The Influence Of (shifts in) Environmental Factors On Land Use Change In The Mekong Delta

Dissertation submitted in partial fulfillment of the requirements for the Degree of Master of Science in Geospatial Technologies

February, 2021

David Simduwa Alsina
dsimduwa@uni-muenster.de

Supervised by:
Judith Verstegen, PhD.
Institute for Geoinformatics
University of Münster

Co-supervised by:
Philip Minderhoud, PhD.
Department of Civil, Environmental
and Architectural Engineering (ICEA)
University of Padova

Co-supervised by:
Ana Costa, PhD.
Information Management School
Universidade Nova de Lisboa

Declaration of Academic Integrity

I hereby confirm that this thesis on *The Influence Of (shifts in) Environmental Factors On Land Use Change In The Mekong Delta* is solely my own work and that I have used no sources or aids other than the ones stated. All passages in my thesis for which other sources, including electronic media, have been used, be it direct quotes or content references, have been acknowledged as such and the sources cited.

Febraury 24, 2021

I agree to have my thesis checked in order to rule out potential similarities with other works and to have my thesis stored in a database for this purpose.

Febraury 24, 2021

Acknowledgements

First of all, I want to thank my supervisor **Prof. Judith Verstegen** who has constantly guided and helped me throughout my thesis. I have learnt so much in this time and can not put in words how valuable this experience has been for me, i am eternally grateful for the support i have received from my supervisor. I would also like to thank my two co-supervisors **Prof. Philip Minderhoud** and **Prof. Ana Costa**, whose comments have helped improve greatly in my work. The work i have done in the last couple of months would not be possible without the support of my supervisors and i am very grateful to have had the opportunity to work with such a great team.

To the teachers and staff in IFGI, UJI and NOVA, thank you so much for all your efforts and support even during a pandemic to make sure my classmates and i receive quality education. I would like to thank my classmates for always being kind and helpful to me throughout my study, I could´t have asked for better colleagues. To my friends **Raül Raga**, **Anu, Janak, Ganesh**, **Emeka Igwe**, **Jeremiah Olowe**,**Abubakar Baralafia**, **Tarek** and many more i am not able to mention here, thank you for being such amazing friends and my support system. I would like to specially thank my best friends **Poshan Niraula** and **David Payares**, for always being there for me, i am blessed beyond words to have someone like you as friends.

To my family, i have no words to explain just how much you mean to me, you are my world and i owe everything to you. To my brothers **Emmanuel** and **Francisco**, thank you for being the best brothers anyone can ask for. To my lovely sisters **Dr. Marianne, Tamara** and **Isabella**, thank you for your unending love and support, I love you. Finally, to my parents, you have been everything to me and i can never thank you enough. I hope that i continue to make you proud always, I love you.

I dedicate this work to my
nephew Nathan Alberto and my niece Elena.
Nothing makes me more proud than being your uncle.
I love you.

Contents

List of Tables

List of Figures

List of Acronyms

AUC	Area under the curve
DEM	Digital Elevation Model
FAO	Food and Agriculture Organization
GIS	Geographic Information System
GSOV	General Statistics Office of Vietnam
GLM	Generalized Linear Model
ROC	Receiver Operator Curve
VMD	Vietnamese Mekong Delta

Abstract

Land use is influenced by factors (drivers) such as elevation, climate, sea level, population growth etc. These factors themselves are not constant and are changing with time. The change in these factors could result in a shift in their influence on land-use change. In land use studies so far, the focus has been on land-use change and not the state of the factors. Thus, this study focuses on the significance of drivers of land-use change between 1988 – 2009 in the Vietnamese Mekong delta. To assess how land-use change has been influenced by the shifts in environmental factors and the relationship between the observed shifts in influence and land-use change. The Vietnamese Mekong delta is one of the largest and most studied delta region in the world due to rapid increase in population and large-scale transformations driven by economic expansion in the last decades.

Five land use types are considered in this study, these are; aquaculture, mixed crops, orchard, rice and urban. Each land use class was modelled using a logistic regression with the following environmental variables; compaction, elevation, population density, distance to river and sea, salinity and subsidence. The study period was divided into four; 1988 – 2006, 1988 – 1996, 1996 - 2006 and 2006 – 2009 to be able to observe a trend in land-use change.

The findings of this study reveal that the significance of the selected environmental variables to land-use change is not constant over the study period. Some variables were not significant to land-use change in one or more periods, became significant in other periods or remained insignificant. The findings in this study provides decision makers with a better understanding on sustainable land-use planning in the study area and similar regions in the world.

Keywords: bootstrap, environmental drivers, land use, land-use change, logistic regression.

Chapter 1

Introduction

1.1 Context and Motivation

Humans interact with the environment in different ways, to sum up in two main points, as a place of abode (a place to live) and as a resource supply (a place to earn a living) (Oldfield & Dearing, 2003). These human – environment interactions can also be seen as how humans use the land and this can have a direct or indirect alteration to the natural state of the environment (Smith, 2013). These alterations can have a global or local effect on the environment. Global warming as a result of emission of greenhouse gasses can be seen as example of a global alteration and occur over a long period of time. A change in land cover such as urban expansion into forest land cover can be seen as an example of a local alteration which can occur over shorter time periods as compared to global warming (Crain et al., 2009). The changes in the environment ultimately plays a part on how the environment is being used. For example, agricultural land use plays vital role in global food sustainability, environmental changes such as droughts, subsidence, rise in sea level etc. have an impact on expansion and production of this land use (Anderson et al., 2020).

Land-use change is influenced by several factors, these factors can be summed up into two main categories, physical (e.g. elevation, climate etc.) and social factors (e.g. social policies etc.) (Briassoulis, 2009). The factors that influence land-use change can operate at different levels (micro or macro) and in some cases are interdependent on one another, for instance an increase in subsidence rates resulting in an increase in influence of sea level rise on land use in coastal regions(Briassoulis, 2009). The factors influencing land-use change are not constant themselves and as such their influence on land-use change can also be affected by their change of state. For example, the influence of climate on agricultural land use is key for expansion and yield of products. The changes in global climate in the last few decades has affected agricultural land use across different regions in the world (Van Meijl et al., 2018; Mendelsohn, 2008).

The shifts in environmental factors and the influence on land-use change has not been given much attention in literature's so far. It remains unclear as to how much the shifts in environmental factors are driving land-use change. Thus, these two points will guide this study, 1) Does the shifts in environmental factors influence the significance of these factors to land use change over time or remain constant, 2) In what direction (positive or negative) are the shifts in environmental variables driving land use change. For example, increased subsidence rates leading to loss of elevation, increases the influence of elevation to land use. (Minderhoud et al., 2019; Minderhoud et al., 2018).

This study considers the following environmental variables; compaction, elevation, population density, distance to river and sea, salinity and subsidence as explanatory variables. The explanatory variables will be modelled using a logistic regression approach with these land use types; aquaculture, mixed crops, orchard, rice and urban.

The focus of this study is on land-use change driven by shifts in environmental variables between 1988 - 2009 in the Vietnamese Mekong delta. If the shifts in environmental factors influence on land-use change, it should be accounted for, to give a better understanding of the factors driving land-use change.

1.2 Related Works

Recently, due to technological advancement in the field of remote sensing, the use of remote sensing and GIS techniques has become very popular, especially in the field of land use studies. It has been widely used in studying land use trends in different parts of the world (Abdi, 2020; Joshi et al., 2016; Adami et al., 2012; Reis, 2008; Chen et al., 2006). Apart from the this, the increase in use of remotely sensed data can also be attributed to the availability of large-scale open source satellite data (Forster, 2010). In a study carried out by Weng (2002), the author also used satellite data from the Landsat 5 TM for a time scale of between 1989 to 1997 to produce three land use maps for Zhujiang delta in China, to monitor the rapid land use change resulting from industrialization and urbanization. The author used a stochastic modelling technique to compare the results for the different time scales. The findings of his research showed that the urban and horticulture land use types increased the most within the time frame the study. It also states that the use of Landsat 5 TM for modelling land use change can be a generally successful.

The Vietnamese Mekong delta has been the center stage for numerous scientific research across multiple disciplines, making it one of the most studied delta region in the world due to its geographical location and pivotal role it plays in the region when it comes to food security and economic importance (Q. H. Nguyen et al., 2020; Minderhoud, 2017; Fujihara et al., 2016; Takagi et al., 2016; Smajgl et al., 2015). These studies have been key to identifying some of the most critical environmental problems in this region, examples include relating land use to land subsidence (Minderhoud et al., 2018), identifying strategies and policies necessary to reduce increase in salinity due to rise in sea level (Smajgl et al., 2015), analyzing the Spatio-temporal dynamics of land cover and land cover in the Vietnamese Mekong delta (Tran, 2015), anthropogenic drivers of relative sea level rise (Parker, 2020), influence of population growth on natural resources (Pech & Sunada, 2008). In a recent study by Minderhoud et al. (2018), the authors used satellite images from Landsat 5 TM for a time frame of 1988 – 2009 to assess the level of land subsidence in the Vietnamese Mekong delta based on the land use type. The output of their research showed that the urban land use contributes the most to land subsidence at a rate of about -18mm per year. This finding can be attributed to the rapid increase in population and urbanization in this region over the last few decades (GSOV, 2019).

Agricultural land use is one of the most important land uses globally, it is the source of livelihood to millions around the world and responsible for food sustainability of the entire world, this makes it an important topic of research, especially in areas faced with climate crisis. The Mekong delta is considered as one of the vulnerable regions to climate change due to it locations and topography, studies have shown an increase in land subsidence due to the impacts of groundwater extraction of between 1 – 4 cm yr-1 and estimates of between 0.35m – 1.4m of land subsidence by the year 2050 at current extraction rates (Erban, Gorelick, Zebker, 2014; Minderhoud et al., 2017). In the research by Smajgl et al., (2015), the author used a river modelling application (MIKE 11) to explore the threat on paddy rice production from rise in sea level and salinization and the came up with recommendations of development of reservoirs and irrigation schemes to support seasons of low flow and construct dykes and gates in strategic locations of rivers around the Mekong basin to prevent flooding and reduce salinization.

Numerous approaches have been developed across multiple scientific disciplines to model land-use change (Briassoulis, 2019; Verburg et al., 2004). The use of statistical and econometric approaches such as regression in land use studies is common (Briassoulis, 2019). The use of logistic regression to model land use change has been employed by some authors. In a research carried out in New castle county in Delware, the authors used a spatial logistic regression approach to model rural - urban land use change from the year 1984 - 1997. The authors used a sampling technique to reduce spatial dependence in the data and conflicts introduced in the results by spatial auto-correlation. The findings of their research showed the use of logistic regression over complex spatial models to model land use change is possible and can yield accurate results (Xie et al., 2005).

1.3 Aim and Objectives

The aim of this study is to investigate the influence of shifts in the following environmental factors; compaction, distance to river and sea, elevation, population density, salinity and subsidence on land-use change in the Vietnamese Mekong delta from 1988 - 2009. The objectives of this study are:

1. To compare the change in land use between 1988 - 2009 and derive where these land use type; aquaculture, mixed crops, orchards, rice and urban have expanded and contracted.

2. To use a regression analysis to assess the significance of compaction, distance to river and sea, elevation, population density, salinity and subsidence on land-use change for the following land use types; aquaculture, mixed crops, orchards, rice and urban.

3. To detect the direction in which the above mentioned environmental factors have driven land-use change for each of the land-use type mentioned above.

1.4 Research Questions

The following research questions will be answered in this study:

1. How can the significance of shifts in environmental variables on land-use change be assessed?
2. How has land-use change been driven by shifts in environmental variables between 1988 - 2009 in the Vietnamese Mekong Delta?

1.5 Study Area

The study is focused in the Vietnamese Mekong Delta (VMD), located in south-western Vietnam. The VMD is one of the largest delta region in the world, with an area of 40,816 km2 and a population of about 18 million people (GSOV, 2019).

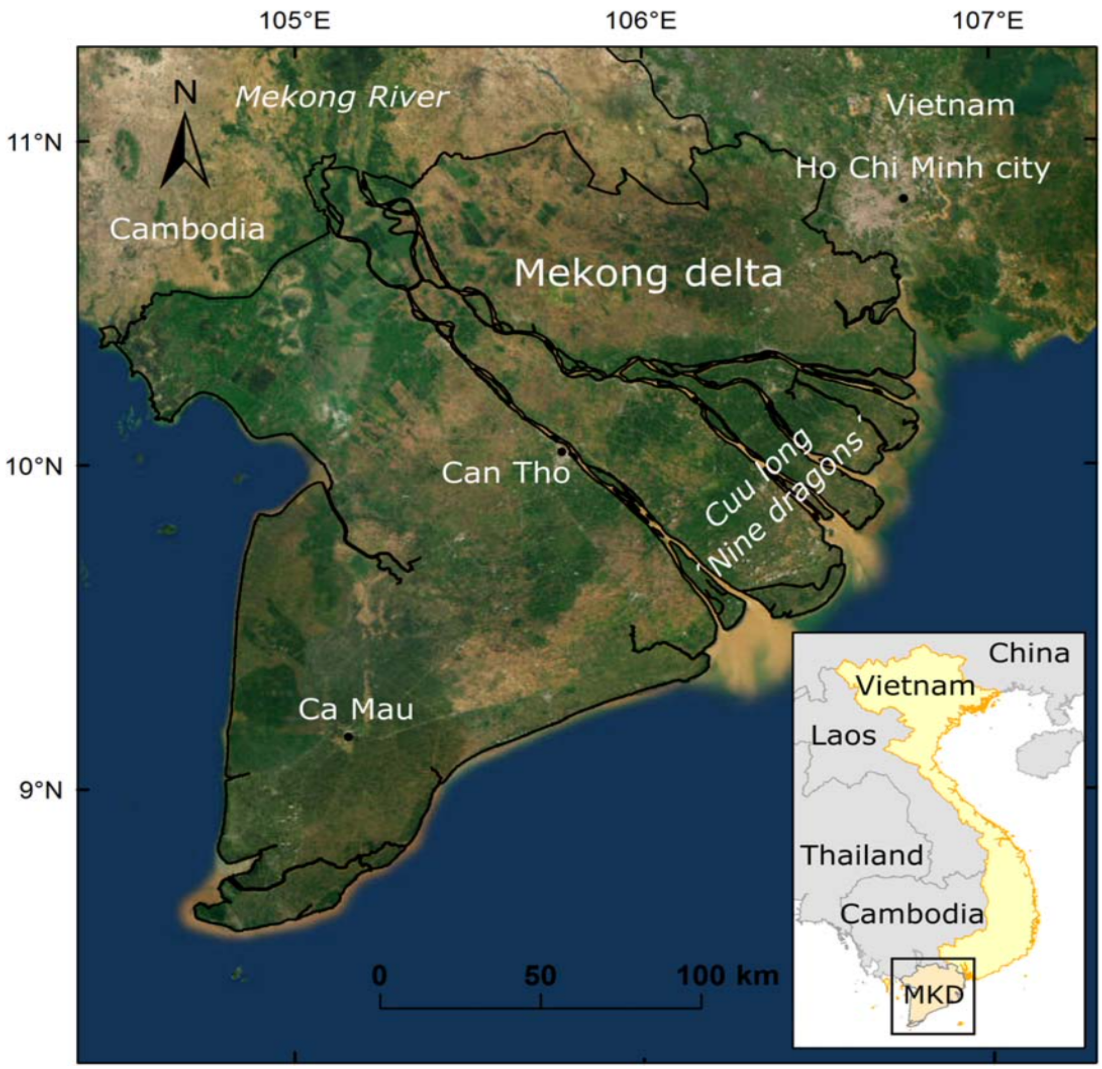

Figure 1.1: Map of the Vietnamese Mekong Delta
(Minderhoud et al., 2020)

The rich fertile soils found in the VMD has attracted more population and large-scale economic expansions driven by agricultural activities, which has led to rapid transformations in land use and land cover (Kuenzer & Knauer, 2013). The VMD is the biggest economic hub in Vietnam, known as the 'rice bowl' because of the agricultural productivity, producing more than 16 million tons of rice every year, with 70% of the population relying on agriculture as a source of living (FAO, 2018). The VMD is one of the most studied delta regions in the world and is considered to be one of the most vulnerable regions to the impact of environmental change such as flooding, land subsidence, sea-level rise etc. (Minderhoud et al., 2018; Fujihara et al., 2016; Takagi et al., 2016; Wassmann et al., 2004).

1.6 Thesis Outline

This thesis is organised as follows: Chapter 2 provides a theoretical background of the methods and concepts used in this thesis. In Chapter 3, the methodology of the model used is described in detail. Chapter 4 presents the data, experiment design and implementation of the model with data. In Chapter 5, the results of the findings are presented and discussed. The last chapter, will be for conclusions, limitations and comments on further studies from the findings in this thesis.

Chapter 2

Theoretical Background

This chapter is meant to serve as the theoretical foundation of the concepts and methods used in the thesis. The first section presents a brief explanation about the concepts of land, land use and factors that influence land use. The second part present the methods used such as logistic regression, bootstrapping, sampling and spatial auto-correlation and discussion about applications of this methods in land use change studies.

2.1 Land and Land Use Definitions

Land is one of the most important natural resource that supports human life. It is the platform that sustains most human activities (Diyer et al., 2013). There are numerous definitions of land in different literature, Cambridge dictionary defines land as "an area of ground, especially when used for a particular purpose such as farming or building". The Food and Agricultural Organization offers a concise definition of land as: *"Land is a delineable area of the earth's terrestrial surface, encompassing all attributes of the biosphere immediately above or below this surface, including those of the near-surface climate, the soil and terrain forms, the surface hydrology (including shallow lakes, rivers, marshes, and swamps), the near-surface sedimentary layers and associated groundwater reserve, the plant and animal populations, the human settlement pattern and physical results of past and present human activity (terracing, water storage or drainage structures, roads, buildings, etc.)"* (FAO, 1995). The UNCCD defines land as "the terrestrial bioproductive system that comprises soil, vegetation, other biota, and the ecological and hydrological processes that operate within the system" (UNCCD, 2017). It is clear what the meaning of land is and from the definitions above, we have an idea as to why land is important to human existence. The focus of our work is not is not merely on what land is but on land use change, thus we will look at how land use is defined in literature in the next section.

The definition of land use, just as that of land, has different definitions in different literatures, Jansen (2006) states that *"the term land use has different meaning across different disciplines"*. Di Gregorio and Jansen (2005) defined land use as *"the arrangements, activities and inputs people undertake in a certain land cover type to produce, change or maintain it"*. To define land use in simplest terms, land use is the purpose in which a land is being utilized (Dickinson & Shaw, 1977).

2.1.1 Land use classes

These land use classes are considered for this study aquaculture, mixed crops, orchard, rice and urban. These five land use classes were chosen because of their relevance in the study area. Each land use class will be discussed in detail and its relevance to the study area in the following paragraphs.

Aquaculture

The FAO defined aquaculture as *"the farming of aquatic organisms, including fish, molluscs, crustaceans, and aquatic plants. Farming implies some form of intervention in the rearing process to enhance production, such as regular stocking, feeding, protection from predators, etc. Farming also implies individual or corporate ownership of the stock being cultivated"* (Edwards & Demaine, 1997)

Aquaculture has quickly become an important land use in the Vietnamese Mekong delta, there has been a huge expansion of fishponds and commercial fish farming along the Mekong river since the late 1990´s, this has led to an increase in commercial fish of 3.6 times to 220, 615 tones in the year 2004 as compared to 1999 (Nguyen Thanh et al., 2007). More recently, as at the year 2018, Vietnam produces about 4 million metric tons of commercial fish per year, an increase of over 20 times more in a period of about 20 years as compared to the late 1990's and a projected further increase of up to 26% by the year 2030 (FAO, 2018). Vietnam is among the top 5 producers and exporters of commercial fish around the world, as at the year 2016, the total amounts of exports from aquacultural products by Vietnam lies at USD 7.3 billion, making it the third largest exporter of aquaculture products in the world after China and Norway (FAO, 2018).

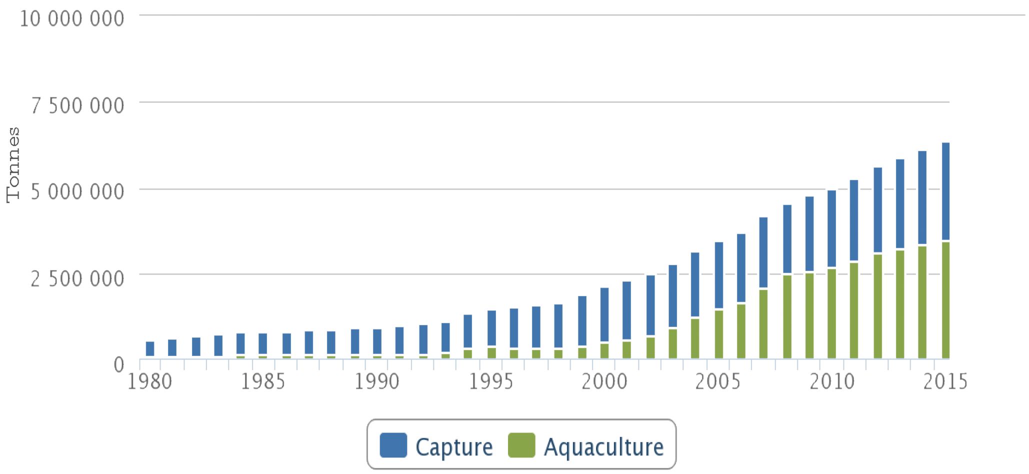

Figure 2.1: Total Capture and Aquaculture Production in Vietnam (FAO, 2018)

Aquaculture and commercial fish production are important part of the land use in the Vietnamese Mekong delta, there are two main species of commercial fish that are produced in this region, the black tiger shrimp (Penaeus monodon) and sutchi catfish (Pangasianodon hypophthalmus) (Nguyen Thanh et al., 2007).

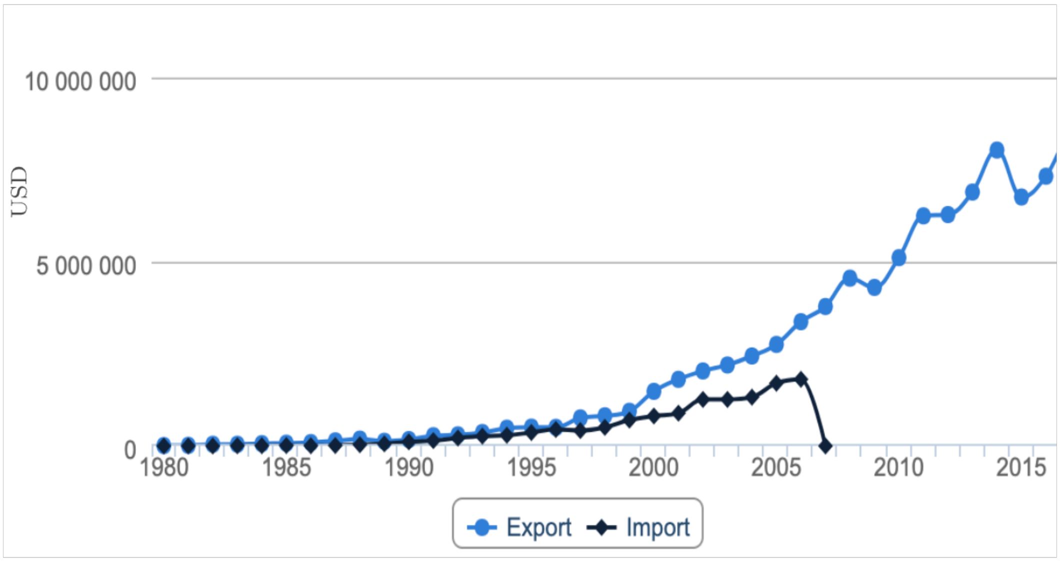

Figure 2.2: Total Imports and Exports of Fish and Fishery Related Products in Vietnam from 1980 - 2015

(FAO, 2018)

Mixed Crops

Mixed crops are land use types where more than one variety of crop is planted. Mixed cropping is an agricultural system that involves planting two or more crop type on the same piece of land. Mixed cropping is a cropping system that is practiced to maximise land productivity (Gliessman, 1985). Mixed cropping system is mostly practiced in tropical regions and the crop types are usually crops with different yield periods. It can be advantageous as more crops can be produced on same piece of land (Gliessman, 1985).

Orchard

The economy of the Mekong delta is mostly supported by agricultural related activities such rice farming, fisheries and fruit farming. Fruit production has grown rapidly with efforts to diversify the agricultural sector and cut down the over dependence on rice production in this region. The main fruits produced are bananas, coconuts and mangos. The exports from fruit products in Vietnam has grown considerably since the mid 1990's, in the year 1995 the revenue generated from fruit exports was USD 56 million while in 2007 it was USD 306 million and in the year 2013 it amounted to almost USD 3 billion (Thanh, Tan, Thu, 2013). The country earned over USD 3.2 billion in exports from fruits and vegetables in the year 2020 and a projected further expansion of 150,000ha of fruit harvesting areas between the year 2020 to 2030 (Vna, 2020). Fruit farming is quickly becoming an important resource for the Vietnamese economy, favorable climatic conditions for fruit growth has made it easy for expansion (Thanh et al., 2013).

Rice

Global food security is of utmost importance, with an increasing global population, the demand for food security has never been higher making it one of the sustainable development goals by the United Nations to end global hunger by the year 2030 (Johnston, 2016). Rice is grown in different parts of the world due to its high resistive character, it can grow in most tropical, sub-tropical and some temperate regions (Portmann et al., 2008). The Asian continent accounts for about 90% of all world rice production (Kuenzer & Knauer, 2013). Vietnam is among the top five rice producing countries in the world, producing an estimated 43 million tons as at the year 2017 (FAO, 2018). Agriculture is key to the Vietnamese economy, contributing to 24% of its GDP and 20% of its total exports (ADB, 2001). Rice production in Vietnam has always been a crucial source of livelihood for the population and has been expanding in the past years, with an average growth rate of 5.6% between 1990 – 1999 (ADB, 2001). Rice can grow under different water conditions, Rice-growing habitats can be grouped into four main classes based on water regime, irrigation, temperature, soil type and topography (Bambaradeniya & Amarasinghe, 2003). These habitats include, upland or dryland rice, rainfed lowland rice, irrigated lowland rice and deep-water and tidal wetland rice (Kuenzer & Knauer, 2013; Bambaradeniya & Amarasinghe, 2003). In the Vietnamese Mekong delta, only two of these rice systems are being practiced, the irrigated lowland and deep-water and tidal wetland systems. The irrigated lowland rice is the most important rice system in the world, accounting for 55% of world rice area and 75% of total rice production. Irrigated lowland rice has better yields because of availability of controlled drainage for almost 80% of the crop life. The duration of rice species can vary depending on the specie and the climatic conditions. The typical life cycle of rice from germination to maturity ranges between 3 – 6 months depending on environmental where it is grown, in the tropical regions it takes about 3 – 4 months (Kuenzer & Knauer, 2013). Five cropping seasons are observed annually in the Vietnamese Mekong delta based on the climatic conditions. The table below show the main cropping season and rice systems practiced.

Season	**Period (Start - End)**	**Cropping system**
Rainy season	July/August - December/January	1x Rain fed rice
Winter - spring	November/December – February/March	2x Irrigated rice
Spring – summer	March/April – May/June	2x Irrigated rice 3x Irrigated rice
Summer – autumn	April/May – July/August	2x Irrigated rice 2x Rain fed rice 3x Irrigated rice
Autumn – winter	July/September – October/December	2x Rain – fed rice

Table 2.1: Rice Cropping Systems and Seasons in the Vietnamese Mekong Delta (Kuenzer & Knauer, 2013)

Urban

The Vietnamese Mekong delta is one of the largest deltaic regions in the world and home to almost 18 million people (GSOV, 2019). The Mekong delta region is the largest economic hub in Vietnam and has attracted a rapid growth in population over the last decades (Tran, 2015). The increase in population has led to rapid urbanization rates in the region (Tran, 2015; Hoang et al., 2008). In the late 1970's after reunification, the Vietnamese economy has suffered from difficulties in inflation, supply and demand balance and rising debt (Jiwon Yun, 2019), the Vietnamese government decided to reform the country's economy, this transition led to a rapid growth in urbanization and industrialization in the country. The transition began in the mid 1980's when the government initiated a socio-economic policy called doi moi, this has since catapulted the country to one of the fastest urbanizing country in the south east Asia, with an estimated 37% of the population being urban. The boost in economic sector has led the country from a low income to a middle income and has seen the overall poverty rate drop drastically from 58% in 1993 to 4.5% in 2015 and a 650% increase in foreign direct investment between 2001 – 2006 (Revilla Diez, 2016; ADB, 2012). The rapid increase in the urban population also implies a rapid change in land use to urban land use, the most affected land use in the Vietnamese Mekong delta is the forest, it has decrease greatly since the 1980's when the Vietnamese economic expansion began, as of the year 1989 the forest area was estimated to be 152,778 ha and by the year 1995 the forest land was estimated to be 116,355ha, a 24% decrease, this reduction in forest land use coincides with the increase in urban land use (Liu et al., 2020). The urban population of the Vietnamese Mekong delta has grown at an average rate of 3.4% per annum as compared to the rural population at 0.4%, although majority of the delta still remains rural with about 76.6% of the delta still considered to be rural (Smith, 2013). The urban growth in this region however comes at price, Minderhoud et al. (2018) showed that the delta is subsiding and the urban land use contributes the most. Vietnam is considered as one of the prone areas to natural disasters because of its low laying nature, with most of the delta almost at sea level (Minderhoud et al., 2019) and its geographic location, exposing it to extreme weather events such as tropical storms and flooding (CFE-DM, 2018). Thus, the need for a sustainable urban land use expansion plan is very important.

2.2 Logistic Regression

Regression methods are among the most widely used statistical techniques in analyzing relationships between a response and one or more explanatory variables (Hosmer & Lemeshow, 2000). logistic regression is also called binomial, multinomial and ordinal logistic regression, based on the scale type of the response variable (Stephenson et al., 2008). Different types of regression methods exist, the application is dependent on the type of data that is being modelled and the expected output. The most commonly used regression methods are linear, log-linear and logistic regression (Xie et al., 2005). Logistic regression is part of the generalized linear models and uses a logit function in its response variable, between 0 and 1 which follows a binomial distribution (Sun Robinson, 2018).

A logistic regression is used to model categorical variables (i.e., change or no change). In this study, we use a logistic regression technique with a binomial output to model how land-use change has been driven by selected environmental variables. Literature has shown that logistic regression outperforms other statistical methods such as Markov chain and survival analysis (Sun & Robinson, 2018) in land use studies. Secondly, based on the focus of our work a categorical output would be necessary to explain if a land-use class changes or not, as compared to a continuous variable like in the case of other regression methods such as linear regression or multiple linear regression.

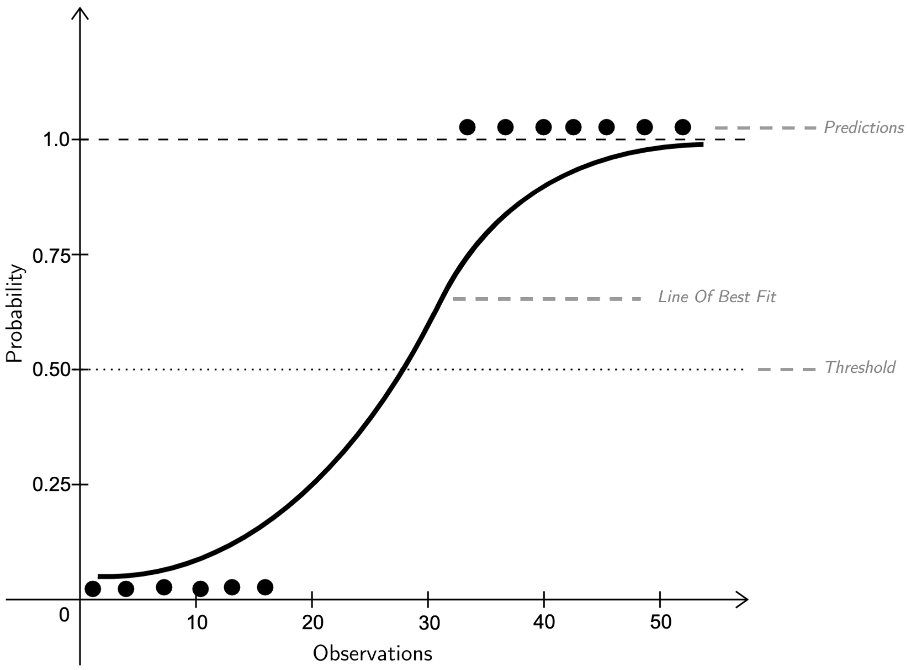

Figure 2.3: Logisitc Regression Diagram

Logistic regression analysis is used across different disciplines because of its flexibility, for example in land use change studies to model the probability of change of a land use from rural to urban (Tayyebi et al., 2010; Xie et al., 2005), in epidemiological studies to test for disease outbreak (Buckeridge et al., 2008). The logistic regression model takes two input variables, the response and the explanatory variable (Peng et al., 2002). The response variable in the model is the variable of interest, that is the variable that is trying to be explained or understood. The explanatory variable are the variables that are used to explain the probability of change of the response variable.

$$p(y) = \frac{1}{1 + e^{-\left(\beta 0 + \sum_{i=1}^{n} \beta ixi\right)}} \tag{2.1}$$

Where:
$p(y)$ is the probability of change of a land use type
x_i are the n explanatory variables
β_0 is the model intercept
β_1 are the coefficient of the explanatory variables (x_i)

The output parameters of the model are used to interpret the relationship between the response and explanatory variable. The model coefficients are the probability of change in the response variable per unit change of the explanatory variable, if all other predictors are constant. The coefficient is denoted with the (β) sign and it can be negative or positive. The model summary provides the significance of each explanatory variable to the probability of change of the response variable using the p-value. The p-value ranges between 0 and 1, the smaller the number the more significant the variable is and larger values indicates less significance. The threshold commonly used in literature's to determine significance is 0.05, which indicates a 95% confidence interval of the significance of the explanatory variable (Speelman, 2014).

2.3 Spatial Auto-Correlation

Spatial auto-correlation can be defined as the degree of similarity of observations within a referenced geographical space (Getis, 2008). In most cases when dealing with geographic data, to account for spatial auto-correlation in the data, two methods are used based on how interactions between cells are being modelled. The first approach is the use of spatial regression models, these models take into account the value of an observation and that of its neighbor by using weight matrix of neighboring observations also known as the spatial lag (Rey et al., 2011; Tu & Xia, 2008; Xie et al., 2005; Anselin, 1995).

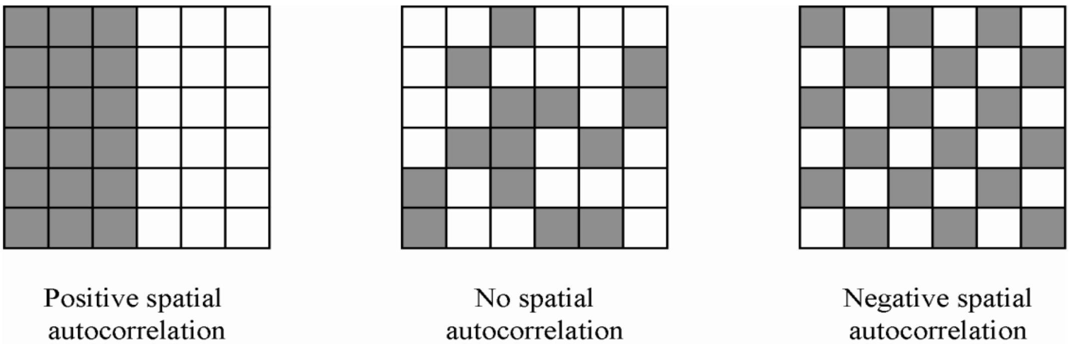

Figure 2.4: Spatial Auto-correlation
(Radil, 2011)

There are two ways of creating spatial weight matrices, distant based and weights based on boundaries. The second approach is the use of a non-spatial model and a sampling technique. This approach is usually used to reduce the spatial dependence in the data by reducing the number of observations and mostly adopted to reduce computing challenges when dealing with large data sets. This approach has been used by different authors in the field of land use studies (Nong & Du, 2011; Tayyebi et al., 2010; Xie et al., 2005).

Spatial auto-correlation in geographical data is common and different methods have been introduced in scientific literature on how to handle auto-correlation in spatial data (Getis, 2010, 2008; Url et al., 1970) . Often referred to as the first law of geography by Waldo Tobler, everything is related everything but near things are more related than distant things (Tobler, 1970). There are two main methods that are often used in literature to calculate the spatial auto-correlation in geographic data, the Moran's Index also known as Moran's I (Moran, 1950) and the Geary contiguity also known also Geary C (Society Statistician, 2016). Both these methods have been long established and developed (Rey et al., 2011; Getis, 2010; Bivand et al., 2009; Getis, 2008; Anselin, 1995). The Moran's I is the most commonly method for calculating spatial auto-correlation, it is calculated using spatial weight matrix, the spatial weight matrix calculates the neighbors of each observation in a data set, this can be done in different ways, by boundary (only takes a pixel into account as a neighbor when they share the same boundary), by distance (takes into account the number of neighbors in a given distance), by k numbers (only takes into account the number of nearest neighbors as defined by the user). The Moran's I equation is as follows:

$$I = \frac{N}{W} \frac{\sum_i \sum_{j\omega ij(x_i-\bar{x})(x_j-\bar{x})}}{\sum_i (x_i - \bar{x})^2} \tag{2.2}$$

Where:
I is the Moran's index,
N is the number of observations *(i & j)*,
W is the sum of all the spatial weights,
ω_{ij} is the spatial weight matrix,
$(x_i - \bar{x})$ is the deviation of the variable of interest,
$(x_j - \bar{x})$ is the deviation of the neighbor variable.

2.4 Bootstrapping Method

Bootstrapping is a random sampling and replacement technique that is used in statistical studies to generate confidence intervals and reduce bias on sampled data (Efron & Tibshirani, 1994). Sampling is a century old statistical approach of inferring or drawing conclusions about a population based on information about a part of it. English merchant John Graunt (1620 – 1674) was the first person to introduce the notion of sampling in his work where he describes a method to estimate the population of London based on partial information (Haan et al., 2009). Sampling methods have been used across numerous scientific disciplines to reduce the observation size which can be due to processing time, hardware requirements and has been used to reduce spatial dependence when dealing with spatial data (Xie et al., 2005). There are different types of sampling methods, the most common sampling methods are random and stratified sampling (Etikan, 2017). Although sampling can be advantageous, it is sometimes criticized because not all observations in a data are accounted for and can sometimes lead to bias or incomplete representation of the entire population. The application of a bootstrap technique is advantageous because it allows for more than one inclusion of a specific observation in a given sample and this can help to reduce the uncertainty that can be introduced when data is sampled (Brownlee, 2019). Bootstrapping is a method whereby samples are constructed by drawing observations one at a time from a data set and returning them to the data sample after they have been collected, an example of a bootstrap sample $x^* = (x_1, x_2, \dots x_n)$ is obtained by randomly sampling n times, with replacement from the original data points $x_1, x_2, \dots x_n$ (Johnson et al., 1989).

2.5 Model Evaluation

There are different statistical methods of validating regression models. The following methods were used to validate the model used for this study; a confusion matrix, Hosmer-Lemershow test, Psuedo R-squared and the Receiver Operator Curve (ROC) graph. These methods will be explained in this section and how they are applied to regression analysis to validate the output.

2.5.1 Confusion Matrix

A confusion matrix is a table that is used to validate prediction of classification models. It is used to test the performance of models by comparing predicted values with known values (Ting, 2010; Provost & Kohavi, 1998). From the confusion matrix, a Kappa coefficient is derived. The Kappa coefficient ranges from 0 to 1 (Landis & Koch, 1977).

The confusion matrix is also known as an error matrix, it has four parameters, True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN). The following statistics can be calculated based on the matrix parameters:

1. The accuracy: The total number of correct predictions divided by total number of observations.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{2.3}$$

2. The True Positive rate (TP) or recall: The number of positive observations that were correctly classified.

$$Recall = \frac{TP}{TP + FN} \tag{2.4}$$

3. The False Positive rate (FP): The number of negative observations that were classified as positive.

$$FP = \frac{FP}{TN + FP} \tag{2.5}$$

4. The True Negative rate (TN): The number of negative observations that were correctly classified as negative.

$$TN = \frac{TN}{TN + FP} \tag{2.6}$$

5. The False Negative rate (FN): The number of positive observations that were classified as negative.

$$FN = \frac{FN}{FN + TP} \tag{2.7}$$

6. The Precision: The number of positive observations that were correct.

$$precision = \frac{TP}{TP + FP} \tag{2.8}$$

	Positive	Negative
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

Figure 2.5: Confusion Matrix

2.5.2 Hosmer Lemeshow Test

The homser lemeshow test is a goodness of fit test for logistic regression models (Hosmer & Lemeshow, 2000). It is used to determine how well the model fits the data. The observations are grouped based on expected probabilities and compared with the observed. The test P-values shows how well the model fits the data, the P-values ranges between o to 1 (Shah & Barnwell, 2003). The smaller values suggest that the model does not fit the data well while larger values suggest that no discrepancies can be found on how the model fits the data.

2.5.3 R-squared

R-squared is a goodness of fit test for regression models, It is derived from the measurement of the cumulative percentage of variation in the dependent variable explained by the explanatory variable (D'Agostino, 2017; Peng et al., 2002). The R-squared value ranges between 0 and 1, with higher values indicating good fit and lower values indicating poor fit. The R-sqaured for non linear models is slightly different from that of linear models. The R-squared for categorical models such as logistic regression are called psuedo R-squared. Over the years different methods have been proposed for calculating R-squared for categorial models, the most commonly used one are the cox and snell's R-squared, Nagelkerke's R-squared and the McFadden's pseudo R-squared. The three mentioned methods are similar, they are calculated based on the log likelihood of the model (Nagelkerke et al., 1991; Cox & Snell, 1989; McFadden et al., 1973).

2.5.4 ROC curve

The receiver operating characteristics (ROC) is a graph that uses varied thresholds to explain the predictive ability of a binary classifier at different thresholds. The graph is generated by plotting the true positive rate against the false positive rate at all classification thresholds (Fawcett, 2006). The true positive rate is derived by dividing the true positive by the true positive plus the false negative observations, as shown in equation 2.4 above. The false positives are the observations that were classified as positive but were negative. The false positive rate is derived by dividing the false positives by the true negatives plus false positive observations, as shown in equation 2.5 above. The area under curve (AUC) was the total area under the ROC curve. Higher AUC's signifies a better classification model (Jones & Athanasiou, 2005; Hanley & McNeil, 1982).

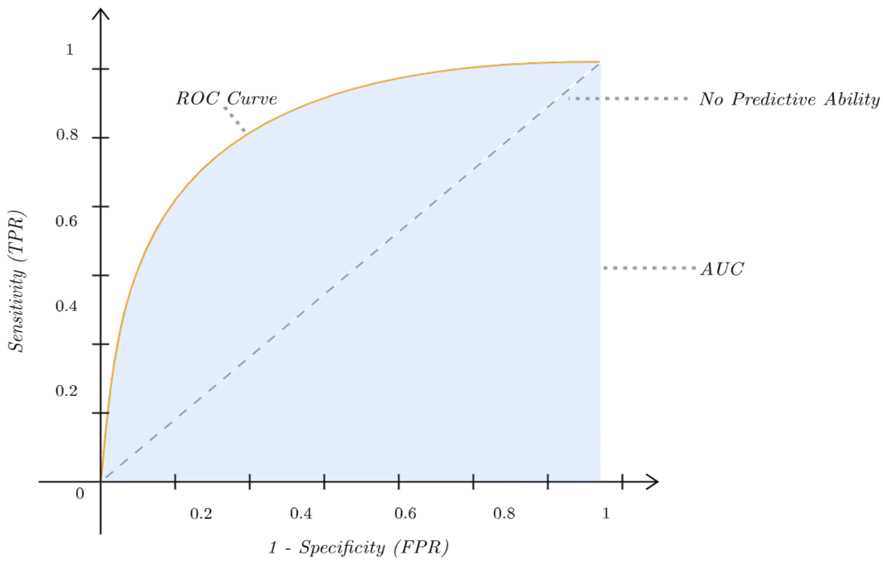

Figure 2.6: Receiver Operating Characteristics Curve

Chapter 3

Methodology

This chapter presents the logistic regression model and describes the methods used in this study. These methods are based on the theories explained in Chapter 2. The chapter is structured as follows, an overview of the model used, the sampling method used, the bootstrap technique, the model inputs, logistic regression and finally the output of the model.

3.1 Overview

This study focuses on the significance of shifts in selected environmental variables and how they have driven land-use change, to assess if their significance remains constant or not between 1988 - 2009 in the Vietnamese Mekong delta. To achieve this, we use regression approach to model the change per land use type (i.e., the response variable) against selected environmental variables (i.e., the explanatory variables). To compute the change per land use class, we use a time series of four land use maps of the Vietnamese Mekong delta, from 1988 to 2009 from the study of (Minderhoud et al., 2018). The explanatory variables are; compaction, distance to river sea, elevation, population density and salinity. These variables are all in raster formats of different scales. All the raster files were re-sampled to a pixel size of a 100 x 100m. Two separate data frames were constructed, one for land use expansion and the other for contraction and each data frame was used in a regression model. Separate models were created to be able to assess which explanatory variables (drivers) are significant to the probability of land use expansion or contraction in the study area within the study period. The data was sampled into two, 70% for training and the remaining 30% for validation. For each model run, a bootstrap approach was used with the training data to run the model for n number of times, this was done to generate confidence intervals of the model output. A post regression analysis was carried out to evaluate how well the model fits the data and the model performance was evaluated based on the predictions. A Hoslem-Lemeshow test was carried out to assess if the model fits the data. The test is evaluated based on the p value, higher p value indicate that there is no evidence that the model fits the data poorly, while lower p values indicate a poor model fit. An ROC curve was created to evaluate the performance of the model based on the correctly predicted observations when compared to the test data. Finally, a pseudo R-squared was calculated based on the model fitted value to see how well the model fits the training data.

Figure 3.1 shows the overview of the model used, the input to the logistic regression model are the land use change classes as response variable and the environmental variable as explanatory. The model output are the probability of change per land use class and that is validate using the test data.

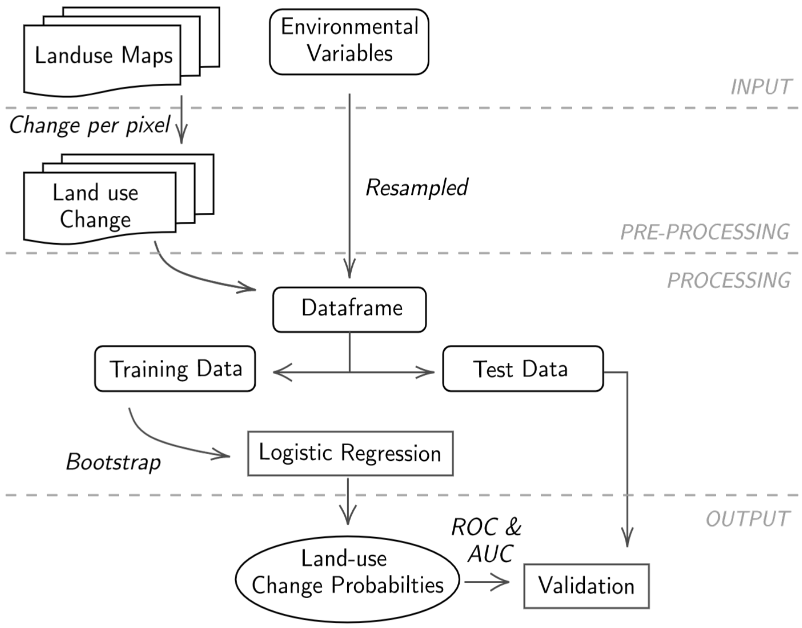

Figure 3.1: Methodological Work-flow

3.2 Model Input

The input to the models are the land-use change per class as response variable and the environmental variables and explanatory variables. As seen in figure 3.1 the land use maps were used to compute the land-use change per class and a data frame with the environmental variables was created and used in the model. The land-use change per class at each period was computed based on the pixel values using the land use maps and used as the response variable in the model. The explanatory variables are compaction, elevation, population density, distance to river and sea, salinity and subsidence of the study area. The details of the data used in the model and pre-processing steps are explained in the next sub-sections

Land-use Change Computation

As an input to the models used, land use change per class is computed by comparing the land use maps of the study area between the period of the study. This is done by comparing two land use maps, one at time step t_0 and the other at t_1. The pixel values are then compared between the two time frames and if a land use type is present in time step t_0 and not in t_1, it has contracted, while if it is present in t_1 but not present in t_0, it has expanded. In the case that a land use remains present in both t_0 and t_1, there is no change and if it is not present in both t_0 and t_1, that land use type has never been in that region.

In figure 3.2 below, an example of rice land-use change is shown. The rice land use is represented by the green pixels, while non-rice is represented by the brown pixels. The red pixels represents pixels where the rice land use has contracted and the orange pixels represents where the rice land use has expanded based on the explanation in the paragraph above. The gray pixels represent pixels that have never been rice land use, while the green pixels represents the pixels that are rice land use in both t_0 and t_1. Two data frames were created, one for land use expansion and the other for land use contraction. For the land use expansion data frame, every pixel where a land use class has expanded will be represented by 1 and where it has not will be 0, the same was done for the land use contraction data frame.

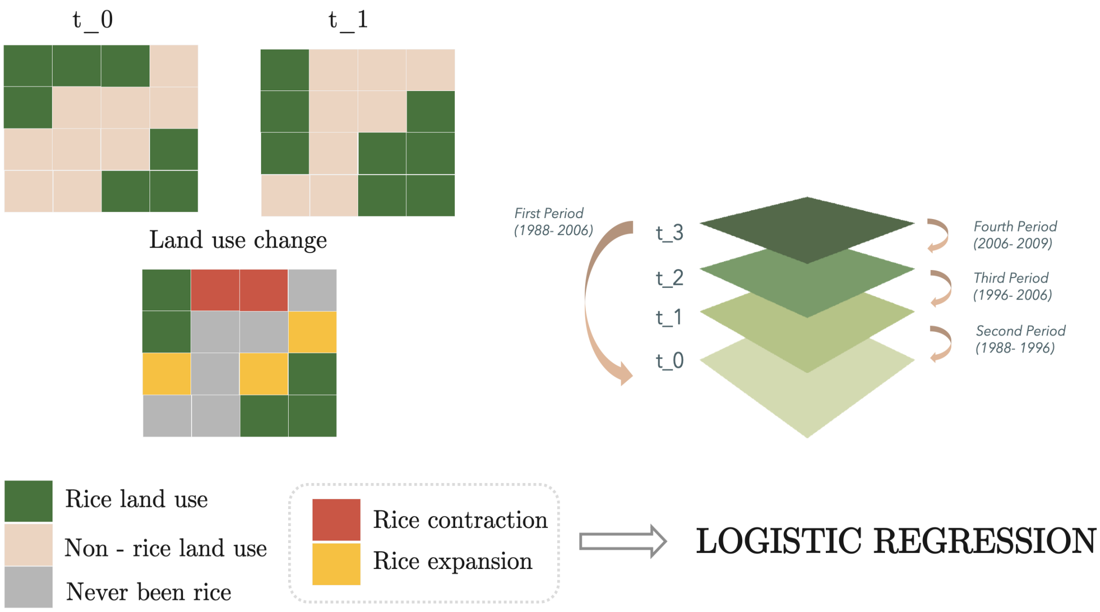

Figure 3.2: Land-use Change Computation Flow Chart

Spatial Re-sampling

The raster data used for this study were acquired at different spatial resolutions and needed to be re-sampled to the same spatial extent for comparative analysis. The data was re-sampled to a pixel size of 100 x 100m. This spatial resolution was chosen because for the purpose of our study, finer resolutions is not necessary and will only increase the processing time.

3.3 Sampling

There are different types of sampling based on the approach in which the data is divided, the most commonly used sampling methods are random and stratified sampling (Wang et al., 2012). We use a combination of stratified and random sampling technique to divide the data set into two, 70% for training and 30% for validation. The use of sampling is important when using non spatial models to model continuous spatial variables such as elevation, salinity etc., to reduce the spatial dependence in the data by reducing the data size. This technique has been used by other authors in land use studies and has proven to be effective and yield favorable results (Xie et al., 2005). It is however important to state that, the use of sampling can introduce some bias or uncertainties in the results as not all the observations are being considered when observations are sampled. To overcome the uncertainties introduced by sampling, employed the use of a bootstrapping approach to build the confidence interval and reduce uncertainties in our results. In the next section we will explain the concept of bootstrapping and how it was applied in our study.

3.4 Bootstrap

The bootstrapping technique as explained in chapter 2.3 above, is a random sampling and replacement technique. It is commonly used to generate confidence intervals in data where observations are sampled. We used the bootstrap approach to run our regression models. The training data was used in a bootstrap regression, a total on 100 iterations were made and a confidence interval was generated based on the iterations. After each iteration, the coefficient and the p value for the explanatory variables were stored. In the end of the bootstrap iterations, the distribution of the coefficients and p values were plot using a histogram plot to be able to visually inspect the concentration of the values on the x-axis. The concentration of the P-values on the x-axis is used to asses the significance of the explanatory variable. The coefficients distribution on the x-axis is used to asses the direction (positive or negative) of the influence of each explanatory variable has on the probability of change of the predictor variable.

3.5 Model Output

The logistic regression model output gives an overview of relationship between the response and the explanatory variables. The output provides the significance of each explanatory variable to the probability of change of the response variable through the P-values. The P-values are scaled between 0 and 1, the closer the P-value is to 0, the more significant the explanatory variable is to the probability of change of the response variable, while the closer to 1 the values are indicates less significance. The threshold commonly used in literature is 0.05 (95%) and below to indicate significance of a variable to the probability of change of the predictor variable. Although this threshold can be different depending on the study done and also in cases with limited observations, lower thresholds such as 0.1 (90%) are used as there is no standard threshold (Fisher, 1950).

The model output provides the coefficient for each explanatory variable used in the model. This explains relationship, that is how a change in the explanatory variable influences the probability of change in the response variable. The relationship between the explanatory and response variable can be positive or negative based on the coefficient. The coefficient value represent the mean change in the response variable for every unit change of the explanatory variable. From the model coefficients we are able to derive which of the explanatory variable has a positive or negative relationship to the probability of land-use change.

Chapter 4

Data and Implementation

This chapter is structured as follows, the data used are described and the source of the data are presented. The links to some of the data sets used are provided in the footnote of the description page below. The preprocessing steps taken are described and finally a flow chart of the steps taken is presented.

4.1 Data Sources

Data	**Data Sources**	**Description of data**
Land use Maps	Minderhoud et al. (2018)	Land use maps of VMD for 1988, 1996, 2006 and 2009
Natural compaction	PhD thesis Minderhoud 2019	
Elevation	Minderhoud et al. (2019)	DEM of the VMD derived from a national topographical map of 2014
Population density	Landscan Database	The LandScan 2004 Global Population Database at 1 km spatial resolution.
Distance to river	Rivers Shapefile	From rivers shapefile of the VMD using GIS
Distance to sea	Sea Shapefile	From sea shapefile of the VMD using GIS
Salinization	Eslami et al. (2021)	Collected along the Hau River in the VMD in 2016 at spatial resolution of 3 km.
Subsidence	Minderhoud (2017)	3D hydrological model derived from geological borehole logs

Table 4.1: Data Sources and Summary

The land use maps used for this study were acquired from the study of Minderhoud et al. (2018) on relating land subsidence to land use in the VMD. In total, four land use maps for the years 1988, 1996, 2006 and 2009 were used. The population density data was acquired from the Landscan database [1]. The data was generated in 2004 as part of the WISDOM project, with a spatial resolution of 1 km. The salinity data [2] was acquired from the supplementary data of the study of Eslami et al. (2021). The salinity data has a spatial resolution of 3km and was created in 2016 by Eslami et al. (2021), along the Hau river in the VMD. The elevation data [3] was gotten from (Minderhoud et al., 2019) supplementary data. The authors interpolated almost 20.000 elevation points derived from a national topographical map of 2014 (scale 1:200,000). The natural compaction data [4] was acquired from the PhD dissertation of Minderhoud, "The sinking mega-delta : Present and future subsidence of the Vietnamese Mekong delta". The subsidence data [5] was acquired from the study of Minderhoud (2017). The authors used a 3D hydrogeological model to simulate the amount of use and extraction of groundwater for 25 years. From the simulation they calculate amount of subsidence induced by groundwater extraction with spatial resolution of 1km^2 in the VMD. The data for distance to river and sea were provided by Philip Minderhoud, PhD. The data was generated using shapefiles of rivers and sea of the VMD.

[1] https://catchmekong.eoc.dlr.de/Elvis/ The LandScan 2004 Global Population Database at 30 arc seconds (1 km. or finer) resolution

[2] https://esurf.copernicus.org/preprints/esurf-2020-109/ This data set contains salinity measurements along the Hau River within the Mekong Delta, Vietnam

[3] https://doi.pangaea.de/10.1594/PANGAEA.902136 Digital Elevation Model (DEM) of the Vietnamese part of the Mekong delta at 500 meter resolution

[4] https://dspace.library.uu.nl/handle/1874/375843 Natural compaction data

[5] https://iopscience.iop.org/article/10.1088/1748-9326/aa7146/meta Subsidence of the VMD derived using a ·D hydrogeological model

4.1.1 Land use maps

This study was carried out using the land use maps of the Vietnamese Mekong Delta for the period of 1988 – 2009. Four land use maps were used for the year 1988, 1996, 2006 and 2009. These land use maps were acquired from the study done by Minderhoud et al. (2018) on relating the impacts of land use to land subsidence in the Vietnamese Mekong delta. Each land use map has 14 land use sub-classes and were further reclassified into 5 land use classes. For the purpose of this study, we only consider these five land use classes; aquaculture, mixed crops, orchard, rice and urban land use.

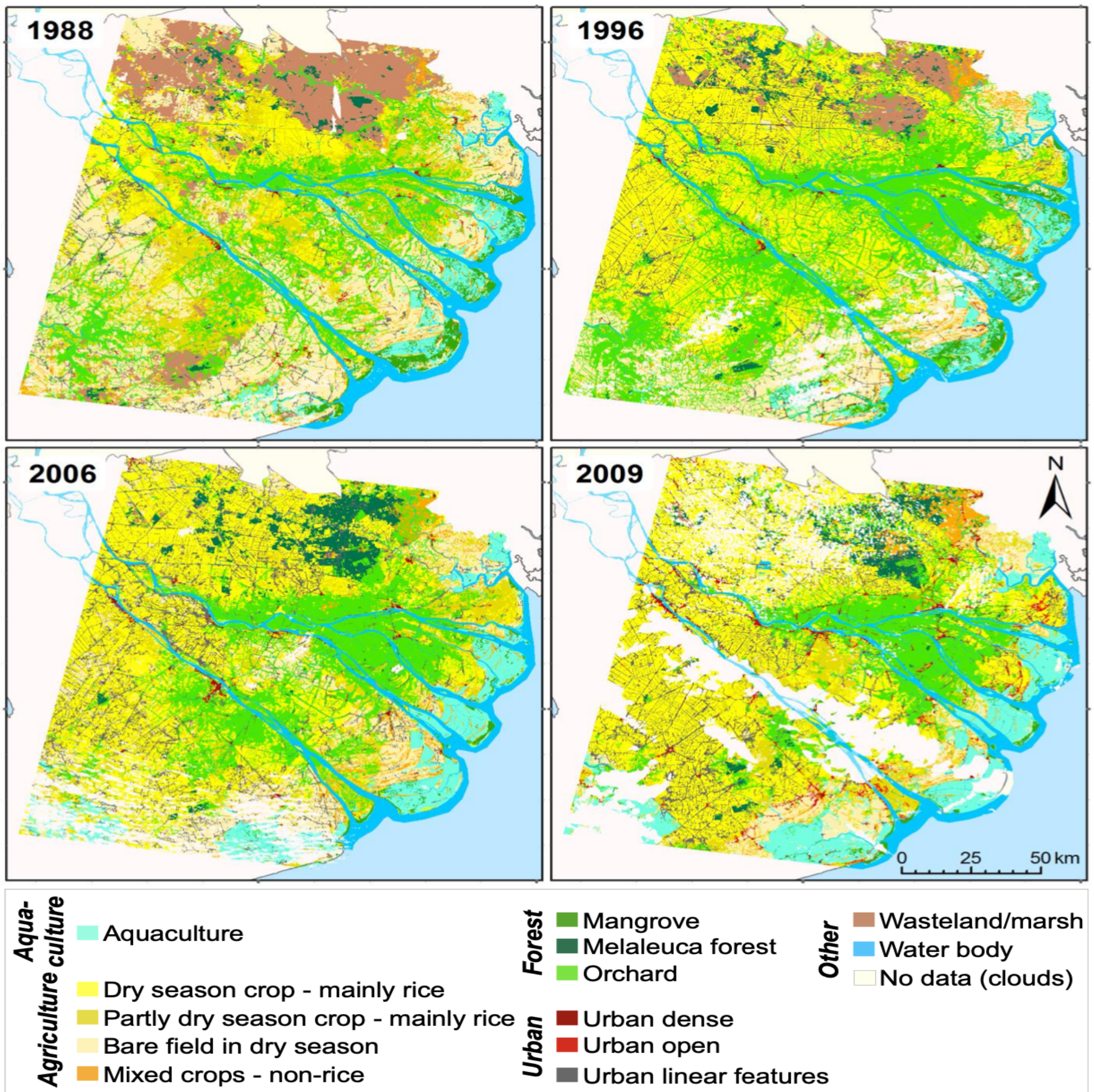

Figure 4.1: Land Use Maps of the Vietnamese Mekong Delta for 1988, 1996, 2006 and 2009

(Minderhoud et al., 2018)

4.2 Data Pre-processing

This study was carried with two separate models, one for land use expansion and the other for contraction and a data frame was created for each model used.

We create a raster stack for the four land use maps used. The land use maps were re-sampled to a spatial resolution of 100 x 100 meters. The re-sampling was done because of the size of the area being studied, in large regions, the land use types will most likely not change over short areas as compared to small regions. Secondly, more pixels will imply longer processing time and more noise in the data, given the aim of our study, this is not needed. From the raster stack created, the land-use change per pixel for each land use type was obtained based on the pixel values. A comparison between the land use maps at the different time stamps (e.g., 1988 and 1996) was done to obtain where each land use type is expanding or contracting within the study area.

To compute the land-use change for each land use type, four periods were derived, the first is the entire period of the study and three intermediate periods. The first period is the overall period of study 1988 - 2006, leaving out 2009 because of missing values from cloud cover in the land use map. The second period is 1998 - 1996, the third period is 1996 - 2006 and the fourth period is 2006 - 2009. We have a total four period for each land use class. For each period, based on the pixel values, we calculate where a land use type is expanding and where it is contracting. Two data frames were created, one for expansion and the other for contraction with each land use type at the four period. The four periods are divided in such a way that from the first period, we are able to derive which of the environmental variables are significant to land-use change or not in the entire period of study. From the three intermediate periods we can observe a trend if the significance of the environmental variable on land-use change is shifting or remains the same.

To calculate the contraction and expansion for each land use class, we compared if a pixel value is present in a land use map at time step t and absent at time step t-1 or vise versa. For example, a pixel of rice land use is present in 1988 and absent in 1996, we take it that the land use has contracted. For the contraction data frame we gave a value of 1 where a land use has contracted and 0 it has not. The same process was carried out for the expansion, we gave a value of 1 where a land use has expanded and 0 where it has not in the data frame.

Finally, we have a data frame for land use expansion and another for contraction, with 1 indicating where a land use has expanded or contracted and 0 where it hasn't in the different periods. We converted the values in the data frames to logical (true or false) to use in our logistic regression model.

Year	**Period**
1988 - 2006	First period
1998 - 1996	Second period
1996 - 2006	Third period
2006 - 2009	Fourth period

Table 4.2: Study Periods Derived from Land Use Change Computed using the Four Land Use Maps

The re-sampled environmental variables were added to the data frames that had been created containing the land-use change classes for land use expansion and contraction. The distribution of the environmental (explanatory) variables were verified to ensure that the data is normally distributed and not skewed. A histogram plot was used to plot the distribution and a summary statistics of the variables were derived to obtain the minimum values of each variable. This is useful to ensure that the model prediction are not biased. In the case where a variable is not normally distributed or has a minimum value of less than zero, we adjusted the minimum values and converted to logarithmic form to reduce data skewedness.

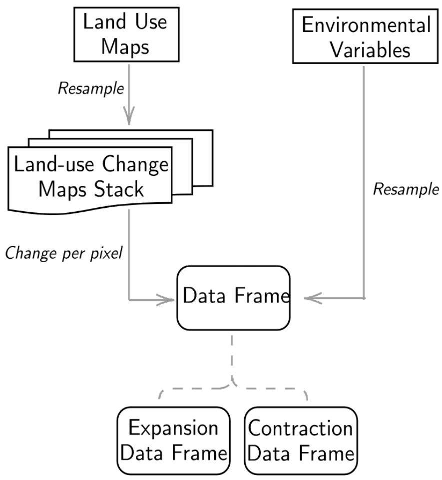

Figure 4.2: Data Preparation Flow Chart

4.3 Model Implementation and Validation

This sections presents how all the previous methods and data were implemented to achieve the findings presented in the next chapter.

We built two models for land use expansion and land use contraction. These models were built based on the assumption that the factors that drive land use expansion might not necessarily be the same (or opposite) as that of land use contraction. The land use expansion data only takes into account observations where a certain land use type has expanded based on the pixel wise comparison between two land use maps at different periods (e.g. present and previous time step). The same process as mentioned in the previous paragraph was repeated for points where a certain land use type has contracted.

In each model used, data preprocessing was done, the preprocessing steps are explained in section 4.2 above. The data was divided into two parts, a training and test data using a random sampling technique. The training data was used to carry out the model iterations, while the test was used to validate the model output. For each model, we used a bootstrap technique on the training to run a regression analysis, a total of hundred iterations were made for each land use type.

The coefficients and p values from the model parameters were saved for each iteration of the model, building a distribution of p values and coefficients per land use class. The mean and standard deviation of the total distribution of p values obtained was done to get the best representation of the distribution. A threshold of 0.1 (90%) is used, based on literature in cases where the observation are not much, lower thresholds are used (Fisher, 1950). A histogram plot was used to visually analyse the distribution of the coefficients and the p values by inspecting where the values are clustered on the x-axis. The clustering of the p values on the x-axis indicate if a value is significant to the probability of change of a land use type if the distribution is clustered closer to zero, while if it is evenly spread out it is less significant. The plots were compared for the different periods to see if there significance of the explanatory variables is shifting between the study period. The coefficients were also plot using a histogram plot to visually inspect the direction (positive or negative) in which the explanatory variable influences the probability of change per land use class. The x-axis of the histogram plots for the coefficient ranges from negative to positive. We are able to see in what direction the explanatory variable is influencing the probability of change of land use based on the clustering of the distribution on the x-axis.

A post regression analysis was carried out using the test data to validate the output of the regression analysis. A confusion matrix was used to calculate the accuracy of each model predictions. The overall accuracy's for each land use type in each model was derived by cross validating the model predictions with the test data. The model accuracy is generated based on the correctly predicted observation divided by the total number of observations. A receiver operating characteristics curve for each land use type in each period was generated to evaluate how well the model predicts at different prediction thresholds.

Chapter 5

Results and Discussions

This chapter presents and discusses the results gotten from the implementation as described in Chapter 4. The first section presents how the significance of each environmental variable on land use expansion and contraction is derived. The second section presents the validation of the results gotten from the models. The third section presents and interprets the overall results for the land use expansion and contraction, followed by discussion of the findings. Finally, the last section describes the limitation of the study and possible further improvements.

5.1 Significant Variables to Land-use Change

To illustrate how the significance of the environmental variables to land-use change were derived, this section presents the coefficient and p values of subsidence on mixed crops expansion.

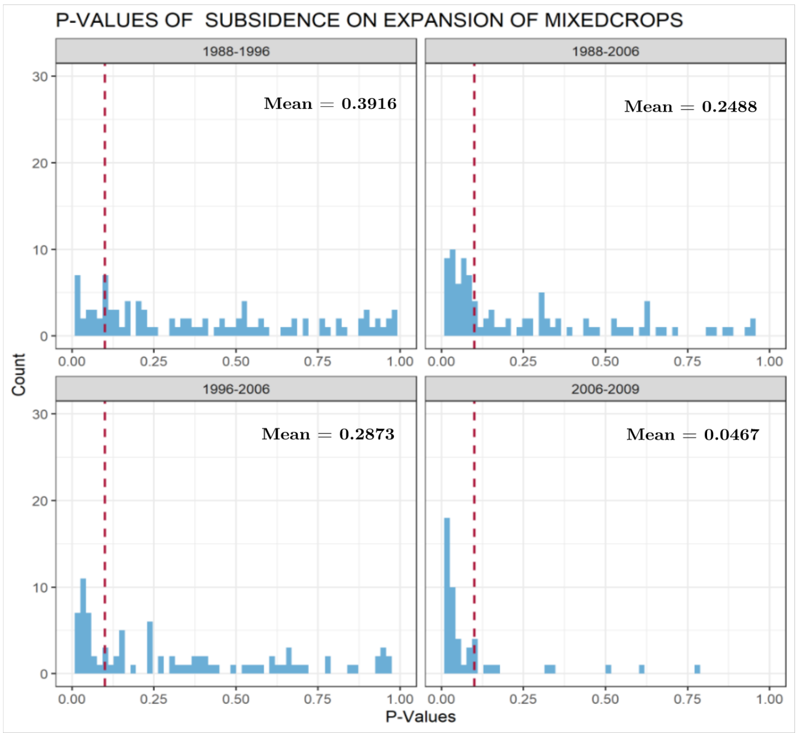

Figure 5.1: P value Distribution of Subsidence on Mixed Crops Expansion

We compare the mean of the p values of each land use between the different periods. Histogram plots showing the distribution of the values were made to visualise these p values.

In this example, we use the influence of subsidence on mixed crops land use expansion in all the time periods of our study. The x-axis of the histogram plot in figure 5.1 shows the p values, while the y-axis shows the count. From figure 5.1 above, we can see the bars are spread out in the first three periods but, are clustered close to zero in the last period. The clustering in the fourth period indicates that subsidence is significant in the last period (2006 - 2009) only, while in the other is not significant.

To evaluate the relationship between the environmental variables and land-use change, we use the model coefficient values. An example can be seen in figure 5.2 below for subsidence on mixed crops expansion.

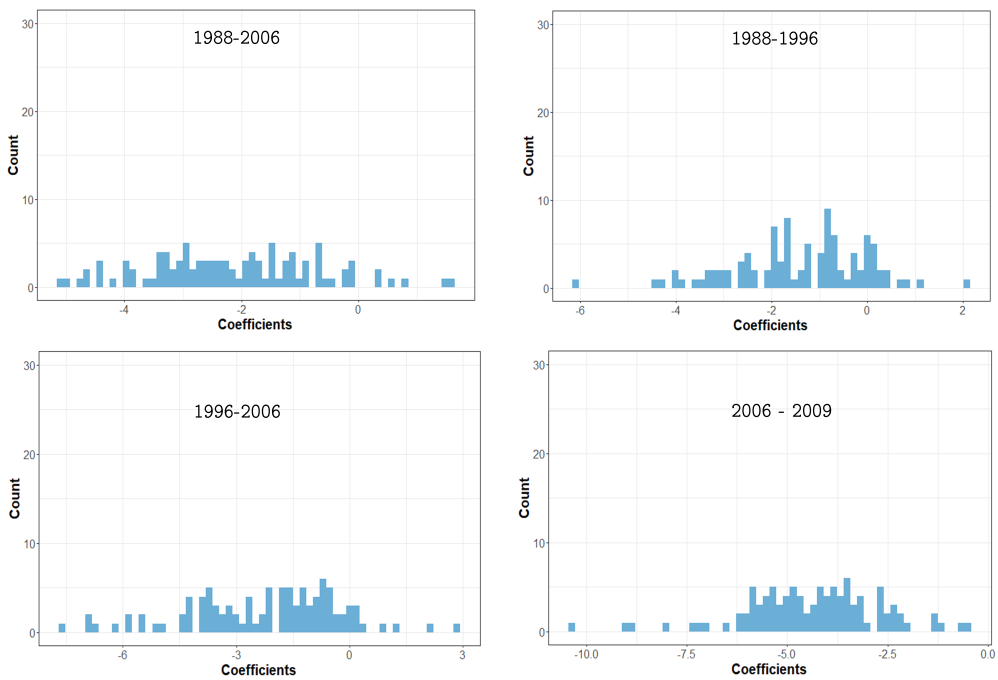

Figure 5.2: Coefficient Distribution of Subsidence on Mixed Crops Expansion

The x-axis of the histogram plot in figure 5.1 shows the p values, while the y-axis shows the count. From figure 5.2 above, we can visually observe that subsidence has a negative relationship to mixed crop expansion. This gives us the direction (positive or negative), how mixed crop land expansion is driven by subsidence. This implies that the more subsidence, the less probability of mixed crop land use expanding.

The same process was repeated for all variables in all land use classes to derive the significant variables and how they drive land-use change. In the next section, the results for all the land use classes are presented and discussed in more detail. The histogram plots for the p values of the significant variables for each land use class are provided in the appendices section of this thesis.

5.2 Environmental Drivers of Land-Use Change

From our regression models, we are able to derive if an environmental variable is significant or not to land-use change. A comparison of the mean distribution of the p values obtained for each explanatory variables shows us if a variable significant to land-use change or not. For each land use class, the mean of the p value of the explanatory variables for entire study period and the three intermediate periods are compared.

Results from our study reveal that the influence of environmental factors that have driven land-use change between 1988 - 2009 in the VMD is not same. There is a shift in significance of these variables on land-use change at different periods. In some cases, the significance of these variables on land-use change shifts, while in others they remain same, either significant or not over the study period. Some variables which are not significant became significant, while others that were significant became insignificant.

In the next two sub-sections, the summary of the results for the land-use contraction and expansion models will be presented and discussed.

5.2.1 Land Use Contraction

In figure 5.3 below, the results for the land use contraction model is presented for each land use type. The columns show the explanatory variables, while the rows show the land use types. The bars in the figure represents if a variable is significant in a period. For example, one bar means the variable is significant to change in land use of that land use type in only one period. The color of the bars represents in which period (1, 2, 3 or 4) a variable is significant.

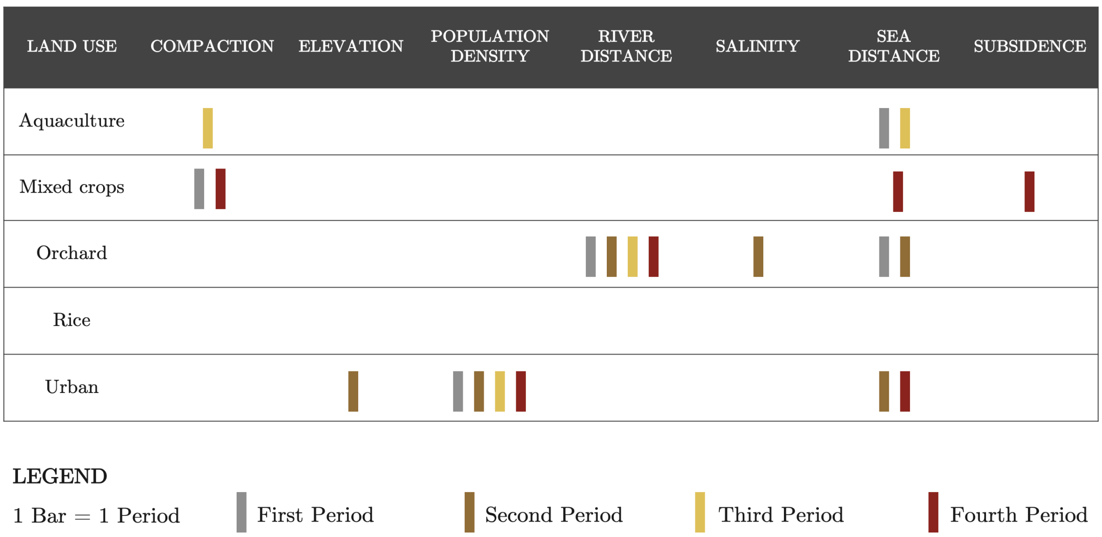

Figure 5.3: Overview of Significant Variables in Land Use Contraction

As seen in the figure 5.3 above, distance to river and population density remain significant to land-use change in orchard and urban land use respectively, throughout the entire study period. Other variables which are not significant become significant, for example sea distance and subsidence become significant to mixed crops contraction only in the last period (2006 -2009). Lastly, variables that were significant ceased being significant, for example sea distance was significant in the first and second period to orchard land use contraction.

The second part of our work, is to evaluate the relationship between the environmental variables and land-use change, how these environmental variables have driven land-use change. We assess this relationship with the coefficient gotten from our models for each explanatory variable on each land use type. In figure 5.4 below, the coefficient for the significant explanatory variables in 5.3 above are shown.

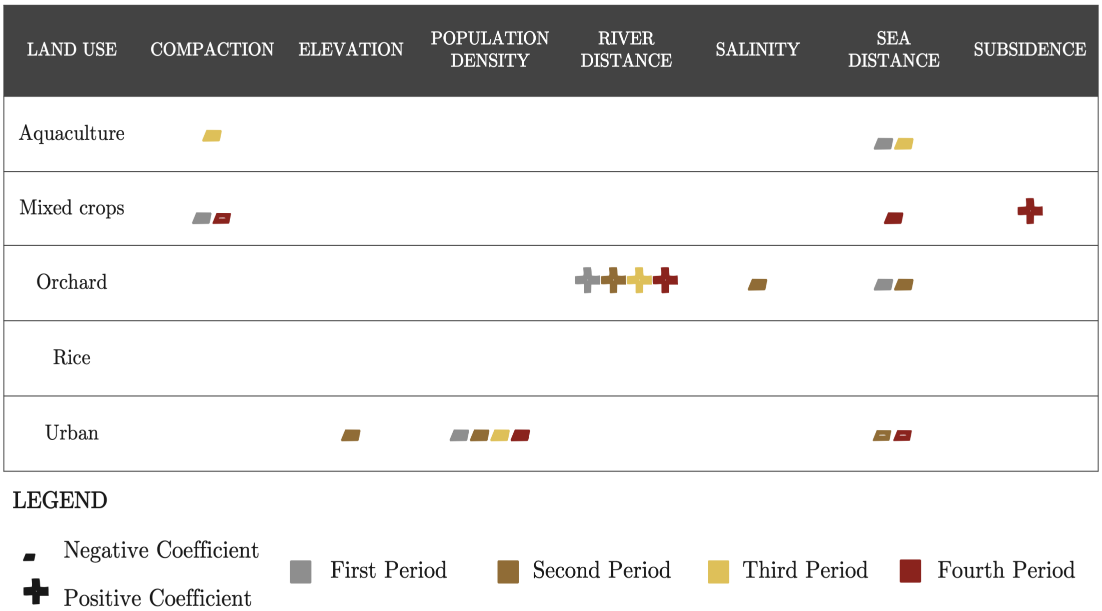

Figure 5.4: Coefficient of Significant Explanatory Variables on Land Use Contraction.

In figure 5.4 above, the minus sign (-) indicates a negative coefficient, while the plus sign (+) indicates a positive coefficient between the explanatory variable and the response variable. The color of each sign indicates the period, just as in figure 5.3 above. From the coefficient we are able to derive the direction of the relationship between the each explanatory and the response variable. For example, population density has a negative relationship with urban land use contraction. The implication of this is, with an increase in population density, the less probability of urban land use contracting if all other factors remain constant. On the other hand, distance to river has a positive relationship to orchard land use contraction, meaning the more the distance to river, the higher the probability of orchard land use contracting.

5.2.2 Land Use Expansion

In figure 5.5 below, the results for the land use expansion model is presented for each land use type. The columns show the explanatory variables, while the rows show the land use types. The bars in the figure represents if a variable is significant in a period . The color of the bar represents in which period (1, 2, 3 or 4) a variable is significant, just as in 5.3 in the previous section.

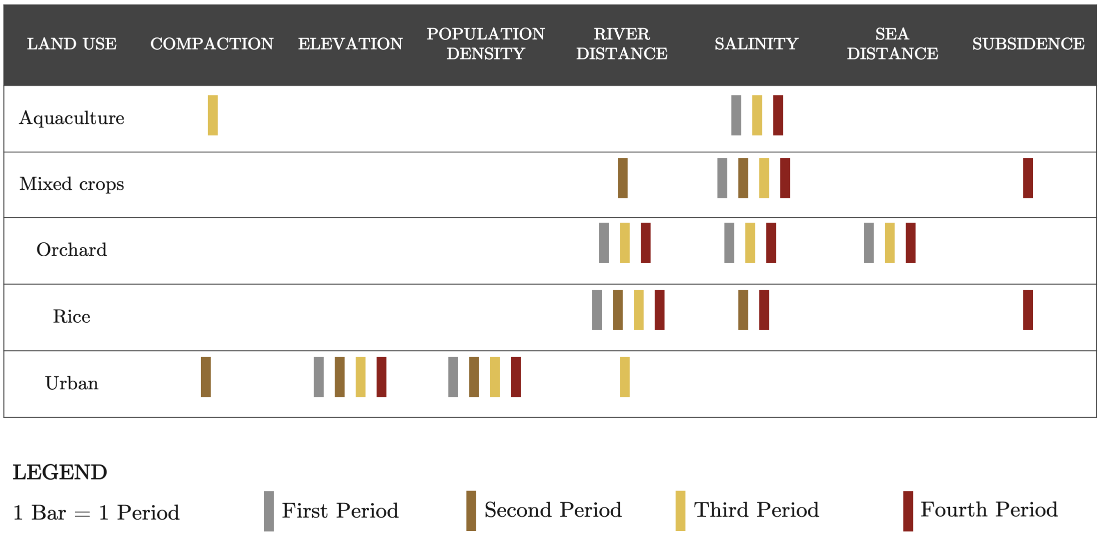

Figure 5.5: Overview of Significant Variables in Land Use Expansion

In figure 5.5 above, more variables are significant to land use expansion than they are to contraction. In some cases, some variables are significant to both expansion and contraction in same periods, for example distance to river. The distance to river is significant in three periods in figure 5.5 for orchard land use expansion as compared to all four periods in figure 5.3 in the same land use contraction. Population density remains the only variable that is significant to both urban contraction and expansion in all four periods. In rice land use expansion, 3 variables are significant, while none was significant to contraction of this land use. This shows that although some variables are significant to both land use contraction and expansion at same or different periods, it is not always the case. A variable can also be only significant to either expansion or contraction of a certain land use type, as seen in the rice land use. The coefficients for each of the explanatory variables on expansion of all land use types are shown in figure 5.6 below.

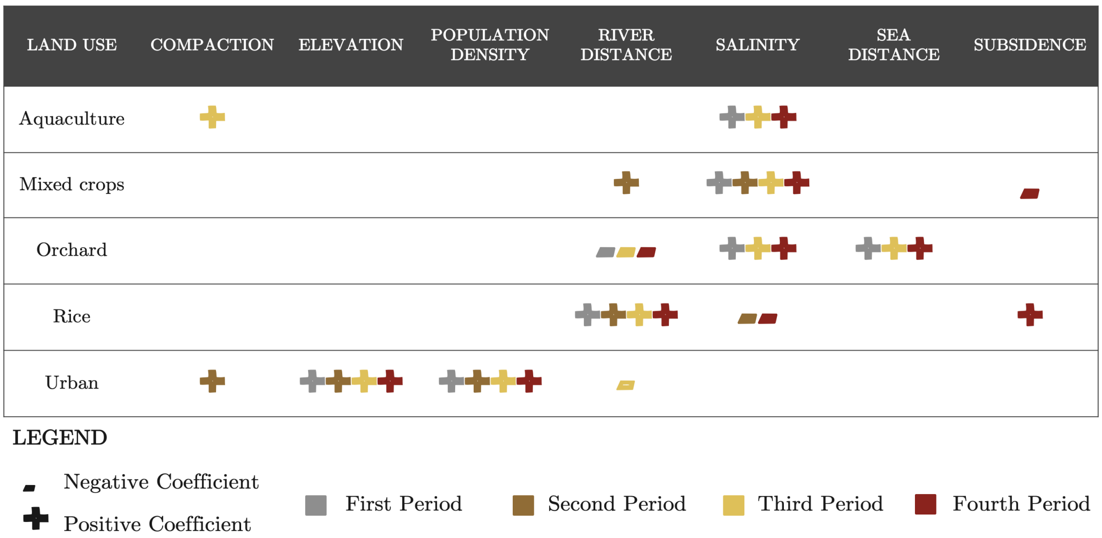

Figure 5.6: Coefficient of Significant Explanatory Variables on Land Use Expansion.

Figure 5.6 above shows the coefficient of the explanatory variables to land use expansion. As in figure 5.4 in previous sub-section, the minus sign indicates a negative coefficient, while the plus sign indicate a positive coefficient. A variable having negative coefficient on a land use type implies that the more that variable increases, the less probability of expansion of that land use type. A positive coefficient means an increase in the variable results to increase in the probability of that land use type expanding. For example, in figure 5.6 above, elevation has a positive influence on urban land expansion, meaning the more the elevation increases, the higher the probability of urban land expanding. The same effect can be seen in population density on urban land expansion.

5.3 Validation

5.3.1 Confusion Matrix

The output of both models used were validated with the test data using the methods explained in section 2.5 above. The overall accuracy's for the model predictions, which is the total number of correct classification divided by total number of observations were computed for each land use class using a confusion matrix. The overall accuracy's are between 48% to 85% in all land use classes for contraction and 57% to 84% for expansion. In the following paragraphs the model prediction accuracy's for each land use type for both models used is presented for every period.

In 5.1 below, the overall accuracy's for each land use class in the first period (1988 - 2006) for both contraction and expansion is presented. In each model for contraction and expansion, the urban land use has the highest accuracy, with 83% and 78% overall accuracy's respectively. This means that for the urban land use contraction, the model correctly predicted 83% of the number of observations where the urban land use changed or not (contracted or not). For example, if the urban contraction has a total of 100 observations, the model correctly predicted 83 observations that have contracted or not. The kappa values represents the level of agreement between the observed and expected values of the model prediction and a higher Kappa values represents better agreement.

	Accuracy		**Kappa**	
Land Use	**Contraction**	**Expansion**	**Contraction**	**Expansion**
Aquaculture	65%	79%	0.32	0.59
Mixed Crops	53%	63%	0.07	0.29
Orchard	71%	81%	0.42	0.62
Rice	62%	72%	0.23	0.44
Urban	83%	78%	0.67	0.56

Table 5.1: Summary of Model Prediction Accuracy 1988 - 2006 (First Period)

	Accuracy		Kappa	
Land Use	**Contraction**	**Expansion**	**Contraction**	**Expansion**
Aquaculture	59%	84%	0.18	0.68
Mixed Crops	53%	67%	0.07	0.35
Orchard	58%	62%	0.17	0.24
Rice	67%	71%	0.35	0.42
Urban	77%	70%	0.54	0.39

Table 5.2: Summary of Model Prediction Accuracy 1988 - 1996 (Second Period)

	Accuracy		Kappa	
Land Use	**Contraction**	**Expansion**	**Contraction**	**Expansion**
Aquaculture	70%	81%	0.39	0.62
Mixed Crops	49%	66%	0.02	0.35
Orchard	67%	65%	0.33	0.29
Rice	53%	69%	0.07	0.39
Urban	85%	73%	0.70	0.47

Table 5.3: Summary of Model Prediction Accuracy 1996 - 2006 (Third Period)

	Accuracy		Kappa	
Land Use	**Contraction**	**Expansion**	**Contraction**	**Expansion**
Aquaculture	57%	81%	0.13	0.62
Mixed Crops	67%	64%	0.34	0.27
Orchard	70%	58%	0.42	0.18
Rice	48%	71%	-0.02	0.42
Urban	57%	57%	0.13	0.16

Table 5.4: Summary of Model Prediction Accuracy 2006 - 2009 (Fourth Period)

5.3.2 Receiver Operating Characteristic Curve

The receiver operating characteristic (ROC) curve for each land use class for both models were created to visually analyze how well the models predicts at all possible classification thresholds. The ROC plots are created based on the correctly classified observations (true positive rate) against the misclassified observations (false positive rate). In our study, the true positives are the observations where a certain land use contracted or expanded and was correctly predicted by the model. The false positives are the observations where a land use did not contract or expand and was predicted falsely by the model.

The plots forming an almost perfect right angle curve to the top left corner of the plot, indicate a better performance in terms of correct classification. Plots closer to the diagonal line indicate a poor classification ability by the model. The figures below show the ROC curves for each land use at all four periods for both contraction and expansion models. The area under curve (AUC) was derived based on the ROC curve, this shows total area under the ROC curve. A higher AUC indicates a better classification performance of the model.

In figure 5.7 for example, the model performed better, with an AUC of 98% in the first period than in the remaining three periods based on the model classification.

Aquaculture Land Use

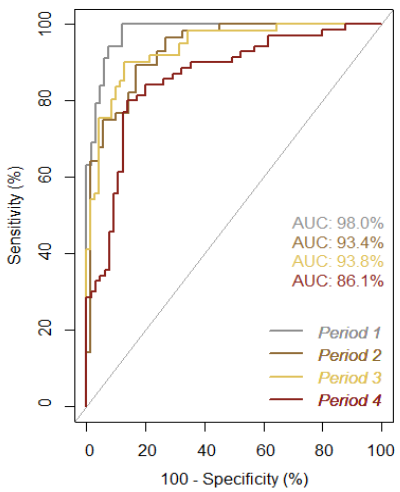

Figure 5.7: AUC and ROC Curve for Aquaculture Contraction

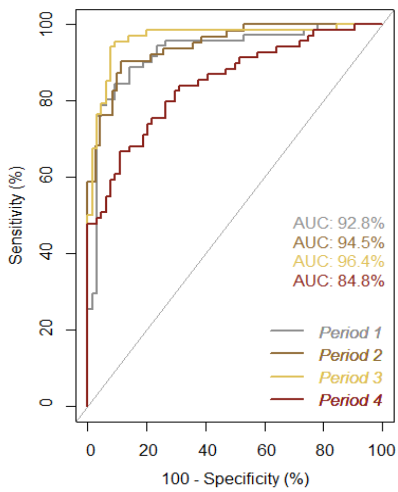

Figure 5.8: AUC and ROC Curve for Aquaculture Expansion

Mixed Crops Land Use

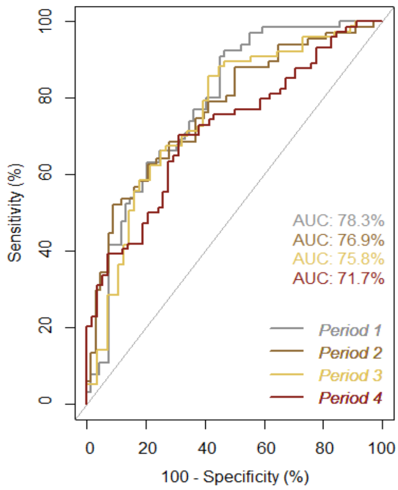

Figure 5.9: AUC and ROC Curve for Mixed Crops Contraction

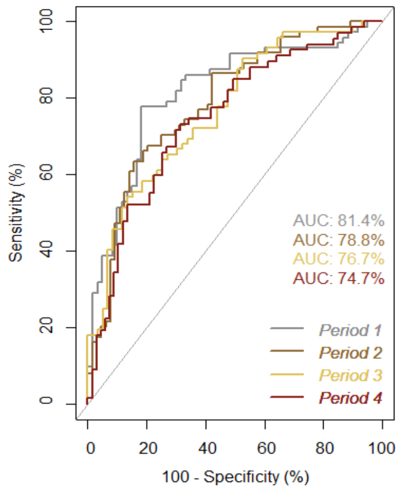

Figure 5.10: AUC and ROC Curve for Mixed Crops Expansion

Orchard Land Use

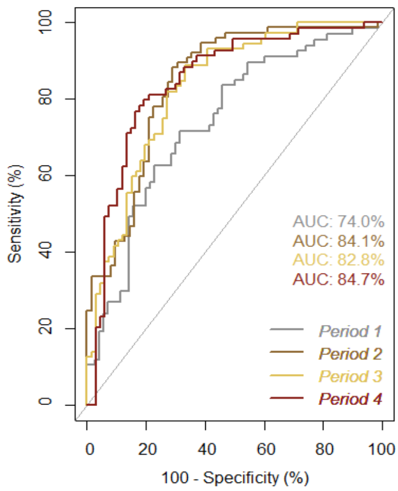

Figure 5.11: AUC and ROC Curve for Orchard Contraction

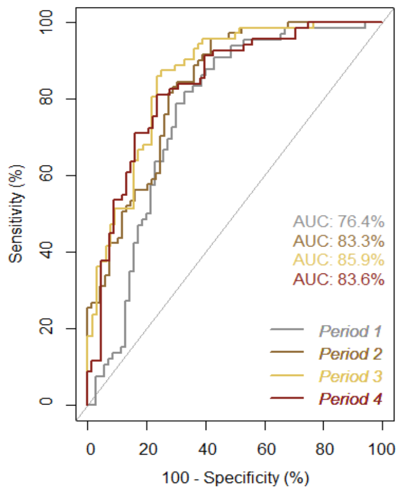

Figure 5.12: AUC and ROC Curve for Orchard Expansion

Rice Land Use

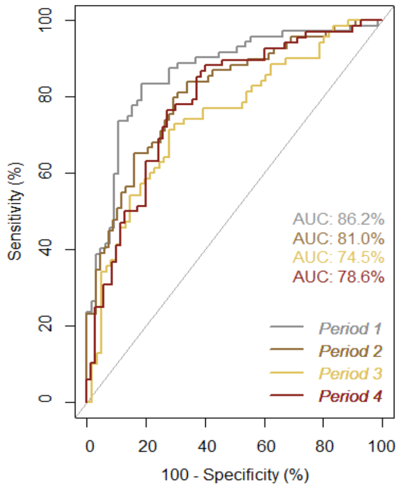

Figure 5.13: AUC and ROC Curve for Rice Contraction

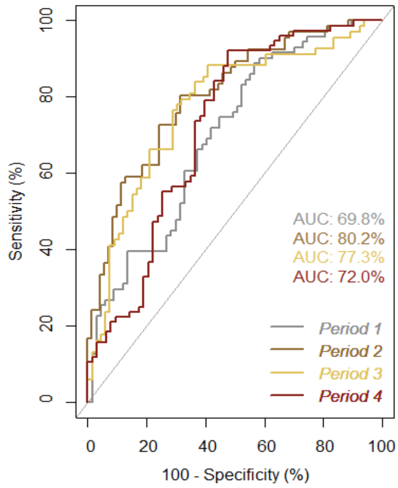

Figure 5.14: AUC and ROC Curve for Rice Expansion

Urban Land Use

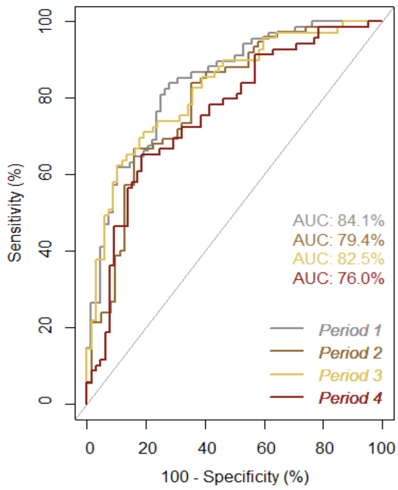

Figure 5.15: AUC and ROC Curve for Urban Contraction

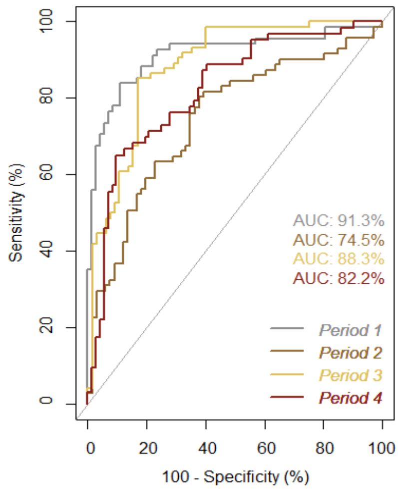

Figure 5.16: AUC and ROC Curve for Urban Expansion

5.3.3 Discussion

Findings from our study reveal that changes in environmental variables influences its significance to land-use change. As seen from the results above, some variables are not significant at some periods but become significant in other periods. For example, the first period (1988 - 2006), gives us an overview of which of the explanatory variables are significant to land-use change in a land-use type over the entire period. Some variables might not be significant in this period but are significant in one or more of the intermediate periods, as seen in the case of subsidence on mixed crops and rice expansions in figure 5.5 above.

Secondly, we observed that more variables are significant in land use expansion in the 3^{rd} (1996 - 2006) and 4^{th} (2006 - 2009) periods than compared to the 2^{nd} period (1988 - 1996). This could be explained by large environmental changes such as agricultural and urban expansion etc. due to economic transformations in the VMD during this period (Yun, 2019; Diez, 2016).

The distance to sea is the most significant variable to land use contraction in the VMD. The distance to sea is significant to land use contraction in at least one period in all land use classes except rice land use between 1988 - 2009. The VMD is a low lying delta region, with most parts of the delta below 2m above sea level and similar to most delta regions around the world, it is exposed to threats from sea level rise (Oppenheimer & Glavovic, 2019; Erban et al., 2014). Based on sea level estimates, the influence of sea distance on land-use change in the VMD is only expected to increase as the sea level increases (Lu & Flavelle, 2019; Oppenheimer & Glavovic, 2019; Takagi et al., 2016).

In the last period of our study, subsidence becomes significant to contraction and expansion of mixed crops and rice expansion. The VMD is the biggest agricultural region in Vietnam, our findings shows that subsidence has a negative impact on mixed crops expansion. Studies have shown that subsidence rates in the VMD is increasing (Minderhoud et al., 2018; Minderhoud, 2017; Erban et al., 2014), implying that future expansion of mixed crops could be at risk if subsidence rates continue to rise and mitigating action is not taken (Minderhoud et al., 2020).

The VMD has seen a rapid changes in land use since the economic reform in the late 1980's, which has attracted more population to this region. Our study shows population density is significant to both urban land use contraction and expansion at all periods. Population density has a positive relationship with urban expansion, implying that as population increases, it is likely that the urban land use will expand. Secondly, elevation is significant and has a positive relationship to urban land use expansion at all periods, implying that urban land use is likely to expand with an increase in elevation if all other factors are constant.

The VMD, is faced with the challenges of sea level rise (Lu & Flavelle, 2019; Oppenheimer & Glavovic, 2019) and increasing rates of subsidence (Minderhoud et al., 2018; Minderhoud, 2017; Erban et al., 2014) as discussed in the previous paragraphs above. Both of these factors contributes to reduction in elevation, which is crucial for an already low lying coastal region like the VMD. Reductions in elevation can be crucial to urban expansion, as seen from our study, the more elevation, the more chances are of urban expansion. Reductions in elevation coupled with increase in population density might put the future expansion of the urban land use in the VMD at risk.

Lastly, rice is one of the most important land use types in the VMD (Kuenzer & Knauer, 2013). Our study shows that salinity is significant to rice expansion, having a negative effect to rice expansion. This implies that with in increase in salinity, it is less likely that rice land use will expand. A report by UNDP (2016) shows that in the year 2016, Vietnam faced one of the worst salinity intrusion in over 90 years (Vu et al., 2018; UNDP, 2016). Based on our findings, increase in salinity would negatively affect the future of rice expansion. Necessary actions such as, reduction in ground water extraction (Minderhoud et al., 2020), planting of resilient crop species (M. T. Nguyen et al., 2019) etc. are needed to avoid reduction in rice expansion.

5.4 Limitations and Future Directions

The limitations and possible future improvements in this work are presented in this section as following:

The main limitation of this study is the limited availability of Land use data. For this study only four land use maps were used between the period of 1988 - 2009. A trend in land-use change might not be adequately observed if the land use data is limited. In order to have a better understanding of the trends in land-use change, more land use information over shorter time intervals could be added.

Secondly, This study did not consider the spatial interactions in the data. A non spatial model with sampling technique was used. A possible future enhancement could be done by considering models such an auto-regressive model which explicitly takes into account the spatial dependencies and auto-correlation. This might improve the results in the cases where the data is spatially correlated.

Finally, this research only focused on the Vietnamese Mekong delta, applying this methods in other related areas with different drivers of land-use change would be good to ascertain how our approach can be expanded.

Chapter 6

Conclusions

This study focused on the influence of environmental variables on land-use change, to understand how changes in these variables has driven land-use change between 1988 - 2009 in the Vietnamese Mekong Delta. To assess how the significance of these environmental variables on land-use change is influenced by shifts in the variables themselves. lastly, to evaluate the relationship between land-use change and changes in the environmental variables.

Four land use maps between 1988 - 2009 were used for this study. In each land use map, five land use classes were considered; aquaculture, mixed crops, orchard, rice and urban. A pixel wise comparison using the four land use maps was done to compute the change in for each land use class. The change was computed by dividing the maps into 4 periods, an overall period and three intermediate periods. The overall period from 1988 - 2006 is the (1^{st} period), excluding 2009 due to missing values from cloud cover. Similarly, the three Intermediate periods are; 1988 – 1996 (2^{nd} period), 1996 - 2006 (3^{rd} period) and 2006 – 2009 (4^{th} period). From the land-use change computed, each land use class expansion and contraction pixels were used to build two logistic regression models, one for land use contraction and another for expansion. In each model, the land-use change (contraction and expansion) was modelled with these six environmental variables; compaction, elevation, population density, distance to river and sea, salinity and subsidence as explanatory variables.

The results of this study shows that shifts in environmental variables have driven land-use change between 1988 - 2009 in the study area. Some variables have driven land-use change through-out the study period, while other variables only became significant to land-use change in the last period. Secondly, the influence shifts in the environmental variables have driven land-use change differently, some variables were significant to both contraction and expansion of a certain land use, while others are only significant in either contraction or expansion. From our study, we can conclude that the significance environmental factors that have driven land-use change is not constant.

In future studies, we recommend adding more land use information, four land-use maps between 1988 - 2009 were used to carry out this study. Adding more land-use maps might help to build a better trend land-use change.

The findings of this study can be useful to decision makers when considering significant factors that have driven land-use change when building a sustainable land use policies in the study area and similar delta regions in the world.

Bibliography

Eslami, S., Hoekstra, P., Kernkamp, H. W. J., Nguyen Trung, N., Do Duc, D., Nguyen Nghia, H., Tran Quang, T., van Dam, A., Darby, S. E., Parsons, D. R., Vasilopoulos, G., Braat, L. & van der Vegt, M. (2021). Dynamics of salt intrusion in the mekong delta; results of field observations and integrated coastal-inland modelling. *Earth Surface Dynamics Discussions*, *2021*, 1–36. https://doi.org/10.5194/esurf-2020-109

Abdi, A. M. (2020). Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience and Remote Sensing*, *57*(1), 1–20. https://doi.org/10.1080/15481603.2019.1650447

Anderson, R., Bayer, P. E. & Edwards, D. (2020). Climate change and the need for agricultural adaptation. *Current Opinion in Plant Biology*, *56*, 197–202. https://doi.org/https://doi.org/10.1016/j.pbi.2019.12.006
Biotic interactions AGRI 2019

Liu, S., Li, X., Chen, D., Duan, Y., Ji, H., Zhang, L., Chai, Q. & Hu, X. (2020). Understanding Land use/Land cover dynamics and impacts of human activities in the Mekong Delta over the last 40 years. *Global Ecology and Conservation*, *22*, e00991. https://doi.org/10.1016/j.gecco.2020.e00991

Minderhoud, Middelkoop, H., Erkens, G. & Stouthamer, E. (2020). Groundwater extraction may drown mega-delta: Projections of extraction-induced subsidence and elevation of the mekong delta for the 21st century. *Environmental Research Communications*, *2*(1), 011005.

Nguyen, Q. H., Tran, D. D., Dang, K. K., Korbee, D., Pham, L. D., Vu, L. T., Luu, T. T., Ho, L. H., Nguyen, P. T., Trang, T. T., Nguyen, D. T., Wyatt, A., van Aalst, M., Tran, T. A. & Sea, W. B. (2020). Land-use dynamics in the Mekong delta: From national policy to livelihood sustainability. *Sustainable Development*, *28*(3), 448–467. https://doi.org/10.1002/sd.2036

Parker, A. (2020). Anthropogenic Drivers of Relative Sea-Level Rise in the Mekong Delta - A Review. *Quaestiones Geographicae*, *39*(1), 109–124. https://doi.org/10.2478/quageo-2020-0009

Briassoulis, H. (2019). Analysis of land use change: Theoretical and modeling approaches.

Brownlee, J. (2019). A gentle introduction to calculating normal summary statistics. https://machinelearningmastery.com/a-gentle-introduction-to-calculating-normal-summary-statistics/

Jiwon Yun. (2019). Vietnam's Politic of a Divided Nation: From the Reunification to DoiMoi (Renovation) and Its Implication for the Korean Peninsula and North Korea. *International Journal of Korean Unification Studies*, *28*(1), 63–92. https://doi.org/10.33728/ijkus.2019.28.1.003

Lu, D. & Flavelle, C. (2019). Rising seas will erase more cities by 2050, new research shows. *New York Times*, *29*.

Minderhoud, Coumou, L., Erkens, G., Middelkoop, H. & Stouthamer, E. (2019). Mekong delta much lower than previously assumed in sea-level rise impact assessments. *Nature Communications*, *10*(1), 1–13. https://doi.org/10.1038/s41467-019-11602-1

Nguyen, M. T., Renaud, F. G., Sebesvari, Z. & Nguyen, D. C. (2019). Resilience of agricultural systems facing increased salinity intrusion in deltaic coastal areas of vietnam. *Ecology and Society*, *24*(4).

Oppenheimer, M. & Glavovic, B. (2019). Chapter 4: Sea Level Rise and Implications for Low Lying Islands, Coasts and Communities. IPCC SR Ocean and Cryosphere. *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate [H.- O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, M. Nicolai, A. Okem, J. Petzold, B. Rama, N. Weyer (eds.)]. In press.*, *Chapter 4*(Final Draft), 1–14. https://doi.org/10.1126/science.aam6284

Yun, J. (2019). Vietnam's politic of a divided nation: From the reunification to doimoi (renovation) and its implication for the korean peninsula and north korea. *International Journal of Korean Unification Studies*, *28*(1).

CFE-DM. (2018). 2018 Vietnam Disaster Management Referance Handbook. *Center for Excellence in Disaster Management and Humanitarian Assistance*. https://www.flickr.com/photos/un%7B%5C_%7Dphoto/6359456941

FAO. (2018). *The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals* (Vol. 29).

Minderhoud, Coumou, L., Erban, L. E., Middelkoop, H., Stouthamer, E. & Addink, E. A. (2018). The relation between land use and subsidence in the Vietnamese Mekong delta. *Science of the Total Environment*, *634*, 715–726. https://doi.org/10.1016/j.scitotenv.2018.03.372

Sun, B. & Robinson, D. T. (2018). Comparison of statistical approaches for modelling land-use change. *Land*, *7*(4). https://doi.org/10.3390/land7040144

Van Meijl, H., Havlik, P., Lotze-Campen, H., Stehfest, E., Witzke, P., Domınguez, I. P., Bodirsky, B. L., van Dijk, M., Doelman, J., Fellmann, T. et al. (2018). Comparing impacts of climate change and mitigation on global agriculture by 2050. *Environmental Research Letters*, *13*(6), 064021.

Vu, D. T., Yamada, T. & Ishidaira, H. (2018). Assessing the impact of sea level rise due to climate change on seawater intrusion in Mekong Delta, Vietnam. *Water Science and Technology*, *77*(6), 1632–1639. https://doi.org/10.2166/wst.2018.038

D'Agostino, R. (2017). *Goodness-of-fit-techniques*. Routledge.

Etikan, I. (2017). Sampling and Sampling Methods. *Biometrics & Biostatistics International Journal*, *5*(6), 215–217. https://doi.org/10.15406/bbij.2017.05.00149

Minderhoud. (2017). Impacts of 25 years of groundwater extraction in mekong delta.

UNCCD. (2017). *Global Land Outlook: Secretariat of the United Nations Convention to Combat Desertification.*

Diez, J. R. (2016). Vietnam 30 years after doi moi: Achievements and challenges. *magazine for "u r economic geography*, *60*(3), 121–133.

Fujihara, Y., Hoshikawa, K., Fujii, H., Kotera, A., Nagano, T. & Yokoyama, S. (2016). Analysis and attribution of trends in water levels in the vietnamese mekong delta. *Hydrological processes*, *30*(6), 835–845.

Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., Jepsen, M. R., Kuemmerle, T., Meyfroidt, P., Mitchard, E. T., Reiche, J., Ryan, C. M. & Waske, B. (2016). A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sensing*, *8*(1), 1–23. https://doi.org/10.3390/rs8010070

Revilla Diez, J. (2016). Vietnam 30 years after Doi Moi: Achievements and challenges. *Zeitschrift fur Wirtschaftsgeographie*, *60*(3), 121–133. https://doi.org/10.1515/zfw-2016-0035

Takagi, H., Thao, N. D. & Anh, L. T. (2016). Sea-level rise and land subsidence: Impacts on flood projections for the Mekong Delta's largest city. *Sustainability (Switzerland)*, *8*(9), 1–15. https://doi.org/10.3390/su8090959

UNDP. (2016). Vietnam drought and saltwater intrusion: Transitioning from emergency to recovery.

Smajgl, A., Toan, T. Q., Nhan, D. K., Ward, J., Trung, N. H., Tri, L. Q., Tri, V. P. & Vu, P. T. (2015). Responding to rising sea levels in the Mekong Delta. *Nature Climate Change*. https://doi.org/10.1038/nclimate2469

Tran, V. B. (2015). Groundwater issues and hydrogeological survey of the Mekong River basin in Vietnam. *Current Status and Issues of Groundwater in the Mekong River Basin.*, 93–121.

Erban, L. E., Gorelick, S. M. & Zebker, H. A. (2014). Groundwater extraction, land subsidence, and sea-level rise in the Mekong Delta, Vietnam. *Environmental Research Letters*, *9*(8). https://doi.org/10.1088/1748-9326/9/8/084010

Speelman, D. (2014). Logistic regression. *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*, *43*, 487–533.

Diyer, M., Namrani, H. & Elkadiri, A. (2013). *Land Use and Land Management Practices in Environmental Perspective.*

Kuenzer, C. & Knauer, K. (2013). Remote sensing of rice crop areas. *International Journal of Remote Sensing*, *34*(6), 2101–2139. https://doi.org/10.1080/01431161.2012.738946

Smith, W. (2013). Agriculture in the central Mekong Delta. *Overseas Development Institute*, (December), 17pp.

Adami, M., Rudorff, B. F. T., Freitas, R. M., Aguiar, D. A., Sugawara, L. M. & Mello, M. P. (2012). Remote sensing time series to evaluate direct land use change of recent expanded sugarcane crop in Brazil. *Sustainability*, *4*(4), 574–585. https://doi.org/10.3390/su4040574

ADB. (2012). *Viet Nam: Urban Sector Assessment, Strategy, and Road Map.*

Wang, J. F., Stein, A., Gao, B. B. & Ge, Y. (2012). A review of spatial sampling. *Spatial Statistics*, *2*(1), 1–14. https://doi.org/10.1016/j.spasta.2012.08.001

Nong, Y. & Du, Q. (2011). Urban growth pattern modeling using logistic regression. *Geo-Spatial Information Science*, *14*(1), 62–67. https://doi.org/10.1007/s11806-011-0427-x

Radil, S. M. (2011). *Spatializing social networks: Making space for theory in spatial analysis* (PhD). University of Illinois at Urbana-Champaign. Retrieved January 23, 2021, from http://hdl.handle.net/2142/26222

Rey, S., Anselin, L., Baltagi, B. H., Anselin, L., Fingleton, B., López-Bazo, E., Quah, D., Barro, R. J., Sala-I-Martin, X., Blanchard, O. J., Hall, R. E., Johnson, P. A., Kar, S., Jha, D., Kateja, A., Johnson, P. A., Rey, S., Baltagi, B. H., Anselin, L., ... Sala-I-Martin, X. (2011). A companion to Theoretical Econometrics Spatial econometrics. *Economics Letters*, *4*(2), 223–251.

Getis, A. (2010). Handbook of Applied Spatial Analysis. *Handbook of Applied Spatial Analysis*, 255–278. https://doi.org/10.1007/978-3-642-03647-7

Tayyebi, A., Delavar, M. R., Yazdanpanah, M. J., Pijanowski, B. C., Saeedi, S. & Tayyebi, A. H. (2010). A spatial logistic regression model for simulating land use patterns: A case study of the shiraz metropolitan area of iran. *Advances in earth observation of global change* (pp. 27–42). Springer.

Ting, K. (2010). Confusion matrix. *Encyclopedia of Machine Learning.*

Bivand, R., Müller, W. G. & Reder, M. (2009). Power calculations for global and local Moran's I. *Computational Statistics and Data Analysis*, *53*(8), 2859–2872. https://doi.org/10.1016/j.csda.2008.07.021

Briassoulis, H. (2009). FACTORS INFLUENCING LAND-USE AND LAND-COVER CHANGE. *LAND USE, LAND COVER AND SOIL SCIENCES*, *1.*

Crain, C. M., Halpern, B. S., Beck, M. W. & Kappel, C. V. (2009). Understanding and managing human threats to the coastal marine environment. *Annals of the New York Academy of Sciences*, *1162*, 39–62. https://doi.org/10.1111/j.1749-6632.2009.04496.x

Haan, M. D., Heij, R. D. & Horsten, M. (2009). The rise of survey sampling. *Discussion paper (09036)*, (09036).

Buckeridge, D. L., Okhmatovskaia, A., Tu, S., O'Connor, M., Nyulas, C. & Musen, M. A. (2008). Predicting outbreak detection in public health surveillance: quantitative analysis to enable evidence-based method selection. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, (February), 76–80.

Getis, A. (2008). A history of the concept of spatial autocorrelation: A geographer's perspective. *Geographical Analysis*, *40*(3), 297–309. https://doi.org/10.1111/j.1538-4632.2008.00727.x

Hoang, X., Dinh, T., Nguyen, T. & Tacoli, C. (2008). Urbanization and rural development in Vietnam 's Mekong Delta: Livelihood Transformations in Three Fruit-Growing Settlements. *Iied*, (May), 73.

Mendelsohn, R. (2008). The impact of climate change on agriculture in developing countries. *Journal of Natural Resources Policy Research*, *1*(1), 5–19.

Pech, S. & Sunada, K. (2008). Population growth and natural-resources pressures in the Mekong River Basin. *Ambio*, *37*(3), 219–224. https://doi.org/10.1579/0044-7447(2008)37[219:PGANPI]2.0.CO;2

Portmann, F., Siebert, S., Bauer, C. & Döll, P. (2008). Global dataset of monthly growing areas of 26 irrigated crops. *Frankfurt Hydrology Paper*, (06), 400.

Reis, S. (2008). Analyzing land use/land cover changes using remote sensing and GIS in Rize, North-East Turkey. *Sensors*, *8*(10), 6188–6202. https://doi.org/10.3390/s8106188

Stephenson, B., Cook, D., Dixon, P., Duckworth, W., Kaiser, M., Koehler, K. & Meeker, W. (2008). Binary response and logistic regression analysis. *available at:< a href=" http://www. stat. wisc. edu/mchung/teaching/MI-A/reading/GLM. logistic. Rpackage. pdf"> http://www. stat. wisc. edu/mchung/teaching/MIA/reading/GLM. logistic. Rpackage. pdf(last access: 30 August 2014).*

Tu, J. & Xia, Z. G. (2008). Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluation. *Science of the Total Environment*, *407*(1), 358–378. https://doi.org/10.1016/j.scitotenv.2008.09.031

Nguyen Thanh, P., Le Xuan, S., Nguyen Quoc, T., Huynh Han, C., Cao Tuan, A. & Nguyen Minh, H. (2007). Economics of aquaculture feeding practices: Viet Nam. *FAO Fisheries Technical Paper*, (505), 183–205.

Chen, X. L., Zhao, H. M., Li, P. X. & Yin, Z. Y. (2006). Remote sensing image-based analysis of the relationship between urban heat island and land use/-

cover changes. *Remote Sensing of Environment*, *104*(2), 133–146. https://doi.org/10.1016/j.rse.2005.11.016

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Jansen, L. J. (2006). Harmonization of land use class sets to facilitate compatibility and comparability of data across space and time. *Journal of Land Use Science*, *1*(2-4), 127–156. https://doi.org/10.1080/17474230601079241

Di Gregorio, A. & Jansen, L. (2005). Chapter 2: Concepts, Definitions and Links. Retrieved December 29, 2020, from http://www.fao.org/3/v8047e/v8047e04.htm%7B%5C#%7Dland%20and%20land%20resources%20http://www.fao.org/3/v8047e/v8047e04.htm%7B%5C#%7Denvironmental%20resources%20and%20natural%20resources

Jones, C. M. & Athanasiou, T. (2005). Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. *The Annals of thoracic surgery*, *79*(1), 16–20.

Xie, C., Huang, B., Claramunt, C. & Chandramouli, C. (2005). Spatial logistic regression and gis to model rural-urban land conversion. *Proceedings of PROCESSUS Second International Colloquium on the Behavioural Foundations of Integrated Land-use and Transportation Models: Frameworks, Models and Applications*, (December 2015), 12–15.

Verburg, P. H., Schot, P. P., Dijst, M. J. & Veldkamp, A. (2004). Land use change modelling: Current practice and research priorities. *GeoJournal*, *61*(4), 309–324. https://doi.org/10.1007/s10708-004-4946-y

Wassmann, R., Hien, N. X., Hoanh, C. T. & Tuong, T. P. (2004). Sea level rise affecting the vietnamese mekong delta: Water elevation in the flood season and implications for rice production. *Climatic change*, *66*(1), 89–107.

Bambaradeniya, C. N. B. & Amarasinghe, F. P. (2003). Biodiversity Associated with the Rice Field Agro-ecosystem in Asian Countries : A Brief Review. *Biodiversity Associated with the Rice Field Agroecosystem in Asian Countries: A Brief Review*, (January 2003), 1–29. http://www.iwmi.cgiar.org/Publications/Working%7B%5C_%7DPapers/working/WOR63.pdf

Oldfield, F. & Dearing, J. A. (2003). The Role of Human Activities in Past Environmental Change, 143–162. https://doi.org/10.1007/978-3-642-55828-3_7

Shah, B. & Barnwell, B. (2003). Hosmer-lemeshow goodness of fit test for survey data.

Peng, C.-Y. J., Lee, K. L. & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The journal of educational research*, *96*(1), 3–14.

Weng, Q. (2002). Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modelling. *Journal of envir-*

onmental management, *64*(3), 273–284. https://doi.org/10.1006/jema.2001.0509

ADB. (2001). Linking the Poor with Rice Value Chains. *Making markets work better for the poor Briefing*, (01), 7.

Hosmer, D. W. & Lemeshow, S. (2000). Applied Logistic Regression.pdf. http://as.wiley.com/WileyCDA/WileyTitle/productCd-0470582472.html

Provost, F. & Kohavi, R. (1998). Guest editors' introduction: On applied research in machine learning. *Machine learning*, *30*(2), 127–132.

Edwards, P. & Demaine, H. (1997). Rural aquaculture: Overview and framework for country reviews. *RAP Publication (FAO).*

Anselin, L. (1995). Local indicators of spatial organization -LISA. *Geographical Analysis*, *27*(2), 93–115.

FAO. (1995). Planning for sustainable use of land resources, i–i. https://doi.org/10.1109/emap.2008.4784209

Efron, B. & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Nagelkerke, N. J. et al. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, *78*(3), 691–692.

Cox, D. R. & Snell, E. J. (1989). *Analysis of binary data* (Vol. 32). CRC press.

Johnson, N., Kotz, S., Efron, B. & Tibshirani, R. J. (1989). lnspection Errors for Attributes in Quality Control An Introduction to the Bootstrap. *J. MillerJ. Hand Inference and Asymptotics O. Barndorff-Nielsen and D.R. Cox*, *153*(56).

Gliessman, S. R. (1985). Multiple cropping systems: A basis for developing an alternative agriculture. *US Congress Off ice of Technology Assessment. Innovative biological technologies for lesser developed countries: workshop proceedings. Congress of the USA. Washington, DC, USA*, 67–83.

Hanley, J. A. & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, *143*(1), 29–36.

Dickinson, G. C. & Shaw, M. G. (1977). What Is 'Land Use'? *Area*, *9*(1), 38–42. http://www.jstor.org/stable/20001161

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159–174.

McFadden, D. et al. (1973). Conditional logit analysis of qualitative choice behavior.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic geography*, *46*(sup1), 234–240.

Url, S., Geography, E., Archive, T. J. & Archive, T. (1970). Spatial Autocorrelation : A Review of Existing and New Measures with Applications Andrew D . Cliff ; Keith Ord Commission on Quantitative Methods . (Jun ., 1970), pp . 269-292 . SPATIAL AUTOCORRELATION : A REVIEW OF EXISTING AND NEW MEASURES WITH APPLIC. *46*, 269–292.

Fisher, R. (1950). Statistical methods for research workers, llth edn oliver boyd.

Moran, P. (1950). Notes on Continuous Stochastic Phenomena Published by : Biometrika Trust Stable URL : http://www.jstor.org/stable/2332142. *Biometrika*, *37*(1), 17–23.

Appendices

Appendix A

Explanatory Data Analysis: Variable correlation plots

	dem	subsidence	compaction	popdensity	riverdistance	seadistance	salinization
aquaculture1	4	5	-16	-3	-11	10	-17
aquaculture2	10	13	13	-7	0	3	-2
aquaculture3	4	1	-10	-16	-16	-7	-7
aquaculture4	4	-2	-11	-6	-10	2	-12
rice1	11	9	3	2	-17	3	6
rice2	13	8	-2	4	-13	3	4
rice3	3	8	7	1	-7	-4	11
rice4	3	3	0	5	-3	-6	13
mixedcrops1	1	-10	-25	-6	14	16	-11
mixedcrops2	0	1	-19	-19	23	19	0
mixedcrops3	-8	-9	-12	-2	0	1	-4
mixedcrops4	3	13	-9	16	-35	-18	16
orchard1	-18	-12	7	9	39	-5	3
orchard2	-17	-20	2	5	16	7	-15
orchard3	-12	-10	10	4	40	-8	11
orchard4	-19	1	19	4	44	10	-6
urban1	-33	-4	-28	-59	23	-22	28
urban2	-33	-2	-19	-48	21	-23	25
urban3	-39	-18	1	-60	40	-12	23
urban4	-21	-6	-9	-33	22	-6	10

Figure A.1: a. Correlation Plot for Land use Contraction and Explanatory Variable

	dem	subsidence	compaction	popdensity	riverdistance	seadistance	salinization
aquaculture1	-17	10	26	-9	3	-25	30
aquaculture2	-11	4	16	-8	-3	-19	18
aquaculture3	-15	11	25	-7	3	-23	28
aquaculture4	-10	7	13	-1	-1	-14	18
rice1	11	-15	-19	14	31	30	-25
rice2	5	-16	-16	11	20	29	-26
rice3	6	-3	-8	15	24	13	-9
rice4	7	7	-4	18	14	19	-19
mixedcrops1	-2	4	5	3	1	-8	11
mixedcrops2	-2	3	7	4	3	-9	10
mixedcrops3	-2	3	4	2	1	-7	9
mixedcrops4	1	-4	-2	-2	11	2	4
orchard1	6	13	-4	7	-18	7	-2
orchard2	8	18	-1	11	-5	5	3
orchard3	3	8	-1	5	-9	5	2
orchard4	4	1	-11	5	-3	6	1
urban1	5	2	1	6	-4	2	-2
urban2	2	1	2	3	-1	0	0
urban3	4	2	1	6	-4	2	-2
urban4	5	2	3	9	-5	-2	3

Figure A.2: b. Correlation Plot for Land use Expansion and Explanatory Variable

Figure A.3: Spearman's Correlation Plots

Appendix B

Histogram Plot of P Values of Significant Variables

B.0.1 Land Use Contraction

The histogram plots of the p values for the significant variables in land use contraction for each land use class presented in the results section above, are shown in this section.

Aquaculture Land Use Contraction

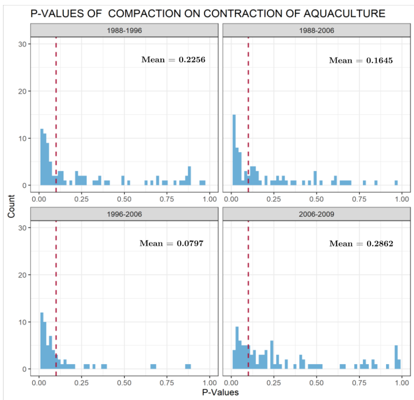

Figure B.1: a. P values of Compaction on Aquaculture Land use Contraction

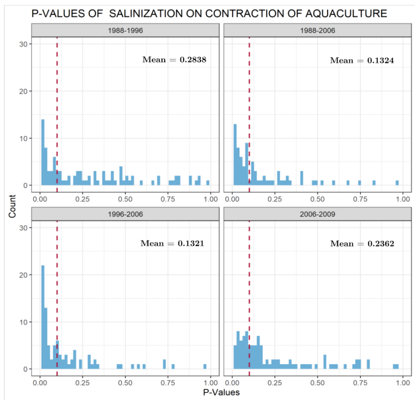

Figure B.2: b. P values of Salinity on Aquaculture Land use Contraction

Figure B.3: P values of Significant Variables on Aquaculture Contraction

Mixed Crops Land Use Contraction

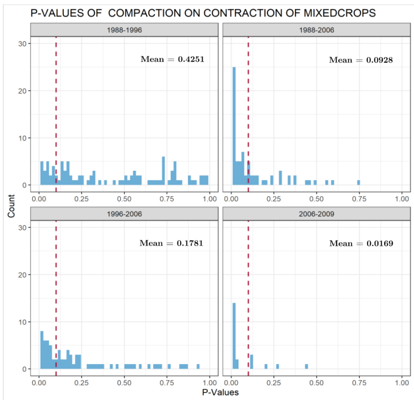

Figure B.4: a. P values of Compaction on Mixed Crops Land use Contraction

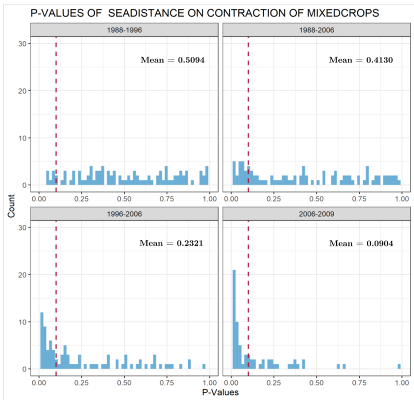

Figure B.5: b. P values of Sea Distance on Mixed Crops Land use Contraction

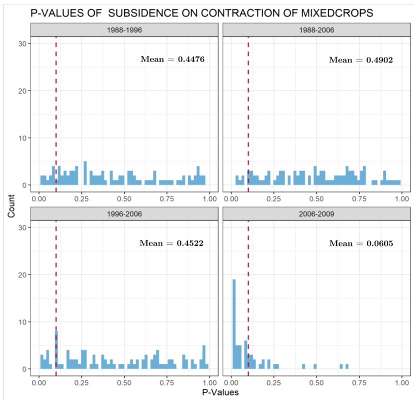

Figure B.6: b. P values of Subsidence on Mixed Crops Land use Contraction

Figure B.7: P values of Significant Variables on Mixed Crops Contraction

Orchard Land Use Contraction

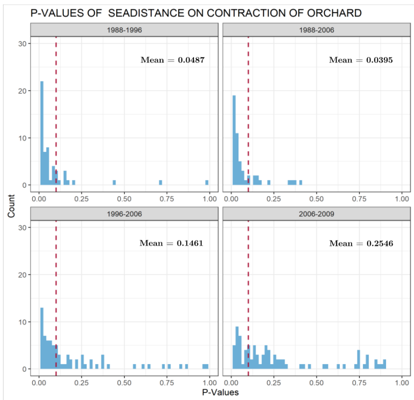

Figure B.8: a. P values of Sea Distance on Orchard Land use Contraction

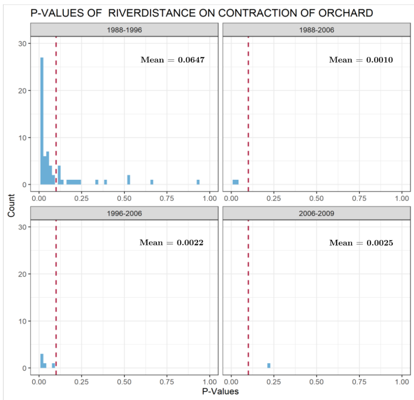

Figure B.9: b. P values of River Distance on Orchard Land use Contraction

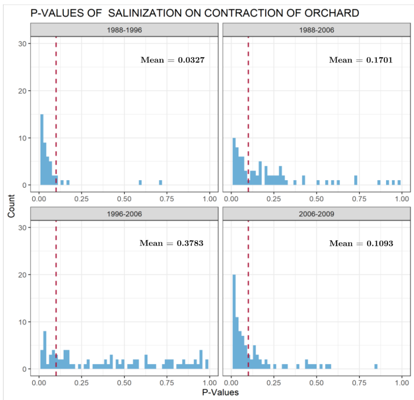

Figure B.10: b. P values of Salinity on Orchard Land use Contraction

Figure B.11: P values of Significant Variables on Orchard Contraction

Urban Land Use Contraction

Figure B.12: a. P values of Elevation on Urban Land use Contraction

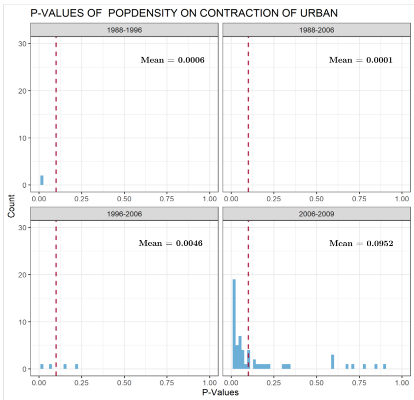

Figure B.13: b. P values of Population Density on Urban Land use Contraction

Figure B.14: b. P values of Sea Distance on Urban Land use Contraction

Figure B.15: P values of Significant Variables on Urban Contraction

B.0.2 Land Use Expansion

The histogram plots of the p values for the significant variables in land use expansion for each land use class presented in the results section above, are shown in this section.

Aquaculture Land Use Expansion

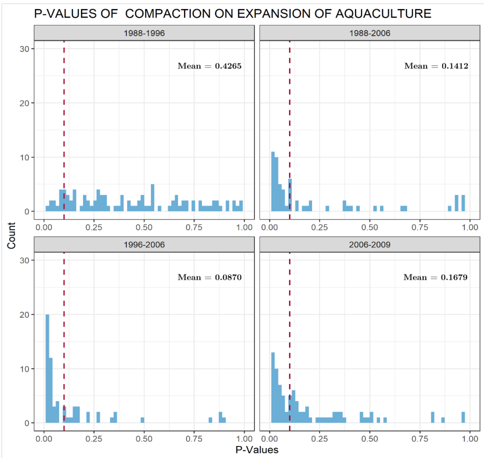

Figure B.16: a. P values of Compaction on Aquaculture Land use Expansion

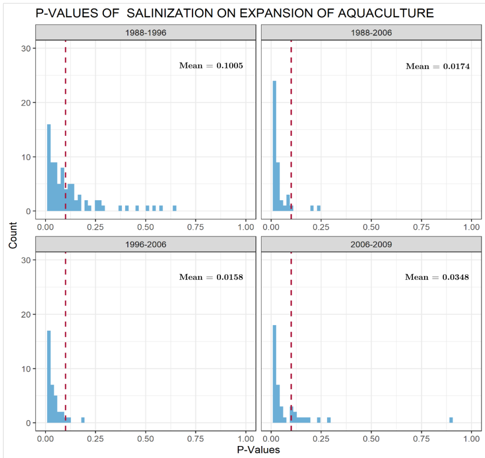

Figure B.17: b. P values of Salinity on Aquaculture Land use Expansion

Figure B.18: P values of Significant Variables on Aquaculture Expansion

Mixed Crops Land Use Expansion

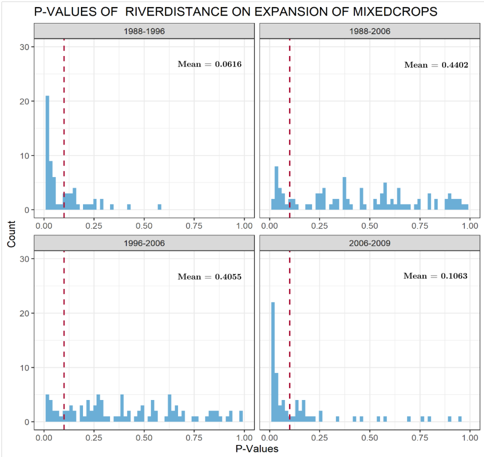

Figure B.19: a. P values of River Distance on Mixed Crops Land use Expansion

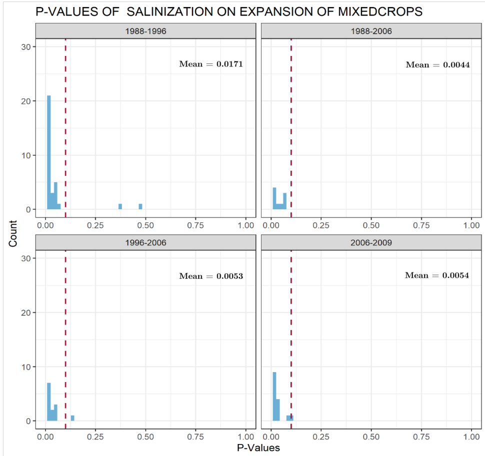

Figure B.20: b. P values of Salinity on Mixed Crops Land use Expansion

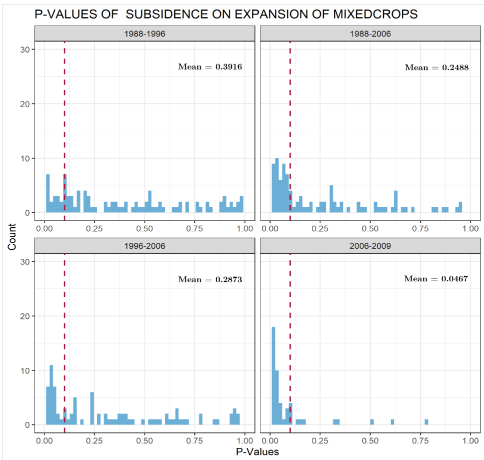

Figure B.21: b. P values of Subsidence on Mixed Crops Land use Expansion

Figure B.22: P values of Significant Variables on Mixed Crops Expansion

Orchard Land Use Expansion

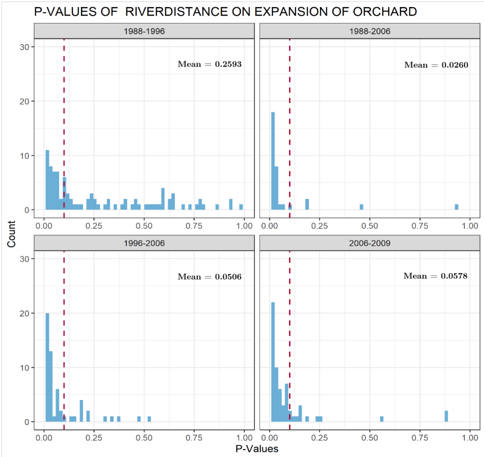

Figure B.23: a. P values of River Distance on Orchard Land use Expansion

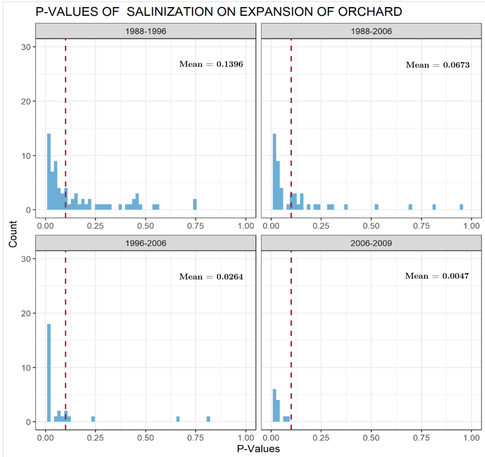

Figure B.24: b. P values of Salinity on Orchard Land use Expansion

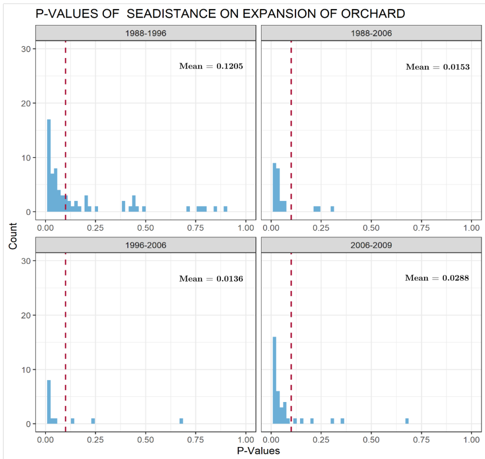

Figure B.25: b. P values of Sea Distance on Orchard Land use Expansion

Figure B.26: P values of Significant Variables on Orchard Expansion

Rice Land Use Expansion

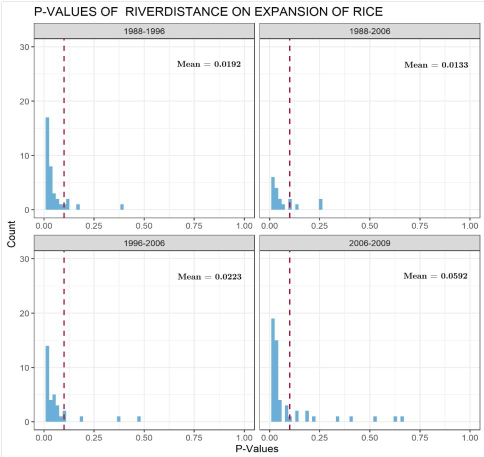

Figure B.27: a. P values of River Distance on Rice Land use Expansion

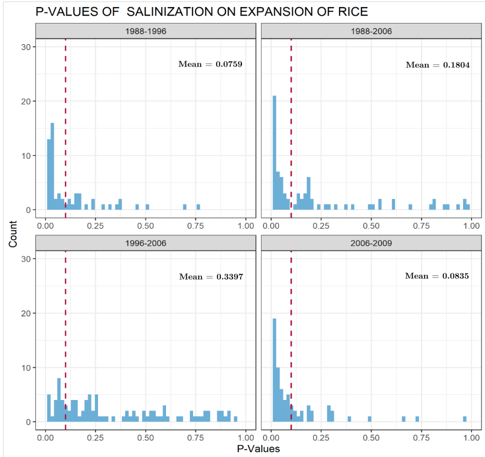

Figure B.28: b. P values of Salinity on Rice Land use Expansion

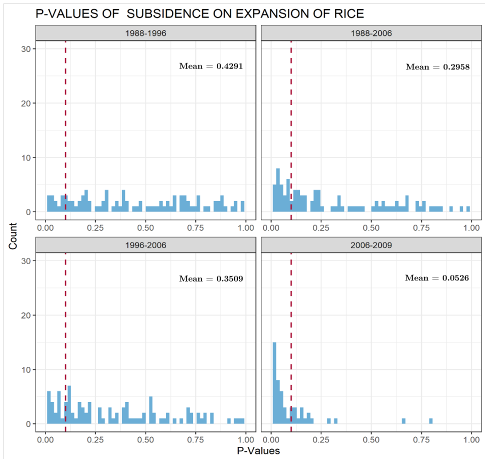

Figure B.29: c. P values of Subsidence on Rice Land use Expansion

Figure B.30: P values of Significant Variables on Rice Expansion

Urban Land Use Expansion

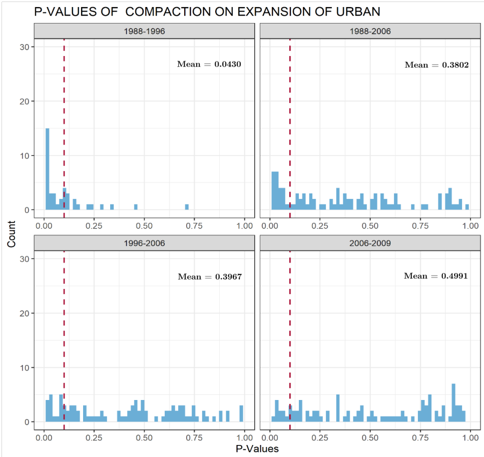

Figure B.31: a. P values of Compaction on Urban Land use Expansion

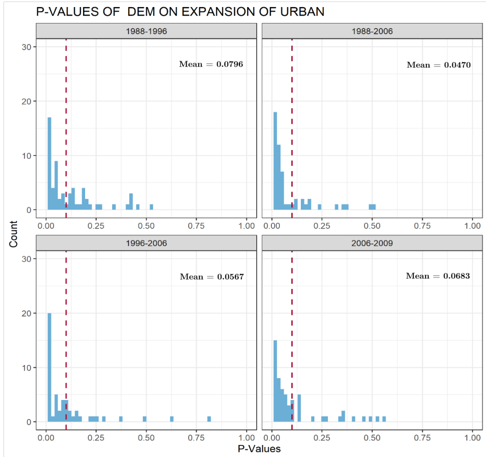

Figure B.32: b. P values of Elevation on Urban Land use Expansion

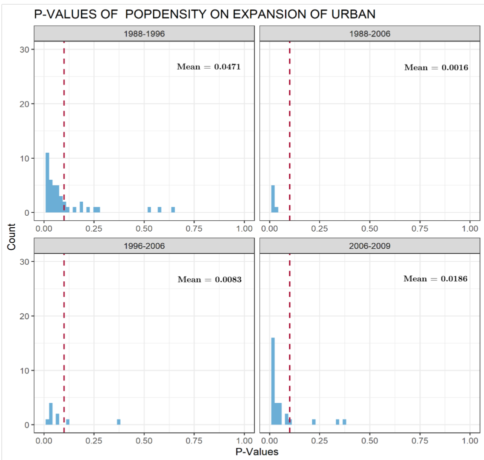

Figure B.33: c. P values of Population Density on Urban Land use Expansion

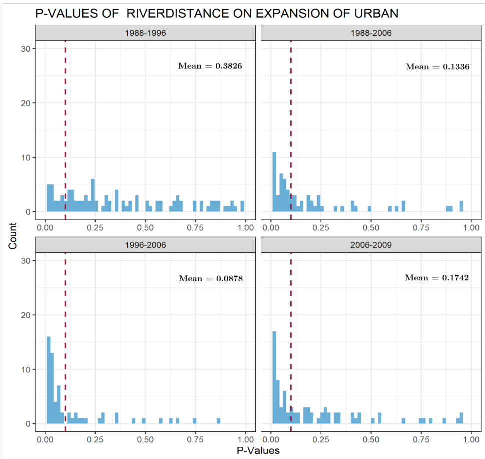

Figure B.34: d. P P values of River Distance on Urban Land use Expansion

Figure B.35: P values of Significant Variables on Urban Expansion