampling points in Wests and Cookies area

Soil sampling points were selected from a binary land-use classification map which was produced by extreme gradient boosting classification from very high spatial resolution Google Earth imagery and Sentinel 2 imagery which described in detail in Chapter 3.

6.2.2 Research framework

The research framework consists of five main steps (Fig. 6.2): (1) collecting surface soil dataset (0-10cm) from the binary land-use classification map; (2) generating predictor indicators from optical (Sentinel 2), synthetic aperture radar (Sentinel 1), and terrain indices derived from ALOS DSM; (3) computing Pearson's correlation analysis and feature selection using genetic algorithm; (4) evaluating the performance of the XGBR model with five different scenarios developed from features derived from S1, S2, and ALOS DSM with 70% of SM measured dataset used for models' training and 30% for models' validation; and (5) comparing the performance of the XGBR model with other ML techniques using optimal features and building the SM dynamics map for the study areas.

The prediction accuracy of ML techniques was tested with different scenarios which were developed based on the level of S-1, S-2, and ALOS DSM data fusion and the results from feature selection and optimization using GA. The five scenarios were presented in Table 3. While SC1 comprises of 22 indicators derived from S-2 and 3 indicators from ALOS-DSM, SC2 includes 27 S1 features, and three features derived from ALOS-DSM. SC3 consists of 22 S-2 predictor variables and 27 S1 variables. SC4 contains the total 52 features generated from both S-1, S-2, and ALOS DSM. The potential of 21 optimal features from GA selection for SM prediction was evaluated in SC5. The scenarios were presented in Table 6.2 below. The aim of scenarios development was to evaluate the impact of the level of different features combinations and the application of feature selection algorithm on how well the SM dynamic prediction went.

Table 6.1 Lists of developed scenarios for soil moisture estimation

Scenario	Data fusion	Number of features
SC1	S-2+DEM	25
SC2	S-1+DEM	30
SC3	S-1+S-2	49
SC4	S-1+S-2+DEM	52
SC5	S-1+S-2+DEM with feature selection	21

6.2.4 Machine learning algorithms

6.2.4.1 Extreme gradient boosting (XGBR)

This algorithm is described in Chapter 3, Section 3.2.4.1.

6.2.4.1 Random forest regression (RFR)

This algorithm is described in Chapter 3, Section 3.2.4.2.

6.2.4.2 Support vector machine (SVM)

This algorithm is described in Chapter 3, Section 3.2.4.3.

6.2.4.3 CatBoost gradient boosting regression (CBR)

This algorithm is described in Chapter 3, Section 3.2.4.4.

6.2.5 Genetic algorithm (GA) for feature selection

This algorithm is described in Chapter 3, Section 3.2.5.

6.2.6 Model performance evaluation

To assess the model performance of the soil moisture estimation, two stardard testing criteria were used to evaluate the performance of ML techniques with various scenarios including: the root mean square error (RMSE), and the coefficient of determination (R^2). These indicators are presented in Chapter 3, Section 3.2.6.

6.3 Results and discussion

6.3.1 Correlation analysis of predictor indicators and measured SM

The relationship between the input features derived from S-1, S-2, ALOS-DSM and measured SM content was computed by Pearson's correlation coefficient method. According to Table 6.3, indicators derived from ALOS imagery have a higher correlation with the SM content compared to other indicators. While DEM and Slope obtained negative correlations, TWI illustrated a positive correlation with the SM. All vegetation indices generated from S-2 demonstrated negative correlation with the SM content. Some these indices revealed higher correlations with the SM content including NDVI, SAVI, and IRECI. Colour index from soil indices had a higher correlation to the SM compared to another SIs. NDWI confirmed a negative and high relationship with the estimation of SM. Regarding to the S-1 derived indicators, most transformation bands obtained weak correlations with the SM content, however; VV, VH, and most GLCM textures from VV confirmed strong relationships with the measured SM.

Table 6.2 Pearson's correlation analysis of input variables and measured SM

Input variables	Correlation coefficient *(r)*	Input variables	Correlation coefficient *(r)*
B2	0.005	VV-VH	0.076
B3	-0.046	VV/VH	-0.045
B4	0.087	VH/VV	0.045
B5	0.064	VH_Contrast	-0.073

Input variables	Correlation coefficient (r)	Input variables	Correlation coefficient (r)
B6	-0.155	VH_Dissimilarity	-0.045
B7	-0.247	VH_Homogeneity	0.022
B8	-0.279	VH_Angular Second Moment	-0.037
B8A	-0.355	VH_Energy	-0.002
B11	0.049	VH_Maximum Probability	-0.009
NDWI	-0.366	VH_Entropy	-0.014
B12	0.125	VH_GLCM Mean	-0.437
RVI	-0.389	VH_GLCM Variance	-0.440
NDVI	-0.402	VH_GLCM Correlation	0.042
GNDVI	-0.249	VV_Contrast	-0.328
NDI45	-0.055	VV_Dissimilarity	-0.382
SAVI	-0.499	VV_Homo-geneity	0.401
MCARI	-0.070	VV_Angular Second Moment	0.332
IRECI	-0.568	VV_Energy	0.352
BI	0.031	VV_Maximum Probability	0.311
BI2	-0.111	VV_Entropy	-0.377
CI	-0.329	VV_GLCM Mean	-0.415
RI	0.142	VV_GLCM	-0.414

Input variables	Correlation coefficient (r)	Input variables	Correlation coefficient (r)
		Variance	
VH	-0.414	VV_GLCM Correlation	0.311
VV	-0.347	DEM	-0.616
(VH+VV)/2	-0.403	Slope	-0.495
VH-VV	-0.083	TWI	0.368

6.3.2 Evaluation and comparison of scenarios and different ML models

The proposed XGBR model was trained and tested with five scenarios which were developed by various features extracted from S-1, S-2 and ALOS DSM (Table 6.4). The SC5 with optimal number of features including seven vegetation indices (NDWI, RVI, NDVI, GNDVI, SAVI, MCARI, IREC1), 11 S-1 derived indicators (VH, VV, MeanVHVV, VV_Contrast, VV_Dissimilarity, VV_Homogeneity, VV_Angular Second Moment, VV_Energy, VV_Maximum Probability, VV_Entropy, VV_GLMC Correlation), and both three variables from ALOS DSM yielded the highest SM estimation accuracy with the highest R2 of 0.891 in the validation phase and the lowest RMSE of 0.875% , followed by SC4 with the maximum number of features extracted from selected sensors.. A combination of S-2 and ALOS DSM derived predictor features illustrated a higher performance than the combination of S1 and DEM and S-1 and S-2 generated indicators.

Three well-known ML techniques including CBR, RFR, and SVM were employed to compare the accurate estimation of the SM content with the proposed XGBR model using multi-source EO data fusion. The comparison of ML techniques was conducted with optimal features derived from S-1, S-2, and ALOS DSM. According to table 6.5, gradient boosting regression algorithms (XGBR) outperformed RFR and SVM. While XGBR achieved a highest prediction accuracy with R^2 = 0.891 and RMSE = 0.875, followed by CBR with R^2 = 0.789 and RMSE = 1.217 and SVM with R^2 = 0.493 and RMSE = 1.850. The RFR produced a lowest prediction performance with R^2 = 0.368 and RMSE = 2.491.

Table 6.3 Performance comparison of ML algorithms on agricultural SM estimation

No	Machine learning model	R^2 testing (30%)	RMSE (%)
1	Extreme gradient boosting regression (XGBR)	0.891	0.875
2	CatBoost gradient boosting regression (CBR)	0.789	1.217
3	Random Forests (RFR)	0.368	2.491
4	Support Vector Machine (SVM)	0.493	1.850

Figure 6.3 presents the scatter plots of the estimated versus measured the SM content from four well-known ML techniques in testing phase. The XGBR yielded a better prediction with optimal variables extracted from these multiple sensors using the genetic algorithm compared to CBR, RFR, and SVM. The proposed model using XGBR and GA indicates an R^2 value of 0.891, showing a higher prediction result compared to recent SM monitoring studies with R^2 reached 0.83 in SM prediction study using S1 and Landsat-7 data in Egypt (Mohamed et al., 2020) and R^2 of 0.72 in surface soil moisture estimation using S1 and S2 in India (Tripathi & Tiwari, 2020).

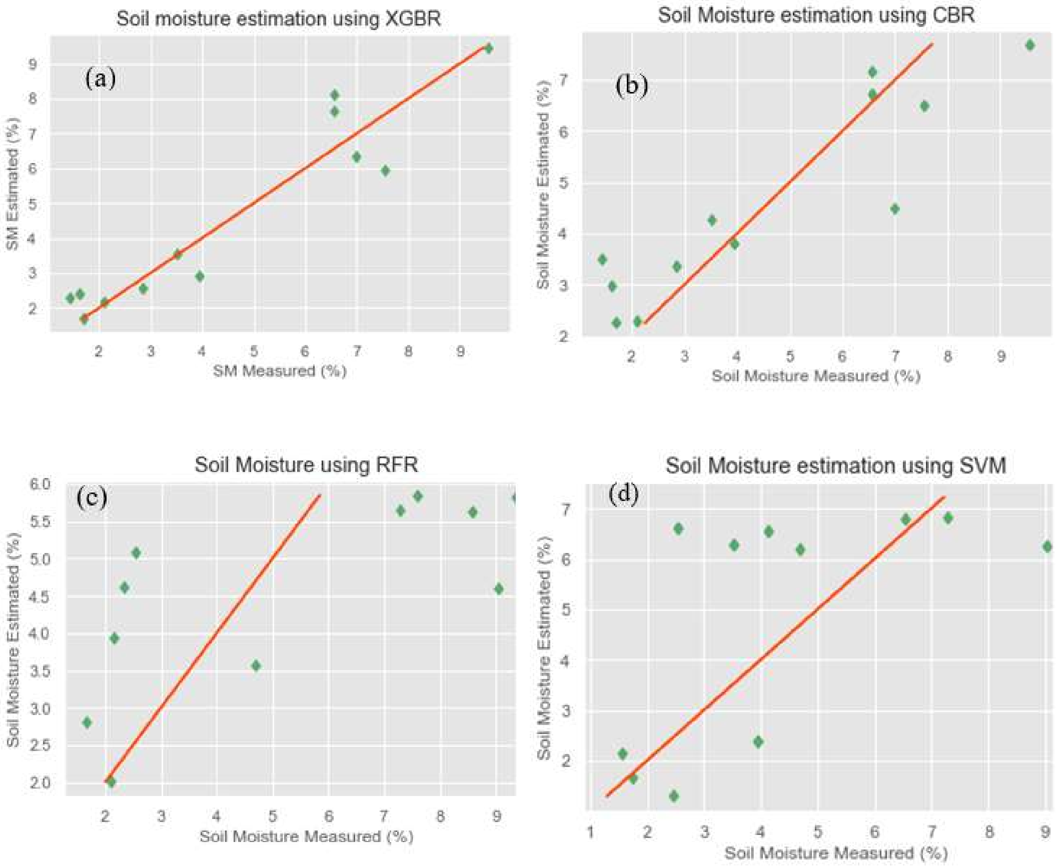


Figure 6.2 Scatter plots of the measured SM and estimated SM using (a) XGBR-GA, (b) CBR, (c) RFR, (d) and SVM.

6.3.3 Spatial distribution patterns of SM maps

Based on scenario 5, the spatial dynamics of SM maps built for the Wests and Cookies areas using S-1, S-2, and ALOS DSM data fusion by the XGBR-GA model are revealed in Fig. 6.4. The XGBR model for the SM prediction in bare-soil pixels obtained the low level of uncertainty and stable prediction capabilities with the low standard deviation value. The proposed moisture prediction model using the XGBR-GA should be calibrated and tested with large-scale earth observation data, over several of land-use types, and various soil-depths.

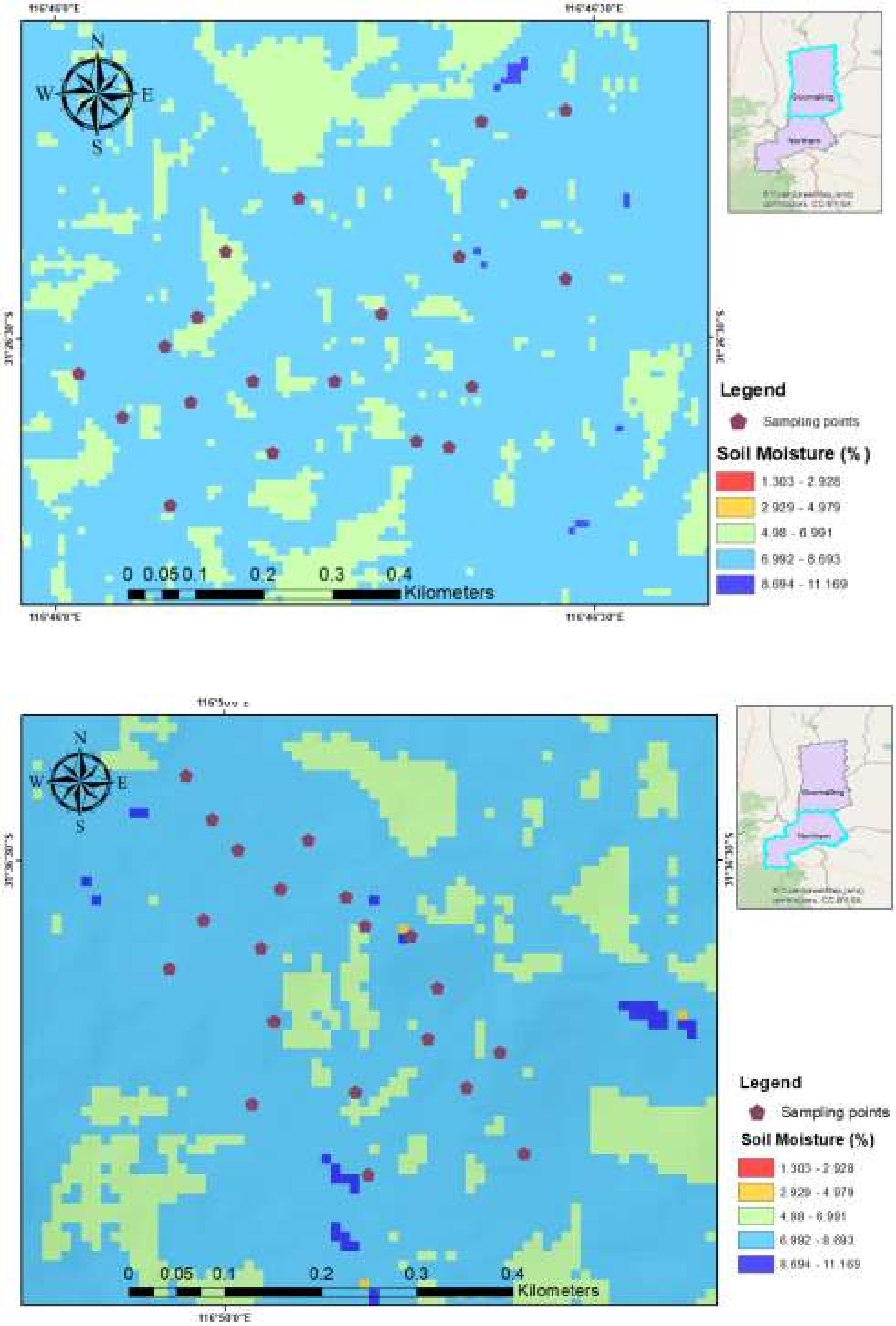


Figure 6.3 Maps of SM content in study areas: (a) Wests and (b) Cookies using XGBR and multiple data fusion.

6.3.4 Relative importance of SM prediction indicators

The estimation accuracy of the SM content has been greatly affected by predictor indicators selection and machine learning algorithm. The higher level of data fusion with optimal feature selection using the GA illustrated better prediction performance for retrieving the SM content. The XGBR had a higher capability to predict the SM pattern. The study also indicated that the GA could help improve the prediction accuracy of the SM estimation which is similar with the results from recent studies using ML models and GA for soil properties estimation (Xie et al., 2015). The successful application of ML models and big data from RS imagery in the SM prediction has been presented in much research at the regional, national and global scale (Carranza et al., 2021; Chaudhary et al., 2021; Cheng et al., 2022; Fang et al., 2021; Ma et al., 2021; Senanayake et al., 2021). The relative importance of optimal features using the GA is presented in Fig. 8. ALOS DSM-derived terrain indices played important roles in the SM prediction. Terrain variables were also mentioned as important indices for the SM prediction in previous studies (Ågren et al., 2021; Leempoel et al., 2015; Zhao et al., 2021). In addition, dual polarization VV, VH, and GLCM textures derived from S-1 are also crucial indices for the SM prediction. The SAR-based prediction indices can improve the estimation of soil moisture (El Hajj et al., 2017; Ma et al., 2020; Zhao et al., 2021). VH was illustrated as the most sensitive index for the SM retrieval in this study. Vegetation indices were selected as optimal features for the SM prediction such as the normalizer difference vegetation index (NDVI), and soil adjusted vegetation index (SAVI) which have been applied for not only vegetation classification, but also further indirectly the SM estimation (Kogan, 1995; Reza et al., 2020). Normalized difference water index (NDWI) from Sentinel 2 also highly correlated with the SM content (Ma et al., 2020). The soil moisture prediction model using the XGBR-GA should be calibrated and tested with large-scale earth observation data, over several of land-use types, and various soil-depths.

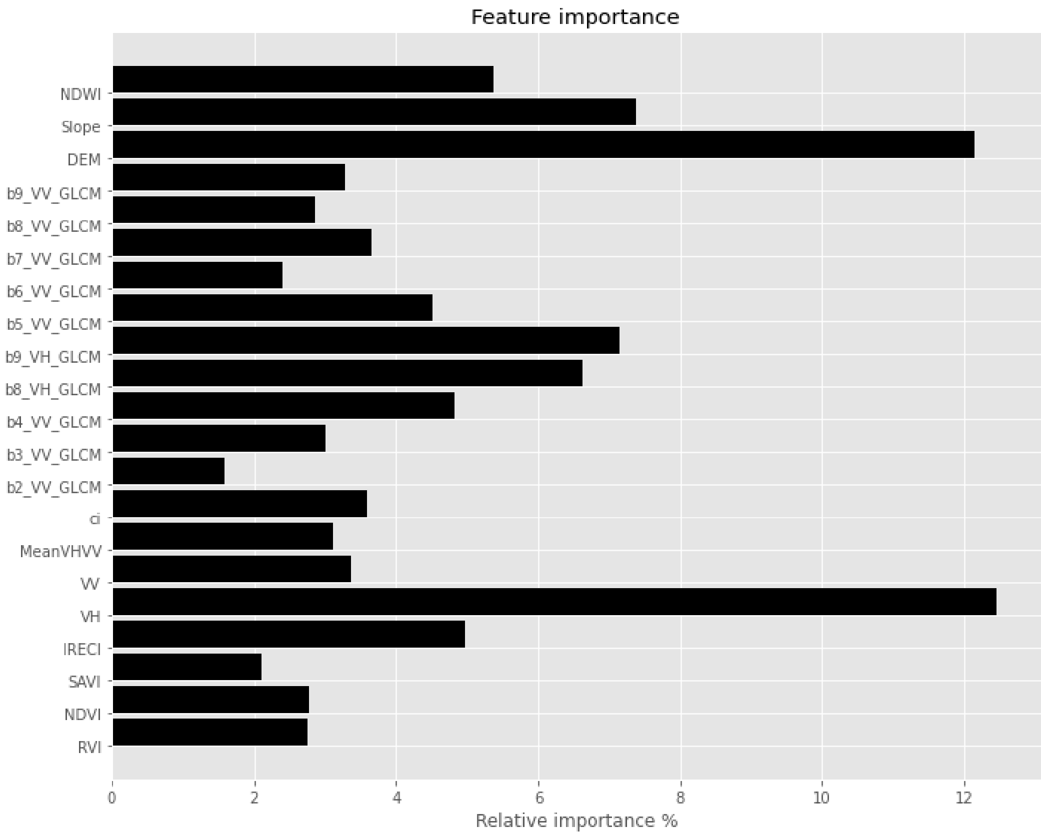


Figure 6.4 Variable importance of optimal features derived from multi-source EO data.

6.4 Conclusions

The present work presented a novel framework using the predictor variables from Sentinel datasets at 10m and ALOS DSM at 30m spatial resolution with a state-of-art machine learning technique (XGBR) and GA for the SM prediction. It is used for estimating the SM content in study sites of Western Australia. It can be seen that the combination of the selected remote sensing dataset illustrated to be very effective for the SM prediction. High level of data fusion and the GA method for optimal features selection showed remarkably better prediction accuracy than single sensor derived features or scenarios without feature optimization. The XGBR model with 21 optimal prediction variables using genetic algorithm approach illustrated the highest prediction performance (R^2=0.891, RMSE= 0.875%). In addition, the proposed XGBR model combined with GA algorithm for variables selected can produce SM maps at 10m spatial resolution using freely remote sensing datasets with a precise accuracy at different scales from field plots to region areas. VH and DEM had the highest relative importance in predicting the SM dynamics. The proposed model should be tested in large-scale areas with various land-use characteristics in further studies. In conclusion, this SM pattern monitoring approach can assist agricultural drought monitoring, the development of appropriate water management strategies, and precision agriculture in terms of climate change.

Chapter 7

A novel intelligence approach for agricultural soil organic carbon prediction

7.1 Introduction

Soil is one of the largest carbon pools in terrestrial ecosystems, and it plays a vital role in the global carbon cycles and care of the ecosystem (Lal, 2008; Zhou et al., 2020b). Agricultural soil organic carbon (SOC) contributes significantly to soil quality, soil fertility, agriculture and greenhouse gas emissions reduction by carbon sequestration in the agricultural SOC stock (Guo et al., 2021; Navarro-Pedreño et al., 2021; Venter et al., 2021). The agricultural SOC depends on land management practices, soil property and differs among rainfall zones (Guo et al., 2021; Six et al., 1998; Venter et al., 2021). Understanding the agricultural SOC distribution spatially is necessary to ensure food security and improve carbon sequestration in soil due to the increasing climate change problems (Gholizadeh et al., 2018). High-precision agricultural SOC data can help local authorities and governments establish appropriate strategies for water management and various farmland activities (Guo et al., 2021). Climate, ecological processes, agricultural production activities, water conditions, soil characteristics, and land management are the key factors greatly influencing agricultural SOC.

The monitoring of agricultural SOC is complex due to the uncertainty of the above factors. Conventional SOC monitoring methods based on field experiments are time- and labour-consuming and subsequently, SOC mapping in large-scale areas is expensive (Forkuor et al., 2017). It is necessary to develop alternative approaches that are more cost-effective and accurate in predicting SOC. Numerous studies have attempted to solve this problem such as developing environmental models to improve the SOC estimation and applying remote sensing sensors to build digital SOC maps (Guo et al., 2021a; Guo et al., 2021b; Ha et al., 2021; He et al., 2021; Le et al., 2021; Mondal et al., 2017; Zhou et al., 2020). While developed SOC prediction models like a Full Carbon Accounting Model (FullCAM) or De-Nitrification De-Composition (DNDC) need a large amount of information from soil type, farming practices, and climate, they have illustrated their limitations in the prediction.

The XGBoost was used in many studies due to its high predictive performance and being an effective supervised learning algorithm for addressing various classification and regression tasks with promising results (Chen and Guestrin, 2016), however; it has not been applied for agricultural SOC monitoring. For these reasons, the present study aims to develop a novel framework using free-of-charge multi-sensor Sentinel 2 and Sentinel 1 with state-of-the-art extreme gradient boosting (XGBoost) to predict agricultural SOC stocks. The specific objectives are to: (1) assess the feasibility of using multi-spectral images and SAR dataset in estimating agricultural SOC; (2) compare the prediction performance of the XGBoost to two other well-known ML techniques (random forest (RF) and support vector machine (SVM)) with various scenarios of data-fusion level in agricultural SOC prediction; and (3) highlight important predictor features in mapping agricultural SOC stock at 10 m spatial resolution. The novel agricultural SOC prediction framework will then be expanded so that relevant stakeholders are aware of the many advantages for agricultural management, climate change mitigation and landholders wanting to make more profit via carbon markets.

Major parts of Chapter 7 was illustrated in were published in peer-reviewed journals:

Nguyen, T.T., Pham, T.D., Nguyen, C.T., Delfos, J., Archibald, R., Dang, K.B., Hoang, N.B., Guo, W., Ngo, H.H. 2022. A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion. *Science of The Total Environment*, **804**, 150187 (IF: **7.963**; SJR: **Q1**).

Nguyen, T.T. 2021. Predicting agricultural soil carbon using machine learning. *Nature Reviews Earth & Environment*, **2**(12), 825-825.

7.2 Materials and methods

7.2.1 Study area

The study sites are the Wests area which belongs to Goomalling shire (latitude coordinate: -31°18'S and longitude coordinate: 116° 49' E), and Cookies area which belongs to Northam shire (latitude: -31° 39' S, and longitude: 116° 39' E). These areas are located in the agricultural lands of Western Australia (WA). The agricultural sector plays an essential role in the WA's economy. Pastoral and cropping are two main agricultural activities in the WA. According to Australian Bureau of Agricultural and Resource Economics, there are three key agricultural climatic zones in Australian, which are High-rainfall, Wheat-sheep, and Pastoral zones (Salim & Islam, 2010). While 95 per cent of gross value of agricultural production in the WA comes from the high-rainfall and wheat-sheep zones, only 5% of agricultural products is produced from pastoral zones. As the agricultural of the WA bases totally on rainfall, the main season for crop production in the WA is from April to October. The rainfall in growing season ranges between 146 to 294 mm (Petersen & Hoyle, 2016).

7.2.2 Soil samples collection

The methods to collect soil samples is described in detail in Chapter 3, Section 3.2.3.

7.2.3 Research framework

The research process includes four main phases (Fig. 7.1): (1) collection of surface soil dataset (0-10cm) based on the binary land-use map; (2) computation of predictor variables from optical (Sentinel 2) and synthetic aperture radar (Sentinel 1) remote sensing data; (3) spatial modelling of agricultural SOC based on advanced machine learning techniques including XGBoost, RF and SVM model; and (4) evaluating the model's performance with 70% of SOC dataset generated for models' training and 30% for models' testing. This was done to select the most accurate model for SOC prediction and mapping the spatial patterns of agricultural SOC.

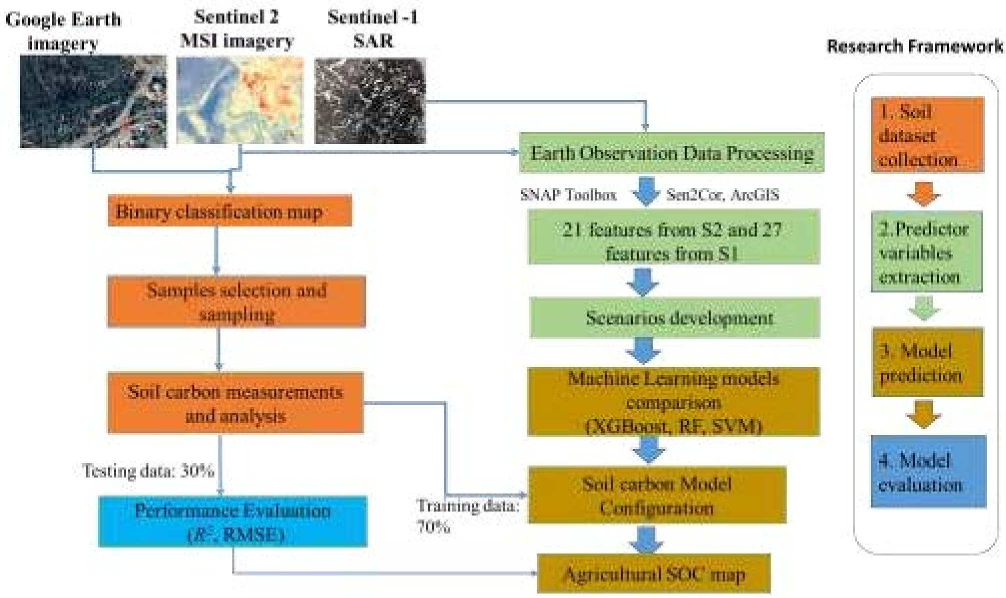


Figure 7.1 A novel established framework of agricultural SOC prediction using multi-sensor data fusion

7.2.4 Remote sensing data acquisition and image processing

7.2.4.1 Data acquisition

Sentinel images were acquired from the Copernicus Open Access Hub from European Space Agency (ESA), which are described in Chapter 3, Section 3.1.1.

7.2.4.2 Image transformation of Sentinel imagery

Sentinel imagery was processed and transformed by SNAP toolbox, which are presented in detail in Chapter 3, Section 3.2.1.1.

7.2.5 Scenarios development

Scenarios were constructed based on the different number of predictor features and the combinations of sensors. While Scenario 1 and Scenario 2 were developed from S-2 derived predictors, Scenario 3 and Scenario 4 were built from S-1 derived predictors. Scenario 1 (SC1) included only 10 features from 10 S-2 bands. Scenario 2 (SC2) consisted of a total of 21 S-2 derived predictors including 10 S-2 bands, 7 VIs bands, and 4 SIs bands. Scenario 3 (SC3) and Scenario 4 (SC4) comprised 7 and 27 predictor features from the S-1 sensor, respectively. Scenario 5 (SC5) included all features based on the combination of S-2 and S-1. The purpose of scenarios development was to assess the impact of the type of predictor variables and the level of different features combinations on how well agricultural SOC prediction went.

7.2.6 Machine learning techniques

7.2.6.1 Extreme gradient boosting (XGBoost)

The extreme gradient boosting technique is presented in Chapter 3, Section 3.2.4.1. In this SOC estimation study, the XGBoost method can use integrated optimization algorithms to tune important hyper-parameters such as the number of trees and the rate of learning to suit a specific dataset. In this study, the best structure with 100 trees, and a learning rate set at 0.5 and gamma value of five was found the highest performance in the XGBoost model.

5.2.6.2 Random forest (RF)

This machine learning algorithm is described in Chapter 3, Section 3.2.4.2. In the current work of the chapter, the RF model with 100 trees and the maximum number of 11 features had the highest performance for this study area.

5.2.6.3 Support vector machine (SVM)

Developed by Cortes and Vapnik (1995), the SVM algorithm is described in Chapter 3, Section 3.2.4.3. In this work, the SVM algorithm with the radial basis function (RBF) kernel and the C value of 10000 was used, and the epsilon value of 0.01 as the best values for tuning hyper-parameters of the SVM model.

7.2.7 Model performance evaluation

To assess the model performance of binary land-use classification, five evaluation criteria have been used including overall accuracy (*OA*), kappa coefficient (*KC*), precision (*P*), Recall (*R*), and F1 score (*F1*) (Chicco & Jurman, 2020; Ha et al., 2021).

For agricultural SOC retrieval, two common validation criteria were employed to assess the performance of machine learning techniques with different scenarios, which are described in Chapter 3, Section 3.2.6.

7.3 Results

7.3.1 Land-use binary mapping

Land-use classification results found by the XGBoost, the RF and the SVM algorithms are indicated in Table 7.2 below. The results present the high accuracy of land-use binary mapping at study sites using the S-2 dataset. The XGBoost algorithm produced the highest accuracy and performed better than the RF and the SVM with 0.94 OA, 0.89 KC, 0.96 P, 0.91 R and 0.93 F1.

Table 7.1 Model's performance of land-use binary mapping using S-2 dataset

No	Machine learning model	*OA*	*KC*	*P*	*R*	*F1*
1	Extreme Boosting (XGBoost)	0.94	0.89	0.96	0.91	0.93
2	Random Forests (RF)	0.92	0.85	0.88	0.87	0.91
3	Support Vector Machine (SVM)	0.86	0.79	0.84	0.82	0.85

The land use binary classification maps were created for the Wests and Cookies area using the XGBboost model using S-2 dataset and Google Earth imagery (Fig. 7.2). The classified map includes only bare soil and vegetation classes. Based on the binary classification maps, the precise locations belonging to the bare-soil pixels were used as a guide for sampling agricultural SOC field collection.

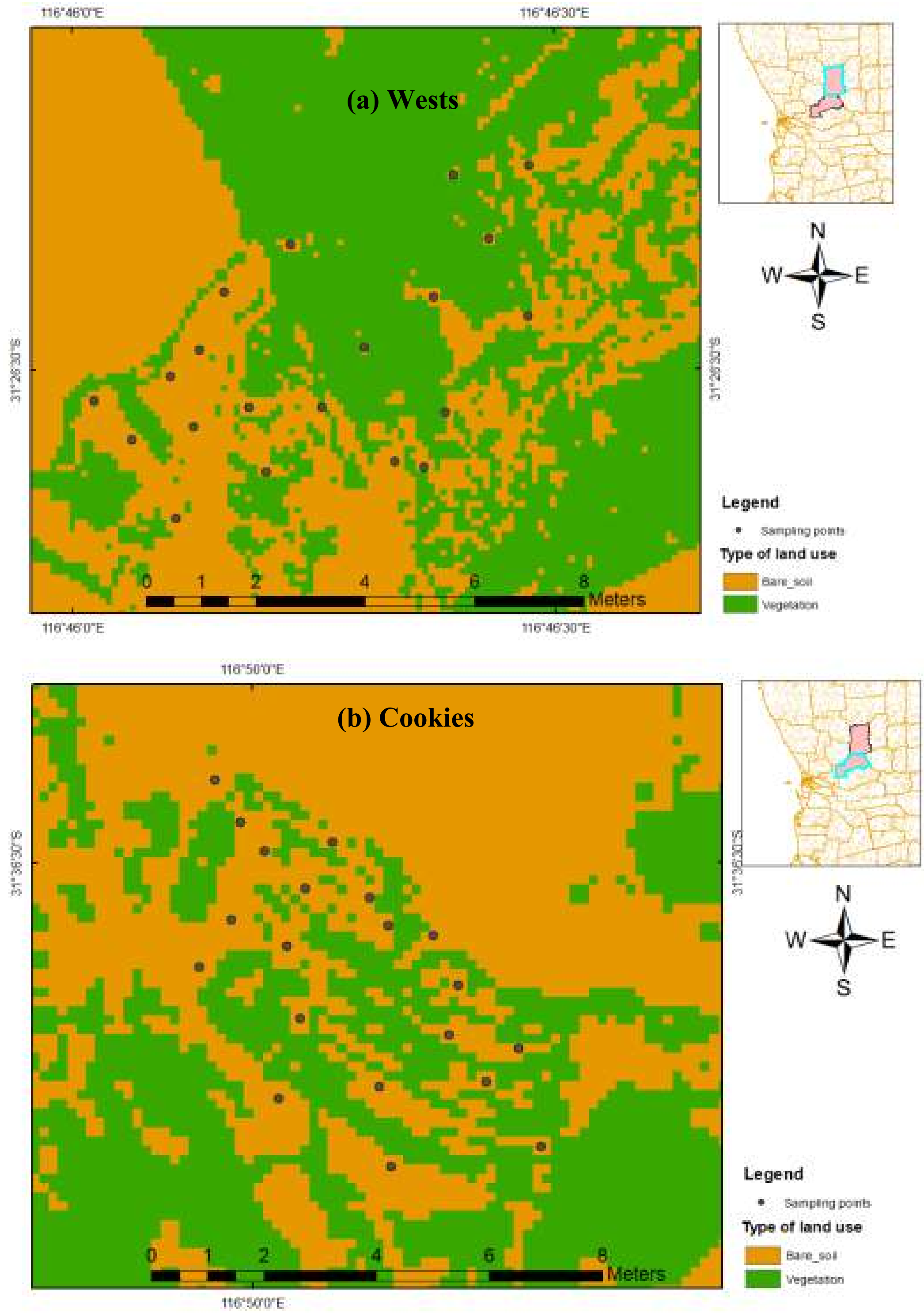


Figure 7.2 Land use binary classification map derived from the XGBoost model using S-2 and sampling points selection: (a) Wests, and (b) Cookies

The correlation coefficient between the input features derived from S2 data, VIs, and SIs with measured agricultural SOC was computed and illustrated in table 7.3. According to Table 7.3, the Ratio Vegetation Index (RVI), the Normalized Difference Vegetation Index (NDVI), and the Soil Adjusted Vegation Index (SAVI) presented the highest correlation with measured agricultural SOC among 21 predictor features derived from the S-2 image. These indices revealed positive correlations with agricultural SOC. In contrast, the lowest correlations were observed between Brightness Index 2 (BI2) and agricultural SOC. Vegetation and Soil Indices confirmed a higher correlation with agricultual SOC than ten S-2 multispectral bands. While vegetation indices illustrated positive correlations with agricultural SOC, most soil indices including BI, CI, and RI demonstrated negative correlations.

Table 7.2 Pearson's correlation analysis of S-2 derived predictor indicators and measured SOC

S2_Bands_Index	Correlation coefficient	S2_VI_BI_Index	Correlation coefficient
B2	-0.056	RVI	0.409
B3	-0.043	NDVI	0.419
B4	-0.162	GNDVI	0.167
B5	-0.131	NDI45	0.116
B6	-0.011	SAVI	0.470
B7	0.059	MCARI	0.088
B8	0.125	IRECI	0.377
B8A	0.170	BI	-0.113

S2_Bands_Index	Correlation coefficient	S2_VI_BI_Index	Correlation coefficient
B11	-0.022	BI2	0.005
B12	-0.025	CI	-0.296
		RI	-0.059

Table 7.3 shows the Pearson's correlation analysis of S-1 derived predictor indicators and measured agricultural SOC. VV, (VV+VH)/2, VH_GLCM Mean, VH_GLCM Variance, VV_Dissimilarity, VV_Homogeneity, VV_Angular Second Moment, VV_Entropy, VV_GLCM Mean, VV_GLCM Variance demonstrated the highest correlation with agricultural SOC compared to other predictor features generated from S-1 data. Most GLCM textures showed strong correlations with agricultural SOC content. Four out of five S-1 SAR transformation bands (VH-VV; VV-VH; VV/VH; and VH/VV) had weak relationships with agricultural SOC.

Table 7.3 Pearson's correlation analysis of S-1 derived predictor indicators and measured SOC

S1 Index	Correlation coefficient	S-1_Index	Correlation coefficient	S-1_Index	Correlation coefficient
VH	0.389	VH_Homogeneity	-0.100	VV_Dissimil-arity	0.417
VV	0.433	VH_Angular Second Moment	-0.047	VV_Homo-geneity	-0.416
(VH+VV)/2	0.439	VH_Energy	-0.083	VV_Angular Second Moment	-0.431
VH-VV	0.251	VH_Maximum Probability	-0.067	VV_Energy	-0.349

VV-VH	-0.243	VH_Entropy	0.106	VV_Maximum Probability	-0.363
VV/VH	-0.118	VH_GLCM Mean	0.434	VV_Entropy	0.432
VH/VV	0.118	VH_GLCM Variance	0.437	VV_GLCM Mean	0.476
VH_Con-trast	0.243	VH_GLCM Correlation	-0.211	VV_GLCM Variance	0.468
VH_Dis sim-ilarity	0.168	VV_Contrast	0.359	VV_GLCM Correlation	-0.328

7.3.2 Evaluation and comparison of scenarios and different ML models

Five scenarios with varied features generated from S-2 and S-1 sensor were tested using the XGBoost technique (Table 7.4). The SC5 with the best possible number of features derived from multi-sensor S-1 and S-2 produced the highest prediction accuracy compared to the others SCs. However, the SC3 with only seven predictor variables from S-1 yielded the worst prediction performance. A combination of S-2 and S-1 derived predictor features showed the highest R^2 of 0.870 in the validation phase and the lowest RMSE of 1.818 tonC/ha.

Table 7.4 Model performance of the XGBoost technique in five scenarios

Scenario (SC)	Number of features	R^2 training (70%)	R^2 validation (30%)	RMSE (Ton C/ha)
SC1	10 features (10 S-2 bands only)	0.713	0.443	3.160
SC2	21 features (10 S-2 bands, 7 bands VIs, and 4 bands SIs)	0.891	0.625	2.370
SC3	7 features (2 bands from dual polarization, 5 SAR	0.559	0.254	3.004

	transformed bands)			
SC4	27 features (2 bands from dual polarization, 5 SAR transformed bands, and 20 bands created from GLMC)	0.998	0.584	2.471
SC5	48 features (21 S-2 bands and 27 S-1-bands)	0.927	0.870	1.818

To compare the effectiveness of the proposed XGBoost model using multi-source EO data fusion, two other well-known ML algorithms were selection for the comparison. The performance of the three ML algorithms on agricultural SOC retrievals are presented in Table 7.5. The SVM model performance in the agricultural SOC prediction was the lowest (R^2 = 0.661) and the RMSE value (4.396 ton/ha) was higher than those produced the XGBoost and the RF model. The XGBoost model with 48 predictor variables derived from a combination of S-2 image and S-1 image yielded the most accurate for agricultural SOC prediction in the validation phases (R^2 = 0.870, and RMSE = 1.818 ton/ha), followed by the RF model (R^2 = 0.724 and RMSE= 2.289 ton C/ha, and the SVM model (R^2 = 0.661 and RMSE= 4.396).

Table 7.5 Performance comparison of ML algorithms on agricultural SOC estimation

No	Machine learning model	R^2 training (70%)	R^2 testing (30%)	RMSE (Ton C/ha)
1	Extreme Boosting (XGBoost)	0.927	0.870	1.818
2	Random Forests (RF)	0.827	0.724	2.289
3	Support Vector Machine (SVM)	0.999	0.661	4.396

Figure 7.3 indicates the scatter plots of the estimated versus measured agricultural soil organic carbon using three well-known ML techniques in testing phase. The

proposed ML models with auxiliary variables from S-2 multispectral imagery and S-1 SAR data can successfully estimate the agricultural SOC. The XGBoost is better at prediction than the RF and SVM.

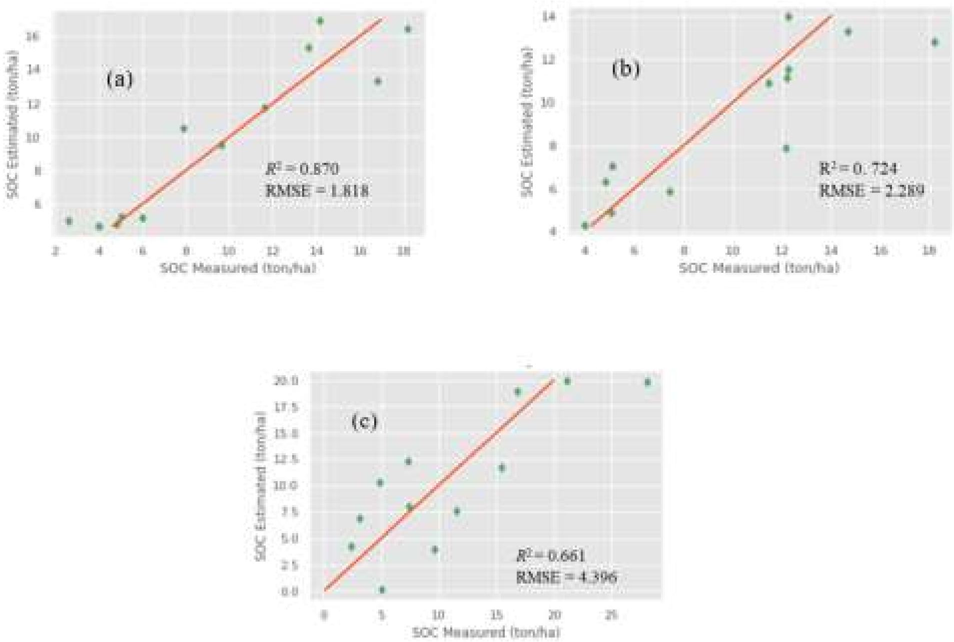
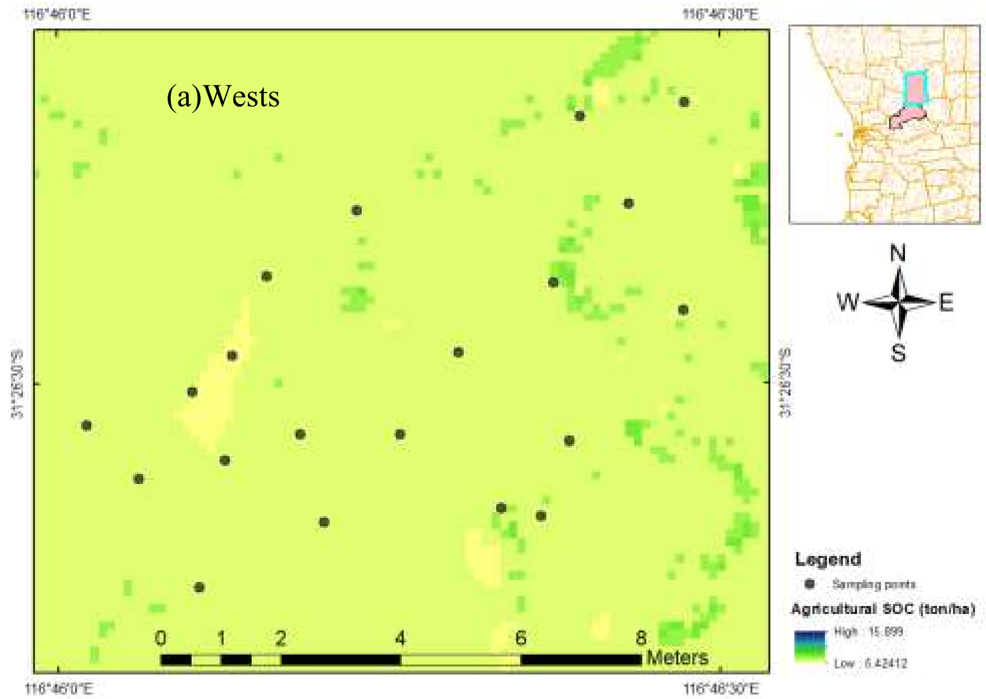


Figure 7.3 Scatter diagrams of the measured SOC and estimated SOC by (a) XGBoost, (b) RF, (c) and SVM.

7.3.3 Spatial distribution patterns of agricultural SOC maps

Based on scenario 5, the spatial distribution of agricultural SOC maps generated for the Wests and Cookies areas using a combination of S-1 and S2 datasets integrated by the XGBoost model are demonstrated in Fig. 7.4. The max, min, mean and standard deviation (SD) values of the predicted agricultural SOC were 15.899 ton C/ha, 5.42 ton C/ha, 6.936 ton C/ha, and 0.45 ton C/ha, respectively. The XGBoost model produced the low level of uncertainty and stable prediction capabilities with the low average value of SD.



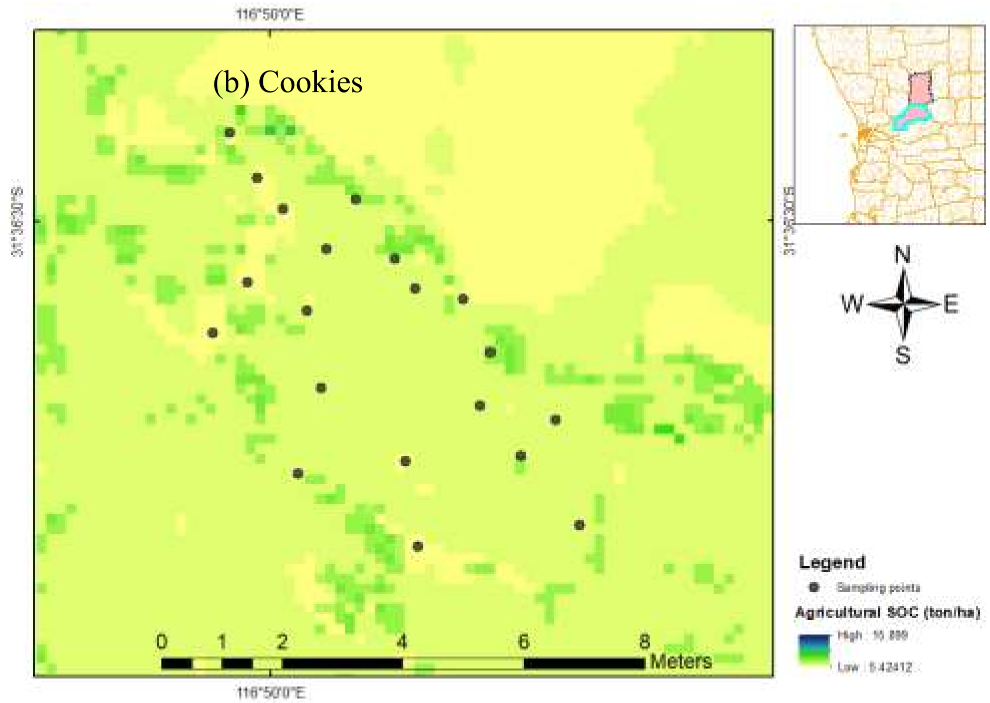


Figure 7.4 Spatial distribution characteristic of agricultural SOC in study areas: (a) Wests (a) and (b) Cookies using the proposed XGBoost combined data fusion.

7.4 Discussion

7.4.1 Performance of agricultural SOC prediction models

The prediction accuracy of agricultural SOC has been greatly influenced by the selection of predictor variables, ML algorithms, and level of data fusion (Table 7.4). The higher level of data fusion with more predictor features derived from Sentinel 2 and Sentinel 1 illustrated better prediction accuracy for retrieving agricultural SOC. This outcome is consistent with what Zhou et al (2020) and Castaldi et al (2019) reported. They indicated that the type of remote sensing data, predictor variables selection and the choice predictive models play important roles in SOC estimation (Castaldi et al., 2019). As well, combining S-2 and S-1 free-of-charge EO data can improve SOC prediction performance. Recent studies also stated that the multi-sensor data fusion has proved to be more effective than the single sensor approach in quantifying SOC for both mangrove SOC stocks and agricultural SOC content (Le et al., 2021; Zhou et al., 2020b).

The XGBoost predictive model is an efficient and effective gradient boosting algorithm which can be applied successfully for predictive modelling in SOC stocks research. The performance of the proposed XGBoost model combined with data fusion in the study performed well and outperformed the two well-known ML algorithms i.e. the RF and the SVM. The XGBoost algorithm is powerful and an advanced ML technique in predicting SOC stocks which is backed up in other recent studies (Ha et al., 2021; Ibrahem Ahmed Osman et al., 2021). The prediction results of the XGBoost in the study shows superior results (R^2 =0.87, RMSE = 1.818 tonC/ha) which are very much higher than the results of other studies noted in Table 1. The proposed framework using the 48 predictor features (10 multispectral bands, 7 vegetation indices, 4 soil indices, 2 bands from dual polarization, 5 SAR transformed bands, and 20 bands created from GLMC) derived from S-1 and S2 combined with the XGBoost ML technique were powerful in agricultural SOC prediction. Importantly, the novel framework developed in this work is able to handle a small number of agricultural SOC samples, reflecting the robustness and cost-effectiveness of the model development for future and long-term agricultural SOC monitoring. However, more studies must be done on more sites, incorporating a wider geographical area.

7.4.2 Relative importance of predictor variables

The successful application of satellite RS images in predicting agricultural SOC has been proved in much research at the regional, national and global scale (Croft et al., 2012; Dvornikov et al., 2021; Hamzehpour et al., 2019; Mirzaee et al., 2016; Paul et al., 2020; Zhou et al., 2020a). However, most studies on this topic concentrated on mapping agricultural SOC based on optical imagery like S-2 imagery, which is due to the close relationship of Sentinel 2 derived indicators and SOC distribution. The present study illustrated that the predictor variables derived from both optical and SAR dataset are effective in estimating agricultural SOC. Similar observations were demonstrated by Yang and Guo (2019) (Yang & Guo, 2019). The relative importance of prediction features is presented in Fig. 7.5. Only 24 variables (10 features derived from S-2 and 14 features derived S-1) out of 48 variables were shown the high relative importance in the agricultural SOC.

Soil Adjusted Vegetation Index (SAVI) was identified as the most important predictor feature for agricultural SOC retrieval. It is due to its high sensitivity to soil characteristics (Huete 1988). The SAVI computed from the NIR and the Red bands also shows the strongest correlation coefficient (0.47) in Table 7.3, reflecting a high sensitivity to soil backgrounds and allowing to quantify the agricultural soil texture and SOC. The result is similar to the finding reported by Xue and Su (2017). The GLCM indicators, and dual polarization VV and VH derived from S-1 are also influential features. The contribution of the predictor variables computed from SAR data on determining agricultural SOC are more significant than S-2 derived variables. This is due to the capture ability of vegetation short-term variation characteristics of the Sentinel 1 sensor. Remarkably, the GLMC textures derived from Sentinel 1 were not previously selected as the predictor features for agricultural SOC prediction. Nonetheless, it can be seen from Fig. 7.5 that GLMC bands from the VV polarization have been illustrated as being satisfactory predictor variables for estimating agricultural soil organic carbon. Future studies focusing on the SAR mechanism on agricultural SOC should be further investigated.

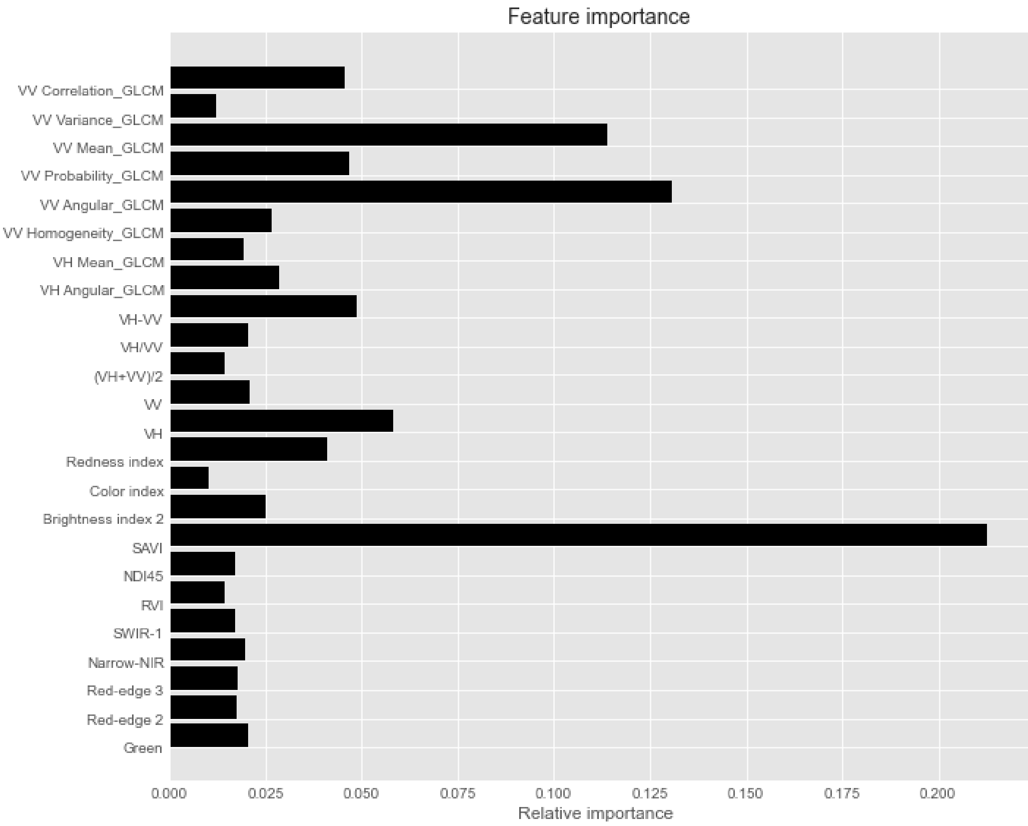


Figure 7.5. Variable importance of optimal features derived from multi-source EO data.

7.5 Conclusions

The present study pioneers the use of predictor features (dual polarizations and transformed bands) from SAR remote sensing imagery (S-1) and the fusion of predictor variables derived from optical remote sensing imagery (S-2) with a state-of-art machine learning technique (XGBoost). It is applied for predicting agricultural SOC in Western Australia. Overall, the combination of S-1 C-band dual polarimetric SAR and optical S2 datasets proved to be very useful for agricultural SOC prediction. High level of data fusion or multi-source sensor derived predictive variables illustrated significantly better prediction performance than a low level of data fusion or single sensor derived features. The proposed XGBoost model using multi-sensor data fusion demonstrated the highest prediction accuracy (R^2=0.870, RMSE= 1.818 ton/ha). In addition, the proposed model is able to derive agricultural SOC maps at 10m spatial resolution on regional scale with a precise accuracy. The binary land-use classification mapping using active learning to select bare soil sampling points and DPGS play important roles in the improvement of agricultural SOC prediction accuracy. Combining ensemble-based learning and active learning can enhance the estimates of agricultural SOC with only a small soil sample dataset. In short, this SOC prediction approach makes possible carbon neutrality for agriculture towards additional revenue via carbon credits.

Chapter 8

Conclusions and recommendations

8.1 Conclusions

Climate change, rapid population growth, and inappropriate regional planning policies in many countries have resulted in large-scale rural water-related problems, such as flood disasters, water pollution, and water shortages. To tackle these issues, the specific water management as employed in the Sponge City concept has operated in China since 2013. The Sponge City concept is emerging as a new type of integrated regional water system, which can address water problems in both urban and rural regions. However, its implementation has encountered a variety of challenges. The lack of models to assist Sponge City planning and implementation, and other water infrastructure strategies is one of the most challenging factors. This thesis aims to address these issues by implementing a new integrated conceptual model concerning rural sustainable water management comprising of: (1) water vulnerability assessment model; (2) soil water model to assist the evaluation of rural water management strategies; and (3) agricultural soil carbon model to support monitoring soil carbon sequestration from various water management practices and different farming measurements. The specific findings of the study are summarized below.

A new approach for water quantity vulnerability assessment based on remote sensing satellite data and GIS ModelBuilder is proposed. The developed approach has three layers: (1) data acquisition mainly from remote sensing datasets and statistical sources; (2) calculation layer based on the integration of GIS-based model and the Intergovernmental Panel on Climate Change's vulnerability assessment framework; and (3) output layer including the indices of exposure, sensitivity, adaptive capacity, water vulnerability, and spatial distribution of remote sensing indicators where these indices at the provincial and regional scales. In total, 27 indicators were incorporated into the case study in Vietnam based on their availability and reliability. The novel approach is based on reliable and updated spatial-temporal datasets (soil water stress, aridity index, water use efficiency, rain use efficiency, and leaf area index), and the incorporation of the GIS-based model. This framework can then be applied effectively for water vulnerability assessment of other regions and countries.

Next generation of soil moisture modelling using advanced machine learning (ML) models, and multi-sensor data fusion including Sentinel-1(S1) C-band dual polarimetric synthetic aperture radar (SAR), Sentinel-2 (S2) multispectral data, and ALOS Global Digital Surface Model (ALOS DSM) to predict precisely soil moisture at 10m spatial resolution across research areas in Australia was presented. A total of 52 predictor variables were generated from S1, S2, and ALOS DSM data fusion, including vegetation indices, soil indices, water index, SAR transformation indices, ALOS DSM derived indices like digital model elevation (DEM), slope, and topographic wetness index (TWI). The field soil data from Western Australia was employed. The performance capability of extreme gradient boosting regression (XGBR) together with the genetic algorithm (GA) optimizer for feature selection for soil moisture prediction in bare lands was examined and compared with various scenarios and ML models.

The proposed model XGBR with 21 desirable features obtained from GA yielded the highest performance (R^2 = 0. 891; RMSE= 0.875%) compared to random forest regression (RFR), support vector machine (SVM), and CatBoost gradient boosting regression (CBR). VH and DEM illustrated the most crucial predictor features for SM estimation using the XGBR. It can be concluded that this innovative method using XGBR, coupled with GA possessing feature from a combination of reliable free-of-charge remotely sensed data from Sentinel and ALOS imagery, can effectively estimate the spatial variability of soil moisture. The described framework can further support precision agriculture and drought resilience programs via better water use efficiency and intelligent irrigation management for crop production.

This work explores the use of Sentinel-1 (S1) C-band dual polarimetric synthetic aperture radar (SAR), Sentinel-2 (S2) multispectral data, and an innovative machine learning (ML) approach using an integration of active learning for land-use mapping and advanced Extreme Gradient Boosting (XGBoost) for robustness of the SOC estimates. Numerous features computed from optical and SAR data fusion were employed to build and test the proposed model's performance. The effectiveness of the proposed machine learning model was assessed by comparing the two well-known algorithms, these being Random Forests (RF) and Support Vector Machine (SVM), to predict agricultural SOC. Results suggest that a combination of S1 and S2 sensors could effectively estimate SOC in farming areas using ML techniques. Satisfactory accuracy of the proposed XGBoost with optimal features was achieved using the highest performance (R^2 = 0.870; RMSE = 1.818 ton C/ha), which outperformed RF and SVM. Thus, multi-sensor data fusion combined with the XGBoost led to the best prediction results for agricultural SOC at 10m spatial resolution. In short, this novel approach could significantly contribute to various agricultural SOC retrieval studies globally.

8.2 Recommendations

For making progress in rural water modelling, several future perspectives are emphasized here:

(1) Uncertainty analysis and the assessment of integrated water modelling should be carried out in parallel with the model's development, testing, and application so that practitioners can rely on the model's output and thereby make good policy decisions. Global Assessment of Modelling Uncertainties could be applied to negate any uncertainties and improve the accuracy of modelling results (Deletic et al., 2012).

(2) The availability of online spatial and attributed data-sharing systems constitutes a fundamental factor in integrated modelling development. The lack of spatial and temporal data creates incomplete knowledge of how integrated models work. Spatial and attribute data-sharing systems can be established based on remote sensing and GIS systems. Remote sensing data should be used much more in the upcoming decades because such data will be very reliable. A better application of remote sensing data might reduce model uncertainties. The spatial data and measurement data should be incorporated into the integrated water management model.

(3) Inter-disciplinary work on integrated rural water modelling research is an essential factor when it comes to building an integrated, comprehensive rural water management model. Water infrastructure implementation in rural areas is a challenging task that requires a multidisciplinary effort to address complex issues and an unforecastable future. A reasonable explanation for the lack of integration

in rural drainage models is that the responsibilities in rural water management have been broken up or isolated from each other (Rauch et al., 2002). Rural water simulation and system analysis include interdisciplinary fields of research such as environment, society, and the economy. A comprehensive rural water management model should incorporate the multi-science field approach. Consequently, reliable simulations of complex interactions in the rural water model will address stakeholders' requirements in solving rural water management problems.

(4) A further effort needs to focus on developing an online and integrated rural water management tool for the best management of rural water operations. This online tool helps decision-makers use the integrated models better. Obtained results from modelling work could enhance the knowledge of how to: firstly, implement and develop integrated rural water management practices; and secondly, estimate the effectiveness of investing in rural water development projects.

(5) There are some areas for rural water model, including the links existing between automatic calibration models and the integrated rural water model to deal with uncertainty and various problems (Elliott and Trowsdale, 2007). In addition, enhancing the adoption of the rural water model by improving the model's communication system is very important (Bach et al., 2014). By being fully informed and disclosing all that is necessary, these model developers will then be able to convince decision-makers about the cost-effectiveness of integrated models and provide accurate and comprehensive results. Further research into the spatial-temporal dynamics of rural rainfall for predictions and improving the spatial simulation of ecological/environmental processes of water strategies is essential (Fletcher et al., 2013; Hou et al., 2019). Finally, an effective rural water management model is likely to support drought, flood, and hydrologic warning systems in releasing more precise information for communities and what they require.

References

A. Guzman, J., Shirmohammadi, A., M. Sadeghi, A., Wang, X., L. Chu, M., K. Jha, M., B. Parajuli, P., Daren Harmel, R., P. Khare, Y., E. Hernandez, J. 2015. Uncertainty Considerations in Calibration and Validation of Hydrologic and Water Quality Models. Transactions of the ASABE, 58(6), 1745-1762.

Ågren, A.M., Larson, J., Paul, S.S., Laudon, H., Lidberg, W. 2021a. Use of multiple LIDAR-derived digital terrain indices and machine learning for high-resolution national-scale soil moisture mapping of the Swedish forest landscape. Geoderma, 404.

Ågren, A.M., Larson, J., Paul, S.S., Laudon, H., Lidberg, W. 2021b. Use of multiple LIDAR-derived digital terrain indices and machine learning for high-resolution national-scale soil moisture mapping of the Swedish forest landscape. Geoderma, 404, 115280.

Ahyerre, M., Chebbo, G., Tassin, B., Gaume, E. 1998. Storm water quality modelling, an ambitious objective? Water Science and Technology, 37(1), 205-213.

Aksoy, S., Yildirim, A., Gorji, T., Hamzehpour, N., Tanik, A., Sertel, E. 2021. Assessing the performance of machine learning algorithms for soil salinity mapping in Google Earth Engine platform using Sentinel-2A and Landsat-8 OLI data. Advances in Space Research.

Alcamo, J., DÖLl, P., Henrichs, T., Kaspar, F., Lehner, B., RÖSch, T., Siebert, S. 2003. Development and testing of the WaterGAP 2 global model of water use and availability. Hydrological Sciences Journal, 48(3), 317-337.

Aldana-Jague, E., Heckrath, G., Macdonald, A., van Wesemael, B., Van Oost, K. 2016. UAS-based soil carbon mapping using VIS-NIR (480–1000nm) multi-spectral imaging: Potential and limitations. Geoderma, 275, 55-66.

Anandhi, A., Kannan, N. 2018. Vulnerability assessment of water resources – Translating a theoretical concept to an operational framework using systems thinking approach in a changing climate: Case study in Ogallala Aquifer. Journal of Hydrology, 557, 460-474.

Angelopoulou, T., Tziolas, N., Balafoutis, A., Zalidis, G., Bochtis, D. 2019. Remote Sensing Techniques for Soil Organic Carbon Estimation: A Review. Remote Sensing, 11(6).

Arshad, M., Guillaume, J., Ross, A. 2014. Assessing the Feasibility of Managed Aquifer Recharge for Irrigation under Uncertainty. Water, 6(9), 2748-2769.

Ashley, R., Lundy, L., Ward, S., Shaffer, P., Walker, L., Morgan, C., Saul, A., Wong, T., Moore, S. 2013. Water-sensitive urban design: opportunities for the UK. Proceedings of the Institution of Civil Engineers - Municipal Engineer, 166(2), 65-76.

aus der Beek, T., Flörke, M., Lapola, D.M., Schaldach, R., Voß, F., Teichert, E. 2010. Modelling historical and current irrigation water demand on the continental scale: Europe. Advances in Geosciences, 27, 79-85.

Bach, P.M., Rauch, W., Mikkelsen, P.S., McCarthy, D.T., Deletic, A. 2014. A critical review of integrated urban water modelling – Urban drainage and beyond. Environmental Modelling & Software, 54, 88-107.

Bär, R., Rouholahnejad, E., Rahman, K., Abbaspour, K.C., Lehmann, A. 2015. Climate change and agricultural water resources: A vulnerability assessment of the Black Sea catchment. Environmental Science & Policy, 46, 57-69.

Besir, A.B., Cuce, E. 2018. Green roofs and facades: A comprehensive review. Renewable and Sustainable Energy Reviews, 82, 915-939.

Beven, K., Binley, A. 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes, 6(3), 279-298.

Breiman, L. 2001. Random Forests. Machine Learning, 45(1), 5-32.

Brown, C.M., Lund, J.R., Cai, X., Reed, P.M., Zagona, E.A., Ostfeld, A., Hall, J., Characklis, G.W., Yu, W., Brekke, L. 2015. The future of water resources systems analysis: Toward a scientific framework for sustainable water management. Water Resources Research, 51(8), 6110-6124.

Brown, R.R., Keath, N., Wong, T.H.F. 2009. Urban water management in cities: historical, current and future regimes. Water Science and Technology, 59(5), 847-855.

Brudermann, T., Sangkakool, T. 2017. Green roofs in temperate climate cities in Europe – An analysis of key decision factors. Urban Forestry & Urban Greening, 21, 224-234.

Brunetti, G., Šimůnek, J., Piro, P. 2016. A comprehensive numerical analysis of the hydraulic behavior of a permeable pavement. Journal of Hydrology, 540, 1146-1161.

Burger, G., Sitzenfrei, R., Kleidorfer, M., Rauch, W. 2014. Parallel flow routing in SWMM 5. Environmental Modelling & Software, 53, 27-34.

Burian, S., Edwards, F. 2002. Historical Perspectives of Urban Drainage. University of Utah, USA.

Campisano, A., Butler, D., Ward, S., Burns, M.J., Friedler, E., DeBusk, K., Fisher-Jeffes, L.N., Ghisi, E., Rahman, A., Furumai, H., Han, M. 2017. Urban rainwater harvesting systems: Research, implementation and future perspectives. Water Res, 115, 195-209.

Candela, A., Freni, G., Mannina, G., Viviani, G. 2011. Receiving water body quality assessment: an integrated mathematical approach applied to an Italian case study. Journal of Hydroinformatics, 14(1), 30-47.

Carle, M.V., Halpin, P.N., Stow, C.A. 2005. PATTERNS OF WATERSHED URBANIZATION AND IMPACTS ON WATER QUALITY1. JAWRA Journal of the American Water Resources Association, 41(3), 693-708.

Carranza, C., Nolet, C., Pezij, M., van der Ploeg, M. 2021. Root zone soil moisture estimation with Random Forest. Journal of Hydrology, 593.

Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., van Wesemael, B. 2019. Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. ISPRS Journal of Photogrammetry and Remote Sensing, 147, 267-282.

Castaldi, F., Palombo, A., Santini, F., Pascucci, S., Pignatti, S., Casa, R. 2016. Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon. Remote Sensing of Environment, 179, 54-65.

Castonguay, A.C., Iftekhar, M.S., Urich, C., Bach, P.M., Deletic, A. 2018. Integrated modelling of stormwater treatment systems uptake. Water Res, 142, 301-312.

Chan, F.K.S., Griffiths, J.A., Higgitt, D., Xu, S., Zhu, F., Tang, Y.-T., Xu, Y., Thorne, C.R. 2018a. “Sponge City” in China—A breakthrough of planning and flood risk management in the urban context. Land Use Policy, in press.

Chan, F.K.S., Griffiths, J.A., Higgitt, D., Xu, S., Zhu, F., Tang, Y.-T., Xu, Y., Thorne, C.R. 2018b. “Sponge City” in China—A breakthrough of planning and flood risk management in the urban context. Land Use Policy, 76, 772-778.

Chandana, D., H., G.M., C., P.K., Hillary, H., Andrea, R., William, S., M., Z.E. 2010. Simulation of Combined Best Management Practices and Low Impact Development for Sustainable Stormwater Management1. JAWRA Journal of the American Water Resources Association, 46(5), 907-918.

Chapin, F.S., Eviner, V.T. 2007. 8.06 - Biogeochemistry of Terrestrial Net Primary Production. in: Treatise on Geochemistry, (Eds.) H.D. Holland, K.K. Turekian, Pergamon. Oxford, pp. 1-35.

Chaudhary, S.K., Srivastava, P.K., Gupta, D.K., Kumar, P., Prasad, R., Pandey, D.K., Das, A.K., Gupta, M. 2021. Machine learning algorithms for soil moisture estimation using Sentinel-1: Model development and implementation. Advances in Space Research.

Che, W., Zhao, Y., Li, J.-q., Wang, W.-l., Wang, J.-l., Wang, S.-s., Gong, Y.-w. 2015. Explanation of Sponge City Development Technical Guide: Basic Concepts and Comprehensive Goals. China Water & Wastewater, 31(8), 1-5.

Chen, H., Ma, J., Wei, J., Gong, X., Yu, X., Guo, H., Zhao, Y. 2018. Biochar increases plant growth and alters microbial communities via regulating the moisture and temperature of green roof substrates. Science of The Total Environment, 635, 333-342.

Chen, T., Guestrin, C. 2016. XGBoost. in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794.

Cheng, M., Li, B., Jiao, X., Huang, X., Fan, H., Lin, R., Liu, K. 2022. Using multimodal remote sensing data to estimate regional-scale soil moisture content: A case study of Beijing, China. Agricultural Water Management, 260.

Chicco, D., Jurman, G. 2020. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Medical Informatics and Decision Making, 20(1), 16.

Christian Amos, C., Rahman, A., Mwangi Gathenya, J. 2016. Economic Analysis and Feasibility of Rainwater Harvesting Systems in Urban and Peri-Urban Environments: A Review of the Global Situation with a Special Focus on Australia and Kenya. Water, 8(4).

Chui, T.F.M., Liu, X., Zhan, W. 2016. Assessing cost-effectiveness of specific LID practice designs in response to large storm events. Journal of Hydrology, 533, 353-364.

Coma, J., Pérez, G., Cabeza, L.F. 2018. Chapter 4.8 - Life Cycle Assessment of Green Roofs. in: Nature Based Strategies for Urban and Building Sustainability, (Eds.) G. Pérez, K. Perini, Butterworth-Heinemann, pp. 341-351.

Cortes, C., Vapnik, V. 1995. Support-vector networks. Machine Learning, 20(3), 273-297.

Cristianini, N., Ricci, E. 2008. Support Vector Machines. in: Encyclopedia of Algorithms, (Ed.) M.-Y. Kao, Springer US. Boston, MA, pp. 928-932.

Croft, H., Kuhn, N.J., Anderson, K. 2012. On the use of remote sensing techniques for monitoring spatio-temporal soil organic carbon dynamics in agricultural systems. Catena, 94, 64-74.

Dang, Tran, Pham, Khoi, Tran, Nguyen. 2019. Exploring Freshwater Regimes and Impact Factors in the Coastal Estuaries of the Vietnamese Mekong Delta. Water, 11(4).

Dardel, C., Kergoat, L., Hiernaux, P., Manuela, G., Mougin, E., Ciais, P., Nguyen, C.-C. 2014. Rain-Use-Efficiency: What it Tells us about the Conflicting Sahel Greening and Sahelian Paradox. Remote Sensing, 6.

Daughtry, C.S.T., Walthall, C.L., Kim, M.S., de Colstoun, E.B., McMurtrey, J.E. 2000. Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. Remote Sensing of Environment, 74(2), 229-239.

David, S. 2014. Water 4.0: the Past, Present, and Future of the World's

Most Vital Resource. Yale University Press, United State.

Davis, A.P., Hunt, W.F., Traver, R.G., Clar, M. 2009. Bioretention technology: Overview of current practice and future needs. Journal of Environmental Engineering, 135(3), 109-117.

Delegido, J., Verrelst, J., Alonso, L., Moreno, J. 2011. Evaluation of Sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. Sensors (Basel, Switzerland), 11(7), 7063-7081.

Deletic, A., Dotto, C.B.S., McCarthy, D.T., Kleidorfer, M., Freni, G., Mannina, G., Uhl, M., Henrichs, M., Fletcher, T.D., Rauch, W., Bertrand-Krajewski, J.L., Tait, S. 2012. Assessing uncertainties in urban drainage models. Physics and Chemistry of the Earth, Parts A/B/C, 42–44, 3-10.

Deng, S., Zhang, X., Shao, Z., Yan, W., Chai, H., He, Q. 2018. An integrated urban stormwater model system supporting the whole life cycle of sponge city construction programs in China. Journal of Water and Climate Change.

Developer, A.W. 2007. Towards sustainable water stewardship. Anglian Water Developer Services.

DHI. 2009. MIKE URBAN - Model Manager. DHI.

Diaz-Granados, M., Rodríguez, J.P., Rodríguez, M.S., Penagos, J.C., Camacho, L.A., Achleitner, S., Maksimovic, C., McIntyre, N. 2009. Towards a Paradigm Shift in Urban Drainage Management and Modelling in Developing Countries. Revista de Ingeniería, 133-150.

Dong, G., Weng, B., Qin, T., Denghua, Y., Wang, H., Gong, B., Bi, W., Xing, Z. 2018. The Impact of the Construction of Sponge Cities on the Surface Runoff in Watersheds, China. Advances in Meteorology, 2018, 9.

Dorogush, A.V., Ershov, V., Gulin, A. 2018. CatBoost: gradient boosting with categorical features support. ArXiv, abs/1810.11363.

Dotto, C.B.S., Kleidorfer, M., Deletic, A., Rauch, W., McCarthy, D.T., Fletcher, T.D. 2011. Performance and sensitivity analysis of stormwater models using a Bayesian approach and long-term high resolution data. Environmental Modelling & Software, 26(10), 1225-1239.

Dubois, A., Teytaud, F., Verel, S. 2021. Short term soil moisture forecasts for potato crop farming: A machine learning approach. Computers and Electronics in Agriculture, 180.

Duong, H.H., Thuc, T., Ribbe, L. 2017. Assessing and Calculating a Climate Change Vulnerability Index for Agriculture Production in the Red River Delta, Vietnam. in: Redefining Diversity & Dynamics of Natural Resources Management in Asia, Volume 2, pp. 27-40.

Dvornikov, Y.A., Vasenev, V.I., Romzaykina, O.N., Grigorieva, V.E., Litvinov, Y.A., Gorbov, S.N., Dolgikh, A.V., Korneykova, M.V., Gosse, D.D. 2021.

Projecting the urbanization effect on soil organic carbon stocks in polar and steppe areas of European Russia by remote sensing. Geoderma, 399.

efBabaeian, E., Paheding, S., Siddique, N., Devabhaktuni, V.K., Tuller, M. 2021. Estimation of root zone soil moisture from ground and remotely sensed soil information with multisensor data fusion and automated machine learning. Remote Sensing of Environment, 260.

El Hajj, M., Baghdadi, N., Zribi, M., Bazzi, H. 2017. Synergic Use of Sentinel-1 and Sentinel-2 Images for Operational Soil Moisture Mapping at High Spatial Resolution over Agricultural Areas. Remote Sensing, 9(12), 1292.

Elliott, A., Trowsdale, S. 2007. A review of models for low impact urban stormwater drainage. Environmental Modelling & Software, 22(3), 394-405.

Escadafal, R. 1989. Remote sensing of arid soil surface color with Landsat thematic mapper. Advances in Space Research, 9(1), 159-163.

Fang, B., Kansara, P., Dandridge, C., Lakshmi, V. 2021. Drought monitoring using high spatial resolution soil moisture data over Australia in 2015–2019. Journal of Hydrology, 594.

Fang, H., Liang, S. 2014. Leaf Area Index Models. in: Reference Module in Earth Systems and Environmental Sciences.

Few, R., Tran, P.G. 2010. Climatic hazards, health risk and response in Vietnam: Case studies on social dimensions of vulnerability. Global Environmental Change, 20(3), 529-538.

Fletcher, T.D., Shuster, W., Hunt, W.F., Ashley, R., Butler, D., Arthur, S., Trowsdale, S., Barraud, S., Semadeni-Davies, A., Bertrand-Krajewski, J.-L., Mikkelsen, P.S., Rivard, G., Uhl, M., Dagenais, D., Viklander, M. 2014. SUDS, LID, BMPs, WSUD and more – The evolution and application of terminology surrounding urban drainage. Urban Water Journal, 12(7), 525-542.

Florinsky, I.V., Eilers, R.G., Manning, G.R., Fuller, L.G. 2002. Prediction of soil properties by digital terrain modelling. Environmental Modelling & Software, 17(3), 295-311.

Flörke, M., Kynast, E., Bärlund, I., Eisner, S., Wimmer, F., Alcamo, J. 2013. Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study. Global Environmental Change, 23(1), 144-156.

Forkuor, G., Hounkpatin, O.K.L., Welp, G., Thiel, M. 2017. High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. PLOS ONE, 12(1), e0170478.

Frampton, W.J., Dash, J., Watmough, G., Milton, E.J. 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. ISPRS Journal of Photogrammetry and Remote Sensing, 82, 83-92.

Freni, G., Mannina, G., Viviani, G. 2009. Uncertainty in urban stormwater quality modelling: the influence of likelihood measure formulation in the GLUE methodology. Sci Total Environ, 408(1), 138-45.

Fu, X., Shao, M., Wei, X., Horton, R. 2010. Soil organic carbon and total nitrogen as affected by vegetation types in Northern Loess Plateau of China. Geoderma, 155(1), 31-35.

Fujita, S. 1997. Measures to promote stormwater infiltration. Water Science and Technology, 36(8), 289-293.

Gago, J., Douthe, C., Coopman, R.E., Gallego, P.P., Ribas-Carbo, M., Flexas, J., Escalona, J., Medrano, H. 2015. UAVs challenge to assess water stress for sustainable agriculture. Agricultural Water Management, 153, 9-19.

Gain, A.K., Giupponi, C., Renaud, F.G. 2012. Climate Change Adaptation and Vulnerability Assessment of Water Resources Systems in Developing

Countries: A Generalized Framework and a Feasibility Study in Bangladesh. Water, 4(2), 345-366.

Gao, B.-c. 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment, 58(3), 257-266.

García, L., Barreiro-Gomez, J., Escobar, E., Téllez, D., Quijano, N., Ocampo-Martinez, C. 2015. Modeling and real-time control of urban drainage systems: A review. Advances in Water Resources, 85, 120-132.

Ge, X., Wang, J., Ding, J., Cao, X., Zhang, Z., Liu, J., Li, X. 2019. Combining UAV-based hyperspectral imagery and machine learning algorithms for soil moisture content monitoring. PeerJ, 7, e6926.

Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., Wolff, E. 2018. Very High Resolution Object-Based Land Use–Land Cover Urban Classification Using Extreme Gradient Boosting. IEEE Geoscience and Remote Sensing Letters, 15(4), 607-611.

Gholizadeh, A., Žižala, D., Saberioon, M., Borůvka, L. 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. Remote Sensing of Environment, 218, 89-103.

Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N. 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sensing of Environment, 58(3), 289-298.

Gomez, C., Viscarra Rossel, R.A., McBratney, A.B. 2008. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. Geoderma, 146(3), 403-411.

Gómez, D., Salvador, P., Sanz, J., Casanova, J.L. 2020. Modelling desert locust presences using 32-year soil moisture data on a large-scale. Ecological Indicators, 117.

Gómez, D., Salvador, P., Sanz, J., Rodrigo, J.F., Gil, J., Casanova, J.L. 2021. Prediction of desert locust breeding areas using machine learning methods and SMOS (MIR_SMNRT2) Near Real Time product. Journal of Arid Environments, 194.

Gujer, W., Krejci, V., Schwarzenbach, R., Zobrist, J. 1982. Von der Kanalisation ins Grundwasser - Charakterisierung eines Regenereignisses im Glattal. Gas-Wasser-Abwasser, 62(7), 298-311.

Guo, L., Sun, X., Fu, P., Shi, T., Dang, L., Chen, Y., Linderman, M., Zhang, G., Zhang, Y., Jiang, Q., Zhang, H., Zeng, C. 2021a. Mapping soil organic carbon stock by hyperspectral and time-series multispectral remote sensing images in low-relief agricultural areas. Geoderma, 398.

Guo, M., Ma, S., Wang, L.-J., Lin, C. 2021b. Impacts of future climate change and different management scenarios on water-related ecosystem services: A case study in the Jianghuai ecological economic Zone, China. Ecological Indicators, 127.

Guo, Y., He, J., Li, S., Zheng, G., Wang, L. 2020. Evaluating the feasibility of GF-1 remote sensing comparison with hyperspectral data for soil organic carbon prediction and mapping. IOP Conference Series: Earth and Environmental Science, 545, 012016.

Ha, N.T., Manley-Harris, M., Pham, T.D., Hawes, I. 2021. The use of radar and optical satellite imagery combined with advanced machine learning and metaheuristic optimization techniques to detect and quantify above ground biomass of intertidal seagrass in a New Zealand estuary. International Journal of Remote Sensing, 42(12), 4712-4738.

Hamzehpour, N., Shafizadeh-Moghadam, H., Valavi, R. 2019. Exploring the driving forces and digital mapping of soil organic carbon using remote sensing and soil texture. Catena, 182.

Hancock, J.T., Khoshgoftaar, T.M. 2020. CatBoost for big data: an interdisciplinary review. Journal of Big Data, 7(1), 94.

Harremoës, P., Madsen, H. 1999. Fiction and reality in the modelling world – Balance between simplicity and complexity, calibration and identifiability, verification and falsification. Water Science and Technology, 39(9), 1-8.

Hatfield, J.L., Dold, C. 2019. Water-Use Efficiency: Advances and Challenges in a Changing Climate. Frontiers in Plant Science, 10(103).

He, M., Xiong, X., Wang, L., Hou, D., Bolan, N.S., Ok, Y.S., Rinklebe, J., Tsang, D.C.W. 2021. A critical review on performance indicators for evaluating soil biota and soil health of biochar-amended soils. J Hazard Mater, 414, 125378.

Heanes, D.L. 1984. Determination of total organic-C in soils by an improved chromic acid digestion and spectrophotometric procedure. Communications in Soil Science and Plant Analysis, 15(10), 1191-1213.

Heusch, S., Kamradt, B., Ostrowski, M. 2010. Simulation of wastewater treatment plant within integrated urban wastewater models. Water Science and Technology, 61(10), 2645-2652.

Hosoda, M., Tokonami, S., Suzuki, T., Janik, M. 2020. Machine learning as a tool for analysing the impact of environmental parameters on the radon exhalation rate from soil. Radiation Measurements, 138.

Hou, J., Mao, H., Li, J., Sun, S. 2019. Spatial simulation of the ecological processes of stormwater for sponge cities. J Environ Manage, 232, 574-583.

Hoyt, P.B. 1966. Chlorophyll-type compounds in soil. Plant and Soil, 25(2), 167-180.

Hu, M., Zhang, X., Siu, Y., Li, Y., Tanaka, K., Yang, H., Xu, Y. 2018. Flood Mitigation by Permeable Pavements in Chinese Sponge City Construction. Water, 10(2).

Huang, J.J., Li, Y., Niu, S., Zhou, S.H. 2014. Assessing the performances of low impact development alternatives by long-term simulation for a semi-arid area in Tianjin, northern China. Water Science and Technology, 70(11), 1740-1745.

Huete, A.R. 1988. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment, 25(3), 295-309.

Ibrahem Ahmed Osman, A., Najah Ahmed, A., Chow, M.F., Feng Huang, Y., El-Shafie, A. 2021. Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. Ain Shams Engineering Journal, 12(2), 1545-1556.

Isundwa, K.F., Mourad, K.A. 2019. The potential for water stewardship partnership in Kenya. Arabian Journal of Geosciences, 12(12).

Iyengar, N.S., Sudarshan, P. 1982. A Method of Classifying Regions from Multivariate Data. Economic and Political Weekly, 17(51), 2047-2052.

Jamali, B., Löwe, R., Bach, P.M., Urich, C., Arnbjerg-Nielsen, K., Deletic, A. 2018. A rapid urban flood inundation and damage assessment model. Journal of Hydrology, 564, 1085-1098.

Jia, H., Wang, Z., Zhen, X., Clar, M., Yu, S.L. 2017. China's sponge city construction: A discussion on technical approaches. Frontiers of Environmental Science & Engineering, 11(4).

Jia, H., Yao, H., Tang, Y., Yu, S.L., Field, R., Tafuri, A.N. 2015. LID-BMPs planning for urban runoff control and the case study in China. Journal of Environmental Management, 149, 65-76.

Jiang, C., Li, J., Li, H., Li, Y., Chen, L. 2017a. Field Performance of Bioretention Systems for Runoff Quantity Regulation and Pollutant Removal. Water, Air, and Soil Pollution, 228(12).

Jiang, Y., Zevenbergen, C., Fu, D. 2017b. Understanding the challenges for the governance of China's "sponge cities" initiative to sustainably manage urban stormwater and flooding. Natural Hazards, 89(1), 521-529.

Jiang, Y., Zevenbergen, C., Ma, Y. 2018. Urban pluvial flooding and stormwater management: A contemporary review of China's challenges and "sponge cities" strategy. Environmental Science & Policy, 80, 132-143.

Jun, K.S., Chung, E.-S., Sung, J.-Y., Lee, K.S. 2011. Development of spatial water resources vulnerability index considering climate change impacts. Science of The Total Environment, 409(24), 5228-5242.

Justice, C.O., Townshend, J.R.G., Vermote, E.F., Masuoka, E., Wolfe, R.E., Saleous, N., Roy, D.P., Morisette, J.T. 2002. An overview of MODIS Land data processing and product status. Remote Sensing of Environment, 83(1), 3-15.

Kamali, M., Delkash, M., Tajrishy, M. 2017. Evaluation of permeable pavement responses to urban surface runoff. Journal of Environmental Management, 187, 43-53.

Karthikeyan, L., Mishra, A.K. 2021. Multi-layer high-resolution soil moisture estimation using machine learning over the United States. Remote Sensing of Environment, 266.

Kerr, Y.H., Waldteufel, P., Wigneron, J., Martinuzzi, J., Font, J., 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), 1729-1735.

Khaleghi, B., Khamis, A., Karray, F.O., Razavi, S.N. 2013. Multisensor data fusion: A review of the state-of-the-art. Information Fusion, 14(1), 28-44.

Khosravi, K., Sartaj, M., Tsai, F.T., Singh, V.P., Kazakis, N., Melesse, A.M., Prakash, I., Tien Bui, D., Pham, B.T. 2018. A comparison study of DRASTIC methods with various objective methods for groundwater vulnerability assessment. Sci Total Environ, 642, 1032-1049.

Kingwell, R., Islam, N., Xayavong, V. 2020. Farming systems and their business strategies in south-western Australia: A decadal assessment of their profitability. Agricultural Systems, 181, 102827.

Kogan, F.N. 1995. Application of vegetation index and brightness temperature for drought detection. Advances in Space Research, 15(11), 91-100.

Kopecký, M., Macek, M., Wild, J. 2021. Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition. Science of The Total Environment, 757, 143785.

Kumar, K., Kozak, J., Hundal, L., Cox, A., Zhang, H., Granato, T. 2016. In-situ infiltration performance of different permeable pavements in a employee used parking lot – A four-year study. Journal of Environmental Management, 167, 8-14.

Lal, R. 2008. Carbon sequestration. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 363(1492), 815-830.

Laurenson, G., Laurenson, S., Bolan, N., Beecham, S., Clark, I. 2013. Chapter Four - The Role of Bioretention Systems in the Treatment of Stormwater. in: Advances in Agronomy, (Ed.) L.S. Donald, Vol. Volume 120, Academic Press, pp. 223-274.

Le Luu, T. 2019. Remarks on the current quality of groundwater in Vietnam. Environmental Science and Pollution Research, 26(2), 1163-1169.

Le, N.N., Pham, T.D., Yokoya, N., Ha, N.T., Nguyen, T.T.T., Tran, T.D.T., Pham, T.D. 2021. Learning from multimodal and multisensor earth observation dataset for improving estimates of mangrove soil organic carbon in Vietnam. International Journal of Remote Sensing, 42(18), 6866-6890.

Lee, D.B. 1973. Requiem for Large-Scale Models. Journal of the American Institute of Planners, 39(3), 163-178.

Lee Joong, G., Heaney James, P. 2003. Estimation of Urban Imperviousness and its Impacts on Storm Water Systems. Journal of Water Resources Planning and Management, 129(5), 419-426.

Leempoel, K., Parisod, C., Geiser, C., Daprà, L., Vittoz, P., Joost, S. 2015. Very high-resolution digital elevation models: are multi-scale derived variables ecologically relevant? Methods in Ecology and Evolution, 6(12), 1373-1383.

Li, H., Ding, L., Ren, M., Li, C., Wang, H. 2017a. Sponge City Construction in China: A Survey of the Challenges and Opportunities. Water, 9(12).

Li, J., Davis, A.P. 2016. A unified look at phosphorus treatment using bioretention. Water Research, 90, 141-155.

Li, J., Zhang, B., Mu, C., Chen, L. 2018a. Simulation of the hydrological and environmental effects of a sponge city based on MIKE FLOOD. Environmental Earth Sciences, 77(2), 32.

Li, Q., Wang, F., Yu, Y., Huang, Z., Li, M., Guan, Y. 2019. Comprehensive performance evaluation of LID practices for the sponge city construction: A case study in Guangxi, China. J Environ Manage, 231, 10-20.

Li, X., Li, J., Fang, X., Gong, Y., Wang, W. 2016. Case Studies of the Sponge City Program in China. in: World Environmental and Water Resources Congress 2016, pp. 295-308.

Li, Z.-p., Long, Y.-q., Tang, P.-q., Tan, J.-y., Li, Z.-g., Wu, W.-b., Hu, Y.-n., Yang, P. 2017b. Spatio-temporal changes in rice area at the northern limits of the rice cropping system in China from 1984 to 2013. Journal of Integrative Agriculture, 16(2), 360-367.

Li, Z., Xu, S., Yao, L. 2018b. A systematic literature mining of Sponge city: Trends, foci and challenges standing ahead. Sustainability (Switzerland), 10(4).

Liang, X. 2018. Integrated economic and financial analysis of China's Sponge City Program for water-resilient urban development. Sustainability (Switzerland), 10(3).

Lijklema, L., Tyson, J.M., Lesouef, A. 1993. Interactions between sewers treatment plants and receiving waters in urban areas: A summary of the Interurba '92 workshop conclusions. Water Science and Technology, 27(12), 1-29.

Liu, A., Goonetilleke, A., Egodawatta, P. 2015. Stormwater Treatment Design. in: Role of Rainfall and Catchment Characteristics on Urban Stormwater Quality, pp. 15-30.

Liu, D. 2016. China's sponge cities to soak up rainwater. Nature, 537, 307.

Liu, H., Jia, Y., Niu, C. 2017. “Sponge city” concept helps solve China’s urban water problems. Environmental Earth Sciences, 76(14), 1-5.

Liu, M., Yang, L., Bai, Y., Min, Q. 2018a. The impacts of farmers’ livelihood endowments on their participation in eco-compensation policies: Globally important agricultural heritage systems case studies from China. Land Use Policy, 77, 231-239.

Liu, Y., Gupta, H., Springer, E., Wagener, T. 2008. Linking science with environmental decision making: Experiences from an integrated modeling approach to supporting sustainable water resources management. Environmental Modelling & Software, 23(7), 846-858.

Liu, Y., Li, T., Peng, H. 2018b. A new structure of permeable pavement for mitigating urban heat island. Sci Total Environ, 634, 1119-1125.

Löwe, R., Urich, C., Sto. Domingo, N., Mark, O., Deletic, A., Arnbjerg-Nielsen, K. 2017. Assessment of urban pluvial flood risk and efficiency of adaptation options through simulations – A new generation of urban planning tools. Journal of Hydrology, 550, 355-367.

Luu, C., von Meding, J., Mojtahedi, M. 2019. Analyzing Vietnam's national disaster loss database for flood risk assessment using multiple linear regression-TOPSIS. International Journal of Disaster Risk Reduction, 40.

Ma, G., Ding, J., Han, L., Zhang, Z., Ran, S. 2021. Digital mapping of soil salinization based on Sentinel-1 and Sentinel-2 data combined with machine learning algorithms. Regional Sustainability, 2(2), 177-188.

Ma, Z., Hu, J., Feng, P., Gao, Q., Qu, S., Song, W., Liu, J. 2017. Assessment of Climate Technology Demands in Chinese Sponge City. Journal of Geoscience and Environment Protection, 05(12), 102-116.

Makropoulos, C.K., Natsis, K., Liu, S., Mittas, K., Butler, D. 2008. Decision support for sustainable option selection in integrated urban water management. Environmental Modelling & Software, 23(12), 1448-1460.

Mangangka, I.R., Liu, A., Egodawatta, P., Goonetilleke, A. 2015. Performance characterisation of a stormwater treatment bioretention basin. Journal of Environmental Management, 150, 173-178.

Mannina, G., Freni, G., Viviani, G., Saegrov, S., Hafskjold, L.S. 2006. Integrated urban water modelling with uncertainty analysis. Water Sci Technol, 54(6-7), 379-86.

Mannina, G., Viviani, G. 2010. An urban drainage stormwater quality model: Model development and uncertainty quantification. Journal of Hydrology, 381(3–4), 248-265.

Mao, X., Jia, H., Yu, S.L. 2017. Assessing the ecological benefits of aggregate LID-BMPs through modelling. Ecological Modelling, 353, 139-149.

Marlow, D.R., Moglia, M., Cook, S., Beale, D.J. 2013. Towards sustainable urban water management: A critical reassessment. Water Research, 47(20), 7150-7161.

Marsalek, J., Barnwell, T.O., Geiger, W., Grottker, M., Huber, W.C., Saul, A.J., Schilling, W., Torno, H.C. 1993. Urban drainage systems: Design and operation. Water Science and Technology, 27(12), 31-70.

Mathieu, R., Pouget, M., Cervelle, B., Escadafal, R. 1998. Relationships between Satellite-Based Radiometric Indices Simulated Using Laboratory Reflectance Data and Typic Soil Color of an Arid Environment. Remote Sensing of Environment, 66(1), 17-28.

McNicoll, G. 2007. United Nations Development Programme: Human Development Report 2006. Beyond Scarcity: Power, Poverty and the Global Water Crisis. Population and Development Review, 33, 198+.

Mei, C., Liu, J., Wang, H., Yang, Z., Ding, X., Shao, W. 2018. Integrated assessments of green infrastructure for flood mitigation to support robust decision-making for sponge city construction in an urbanized watershed. Sci Total Environ, 639, 1394-1407.

Mentens, J., Raes, D., Hermy, M. 2006. Green roofs as a tool for solving the rainwater runoff problem in the urbanized 21st century? Landscape and Urban Planning, 77(3), 217-226.

Michalski, A., Czajewski, J. 2004. The accuracy of the global positioning systems. IEEE Instrumentation & Measurement Magazine, 7(1), 56-60.

Mirzaee, S., Ghorbani-Dashtaki, S., Mohammadi, J., Asadi, H., Asadzadeh, F. 2016. Spatial variability of soil organic matter using remote sensing data. CATENA, 145, 118-127.

Mondal, A., Khare, D., Kundu, S., Mondal, S., Mukherjee, S., Mukhopadhyay, A. 2017. Spatial soil organic carbon (SOC) prediction by regression kriging using remote sensing data. The Egyptian Journal of Remote Sensing and Space Science, 20(1), 61-70.

Moriasi, D.N., Arnold, J.G., Liew, M.W.V., Bingner, R.L., Harmel, R.D., Veith, T.L. 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Transactions of the ASABE, 50(3), 885-900.

Morisette, J.T., Privette, J.L., Justice, C.O. 2002. A framework for the validation of MODIS Land products. Remote Sensing of Environment, 83(1), 77-96.

Murphy, P.N.C., Ogilvie, J., Castonguay, M., Zhang, C.-f., Meng, F.-R., Arp, P.A. 2008. Improving forest operations planning through high-resolution flow-channel and wet-areas mapping. The Forestry Chronicle, 84(4), 568-574.

Muthanna, T.M., Viklander, M., Blecken, G., Thorolfsson, S.T. 2007. Snowmelt pollutant removal in bioretention areas. Water Research, 41(18), 4061-4072.

Navarro-Pedreño, J., Almendro-Candel, M.B., Zorpas, A.A. 2021. The Increase of Soil Organic Matter Reduces Global Warming, Myth or Reality? Sci, 3(1), 18.

Ngo, T.T., Le, N.T., Hoang, T.M., Luong, D.H. 2018. Water Scarcity in Vietnam: a Point of View on Virtual Water Perspective. Water Resources Management, 32(11), 3579-3593.

Nguyen, T.G., de Kok, J.L., Titus, M.J. 2007. A new approach to testing an integrated water systems model using qualitative scenarios. Environmental Modelling & Software, 22(11), 1557-1571.

Nguyen, T.T., Ngo, H.H., Guo, W., Wang, X.C., Ren, N., Li, G., Ding, J., Liang, H. 2018. Implementation of a specific urban water management - Sponge City. Sci Total Environ, 652, 147-162.

Nikolakopoulos, K.G. 2020. Accuracy assessment of ALOS AW3D30 DSM and comparison to ALOS PRISM DSM created with classical photogrammetric techniques. European Journal of Remote Sensing, 53(sup2), 39-52.

Nkwunonwo, U.C., Whitworth, M., Baily, B. 2016. Review article: A review and critical analysis of the efforts towards urban flood risk management in the Lagos region of Nigeria. Natural Hazards and Earth System Sciences, 16(2), 349-369.

Norrman, J., Sparrenbom, C.J., Berg, M., Nhan, D.D., Nhan, P.Q., Rosqvist, H., Jacks, G., Sigvardsson, E., Baric, D., Moreskog, J., Harms-Ringdahl, P., Hoan, N.V. 2008. Arsenic mobilisation in a new well field for drinking water production along the Red River, Nam Du, Hanoi. Applied Geochemistry, 23(11), 3127-3142.

Pahl-Wostl, C. 2007. Transitions towards adaptive management of water facing climate and global change. Water Resources Management, 21(1), 49-62.

Paul, S.S., Coops, N.C., Johnson, M.S., Krzic, M., Chandna, A., Smukler, S.M. 2020. Mapping soil organic carbon and clay using remote sensing to predict soil workability for enhanced climate change adaptation. Geoderma, 363.

Petersen, E.H., Hoyle, F.C. 2016. Estimating the economic value of soil organic carbon for grains cropping systems in Western Australia. Soil Research, 54(4), 383-396.

Pham, T.D., Yokoya, N., Nguyen, T.T.T., Le, N.N., Ha, N.T., Xia, J., Takeuchi, W., Pham, T.D. 2020. Improvement of Mangrove Soil Carbon Stocks

Estimation in North Vietnam Using Sentinel-2 Data and Machine Learning Approach. GIScience & Remote Sensing, 58(1), 68-87.

Plummer, R., de Loë, R., Armitage, D. 2012. A Systematic Review of Water Vulnerability Assessment Tools. Water Resources Management, 26(15), 4327-4346.

Prasad, R., Deo, R.C., Li, Y., Maraseni, T. 2018. Soil moisture forecasting by a hybrid machine learning technique: ELM integrated with ensemble empirical mode decomposition. Geoderma, 330, 136-161.

Qin, H.P., Li, Z.X., Fu, G. 2013. The effects of low impact development on urban flooding under different rainfall characteristics. J Environ Manage, 129, 577-85.

R Brown, R., Keath, N., Wong, T. 2009. Urban Water Management in Cities: Historical, Current and Future Regimes.

Randelovic, A., Zhang, K., Jacimovic, N., McCarthy, D., Deletic, A. 2016. Stormwater biofilter treatment model (MPiRe) for selected micro-pollutants. Water Research, 89, 180-191.

Rauch, W., Bertrand-Krajewski, J.L., Krebs, P., Mark, O., Schilling, W., Schütze, M., Vanrolleghem, P.A. 2002. Deterministic modelling of integrated urban drainage systems. in: Water Science and Technology, Vol. 45, pp. 81-94.

rauch, W., Seggelke, K., Brown, R., Krebs, P. 2005. Integrated Approaches in Urban Storm Drainage: Where Do We Stand? Environmental Management, 35(4), 396-409.

Rauch, W., Urich, C., Bach, P.M., Rogers, B.C., de Haan, F.J., Brown, R.R., Mair, M., McCarthy, D.T., Kleidorfer, M., Sitzenfrei, R., Deletic, A. 2017. Modelling transitions in urban water systems. Water Research, 126, 501-514.

Ren, N., Wang, Q., Wang, Q., Huang, H., Wang, X. 2017. Upgrading to urban water system 3.0 through sponge city construction. Frontiers of Environmental Science & Engineering, 11(4), 9.

Reza, H., Davoud, Z., Mohammad Reza, N., Bakhtiar, F., Mehdi, R. 2020. Modification on optical trapezoid model for accurate estimation of soil moisture content in a maize growing field. Journal of Applied Remote Sensing, 14(3), 1-19.

Ritchie, H., Roser, M. (2020). Water Use and Stress Our world in Data.

Robinson, D.A., Campbell, C.S., Hopmans, J.W., Hornbuckle, B.K., Jones, S.B., Knight, R., Ogden, F., Selker, J., Wendroth, O. 2008. Soil Moisture Measurement for Ecological and Hydrological Watershed-Scale Observatories: A Review. Vadose Zone Journal, 7(1), 358-389.

Rouse, J., Haas, R.H., Schell, J.A., Deering, D. 1973. Monitoring vegetation systems in the great plains with ERTS.

Roy, A.H., Wenger, S.J., Fletcher, T.D., Walsh, C.J., Ladson, A.R., Shuster, W.D., Thurston, H.W., Brown, R.R. 2008. Impediments and solutions to sustainable, watershed-scale urban stormwater management: lessons from Australia and the United States. Environ Manage, 42(2), 344-59.

Running, S.W., Thornton, P.E., Nemani, R., Glassy, J.M. 2000. Global Terrestrial Gross and Net Primary Productivity from the Earth Observing System. in: Methods in Ecosystem Science, (Eds.) O.E. Sala, R.B. Jackson, H.A. Mooney, R.W. Howarth, Springer New York. New York, NY, pp. 44-57.

Rycewicz-Borecki, M., McLean, J.E., Dupont, R.R. 2017. Nitrogen and phosphorus mass balance, retention and uptake in six plant species grown in stormwater bioretention microcosms. Ecological Engineering, 99, 409-416.

Sadeghfam, S., Hassanzadeh, Y., Nadiri, A.A., Zarghami, M. 2016. Localization of Groundwater Vulnerability Assessment Using Catastrophe Theory. Water Resources Management, 30(13), 4585-4601.

Sailor, D.J. 2008. A green roof model for building energy simulation programs. Energy and Buildings, 40(8), 1466-1478.

Šakić Trogrlić, R., Rijke, J., Dolman, N., Zevenbergen, C. 2018. Rebuild by Design in Hoboken: A Design Competition as a Means for Achieving Flood

Resilience of Urban Areas through the Implementation of Green Infrastructure. Water, 10(5).

Salim, R.A., Islam, N. 2010. Exploring the impact of R&D and climate change on agricultural productivity growth: the case of Western Australia*. Australian Journal of Agricultural and Resource Economics, 54(4), 561-582.

Salvadore, E., Bronders, J., Batelaan, O. 2015. Hydrological modelling of urbanized catchments: A review and future directions. Journal of Hydrology, 529, Part 1, 62-81.

Santhi, C., Arnold, J.G., Williams, J.R., Dugas, W.A., Srinivasan, R., Hauck, L.M. 2001. VALIDATION OF THE SWAT MODEL ON A LARGE RWER BASIN WITH POINT AND NONPOINT SOURCES1. JAWRA Journal of the American Water Resources Association, 37(5), 1169-1188.

Sargent, R.G. 1991. Simulation model verification and validation. 1991 Winter Simulation Conference Proceedings., 8-11 Dec. 1991. pp. 37-47.

Schaffer, D., Vollmer, D. 2010. Pathways to urban sustainability: Research and development on urban systems.

Schellart, A.N., Tait, S.J., Ashley, R.M. 2010. Towards quantification of uncertainty in predicting water quality failures in integrated catchment model studies. Water Res, 44(13), 3893-904.

Schmidt, A., Mainwaring, D.B., Maguire, D.A. 2020. Development of a tailored combination of machine learning approaches to model volumetric soil water content within a mesic forest in the Pacific Northwest. Journal of Hydrology, 588.

Scholz, M. 2006. Best management practice: A Sustainable Urban Drainage System management case study. Water International, 31(3), 310-319.

Scholz, M., Uzomah, V.C. 2013. Rapid decision support tool based on novel ecosystem service variables for retrofitting of permeable pavement systems in the presence of trees. Science of The Total Environment, 458-460, 486-498.

Schönauer, M., Väätäinen, K., Prinz, R., Lindeman, H., Pszenny, D., Jansen, M., Maack, J., Talbot, B., Astrup, R., Jaeger, D. 2021. Spatio-temporal prediction of soil moisture and soil strength by depth-to-water maps. International Journal of Applied Earth Observation and Geoinformation, 105.

Schütze, M., Butler, D., Beck, M.B. 1999. Optimisation of control strategies for the urban wastewater system — An integrated approach. Water Science and Technology, 39(9), 209-216.

Sedlak, D. 2014. Water 4.0

The Past, Present, and Future of the World's Most Vital Resource. Yale University Press.

Senanayake, I.P., Yeo, I.Y., Walker, J.P., Willgoose, G.R. 2021. Estimating catchment scale soil moisture at a high spatial resolution: Integrating remote sensing and machine learning. Science of The Total Environment, 776.

Seppelt, R., Müller, F., Schröder, B., Volk, M. 2009. Challenges of simulating complex environmental systems at the landscape scale: A controversial dialogue between two cups of espresso. Ecological Modelling, 220(24), 3481-3489.

Shafique, M., Kim, R., Rafiq, M. 2018. Green roof benefits, opportunities and challenges – A review. Renewable and Sustainable Energy Reviews, 90, 757-773.

Shao, W., Zhang, H., Liu, J., Yang, G., Chen, X., Yang, Z., Huang, H. 2016. Data Integration and its Application in the Sponge City Construction of CHINA. Procedia Engineering. pp. 779-786.

Sheffield, J., Wood, E.F., Pan, M., Beck, H., Coccia, G., Serrat-Capdevila, A., Verbist, K. 2018. Satellite Remote Sensing for Water Resources Management: Potential for Supporting Sustainable Development in Data-Poor Regions. Water Resources Research, 54(12), 9724-9758.

Shi, T., Zhang, X., Du, H., Shi, H. 2015. Urban water resource utilization efficiency in China. Chinese Geographical Science, 25(6), 684-697.

Six, J., Elliott, E.T., Paustian, K., Doran, J.W. 1998. Aggregation and Soil Organic Matter Accumulation in Cultivated and Native Grassland Soils. Soil Science Society of America Journal, 62(5), 1367-1377.

Sriwastava Ambuj, K., Tait, S., Schellart, A., Kroll, S., Dorpe Mieke, V., Assel Johan, V., Shucksmith, J. 2018. Quantifying Uncertainty in Simulation of Sewer Overflow Volume. Journal of Environmental Engineering, 144(7), 04018050.

Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Lioy, R., Hoffmann, L., van Wesemael, B. 2010. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. Geoderma, 158(1), 32-45.

Stovin, V. 2010. The potential of green roofs to manage Urban Stormwater. Water and Environment Journal, 24(3), 192-199.

Swain, S., Wardlow, B.D., Narumalani, S., Rundquist, D.C., Hayes, M.J. 2013. Relationships between vegetation indices and root zone soil moisture under maize and soybean canopies in the US Corn Belt: a comparative study using a close-range sensing approach. International Journal of Remote Sensing, 34(8), 2814-2828.

Taugourdeau, S., le Maire, G., Avelino, J., Jones, J.R., Ramirez, L.G., Jara Quesada, M., Charbonnier, F., Gómez-Delgado, F., Harmand, J.-M., Rapidel, B., Vaast, P., Roupsard, O. 2014. Leaf area index as an indicator of ecosystem services and management practices: An application for coffee agroforestry. Agriculture, Ecosystems & Environment, 192, 19-37.

Thanh Thi Pham, N., Nong, D., Raghavan Sathyan, A., Garschagen, M. 2020. Vulnerability assessment of households to flash floods and landslides in the poor upland regions of Vietnam. Climate Risk Management, 28.

Thorndahl, S., Beven, K.J., Jensen, J.B., Schaarup-Jensen, K. 2008. Event based uncertainty assessment in urban drainage modelling, applying the GLUE methodology. Journal of Hydrology, 357(3-4), 421-437.

Thornes, J.E. 2002. IPCC, 2001: Climate change 2001: impacts, adaptation and vulnerability, Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change, edited by J. J. McCarthy, O. F. Canziani, N. A. Leary, D. J. Dokken and K. S. White (eds). Cambridge University Press, Cambridge, UK, and New York, USA, 2001. No. of pages: 1032. Price: £34.95, ISBN 0-521-01500-6 (paperback), ISBN 0-521-80768-9 (hardback). International Journal of Climatology, 22(10), 1285-1286.

Trabucco, A. 2010. High- Resolution Global Soil-Water Balance Explicit for Climate-Standard Vegetation and Soil Conditions. Consortium for Spatial Information.

Trabucco, A., J.Zomer, R. 2018. Global Aridity Index and Potential Evapo-Transpiration (ET0) Climate Database v2.

Trowsdale, S.A., Simcock, R. 2011. Urban stormwater treatment using bioretention. Journal of Hydrology, 397(3–4), 167-174.

Tscheikner-Gratl, F., Bellos, V., Schellart, A., Moreno-Rodenas, A., Muthusamy, M., Langeveld, J., Clemens, F., Benedetti, L., Rico-Ramirez, M.A., de Carvalho, R.F., Breuer, L., Shucksmith, J., Heuvelink, G.B.M., Tait, S. 2019. Recent insights on uncertainties present in integrated catchment water quality modelling. Water Res, 150, 368-379.

Tucker, C.J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment, 8(2), 127-150.

Tuia, D., Volpi, M., Copa, L., Kanevski, M., Munoz-Mari, J. 2011. A Survey of Active Learning Algorithms for Supervised Remote Sensing Image Classification. IEEE Journal of Selected Topics in Signal Processing, 5(3), 606-617.

Tuller, M., Babaeian, E., Jones, S., Montzka, C., Vereecken, H., Sadeghi, M. 2019. The paramount societal impact of soil moisture. Eos, 100.

Urich, C., Bach, P.M., Sitzenfrei, R., Kleidorfer, M., McCarthy, D.T., Deletic, A., Rauch, W. 2013. Modelling cities and water infrastructure dynamics. Proceedings of the Institution of Civil Engineers - Engineering Sustainability, 166(5), 301-308.

van de Meene, S.J., Brown, R.R., Farrelly, M.A. 2011. Towards understanding governance for sustainable urban water management. Global Environmental Change, 21(3), 1117-1127.

Vanrolleghem, P.A., Rosen, C., Zaher, U., Copp, J., Benedetti, L., Ayesa, E., Jeppsson, U. 2005. Continuity-based interfacing of models for wastewater systems described by Petersen matrices. in: Water Science and Technology, Vol. 52, pp. 493-500.

Vaudour, E., Gomez, C., Fouad, Y., Lagacherie, P. 2019. Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. Remote Sensing of Environment, 223, 21-33.

Venter, Z.S., Hawkins, H.J., Cramer, M.D., Mills, A.J. 2021. Mapping soil organic carbon stocks and trends with satellite-driven high resolution maps over South Africa. Sci Total Environ, 771, 145384.

Vohland, M., Ludwig, M., Thiele-Bruhn, S., Ludwig, B. 2017. Quantification of Soil Properties with Hyperspectral Data: Selecting Spectral Variables with Different Methods to Improve Accuracies and Analyze Prediction Mechanisms. Remote Sensing, 9(11).

Voinov, A., Shugart, H.H. 2013. 'Integronsters', integral and integrated modeling. Environmental Modelling & Software, 39, 149-158.

Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., Davies, P.M. 2010. Global threats to human water security and river biodiversity. Nature, 467(7315), 555-561.

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldaña, J., de Rosnay, P., Jann, A., Schneider, S., Komma, J., Kubu, G., Brugger, K., Aubrecht, C., Züger, J., Gangkofner, U., Kienberger, S., Brocca, L., Wang, Y., Blöschl, G., Eitzinger, J., Steinnocher, K. 2013. The ASCAT Soil Moisture Product: A Review of its Specifications, Validation Results, and Emerging Applications. Meteorologische Zeitschrift, 22(1), 5-33.

Wang, B., Waters, C., Orgill, S., Gray, J., Cowie, A., Clark, A., Liu, L. 2018a. High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. Sci Total Environ, 630, 367-378.

Wang, J., Zhang, P., Yang, L., Huang, T. 2016. Cadmium removal from urban stormwater runoff via bioretention technology and effluent risk assessment for discharge to surface water. Journal of Contaminant Hydrology, 185–186, 42-50.

Wang, M., Sweetapple, C., Fu, G., Farmani, R., Butler, D. 2017a. A framework to support decision making in the selection of sustainable drainage system design alternatives. J Environ Manage, 201, 145-152.

Wang, Q., Gong, Y.-w., Zhang, W.-w., Li, J.-q., Zhang, X.-b. 2017b. Operation and Maintenance of Source Runoff Control Facilities. Zhongguo ji Shui pai Shui = China Water & Wastewater(7).

Wang, S., Lin, X., Yu, H., Wang, Z., Xia, H., An, J., Fan, G. 2017c. Nitrogen removal from urban stormwater runoff by stepped bioretention systems. Ecological Engineering, 106, Part A, 340-348.

Wang, X., Li, T., Yang, X., Zhang, T., Liu, Z., Guo, E., Liu, Z., Qu, H., Chen, X., Wang, L., Xiang, H., Lai, Y. 2018b. Rice yield potential, gaps and constraints during the past three decades in a climate-changing Northeast China. Agricultural and Forest Meteorology, 259, 173-183.

Warner, D.L., Guevara, M., Callahan, J., Vargas, R. 2021. Downscaling satellite soil moisture for landscape applications: A case study in Delaware, USA. Journal of Hydrology: Regional Studies, 38.

Wijesiri, B., Egodawatta, P., McGree, J., Goonetilleke, A. 2016. Assessing uncertainty in stormwater quality modelling. Water Research, 103, 10-20.

Woods Ballard, B., Wilson, S., Udale-Clarke, H., Illman, S., Scott, T., Ashley, R., Kellagher, R. 2015. The SuDS Manual, C753. CIRIIA.

Wu, W.-H., Wang, S.-S., Li, J.-Q., Xin, J.-G., Zhang, Z.-M., Liu, C. 2017. Discussion on Key Technical Issues of the Sponge City Implementation Program:A Case Study of Ningbo City. China Water & Wastewater(6).

Wu, Y. 2015. Sponge City Design: Concept, Technology & Case Study. Jiangsu: Phoenix Science Press.

Xia, J., Shi, W., Wang, Q., Zou, L. 2017a. Discussion of several hydrological issues regarding sponge city construction. Water Resources Protection, 33(1), 1-8.

Xia, J., Zhang, Y., Xiong, L., He, S., Wang, L., Yu, Z. 2017b. Opportunities and challenges of the Sponge City construction related to urban water issues in China. Science China Earth Sciences, 60(4), 652-658.

Xie, H., Zhao, J., Wang, Q., Sui, Y., Wang, J., Yang, X., Zhang, X., Liang, C. 2015. Soil type recognition as improved by genetic algorithm-based variable selection using near infrared spectroscopy and partial least squares discriminant analysis. Scientific Reports, 5(1), 10930.

Xu, S., Zhao, Y., Wang, M., Shi, X. A comparison of machine learning algorithms for mapping soil iron parameters indicative of pedogenic processes by hyperspectral imaging of intact soil profiles. European Journal of Soil Science, n/a(n/a).

Xu, W., Yu, J. 2017. A novel approach to information fusion in multi-source datasets: A granular computing viewpoint. Information Sciences, 378, 410-423.

Xue, J., Su, B. 2017. Significant remote sensing vegetation indices: A review of developments and applications. Journal of Sensors, 2017.

Yang, R.-M., Guo, W.-W. 2019. Modelling of soil organic carbon and bulk density in invaded coastal wetlands using Sentinel-1 imagery. International Journal of Applied Earth Observation and Geoinformation, 82, 101906.

Yu, M.M., Zhu, J.W., Gao, W.F., Xu, D.P., Zhao, M. 2017. Urban permeable pavement system design based on "sponge city" concept. IOP Conference Series: Earth and Environmental Science, 82.

Zhang, K., Chui, T.F.M. 2019. Linking hydrological and bioecological benefits of green infrastructures across spatial scales – A literature review. Science of The Total Environment, 646, 1219-1231.

Zhang, S., Zevenbergen, C., Rabé, P., Jiang, Y. 2018a. The Influences of Sponge City on Property Values in Wuhan, China. Water, 10(6).

Zhang, X., Shen, L., Wu, Y. 2011. Green strategy for gaining competitive advantage in housing development: a China study. Journal of Cleaner Production, 19(2), 157-167.

Zhang, Y., Lu, H., Luo, S., Sun, Z., Qu, W. 2017a. Human-Scale Sustainability Assessment of Urban Intersections Based upon Multi-Source Big Data. Sustainability, 9(7).

Zhang, Y., Xiao, X., Wu, X., Zhou, S., Zhang, G., Qin, Y., Dong, J. 2017b. A global moderate resolution dataset of gross primary production of vegetation for 2000-2016. Sci Data, 4, 170165.

Zhang, Z., Szota, C., Fletcher, T.D., Williams, N.S.G., Werdin, J., Farrell, C. 2018b. Influence of plant composition and water use strategies on green roof stormwater retention. Science of The Total Environment, 625, 775-781.

Zhao, Y.-B., Yang, M.-Z., Ni, H.-G. 2018. An emergy-GIS method of selecting areas for sponge-like urban reconstruction. Journal of Hydrology, 564, 640-650.

Zhao, Z., Yang, Q., Ding, X., Xing, Z. 2021. Model Prediction of the Soil Moisture Regime and Soil Nutrient Regime Based on DEM-Derived Topo-Hydrologic Variables for Mapping Ecosites. Land, 10(5).

Zhou, T., Geng, Y., Chen, J., Liu, M., Haase, D., Lausch, A. 2020a. Mapping soil organic carbon content using multi-source remote sensing variables in the Heihe River Basin in China. Ecological Indicators, 114.

Zhou, T., Geng, Y., Chen, J., Pan, J., Haase, D., Lausch, A. 2020b. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. Sci Total Environ, 729, 138244.

Zhou, T., Geng, Y., Chen, J., Pan, J., Haase, D., Lausch, A. 2020c. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. Science of The Total Environment, 729, 138244.

Zhou, T., Geng, Y., Ji, C., Xu, X., Wang, H., Pan, J., Bumberger, J., Haase, D., Lausch, A. 2021a. Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images. Sci Total Environ, 755(Pt 2), 142661.

Zhou, T., Geng, Y., Ji, C., Xu, X., Wang, H., Pan, J., Bumberger, J., Haase, D., Lausch, A. 2021b. Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images. Science of The Total Environment, 755, 142661.

Zischg, J., Rogers, B., Gunn, A., Rauch, W., Sitzenfrei, R. 2019. Future trajectories of urban drainage systems: A simple exploratory modeling approach for assessing socio-technical transitions. Sci Total Environ, 651(Pt 2), 1709-1719.

Zomorodian, M., Lai, S.H., Homayounfar, M., Ibrahim, S., Fatemi, S.E., El-Shafie, A. 2018. The state-of-the-art system dynamics application in integrated water resources modeling. Journal of Environmental Management, 227, 294-304.

Zuo, Q. 2016. Water science issues in sponge city construction. Water Resources Protection, 32(4), 21-26.

Appendix

Check correlation between soil moisture and 21 bands from Genetic Algorithm optimizer using Pearson method

Coefficient of cor. between SM and RVI
-0.38912423970947996
Coefficient of cor. between SM and NDVI
-0.40155658771960934
Coefficient of cor. between SM and GNDVI
-0.24923957987188916
Coefficient of cor. between SM and SAVI
-0.4996274357238523
Coefficient of cor. between SM and MCARI
-0.07048246430859731
Coefficient of cor. between SM and IRECl
-0.5684716701919917
Coefficient of cor. between SM and VH
-0.4135030521838582
Coefficient of cor. between SM and VV
-0.3466834725915791
Coefficient of cor. between SM and MeanVHVV
-0.40313845774971785
Coefficient of cor. between SM and b1_VV_GLCM
-0.32750946927696445
Coefficient of cor. between SM and b2_VV_GLCM
-0.3817347601647576
Coefficient of cor. between SM and b3_VV_GLCM
0.40077800241505585
Coefficient of cor. between SM and b4_VV_GLCM
0.3321303804303876
Coefficient of cor. between SM and b5_VV_GLCM
0.35191870985308066
Coefficient of cor. between SM and b6_VV_GLCM
0.3111939550945679
Coefficient of cor. between SM and b7_VV_GLCM
-0.37735842031841016
Coefficient of cor. between SM and b10_VV_GLC
0.3106859627766614
Coefficient of cor. between SM and DEM
-0.61578668107054
Coefficient of cor. between SM and Slope
-0.4954060408813305
Coefficient of cor. between SM and TWI
0.3679195923006985
Coefficient of cor. between SM and NDWI
-0.3661351269701774