HYDRO-GEOMORPHOLOGICAL ATTRIBUTES AND DISTRIBUTION OF URBAN LAND AND POPULATION IN RIVER BASINS

A Dissertation

by

YING LIU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Inci Güneralp
Committee Members,	Burak Güneralp
	Huilin Gao
	Anthony Filippi
	John Giardino
Head of Department,	Ronald Kaiser

August 2016

Major Subject: Water Management and Hydrological Science

ABSTRACT

Global urbanization is a major trend in the 21st century and imposes a significant impact on the hydrologic cycle, climate, and biodiversity. In particular, Asia and Africa are expected to experience the fastest urbanization rate in the following decade, manifested by impervious land cover construction and urban population growth. Therefore, accurate information on spatial distribution of urban land and population at the present stage can attribute to a better understanding of urbanization processes in the future. Meanwhile, the knowledge of hydrological responses on the land use and land cover (LULC) change can provide valuable insights into sustainable development strategies.

This dissertation aims to answer three fundamental questions: (1) How do natural environmental factors that relate to water resource, climate, and geomorphological attributes constrain the distribution of current urban land and population? (2) How will the global urban growth in the near future affect the urban exposure to natural disasters such as fluvial flood, drought, and ecosystem degradation? (3) How will land change impact hydrologic processes within a river basin?

To solve the research questions, studies presented in this dissertation are organized into three parts. Part one analyzes the spatial distribution of urban ratio, population density, and urban population density in 11 river basins in Asia and Africa, considering average annual precipitation, surface freshwater availability, and access to coastal zone as three influencing factors. Then, a set of regression models is conducted for the Yangtze River Basin as a more comprehensive investigation. Part two examines the global and regional patterns of urban growth from 2000 to 2030 as well as the change of urban area's

exposure to floods and droughts. Part three is an assessment of streamflow in the Chao Phraya Basin based on different precipitation and LULC scenarios. The Soil and Water Assessment Tool (SWAT) is used to develop the hydrological model.

The results reveal that higher urban ratio and human concentration occur in the vicinity of a stream network and a coastal zone; while precipitation does not effectively influence the distribution of urban land and population as expected. The emerging coastal metropolitan regions in Africa and Asia will be larger than those in the developed countries and will have larger areas exposed to flood and drought. The case study of Chao Phraya River Basin demonstrates that land change will increase both the risk of drought and flood hazards.

DEDICATION

I dedicate this work to my loving parents.

ACKNOWLEDGEMENTS

I would like to acknowledge the inspirational instruction and guidance from my advisor, Dr. Inci Güneralp. Her insightful understanding of the research was always the source of power that initiated and expanded my research work. She is knowledgeable in hydrology and geomorphology, so I obtained a tremendous amount of know-how just by working with her on all these projects. She was strict on the quality of the research, meanwhile encouraged me to explore. With her guidance, support, and encouragement, I was able to push my way through numerous obstacles along the way.

I also want to express my sincere gratitude to my committee members. Dr. Burak Güneralp directed me in research on spatial analysis. His background and expertise in urbanization supplied solid support for me. Dr. Anthony Filippi helped me to establish links from GIS to hydrological modeling. Dr. John Giardino taught me to form a good habit of writing. Dr. Huilin Gao provided valuable comments to this dissertation.

I want to thank all my "academic siblings" in the fluvial geomorphology team: Billy Hales, Cesar Ricardo Castillo, and Mingde You. We shared happiness and sadness. Having them as friends has made my life much easier.

Finally, I would like to thank my family in China. Their love and trust are important for me to finish my study.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

Page

LIST OF TABLES

Page

Page

CHAPTER I

INTRODUCTION

1.1 Background

Since time immemorial, humans always attempt to change landscapes to improve the amount, quality, and security of natural resources that are critical to their living standards. Through the increased use of innovation, human populations have increased their ability to derive resources from the environment.

Land change increased markedly after the Industrial Revolution, both in terms of extent and intensity. Then, the transition from an agricultural to an industrial society initiated rapid-paced urbanization, with a large-scale migration of people from rural places to cities. Nowadays, there is a consensus that urbanization is one of the major trends in the 21st century. At the beginning of this millennium, 2.8 billion people, or 47% of the global population, lived in the urban areas (UNWUP 2005). Given the rapid pace of urbanization and overall growth of the world's population, by 2030, two-thirds of the global population would live in cities (UNDESA, 2014), which means 1.7 billion people are expected to become new urban residents within only 30 years. Meanwhile, urban land is forecasted to increase by approximately 1.2 million km^2 by 2030 (Güneralp & Seto, 2013; Seto et al., 2012).

For the past few decades, most of the demographic transition and land change brought about by urbanization has been happening in developing countries. However, due to the lack of infrastructure and financial support, such developing countries have limited

capacity to cope with the rapid change. This situation would inevitably have consequences for urban residents' well-being, including their vulnerability to various natural and man-made hazards (Srinivasan et al., 2013; Vörösmarty et al., 2000) and scarcity of various natural resources foremost being water resource scarcity (Harhay, 2011; McDonald, Douglas et al., 2011; McDonald, Green, et al., 2011). Therefore, the challenge of maintaining human well-being is great.

By contrast, from an environmental perspective, urban expansion imposes a significant stress on natural resources, especially on water resources and the local ecosystem. Previous studies about urban impact on the hydrological cycle have reached common conclusions: conversion of land use from original states to impervious cover leads to a lower evapotranspiration (ET) rate and a corresponding increasing runoff in urban regions during a precipitation event, thus restricts groundwater recharge and exacerbates flood risks (Claessens et al., 2006; Cuo et al., 2008; Hollis, 1975; Nagy et al., 2012). This alteration may further result in negative implications to biodiversity and ecosystem services (Krause et al., 2004; McBride & Booth, 2005).

Even though urban growth after industrialization is mainly driven by economic factors rather than mere natural resources, it can be treated as part of the long trajectory of human habitat development; hence, theories and methods applied in the research of pre-industrial cities are still available for current study. When considering cities as an engine of economic growth, it is reasonable to hypothesize that urban land expansion follows a path-dependent manner over an extended period, similar to the development of economies. Early cities were formed mainly based on natural advantages such as adequate freshwater

resource, moderate climate and suitable topography attributes. When people had a surplus of food to trade, new cities were developed along rivers, because the shipment of goods by boat was a dominant form of transportation until the early to mid-19th century, when canals and railways were developed. After the Maritime Silk Road was formed, coastal zone also became an optimum place to settle down, because the overseas trade provided better infrastructures and living standards. Menze and Ur (2012) mapped human settlements accumulated over eight millennia in Northern Mesopotamia and indicated the settlement's size increases exponentially as the precipitation and the Strahler order of river segment increases. Woodroffe et al. (2006), Syvitski and Saito (2007), and James P. M. Syvitski et al. (2009) showed many of the world's largest deltas are still dense-populated currently. These findings prove that urban development is process-dependent, previous locational formations. Once natural or economic advantages helped determine the location of early cities, the later development will take place around the focal point.

Development of remote sensing techniques enables deeper and more accurate analysis of urban spatial distribution. Detailed spatial and temporal information of urban morphology, infrastructure, and land use/land cover patterns are provided. Based on the remote sensing data, Small and Cohen (2004) concluded that the distribution of global population and land area are functions of individual parameters such as proximity to the coast or annual precipitation. Miyamoto et al. (2011) analyzed 109 river basins in Japan and discovered that human population and land use in Japan also have scale-invariant fractal structures based on stream's Strahler order, similar to the laws of drainage

composition. All these research prove that natural environmental factors still play a major role in urban land and population distribution.

Most of the previous works concentrated on impacts of urbanization on the hydrological, geomorphological or ecological system in isolation. Moreover, limited quantitative evidence is shown for the regularity and uniformity of urban land and population distribution at a river basin scale, and research seldom mentioned the constraints of environmental factors on urban development. Herein, I take a long-term view and analyze the potential relationships between urban land and population distribution and hydro-geomorphological attributes of river basins. Results of present research can address sources of urban land exposure and vulnerability to natural hazards and offer valuable suggestions for decision-making on urban planning and water management.

1.2 Objectives

The overall goal of this dissertation is two-fold: 1) to provide a new interpretation of urbanization pattern based on natural indicators, especially the ones with respect to water resources; and 2) to obtain a better understanding of the hydrological responses to land change. The specific objectives are:

1. Develop a stream network dataset for selected major river basins where urbanization rate are expected to be the fastest till 2030.

2. Investigate the relationship between hydro-geomorphological attributes and the current distribution of urban land using spatial analysis and linear regression models.
3. Examine the global and regional patterns of urban growth from 2000 to 2030 and the change of urban area's exposure to floods and droughts.
4. Build a hydrological model to simulate the responses of streamflow to land use/land cover change in one of the selected river basins.

1.3 Organization of dissertation

The entire dissertation consists of six chapters. After this introduction chapter, Chapter II estimates the changing exposure of urban infrastructure to floods and droughts resulted in urban land expansion from 2000 to 2030. Chapter III performs spatial analysis on the urban land and population distribution in 11 major river basins in Asia and Africa, in order to test a hypothesis that, at a river basin scale, limitations on surface water availability and local topography will affect urban development. Chapter IV focuses on developing linear regression models on urban land distribution and the hydro-geomorphological attributes. Both the original least square (OLS) and the geographically weighted regression (GWR) methods are employed. A hydrological model built up using SWAT for the Chao Phraya River Basin in Southeast Asia is presented in Chapter V, addressing the impact of land change on precipitation-runoff generation. The major conclusions, limitations, and recommendations for future work are given in Chapter VI.

CHAPTER II

EXPOSURE OF URBAN AREAS TO WATER-RELATED HAZARDS: A FORWARD-LOOKING PERSPECTIVE*

2.1 Introduction

Contemporary urbanization will have far-reaching consequences for sustainability and human well-being. By 2030, it is estimated that almost 5 billion people (60% of the world's population) will live in cities compared to 2.9 billion in 2000 (47%) (United Nations, 2012). Where this urbanization will take place, which urban areas will grow the fastest or will end up the largest can significantly affect the number of people and infrastructure exposed to various hazards. In particular, water-related hazards are the most ubiquitous of all because, for ages, transportation and communication advantages together with abundant food mean that cities have historically been located on or near water bodies, typically on riverine or coastal flood plains.

Globally, about 10 percent of the world's population and 13 percent of the world's urban population live in the low-elevation coastal zones that cover 2 percent of world's land area (Gordon McGranahan et al., 2007). In addition, twenty-one of the 33 megacities are located on or near the coast (Klein et al., 2003). Already coastal and fluvial floods cause significant damage to infrastructure and affect the livelihoods of millions every year.

*Reprinted from Global Environmental Change, Güneralp, B., Güneralp, İ., & Liu, Y, Changing global patterns of urban exposure to flood and drought hazards, 31, 217-225, Copyright (2015), with permission from Elsevier.

Water scarcity poses a different kind of water-related hazard. While most urban populations are located along coastal areas, urbanization in drylands can also be significant (Maestre et al., 2012; Sherbinin et al., 2012). G. McGranahan et al. (2005) estimated that, as of the year 2000, 45% of total dryland population was urban and urban land in drylands was twice as large as in the coastal zone. Particularly exposed to water scarcity are those cities located in arid regions of the world (Gober, 2010). McDonald, Douglas, et al. (2011) identified that over half a billion urban residents face chronic water availability challenges and over a billion live in cities with inadequate water delivery infrastructure. The increase in urban population and infrastructure in drylands also intensify the water-use conflicts with agriculture (Portnov & Paz, 2008).

The interaction between continued socio-economic changes and climatic changes will further increase the exposure and vulnerability of people and assets to water-related hazards in the near future. Population and economic assets exposed to a 1 in 100-year coastal flood event by 2070 in 136 port cities across the world are expected to increase more than three and ten-fold, respectively (Hanson et al., 2011). A recent global assessment estimated that economic exposure to coastal and fluvial flooding could increase from 27 trillion USD in 2010 to 80 trillion USD in 2050 if based on land use and from 46 trillion USD to 158 trillion USD if based on population (Jongman et al., 2012). On the other hand, the number of urban residents in perennial water shortages is forecasted to increase more than five times to 160 million by 2050 due for the most part to population growth but also as a result of climate change (McDonald, Green, et al., 2011).

In the context of water-related risks, this means a high concentration of people and infrastructure that will be exposed to floods or droughts or both (de Sherbinin et al., 2007; Hanson et al., 2011; McDonald, Douglas, et al., 2011). However, none of these studies took into account the increase in the extents of the urban areas. They also typically focus on the change solely in population or economic activity in usually a subset of urban areas to quantify the evolution of their exposure to one or more water-related hazards. So far, a detailed spatially-explicit investigation of the increasing footprint of urban areas that may be exposed to a wider range of water-related hazards has been lacking.

This study focuses on the potential increase in the exposure of urban areas to water-related hazards due to socio-economic factors that drive urban expansion. Building on the first global urban land expansion forecasts (Güneralp & Seto, 2013; Seto et al., 2012), we focus on potential urban expansion in areas with high exposure to drought or flood or to both. Since these issues are intimately related to hydrology we look at urban expansion patterns across the river basins as well as by geographical regions to explore two questions:

1. What was the distribution of urban land in 2000 across the world's watersheds and which watersheds are expected to undergo the most urban expansion in the near future?

2. Which geographical regions and watersheds around the world will experience the highest amounts and rates of urban expansion in lands particularly prone to droughts, flooding or both?

2.2 Methods and data

We use the only available spatially-explicit global forecasts of urban expansion (Güneralp & Seto, 2013; Seto et al., 2012). The base urban map for the forecasts comes from the Collection 5 MODIS Land Product (Schneider et al., 2009), which reflects the global land cover around the year 2000; the forecasts are for the year 2030. These forecasts take into account both the uncertainty in the population and economic growth (two major drivers of urbanization) and varying suitability of each land parcel for urban expansion. Therefore, in the resulting urban expansion forecasts, each land ends up with a certain probability to become urban. In this study, we focus on the areas with high probability to become urban following Seto, et al. (2012), that is, those areas that have >75% probability to undergo urban expansion by 2030.

We use the data set of major river basins in vector form, developed by the Food and Agriculture Organization (FAO, 2009). This vector dataset is derived from HydroSHEDS, a more recent elevation product with a higher spatial resolution (Lehner et al., 2008) than Hydro1k (U.S. Geological Survey, 2000). Compared to Hydro1k, HydroSHEDS provides more accurate delineation of the river basins in flat regions such as coastal zones. The FAO dataset provides complete global coverage including areas north of the original HydroSHEDS extent, i.e., 60° N, obtained by merging HydroSHEDS with the HYDRO1k basin layer. Importantly, the FAO dataset provides detailed information of each river basin, including the sub-basin boundaries and their names. A recent meta-analysis on urban expansion indicated that urban land expansion has been occurring faster in this zone than in other areas around the world (Seto et al., 2012). To

track forecasted urban expansion in flood-prone areas, we use the flood frequency map created at the Center for Hazards + Risk Research (CHRR) and the Socio-Economic Data and Applications Center (SEDAC) at the Center for International Earth Science Information Network, CIESIN (Trabucco & Zome, 2009) based on data compiled by the Dartmouth Flood Observatory (DFO, 2014). The dataset reflects the relative distribution and frequency of global flood hazard based on recorded incidences of extreme flood events between 1985 and 2003. The gridded flood frequency map is at a 2.5-minute resolution and consists of 10 classes where larger grid cell values denote higher the relative frequency of flood occurrence. We further aggregate these classes into high, medium, and low-frequency floods and focus on the high-frequency floods in our analysis. Droughts are typically associated with aridity (Dai, 2010; Seager et al., 2007). Therefore, to track forecasted urban expansion in drought-prone areas, we use the Global Aridity Index dataset from the Consortium for Spatial Information at the Consultative Group on International Agricultural Research (CGIAR-CSI) (CGIAR-CSI, 2014; Zomer et al., 2008). We particularly focus on urban areas within hyperarid, arid, and semiarid lands.

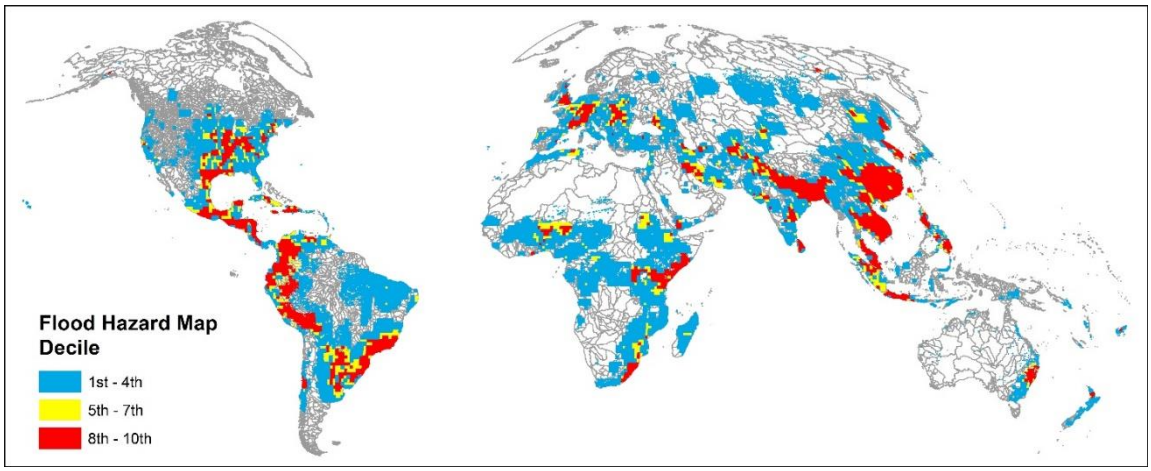


Figure 2.1 Low-, medium-, and high-frequency (respectively, 1st–4th, 5th–7th, and 8th–10th deciles of the distribution) flood zones based on the data compiled by the Dartmouth Flood Observatory (DFO, 2014).

We first analyze the distribution of urban land across the world's watersheds in 2000 and as forecasted in 2030. The spatial analyses to identify the extent of urban land cover in areas prone to water-related hazards involved intersecting the urban extent maps for the year 2000 and 2030 with the flood frequency (Figure 2.1) and global aridity index maps (Figure 2.2). This also allowed us to assess the expected change in the exposure of urban areas to fluvial and coastal flooding as well as droughts. We compare the forecasted change in exposure of urban areas to flooding and drought at a watershed level as well as regionally and globally. All analyses are conducted at the resolution of the urban expansion forecasts, 25km^2.

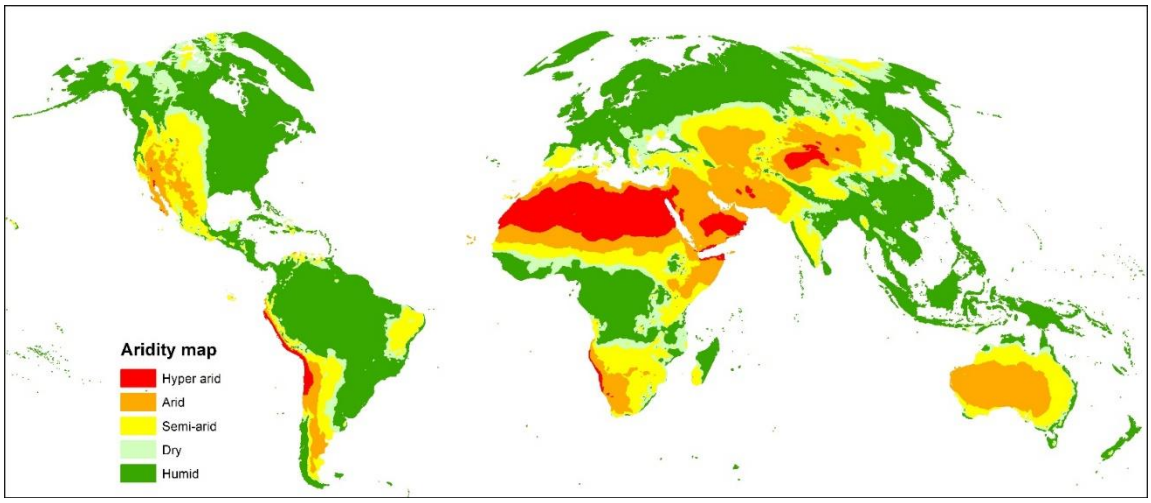


Figure 2.2 Global aridity index map (Zomer et al., 2008 and CGIAR-CSI, 2014).

2.3 Results

Studies show that there is a surprisingly well-defined relationship between percent impervious cover in a watershed and the level of degradation (Arnold & Gibbons, 1996; Schueler et al., 2009). Around a threshold of 10% ± 5% impervious cover in a watershed, signs of degradation in water quality and hydrology in the form of increased flooding risk and reduced aquatic habitat quality become noticeable and progressively become more severe as the percent impervious cover increases. This relationship is, however, shown to hold for much smaller watersheds than those we consider in our study (Schueler et al., 2009). Therefore, we loosely follow these findings in presenting our findings on percent urban cover across the world's watersheds simply to point at the highest geographical concentrations of watersheds likely to be stressed due to relatively large percentages of impervious surface cover (Figure 2.3).

We find that, even by the year 2000, the initial year of our study, several watersheds along the Chinese coast and in most of the rest of East Asia were significantly urbanized (Figure 2.3a). In the developed world, many watersheds in Western Europe, northeastern U.S. along with California and Florida were also highly urbanized. These results are broadly in agreement with those reported in Elvidge et al. (2007) who estimated impervious surface cover in watersheds around the world based on the satellite observed nighttime lights and population counts.

By 2030, the proportion of watersheds 10% or more urbanized is expected to increase from 3% to 8% (Figure 2.3b; Figure 2.4). The largest percentages of watersheds that are expected to have over 10% urban cover will still be in the developed regions of the world, specifically in Eastern Asia and Western Europe (Figure 2.4). Nevertheless, in large parts of Africa and Asia including China and India, the percentage of watersheds with significant urban cover will increase dramatically. In particular, watersheds that are expected to be over 10% urbanized by 2030 are concentrated certain localities in Asia, Africa, and South America. These are along the seaboard of China, southwestern shores of India and in Brahmaputra basin in Asia; the mouth of Niger River, and northern shores of the Lake Victoria, in Africa; and the northwest coast of South America. Some –but not all– of these watersheds belong to the larger basins such as the Niger and Yangtze. These basins are notable for their mega-deltas that house an increasingly larger share of their population in urban areas (Seto, 2011).

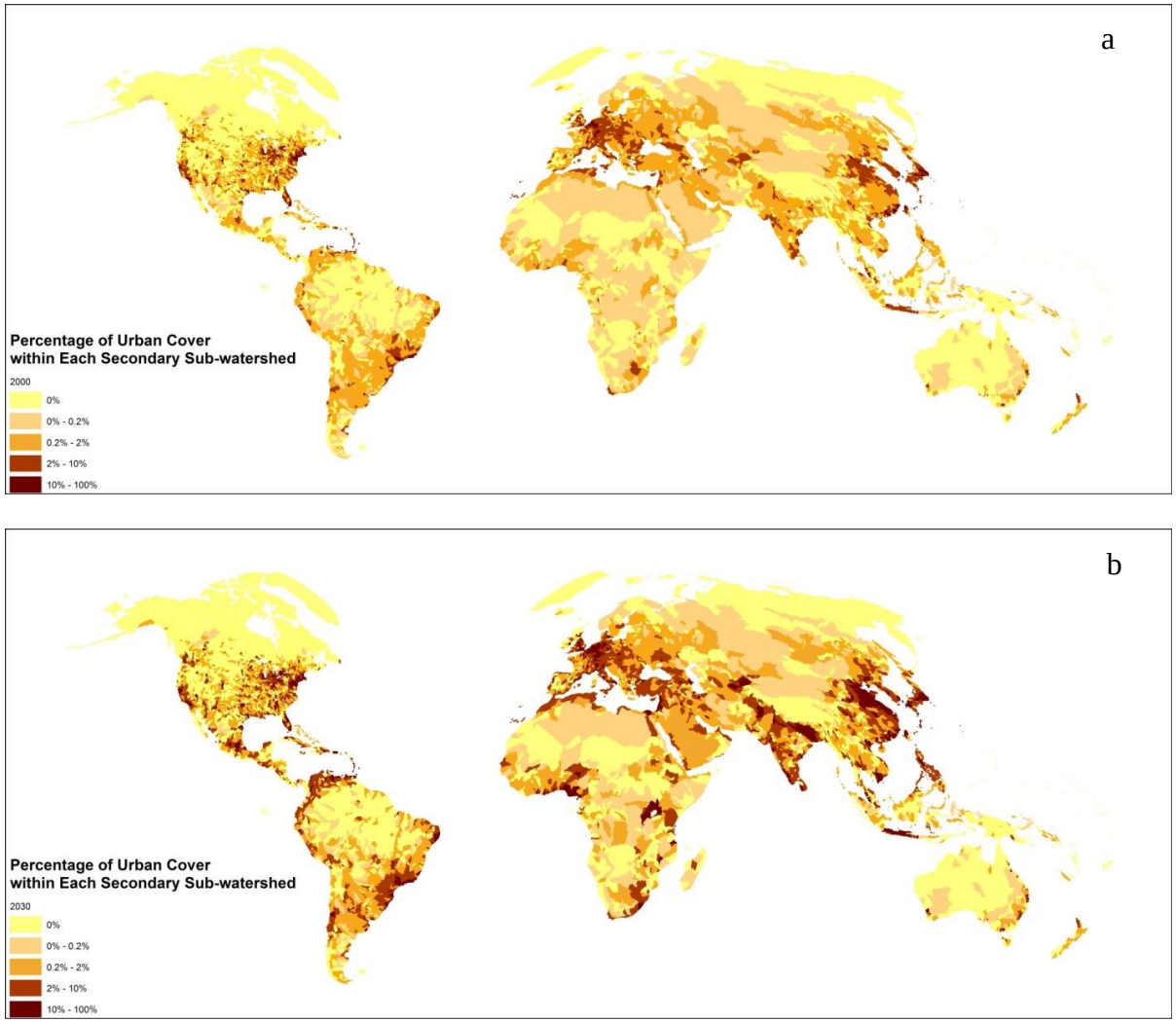


Figure 2.3 Percentages of basin areas occupied by urban land in a) 2000 and b) 2030.

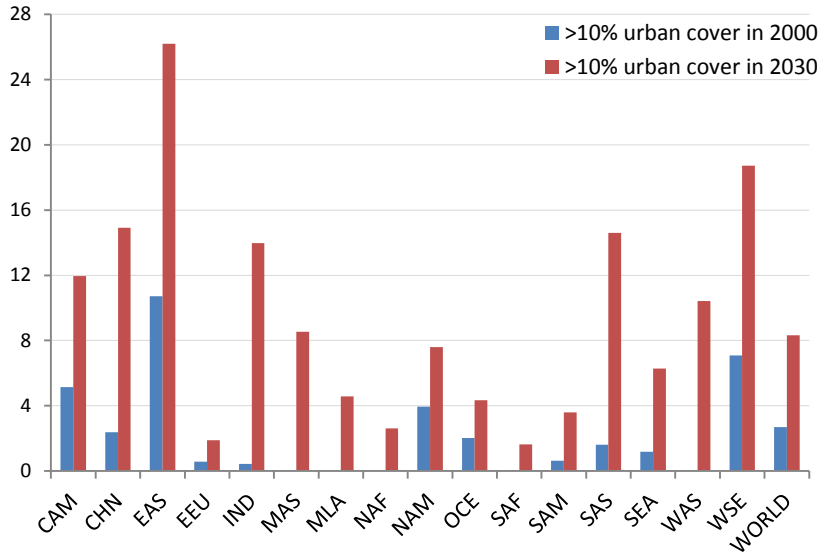


Figure 2.4 Percentages of basins with 10% or more urban land cover in 2000 and 2030.

2.3.1 Urban expansion in flood-prone areas

In 2000, about 30% of the global urban land (nearly 200,000 km^2) was located in the high-frequency flood zones (Table 2.1). The urban areas with high flood risk were mostly concentrated within watersheds of Rhône, Danube, and Seine rivers in Western Europe (Figure 2.6a). In the developing world, over half of all the urban land in Southern and Southeast Asia was already within the high-frequency flood zones. Specifically, large concentrations of urban land within watersheds that are prone to high-frequency floods were located within the Chao Phraya Basin in Thailand, Indus and Ganges in India and Pakistan, Yangtze and Pearl River Deltas in China as well as across the Korean peninsula. Sao Paolo urban agglomeration in South America was also largely located within the high-frequency flood zone.

By 2030, most of the urban growth, over 500,000 km^2, will take place in the high-frequency flood zones, thus in highly exposed locations around the world (Figure 2.5a and Table 2.1). Thus, the percentage of urban land will reach 40% by 2030. Almost everywhere except Eastern Asia and North Africa, there will be more urban expansion in high-frequency flood zones than elsewhere (Table 2.1). South America and China will continue to be among those with the largest urban land in the high-frequency flood zones in 2030. By 2030, both regions are expected to have over 80% of urban land under high-frequency flood risk. The largest amounts of increase are expected in the watersheds of Indo-Gangetic plain, Mekong Delta, and Yellow and Yangtze deltas in Asia; and Port Harcourt area, around Lake Victoria in Midlatitudal Africa, and around Durban in South Africa (Figure 2.6b). Notably, the largest percentage increase in urban lands faced with high-frequency flood risk is expected in Midlatitudal Africa (Table 2.1), mostly due to urban expansion around Lake Victoria (Figure 2.6).

Table 2.1 The amount and percentage urban land in high-frequency flood zones in 2000 and 2030. The increase in percentage indicates that more urban expansion is expected in high flood zone than elsewhere.

Region	Urban extent in high-frequency flood zones		Percent urban in high-frequency flood zones		Urban extent in hyperarid, arid, semiarid zones		Percent urban in hyperarid, arid, semiarid zones		Total urban extent	
	2000	**2030**	**2000**	**2030**	**2000**	**2030**	**2000**	**2030**	**2000**	**2030**
CAM	4,897	23,091	36	42	5,400	19,725	40	36	13,500	54,525
CHN	27,579	107,055	34	36	19,875	60,850	25	20	80,525	300,225
EAS	9,893	16,569	35	29	0	0	0	0	28,075	57,875
EEU	6,471	7,059	9	10	7,725	7,975	11	11	70,275	74,250
IND	15,903	120,556	52	88	13,875	34,250	46	25	30,400	137,675
MAS	550	2,300	3	6	16,325	39,975	99	98	16,500	40,725
MLA	2,092	56,546	11	28	4,475	32,525	23	16	19,675	199,800
NAF	687	2,542	5	4	11,925	54,100	89	90	13,350	60,225
NAM	34,968	67,088	27	27	29,525	48,025	23	19	130,500	248,675
OCE	1,569	3,133	15	16	1,700	2,600	16	13	10,450	20,150
SAF	786	7,036	9	27	5,675	15,125	67	58	8,425	25,900
SAM	35,412	107,290	44	50	19,125	40,400	24	19	80,025	214,075
SAS	7,967	65,282	49	74	13,675	64,725	84	73	16,250	88,650
SEA	18,725	78,921	69	82	225	875	1	1	27,275	96,725
WAS	720	3,695	3	4	19,400	67,400	72	75	26,800	89,425
WSE	27,274	55,004	34	36	4,275	9,275	5	6	80,800	154,400
WORLD	195,493	723,167	30	39	173,200	497,825	27	27	652,825	1,863,300

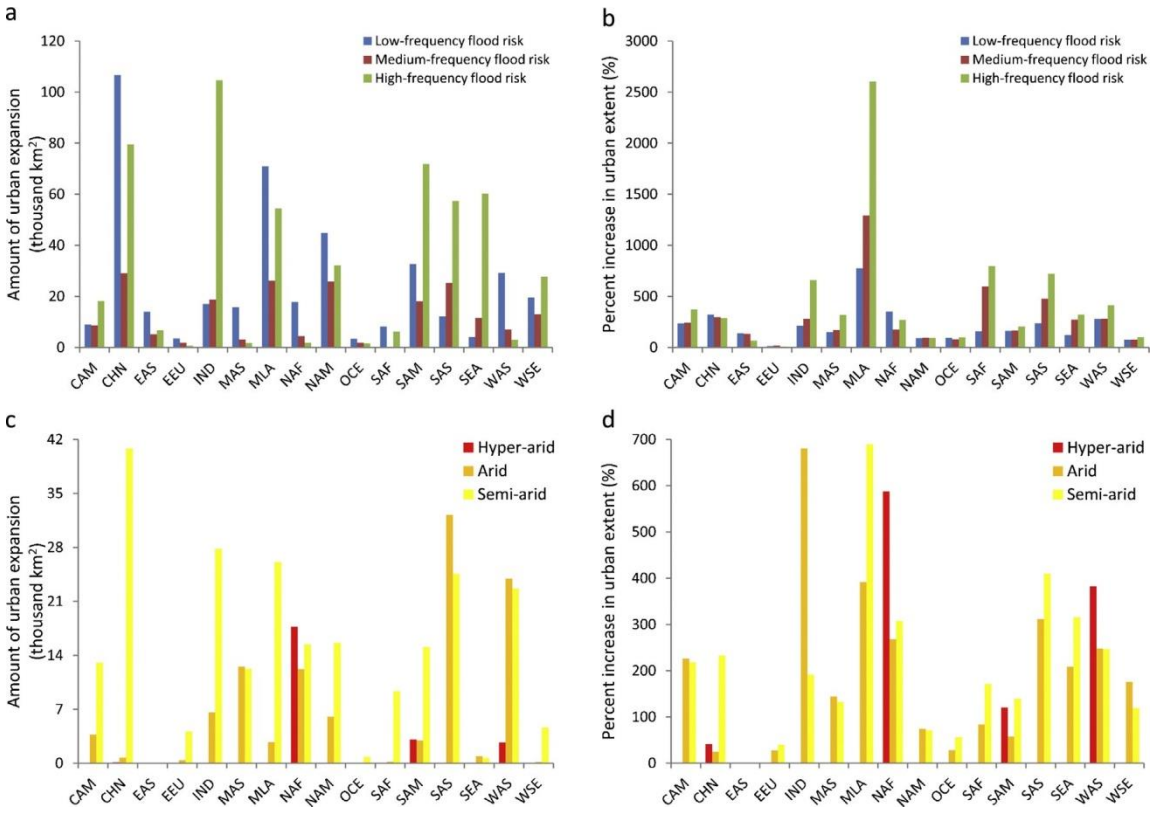


Figure 2.5 The forecasted amount of urban expansion (a) and percent increase in urban extent (b) in the three flood frequency zones by geographical region. The forecasted amount of urban expansion (c) and percent increase in urban extent (d) in the most-severe three aridity zones by geographical region.

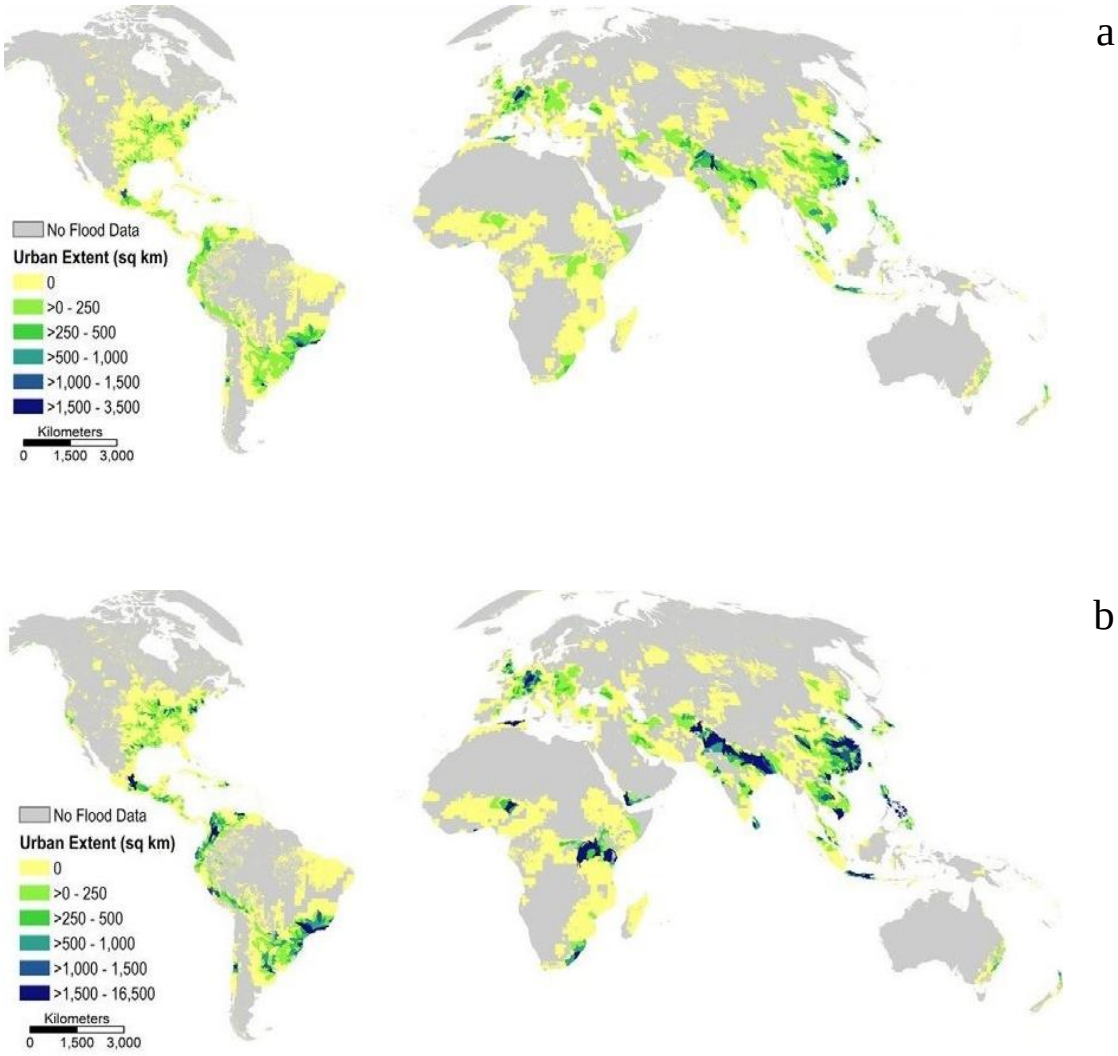


Figure 2.6 Urban land within the high-frequency flood zones by watershed in a) 2000 and b) 2030.

2.3.2 Urban expansion in drought-prone areas

Although semiarid, arid, or hyperarid zones, collectively drylands, cover about 40% of the global land surface, only slightly over one-fourth of all urban land was to be found in these zones in 2000. As in the case of coastal urbanization, urban land cover tended to be concentrated in certain locations within the drylands due to the presence of

large urban centers. In particular, the dryland watersheds with the largest urban extents were in the Arabian Peninsula, Indus valley, and in Central Asia (Figure 2.7a) with, respectively, Riyadh, Lahore and Faisalabad, and several cities in Uzbekistan accounting for most of the urban land cover.

By 2030, a third of forecasted global urban expansion is expected to be in drylands (Table 2.1). In many regions, the percent of urban land within drylands is expected to decrease; globally, however, the percentage of urban land in drylands remain the same. Importantly, the amount of urban land within these zones in certain parts of the world will increase substantially. In particular, most urban expansion is expected in semi-arid regions of China and South Asia (Figure 2.5c; Figure 2.7b). While some of the growth will be around the existing urban centers in the Arabian Peninsula, in the Indus River valley, and in Central Asia, new concentrations of urban land will also emerge in locations that had relatively small urban extents in 2000. For example, the semiarid Kano region of Nigeria is forecasted to experience substantial urban expansion as is China to the north of the country. However, the most urban expansion in hyper-arid regions is expected in Northern Africa, a region with burgeoning populations but also one of the most vulnerable to water scarcity. In fact, we expect that the urban expansion on the desert-like landscapes of the region will increase six times in 30 years (Figure 2.5d). Importantly, urban expansion in watersheds along the Nile river accounts for most of the urban expansion in Northern Africa (Figure 2.7).

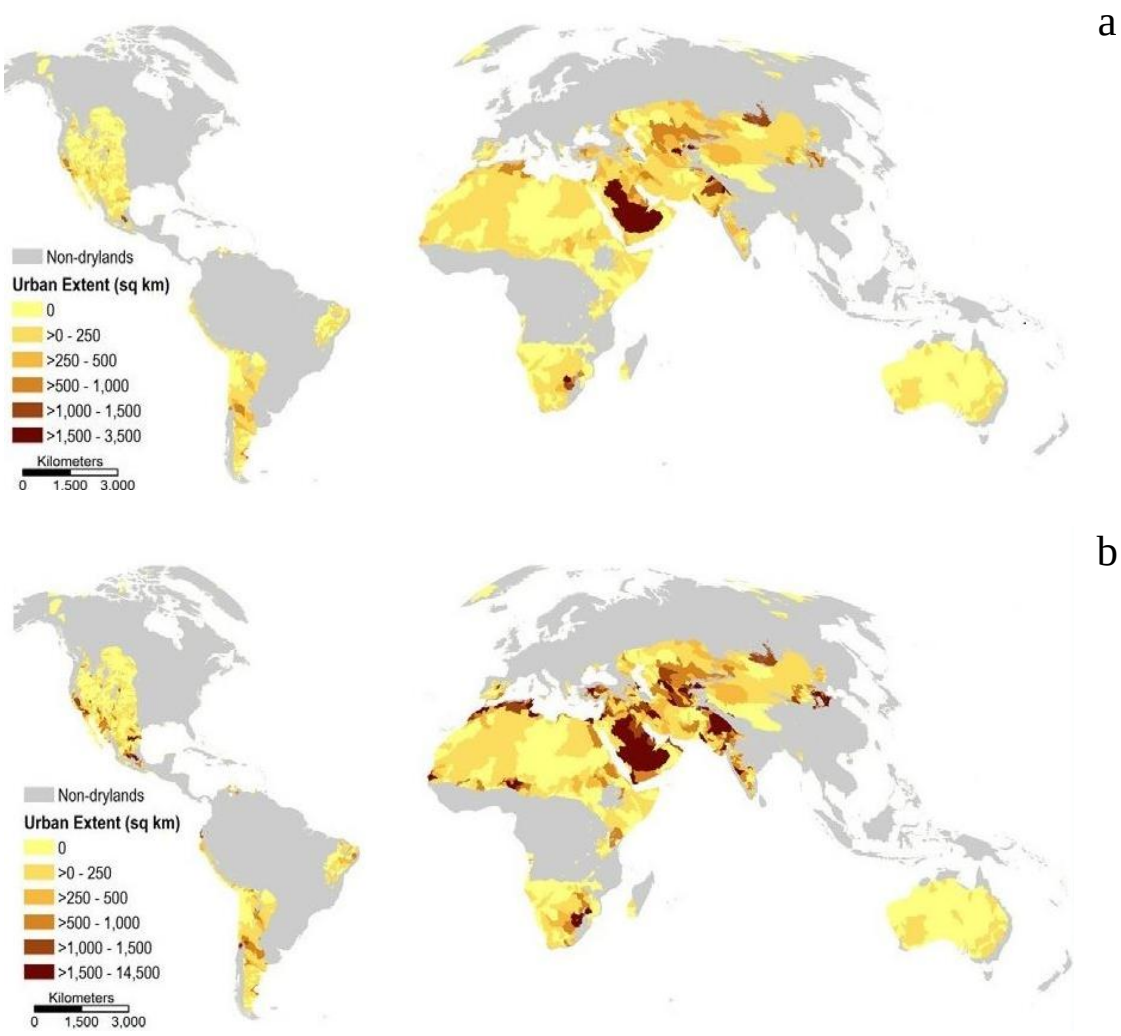


Figure 2.7 Urban land within the arid lands by watershed in 2000 and 2030.

2.3.3 Combined exposure to multiple water-related hazards

Certain places can face hazards due to both flooding and drought. In particular, Southern Asia (including India) and South America used to have the most urban land in 2000 in areas exposed to both frequent floods and to recurrent droughts (Table 2.2). By 2030, midlatitudinal Africa is expected to join these two regions in having the largest

urban extents exposed to both floods and drought. However, the only urban areas on the earth that are both located in a hyperarid climate and exposed to high-frequency floods will still be in South America in 2030 as was in 2000. These urban areas, which is Lima metropolitan area in Peru as well as other smaller cities to the north and south along the coast, are forecasted to increase in size twice to 1500km^2 by 2030.

Table 2.2 The extent and percentage of urban land exposed to multiple water-related hazards by geographical region in 2000 and 2030.

	Urban in both High-frequency flood zone and Drylands (km^2)	
Year / Region	**2000**	**2030**
CAM	2,500	7,750
CAS	2,100	6,500
CHN	1,575	4,100
EAS	0	0
EEA	450	450
IND	6,375	18,425
MLA	1,500	15,800
NAF	2,275	5,625
NAM	4,975	9,100
OCE	250	550
SAF	50	450
SAM	6,050	14,500
SAS	8,275	47,025
SEA	0	0
WAS	1,500	7,850
WSE	125	475
WORLD	38,000	138,600

Another region where we forecast substantial increase in urban extents exposed to multiple water-related hazards is Western Asia. For example, the Tigris-Euphrates delta has an arid climate, but large parts of the delta are also prone to frequent floods. The amount of urban land facing high flood risk while located in arid climates of Southern Asia will increase over five times to 9,500 km^2 between 2000 and 2030. Most of this expansion will be in Indus valley in Pakistan.

2.4 Discussion

Our study is the first global assessment of the urban expansion expected to take place in places that are the most prone to water-related hazards. It complements similar studies focused on future urban exposure to various water-related hazards based on demographic changes (McDonald, Douglas, et al., 2011; McGranahan et al., 2007) or economic development (Hallegatte et al., 2013; Hanson et al., 2011). While broadly in agreement with the findings of those studies in terms of some of the hotspots of urban such as Shanghai (de Sherbinin et al., 2007) or Mekong (Storch & Downes, 2011), our study highlights the expected change in the exposure of urban areas to water-related hazards from a land-change perspective. Our focus on the expansion of urban footprint, which is left out in previous analyses, is important because urban exposure and vulnerability are defined by the layout of the cities among other factors and second, urbanization can alter hydrological dynamics through land change.

The distribution of urban land in 2000 and 2030 among the world's watersheds highlights those watersheds that are likely to experience further negative hydrologic and

ecological effects due to urban expansion (Figure 2.3). In general, geographical clustering of several watersheds with large proportions of urban cover coincides with the large urban agglomerations. Beyond that, however, our watershed-based approach also uncovers more nuanced geographical patterns of those watersheds that are expected to be the most urbanized by 2030. For example, while forecasted urban expansion in South America is relatively less than those forecasted for Africa and Asia (Table 2.2), most watersheds in the former are expected to be urbanized at comparative levels to those in the latter. Likewise, while more urban expansion is forecasted in the northern part of India, from a watershed perspective, the intensity of urban expansion is more evenly spread across the country and the intensity –in terms of percent of watershed that becomes urban– along the southwest coast of the country part is similar to that in its north.

In many places around the world, urbanization seems to be the major dynamics behind increasing exposure and vulnerability of people and assets to various natural hazards (Dawson et al., 2011; de Moel et al., 2011; Storch & Downes, 2011). Although the most dire effects of climate change are unlikely to materialize within the timeframe of our study, the socio-economic processes driving urban expansion will be enough by themselves to increase the exposure and vulnerability of urban areas to existing flood and drought hazards.

Availability of water for human consumption, mediated by several factors, has been declining (Gleick & Palaniappan, 2010; Srinivasan et al., 2012). In addition to those in developed regions such as North America and Australia (McDonald, Douglas, et al., 2011), cities have also been growing across drylands in less developed and even politically

unstable parts of the world such as the Middle East and North Africa (Portnov & Safriel, 2004; Stewart, 2002). Continuing urbanization in arid and hyper-arid regions will lead to increased pressure on already scarce water resources. The consequences of this will depend on how the societies and governments respond to these changes. In countries located in arid regions, the national governments can implement ambitious plans to bring water to their growing cities if they have access to sufficient financial resources (Kuwairi, 2006). In general, meeting rapidly rising demand for water in fast-growing cities is a challenge wherever the governments have limited institutional or financial resources for building infrastructure and other means of delivering water to residents (McDonald, Douglas, et al., 2011). In many cities across Southeast Asia, a significant number of the residents lack access to potable water (Daniere et al., 2005) even though the region is endowed with abundant freshwater. There may be geopolitical implications as well. For example, the arid and drought-prone Euphrates-Tigris watershed is forecasted to experience even more water-stress with geopolitical implications as upstream countries retain more of the dwindling freshwater leaving less for those downstream (Bozkurt & Sen, 2013). Urbanization is progressing in these water-deprived regions of the world thanks to the financial and technological prowess that make such urbanization possible. Predicted impacts of climate change will likely elevate the water insecurity in these urbanizing drylands, but the impacts will not be felt at the same frequency or intensity everywhere around the world (Gerten et al., 2013; Schewe et al., 2014; Woodruff et al., 2013).

We forecast that the percentage of urban lands in arid lands will decrease suggesting a trend away from arid lands. This is in line with findings based purely on historical time series population distribution grids (Sherbinin et al., 2012). Nevertheless, the absolute size of urban lands in such drought-prone areas will increase, in many regions more than twice (Table 2.2). Even in Southern Asia where we forecast the largest proportional decreases in urban land in the arid lands, the amount of urban land is still expected to increase dramatically. In general, the forecasted urban expansion in arid zones tends to be around major urban centers located near large freshwater sources such as Multan and Lahore in Pakistan. Nevertheless, both cities are located in the water-scarce parts of the Indus basin (Babel & Wahid, 2008), and increasing water demand in these urban areas will put even more strain on scarce water resources in the region. Likewise, although urban expansion along the Nile river accounts for most of the urban expansion in Northern Africa that is no reprieve from water scarcity since, in Egypt, water demand already exceeds the supply sourced mainly from the Nile (Elsaeed, 2012).

In this study, we focus on direct land change due to urbanization. However, the effects of urban on water flow and water quality is not simply a matter of increasing impervious surface cover. Direct or indirect effects of urbanization via land use and land cover change, bank protection, channelization, and other means all contribute to changes in the flow patterns through their alteration of river migration patterns, widening, deepening, scouring of river beds, straightening, and instability. Geomorphic influences such as these can have serious consequences in terms of increased peak discharges that lead to increased risk of flooding (Bilskie et al., 2014; James P. M. Syvitski et al., 2009).

In addition, they will change in sediment rates (e.g., starvation or excess flux toward coast) resulting in more problems downstream (even upstream). For example, development across the Pearl River Basin but especially on and near its coastal portion resulted in significant changes in the hydrology of the delta and increased the exposure of the cities across the delta to fluvial and coastal flooding (F. Liu et al., 2014).

Creating reliable spatially explicit forecasts of changes in both urban land and urban population at regional as well as global scales is a significant challenge. A major constituent of this challenge is accurately predicting population movements to be triggered by –often interacting– environmental, socioeconomic, and political factors (Black et al., 2011). There are promising advances in both understanding the likely mechanisms behind future mobility of populations in the future (Foresight: Migration and Global Environmental Change, 2011) and modeling the spatio-temporal patterns of these changes (Jones & O'Neill, 2013).

A related challenge in research on global environmental change is the incorporation of feedbacks among the socio-economic and biophysical systems (Güneralp & Seto, 2008). Over the past couple of decades, there have been numerous studies on the predicting hydrological changes due to climatic changes and the likely impacts on populations (Döll et al., 2003; Van Vliet et al., 2013; Wisser et al., 2010) including more recent developments in flood inundation modeling at continental to global scales (Schumann et al., 2013; Ward et al., 2013; Yamazaki et al., 2011). In the context of water-related hazards there is a need for more comprehensive, systems-oriented studies that account for humans' impact on water and climate (Maestre et al., 2012; Thompson et al.,

2013). The influence of human beings in the form of water abstraction, land change, and large infrastructure building on the regional and global hydrology are largely absent in the current studies even though their individual or combined impacts may amount to significant alteration in hydrological dynamics. Incorporation of these interactions would lead to a much more informed understanding of the exposure but also the vulnerability of urban areas and how they will change in the future.

2.5 Conclusion

Our findings shed light on the geographical distribution of the most exposed urban locations to various water-related hazards at continental and global scales around the year 2000. They also suggest how this geographical distribution is likely to shift to less-developed parts of the world over the first three decades of the 21st century. Our watershed-based approach allows for assessing the geographical spread of the land area that urban cover influences/interact with from a hydrology perspective. Depending on the geographical distribution of future urban expansion on the coast, along flood plains, or in arid areas and unless proper adaptive strategies are adopted, the exposure of the urban inhabitants and urban infrastructure to such water-related hazards will continue to increase.

CHAPTER III

PRECIPITATION, GEOMORPHOLOGY, AND BASIN-WIDE DISTRIBUTION OF URBAN LAND AND POPULATION: EVIDENCE FROM ASIA AND AFRICA

3.1 Introduction

Global urbanization, which is the most significant trend in the 21st century, manifests itself in rapid rates of population growth and urban land expansion (McDonald, 2008; Turner et al., 2004; Wang et al., 2012). At the beginning of this millennium, 2.8 billion people or 47% of the global population lived in cities (United Nations, 2006). By the year 2030, about 1.7 billion additional urban residents are expected to raise the global urban population ratio to 60% (United Nations, 2014). Meanwhile, the global urban land is forecasted to increase by approximately 1.2 million km^2 (Güneralp & Seto, 2013; Seto et al., 2012).

Notably, most of the demographic transition and land use alteration is happening in developing countries. However, due to the lack of infrastructure and financial support, such developing countries have limited capacity to cope with the rapid changes. This situation will inevitably increase the vulnerability of urbanites to natural hazards (Srinivasan et al., 2013; Vörösmarty et al., 2000) and exacerbate water resource scarcity (Harhay, 2011; McDonald, Douglas, et al., 2011).

From a natural environmental perspective, the intensive anthropogenic processes pose great strains on the natural environment, including inferences on hydrologic cycles (Foley et al., 2005), biodiversity (Czech, 2004; Rebelo et al., 2011), biogeochemical

dynamics (Flintrop et al., 1996), and size and arrangements of natural habitats (Grimm et al., 2000). In this context, sustainable urban development should be pursuit, aiming at reducing the vulnerability of human beings and conservating natural environment. Therefore, building development strategies and providing policy-relevant guidance require a detailed and accurate evidence on the spatial distribution of existing urban lands and population.

Historically, civilizations flourished around rivers, and then first cities grew out of natural advantages in raw material supply or transport provided by the beneficial hydro-geomorphologic attributes. Scholars believe that the first real cities emerged in southern Mesopotamia (present-day Iraq), Egypt, the Indus Valley, and northern China were all along rivers. These places offered adequate fresh water and fertile soils, which made cropping and the surplus of food available (Fernández-Armesto, 2001; Wood, 1992). When people once lived on farming had a surplus to trade, new cities were developed along rivers because the shipment of goods by boat was a dominant form of transportation until railways were developed (Hylton, 2007). Menze and Ur (2012) mapped human settlements accumulated over eight millennia in Northern Mesopotamia and indicated the settlement's size increases exponentially as the precipitation and the Strahler order of river segment increases.

In addition to river valleys, early civilizations also occurred in the deltas of large rivers. In fact, the expansion of low-lying Holocene deltas in semiarid areas, such as deltas of the Nile River and Mesopotamia, is considered to trigger the appearance of urban centers, because the seasonally flooded plains provide adequate water resources and

nutrients for agriculture (Rusiecki & Warne, 1993). In contrast, seasonal flooding in tropical Asia deltas led to inaccessibility, which reduced the attraction for humankind (Büdel, 1966). However, well-preserved archeological evidence was reported showing that deltas of the Yangtze River, the Mekong River, and the Indus River were occupied by human dated to around 6000 BP. Recent studies also show that human population is concentrated along or near coasts (J. P. M. Syvitski et al., 2009; Woodroffe et al., 2006). According to Hinrichsen (1999), in the year 1998, over 50% of the global population — about 3.2 billion people — lived and worked within 200 kilometers of a coast. These conclusions demonstrate that hydro-geomorphologic attributes are crucial in constraining urban settlement locations and population distributions: Natural advantages helped determine the location of early cities, and the later development is commonly observed around these initial locations.

Since water is vital for both human and natural communities, and it is at the core of sustainable development, the hydro-geomorphologic attributes related to water resources are more fundamental than other factors. Unfortunately, in most developing countries water resource management receives a relatively low political priority compared with other development sectors, especially health and education (United Nations, 2013). In this study, we hypothesize that, at a river basin scale, limitations on surface water availability and local topography will affect urban development, expressing in the form of urban land and population distribution. We choose Asia and Africa as the study area because these two continents are expected to undertake the fastest urbanization processes in the following decades.

Basins of 11 major rivers in Asia and Africa were selected specifically (Figure 3.1), including the Yangtze River Basin, the Pearl River Basin, the Red River Basin, the Mekong River Basin, the Chao Phraya Basin, the Irrawaddy Basin, the Ganges–Brahmaputra Basin, the Indus River Basin, the Tigris–Euphrates Basin, the Nile River Basin, and the Niger River Basin. These 11 river basins constitute a diverse analysis sample group, in terms of their associated climates (including tropical, arid, and temperate), drainage areas, and economic development conditions.

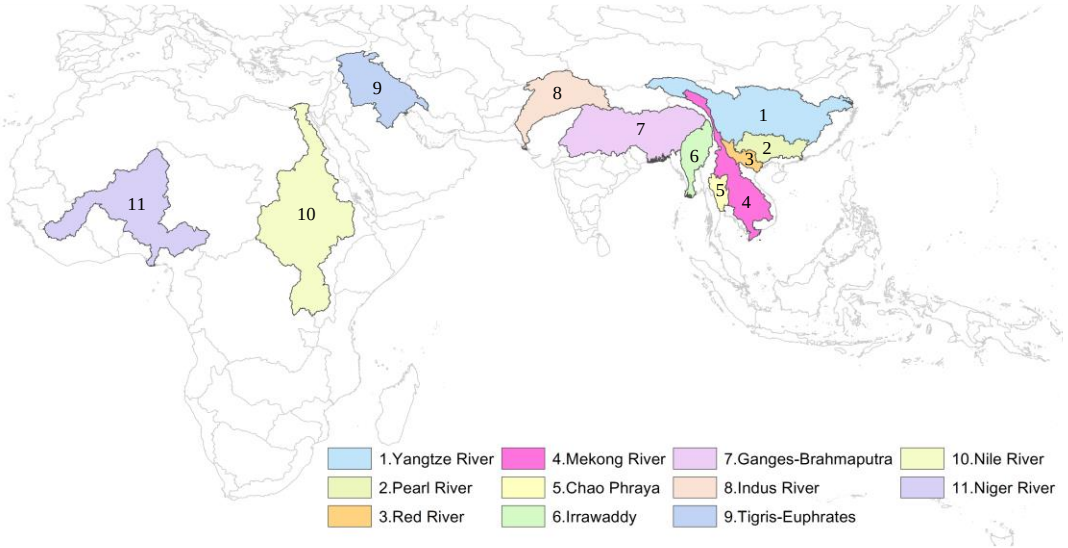


Figure 3.1 Map of the 11 major river basins in Asia and Africa as the study areas.

Then we conduct spatial analysis with a variety of metrics, to investigate how hydro-geomorphologic attributes influence the existing urban land distribution condition. The analysis method focuses on three factors: average annual precipitation, surface water availability, and access to the coast. We innovatively quantify the latter two factors as

kernel density of stream network and effective distance to the nearest shoreline, respectively. The specific research questions we focused on were:

1. What is the relationship between hydro-geomorphologic attributes and the distributions of current urban land and population?
2. Is there a common pattern of urban settlement and population distributions among the 11 fast urbanizing river basins?
3. What are the challenges brought by the distribution patterns within these developing regions on the sustainable development and what are the potential solutions?

3.2 Materials and methods

Recent advances in geographic information system (GIS) techniques have facilitated research on spatial analysis. In the present study, analysis of urban land and population distribution was implemented mainly through ArcGIS software. We employed five datasets, which can be formed into two groups. The first group of datasets was applied to delineate or calculate the hydro-geomorphologic attributes of the 11 river basins. We chose a digital elevation model (DEM) provided by "EarthEnv-DEM90" (Robinson et al., 2014) to serve as the raw terrain data. EarthEnv-DEM90 compiled ASTER GDEM2 (Gesch et al., 2012) and CGIAR-CSI SRTM v4.1 (Jarvis et al., 2008) by means of seamlessly merging, filling voids, and reducing terrain noise of the input datasets; and yielded a new dataset with higher quality at a three arc-second resolution. The later process including watershed delineation, land surface's slope determination, and calculation of

stream network's kernel density as well as the effective distance to the coast, were all dependent on the EarthEnv-DEM90 basemap.

A precipitation dataset developed by the Climate Research Unit (CRU) was used to represent the regional hydrological attribute. The CRU CL 2.0 dataset (New et al., 2002) consists average monthly precipitation over global land areas except for Antarctica, at a five arc-minute resolution (i.e. about 9.25 km in distance near the equator). This raster dataset is constructed based on the observed precipitation quantity from 1961 to 1990. Then, average annual precipitation was calculated for each grid as a factor in urban development.

The access to coast from a random point in a river basin is another natural factor that influencing local people's activity and urban development. In this research, we quantified the access to coast factor as the least cost distance from one point to the shoreline, and a World Vector Shorelines (WVS) from the Global Self-consistent, Hierarchical, High-resolution Geography Database (GSHHG) was employed. Five levels of generalization are available in this shoreline dataset (Wessel & Smith, 1996): full, high, intermediate, low and crude. The "full" version has the original resolution at a 1:250,000 scale, which is the highest in the series. The successive level has an 80% reduction in size and quality comparing to the former one. Considering the accuracy, study scale, and computing time, we used the full resolution data.

Information on water availability, as another key environmental factor, is often lacking or unreliable in the 11 developing regions. Therefore, stream network's kernel density weighted by Strahler order serves as a proxy of the surface freshwater availability

in this analysis. We assume that the region with higher kernel density of stream network has higher water quantity. The other advantage of this quantifying method is, the effect of the whole river system rather than that of an individual river segment is taken into consideration.

The second group of data sets illustrates urban footprints and the corresponding populations. The urban extent map comes from a widely used dataset created from Collection 5 Moderate Resolution Imaging Spectroradiometer (MODIS) observation (Schneider et al., 2009). This dataset depicts global urban land around the year 2001-2002 at a 500-m resolution and represents a high quality of spectral measurements. Compared to earlier global urban maps (Arino et al., 2007; Bartholomé & Belward, 2005; Danko, 1992), this dataset shows advances in accuracy and consistency of urban's definition and scale. It characterizes urban regions as places that are dominated by the built environment (Schneider et al., 2010): if non-vegetative, human-constructed landscape covers more than 50% of a pixel's area, the pixel is classified as urban.

The population dataset Global Rural-Urban Mapping Project (GRUMP) is provided by Center for International Earth Science Information Network (CIESIN). This raster data represents population density circa 2000 at 30 arc-second resolution (about 1 km near the equator). The population density is in a unit of persons per square kilometers and is adjusted to match UN totals.

During the data processing, the precipitation data layer was resampled to 10 km resolution due to the relatively coarse resolution; while application of other geographic layers was conducted at a resolution of 500 ×500 m^2, same as the original urban map. The

Goode Homolosine projection was used because the 11 river basins are spread, and the projection's equal-area property is appropriate for spatial distribution.

3.2.1 Watershed delineation

We created a new dataset of the 11 river basins' watersheds and stream networks instead of using the two most popular existing datasets, HydroSHEDS (Lehner & Grill, 2013) and Hydro1k (USGS, 2000) because neither of them meets the present research's needs. Hydro1k is not accurate in the relatively flat regions as coastal zones; while HydroSHEDS does not show the specific drainage area (i.e. sub-basin) for every stream segment. Moreover, these two datasets do not include information about stream's Strahler order.

The watershed delineation processes were implemented by the ArcGIS software (Figure 3.2). First, small imperfections in the DEM were removed by filling sinks. Next, the flow direction was computed using eight-direction (D8) method (O'Callaghan & Mark, 1984), and the flow accumulation was determined. Then, we needed to specify a threshold constant of critical support area, which defines the minimum accumulation area for a permanent stream feature and identifies the sources (i.e. beginning point) of a stream network.

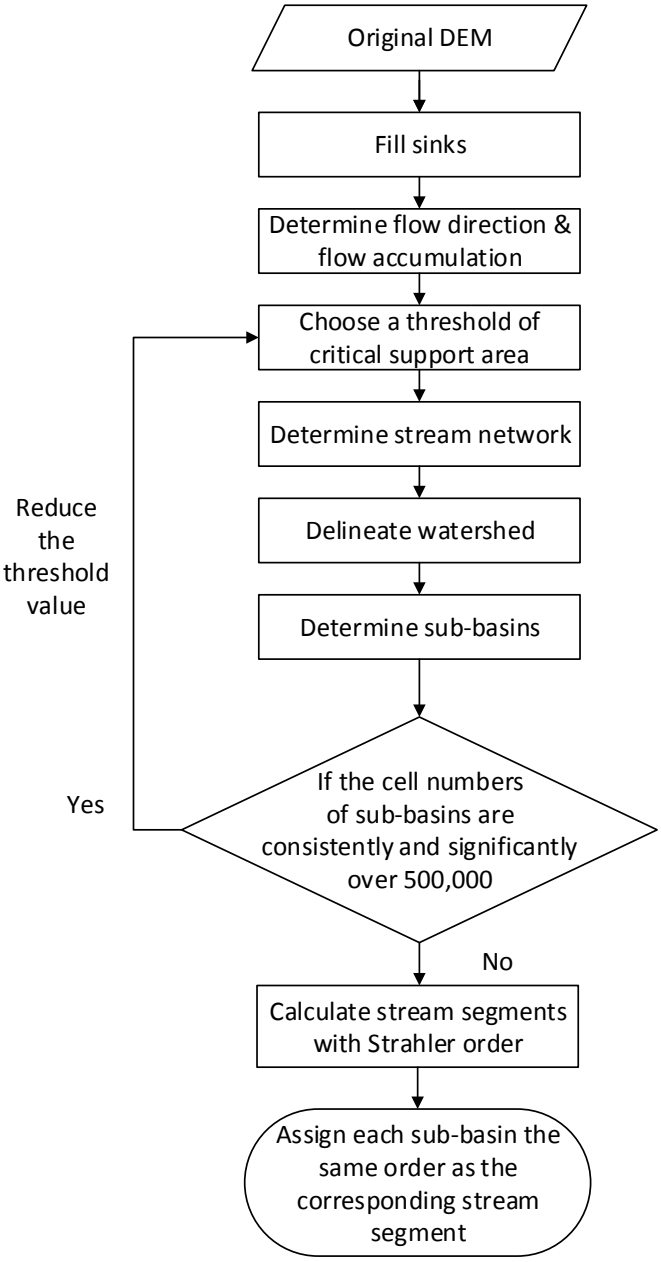


Figure 3.2 Watershed delineation processes

According to the Fast Watershed Delineation (FWD) method (Djokic et al., 1997), two criteria were employed to choose a proper threshold of critical support area: 1) a threshold value should no larger than 1/500 the number of cells for a basin. 2) the pixel

number of each resulted sub-basin in a watershed should less than 500,000. Taking computational time and model accuracy into consideration, we selected a uniform threshold constant as 300 km^2 for all the 11 river basins. This method can avoid scale issue and enable the later comparisons between different river basins' results. Last, each stream segment and the corresponding sub-basin were assigned a Strahler order according to Horton's law.

3.2.2 Strahler order and Horton's law

Strahler order is developed by Horton (1945) and Strahler (1952, 1957) to define the stream size based on a hierarchy of tributaries. It is a complex, multidimensional characteristic that includes stream dimensions, discharge, stream length, number of tributaries, and contributing drainage area. One can assign Strahler order to all segments of a stream network according to following rules: 1) All links without any tributaries (i.e. stream segments originate at sources) are assigned as the first order. 2) For a node joins two tributaries with order i and j respectively, the subsequent downstream river segment's order k could be:

$$k = \begin{cases} i+1 & (i = j) \\ \max(i, j) & (i \neq j) \end{cases} \tag{3-1}$$

Table 3.1 Parameters of Horton's law showing geometric relationships for a stream network.

Parameter	Symbol	Definition
Bifurcation ratio	R_b	N_{w-1}/N_w , where *N* is the number of streams of order *w*
Length ratio	R_l	L_w/L_{w-1} , where *L* is the mean stream length of order *w*
Area ratio	R_a	A_w/A_{w-1} , where *A* is the mean total contributing area of order *w*

The geomorphology of a stream network is always characterized regarding Horton's law. The definitions of the three related parameters shown in Table 3.1can be transformed into the following equations:

$$\log_{10} N_w = b_{0b} - \log_{10} R_b \cdot w \tag{3-2}$$

$$\log_{10} L_w = b_{0l} + \log_{10} R_l \cdot w \tag{3-3}$$

$$\log_{10} A_w = b_{0a} + \log_{10} R_a \cdot w \tag{3-4}$$

where b_0 is the vertical axis intercept. Obviously, semilogarithmic plots of stream segment numbers, mean stream lengths, and mean contributing drainage areas versus Strahler order should be nearly linear. The slopes of these three linear equations are the common logarithm of average R_b, R_l and R_a, respectively. Therefore, we applied least squares regression to approximate three parameters of Horton's law. Figure 3.3 displays an example of resulted stream network with Strahler orders and the three parameters related to Horton's law for the Nile River Basin.

Watershed delineation of the Nile River and (inset) the semilogarithmic plots of the number of streams (blue circles), mean stream length (red square), and mean contributing drainage area (green triangles) related to Horton's law. The bifurcation ratio can be obtained by the slope of the fitted line in blue through $R_b = 10^{-slope}$. The length ratio (R_l) and the drainage area ratio (R_a) can be achieved by a similar mathematical form but with positive exponents.

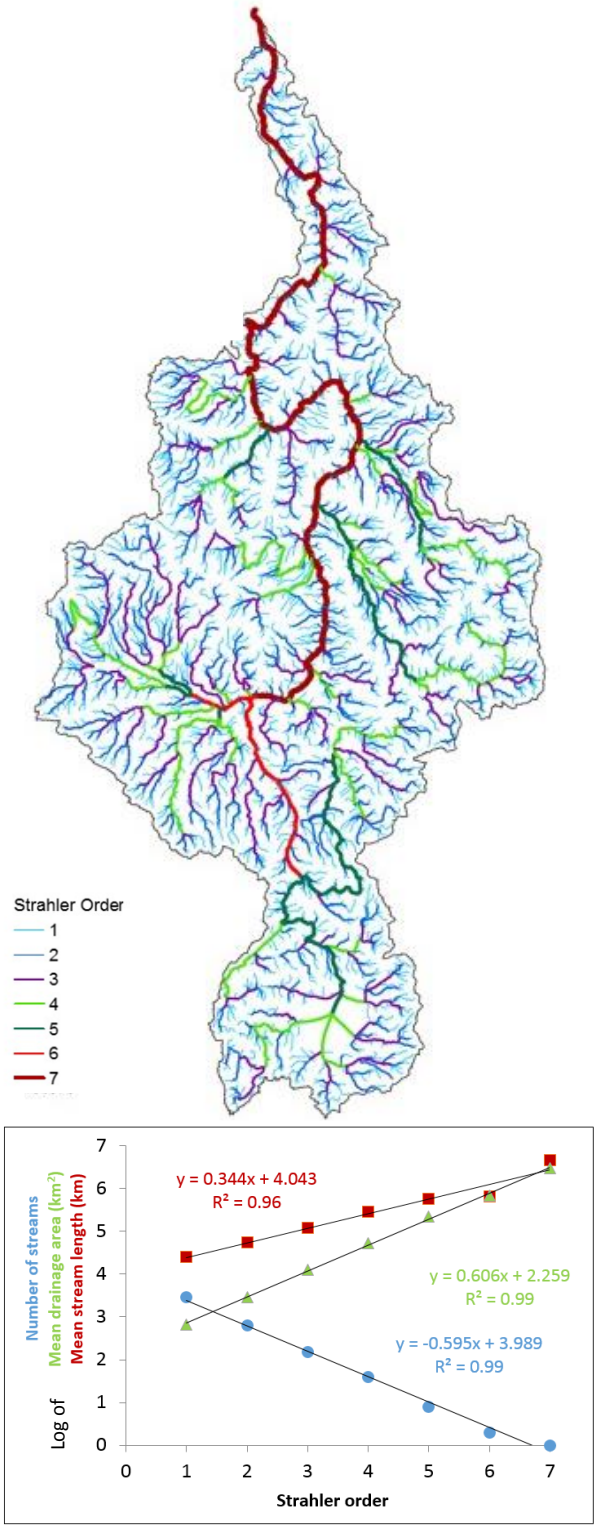


Figure 3.3 Stream network of the Nile River with Strahler orders and the three Horton's law parameters

3.2.3 Kernel density of stream network (KDSN)

Kernel density estimation (KDE) is a non-parametric statistical method, which transforms a sample of geographical points or lines into a smoothed continuous surface. Although KDE is seldom used in hydrological and geomorphological research, the technique is extensively applied within the domain of wildlife ecology (Seaman & Powell, 1996; Steiniger & Hunter, 2013) and urban analysis (Kloog et al., 2009). The basic bivariate point-based kernel density function is described as follows:

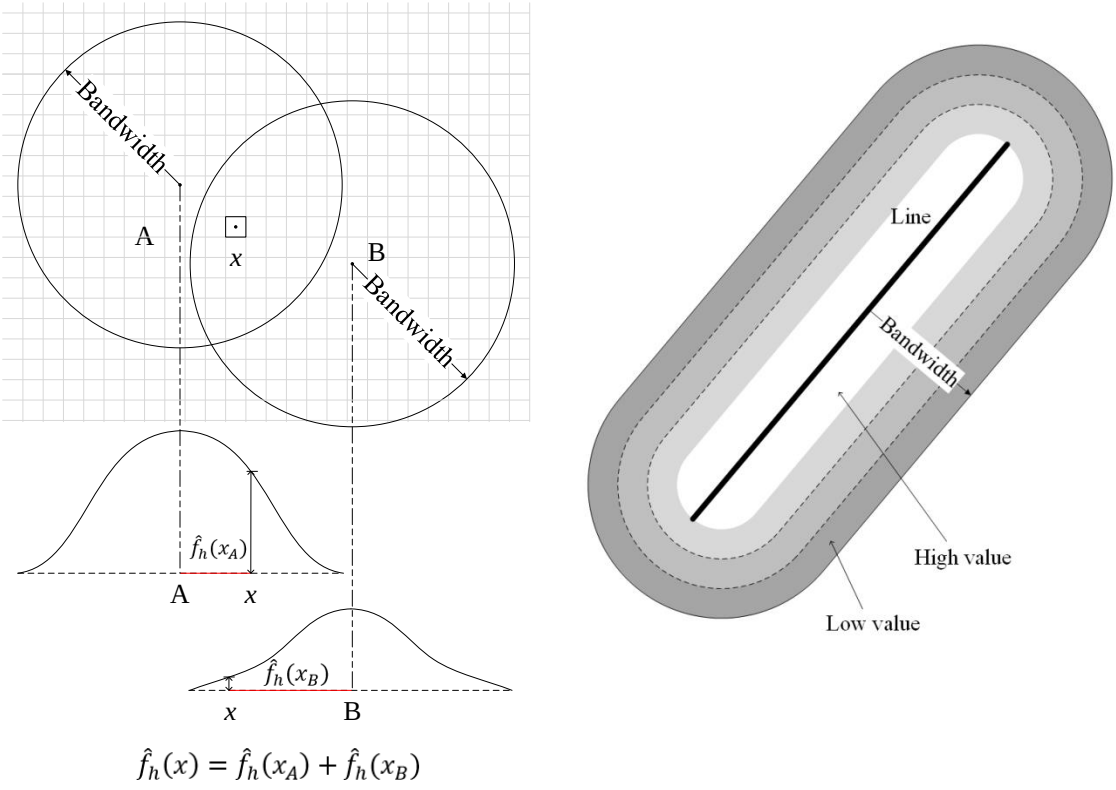
$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{3-5}$$

where *n* is the number of observations in an independent and identically distributed sample $(X_1, X_2, X_3 \dots, X_n)$, h is the bandwidth, $K(u)$ is the kernel function, $x - X_i$ is the distance between an estimated point x to sample point X_i. The kernel function $K(u)$ is required to satisfy two conditions:

$$K(u) \geq 0 \quad and \quad \int_R K(u)du = 1 \tag{3-6}$$

The essence of KDE calculation is local averaging (Figure 3.4a). The KDE result for a sample point (point A or point B) is a smoothly curved circular surface. The yielded circular neighborhood has an origin at the sample point and a radius same as the pre-defined bandwidth. The volume of the surface is determined by the sample point's attribute value or equals one if the sample point only represents a location without a numerical attribute. A neighboring sample point closer to the estimated point x weights more and assigns higher density at the point. Through ArcGIS, the KDE result is always

expressed in a form of a raster layer. The final density value at x is calculated by summing up all the kernel surfaces values that overlaying the grid cell. In summary, the KDE result illustrates the intensity of individual observation over space.



a b

Figure 3.4 Kernel density estimation based on a) point and b) line

The line-based KDE is an extension of the point-based KDE (Figure 3.4b), and the result changes correspondingly into a smoothly curved density surface fitted over a line. The density value is at its peak on the line and diminishes with increasing distance from the line. When the distance excesses the pre-defined bandwidth, the density value reaches

zero. The volume under the surface equals the product of line length and the weighted factor. Similar to the point-based KDE, the density value of the line-based KDE at an estimated point is the total of kernel surfaces values that overlaying the grid cell.

In this research, kernel density of stream network weighted by Strahler order was treated as a proxy for the surface freshwater availability. However, before conducting the calculation, we needed to address the bandwidth selection issue. The value of bandwidth determines the smoothing degree of KDE: a small bandwidth will yield a spiky distribution result, which is hard to interpret; a large bandwidth may over-smooth the data, and skim certain trends or structures. Although consensus on the choice of an appropriate bandwidth does not exist, there are five most well known and extensively used theoretical method: The Normal rule of thumb (Silverman, 1986), the Silverman's rule of thumb (Silverman, 1986), the least square cross-validation method (Bowman, 1984; Rudemo, 1982), the biased cross-validation method (Scott & Terrell, 1987), and the plug-in method (Nadaraya, 1974; Woodroofe, 1970).

Unfortunately, since the stream network data we created and used as the KDE input are lines, the theoretical methods mentioned above are not practical. Some scientists suggest visual inspection method, by examining several density plots derived from the same sample but using a set of different bandwidths, to gain greater insight into the data. Nevertheless, considering the number of river basins involved, this empirical method is redundant and unreliable. Hence, we designed a new method to determine the bandwidth of stream network KDE. First, a neighborhood distance band was calculated for each stream network. The result returned a list of Euclidean distances between each stream

segment's centroid and its nearest neighbor's centroid. Then the maximum value of the distances was designated as the bandwidth. This method guarantees each stream segment has, at least, one neighbor within the selected bandwidth. Therefore, every grid's kernel density value is larger than 0, and the variation of the results among all 11 the river basins is limited.

3.2.4 Effective distance to the coast

Previous studies always use simple Euclidean distance to represent the proximity to the coast. However, this method does not consider the complexity of topography. In this study, we employed a least-cost model to include the landscape and behavioral aspects. Comparing to the Euclidean distance, the least-cost distance is more realistic by involving an impedance factor, which shows the resistance per unit distance.

The least-cost modeling was conducted through the spatial analyst "cost distance" extension of ArcMap. Cost distance method is a 3×3 cell neighborhood algorithm, which allows eight potential moving directions including the ones along diagonals. The model requires two GIS layers as input – a source layer and a resistance layer. The source layer indicates the movement destination. The resistance layer determines the cost value, the geographical extension of departure points, and the moving direction. In this research, the source layer was the GSHHG shoreline vector. The resistance layer was weighted by the topographic slope of every 1×1 km2 square area in a river basin. In order to prepare the resistance layer, a raster of slope in percent was first derived from the DEM. Then the

slope values were reclassified, and each grid cell was assigned a resistance value according to the slope's class (Table 3.2)

Table 3.2 Reclassification of slope

Percent of slope	Degree range	Resistance value
0.00 - 1.00	0.00 - 0.57	1
1.00 - 5.00	0.57 - 2.86	$\sqrt{s_p^2 + 1}$
>5.00	>2.86	s_p

The outcome of cost distance is the least accumulated resistance value from a certain location to the destination, denoting the difficulty of traveling. The least-cost value has the same unit as the Euclidean distance (m). To avoid confusion, we use the term "effective distance to the coast" to represent the calculated least cost distance.

3.2.5 Urban settlement patterns and population distribution

A series of spatial analysis were performed to characterize the typical patterns of urban settlement and population distribution in the 11 river basins. First, within each study area, the percentage of a river basin's land surface that contains 80% of the population was calculated. This measure can identify river basins with highly focal population distribution. Second, we took average annual precipitation, KDSN, and effective distance to the coast to represent regional climate, surface freshwater availability, and access to the coastal zone, respectively. To seek the relationship between hydro-geomorphologic

attributes and the distributions of current urban land and population, the three factors served as key natural influences, and their combined impact was emphasized. We considered that KDSN has a higher priority than the other two factors because KDSN represents not only geomorphologic attributes but also potential water resource capacity. Meanwhile, precipitation amount and effective distance to the coast may have a linear relationship, due to topographic gradients. Therefore, only two sets of influential factor combination were formed: 1) average annual precipitation and KDSN, and 2) effective cost distance to the coast and KDSN. Every influencing factor was reorganized into five classes according to the quintile values. In this way, each class of the factor covers 1/5 of a river basin's area, and every river basin was divided unevenly into 25 patches. Then, three indices – urban ratio, population density, and urban population density – were used to reflect the urban condition. For each river basin, we express the KDSN's classes as i, and the precipitation's classes/effective cost distance's classes as *j*, where *i* and *j* vary from 1 to 5, then the three urban indices can be defined as:

$$Urban\ ratio(i,j) = \frac{Urban\ area\ (i,j)}{Total\ area\ (i,j)} \tag{3-7}$$

$$Population\ density(i,j) = \frac{Total\ population\ (i,j)}{Total\ area\ (i,j)} \tag{3-8}$$

$$Urban\ population\ density(i,j) = \frac{Urban\ population\ (i,j)}{Urban\ area\ (i,j)} \tag{3-9}$$

These three indices were normalized by the corresponding patch's area to avoid potential biases.

3.3 Results

3.3.1 Demographic and geomorphologic attributes

The 11 river basins form a group of diverse geomorphologic and demographic attributes (Table 3.3). According to our computation, the Nile River has the highest seventh Strahler order and the associated largest drainage area. However, the low population of 149.1 million people at 50 persons per square kilometers (basin-wide) makes this river basin less dense compared to the basins of the Pearl River, the Red River, and the Chao Phraya. The latter three river basins are more industrialized although they only yielded the fifth Strahler order and are one order of magnitude smaller in area than the Nile River. The sixth-Strahler-ordered river basins are intermediate in size between the seventh and the fifth. The Ganges-Brahmaputra Basin supports the largest population with 510.1 million people at 328 persons per square kilometers, which is also the densest.

Geomorphological metrics relevant to Horton's law are listed in the last three column in Table 3.3. All relationships used to compute these metrics through regression model yielded R^2 values greater than 0.93. It is worth noticing that a low bifurcation ratio can be a signal of a high flooding risk as the water will restrain in lower order channels due to the limited discharge capacity of the successive higher order streams. In this case, the Irrawaddy and the Red River have relatively higher fluvial flood risk compared to the other considering rivers.

Table 3.3 Major demographic and geomorphologic attributes of 11 river basins

River	Drainage area (10^5 km^2)		Population (millions)	Basin-wide population density (persons/km^2)	Max Strahler Order Ω	Bifurcation ratio R_b	Length ratio R_l	Drainage area ratio R_a
	FAO	New-derived						
Nile River	30.7	30.4	149.1	50	7	3.9	2.2	4.0
Niger River	21.4	21.1	71.0	34	6	4.5	2.3	5.2
Yangtze River	17.9	17.0	370.2	219	6	4.2	2.4	4.6
Ganges-Brahmaputra	16.7	15.8	510.1	328	6	4.2	2.2	4.6
Tigris-Euphrates	9.2	9.0	50.3	57	6	3.8	1.9	4.1
Indus River	8.7	8.5	151.5	184	6	3.7	2.1	4.2
Mekong River	8.0	7.6	50.1	66	6	3.8	2.0	4.2
Irrawaddy	4.0	3.7	25.2	68	6	3.2	1.9	3.6
Pearl River	4.1	4.0	72.8	183	5	4.4	2.4	5.0
Red River	1.6	1.5	29.2	188	5	3.5	1.9	3.9
Chao Phraya	1.5	1.5	18.3	119	5	3.6	1.8	3.9
Typical range (Kirchner, 1993)						(3, 5)	(1, 3)	(3, 6)

We employed a widely used dataset of global major river basins developed by the Food and Agriculture Organization (FAO, 2009) to validate the new derived river basins. We also accessed the accuracy of the delineated stream networks by comparing the three Horton's law ratios to the typical ranges. The minor differences in drainage area between two datasets are mainly attributed to the discrepant definitions of a river basin. In our method, the tide-driven delta area was not considered as a constituent part of a basin. Therefore, the areas of the newly derived river basins are lower than their counterparts in the FAO dataset. The comparison results confirmed that the delineated river basins and stream networks are theoretically correct and accurate for further analysis.

3.3.2 Kernel density of stream network (KDSN)

Using stream segments' lengths as multi-cores, the KDSN for each river basin yielded in a form of continuous raster surface, with a cell size of 1 km (Figure 3.5). Each cell's value represents the estimated cumulative density of stream segments (km/km^2) that located within the search radius (same as the predefined bandwidth). A relatively greater value of KDSN occurs along the main channel with a higher Strahler order, or at a place where stream segments are densely distributed. Therefore, warm colors such as reds and oranges on the KDSN maps indicate higher accessibility of surface freshwater.

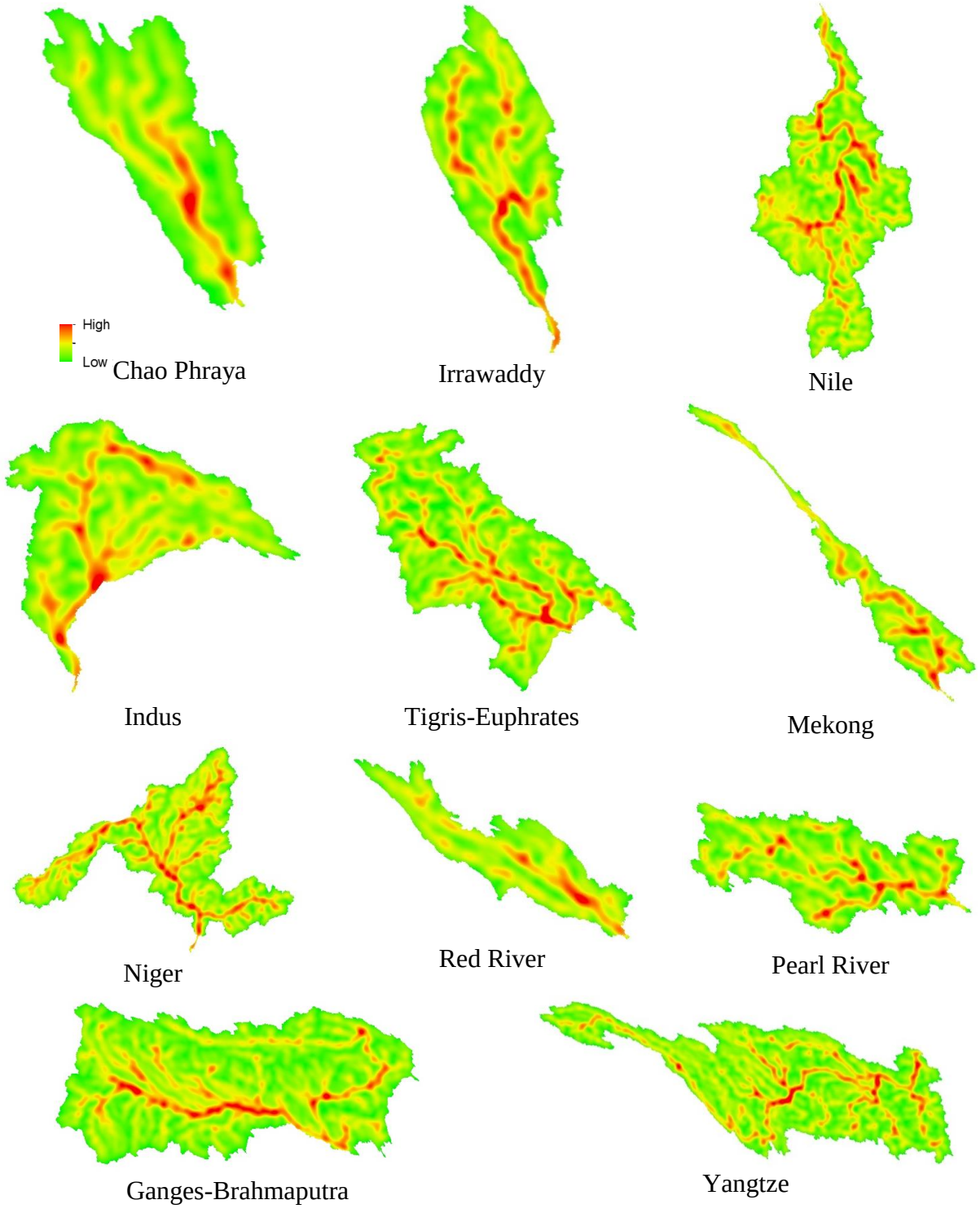


Figure 3.5 Kernel density of stream network for each river basin (the cumulative stream length per km^2).

As shown in Figure 3.5, large KDSN values occur more frequent in the middle and lower catchments of a river basin, because streams in the upper catchment, most of which are minor tributaries, are relatively dispersed and inferior in Strahler order.

The results of KDSN are sensitive to the specified bandwidths. While we only conducted an empirical method to select bandwidths instead of a mathematical approach, the chosen bandwidths still served to illustrate clusters of stream networks clearly.

3.4 Urban land and population distribution

The spatial patterns urban land and population in the 11 river basins were characterized using different spatial metrics. Figure 3.6 shows the configurations of basin-wide population distribution. Among all the 11 developing regions, the Nile River Basin has the most highly focal population distribution, with 80% of the residents living in less than 15% of the territory. The Tigris-Euphrates Basin and the Niger River Basin, the climates of where are similar to the Nile River Basin as arid to semi-arid, also show high population concentration. On the contrary, the population distributions in the Pearl River Basin and the Chao Phraya Basin are both dispersed. At the same time, these two river basins also have the higher urban area ratios, which indicate higher industrialized and levels.

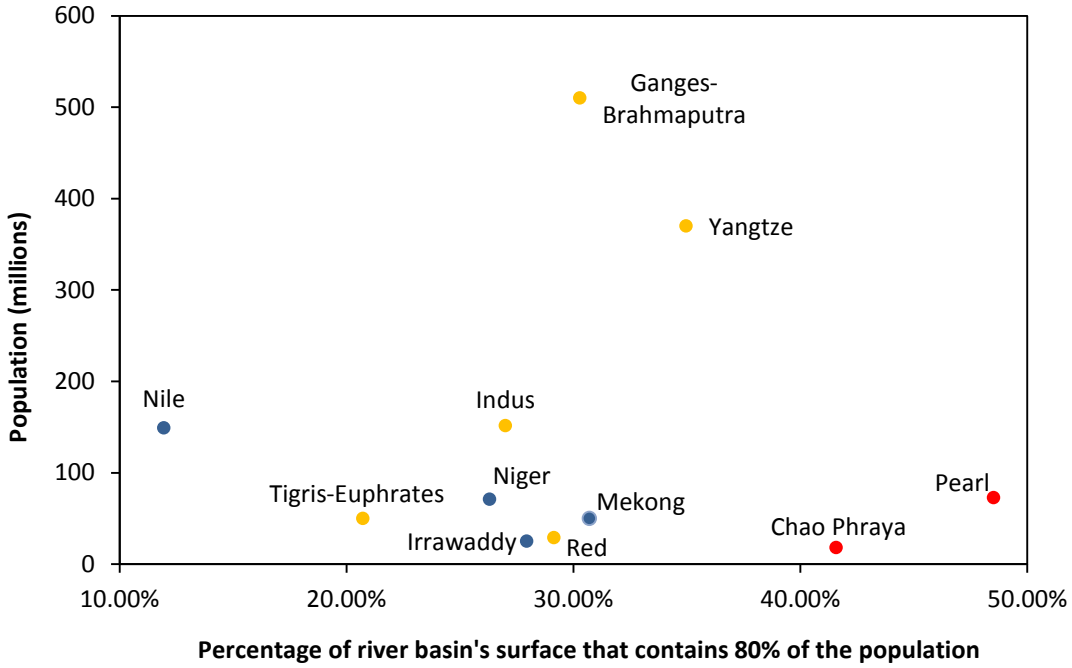


Figure 3.6 Scatterplot of the percentage of basin's surface that contains 80% of the population versus the total population within the basin.

Two sets of bivariate histograms shown in Figure 3.7 and Figure 3.8 illustrate the correlations between hydro-geomorphologic attributes (i.e. average annual precipitation, KDSN, and effective distance to the coast) and the urban indices (i.e. urban ratio, population density, and urban population density). The combination of precipitation and KDSN factors focus on the impact of water availability and accessibility; while the combined influence of effective distance and KDSN embodies the effect of geomorphologic attributes. Each bivariate histogram is a topologic transformation of the map beneath it.

For both sets of the bivariate histograms, the Chao Phraya Basin, the Pearl River Basin and the Red River Basin display patches with higher urban ratios, which indicates

higher development situations. Meanwhile, the maximum value of the urban ratio in the Mekong River Basin and the Niger River basin are one order of magnitude lower than the other basins' values. The result is also in accordance with the fact that these two river basins are less developed.

Although these geospatial analysis results show a significant degree of variation, regular trends still exist among the 11 river basins. Next, we will describe the commonalities and the exceptions present in each type of analysis in detail.

3.4.1 Distribution based on average annual precipitation and KDSN

Figure 3.7 shows joint distribution based on average annual precipitation and KDSN, the largest values of the urban ratio and population density are distinct on the right side. This pattern shows people (in both rural or urban area) prefer to live in a region close to the streams, where surface freshwater is more available and easier to access.

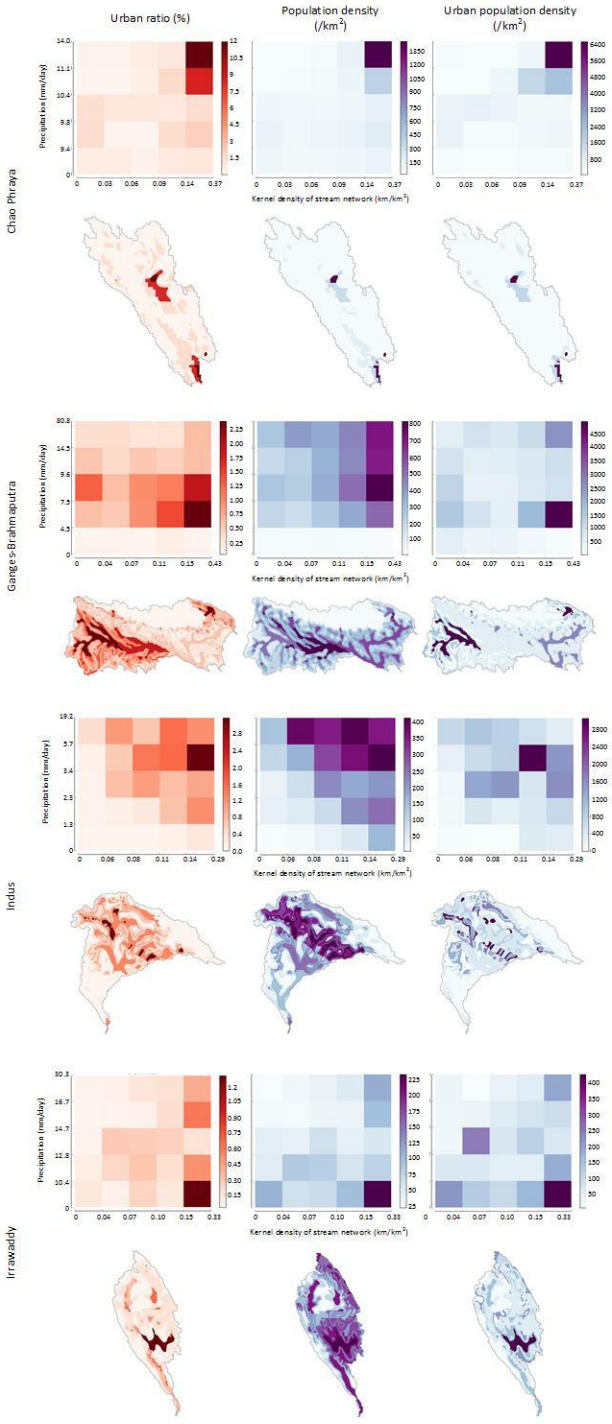


Figure 3.7 Bivariate histogram of the urban ratio, population density, and urban population density with given average annual precipitation and kernel density of stream network.Joint distribution based on average annual precipitation and KDSN

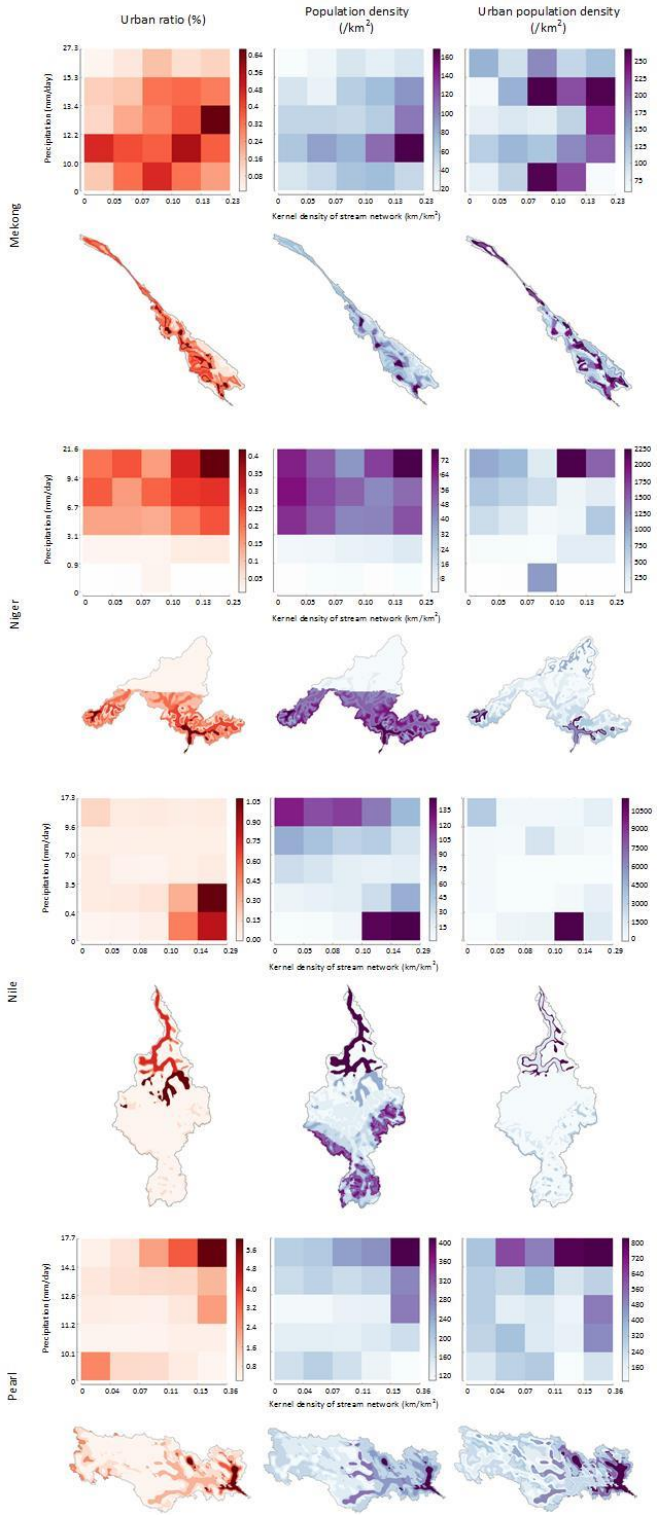


Figure 3.7 Continued

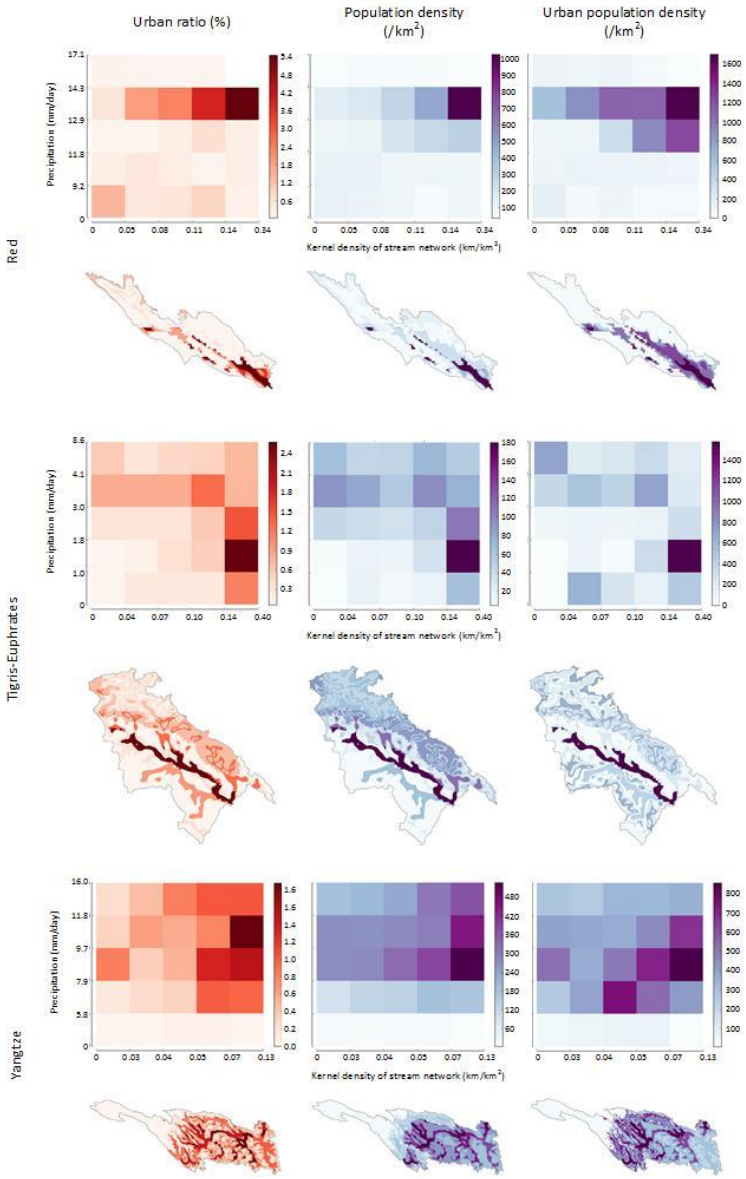


Figure 3.7 Continued

However, given a specified class of KDSN, the urban condition indices do not show obvious trends with the increase of average annual precipitation. A general increasing trend, with the highest values of urban condition indices occurring in

precipitation-5 class, characterized the Chao Phraya Basin, the Pearl River Basin, and the Niger River Basin. On the contrary, people prefer to live in the precipitation-1 class regions in the Irrawaddy Basin. The rest river basins present the darkest pixels of bivariate histograms in the intermediate precipitation categories.

The "appropriate precipitation" varies significantly among the 11 river basins, ranging from 3.4-5.7 mm/day in the Indus River Basin to 9.4-21.6 mm/day in the Niger River Basin. These pieces of evidence imply that precipitation, as an individual factor, does not influence urban land and population distribution adequately.

To explain this phenomenon, we need to put the 11 river basins into two groups. The first group contains all river basins except the Nile River Basin and the Tigris-Euphrates Basin. These nine regions are affected by tropical or subtropical monsoons. For monsoon precipitation, in general, the higher the average annual amount, the greater the range of seasonal variation throughout a year. In this case, people who dwell in regions with moderate annual precipitation may have less probability exposing to extreme precipitation events and less risk of floods and droughts.

The Nile River Basin and the Tigris-Euphrates River Basin composite the second group. These two basins have lower annual average precipitations in our 11 study areas, owing to the semi-arid to arid climate. Also, the direction of precipitation gradient in these two river basins are opposite to the gradient direction of their main channels. The Nile River flows south to north with the annual precipitation decreasing towards the north. However, more than 35% of the total population in this basin clusters in Egypt, a country located in the most north. As the Nile River across the Sahara desert after its headwater,

the upper and middle reaches are not fit for human habitats, even though the precipitation there is higher than the lower reach. After the river's passage of Cairo, the green band clinging to the Nile expands abruptly into the broad delta, where more arable land is available, and more urban land can be developed based on adequate food supply, upstream water resource capacity, and stable population. Similarly, the Tigris and the Euphrates rivers both originate in Turkey and receive their majority discharges in the mountainous, humid upper catchments. Then the twin rivers flow to the Mesopotamian plain in Iraq, where is arid, flat, and relatively more densely populated. According to UNEP (2001) and the study of Al-Ansari (2013), Turkey contributes 78.1% of the total discharge in the Tigris-Ehpurates River Basin, while Iraq only adds 8.1% of the water resource. Again, in this basin, precipitation does not impose a powerful constraining impact on urban development and population growth.

In summary, precipitation has less weight on affecting the urban distribution comparing to geomorphologic attributes. In these two river basins, Cairo and Baghdad are typical cities that almost wholly depend on external water resources. Their beneficial geomorphologic locations bring adequate water resources from the upstream and compensate the local climatic limitation.

3.4.2 Distribution based on cost distance and KDSN

Figure 3.8 exhibits that most of the 11 river basins have urban land or population clusters in the first category of cost distance; the largest values of urban indices also occur in the pixels with high KDSN values. These two regulations suggest the existing urban

lands tend to build on flat, low-lying landscapes such as in coastal zones and along navigable rivers.

It is apparent among all the 11 river basins that big cities are rare when cost distance extends 1500×10^3 km. However, exceptions exist in the Chao Phraya River Basin and the Yangtze River Basin. In the bivariate histogram of Chao Phraya's urban ratio, the dark pixel on the upper right of the square plot represents Chiang Mai, the fourth largest city in Thailand by population. This city is situated amongst the highest mountains in Thailand, as well aside the Ping River, which is a major tributary of the Chao Phraya River.

In the Pearl River Basin and the Nile River Basin, the maximum urban population density occurs in a lower KDSN class comparing to the urban ratio, while the category of cost distance remains. Simultaneously, the largest total population density stays in the same pixel location as the urban ratio. This phenomenon reflects that most of the floodplains are occupied by farmlands instead of developed regions. The buffer between the most populated urban lands and the river channels not only reduces the cost of flood inundation damage but also allows the floodplain to provide beneficial functions such as wild species' habitats and discharge zones of groundwater.

The urban land and population distribution patterns of the Mekong River Basin and the Niger River Basin are not as regular as others. Plausible explanations for these inconsistencies include, first, the urban ratios of these two river basins are one order of magnitude lower than that of other basins, which indicates these two basins are less developed and remain primarily agricultural. So the distribution of urban land is patchy.

Second, as two transboundary basins, Mekong and Niger are strongly affected by different developing rates and policies among several countries within them. Although the Mekong riparian countries enjoy abundant water resources, availability varies widely by region and by season, due to the monsoon rainfall pattern.

Apart from the reasons shared with the Mekong River Basin on irregular urban distribution patterns, the Niger River Basin is notorious for its unplanned urban development. In the Niger River Basin, the vegetation of delta includes rainforest and mangrove swamp, where neither drainage conditions nor accessibility are good. So urban lands are located in the interior delta. Communities throughout the region suffer from a weak infrastructure for the efficient and effective delivery of water supply and access to potable water. The vast majority of the settlements depend on springs and wells for their water supply instead of surface freshwater. Moreover, the establishment of Abuja as the new capital of Nigeria in 1991 affected the urban lands and population distribution within the Niger River Basin. The Niger River Delta is mainly situated in Nigeria, but Abuja, the third largest cities in the country with three million population, is neither proximate to the mainstem nor the coast.

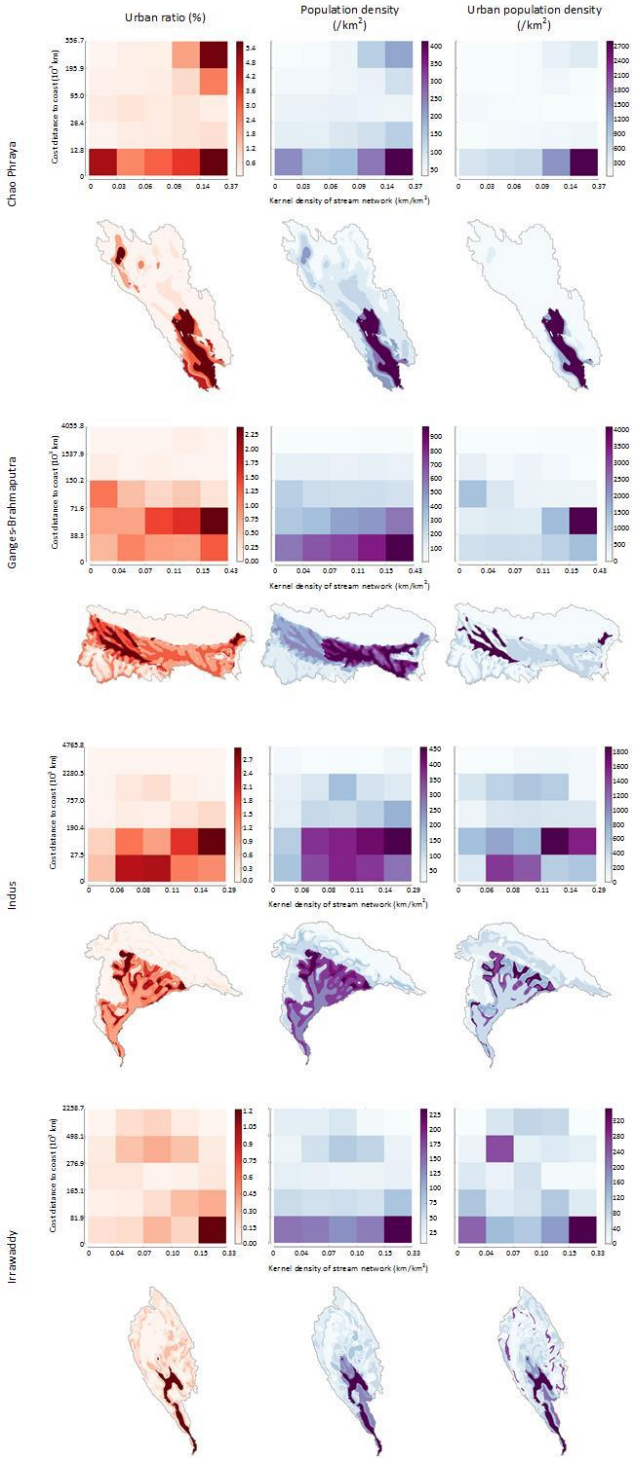


Figure 3.8 Bivariate histogram of the urban ratio, population density, and urban population density with given cost distance to coast and kernel density of stream network.

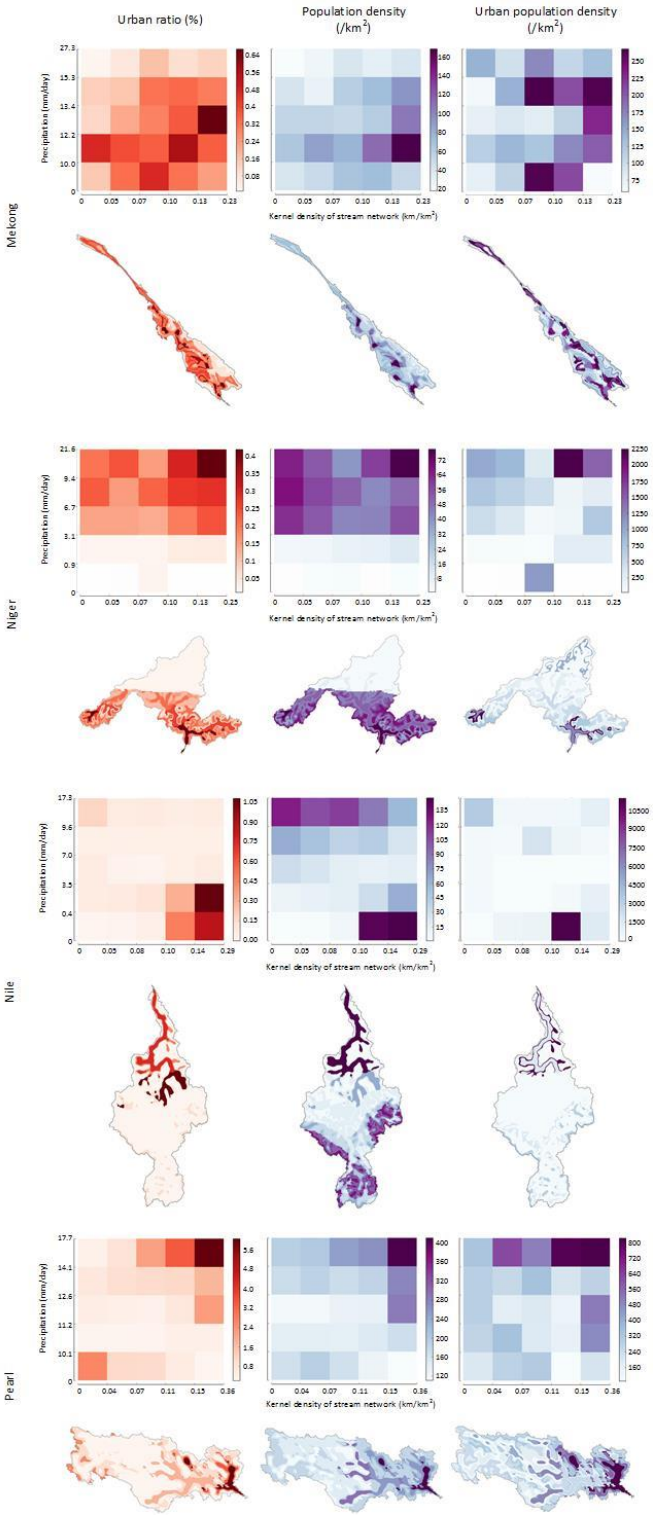


Figure 3.8 Continued

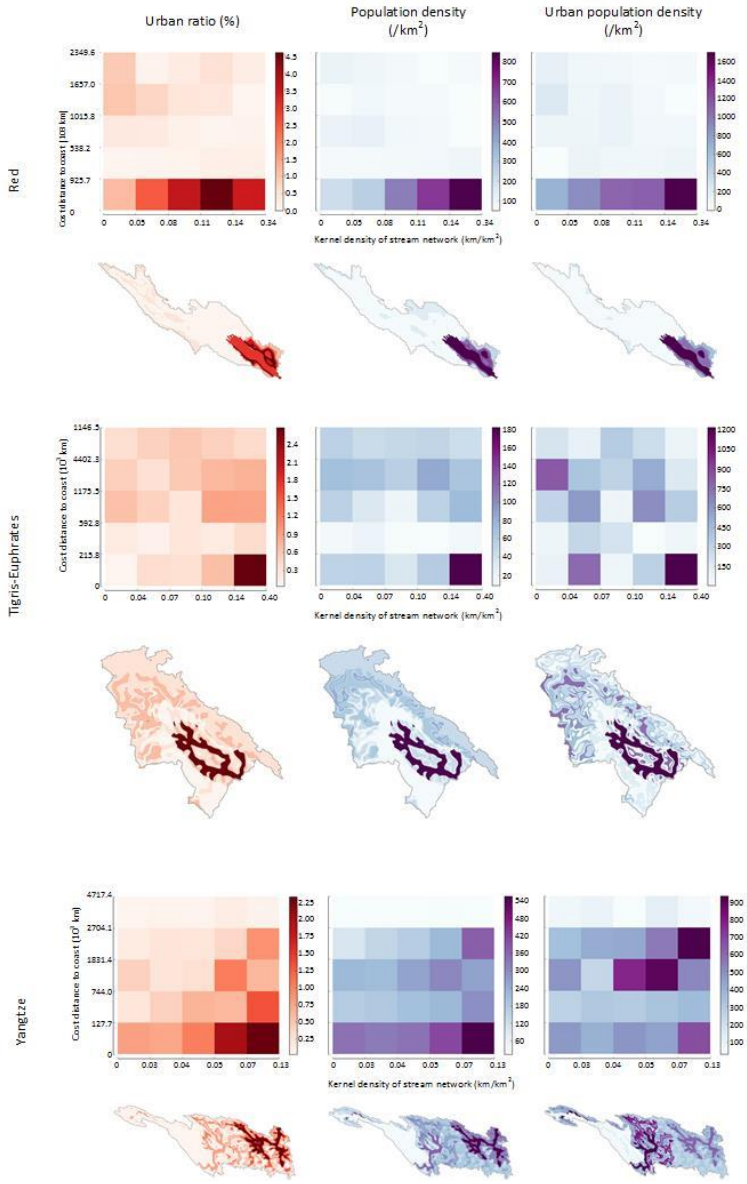


Figure 3.8 Continued

3.5 Discussion

Contemporary, detailed information on the extent of urban land and population across a river basin is important for understanding a variety of environmental issues, for

building sustainable development strategies and providing policy recommendations. The present study innovatively used kernel density and cost distance method to represent the hydro-geomorphologic attributes, based on which the spatial distribution pattern of urban land and population in 11 major river basins of Asia and Africa were analyzed. The results suggest that the urban condition indices – expressed in terms of urban ratio, population density, and urban population density – is positively correlated with the kernel density of streams and negatively correlated to the effective distance to the coast, which means cities are more concentrated in coastal zones and along streams with higher Strahler orders. This conclusion is in the lines of earlier literature (Ceola et al., 2015; Menze & Ur, 2012). Concurrently, average annual precipitation is proved not sufficient to explain the spatial distribution of the three indices. This finding is consistent with a previous analysis conducted by Small and Cohen (2004) on the global population distribution, which concluded that people are more localized associated with physiographic parameters than the climatic parameters.

There is one consensus that in the near future, population growth within flat and low-lying regions like deltas and floodplains of rivers will be faster than other places due to either demographic migration (Brettell & Hollifield, 2014; Hugo & Bardsley, 2014; Seto, 2011; Skeldon, 2014) or urban expansion based on current extent (Cohen, 2006; Pijanowski et al., 2002; Zhou et al., 2015), or both. The rapid demographic transition trend will bring myriad water challenges to 11 study river basins, which include but not limited to three aspects: Increasing urban flood risk, inadequate access to safe water, and degradation of ecosystem services. Additionally, the Nile River, the Niger River, the

Ganges-Brahmaputra, the Mekong River, and the Irrawaddy flow through 16 least developed countries in total, which make the associated river basins' sustainable development more difficult considering the lack of human, institutional and financial capital in those countries.

With most densely populated cities located in coastal zones and along rivers, nearly 90% of all natural disasters in urban areas are water related (Jha et al., 2012). In conjunction of more frequent natural hazards resulted from sea level rising, soil subsidence and climate change, the vulnerability of urban areas is growing, not only regarding humans but also vital infrastructures. Meanwhile, the conventional management strategies in terms of technical approaches are inadequate. Sivapalan et al. (2012) proposed a new science called socio-hydrology to have a better understanding of dynamics and co-evolution of the human-water system. This new field aims at increasing a system's self-organization ability by treating humans as an internal component.

Another major concern is to provide adequate access to safe water for the unchecked growing population in the urban area. Although the percentages of the urban population using improved drinking water sources in all of the countries within the 11 river basin are above 80% in 2015 (2015 MDG report), the coverages of piped water on premises, which is preferred option for urban areas, is still low except in Eygpt (100%). Therefore, access to safe drinking water sources continues to be a major problem in these developing regions. Moreover, the sanitation condition in some of the countries fails to keep pace with the urbanization rate, thus dramatically reduces the safe water availability. Indeed, all of the countries in the Niger River Basin did not meet the Millennium

Development Goal (MDG) target by 2015 to halve the proportion of people without sustainable access to safe drinking water and basic sanitation(WHO, 2015) since limited progress had been made with respect to sanitation (2015 MDG report). In general, most of the people without access to safe water and sanitation in urban areas are slum dwellers.

Urban water supply depends more on surface water than groundwater. However, for large cities located in the arid and semi-arid regions, such as Cairo in the Nile River Basin, and Ouagadougou in the Niger River Basin (Struckmeier, et al., 2008), groundwater is the primary water resource. In the humid area like the Yangtze River Basin, groundwater also serves as a buffer when surface water scarcity occurs during the prolonged droughts. Unfortunately, in the developing world, oversight of groundwater is often non-existent. Iraq and India have pumped excessive groundwater and caused dramatic water table decrease. Groundwater also has inherent limitations. In general, the groundwater is fragile to anthropogenic contaminants coming from agriculture, landfill and mining. Investigations by the Chinese Ministry of Water Resources revealed that 64% of cities have seriously polluted drinking water from groundwater sources (World Bank 2007). In some places, groundwater even contains dangerous chemicals naturally. Like in Bangladesh, high concentrations of arsenic have been found in tube wells in 61 out of 64 districts.

McDonald, Green, et al. (2011) estimated per-capita water availability by the year 2050 for major cities in developing countries and revealed that about 94 million urban dwellers in Asia and 66 million in Africa will have to deal with perennial water shortage. A large number of cities in rapidly urbanizing China and India will confront seasonal water

scarcity, with insufficient flows occurring in at least one month of the year. As easily available surface water and groundwater sources have been depleted in many urbanized areas, cities have to develop new water sources to tackle water shortage problems. Rainwater harvesting, which usually collects water from rooftop gutters, can be used to supplement water supply. It can also help to enhance the recharge of groundwater and reduce storm flood risk. Likewise, recycled waste water can satisfy most nonpotable water demands.

Concentrated cities and population combined with rapid urbanization rate in the coastal zones and areas proximate to stream networks in all of the 11 river basins point to massive infrastructure investment related to flood protection and water supply regulation, otherwise the old, low-coverage and poorly maintained infrastructure may not burden the new water management systems. Construction of flood control structures needs to take socio-hydrology into account. Though infrastructure like levees can avoid frequent and minor flooding, they may surprisingly increase flood risk by resulting in rare, catastrophic flood disasters (Werner & McNamara, 2007). Sivakumar (2012) indicates that supplementary methods such as enhancing insurance provision and raising of people's consciousness of floods are more helpful to transfer a vulnerable system into a resilient one. Construction of water supply facilities calls for the alignment of water distribution network, sanitation improvement, and reduction of unaccounted-for water. As mean unaccounted-for water (mainly due to leakage along the distribution system) rates in large cities of Asia and Africa are 42% and 39%, respectively(WHO 2000), upgraded

infrastructure with less leakage can substantially increase safe water availability, reduce chances of drinking water contamination and avoid unnecessary new facility construction.

In addition to being at risk of natural hazard, the concentrated and rapid growing populations in coastal zones and proximate to stream segments can be a burden on ecosystems. Actually, some anthropogenic activities aiming at increasing urban resilience, such as infrastructure development, are also primary drivers of ecosystem degradation. Fortunately, the "natural infrastructure" can be applied to solve the dilemma and build a win-win situation for both human populations and the environment. Natural infrastructure is the nature's equivalent to built infrastructure through the strategical use of ecosystem services. Human and natural infrastructure can mutually reinforce, and their coordinated management increases benefits by reducing costs (compared to built infrastructure). Natural infrastructure solution such as wetland restoration/construction and permeable pavement system in an urban area can not only moderate riverine flood but also help to increase available water sources through collecting, storage, and purification processes. For the coastal flood, protection of mangroves, reefs, and coastal marshes and dunes can effectively increase cities' resilience to storms. Therefore, maintaining a beneficial and balanced mix between built and natural infrastructure is an important step to achieving sustainable urban development.

As a matter of fact, whether the urban development is sustainable regarding water resources is also highly dependent on the regional integration and cooperation due to the transboundary nature of most of the 11 considered river basins. A more resilient regional

social system requires coherent and equitable water resource management strategies on a river basin scale, which meet both regional and national needs.

There are several limitations in this study that should be noted. First, the influencing factors we considered in this research do not include social, economic, and historical attributes. For example, transportation infrastructure like road network is also a primary influencing factor on the spatial distribution of urban lands and population (Güneralp & Seto, 2013). Second, we do not include detailed groundwater information concerning to water resource. However, the analysis results involve the influence of groundwater implicitly. Surface water and ground water are connective entities through recharge and baseflow; it is hard to separate them apart. Since the precipitation factor used in this analysis is a critical replenishment method of groundwater, regions with higher precipitation amount potentially have higher groundwater storage. Third, optimizing kernel density's bandwidth remains an irresolvable problem.

3.6 Conclusion

In this paper, we investigated the geographical distribution of urban land and associated population in 11 river basins of Asia and Africa, based on the condition in the year 2000. We selected average annual precipitation, freshwater availability, and accessibility of coast as three natural indicators, and employed kernel density and cost distance for the first time to represent proximity to stream network and the nearest shoreline, respectively. The approach allowed us to identify the impact of hydro-geomorphologic attributes, which are related to water resources, on historical urban site

selection and development. Interestingly, although the 11 major river basins vary in climate, landscape, and economic condition, their analyses results show a common pattern: cities and population are more concentrated in coastal zones and across floodplains of streams that have higher Strahler orders. The results also indicated that although the dependence of urban development on natural resources has decreased with the growing technology, freshwater accessibility still plays a significant role.

According to the analyses results, urban land and population distribution patterns are patchy in the 11 river basins. This phenomenon places great challenges on sustainable development. Since water is the main factor in tackling the issues, planning and financing both built and natural infrastructure in cities, particularly those located in coastal zones and across floodplains, can cost-effectively make improvement in urban resilience, human well-beings and ecosystem healthy. Although the river basins considered in this paper only consist 16% of land area in Asia and 17% in Africa, they are typical enough to represent the vast developing regions within these two continents.

CHAPTER IV

SPATIAL ANALYSIS ON URBAN LAND INTENSITY IN THE YANGTZE RIVER BASIN USING GEOGRAPHICAL WEIGHTED REGRESSION

4.1 Introduction

With the dramatic economic growth, rapid urbanization rate, and the densest population in the world, China is inevitably facing environmental problems. For example, natural hazards, water quality deterioration, and water shortage already seriously declined living standard and restricted the regional economic development. The key challenge that resource planners and decision makers must confront is how to keep a balance between urban development and environmental conservation.

Most of the previous studies assumed that urban development was mainly based on economic factors such as Gross Domestic Product (GDP), proximity to roads, and investment in capital construction, while the constraints from natural environment received less attention. Under such circumstances, interpreting the relationships between urban land's spatial characteristics and the influential environmental factors is imperative for resource management and sustainable urban development.

The primary statistical method to analyze correlation relationships is linear regression, among which the ordinary least square (OLS) regression model is the most widely applied in many fields. Unfortunately, most urban study literature using OLS method only investigated the urbanization impact on the natural environment. Research seldom explained the constraints from resource capacity on urban development. Menze

and Ur (2012) set an OLS model for the long-term settlement in northern Mesopotamia in a 23,000 km2 area and found the settlement site volume is correlated with the local precipitation and stream drainage size. Zhang et al. (2014) applied a multilinear regression model for China's eastern coastal zone and concluded that the effects of topographic factors (i.e. elevation, slope, and aspect) on settlement locations are stronger in the mountainous area. These two studies showed that environmental factors play important roles in determining both the existing urban land's location and the urban development potential in the near future.

However, OLS has obvious disadvantages on spatial analysis. The OLS model can only generate a global relationship that reflects the average condition but consequently hides some local-specific relationships. In addition, OLS has three basic assumptions: the residuals should be normally distributed, have constant variance (homoscedasticity), and be uncorrelated with each other (independence). Yet violations of the latter two assumptions, which are common for both urban and environmental data, will lead to spatial nonstationarity and autocorrelation issues. Thus, the model's efficiency will be reduced, and the result interpretation may be misleading.

In light of the OLS regression's drawbacks in dealing with spatial data, we employ a more sophisticated model, geographically weighted regression (GWR), to capture the spatial heterogeneity in the relationships between urban land distribution and influential environmental factors. GWR is similar to OLS in methodological aspects, but GWR can generate local regression results (e.g. local parameter estimates, residual, and adjusted R^2) for every sample point. GWR also contains a kernel function to quantify the neighboring

observations' weight. The kernel function reflects the first law of geography (Tobler, 1970) by assuming that a nearby observation has a greater impact on the studied point than an observation that is far away.

Given the above-mentioned advantages, GWR has become popular in applied geographical research in the environmental (Foody, 2003; Jaimes et al., 2010; Li et al., 2010; Tu, 2011), social (Hanham et al., 2009; Ogneva-Himmelberger et al., 2009), and urban (Gao & Li, 2011) research fields.

This study aims to discover the spatially varying relationships between urban land distribution and selected environmental factors using the GWR method, with the Yangtze River Basin as a case study. Specific objectives of this study were to 1) investigate the urban land spatial characteristics in the Yangtze River Basin; 2) determine the environmental factors that constrain urban development, in particular the factors related to hydro-geomorphologic attributes; 3) build OLS and GWR models to explore the relationships between urban land intensity and environmental indicators; 4) examine the GWR's advantages over OLS; and 5) discuss general management implications.

4.2 Study area

The Yangtze River is the third longest river in the world and the longest river in China. It originates from the Tibetan Plateau and drains into the East China Sea, extending more in latitudinal than in longitudinal direction (Figure 4.1). The river is over 6300 km

long with a 1.8 million km^2 drainage area. According to geology, climate, and the river's geomorphology, the basin can be divided into upper, middle, and lower reaches.

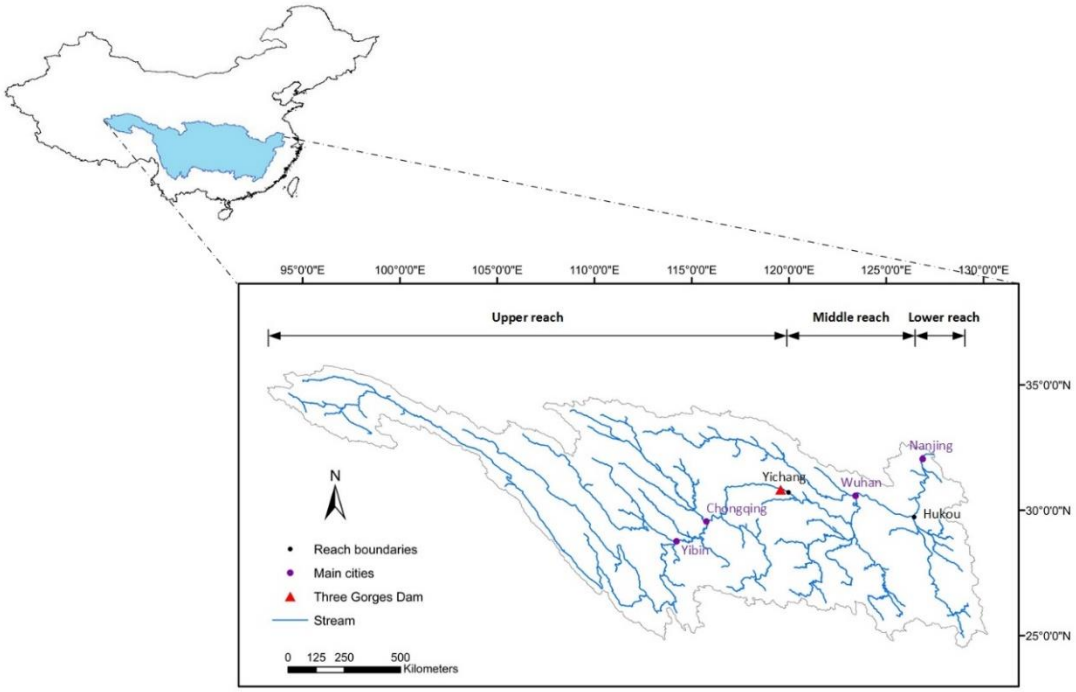


Figure 4.1 Location of the Yangtze River Basin and major cities along the streams

The upper reach of the Yangtze is about 4300 km long from the headwater to Yichang city. All the major tributaries within this section originate from the Tibetan Plateau, where the average elevation is 4500 m. As the river snakes through the V-shaped valley, it descends more than 1300 m before reaching a major city Yibin. Such a steep hydraulic gradient provides a favorable condition for hydropower generation. Constrained by the topographical attributes, city size in the upper reach is generally small, with

Chongqing and Chengdu as exceptions. These two adjacent large cities along with several nearby prefectural-level cities compose a metro region in the Sichuan Basin.

The middle reach starts in Yichang and ends at Hukou. Along the 950 km mainstem, three major tributaries join and contribute a massive amount of discharge. The slope decreases to $2\text{-}3\times10^{-5}$, and the river channels develop into a typical meandering pattern with many cutoffs, making this section the most vulnerable region to fluvial flood hazards in the basin. Wuhan is the largest city by population and economic size in the middle reach. The city is also well-known as the country's geographic heart and a transportation hub.

The final 930 km stream below Hukou constitutes the lower reach, with an average slope as 1.0×10^{-5}. In this region, four out of China's five main freshwater lakes contribute their water to the Yangtze River. The lower reach is also the most densely populated and highly urbanized area in China. The Yangtze delta, in particular, covers only 2.1% of China's territory and is home to 6% of the population, but generates about 23% of the GDP (China., 2011).

Although the delta area is important, that region is partially exclusive in this research for both technical and theoretical reasons. On one hand, since the Yangtze delta is tide-dominated, neither the most popular applied hydrographic datasets (e.g. HydroSHEDS and HYDRO1k) nor the watershed delineation methods used in hydrological models can accurately capture the delta area as a component of the river basin. On the other hand, the highly policy-driven urban development in Yangtze delta is

significantly different from other areas in the basin. Hence, Shanghai city and part of Jiangsu Province are not included in the study area.

As a humid subtropical monsoon climate dominates the Yangtze River Basin, the regional precipitation is intensive during each summer and often leads to substantial fluvial flooding. Furthermore, Northwestern Pacific tropical cyclones (typhoons) often affect the delta area by raising the sea level due to storm surge. In extreme cases, typhoon-generated storms came together with annual intensive precipitation and resulted in extensive life and property losses.

4.3 Materials and methods

The spatial analysis in this research includes five steps: 1) choose appropriate urban index and potential impact factors as the dependent variable and explanatory variables, respectively; 2) determine if variable transformation is necessary; 3) set a proper specific model for the linear relationship; and 4) run an OLS regression for model validation and improve the model by GWR method; 5) investigate the spatial variation of relationships between the urban index and selected environmental factors.

Several previous studies quantified regression model's variables by calculating the averaged values based on sub-regions. However, since the study focused on a river basin scale, such variable quantification method is improper. The procedure of dividing study area into sub-regions, whether according to administrative unit or catchment boundary, is arbitrary and breaks the first law of geography (Tobler, 1970): "*Everything is related to everything else, but near things are more related than distant things.*" The violation is

evident when two adjacent points with similar attributes are located in two sub-regions that have significantly different average values. To avoid such a problem, we employed the kernel density estimation (KDE) technique to interpolate dispersed distributed physical terms. KDE can transform a set of geographically referenced data into a continuous surface, showing the individual observation intensity over space. The KDE can result in a continuous raster that covers the whole study area and illustrates spatial differences. Meanwhile, the GWR method was applied to investigate the spatial relationships between urban land distribution and natural determinants.

4.3.1 Dependent variable

The KDE method was applied to quantify the urban intensity in the Yangtze River Basin as the dependent variable. As the theory of KDE has been introduced detailed in Chapter III, only the main application procedures including data pre-processing and bandwidth selection are described here.

The urban extent data circa 2000 in the Yangtze River Basin (Seto et al., 2012) was derived from a widely used dataset created from Collection 5 Moderate Resolution Imagine Spectroradiometer (MODIS) at a 1 km resolution (Figure 4.2). The urban extent raster was first converted into points, which serve as the KDE input. Thus, each point represents a 1 km^2 urban area. Then the KDE was conducted through the ArcGIS Spatial Analyst module (Johnston & MacCoy, 2001).

The bandwidth selection during the KDE process exhibits a strong influence on the result by controlling the smoothing degree. In most major software packages, including

the ArcGIS, the normal rule of thumb (NROT) that popularized by Silverman (1986) are implemented to determine optimal bandwidth. The NROT method differentiates the kernel density equation (Equation 3-5) with respect to bandwidth (*h*) and sets the derivative equal to zero. Consequently, the solution of bandwidth can be expressed as:

$$h_{AMISE} = \left[\frac{R(K)}{\mu_2(K)^2 R(f'')}\right]^{1/5} n^{-1/5} \tag{4-1}$$

where K is the kernel, $R(K) = \int K(x)^2 dx$, $\mu_2(K) = \int x^2 K(x)\, dx$, f'' is the second derivative of kernel density equation *f*, and *n* is the sample size. In Equation (4-1), only the second derivative term f'' is unknown. So the NROT further estimates f'' by a data-based method: let σ represent the standard deviation of a sample, then f'' can be assigned as a $[N(0, \sigma^2)]$ density noted as $f'' = \emptyset_{\sigma^2}$. When a Gaussian kernel is applied and $f'' = \emptyset_{\sigma^2}$ is placed in Equation (4-1), the optimal choice for bandwidth turns to be

$$h_{NROT} = 1.06\sigma n^{-1/5} \tag{4-2}$$

In ArcGIS, the bandwidth determination by the NROT method works as follows: First, the spatial mean center of the input points is identified, and the distances from the mean center to the rest points are calculated. Then, the median of the distance set is figured out (D_m) and the standard distance (*SD*) is calculated as:

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n} + \frac{\sum_{i=1}^{n}(y_i - \bar{Y})^2}{n}} \tag{4-3}$$

where x_i and y_i are the coordinates for feature ii, $(\bar{X}, \bar{Y})$ represents the mean center for the features, and n is the sample size. Finally, the bandwidth can be selected based on the following criteria:

$$h = 0.9 \times \min\left(SD, \sqrt{\frac{1}{\log 2}} \times D_m\right) \times n^{-1/5} \tag{4-4}$$

where the min(·) represents the smaller value of the inside two options.

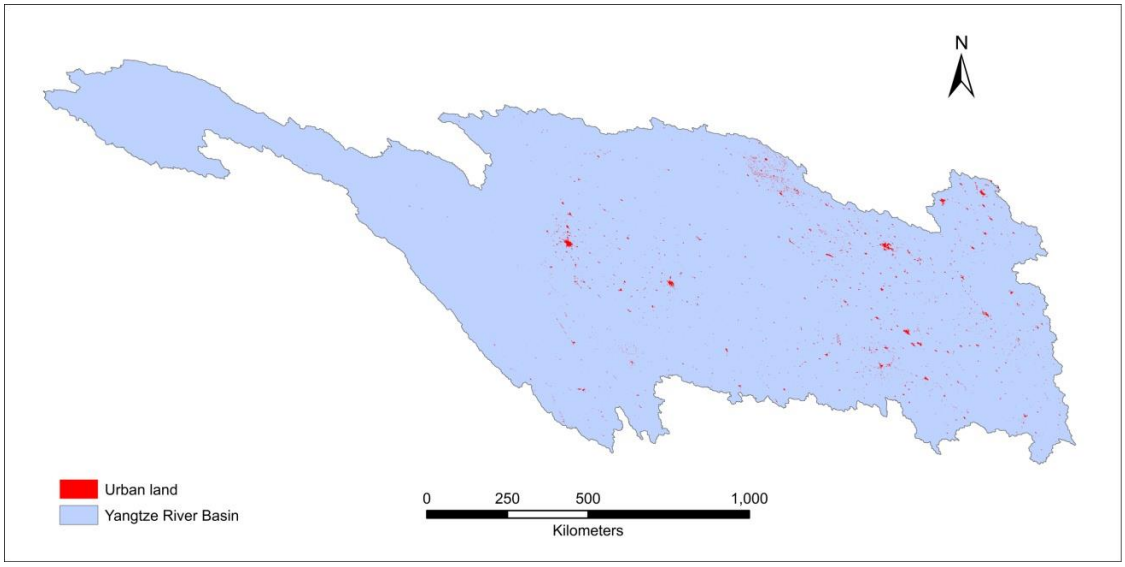


Figure 4.2 Urban land distribution in the Yangtze River Basin

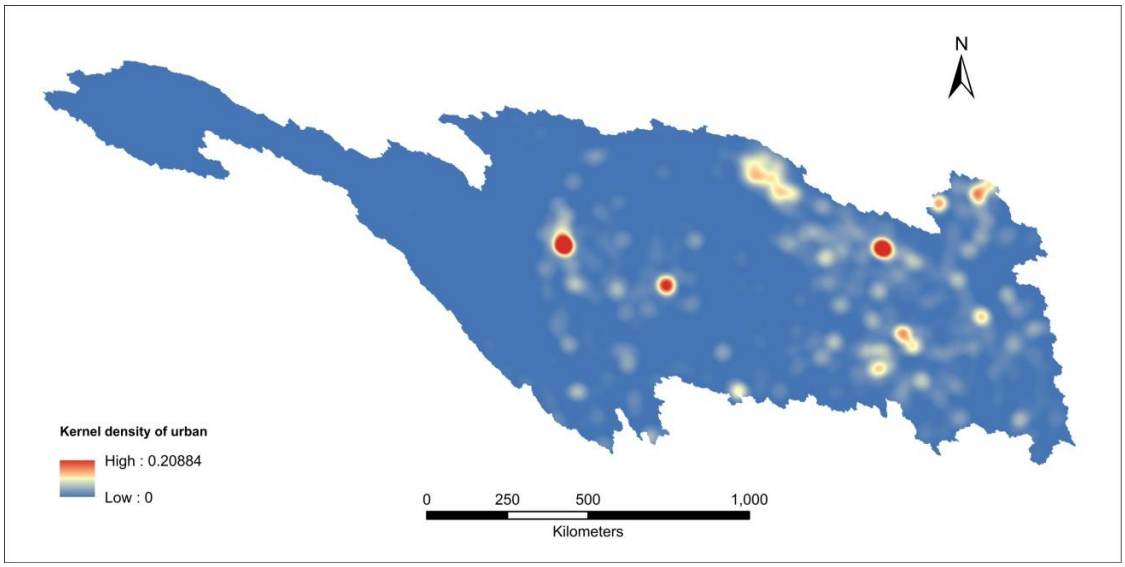


Figure 4.3 Urban intensity resulted from KDE in the Yangtze River Basin

According to the above-mentioned approach, the bandwidth for urban points in the Yangtze River Basin was selected as 42 km. One point can only influence the urban intensity in a circular region with a 42 km radius. And the attribute value of each cell (1 ×1 km^2) in the continuous raster (Figure 4.3) equals to the sum of all influential urban points' KDE.

4.3.2 Explanatory variables

The pattern, process, and spatial relationships are fundamental issues in geography. Urban development interacts intensively with associated natural attributes (Gustafson, 1998). Several previous studies indicate situation-specific geographical factors can critically drive/constrain urban settlement development. J. Y. Liu et al. (2005) analyzed urban expansion model in China during 1990-2000 and identified several driving forces based on a regression model involving 13 mega cities in China. The result shows

that the plain land with lower gradients is always more favorable for urban land expansion. Similarly, Li et al. (2011) explored the human settlements spatial distribution in Chongqing within the Three Gorges Reservoir Area and found the distribution is closely related to terrain factors such as slope.

According to the previous studies conclusions and the result of Chapter III, we selected four explanatory variables for the regression model: elevation (m), slope (percent), effective distance to the coast (km), and kernel density of stream network (km/km^2).

4.3.2.1 Elevation and slope

Elevation and slope have particular significance for determining the urban settlement location in the basin context. Comparing to hilly areas where steep topographic gradients are common, flat and low-lying regions provide more connectivity opportunity, lower construction cost, and fewer risks of natural hazards including landslide and earthquake.

In this research, the topographical information was sourced from a digital elevation model (DEM) provided by EarthEnv-DEM90 (Robinson et al., 2014). To match the urban extent map, the DEM was resampled from the original 90 m resolution to a 1 km resolution (Figure 4.4). Then the slope of each raster cell was calculated in percent rise unit.

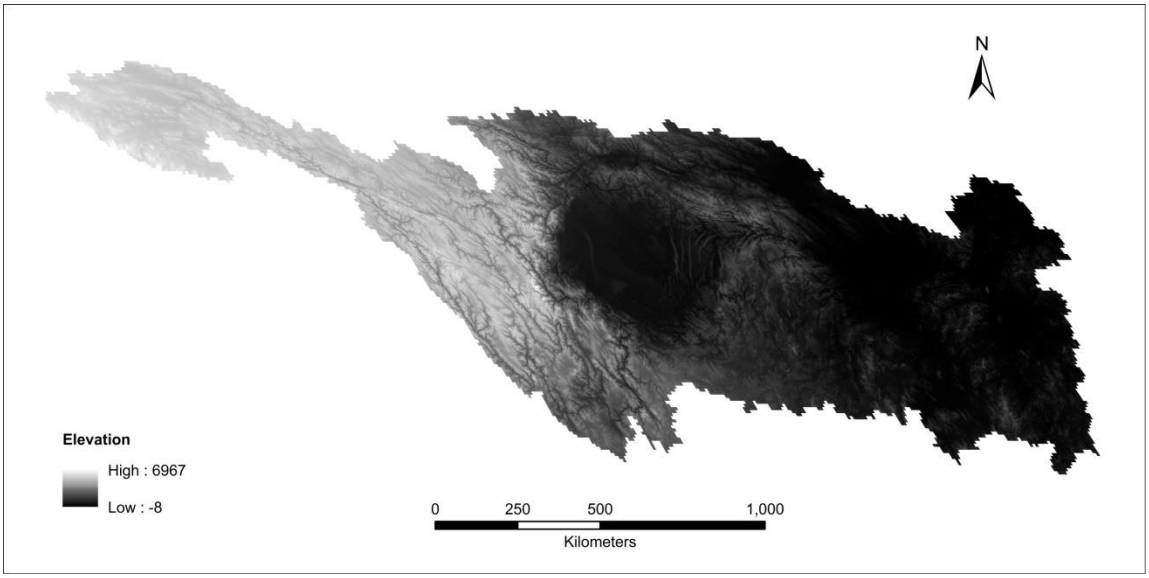


Figure 4.4 Digital elevation model (DEM) of the study area.

4.3.2.2 Effective distance to the coast

In the urban development history of the Yangtze River Basin, access to the coast influences the urban settlement location significantly for both environmental and political reasons. Since 1978, several coastal cities have thrived under the reform and opening up policy. The low-cost, export-oriented manufacturing centered in the coastal urban agglomeration has generated wealth and offered lots of job opportunities, leading to a population migration from the interior to the coast and boosting the development of the existing coastal cities. At the same time, cities that have easier access to the coast, such as Wuhan and Nanjing, also developed more rapidly than the interior cities. Therefore, the coastal proximity was considered as another explanatory variable to explain the urban land intensity within the basin.

Table 4.1 Slope reclassification and the corresponding resistant values

Slope (%)	Slope degree range	Resistant value
0.00 - 1.00	0.00 - 0.57	1
1.00 - 2.15	0.57 - 1.43	2
2.15 - 4.64	1.43 - 2.66	3
4.64 - 10.00	2.66 - 5.71	4
10.00 - 21.50	5.71 - 12.13	5
21.50 - 46.40	12.13 - 24.89	6
46.40 - 100.00	24.89 - 45.00	7

Rather than using simple Euclidean distance, this study quantified the coastal proximity by the cost distance method, which requires two input feature layers in the ArcGIS spatial analyst module: A source feature performs as the destination, and a cost layer shows resistant values in movement. In this research, the source layer was a vector shoreline of Southeast China provided by the Global Self-consistent Hierarchical High-resolution Shorelines (GSHHS). And the data with the highest resolution was applied among the five different versions (Wessel & Smith, 1996). The resistant value in the cost layer was determined by each 1×1 km^2 cell's slope. At first, the slope values were reclassified according to the criteria listed in Table 4.1. Then resistant values (Figure 4.5) were assigned corresponding to the seven slope classes (Table 4.1). The output raster of cost distance calculation (Figure 4.6) shows the weighted distance from each cell's centroid to the nearest point on shoreline through the least-cost path. With a purpose to avoid confusion, the result was named effective distance to the coast.

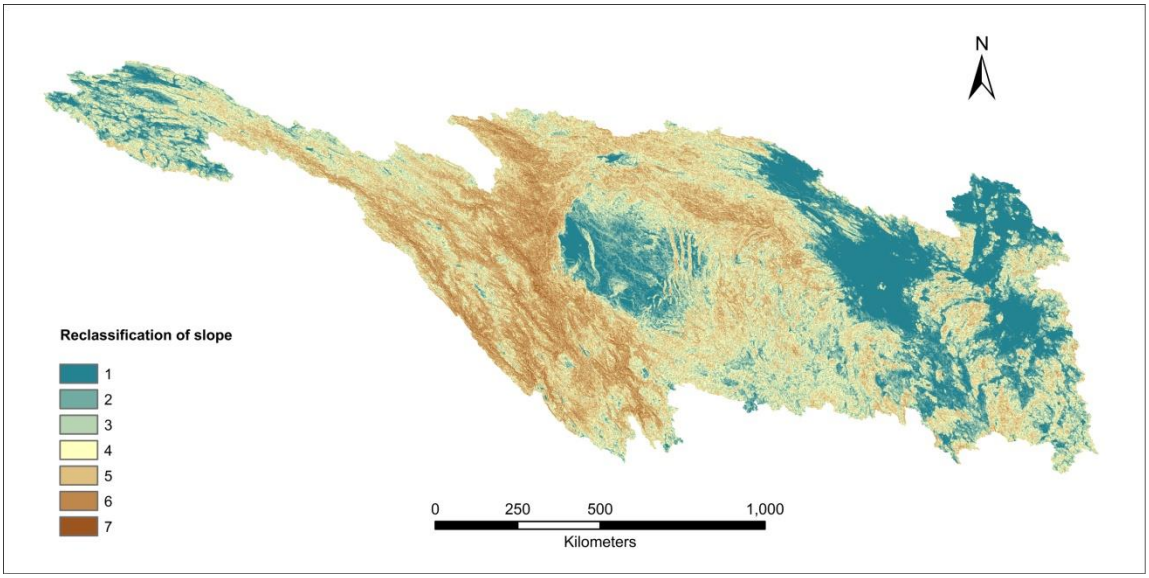


Figure 4.5 Cost layer determined by the slope values.

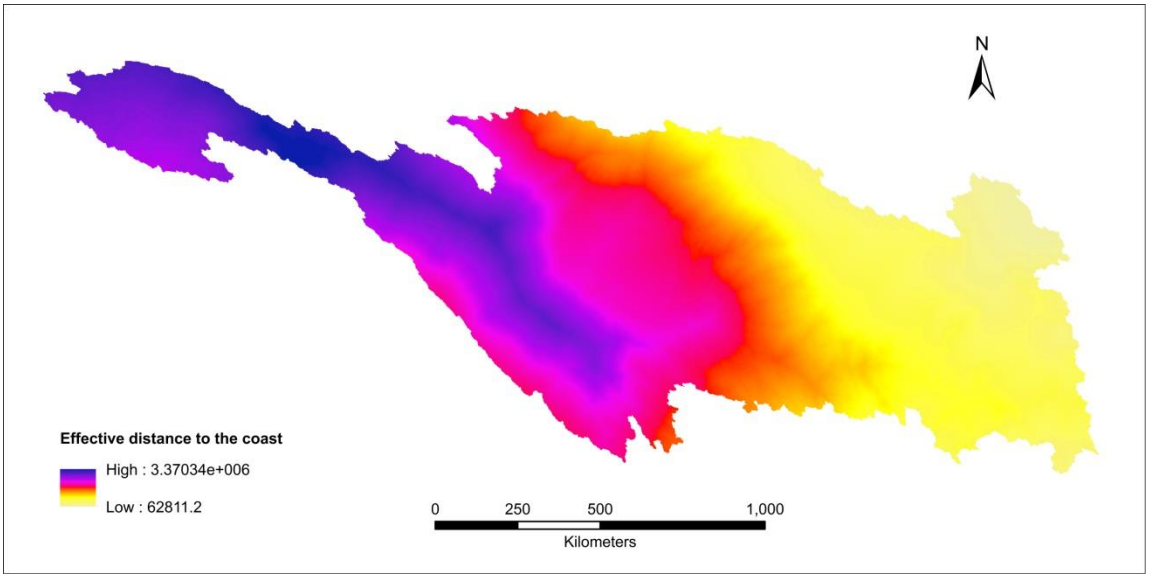


Figure 4.6 Result of the cost distance calculation showing the effective distance from each point to the coast.

4.3.2.3 Kernel density of stream network (KDSN)

Proximity to the Yangtze River's channel exerts environmental and economic influences on urbanization process. A location closer to a stream segment always has easier access to fresh surface water, which is the most important resource for human life, ecological health, food security, industrialization and urban development.

Furthermore, the Yangtze River's 2800-km-long navigable channel serves as a central superhighway for goods and people, which connects China's less-developed interior regions to the coast and between each other. Cities along the Yangtze are inherently more viable as manufacturing and trade hubs because of the beneficial geographic attributes: Waterway transport costs 1/30 as much as transport by road, or 1/3 by railway.

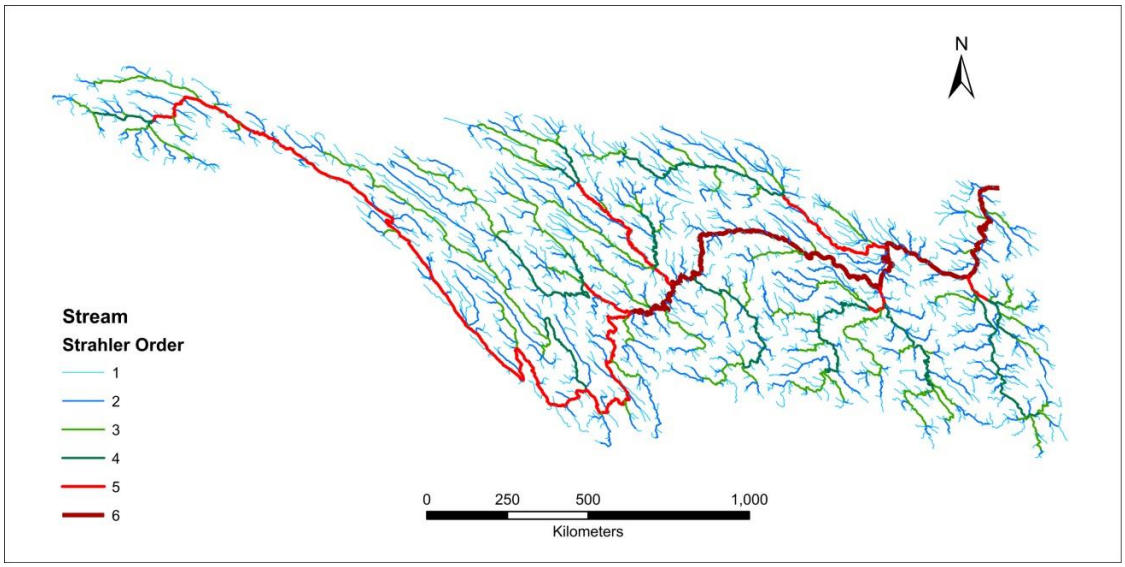


Figure 4.7 Stream network of the Yangtze with Strahler order

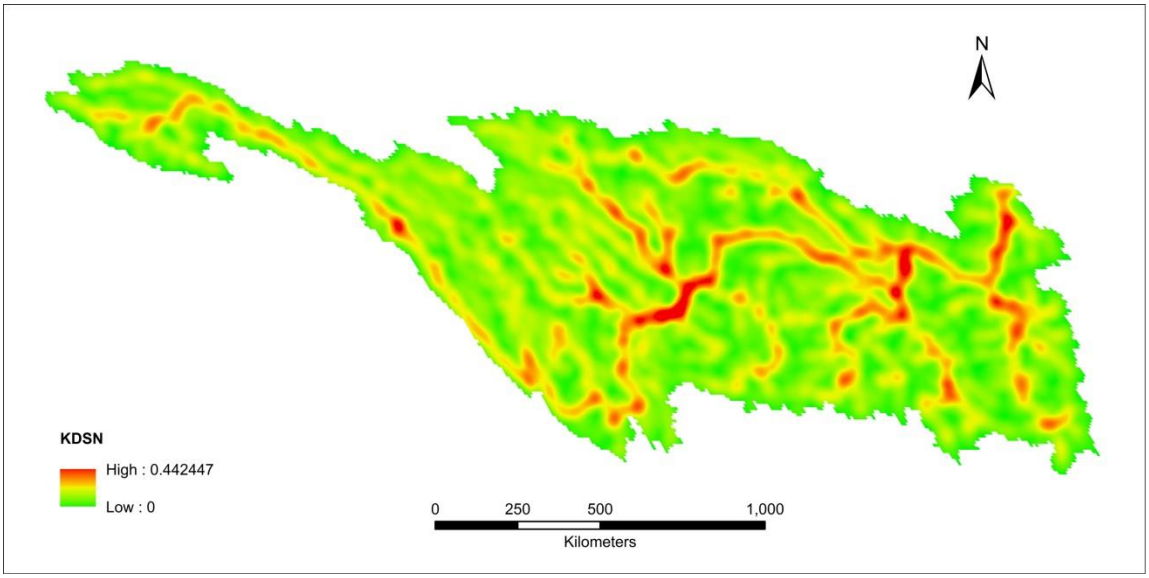


Figure 4.8 Map of KDSN in the Yangtze River Basin

Some previous studies represented the proximity to river by horizontal Euclidean distance from a location to the nearest permanent stream. However, this method assumes that the demanded water resource can only be derived from the nearest stream segment, regardless of the drainage area or flow capacity. Actually, the Euclidean distance method fails to reflect the combined influences when a study area is on a plain created by two or more stream segments. Hence, in this study, the proximity to river was quantified as the kernel density of Yangtze's stream network (Figure 4.7) weighted by the Strahler order. In this circumstances, a point with a larger KDSN value in the resulted continuous surface layer (Figure 4.8) indicates the place is located either in a stream densely distributed region or near a stream segment with a higher Strahler order. In addition, as the stream segment with a higher Strahler order always has more joined tributaries, it is reasonable to assume that a higher Strahler order implies more discharge or more fresh surface water available. Detailed calculation procedures have been described in Chapter III.

4.3.3 Sample drawing and variable transformation

Three steps were conducted to obtain the sample data for regression models. First, 2000 random points were created within the Yangtze River Basin. The minimum allowed distance between any two points was set as 5 km to avoid cluster. Second, both the dependent and explanatory variable values were extracted from raster layers to the 2000 points. At this point, abbreviations instead of variables' full names were assigned (Table 4.2). At last, 1512 points were used to set up the regression model after removing the cases with KDUrb = 0 and the outliers (Figure 4.9). Here the outlier is defined as a sample value which is 1.5 times the interquartile range larger than the third quartile or smaller than the first quartile.

Table 4.2 Abbreviation and description of model variables

Variables	Description
KDUrb	Kernel density of urban land (Dependent variable)
Elevation	Elevation obtained from DEM (Explanatory variable)
Slope	Slope calculated from DEM (Explanatory variable)
EDTC	Effective distance to the coast (Explanatory variable)
KDSN	Kernel density of stream network (Explanatory variable)

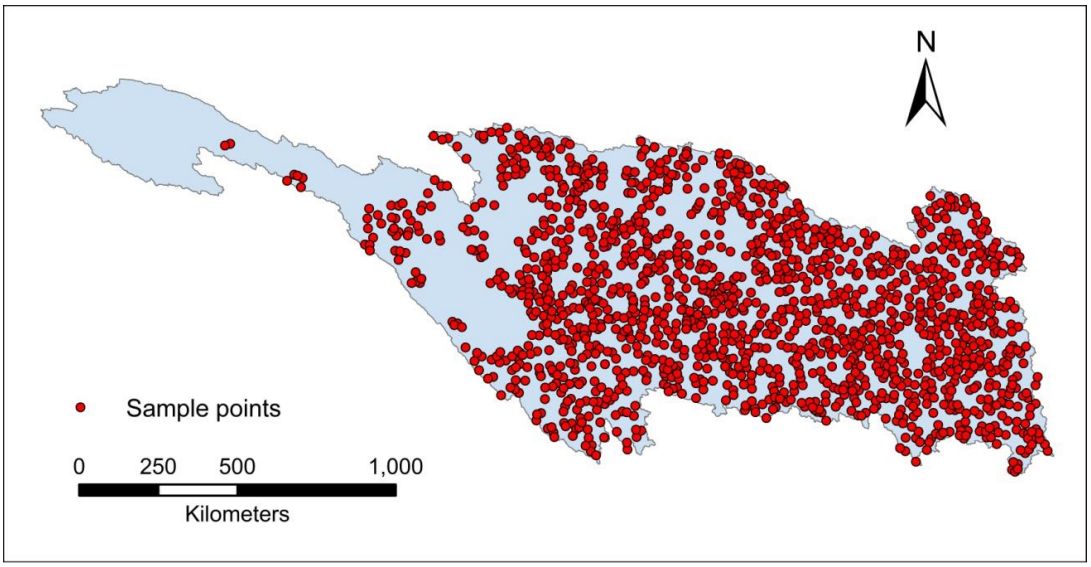


Figure 4.9 Distribution of sample points drawn for the regression.

Scatterplots of KDUrb versus each candidate explanatory variable were graphed (Figure 4.10) to check whether linear relationships exist. Since the trends in KDUrb versus EDTC and KDUrb versus KDSN are not obvious, keeping the original variable metrics will lead to significant variances in the regression model. Furthermore, the histogram of KDUrb (Figure 4.11 left) indicates the dependent variable is also highly skewed. Although OLS model does not require variables to be normally distributed, the natural log transformation was performed on the KDUrb, EDTC, and KDSN variables. Consequently, the transformed dependent variable closely followed a normal distribution and got better linear relationships with the new set of explanatory variables (Figure 4.11).

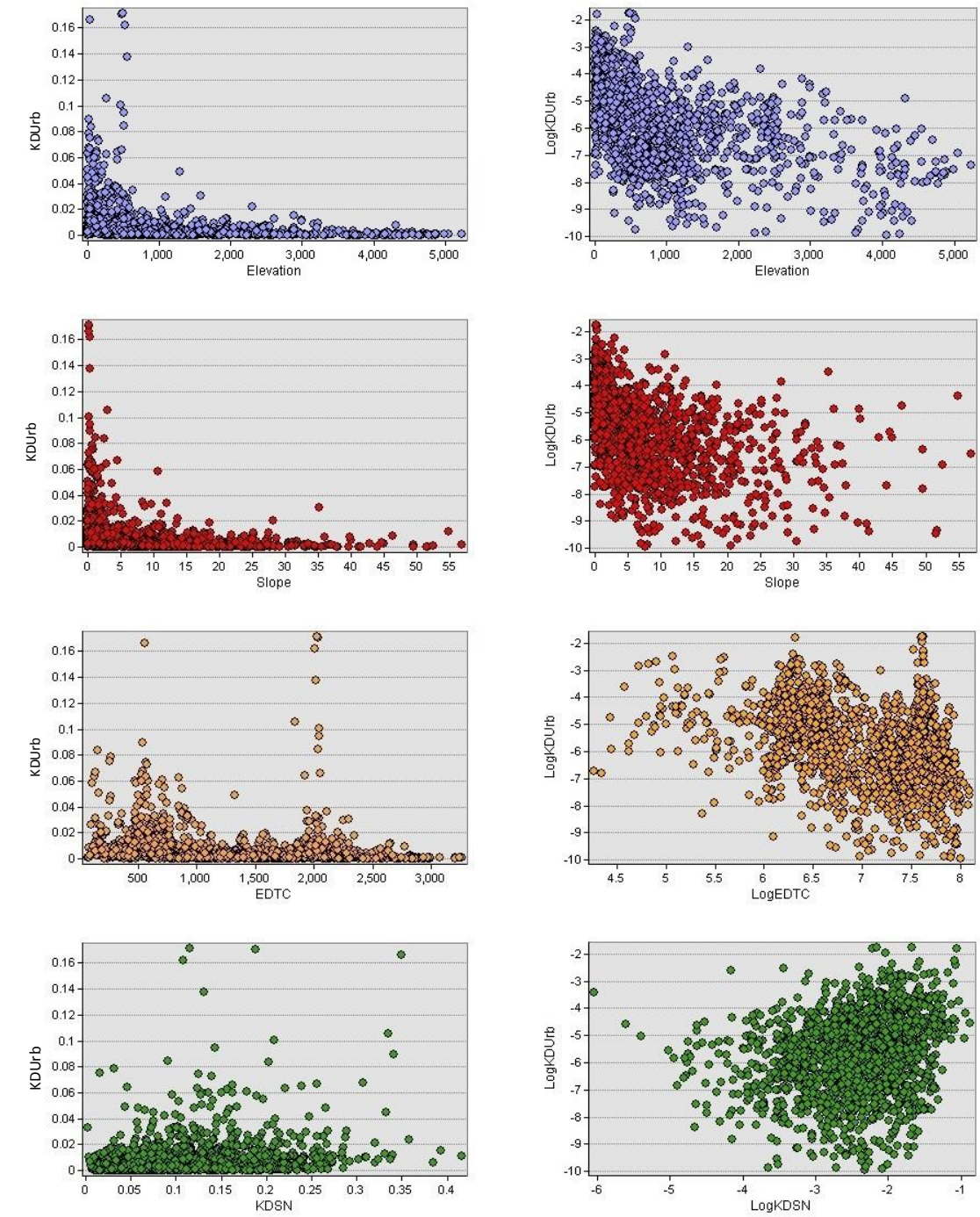


Figure 4.10 Scatter plots of the KDUrb versus each explanatory variable

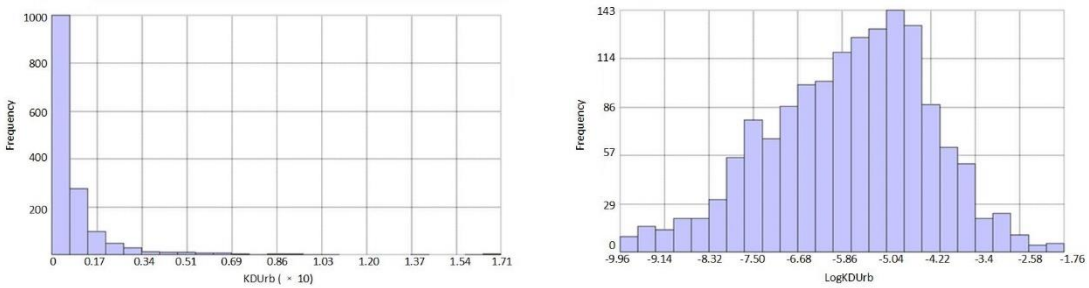


Figure 4.11 Histograms of dependent variable with (left) original format and (right) log-transformed format

To summarize, the dependent variable is the kernel density of urban land (KDUrb); the candidate explanatory variables include elevation, slope, natural logarithm of effective distance to the coast (LogEDTC), and natural logarithm of stream network's kernel density (LogKDSN). The descriptive statistics of each variable are shown in Table 4.3.

Table 4.3 Descriptive statistics of model variables

Variables	Min	Mean	Max	Std Dev
LogKDUrb	-9.956	-5.731	-1.764	1.487
Elevation	3.000	919.747	5235.000	1071.762
Slope	0.000	7.764	56.875	8.720
LogEDTC	4.267	6.925	8.090	0.707
LogKDSN	-6.046	-2.466	-0.877	0.736

After selecting the sample data, a Gettis-Ord Gi* statistic was conducted to determine whether and where points with high or low LogKDUrb values cluster spatially. The Hot Spot Analysis Tool in the ArcGIS was applied by analyzing each point within the

neighborhood context. A threshold searching distance was set as 42 km, identical to the predefined bandwidth for KDUrb.

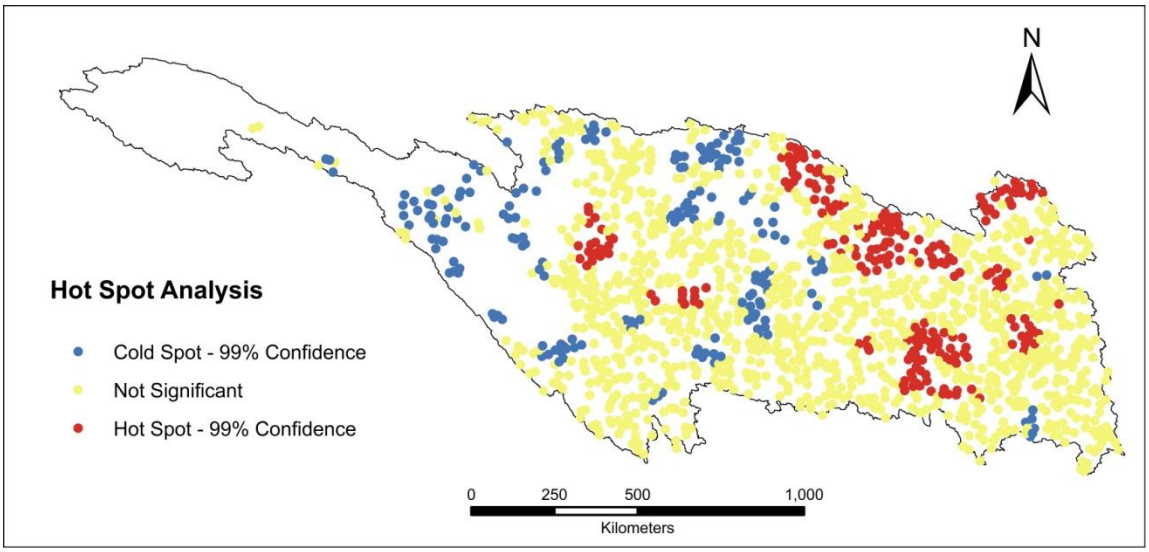


Figure 4.12 Hot spot analysis on spatial statistical trends of high and low urban intensity clumping. Statistical significance measured by the z-score, standard deviations

In the resulting map (Figure 4.12), two cluster patterns were identified as hot spots and cold spots respectively. A red point shows a hot spot at the 99% confidence level, meaning a point not only has a relatively large attribute value but also is surrounded by other points with high values. On the contrary, a blue point or a cold spot represents a low-valued point with neighbors that also have low attribute values. The rest points displayed in yellow are not included in any statistically significant clusters.

There are obvious regional differences in urban intensity across the Yangtze River Basin. Areas with high development intensity include Chongqing-Chengdu district, Xiangfan-Yichang-Wuhan triangle megaregion, Changsha-Zhuzhou-Xiangtan urban

agglomeration, and some provincial capitals such as Nanchang, Hefei, and Nanjing. A majority of the hot spots are located in the middle and lower reaches. There are also low development intensity areas such as the source area, Hengduan Mountains, Qinling Mountains, Wuling Mountains and Nanling Mountains. In general, hot spots are located in the low-flat plains; while cold spots all occur at locations more than 1300 m above the average sea level. This phenomenon indicates elevation factor constrains regional urban development significantly in the Yangtze River Basin.

In summary, the hot spot analysis results serve as strong evidence that the KDE method did a decent work on capturing the urban land distribution characteristics. The visualized map also offered clues about the phenomenon's underlying causes in relation to environmental factors, which enhanced the credibility of the explanatory variables.

4.3.4 Regression analysis processes

The regression analysis contains three main steps. First, exploratory regression was performed to test whether all selected environmental factors contribute to influence the urban land intensity based on linear relationships, then a properly specified model was determined. Second, an OLS regression was run to validate the overall significance of the model. At last, a GWR was set up using the same dependent and explanatory variables as the OLS model to show the spatial varying relationships. The analysis was performed through two software packages: ArcGIS is responsible for implementing the exploratory regression, OLS global analysis, and data visualization; the software GWR 4.0 was employed to carry out local regression models.

4.3.4.1 Exploratory regression

The exploratory regression builds models using all possible combinations of the candidate explanatory variables, assesses which models pass the necessary statistics checks and determine the proper specific model by three metrics: adjusted R^2, bias-corrected Akaike's information criterion (AICc) and variance inflation factor (VIF).

The adjusted R^2 is a modified version of R^2, indicating how well the observed data fit a statistical model. The adjusted R^2 considers sample size and only increases when an additional explanatory variable improves the model more than expected by chance. So the adjusted R^2 value is always lower than the R^2, and in some cases, the adjusted R^2 value is even negative. Higher adjusted R^2 implies that the regression model can account for more dependent variable variance.

Another important diagnostic for assessing model is the AICc. The AICc value is an indicator for comparing regression models with the same dependent variable but different explanatory variables. Generally, a model with a smaller AICc value means a better fit to the observed data, and the associated combination of explanatory variables are closer to the reality. Differences in AICc values that are less than three are inconsequential.

The VIF value measures redundancy among explanatory variables (i.e. multicollinearity). As a rule of thumb, an explanatory variable with a VIF larger than 7.5 is of concern and indicates at least two explanatory variables are telling the same story. In this case, the explanatory variable with the highest VIF value should be removed from the regression model.

Eventually, among the ones that have passes the multicollinearity check, a model with the lowest AICc and a higher adjusted R^2 and can be selected as the proper specific regression model.

4.3.4.2 OLS

OLS is a common technique applied in multivariate analysis. For a sample that contains *i* observations, OLS model is written as:

$$y_i = \beta_0 + \sum_k \beta_k x_{ki} + \varepsilon_i \quad (4\text{-}5)$$

where y is the dependent variable, β_0 is the intercept, k is the number of explanatory variables, β_k is the coefficient of the explanatory variable x_k, and ε is a random error (also called a residual), which indicates the difference between the dependent variable's observed value and predicted value.

OLS has three principal assumptions on residuals. First, the residuals are normally distributed with a mean of zero. Second, the residuals should be homoscedastic, which means the residual variance is spatially constant. Third, the residuals are statistically independent. In other words, there should have no correlation between consecutive residuals. Violation of these assumptions may lead to inaccurate estimations that are inefficient and biased. However, in reality, spatial data often fail to achieve all the assumptions owing to the intrinsic attributes: spatial autocorrelation and spatial nonstationarity.

Spatial autocorrelation pertains to the non-random pattern of attribute values among closely-spaced observations. Two general autocorrelation forms exist: positive

form reflects value similarity or cluster; while negative one reflects dissimilarity and dispersed pattern. This phenomenon violates the independence assumption.

Spatial nonstationarity indicates a lack of stability on the relationships over the study area, which is potentially owing to imperfect data collection (Brunsdon et al., 1998a) or systematic biological, environmental and physical processes in the natural environment (Ecker et al., 2013; Gregersen et al., 2013; Legendre & Legendre, 1998) . Nonstationarity violates the homoscedasticity assumption and leads to biased parameter estimates and misleading significance tests. These drawbacks of OLS in dealing with spatial data are the foundation for developing new models fit for spatial analysis such as GWR.

4.3.4.3 GWR

Local regression and smoothing techniques were the original developments in the GWR (Brunsdon et al., 1998b; Fotheringham et al., 1998). Then the model was further improved with the development of statistical measures such as maximum likelihood kernel bandwidth selection (Páez et al., 2002). Recently, GWR gains increasing popularity in many disciplines due to its ability to capture both spatial dependency and spatial autocorrelation.

GWR extends traditional regression framework by allowing local rather than global parameters to be estimated. So the GWR model can be expressed as:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ki} + \varepsilon_i \qquad (4\text{-}6)$$

where (u_i, v_i) is the coordinates of the i_{th} point in space, $\beta_0(u_i, v_i)$ and $\beta_k(u_i, v_i)$ represent the intersect and coefficient at point i, respectively. Since the local coefficients have a greater flexibility over space, GWR is expected to provide a better model fit than OLS in residuals.

Using the best combination of explanatory variables, the GWR model was conducted by GWR 4.0 software under the fixed Gaussian kernel type and AICc minimization method. The optimal bandwidth was calculated by the software's iteration algorithm.

4.4 Results

4.4.1 Exploratory regression result

Table 4.4 Results of exploratory regression

Number of variable(s)	Combination	Adj R^2	AICc
1	Elevation	0.27	5022.16
	Slope	0.19	5168.45
	LogEDTC	0.15	5246.11
2	Elevation & Slope	0.30	4963.87
	Elevation & LogKDSN	0.30	4964.87
	Elevation & LogEDTC	0.27	5015.02
3	Elevation, Slope & LogKDSN	0.32	4904.49
	Elevation, LogKDSN & LogEDTC	0.31	4945.95
	Elevation, Slope & LogEDTC	0.30	4959.91
4	Elevation, Slope, LogEDTC & LogKDSN	0.33	4890.47

Table 4.4 summarizes the output of exploratory regression and groups the tested models by the number of explanatory variables. Only the three models with the highest

adjusted R^2 values are listed for each category, and the associated AICc values are also shown. As a rule of thumb, the combination of explanatory variables with the maximum adjusted R^2 and the minimum AICc is granted as the best model. Therefore, the subset including all four candidate explanatory variables performs better than the rest, and the specified model determined for the OLS regression can be expressed as:

$$\log(\text{KDUrb}) = \beta_0 + \beta_1 \cdot \text{Elevation} + \beta_2 \cdot \text{slope} + \beta_3 \cdot \log(\text{EDTC}) + \beta_4 \cdot \log(\text{KDSN}) + \varepsilon \quad (4\text{-}7)$$

Furthermore, the VIF values of the four explanatory variables (Table 4.5) indicate the multicollinearity issue does not exist in the specified model.

Table 4.5 VIF result of each explanatory variable

Explanatory variable	VIF
Elevation	2.245
Slope	1.538
LogEDTC	1.829
LogKDSN	1.042

4.4.2 OLS result

4.4.2.1 Evaluation of explanatory variables

According to both the probability and the robust probability results of the t-test (Table 4.6), all explanatory variables are statistically significant in the multilinear regression at the 99% confidence level (p-value < 0.01). Here the robust probabilities are accurate even when the relationships among the spatial data are nonstationary. In other

words, when the relationship is nonstationary, only the robust probabilities can serve as credible parameters to tell if an explanatory variable is statistically significant.

Table 4.6 Summary of OLS results

Variable	Coefficient $(\hat{\beta})$	StdError	t-value	Prob	Robust_t	Robust_Prob
Intercept	-2.494	0.424	-5.887	< 0.0001	-6.094	< 0.0001
Elevation	-0.0004	0.000	-9.858	< 0.0001	-10.218	< 0.0001
Slope	-0.034	0.004	-7.643	< 0.0001	-6.297	< 0.0001
LogEDTC	-0.240	0.060	-4.009	< 0.0001	-4.089	< 0.0001
LogKDSN	0.371	0.043	8.540	< 0.0001	8.347	< 0.0001

The coefficient shows the strength and type of the relationship between an explanatory variable and the dependent variable. In the OLS model, the elevation, slope, and LogEDTC factors were assigned negative coefficients, indicating that the KDUrb decreases as the values of elevation and slope increase, and when the region is farther from the coast. The KDSN factor's positive coefficient indicates that more urban agglomerations are located proximate to river channels.

The coefficient also reflects the expected change in the dependent variable for every one unit change in a specific explanatory variable, holding all other variables constant. The relationships of the elevation and slope factors to the dependent variable are based on a log-linear model. Since coefficients of these two factors are small, we can assume that $e^{\hat{\beta}} \approx 1 + \hat{\beta}$ and conclude directly according to the coefficients that one-meter increase in elevation will decrease the KDUrb by 0.04%, and one percent increase in slope corresponds to an expected 3.4% decrease in KDUrb.

The other two factors, KDSN and EDTC, were both log-transformed along with the dependent variable, the expected change of KDUrb based on one unit increase in these two explanatory variables (all unlogged) cannot be shown directly in Table 4.6. In this case, a 1% increase in an explanatory variable will get a $\left(1 - e^{\widehat{\beta}\cdot\log(1.01)}\right)$ proportional change in the dependent variable. Therefore, a 1% increase in the effective distance to the coast will decrease the KDUrb by 0.2%. Similarly, a 1% increase in the kernel density of stream network would yield a 0.4% increase in the KDUrb.

The Pearson correlations between KDUrb and explanatory variables are shown in Table 4.6. Elevation and slope factors exhibit negative correlations while KDSN has a positive correlation with the dependent variable. Generally, more urban settlements are located in the lower flat region rather than areas with high elevation or large topographical gradient. Meanwhile, urban agglomeration occurs more frequently near streams.

In a summary, there is no redundancy existing among the explanatory variables, and each explanatory variable is associated with a statistically significant coefficient that matches with theory and common sense. In this case, all these four natural factors are important to explain the urban land distribution in the Yangtze River Basin. Primary linear relationships between each explanatory variable and the dependent variable are shown in Figure 4.13

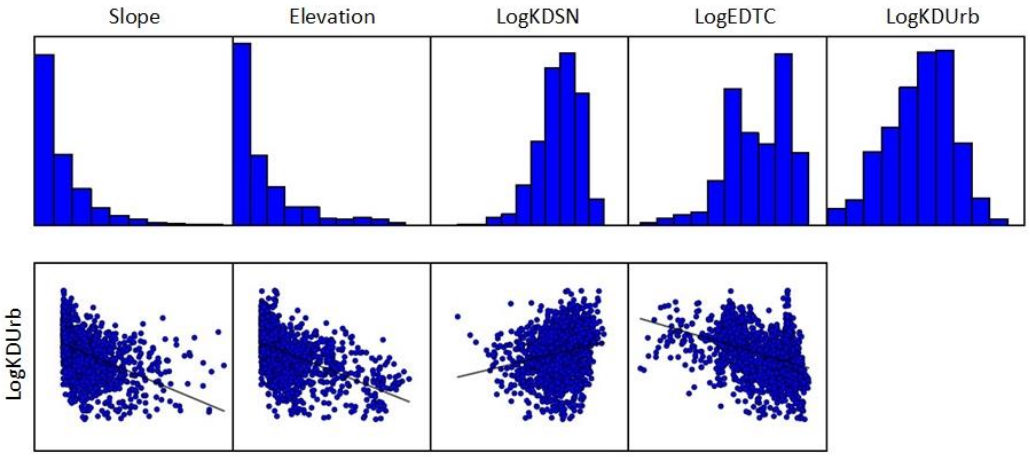


Figure 4.13 Variable distributions and relationships

4.4.2.2 Model performance

Model significance

In the OLS model, the adjusted R^2 is 0.33, indicating that the selected natural environmental factors together explain approximately 33% of the variation in the dependent variable.

Table 4.7 OLS diagnostics

Number of observations	1512
Number of variables	4
AICc	4890.41
R^2	0.332
Adjusted R^2	0.331
Joint F-Statistic: 188.059	Prob(>F), (4,1507) degrees of freedom: 0.0000*
Joint Wald Statistic: 757.4	Prob(>chi-squared), (4) degrees of freedom: 0.0000*
Koenker (BP) Statistic: 34.568	Prob(>chi-squared), (4) degrees of freedom: 0.00001*
Jarque-Bera Statistic: 15.861	Prob(>chi-squared), (2) degrees of freedom: 0.00036*

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

Both the Joint F-Statistic and the Joint Wald Statistic can test the overall model's statistical significance. The two tests have a same null hypothesis: the explanatory variables in the regression model are not effective. For a 99% confidence level, a p-value that is smaller than 0.01 indicates a statistically significant model. The choice between these two overall model tests is based on the Koenker (BP) statistic, which will be described later. When the Koenker (BP) test is not significant, the Joint F-Statistic is convincing. But when the Koenker (BP) test is significant, only the Joint Wald Statistic can be trusted. According to the values of the Joint F-Statistic and the Joint Wald Statistic listed in Table 4.7, both tests are significant at the 99% confidence level. So the overall OLS model is statistically significant.

Tests of assumptions on residuals

The output of OLS in ArcGIS also returned a map of the standardized residuals (Figure 4.14) representing the regression model's over- and underpredictions. Red points show locations where actual values of dependent variance are higher than the estimations; blue points' actual values are lower than the model's prediction.

It is easy to figure out the most over-predicted points, whose residuals are 2.5 times of standard deviation less than the mean, spatially overlap some cold spots of the LogKDUrb. And points with high dependent variable values were in general underestimated. This phenomenon implies that the global analysis result of OLS can only reflect the average condition, but ignore some local-specific relationships. The low adjusted R^2 yielded by OLS is also attributed to the high variance of residuals.

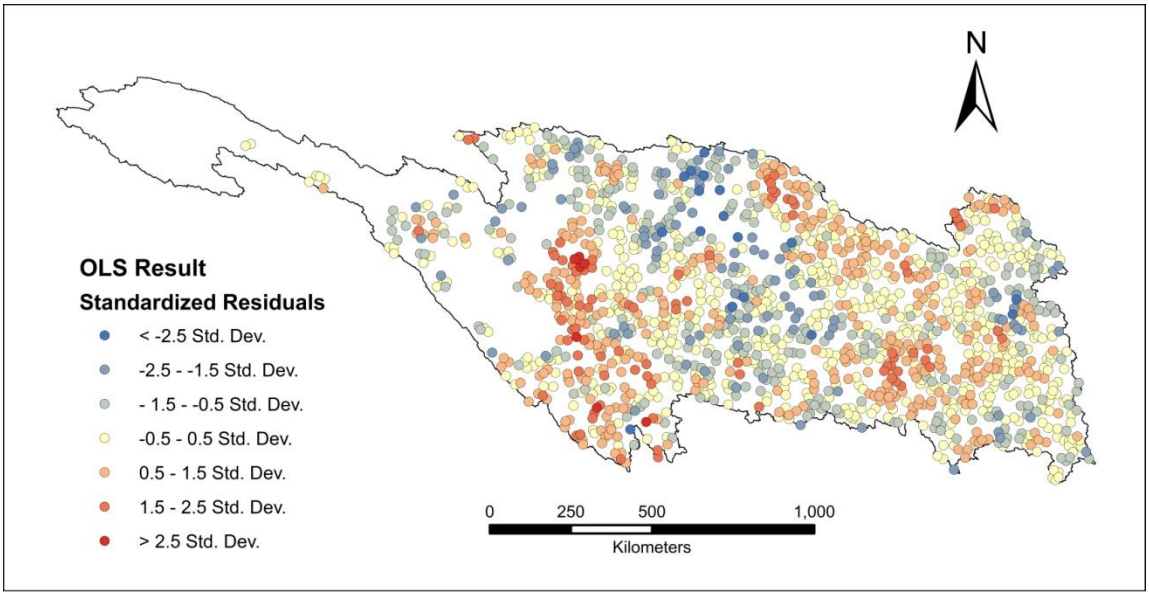


Figure 4.14 Map of OLS residuals

Testing for bias

The Jarque-Bera statistic tests whether the OLS model is biased, with a null hypothesis that the residuals are normally distributed. In other words, a model is considered unbiased if the result shows the Jarque-Bera test is not statistically significant. In this study, the Jarque-Bera's p-value is smaller than 0.01, indicating the residuals are not normally distributed at a 99% confidence level. However, the value of Jarque-Bera statistic is highly influenced by the heteroscedasticity of spatial data. So the violation of residuals normal distribution can be triggered by spatial autocorrelation. In order to avoid the confusion, two alternative assessments focusing solely on the residuals were provided to determine if the OLS is biased.

First, a histogram of the standardized residuals was constructed and compared to the normal distribution curve. Here the standardized residual is a transformation from the

actual residuals. The two kinds of residuals show the same distribution, but the standard residuals have a mean of zero and a standard deviation of one. As shown in Figure 4.15, the histogram of standardized residuals resembles the classic bell-shaped curve. So the residuals are approximately normally distributed.

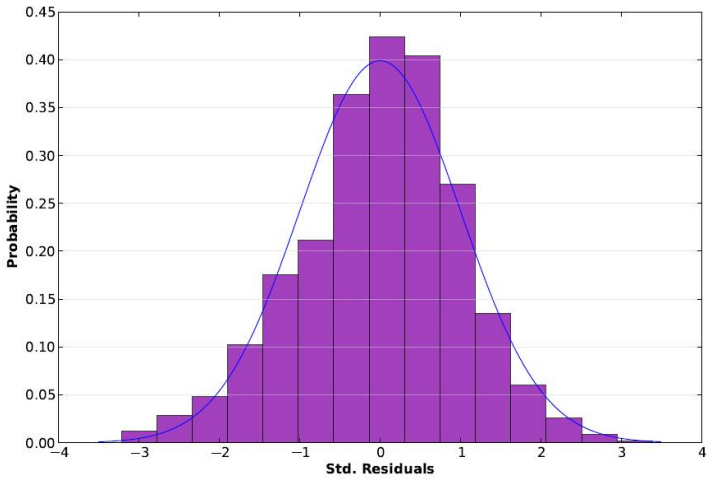


Figure 4.15 Histogram of standardized residuals yielded from OLS model

In order to get more persuasive evidence, a quantile-quantile (Q-Q) plot was created to validate the normal distribution for the standard residuals (Figure 4.16). Generally, a Q-Q plot is more powerful than the common technique of comparing histograms. The basic idea of Q-Q plot is to compute the theoretically expected standard residual for each data point based on normal distribution, and then plot the quantiles of actual standardized residuals versus the quantiles of theoretical standardized residuals. If the data follow the normal distribution, then the points on the Q-Q plot should fall approximately on the diagonal reference line.

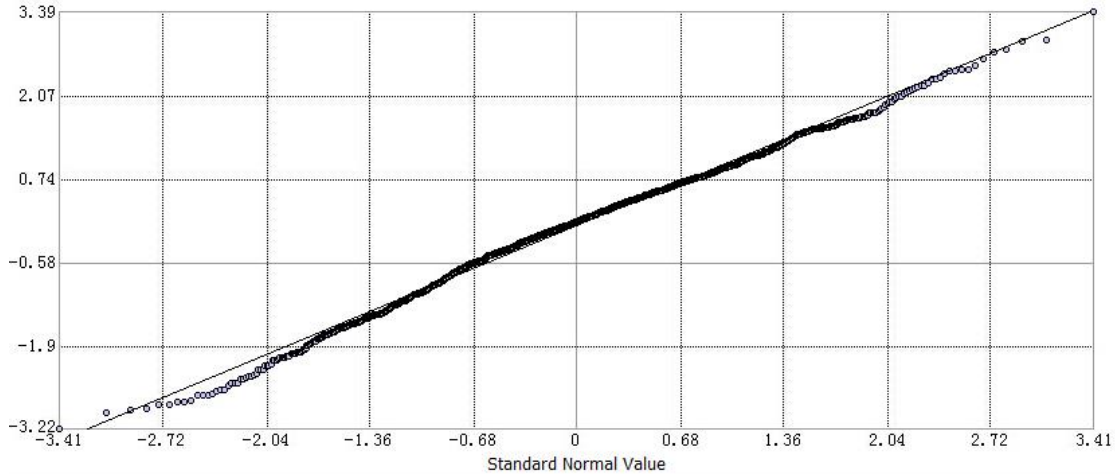


Figure 4.16 A normal Q-Q plot of actual standardized residuals versus the quantiles of theoretical standardized residuals

The set of tests indicates that although the Jarque-Bera statistic is significant, the distribution of residuals in the OLS model is actually normal distributes. Thus, there is small opportunity that the OLS is biased. Furthermore, the conflict Jarque-Bera statistic result may probably be caused by the spatial autocorrelation of the residuals.

Testing for homoscedasticity and stationary

The Koenker (BP) statistic is a test to determine whether the explanatory variables in the model have a consistent relationship to the dependent variable in geographic space. The model's consistency in geographic space declares that modeled spatial processes have the same performance across the whole study area (i.e. stationary). In the OLS result, the Koenker p-value is smaller than 0.01, declaring the linear relationships vary across the study area at the 1% significant level and are therefore nonstationary. Generally, such a

regression model with statistically significant nonstationarity is a good candidate for GWR analysis.

Testing for independence of residuals

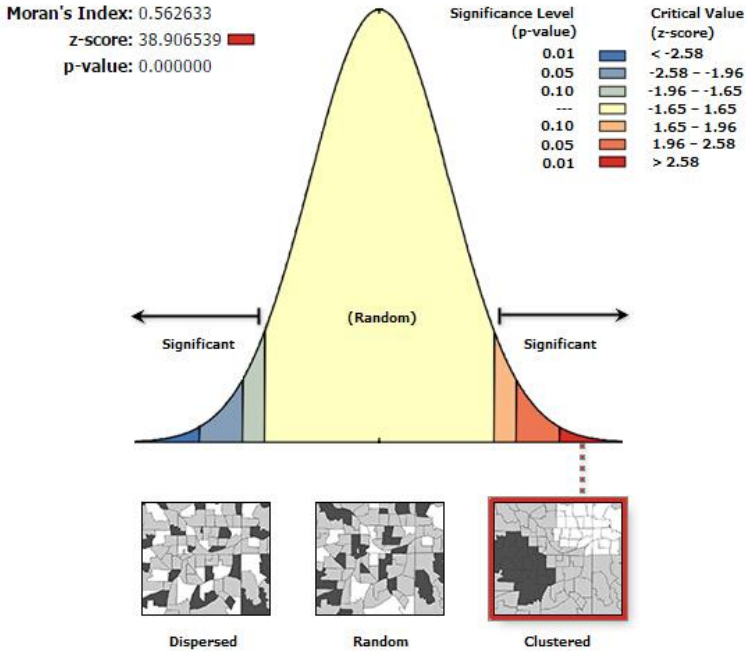


Figure 4.17 Global Moran's I result of the OLS standardized residuals.

Overall, the performance of the OLS regression and associated statistic tests proved that the specified model is able to explain part of the linear relationships between urban land distribution and the selected environmental factors. Development constraints of the Yangtze River Basin include the local elevation and slope, the proximity to the river channels, and the access the coastal zone. However, as expected, the residuals of the OLS

model violated assumptions and show significant spatial nonstationarity and spatial autocorrelation of the data. In this case, it is appropriate to set up a GWR model with the same dependent and explanatory variables.

4.4.3 GWR result

Different from a single set of constant values across space yielded from OLS, GWR produced a series of parameter estimates and model statistics at each sample point. Maps of the local R^2 and the coefficient for each explanatory variable obtained from the GWR model were created respectively (Figure 4.16 - Figure 4.18) to detect the spatial varying relationships between LogKDUrb and associated natural environmental factors.

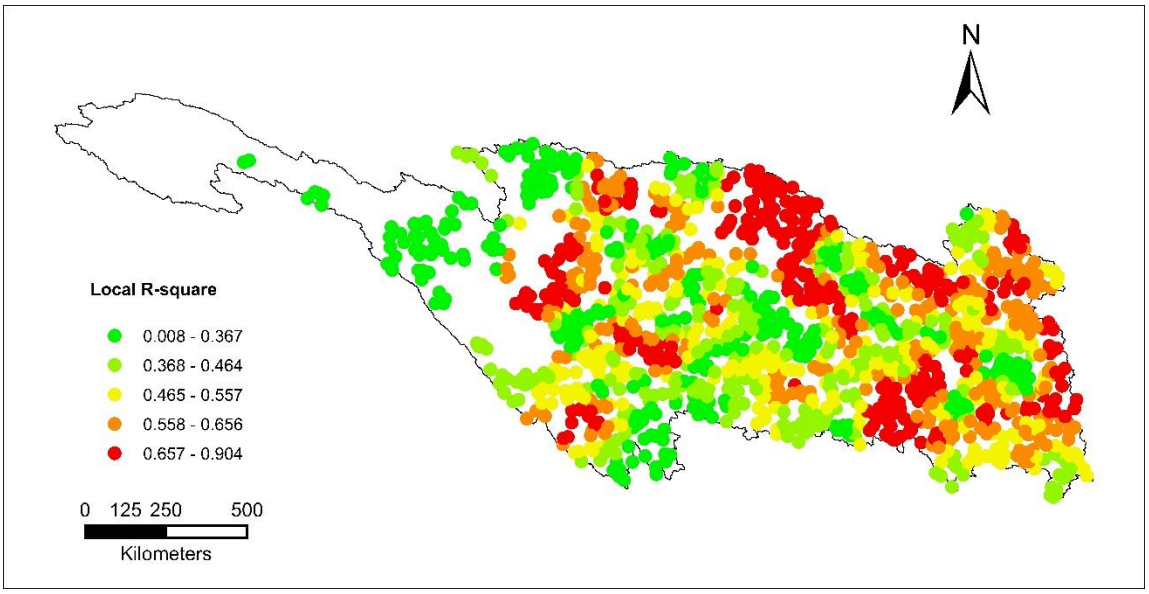


Figure 4.18 Spatial variation of local R-square from the GWR model.

Local R^2 is an important diagnostic indicates how well the local regression model fits the observation at each sample point. In the upper and middle reaches of the Yangtze River, regions with relatively large local R^2 values are buffers between mountainous areas and plains, including the outside circle of Sichuan Basin and the southwest side of Dabie Mountain. The flat lower reach expresses higher local R^2 in general than the western part, with some exceptions at the locations of several major lakes. Overall, 1276 out of 1511 sample points have local R^2 values that are larger than the global R^2 yielded from OLS. The proportion is more than 84%.

Cook's distance (Cook, 1977) is another important diagnostic statistic produced by GWR, showing a particular observation's influence. Usually, a Cook's distance value larger than one is of concern, indicating the associated observation is influential due to its large standardized residual, or significant leverage or both. According to the result of the GWR model in this research, all the Cook's distances are far less than one.

Maps of coefficients obtained from GWR model provide a simple way to detect the spatial varying relationships between urban land intensity and associated environmental factors. Owing to the spatial nonstationarity nature of the data used in this analysis, the GWR coefficients for all the four explanatory variables express significant sign change across the Yangtze River Basin. In other words, the coefficients are positive in part of the study area and negative in the rest part. Higher absolute values of coefficient indicate a greater influence of the particular explanatory variable on the dependent variable, whereas lower values mean lesser influence for the particular variable. The categories of coefficients in the following four maps are classified by their quantiles. Since

the change of sign substantially increases the complexity, the interpretation of GWR results mainly focuses on the local coefficients that share the same sign as the results of OLS.

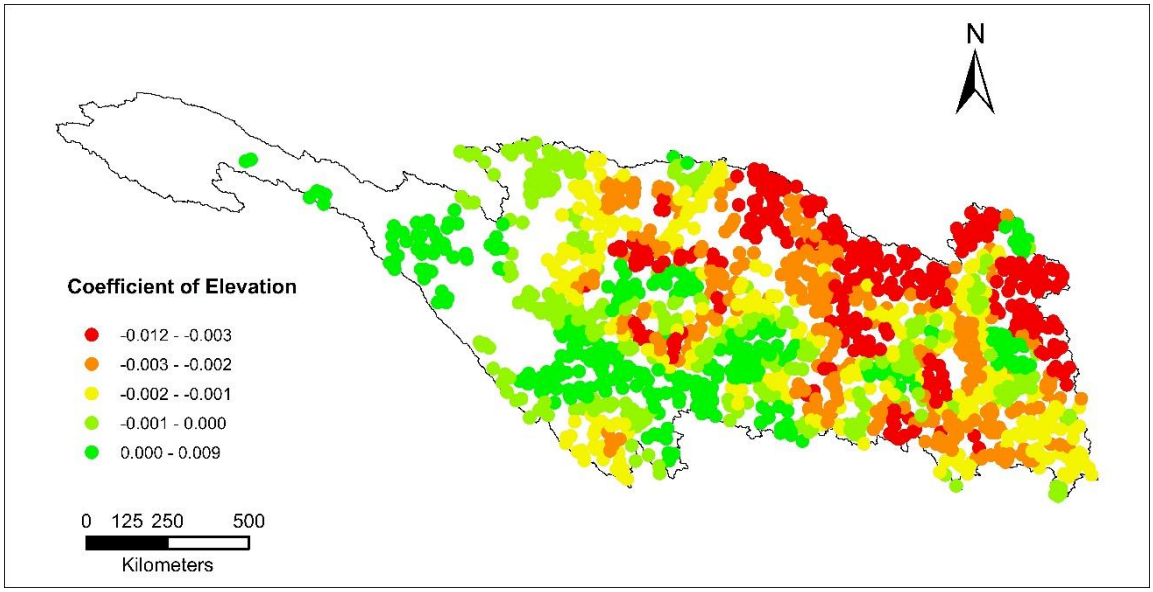


Figure 4.19 Coefficient of elevation factor yielded from the GWR model.

The elevation factor shows strong negative correlation with the dependent variable in most of the region downstream the Three Gorges Dam and in an orbit-shaped region around the Sichuan Basin (Figure 4.19). This kind of relationship means urban land are more likely to locate in an area with a lower elevation. It is interesting that in the broad western part from the headwater to Yibin, coefficients of elevation are positive. One potential cause of this phenomenon is, in the mountainous area, urban settlements are always located in plains that have higher elevations; while the river valleys, which are

low-elevated, are too narrow for development or exposure to natural hazards such as landslide and fluvial flood.

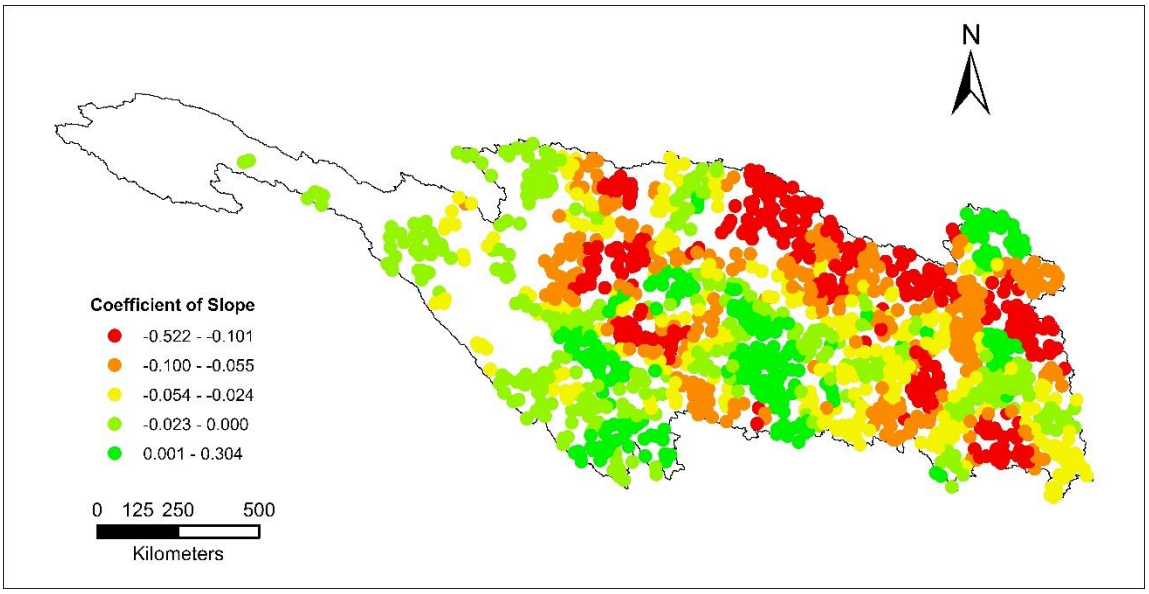


Figure 4.20 Coefficient of slope factor yielded from the GWR model.

The map of slope's coefficient (Figure 4.20) is similar to that of elevation's coefficient. Regions showing a significant negative correlation between slope and LogKDUrb include a band between 103E and 109E, the Sichuan Basin, the Jianghan Plain, and the Dongting Plain. These regions all have relatively low elevations.

The warm-colored points in Figure 4.21 reflect regions where kernel density of urban land increases as the effective distance to the coast decreases. It is noteworthy that such points are along several main tributaries of the Yangtze River. From left to the right of the map, red-point clusters depict the Min River, Wu River, Han River, Xiang River and Gan River, respectively. This found proves that although the land transportation has

been highly developed and become more popular, navigation of the Yangtze River still plays an important role in urban development, especially for some interior cities that are far from the coastal zone.

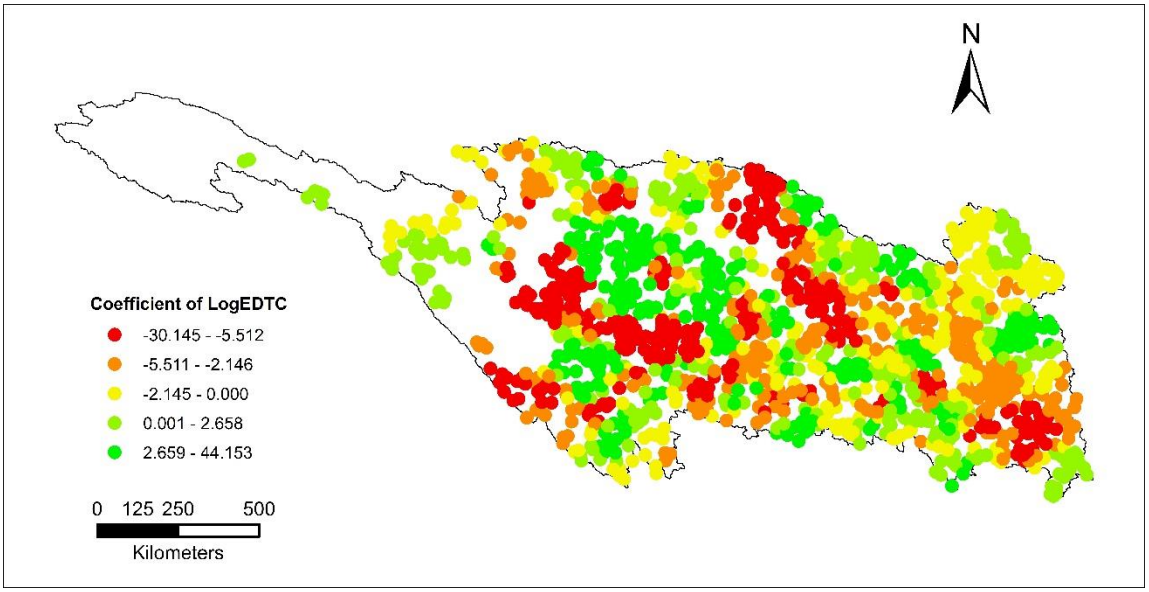


Figure 4.21 Coefficient of LogEDTC yielded from the GWR model.

Sample sites showing significant positive correlations between urban land distribution and proximity to stream network are all found adjacent to the basin's boundary (Figure 4.22). This pattern implies that in the drainage area of streams with low Strahler orders, proximity to the river channel imposes strong influence on urban location and development. However, as the Strahler order increases, the influence declined, probably because the mainstem with more discharge raises the risk of fluvial flood and weakens people's willing to reside nearby.

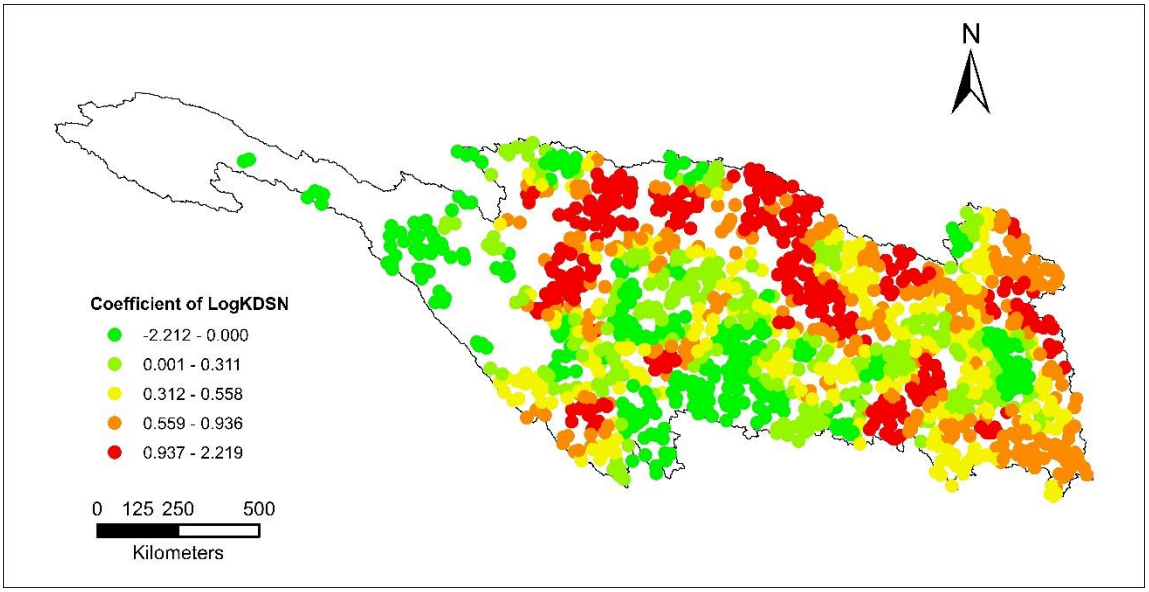


Figure 4.22 Coefficient of LogKDSN yielded from the GWR model.

The GWR coefficient maps illustrate how relationships between the dependent variable and each explanatory variable fluctuate geographically.

4.4.4 Comparisons between OLS and GWR models

4.4.4.1 Comparison of model performances

We compared values of AICc and Adjusted R2 from GWR with those from OLS to detect whether the GWR have better model performances. Table 4.8 lists the results of every explanatory variable and the combination of four. In all cases, the GWR models returned a larger adjusted R2 and a smaller AICc than the corresponding OLS models, so the local model outperformed the global model by allowing varying relationships across the Yangtze.

Table 4.8 Comparison of adjusted R^2 and AICc between OLS and GWR models

Variables	Adjusted R^2		AICc	
Elevation	R^2_O	0.268	AIC_O	5022.161
	R^2_G	0.632	AIC_G	4109.929
Slope	R^2_O	0.194	AIC_O	5168.455
	R^2_G	0.613	AIC_G	4191.485
LogEDTC	R^2_O	0.152	AIC_O	5246.112
	R^2_G	0.621	AIC_G	4119.208
LogKDSN	R^2_O	0.044	AIC_O	5426.623
	R^2_G	0.656	AIC_G	4024.207
All	R^2_O	0.331	AIC_O	4890.466
	R^2_G	0.698	AIC_G	4063.461

Symbols with the subscript "O" represent results of OLS models; symbols with the subscript "G" represent results of GWR models.

It is noteworthy that when turning to GWR models, the single explanatory variable with the largest adjusted R^2 changed from elevation factor to the LogKDSN, which means the indicator related to the proximity of stream network has the highest degree of spatial variance.

4.4.4.2 Comparison of spatial autocorrelations

Global Moran's I of standardized residuals from the OLS and GWR were computed to compare the two models' ability to deal with spatial autocorrelation issue. The results (Table 4.9) show that the GWR models produced smaller global Moran's I than OLS models with the same explanatory variable(s), indicating that GWR models improve the reliability of the relationships by reducing the spatial autocorrelations of residuals.

Table 4.9 Moran's I results of standardized residuals from OLS and GWR models

Explanatory variables	I_O	p-value	I_G	p-value
Elevation	0.581		0.207	
Slope	0.534		0.220	
LogEDTC	0.610	< 0.0001	0.257	< 0.0001
LogKDSN	0.655		0.236	
All	0.563		0.196	

Symbols with the subscript "O" represent results of OLS models; symbols with the subscript "G" represent results of GWR models.

4.5 Discussion

The Yangtze River Basin is a huge system. In this case, comprehensive plans of the basin's development target, strategy, and development models should be conducted. The planning should respect both the laws of nature and socioeconomic development. This paper investigated spatial relationships between urban land distribution and hydro-geomorphologic attributes by multilinear regression. The results can serve as a scientific basis and a decision-making references for sustainable development of the whole basin.

Three policy suggestions are proposed: (1) Implement different regional development policies according to various natural constraints. Meanwhile, resource-saving and environmental-friendly development modes deserve an attention with the first priority. For example, Chongqing will be built into an important pivot for opening up the western region of China based on the city's geographical, economical and political advantages. However, the vast landmass around Chongqing district is a relatively lagging region within China's interior. In this case, port logistics should be enhanced to connect

the “wild west” with the coastal zone along the Yangtze River. Moreover, the region around Chongqing is environmentally sensitive and fragile, problems such as water shortage, pollution, and ecological degradation may arise due to population growth followed by increasing resource consumption. Therefore, population migration and layout are reasonable options for local development. On the contrary, urban development in the middle reach of the Yangtze River should focus on maintaining water quality and reducing flood risk.

(2) Seek balance in the human-flood system in the urban area adjacent to streams. Since most of the existing high urban intensity regions are located in the river’s floodplains, successful flood management strategy is essential to reduce the risk of life and property loss. Apart from infrastructure construction, enhancing people’s awareness of socio-hydrology by education and by providing flood insurance is also indispensable. Unfortunately, as far as the present research was conducted, China does not have an adequate flood insurance system. So, the basin management concept needs to be updated. Accordingly, urban expansion and the configuration of new urban areas should be managed through spatial planning.

(3) Strengthen the communication and cooperation between different regions, in order to promote integrated basin development and environment protection. Apart from important ecological corridors, Yangtze River channels also serve as economic corridors. According to our spatial analysis on hot spots of KDUrb, regions with high urban intensity form a cross structure in the Yangtze River Basin (Figure 4.23). On the east-west axis, the river channels connect the urban/industrial zones from Chongqing-Chengdu district to

Nanjing. From north to south direction, the Han River and the Xiang River, as two main tributaries of the Yangtze, connect urban agglomerations around Wuhan and Changsha to the mainstem. Therefore, navigation functions of the Yangtze River stream network should be strengthened to promote the Yangtze River economic belt development strategy.

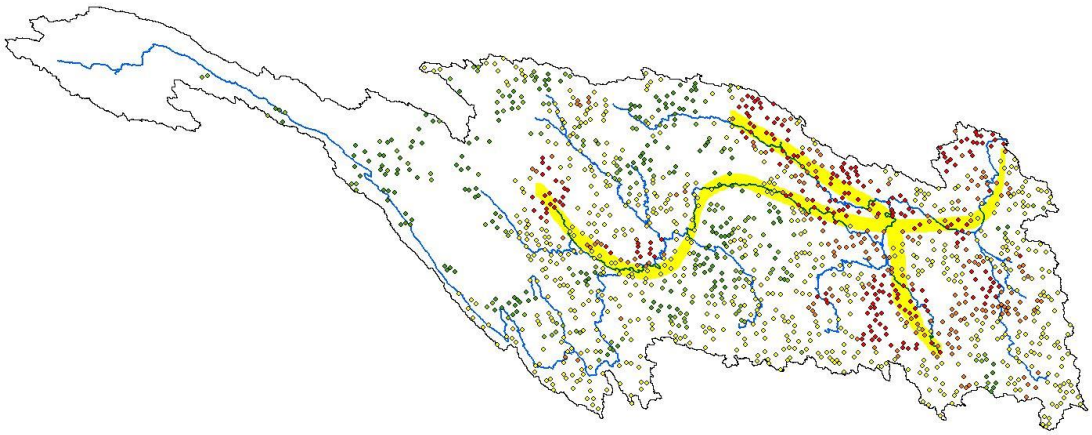


Figure 4.23 Structure of high urban intensity regions in the Yangtze River Basin

4.6 Conclusion

This study set up linear regression models to investigate the relationships between hydro-geomorphological attributes and urban land distribution in the Yangtze River Basin. The OLS and GWR methods were applied to build a global model and a local model, respectively. Although the determinants of urban land's location and distribution are complicated issues that rarely have simple answers, the result of the present research provides a new way to identify the influential natural factors. The result is also an

extension of the research described in Chapter III. To our knowledge, this study is the first reported to apply GWR in examining the impact of hydro-geomorphologic attributes on urban intensity using linear regression models, and it is the first time to employ the kernel density of stream network as an explanatory variable.

The result OLS model revealed that environmental constraints of urban development include the local elevation, slope, effective distance to the coast, and the proximity to stream networks. However, the residuals yielded from the global model violated the assumptions of constant variance and independence, indicating that a local regression method such as GWR was necessary to solve the spatial nonstationarity and spatial autocorrelation issues.

The GWR model made noticeable improvements in model performance over the OLS model, proved by the comparisons of the two model's corresponding adjusted R^2 and AICc values. The GWR model also improved the reliability of the linear relationships in three ways: 1) increasing the local predictive power of the regression; 2) reducing autocorrelation in the residuals; and 3) relaxing the assumption of stationary regression coefficients. The visualized GWR results of local R^2 and coefficients highlighted the great spatial variations in the impacts of hydro-geomorphologic attributes on the urban intensity.

The results of this study suggest that GWR technique can serve as a useful tool in socio-hydrological research at a river basin scale because GWR model can explain the spatial nonstationary relationships caused by the nature of both the environmental and anthropogenic factors.

CHAPTER V

THE IMPACT OF LAND CHANGE ON PRECIPITATION-RUNOFF GENERATION PROCESSES IN THE CHAO PHRAYA RIVER BASIN

5.1 Introduction

The transformation of land to provide more food and settlements for the growing population has been one of the most obvious anthropogenic impacts on the global environment. Land change mainly included the replacement of natural vegetation into crops, urban land, and secondary vegetation (e.g. regenerated grassland after deforestation).

The anthropogenic conversion of the Earth's land surface has occurred rapidly over recent decades and disrupted the hydrological cycle. From the perspective of social development, the alterations in precipitation-runoff processes, especially the streamflow patterns, have an inverse influence on human activities in terms of droughts and floods.

There is a large volume of published studies describing the linkages between land use/land cover (LULC) and streamflow. In the early stage, most of the research focused on controlled experimental manipulations involving small drainage areas (less than 1 hectare). Bosch and Hewlett (1982) reviewed 94 watersheds all over the world and summarized that deforestation will result in higher streamflow while declined streamflow always follows the increase in forest cover. In the same vein, Sahin and Hall (1996) applied fuzzy regression analysis on 145 catchments and concluded that the runoff was

increased by 10 mm/y after a 10% deforestation and by 213 mm/y if the forest was entirely removed.

During the last two decades, the development of remote sensing and Geographical Information System (GIS) techniques have made the large-scale analysis possible. Van der Weert (1994) compared streamflow of the 4133 km^2 Citarum River Basin in Indonesia between the periods 1922–1929 and 1979–1986. He pointed out that although the average annual precipitation stayed almost the same in the two time periods, the corresponding average streamflow increased 11% due to the expansion of cropland and human settlement. Later, Cheng (1999) claimed that in several river basins (25,500 – 66,625 km^2) of Southwest China, the abruptly increased annual discharge was caused by logging. Guo et al. (2008) conducted a study on the Poyang Lake Basin in the lower reach of the Yangtze River and concluded that the land change had a moderate impact on annual streamflow but a strong influence on seasonal streamflow.

More recent LULC research has focused on considering urbanization as a source of hydrological alterations (Paul and Meyer, 2001). Urbanization behaves as an extreme LULC mode showing as replacements of natural vegetation with impervious surfaces. Thus far, previous studies have identified five major effects of urbanization: (1) a higher ratio of precipitation turns to surface runoff, associated with declines in evapotranspiration and infiltration; (2) the lag time between precipitation and runoff generation is shortened; (3) the instream discharge always has a faster and larger peak flow; (4) the baseflow from groundwater to recharge streamflow decreases significantly; and (5) water quality is

degraded by effluent discharges and impose negative implications to biodiversity and ecosystem services (Shaw, 1994).

In this study, we examined streamflow responses to land change in the Chao Phraya River Basin, Thailand. The Soil and Water Assessment Tool (SWAT) was applied to build a hydrological model for this flood-prone area. SWAT was first developed in the early 1990s to assess water balance and pollutant loadings under various land use practices within agricultural watersheds (Neitsch et al. 2001). The model is semi-distributed, physically based, and computationally efficient. SWAT simulates the hydrological cycle in a watershed through a set of physical processes that can be grouped into a land phase and a channel routing phase. Additionally, outputs of daily, monthly, and yearly time steps are available for different temporal requirements. The development of SWAT as an extension of the ArcGIS software has increased its flexibility. The model's strength in estimating long-term impact of land-use change on streamflow also makes it particularly suitable for the present study.

Objectives of this study include: (1) set up a SWAT model of Chao Phraya Basin; (2) analyze the land change between the year 2000 and 2010 and evaluate the impact on instream discharge; (3) quantify the contribution of land change on the 2011 flood hazard that followed an insignificant precipitation increase.

5.2 Study area

The Chao Phraya River drains an approximate 1.5×10^5 km^2 watershed (Figure 5.1) from north to south in central Thailand, covering 30% of the country's territorial area.

Asian monsoon climate dominates the river basin, inducing distinct dry and wet seasons. In the dry season from November to April, precipitation is rare, and the average temperature is 18 ℃. On the contrary, more than 90% of the annual precipitation (~ 1200 mm/y) falls in the wet season when the average temperature reaches 25℃.

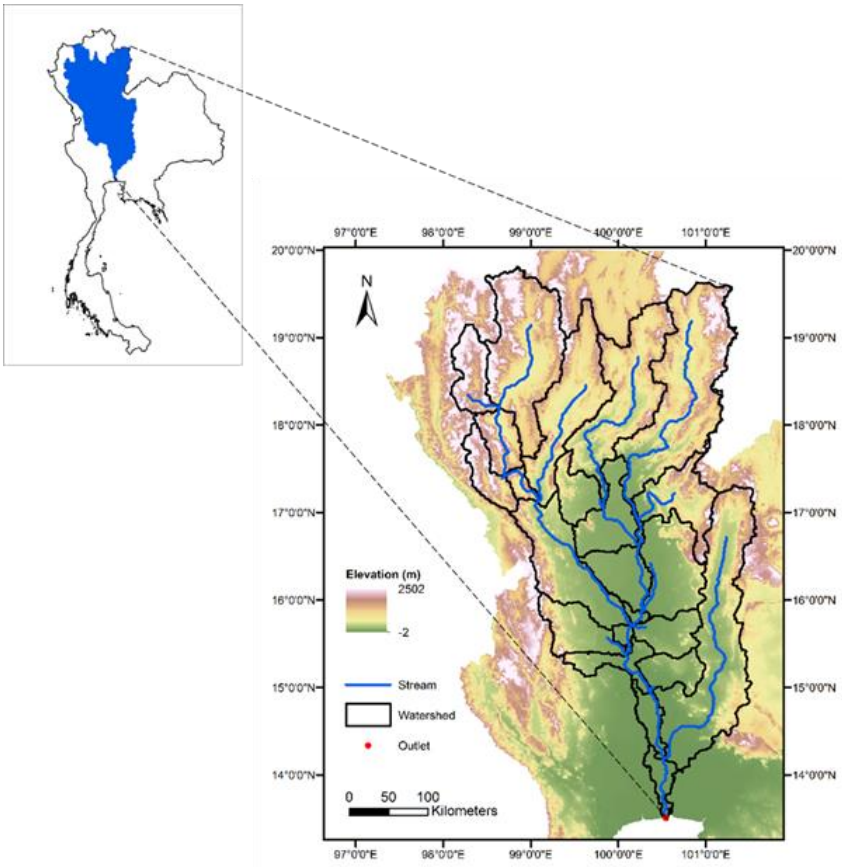


Figure 5.1 Location and delineated watershed of Chao Phraya Basin

The Chao Phraya River Basin is home to about 23 million population, or 40% of Thailand's citizens. Two cities in the study area have more than 100,000 residents: One is

Chiang Mai in northern Thailand, the other is the coastal mega-city Bangkok, which is the capital and the largest city of Thailand. A considerable spatial variation exists in rural and urban population distribution in the river basin. Within the upper reach, urban residents make up less than 10% of the population, while in the lower reach, about 55% people live in cities. Farmland occupies half of the river basin. The main crops are rice, sugarcane, and corn.

The Chao Phraya River Basin is no stranger to flooding. As the monsoon is the main contributor of precipitation in this region, tropical storms also enhance precipitation during the wet season. In recent years, flood hazards occurred more frequently even though the precipitation did not significantly increase. The flood in 2011 caused about USD 45.7 billion economic damage to Bangkok and surrounding areas (World Bank, 2012). However, the total precipitation from May to October in 2011 was only 400 mm larger than the average monsoon season precipitation. Experts highlighted that the unprecedented flood following an insignificant increase in precipitation was mainly caused by the decline of flood retention areas due to the rapid urbanization and the intensification of agriculture.

5.3 Materials and methods

SWAT requires a set of input data including a digital elevation model (DEM), LULC maps, soil properties, and meteorological data, respectively. All of the digital data have to share the same projected coordinate system. To minimize distortion, the Indian_1975_UTM_Zone_47N projected coordinates was selected.

SWAT first delineates the watershed boundary and stream networks based on the DEM, and then each depicted sub-basin is further divided into several hydrologic response units (HRUs) based on land use, soil, and slope characteristics. Next, historical meteorological data are input as forcing of the hydrological cycle. At last, water fluxes in both the land phrase and the channel routing phase are simulated. Detailed implement procedures and corresponding data pre-processing are described as follows.

5.3.1 Topography and watershed delineation

The DEM used in this study was obtained from EarthEnv-DEM90 (Robinson et al., 2014). With an aim to match other data layers, the DEM was resampled from the original three arc second (~ 90m) resolution to 1000 m resolution. Then, an automatic watershed delineation process was conducted. The critical support area threshold was set as 3000 km^2 and yielded 24 sub-basins for the Chao Phraya Basin.

5.3.2 Land cover and land use

The LULC maps of the year 2000 and 2010 were derived by merging two datasets: the Collection 5 MODIS Land Cover Type Product (MCD12Q1) and an urban expansion map provided by the World Bank (Schneider, et al., 2009). The MCD12Q1 dataset contains spatially aggregated LULC information for each year in the period 2001–2012 and identifies 17 land cover classes. However, the time series of LULC shows a decreasing trend in the urban land area, which is contrary to the fact. Since regional hydrological processes are highly sensitive to urban expansion, the World Bank urban map showing

conditions in the year 2000 and 2010 were utilized as baselines to correct the MODIS product. Consequently, two new LULC layers were synthesized at a resolution of 500 m. The new layers inherited attributes from the two original datasets and represented the c2000 and c2010 situations. Later, the layer showing 2000 conditions was used to set up the hydrological model while the 2010 LULC layer was applied in the scenario analysis.

Table 5.1 Pairing of LULC classes from MODIS to SWAT

MODIS LULC class*	SWAT LULC	SWAT code
Water	Water	WATR
Evergreen Needleleaf Forest	Forest-Evergreen	FRSE
Evergreen Broadleaf Forest		
Deciduous Broadleaf Forest	Forest-Deciduous	FRSD
Mixed Forest	Forest-Mixed	FRST
Closed Shrublands	Range-Brush	RNGB
Open Shrublands		
Woody Savannas		
Savannas	Range-Grasses	RNGE
Grasslands		
Permanent wetlands	Wetlands-Mixed	WETL
Croplands	Agricultural Land-Row Crops	AGRR
Cropland/Natural Vegetation Mosaic		
Urban and Built-up	Residential-High Density	URHD
Barran or Sparsely Vegetated	Barren	BARR

*The snow and ice class does not exist in the Chao Phraya River Basin.

Then, the MODIS LULC classes were converted to the SWAT classes (Table 5.1), and the LULC layer was reorganized with a new attribute table. The resulting map (Figure 5.2) shows that forests dominate the upper reach, grassland and agriculture land are

prevalent in the middle reach, and urban land agglomerations locate in the lower reach, especially in the coastal zone. In terms of areal coverage, the study area mainly consists of agricultural land, evergreen forest, and shrubland around the year 2000.

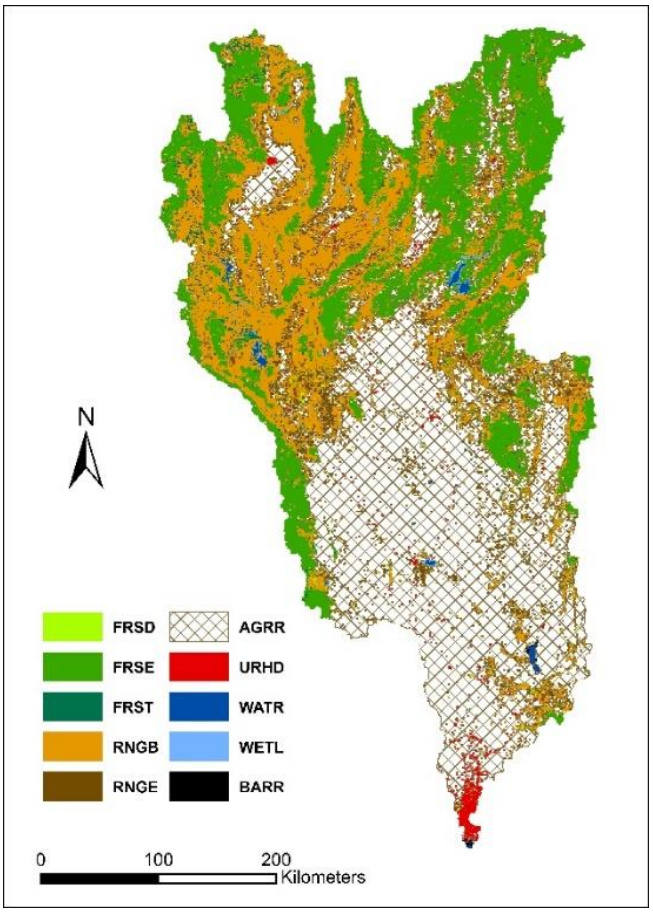


Figure 5.2 LULC condition in 2000 based on the SWAT LULC class system.

5.3.3 Soil properties

Since the default soil data embedded in the SWAT model is only fit for the U.S., a new soil attribute dataset for Chao Phraya Basin was created and appended to the existing SWAT database.

The soil data preparation was completed based on the Harmonized World Soil Database (HWSD). The HWSD provides a 30 arc-second raster map, which links to harmonized parameters in terms of soil units. The soil classification is in accordance with 1:5 000 000 scale Soil Map of the World (FAO, 19711981), and the soil parameter data were developed by the World Inventory of Soil Emission Potential (WISE) project. Figure 5.3 illustrates the 11 soil types and their distribution in the Chao Phraya River Basin. However, some of the required soil properties are not directly provided, calling for a series of extra calculations.

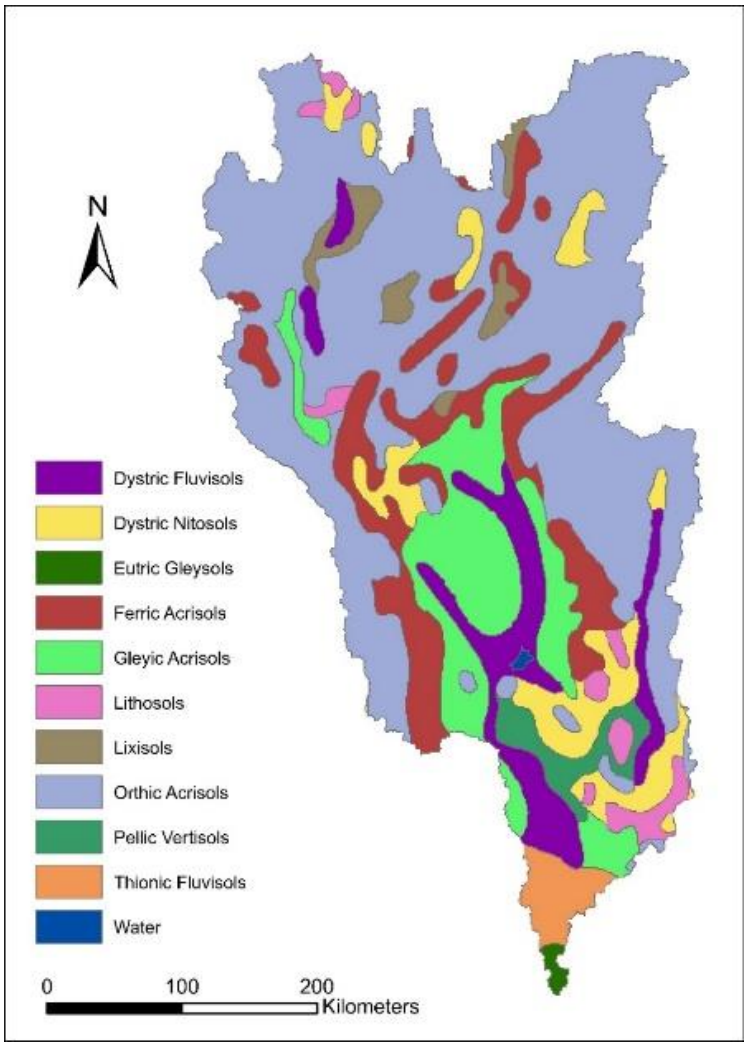


Figure 5.3 Soil map of the Chao Phraya River Basin

According to Table 5.2, five required soil parameters are not provided by the HWSD. Among these unknown parameters, the available water capacity (SOL_AWC) and saturated hydraulic conductivity (SOL_K) can be calculated based on each soil's texture and organic carbon content, using the Soil Water Characteristics module of the Soil SPAW-Plant-Air-Water (SPAW) Program. Soil albedo is the ratio of the incident

radiation that is reflected from the ground surface. It can be calculated from a linear regression equation:

$$SOL\_ALB = 0.2227 \times e^{-1.8672 \times SOL\_CBN} \quad (5\text{-}1)$$

Table 5.2 SWAT required soil attributes

Variable	Description	Unit	Offered by HWSD
SOL_ZMX	Maximum rooting depth of soil profile	mm	Yes
SOL_Z	Depth from soil surface to bottom of layer	mm	Yes
SOL_D	Moist bulk density	g/cm^3	Yes
SOL_CBN	Organic carbon content	% soil weight	Yes
SOL_CLAY/SILT/SAND	Clay/Silt/Sand content	% soil weight	Yes
SOL_ROCK	Rock fragment content	% total weight	Yes
SOL_AWC	Available water capacity of the soil layer	mm H_2O/mm soil	No
SOL_K	Saturated hydraulic conductivity	mm/hr	No
HYDGRP	Soil Hydraulic Group	--	No
SOL_ALB	Moist soil albedo	--	No
USLE_K	Universal Soil Loss Equation soil erodibility (K) factor	--	No

The USLE_K is a soil erodibility factor, which is sensitive to soil's texture, the organic carbon content, soil structure, and soil permeability. Williams (1995) proposed an equation to calculate the soil erodibility factor (*K*):

$$K = f_{csand} \cdot f_{cl-si} \cdot f_{orgc} \cdot f_{hisand} \quad (5\text{-}2)$$

The four factors in Equation (5-2) are calculated as:

$$f_{csand} = 0.2 + 0.3 \times exp\left[-0.256 \times \text{SOL\_SAND} \times \left(1 - \frac{\text{SOL\_SILT}}{100}\right)\right] \quad (5\text{-}3)$$

$$f_{cl-si} = \left(\frac{\text{SOL\_SILT}}{\text{SOL\_CLAY} + \text{SOL\_SILT}}\right)^{0.3} \quad (5\text{-}4)$$

$$f_{orgc} = 1 - \frac{0.0256 \times \text{SOL\_CBN}}{\text{SOL\_CBN} + exp(3.72 - 2.95 \times \text{SOL\_CBN})} \quad (5\text{-}5)$$

$$f_{hisand} = 1 - \frac{0.7 \times \left(1 - \frac{\text{SOL\_SAND}}{100}\right)}{\left(1 - \frac{\text{SOL\_SAND}}{100}\right) + exp\left[-5.51 + 22.9 \times \left(1 - \frac{\text{SOL\_SAND}}{100}\right)\right]} \quad (5\text{-}6)$$

In the SWAT model, soils are classified into four hydrologic groups (A, B, C, and D) based on infiltration characteristics. Soils in group A have the lowest runoff potential caused by their high infiltration rates. Then, each following group has a lower average infiltration rate and corresponding higher precipitation-runoff ratio. For example, the infiltration rate of Group B is lower than Group A. Detailed classification criteria are provided by the U.S. Natural Resource Conservation Service (NRCS, 1996) Reorganized soil properties and each soil type's area ratio were listed in Table 5.3. The required attributes of soils in the Chao Phraya River Basin were appended to the SWAT "usersoil" database.

Table 5.3 Soil types and properties in the Chao Phraya Basin

Soil type	Depth of soil layer	Composition			Rock fragment (%)	Organic carbon (%)	Bulk density (g/cm^3)	Available water capacity (mm/mm)	Saturated hydraulic conductivity (mm/hr)	Moist soil albedo	K factor	Hydraulic Group	Areal coverage (%)
		Clay (%)	Silt (%)	Sand (%)									
Ferric Acrisols	1000	10	12	78	4	0.61	1.58	0.03	53.09	0.07	0.13	A	12.30%
Gleyic Acrisols		23	43	34	9	0.83	1.38	0.14	7.87	0.05	0.18	B	11.71%
Orthic Acrisols		24	27	49	10	1	1.4	0.11	9.14	0.04	0.17	B	52.44%
Eutric Gleysols		52	31	17	7	1.25	1.22	0.12	1.52	0.02	0.16	D	0.30%
Lithosols		23	34	43	26	1.4	1.39	0.11	9.40	0.02	0.17	C	2.50%
Dystric Fluvisols		20	41	39	4	0.9	1.41	0.14	11.68	0.04	0.18	A	7.43%
Lixisols		45	19	36	4	0.99	1.28	0.12	1.02	0.04	0.14	D	2.34%
Thionic Fluvisols		58	34	8	4	2.59	1.19	0.10	3.30	0.00	0.21	B	1.94%
Dystric Nitosols		55	23	22	1	1.32	1.22	0.12	0.76	0.02	0.14	C	6.81%
Pellic Vertisols		56	24	20	2	1.2	1.22	0.12	0.76	0.03	0.14	D	2.14%
Water		0	0	0	0	0	0	0.00	0.00	0.23	0.01	D	0.09%

5.3.4 Meteorological data

Meteorological data provide the moisture and energy inputs that control the water balance in a river basin. Required meteorological input files in the SWAT model include daily maximum and minimum temperatures, precipitation, solar radiation, relative humidity, and wind speed. To set up the hydrological model for the Chao Phraya Basin, daily temperature, and precipitation data at ten stations were obtained from the Global Historical Climatology Network (GHCN) database (Figure 5.4). These two daily observed data are available from the year 1950 to 2016 with an interruption in 1999.

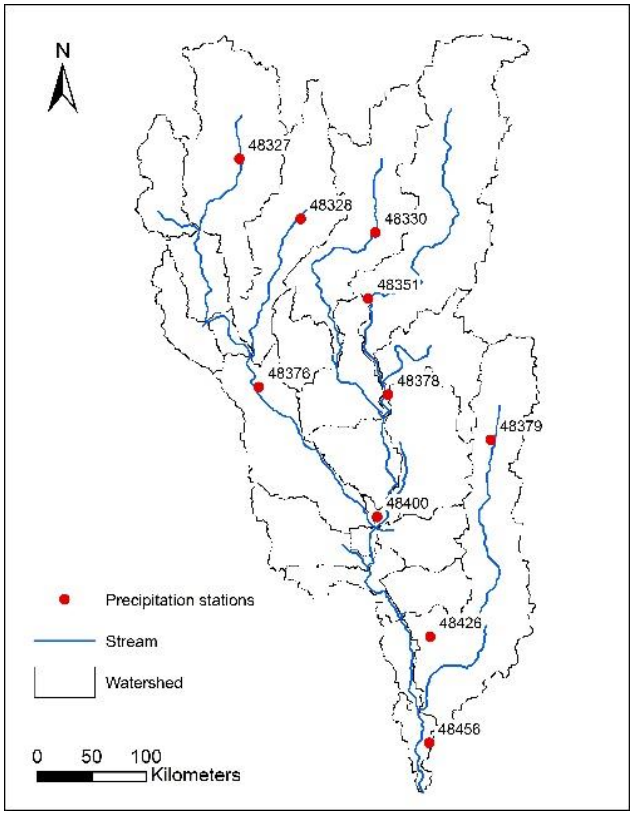


Figure 5.4 Map of precipitation stations in the Chao Phraya River Basin

Since the lack of long-term measured solar radiation, relative humidity, and wind speed data for the Chao Phraya Basin, an alternative database provided by the National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR) was employed. The CFSR data are available for each hour from 1979 to 2014 at a 38-km resolution. For the Chao Phraya River Basin, 129 data points were selected, and the associated meteorological data were aggregated to daily (Figure 5.5).

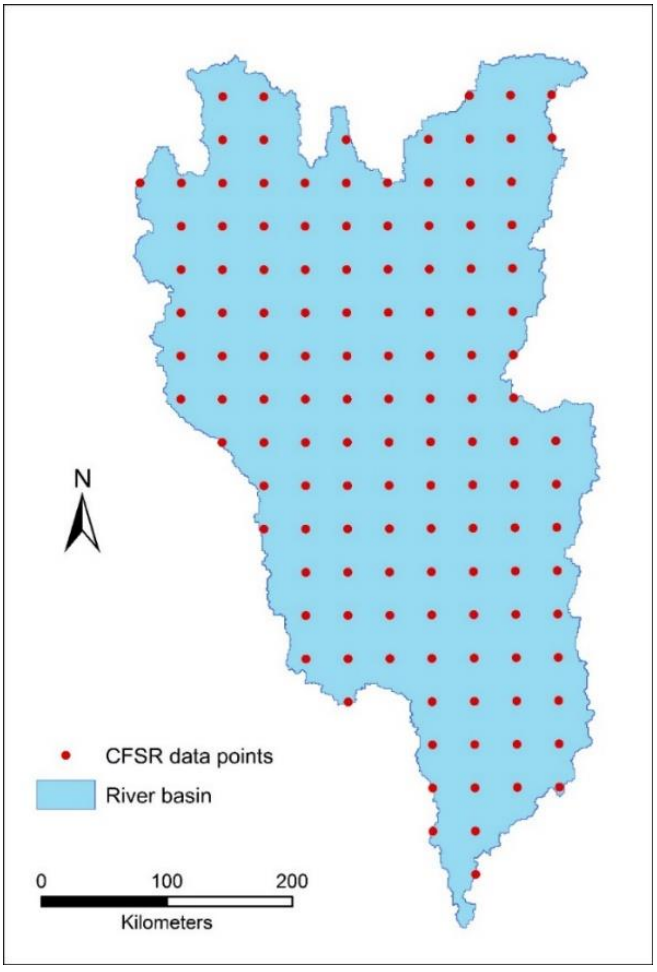


Figure 5.5 Map showing the CFSR data points selected for the Chao Phraya Basin

5.3.5 Streamflow discharge data

The measured daily streamflow discharge data were obtained from the GAME-T2 Data Center to evaluate the SWAT model by calibration and validation. Observed streamflow at the Nakhon Sawan Station (drainage area: 110,569km^2) was used, because this hydrological station is located at the junction of the Ping and Nan Rivers (Figure 5.6), with daily runoff records from 1955 to 2000. No other downstream hydrological stations provide such a long-term observation data.

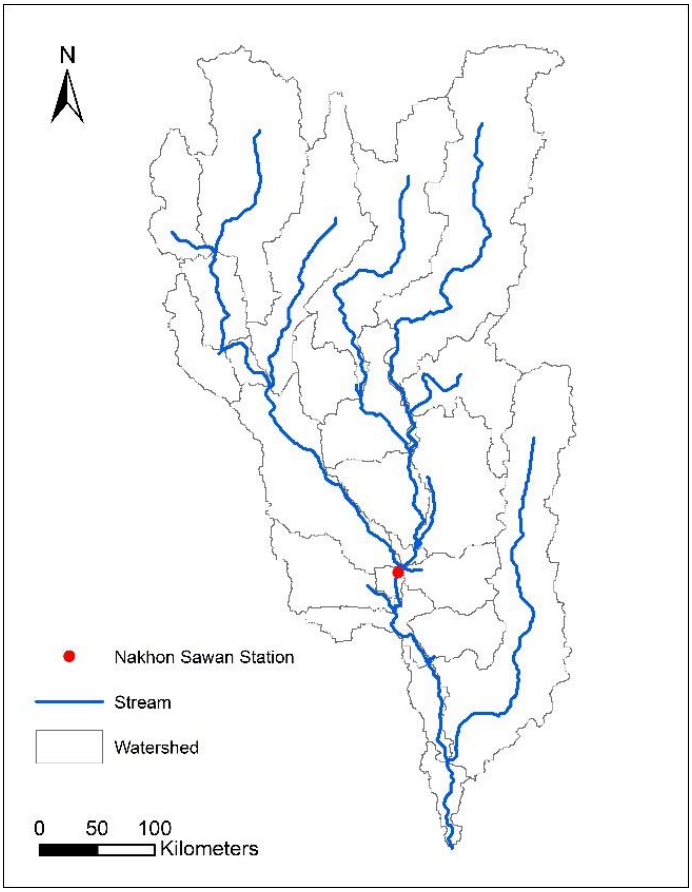


Figure 5.6 Location of the Nakhon Sawan gauge station

5.3.6 SWAT model setup

After inputting LULC and soil property files successfully, each depicted sub-basin was further divided into several hydrologic response units (HRUs). HRU is the smallest spatial unit in the SWAT model. Within a sub-basin, each HRU has the homogeneous LULC, soil, and slope according to user-defined thresholds. For the sake of HRU generation, two slope classes (0 to 1%, and greater than 1%) were created, and the defined thresholds for land use, soil, and slope were 10%, 5%, and 10%, respectively. Then, the SWAT model yielded 222 HRUs in the Chao Phraya Basin.

For each HRU, SWAT simulates the water movement during a precipitation event and estimates the hydrological components separately. As precipitation descends, an amount of raindrops/snows may be intercepted by the vegetation canopy instead of reaching the soil. The amount of water held by canopy is a function of the density of plant cover and the morphology of plant species defined by the leaf area index (LAI). Then part of the water falling on the ground will infiltrate into the soil profile. When the field capacity of a soil layer is exceeded and the layer below is not saturated, the percolation occurs downward through a soil profile, and the process may deliver soil water to the aquifer. Simultaneously, an inflow takes place within the unsaturated soil profile in the lateral direction.

When precipitation rate is larger than infiltration rate, the surface runoff will occur along a sloping surface. Runoff moves relatively quickly toward a stream channel and contributes to short-term stream response. SWAT computes surface runoff volumes by a modified USDA Soil Conservation Service (SCS) curve number method (USDA Soil

Conservation Service, 1972). SCS precipitation-runoff relationship is an empirical function described as:

$$Q_{\text{surf}} = \frac{R_{\text{day}} - I_{\text{a}}}{R_{\text{day}} - I_{\text{a}} + S} \tag{5-7}$$

where Q_{surf} is the accumulated runoff or excess rain water (mm), R_{day} represents the daily rainfall (mm), I_{a} the initial abstractions which include surface storage, the interception, and infiltration prior to runoff (mm), and S is the retention parameter (mm). The retention parameter varies spatially due to changes in topography, soils, land-use, and management. It also varies temporally related to changes in soil water content. This parameter is defined as

$$S = 25.4\left(\frac{1000}{CN} - 10\right) \tag{5-8}$$

where *CN* is the non-dimensional curve number, which is determined by soil types and land use.

Evapotranspiration (ET) is the primary water loss mechanism in a river basin. ET includes evaporation from water bodies, soil, and canopy interception; and transpiration through plants. In this research, actual ET was calculated based on the method developed by Ritchie (1972), and potential ET was derived from the Penman-Monteith method.

Once SWAT determined the water loading entered the channel, the loading was routed through the stream network. As water flows downstream, a portion may be lost due to evaporation. On the other hand, base flow originating from groundwater may contribute to streams and increase the discharge. Another supplement comes from the

precipitation that is directly falling on the stream network. In this research, a variable storage coefficient method developed by Williams (1969) was applied to complete the flow routing.

5.3.7 Model calibration and validation

The overall 20-year (1979-1998) estimation period was split into a three-year (Jan. 1979 to Dec. 1981) warming up section, a 14-year (Jan. 1982 to Dec. 1995) calibration section, and a three-year (Jan. 1996 to Dec 1998) validation section. The 2000 LULC was applied in both the calibration and validation because earlier reliable LULC data was unavailable.

Model parameters were calibrated based on the average monthly discharge at the Nakhon Sawan Station. The calibration and validation processes were implemented by the Sequential Uncertainty Fitting (SUFI-2) algorithm (Abbaspour et al., 2004, 2007) in the SWAT-CUP software. Both pair-wise Goodness-of-fit (GOF) indicators and graphical techniques were referenced to compare the monthly averaged simulated and observed discharges. Three primary GOF indicators were used in this analysis: Coefficient of determination (R^2), Nash-Sutcliffe efficiency (NSE), and percent bias (PBIAS). Among these three indicators, R^2 is a standard regression statistic determining the strength of a linear relationship; NSE is a dimensionless statistic, and RSR and PBIAS are error indices quantifying the deviation between observed and measured values in the interested data's unit. Thus, the three indicators cover all categories of quantitative statistics and are

comprehensive enough to represent the model's performance. Description of each indicator appears subsequently.

In the calibration and validation processes, the coefficient of determination (R^2) shows the degree of the linear relationship between simulated and measured streamflow. The indicator can be calculated as:

$$R^2 = \frac{\left[\sum_i (Q_{m,i} - \bar{Q}_m)(Q_{s,i} - \bar{Q}_s)\right]^2}{\sum_i (Q_{m,i} - \bar{Q}_m)^2 \sum_i (Q_{s,i} - \bar{Q}_s)^2} \quad (5\text{-}9)$$

where Q_m and Q_s stand for measured and simulated discharges, respectively, and the bar stands for average value. *i* is the total number of observations.

R^2 ranges from 0 to 1. A higher R^2 value indicates stronger linear relationship and less error variance. Typically, an R^2 value greater than 0.5 is acceptable (Santhi et al., 2001, Van Liew et al., 2003).

The Nash-Sutcliffe efficiency (NSE) determines the relative magnitude of the residual variance compared to the measured data variance (Nash and Sutcliffe, 1970). This indicator is considered as the best objective function for reflecting the overall fit of a hydrograph Sevat and Dezetter (1991). NSE is computed as:

$$NSE = 1 - \frac{\sum_i (Q_m - Q_s)_i^2}{\sum_i (Q_{m,i} - \bar{Q}_m)^2} \quad (5\text{-}10)$$

NSE ranges between negative infinity and one, and the value one indicates the optimal condition. Values between 0.0 and 1.0 are viewed as an acceptable performance, while values less than 0 shows that the mean observed value is a better predictor than the simulated value, which indicates unacceptable performance.

Percent bias (PBIAS) implies the model's accurate in simulation by calculating the average tendency of the differences between simulated data and their observed counterparts (Gupta et al., 1999). PBIAS is expressed as:

$$PBIAS = 100 \times \frac{\sum_{i=1}^{n}(Q_m - Q_s)_i}{\sum_{i=1}^{n} Q_{m,i}} \quad (5\text{-}11)$$

A positive PBIAS indicates underestimation, which means the estimated values are less than the measured values in general. A negative value of PBIAS indicates model overestimation bias. Model performance can be judged based on general performance ratings (Table 5.4).

Table 5.4 General performance ratings for recommended statistics for a monthly time step

Performance Rating	R^2	NSE	PBIAS(%)
Very good	$R^2 \geq 0.90$	$0.75 \leq NSE \leq 1.00$	$\lvert PBIAS \rvert < 10$
Good	$0.6 < R^2 \leq 0.9$	$0.65 < NSE \leq 0.75$	$10 \leq \lvert PBIAS \rvert \leq 15$
Satisfactory	$0.4 < R^2 \leq 0.6$	$0.50 < NSE \leq 0.65$	$15 \leq \lvert PBIAS \rvert \leq 25$
Unsatisfactory	$R^2 \leq 0.4$	$NSE \leq 0.50$	$\lvert PBIAS \rvert \geq 25$

While the quantitative statistics figure out the best estimation of the parameter set (Table 5.5), the graphical technique in the SUFI2 method takes parameter's uncertainty into consideration. And the uncertainty is quantified as the 95% prediction uncertainty (95PPU) calculated at the 2.5% and 97.5% levels of the cumulative distribution function of the output variables. Latin hypercube sampling is used to draw independent parameter sets (Abbaspour et al., 2007). Two indices, P-factor and R-factor, are associated with the

95PPU. P-factor is the percentage of observed data enveloped by the 95PPU, and R-factor is the thickness. For discharge, a P-factor larger than 70% with an R-factor around one is satisfactory.

Sensitive analysis of the selected parameters is also conducted during the calibration process. Parameter's global sensitivity is determined by the ordinary least square method, and a t-test is then used to identify the relative significance of each parameter.

Table 5.5 SWAT parameters related to discharge for model calibration

Parameter*	Description	Initial input range	
		Min	Max
r__CN2.mgt	Soil Conversion Service (SCS) runoff curve number for moisture condition 2	-0.6	0.6
r__HRU_SLP.hru	Average slope steepness (m/m)	-0.6	0.6
r__OV_N.hru	Manning's "n" value for overland flow	-0.6	0.6
r__SLSUBBSN.hru	Average slope length (m)	-0.6	0.6
r__SOL_AWC(1).sol	Available water capacity of the soil layer (mm H_2O/mm soil)	-0.6	0.6
r__SOL_BD(1).sol	Moist bulk density (g/cm^3)	-0.6	0.6
r__SOL_K(1).sol	Saturated hydraulic conductivity (mm/hr)	-0.6	0.6
v__ALPHA_BF.gw	Baseflow alpha factor (1/day)	0	1
v__ALPHA_BNK.rte	Baseflow alpha factor for bank storage (days)	0	1
v__CANMX.hru	Maximum canopy index	0	100
v__CH_K2.rte	Effective hydraulic conductivity in main channel alluvium (mm/h)	0.01	500
v__CH_N2.rte	Manning's "n" value for the main channel	0	0.5
v__EPCO.hru	Threshold depth of water in the shallow aquifer to percolation to the deep aquifer (mm H_2O)	0	1
v__ESCO.hru	Soil evaporation compensation factor	0	1
v__GW_DELAY.gw	Groundwater delay time (days)	30	500
v__GW_REVAP.gw	Groundwater “revap” coefficient	0.02	0.2
v__GWQMN.gw	Threshold depth of water in the shallow aquifer required for return flow to occur (mm)	0	500
v__REVAPMN.gw	Threshold depth of water in the shallow aquifer for percolation to the deep aquifer (mm H_2O)	0	500
v__SURLAG.bsn	Surface runoff lag coefficient	0.05	24

*“a_”,“v_” and “r_” means an absolute increase, replacement, and a relative change in the initial parameter values, respective

5.4 Results

5.4.1 Calibrated parameters

The top nine parameters with sensitivity indices (p-value) less than 0.05 are shown in Table 5.6 with the optimal values. The smaller the p-value, the more sensitive the parameter. Among all the sensitive parameters, CN2 and CANMX determine the amount of water that turns into runoff; ESCO controls the evaporation process; the rest factors are all related to groundwater. After updating the associated input files, the SWAT model was calibrated using the optimized parameters.

Table 5.6 Nine calibrated parameters with the highest sensitivities

Parameter	Optimal value	P-Value of the t-test for sensitivity
r__CN2.mgt	-0.15	0.00
v__GW_DELAY.gw	459	0.00
v__ESCO.hru	0.27	0.00
r__SOL_AWC(..).sol	-0.27	0.00
v__REVAPMN.gw	109	0.00
v__GWQMN.gw	108	0.00
v__ALPHA_BNK.rte	0.61	0.00
v__GW_REVAP.gw	0.20	0.00
v__CANMX.hru	100	0.04

5.4.2 Model performance evaluation

Hydrographs of the simulated and observed average monthly streamflow (Figure 5.7) show the model performance in calibration and validation periods, respectively. As mentioned before, the first three years of estimation (1979 – 1981) were used as the model's spin up time, so the results for this time period were not displayed.

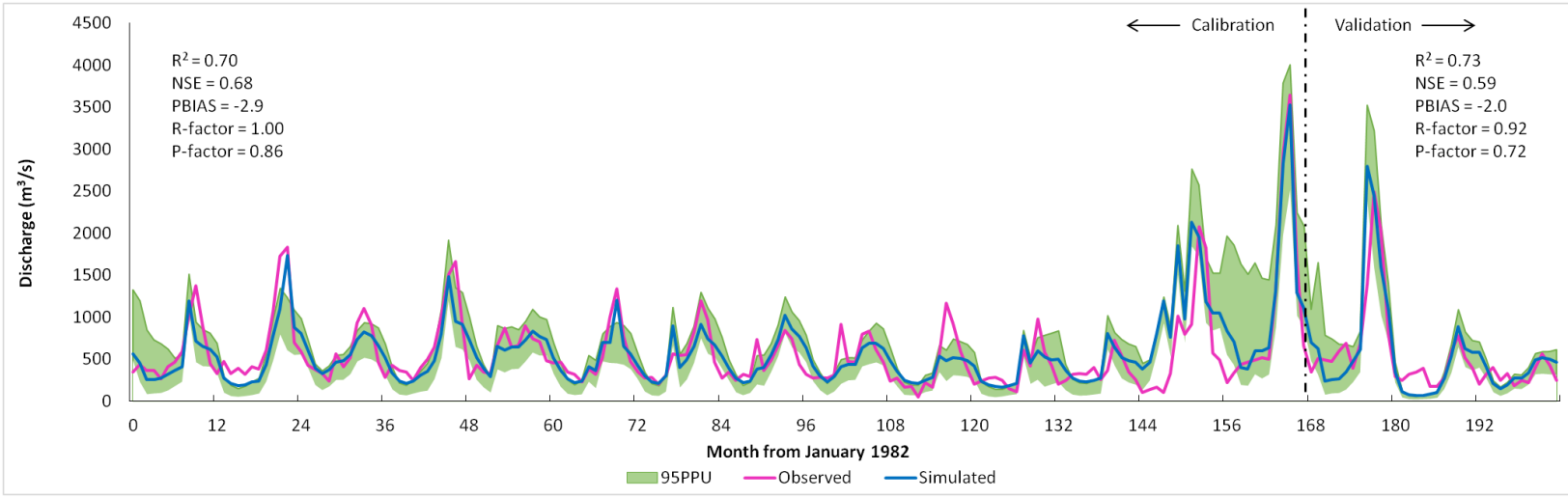


Figure 5.7 Observed (pink line) and SWAT model simulated streamflow (blue line) of the Chao Phraya River Basin for the calibration period of 1982–1995 and the validation period of 1996–1998.

In the calibration period, the SWAT model reproduced the peaks and recession curves of the observed data efficiently from 1982 to 1989 and successfully replicated historical flood events in the year 1983 and 1995. However, the estimation results do not match the observation data well from the year 1990 to 1994. Underestimates of streamflow are evident from 1990 to 1992. Major discrepancies between observed and simulated streamflow occur in the wet seasons, when none of the three peak flows were estimated accurately. On the contrary, overestimate takes place throughout the year of 1994. The validation results from 1997 to 1999 produced NSE = 0.59. This slight decrease of the accuracy, compared with the calibration period, is reflected primarily in the dry seasons of these three years.

Rational explanations of all the discrepancies include 1) some management processes in the river basin are unknown or unaccountable due to the lack of information, for example, the dam operation, water transfer, and irrigation; 2) the quantity of input precipitation data is not guaranteed.

As shown in Figure 5.7, most of the observation data are bracket by the 95PPU (86% during the calibration period and 72% during the validation period), indicating SUFI-2 is capable of capturing the observations during both calibration and validation periods. The 95PPU is quite suitable to bracket the observations in 1985, 1988 and 1989, while it is somehow slightly overestimated in 1986, 1987 and 1990, especially in the recession part.

Given the spatial scale of the study area and the quality of original input data, in particular the meteorological data, the calibrated model performed satisfactorily. The

calibration and validation results ensured that the SWAT model with optimized parameters was accurate enough for analyzing the responses of streamflow to land change.

5.4.3 Scenario analysis

The scenario analysis was designed to estimate the effect of land change on streamflow in the Chao Phraya River Basin. To quantify the effect, two LULC datasets were applied to reflect the conditions in 2000 and 2010 respectively, while the meteorological conditions in the year 2011, which caused the worst flood in the basin, stayed the same when running the calibrated SWAT model.

Table 5.7 Areal coverage change of each LULC class between the year 2000 and 2010

SWAT LULC	**Areal coverage**		**Change in area (km^2)**	**Relative change (%)**
	2000 condition (km^2)	**2010 condition (km^2)**		
AGRR	66028	67772	1744	2.64
FRSE	34511	25310	-9201	-26.66
RNGB	33870	45621	11751	34.69
RNGE	7174	966	-6208	-86.53
FRST	2279	3160	881	38.66
URHD	1477	2432	955	64.66
Other	1318	1396	78	5.92

Table 5.7 shows the land change between the year 2000 and 2010. During the decade, extensive loss of evergreen forest coincided with the shrubland expansion in the upper reach. Meanwhile, a significant area of grassland also degraded into shrubland.

Urban development, which mainly took place on the periphery of Chiang Mai and Bangkok, replaced substantial areas of grassland and cropland into impervious surface.

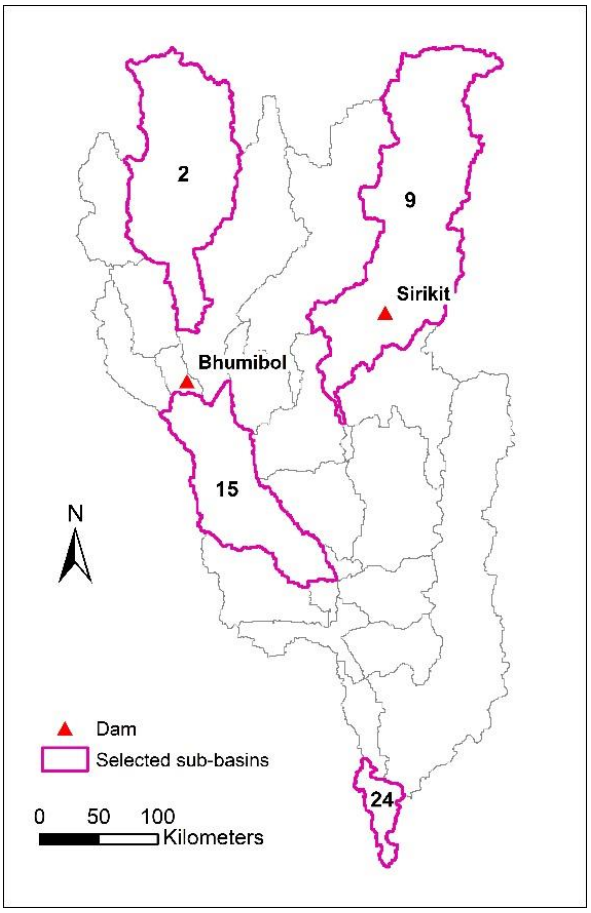


Figure 5.8 Four selected sub-basins to represent results of scenario analysis

Four sub-basins were selected to represent the streamflow responses under typical land change from 2000 to 2010 (Figure 5.8). Sub-basin 2 covers the upper catchment of the Ping River, which is a major tributary of the Chao Phraya. This 1.5×10^4 km^2 sub-basin also contains Chiang Mai city, the second largest city within the study area. The Bhumibol Dam is about 70.5 km downstream the sub-basin's outlet. According to Table 5.8 urban land in sub-basin 2 doubled its size from 2000 to 2010, with a decrease of 155 km^2

cropland. At the same time, about 300 km^2 grassland degraded to shrubland; while more than 70 km^2 grassland turned into forest.

Table 5.8 Land change matrix from 2000 to 2010 in sub-basin 2

LULC in 2010 / LULC in 2000	AGRR	FRSD	FRSE	FRST	RNGB	RNGE	URHD	WETL
AGRR	2296	1	39	4	359	18	155	0
FRSD	7	1	1	0	15	0	0	0
FRSE	30	0	3868	86	361	1	0	2
FRST	27	1	123	58	328	2	0	1
RNGB	399	16	554	402	4771	31	1	2
RNGE	280	2	3	4	293	18	0	0
URHD	0	0	0	0	0	0	97	0
WETL	0	0	8	0	2	0	0	12

Sub-basin 9 is the drainage area of the Nan River, another major upstream tributary of the Chao Phraya. The Sirikit Dam is located in this region. Comparison of LULC conditions in 2000 and 2010 (Table 5.9) presents a significant conversion from forest to shrubland. The change was mainly caused by deforestation due to the prosperity of logging industry.

Table 5.9 Land change matrix from 2000 to 2010 in sub-basin 9

LULC in 2010 / LULC in 2000	AGRR	FRSD	FRSE	FRST	RNGB	RNGE	URHD	WATR	WETL
AGRR	3016	11	199	27	1241	34	61	0	2
FRSD	2	1	1	1	11	0	0	0	1
FRSE	381	21	6599	305	4212	1	0	0	8
FRST	13	0	42	35	113	0	0	0	3
RNGB	646	9	218	72	2759	21	1	0	4
RNGE	425	1	16	5	293	16	0	0	1
URHD	0	0	0	0	0	0	58	0	0
WATR	0	0	0	3	1	0	0	124	19
WETL	0	0	5	37	22	0	0	3	89

Sub-basin 15 is a transition zone from mountainous region to the low-flat plains. Dominant land change type in this area is the degradation of grassland to shrubland (Table 5.10)

Table 5.10 Land change matrix from 2000 to 2010 in sub-basin 15

LULC in 2010 / LULC in 2000	AGRR	FRSD	FRSE	FRST	RNGB	RNGE	URHD	WETL
AGRR	4892	8	37	15	185	37	48	14
FRSD	16	1	1	5	46	5	0	0
FRSE	46	4	1215	62	338	2	0	1
FRST	29	1	4	3	51	1	0	0
RNGB	258	39	22	98	1296	66	0	3
RNGE	420	8	2	16	267	73	0	0
URHD	0	0	0	0	0	0	64	0
WETL	0	0	0	0	0	0	0	0

Sub-basin 24 contains the outlet of the entire Chao Phraya River and part of the Bangkok city. Because of the advantages of geographical location, such as adequate

freshwater resources, flat topography, and proximity to coast, this sub-basin is more developed than the rest region in the study area. Transitions from cropland to urban land are the dominant land change (Table 5.11).

Table 5.11 Land change matrix from 2000 to 2010 in sub-basin 24

LULC in 2010 / LULC in 2000	AGRR	BARR	FRSD	RNGB	RNGE	URHD	WATR	WETL
AGRR	792	0	3	2	2	126	0	12
BARR	0	4	0	0	0	0	5	13
FRSD	3	0	0	0	0	1	0	0
RNGB	14	0	0	1	1	1	0	4
RNGE	10	2	1	1	0	3	0	11
URHD	0	0	0	0	0	661	0	0
WATR	0	0	0	0	0	1	12	0
WETL	2	2	0	1	0	1	0	6

Two sets of average monthly streamflow values yielded from the model simulation for each sub-basin were used in scenario analysis (Table 5.12). By comparing the two time series, one can conclude that the average monthly streamflow from all four sub-basins decreased about 2.5% in the dry season from 2000 to 2010. On the contrary, more than 5% increases in averaged monthly streamflow occurred during the wet season.

Table 5.12 Streamflow responses to land change in four selected sub-basins

No. of sub-basin	Average monthly discharge in 2000		Average monthly discharge in 2010		% change in average discharge	
	Dry season	Wet season	Dry season	Wet season	Dry season	Wet season
2	50.41	208.89	49.24	219.43	-2.33	5.04
9	89.91	424.53	87.28	444.73	-2.93	4.76
15	210.57	848.18	204.47	900.48	-2.9	6.17
24	616.74	2980.33	600.70	3149.83	-2.6	5.69

5.5 Discussion and conclusion

This study was carried out to analyze hydrological responses to land change in the Chao Phraya River Basin, Thailand. In order to accomplish this objective, simulations with different LULC conditions of the year 2000 and 2010 were presented by using the SWAT model. The typical climatic data in 2011, which caused the worst catastrophic flood in Thailand's history, were used as the meteorological input of the model.

Four sub-basins with different dominant land change types and spatial characteristics were selected to represent the whole river basin. According to the results, predictions can be made that in the near future, the risk of drought will be higher during the dry season in the study area, especially in some upstream sub-basins that receive less benefit from dam operations. Meanwhile, devastating floods will take place more frequently at the lower reach of the river basin in the wet season, even when the return period of precipitation is short.

The increasing streamflow caused by land change in upstream tributaries like the Ping River and the Nan River will add pressure on the two main dams. At the same time,

rapid urbanization and the intensification of agriculture in lower Chao Phraya River Basin will further increase the exposure and the vulnerability towards floods. Since the limited discharge capacity through the city of Bangkok in the low elevation coastal zone has created a bottleneck for the upstream flood drainage, losses from flood will continue to rise.

Short-term engineering flood protection approaches, such as building dams and dikes, may mitigate water shortage and reduce the risk of high-frequent floods. But the built constructions also make people over-confident and thus more vulnerable to floods with long return periods. Acceleration of the hydrological cycle, subsidence caused by groundwater overdraw, and sea-level rise resulting from climate change could worsen the problem. In this circumstance, the existing city area needs to be redesigned to make space for water so that the natural processes of flooding occur with minimized damage. This could be done through "green engineering" solutions and more integrated flood management. Meanwhile, models in socio-hydrology should be developed to further the understanding of water in a cultural, social, economic and political sense.

As future research based on this study, more detailed analysis of flood in the lower reach of Chao Phraya River Basin should be implemented, considering management approaches such as reservoir operations and irrigation. Floodplain inundation will be simulated then by combining hydrologic and hydraulic model.

CHAPTER VI

CONCLUSION

This research investigates the interactions between natural environment and humans based on global, continental, and river basin scales, respectively. A probabilistic forecast of global urban expansion suggests that even without considering the climate change, the urban areas exposed to flood and drought hazards will increase considerably due to the spread of urban extents, particularly in Asia and Africa.

The spatial analysis of urban land and population distributions in 11 selected major river basins in Asia and Africa demonstrates that the existing cities and current population are more concentrated in coastal zones and across floodplains of streams that have higher Strahler orders. The results also indicate that although the dependence of urban development on natural resources has decreased with the growing technology, natural environment factors still play important roles when people determine their living locations.

To further scrutinize the environmental influences on urban land distribution, a case study on the Yangtze River Basin is performed by means of linear regression models. Treating elevation, slope, kernel density of stream network, and effective distance to the coast as four explanatory variables, results from both the Ordinary Least Squares and Geographically Weighted Regression models show that urban agglomerations are highly associated with selected environmental factors.

Finally, a hydrological model of the Chao Phraya River Basin, where large-scale floods are a recurring phenomenon, is set up using SWAT. Scenario analysis indicates that

regardless of water management, the land change from 2000 to 2010 increased the risk of both floods and droughts in the river basin based on the current climatic characteristics.

Interactions between humans and the environment are occurring on a scale that has never previously been seen. Increased urbanization, together with population growth and increased standard of living, often significantly affect the hydrological system. Spiraling demands for resources, such as freshwater, are significantly impacting riverine and coastal areas. Meanwhile, the affected natural environment, especially the altering hydrological processes due to land change, makes people more exposed to natural hazards such as floods and droughts.

It is against this background that studies in this dissertation contribute to our understanding of environmental constraints on urban development, but of how hydrological processes will respond to human-induced land change. These findings can be used to communicate on decision making and hazard prevention effectively to resource managers, planners, government, and most important, to the public.

The limitations of the analyses in this dissertation are as follows: First, all the analyses are based on the current climate conditions. The influence of climate change is not taken into account. Although climate change is projected to result in rising sea levels and increasing frequency and intensity of extreme events in many parts of the world, my research period only spans 30 years (i.e. from 2000 to 2030), and it is reasonable to assume that the near future climate will not significantly change from the current conditions. Second, delineated watersheds do not include coastal zones created by tidal channels. In this case, some big cities located in the coastal area, like Shanghai in China, are not

contained in the study area even though, it has been widely accepted that most of the large cities of Asia and Africa are clustering in coastal zones. The exclusion of these urban centers may lead to a bias in the spatial distribution of urban land and population in the Basins analyzed. Third, analyses about fresh water availability are mainly based on surface water, while the quantity of groundwater is not considered.

REFERENCES

Al-Ansari, N. A. (2013). Management of Water Resources in Iraq: Perspectives and Prognoses. *Engineering*, *5*(08), 667.

Arino, O., Gross, D., Ranera, F., Leroy, M., Bicheron, P., et al. (2007). GlobCover: ESA service for Global land cover from MERIS. *Igarss: 2007 Ieee International Geoscience and Remote Sensing Symposium, Vols 1-12*, 2412-2415.

Arnold, C. L. J., & Gibbons, J. C. (1996). Impervious surface coverage: the emergence of a key environmental indicator. *Journal of the American Planning Association, 62*(2), 243-258.

Büdel, J. (1966). Deltas: a basis of culture and civilization. *Scientific Problems of the Humid Zone Deltas and their Implications: Proceedings of the Dacca Symposium.*, 295-300.

Babel, M. S., & Wahid, S. M. (2008). Freshwater under threat: South Asia.*United Nations Environment Program, Nairobi, Kenya*.

Bartholomé, E., & Belward, A. (2005). GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing, 26*(9), 1959-1977.

Bilskie, M. V., Hagen, S. C., Medeiros, S. C., & Passeri, D. L. (2014). Dynamics of sea level rise and coastal flooding on a changing landscape. *Geophysical Research Letters*. doi:10.1002/2013gl058759

Black, R., Adger, W. N., Arnell, N. W., Dercon, S., Geddes, A., et al. (2011). Migration and global environmental change. *Global Environmental Change, 21, Supplement 1*(0), S1-S2. doi:10.1016/j.gloenvcha.2011.10.005

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika, 71*(2), 353-360.

Bozkurt, D., & Sen, O. L. (2013). Climate change impacts in the Euphrates-Tigris Basin based on different model and scenario simulations. *Journal of Hydrology, 480*, 149-161.

Brettell, C. B., & Hollifield, J. F. (2014). Migration theory: Talking across disciplines: Routledge.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. (1998a). Spatial nonstationarity and autoregressive models. *Environment and Planning A, 30*(6), 957-973.

Brunsdon, C., Fotheringham, S., & Charlton, M. (1998b). Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician), 47*(3), 431-443.

Ceola, S., Laio, F., & Montanari, A. (2015). Human‐impacted waters: New perspectives from global high‐resolution monitoring. *Water Resources Research, 51*(9), 7064-7079.

Cohen, B. (2006). Urbanization in developing countries: Current trends, future projections, and key challenges for sustainability. *Technology in society, 28*(1), 63-80.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics, 19*(1), 15-18.

Czech, B. (2004). Urbanization as a threat to biodiversity: Trophic theory, economic geography, and implications for conservation land acquisition. *Policies for Managing Urban Growth and Landscape Change: A Key to Conservation in the 21st Century*, *265*, 8-13.

Döll, P., Kaspar, F., & Lehner, B. (2003). A global hydrological model for deriving water availability indicators: Model tuning and validation. *Journal of Hydrology, 270*(1-2), 105-134.

Dai, A. (2011). Drought under global warming: a review. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(1), 45-65.

Daniere, A., Takahashi, L., NaRanong, A., & Lan, V. T. N. (2005). Social capital and urban environments in Southeast Asia: Lessons from settlements in Bangkok and Ho Chi Minh City. *International Development Planning Review, 27*(1), 21-58.

Danko, D. M. (1992). The digital chart of the world project. *Photogrammetric Engineering and Remote Sensing, 58*(8), 1125-1128.

Dawson, R. J., Ball, T., Werritty, J., Werritty, A., Hall, J. W., et al. (2011). Assessing the effectiveness of non-structural flood management measures in the Thames Estuary under conditions of socio-economic and environmental change. *Global Environmental Change, 21*(2), 628-646.

de Moel, H., Aerts, J. C. J. H., & Koomen, E. (2011). Development of flood exposure in the Netherlands during the 20th and 21st century. *Global Environmental Change, 21*(2), 620-627.

de Sherbinin, A., Schiller, A., & Pulsipher, A. (2007). The vulnerability of global cities to climate hazards. *Environment and Urbanization, 19*(1), 39-64.

Dartmouth Flood Observatory (2014). Space-based Measurement and Modeling of Surface Water for Research, Humanitarian, and Water Management Applications. *Online at http://floodobservatory.colorado.edu/* (accessed 01.01.14)

Djokic, D., Ye, Z., & Miller, A. (1997). *Efficient watershed delineation using ArcView and spatial analyst.* Paper presented at the Proceedings, 17th Annual ESRI User Conference, San Diego, CA.

Ecker, M. D., De Oliveira, V., & Isakson, H. (2013). A note on a non-stationary point source spatial model. *Environmental and Ecological Statistics, 20*(1), 59-67.

Elsaeed, G. (2012). Effects of Climate Change on Egypt's Water Supply. In H. J. S. Fernando, Z. Klaić, & J. L. McCulley (Eds.), *National Security and Human Health Implications of Climate Change* (pp. 337-347): Springer Netherlands.

Elvidge, C. D., Tuttle, B. T., Sutton, P. S., Baugh, K. E., Howard, A. T., et al. (2007). Global distribution and density of constructed impervious surfaces. *Sensors, 7*(9), 1962-1979.

Fernández-Armesto, F. (2001). *Civilizations: culture, ambition, and the transformation of nature*: Simon and Schuster.

Flintrop, C., Hohlmann, B., Jasper, T., Korte, C., Podlaha, O. G., et al. (1996). Anatomy of pollution: Rivers of north Rhine-Westphalia, Germany. *American Journal of Science, 296*(1), 58-98.

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., et al. (2005). Global consequences of land use. *Science, 309*(5734), 570-574.

Foody, G. M. (2003). Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI-rainfall relationship. *Remote Sensing of Environment, 88*(3), 283-293. doi:10.1016/j.rse.2003.08.004

Foresight, U. K. (2011). Foresight: Migration and Global Environmental Change: Final Project Report. *The Government Office for Science, London*.

Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A, 30*(11), 1905-1927. doi:10.1068/A301905

Güneralp, B., & Seto, K. C. (2008). Environmental impacts of urban growth from an integrated dynamic perspective: A case study of Shenzhen, South China. *Global Environmental Change, 18*, 720–735.

Gao, J. B., & Li, S. C. (2011). Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using Geographically Weighted Regression. *Applied Geography, 31*(1), 292-302. doi:10.1016/j.apgeog.2010.06.003

Gerten, D., Lucht, W., Ostberg, S., Heinke, J., Kowarsch, M., et al. (2013). Asynchronous exposure to global warming: freshwater resources and terrestrial ecosystems. *Environmental Research Letters, 8*(3), 034032.

Gesch, D., Oimoen, M., Zhang, Z., Meyer, D., & Danielson, J. (2012). Validation of the Aster Global Digital Elevation Model Version 2 over the Conterminous United States. *Xxii Isprs Congress, Technical Commission Iv, 39-B4*, 281-286.

Gleick, P. H., & Palaniappan, M. (2010). Peak water limits to freshwater withdrawal and use. *Proceedings of the National Academy of Sciences*, -. doi:10.1073/pnas.1004812107

Gober, P. (2010). Desert urbanization and the challenges of water sustainability. *Current Opinion in Environmental Sustainability, 2*(3), 144-150.

Gregersen, I. B., Madsen, H., Rosbjerg, D., & Arnbjerg‐Nielsen, K. (2013). A spatial and nonstationary model for the frequency of extreme rainfall events. *Water Resources Research, 49*(1), 127-136.

Grimm, N. B., Grove, J. G., Pickett, S. T., & Redman, C. L. (2000). Integrated Approaches to Long-Term Studies of Urban Ecological Systems Urban ecological systems present multiple challenges to ecologists—pervasive human impact and extreme heterogeneity of cities, and the need to integrate social and ecological approaches, concepts, and theory. *Bioscience, 50*(7), 571-584.

Güneralp, B., & Seto, K. C. (2013). Futures of global urban expansion: uncertainties and implications for biodiversity conservation. *Environmental Research Letters, 8*, 014025.

Gustafson, E. J. (1998). Quantifying landscape spatial pattern: what is the state of the art? *Ecosystems, 1*(2), 143-156.

Hallegatte, S., Green, C., Nicholls, R. J., & Corfee-Morlot, J. (2013). Future flood losses in major coastal cities. *Nature Clim. Change, 3*(9), 802-806. doi:10.1038/nclimate1979

Hanham, R., Hoch, R. J., & Spiker, J. S. (2009). The Spatially Varying Relationship Between Local Land-Use Policies and Urban Growth: A Geographically Weighted Regression Analysis. *Planning and Socioeconomic Applications, 1*, 43-56. doi:10.1007/978-1-4020-9642-6_4

Hanson, S., Nicholls, R., Ranger, N., Hallegatte, S., Corfee-Morlot, J., et al. (2011). A global ranking of port cities with high exposure to climate extremes. *Climatic Change, 104*(1), 89-111.

Harhay, M. O. (2011). Water stress and water scarcity: A global problem. *American Journal of Public Health, 101*(8), 1348-1349.

Hinrichsen, D. (1999). The coastal population explosion. In *Trends and Future Challenges for US National Ocean and Coastal Policy: Proceedings of a Workshop* (Vol. 22, pp. 27-29). Washington, DC: NOAA, January 22, 1999.

Horton, R. E. (1945). Erosional Development of Streams and Their Drainage Basins - Hydrophysical Approach to Quantitative Morphology. *Geological Society of America Bulletin, 56*(3), 275-370. doi:10.1130/0016-7606(1945)56[275:Edosat]2.0.Co;2

Hugo, G., & Bardsley, D. K. (2014). Migration and environmental change in Asia *People on the Move in a Changing Climate* (pp. 21-48): Springer.

Hylton, S. (2007). *The Grand Experiment: The Birth of the Railway Age, 1820-1845*: Ian Allan.

Jaimes, N. B. P., Sendra, J. B., Delgado, M. G., & Plata, R. F. (2010). Exploring the driving forces behind deforestation in the state of Mexico (Mexico) using geographically weighted regression. *Applied Geography, 30*(4), 576-591. doi:10.1016/j.apgeog.2010.05.004

Jarvis, A., Reuter, H., Nelson, A., & Guevara, E. (2008). Hole-filled seamless SRTM data V4, International Centre for tropical Agriculture (CIAT).

Jha, A. K., Bloch, R., & Lamond, J. (2012). *Cities and flooding: a guide to integrated urban flood risk management for the 21st century*: World Bank Publications.

Jones, B., & O'Neill, B. C. (2013). Historically grounded spatial population projections for the continental United States. *Environmental Research Letters, 8*(4), 044021.

Jongman, B., Ward, P. J., & Aerts, J. C. J. H. (2012). Global exposure to river and coastal flooding: Long term trends and changes. *Global Environmental Change, 22*(4), 823-835. doi:10.1016/j.gloenvcha.2012.07.004

Kirchner, J. W. (1993). Statistical inevitability of Horton's laws and the apparent randomness of stream channel networks. *Geology, 21*(7), 591-594.

Klein, R. J. T., Nicholls, R. J., & Thomalla, F. (2003). The resilience of coastal megacities to weather-related hazards: a review. In A. Kreimer, Arnold, M.,

Carlin, A. (Ed.), *Building Safer Cities: The Future of Disaster Risk* (pp. 101-120). Washington, DC, USA: World Bank.

Kloog, I., Haim, A., & Portnov, B. A. (2009). Using kernel density function as an urban analysis tool: Investigating the association between nightlight exposure and the incidence of breast cancer in Haifa, Israel. *Computers Environment and Urban Systems, 33*(1), 55-63. doi:10.1016/j.compenvurbsys.2008.09.006

Kuwairi, A. (2006). Water mining: The Great man-made River, Libya. *Proceedings of the Institution of Civil Engineers: Civil Engineering, 159*(1 SPEC. ISS.), 39-43.

Legendre, P., & Legendre, L. (1998). Numerical Ecology, Volume 24, (Developments in Environmental Modelling).

Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes, 27*(15), 2171-2186. doi:10.1002/hyp.9740

Li, S. C., Zhao, Z. Q., Xie, M. M., & Wang, Y. L. (2010). Investigating spatial non-stationary and scale-dependent relationships between urban surface temperature and environmental factors using geographically weighted regression. *Environmental Modelling & Software, 25*(12), 1789-1800. doi:10.1016/j.envsoft.2010.06.011

Li, Y. C., Liu, C. X., Zhang, H., & Gao, X. (2011). Evaluation on the human settlements environment suitability in the Three Gorges Reservoir Area of Chongqing based on RS and GIS. *Journal of Geographical Sciences, 21*(2), 346-358. doi:10.1007/s11442-011-0849-2

Liu, F., Yuan, L., Yang, Q., Ou, S., Xie, L., et al. (2014). Hydrological responses to the combined influence of diverse human activities in the Pearl River delta, China. *CATENA, 113*(0), 41-55. doi:10.1016/j.catena.2013.09.003

Liu, J. Y., Zhan, J. Y., & Deng, X. Z. (2005). Spatio-temporal patterns and driving forces of urban land expansion in china during the economic reform era. *Ambio, 34*(6), 450-455. doi:10.1639/0044-7447(2005)034[0450:Spadfo]2.0.Co;2

Maestre, F. T., Salguero-Gómez, R., & Quero, J. L. (2012). It is getting hotter in here: Determining and projecting the impacts of global environmental change on drylands. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1606), 3062-3075.

McDonald, R. I. (2008). Global urbanization: can ecologists identify a sustainable way forward? *Frontiers in Ecology and the Environment, 6*(2), 99-104.

McDonald, R. I., Douglas, I., Revenga, C., Hale, R., Grimm, N., et al. (2011). Global Urban Growth and the Geography of Water Availability, Quality, and Delivery. *AMBIO: A Journal of the Human Environment, 40*(5), 437-446. doi:10.1007/s13280-011-0152-6

McDonald, R. I., Green, P., Balk, D., Fekete, B. M., Revenga, C., et al. (2011). Urban growth, climate change, and freshwater availability. *Proceedings of the National Academy of Sciences, 108*(15), 6312-6317. doi:10.1073/pnas.1011615108

McGranahan, G., Balk, D., & Anderson, B. (2007). The rising tide: assessing the risks of climate change and human settlements in low elevation coastal zones. *Environment and urbanization, 19*(1), 17-37.

McGranahan, G., Marcotullio, P. J., Bai, X., Balk, D., Braga, T., et al. (2005). Urban Systems. In R. Hassan, R. Scholes, & N. Ash (Eds.), *Ecosystems and Human Well-Being: Current State and Trends* (pp. 797-825): Island Press.

Menze, B. H., & Ur, J. A. (2012). Mapping patterns of long-term settlement in Northern Mesopotamia at a large scale. *Proceedings of the National Academy of Sciences of the United States of America, 109*(14), E778-E787. doi: 10.1073/pnas.1115472109

Miyamoto, H., Hashimoto, T., & Michioku, K. (2011). Basin-wide distribution of land use and human population: stream order modeling and river basin classification in Japan. *Environmental Management, 47*(5), 885-898.

Nadaraya, E. (1974). On the integral mean square error of some nonparametric estimates for the density function. *Theory of Probability & Its Applications, 19*(1), 133-141.

New, M., Lister, D., Hulme, M., & Makin, I. (2002). A high-resolution data set of surface climate over global land areas. *Climate Research, 21*(1), 1-25. doi: 10.3354/Cr021001

O'Callaghan, J. F., & Mark, D. M. (1984). The extraction of drainage networks from digital elevation data. *Computer vision, graphics, and image processing, 28*(3), 323-344.

Ogneva-Himmelberger, Y., Pearsall, H., & Rakshit, R. (2009). Concrete evidence & geographically weighted regression: A regional analysis of wealth and the land cover in Massachusetts. *Applied Geography, 29*(4), 478-487. doi:10.1016/j.apgeog.2009.03.001

Páez, A., Uchida, T., & Miyamoto, K. (2002). A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific

kernel bandwidths and a test for locational heterogeneity. *Environment and Planning A, 34*(4), 733-754.

Pijanowski, B. C., Brown, D. G., Shellito, B. A., & Manik, G. A. (2002). Using neural networks and GIS to forecast land use changes: a land transformation model. *Computers, environment and urban systems, 26*(6), 553-575.

Portnov, B. A., & Paz, S. (2008). Climate change and urbanization in arid regions. *Annals of Arid Zone, 47*(3-4), 457-471.

Portnov, B. A., & Safriel, U. N. (2004). Combating desertification in the Negev: dryland agriculture vs. dryland urbanization. *Journal of Arid Environments, 56*(4), 659-680. doi:10.1016/s0140-1963(03)00087-9

Rebelo, A. G., Holmes, P. M., Dorse, C., & Wood, J. (2011). Impacts of urbanization in a biodiversity hotspot: Conservation challenges in Metropolitan Cape Town. *South African Journal of Botany, 77*(1), 20-35. doi:10.1016/j.sajb.2010.04.006

Robinson, N., Regetz, J., & Guralnick, R. P. (2014). EarthEnv-DEM90: A nearly-global, void-free, multi-scale smoothed, 90m digital elevation model from fused ASTER and SRTM data. *Isprs Journal of Photogrammetry and Remote Sensing, 87*, 57-67.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 65-78.

Rusiecki, V. K., & Warne, S. A. (1993). Synthesis of N-Alpha-Fmoc-N-Epsilon-Nvoc-Lysine and Use in the Preparation of Selectively Functionalized Peptides. *Bioorganic & Medicinal Chemistry Letters, 3*(4), 707-710. doi: 10.1016/S0960-894x(01)81259-2

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., et al. (2014). Multimodel assessment of water scarcity under climate change. *Proceedings of the National Academy of Sciences, 111*(9), 3245–3250. doi:10.1073/pnas.1222460110

Schneider, A., Friedl, M. A., & Potere, D. (2009). A new map of global urban extent from MODIS satellite data. *Environmental Research Letters, 4*(4). doi:10.1088/1748-9326/4/4/044003

Schneider, A., Friedl, M. A., & Potere, D. (2010). Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions'. *Remote Sensing of Environment, 114*(8), 1733-1746. doi:10.1016/j.rse.2010.03.003

Schueler, T., Fraley-McNeal, L., & Cappiella, K. (2009). Is Impervious Cover Still Important? Review of Recent Research. *Journal of Hydrologic Engineering, 14*(4), 309-315. doi:10.1061/(ASCE)1084-0699(2009)14:4(309)

Schumann, G. J. P., Neal, J. C., Voisin, N., Andreadis, K. M., Pappenberger, F., et al. (2013). A first large-scale flood inundation forecasting model. *Water Resources Research, 49*(10), 6248-6257.

Scott, D. W., & Terrell, G. R. (1987). Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association, 82*(400), 1131-1146.

Seager, R., Ting, M., Held, I., Kushnir, Y., Lu, J., et al. (2007). Model Projections of an Imminent Transition to a More Arid Climate in Southwestern North America. *Science, 316*(5828), 1181-1184. doi:10.1126/science.1139601

Seaman, D. E., & Powell, R. A. (1996). An evaluation of the accuracy of kernel density estimators for home range analysis. *Ecology, 77*(7), 2075-2085. doi:10.2307/2265701

Seto, K. C. (2011). Exploring the dynamics of migration to mega-delta cities in Asia and Africa: Contemporary drivers and future scenarios. *Global Environmental Change-Human and Policy Dimensions, 21*, S94-S107. doi:10.1016/j.gloenvcha.2011.08.005

Seto, K. C., Fragkias, M., Güneralp, B., & Reilly, M. K. (2011). A meta-analysis of global urban land expansion. *PloS ONE, 6*(8), e23777.

Seto, K. C., Güneralp, B., & Hutyra, L. R. (2012). Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences of the United States of America, 109*(40), 16083-16088.

Sherbinin, A. d., Levy, M., Adamo, S., MacManus, K., Yetman, G., et al. (2012). Migration and risk: net migration in marginal ecosystems and hazardous areas. *Environmental Research Letters, 7*(4), 045602.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis* (Vol. 26): CRC press.

Sivakumar, B. (2012). Socio-hydrology: not a new science, but a recycled and re-worded hydrosociology. *Hydrological Processes, 26*(24), 3788-3790. doi:10.1002/hyp.9511

Sivapalan, M., Savenije, H. H. G., & Bloschl, G. (2012). Socio-hydrology: A new science of people and water. *Hydrological Processes, 26*(8), 1270-1276. doi:10.1002/hyp.8426

Skeldon, R. (2014). *Migration and development: A global perspective*: Routledge.

Small, C., & Cohen, J. E. (2004). Continental physiography, climate, and the global distribution of human population. *Current Anthropology, 45*(2), 269-277.

Srinivasan, V., Lambin, E. F., Gorelick, S. M., Thompson, B. H., & Rozelle, S. (2012). The nature and causes of the global water crisis: Syndromes from a meta-analysis of coupled human-water studies. *Water Resource Research, 48*(10), W10516. doi:10.1029/2011wr011087

Srinivasan, V., Seto, K. C., Emerson, R., & Gorelick, S. M. (2013). The impact of urbanization on water vulnerability: A coupled human-environment system approach for Chennai, India. *Global Environmental Change, 23*(1), 229-239.

Steiniger, S., & Hunter, A. J. S. (2013). A scaled line-based kernel density estimator for the retrieval of utilization distributions and home ranges from GPS movement tracks. *Ecological Informatics, 13*, 1-8. doi:10.1016/j.ecoinf.2012.10.002

Stewart, D. (2002). Middle East Urban Studies II: Growth, Environment, and Economic Development. *Urban Geography, 23*(4), 388-394.

Storch, H., & Downes, N. K. (2011). A scenario-based approach to assess Ho Chi Minh City's urban development strategies against the impact of climate change. *Cities, 28*(6), 517-526.

Strahler, A. N. (1952). Hypsometric (Area-Altitude) Analysis of Erosional Topography. *Geological Society of America Bulletin, 63*(11), 1117-&. doi: 10.1130/0016-7606(1952)63[1117:Haaoet]2.0.Co;2

Strahler, A. N. (1957). Quantitative analysis of watershed geomorphology.*Eos, Transactions American Geophysical Union, 38*(6), 913-920.

Struckmeier, W. et al. Groundwater Resources of the World (1:25,000,000) (BGR & UNESCO World-wide Hydrogeological Mapping and Assessment Programme, 2008).

Syvitski, J. P., & Saito, Y. (2007). Morphodynamics of deltas under the influence of humans. *Global and Planetary Change, 57*(3), 261-282.

Syvitski, J. P. M., Kettner, A. J., Overeem, I., Hutton, E. W. H., Hannon, M. T., et al. (2009). Sinking deltas due to human activities. *Nature Geoscience, 2*(10), 681-686. doi:10.1038/NGEO629

Thompson, S. E., Sivapalan, M., Harman, C. J., Srinivasan, V., Hipsey, M. R., et al. (2013). Developing predictive insight into changing water systems: Use-inspired hydrologic science for the anthropocene. *Hydrology and Earth System Sciences, 17*(12), 5013-5039.

Trabucco, A., & Zomer, R. J. (2009). Global aridity index (global-aridity) and global potential evapo-transpiration (global-PET) geospatial database.*CGIAR Consortium for Spatial Information*.

Tu, J. (2011). Spatially varying relationships between land use and water quality across an urbanization gradient explored by geographically weighted regression. *Applied Geography, 31*(1), 376-392. doi:10.1016/j.apgeog.2010.08.001

Turner, W. R., Nakamura, T., & Dinetti, M. (2004). Global urbanization and the separation of humans from nature. *Bioscience, 54*(6), 585-590.

UNEP. (2001). The Mesopotamian Marshlands: Demise of an Ecosystem.*Division of Early Warning and Assessment, United Nations Environment Program (UNEP) Nairobi, Kenya, 46*.

United Nations. (2011). *World urbanization prospects: the 2005 revision*. United Nations Publications.

United Nations. (2012). World urbanization prospects, the 2011 revision: Highlights. *New York. Online at http://esa. un. org/unpd/wup/pdf/WUP2011_Highlights. pdf*. (access 05.14.2015)

United Nations. (2013). World population prospects: the 2012 revision. *Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, New York*. United Nations. (2014). World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER. A/352). *New York, United*.

USGS. (2000). HYDRO1k Elevation Derivative Database. *Online at http://edc.usgs.gov/products/elevation/gtopo30/hydro/* (access 11.22.2013).

Vörösmarty, C. J., Green, P., Salisbury, J., & Lammers, R. B. (2000). Global water resources: Vulnerability from climate change and population growth. *Science, 289*(5477), 284-288.

Van Vliet, M. T. H., Franssen, W. H. P., Yearsley, J. R., Ludwig, F., Haddeland, I., et al. (2013). Global river discharge and water temperature under climate change. *Global Environmental Change, 23*(2), 450-464.

Wang, H., He, Q., Liu, X., Zhuang, Y., & Hong, S. (2012). Global urbanization research from 1991 to 2009: A systematic research review. *Landscape and Urban Planning, 104*(3), 299-309.

Ward, P. J., Jongman, B., Weiland, F. S., Bouwman, A., Van Beek, R., et al. (2013). Assessing flood risk at the global scale: Model setup, results, and sensitivity. *Environmental Research Letters, 8*(4).

Werner, B., & McNamara, D. (2007). Dynamics of coupled human-landscape systems. *Geomorphology, 91*(3), 393-407.

Wessel, P., & Smith, W. H. F. (1996). A global, self-consistent, hierarchical, high-resolution shoreline database. *Journal of Geophysical Research-Solid Earth, 101*(B4), 8741-8743. doi:10.1029/96jb00104

WHO, U. (2015). Progress on Sanitation and Drinking Water: 2015 Update and MDG Assessment. *Geneva: World Health Organization*.

Wisser, D., Fekete, B. M., Vörösmarty, C. J., & Schumann, A. H. (2010). Reconstructing 20th century global hydrography: A contribution to the Global Terrestrial Network- Hydrology (GTN-H). *Hydrology and Earth System Sciences, 14*(1), 1-24.

Wood, F. B. (1992). Climate Cycles and the Development of Civilization. *General Systems Approaches to Alternative Economics and Values, Vols 1 and 2*, 203-212.

Woodroffe, C. D., Nicholls, R. J., Saito, Y., Chen, Z., & Goodbred, S. L. (2006). Landscape variability and the response of Asian megadeltas to environmental change *Global change and integrated coastal management* (pp. 277-314): Springer.

Woodroofe, M. (1970). On choosing a delta-sequence. *The Annals of Mathematical Statistics*, 1665-1671.

Woodruff, J. D., Irish, J. L., & Camargo, S. J. (2013). Coastal flooding by tropical cyclones and sea-level rise. *Nature, 504*(7478), 44-52. doi:10.1038/nature12855

Yamazaki, D., Kanae, S., Kim, H., & Oki, T. (2011). A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resources Research, 47*(4), W04501. doi:10.1029/2010wr009726

Zhang, Z. H., Xiao, R., Shortridge, A., & Wu, J. P. (2014). Spatial Point Pattern Analysis of Human Settlements and Geographical Associations in Eastern Coastal China - A Case Study. *International Journal of Environmental Research and Public Health, 11*(3), 2818-2833. doi:10.3390/ijerph110302818

Zhou, Y. Y., Smith, S. J., Zhao, K. G., Imhoff, M., Thomson, A., et al. (2015). A global map of urban extent from nightlights. *Environmental Research Letters, 10*(5). doi:10.1088/1748-9326/10/5/054011

Zomer, R. J., Trabucco, A., Bossio, D. A., & Verchot, L. V. (2008). Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agriculture, Ecosystems & Environment, 126*(1–2), 67-80. doi:10.1016/j.agee.2008.01.014