# Data-driven Analysis of Potential Impacts of Land-use/cover Change on Water Resources in Coastal Watersheds: Perspectives from Non-stationarity and Nonlinearity

Dissertation

Zur Erlangung des Doktorgrades
von der Mathematisch-Naturwissenschaftlich Fakultät der
Christian-Albrechts-Universität zu Kiel

vorgelegt von
Zhenyu Zhang M. Sc.

Kiel, 2023

Erste Gutachterin: Prof. Dr. Nicola Fohrer

Zweite Gutachterin: Prof. Dr. Eileen Eckmeier

Tag der mündlichen Prüfung: 26 May 2023

# Abstract

Studies of land-water interactions are important for water management. The properties of land surfaces determine the quantity and quality of water resources. Land use changes can modify streamflow regimes and exacerbate water pollution. Thus, a better understanding of the quantity and quality of water resources in changing environments is critical for watershed management, especially in coastal watersheds, because they provide important ecosystem services.

Understanding how land use/cover change (LUCC) impacts regional water resources is also of great concern for water security. The assumption of stationarity in hydrology is questioned in a changing environment as the theoretical basis of water resource management and the linkage between land use and water quality is scale-dependent and nonlinear. Thus, the aims of this study are to identify the spatial driving factors that control LUCC patterns, to investigate the impacts of human activities on hydrological extremes, and to explore the nonlinear relationship between LUCC and water quality in coastal watersheds.

The driving factors of LUCC were evaluated using a machine learning-based CA-Markov model developed in this study. The results illustrated that changes in each category of land use were well calculated, with average AUC (Area Under Curve) values of 0.999 and 0.916 for the training and testing periods, respectively. We found out that urbanization is highly related to the distribution of population and GDP, and spatial variations of topography factors influence the distribution of woodland. The proposed method identifies the driving forces of LUCC and can therefore be used for sustainable land management in coastal watersheds.

Besides the LUCC, the climate variability may change the regimes of streamflow. Thus, the non-stationary methods have been employed to understand the relationship between hydrological extremes, human activities and climate variability. The results reveal that amplified streamflow according to precipitation may be easier to observe at shorter time scales, and human activities may potentially amplify streamflow extremes in the watersheds, especially, in small-scale watersheds. To further understand the coupling effects of climate change and human activities, the relationship between LUCC and regional climate change was analyzed. We found that accelerating urbanization in the Minjiang River watershed modified climate extremes. The influence of climate variability on water resources is intensified by human activities.

To further understand the effects of LUCC on water quality at different developing stages, we selected coastal watersheds in difference regions as the study areas and analyzed them using the interpretable machine learning. We found that the land use category with the highest influence on water quality changes with development stages, although agricultural activities

still played a major role in water pollution at all stages. In China, pollution from urbanization still constitutes an important factor controlling the water quality in watersheds. In contrast, because of the high rate of domestic wastewater treatment in the USA, urbanization has contributed only slightly to water quality degeneration in coastal watersheds. Additionally, climate variability is a cause of a nonlinear relationship between LUCC and water quality. It amplifies the export of nutrients from nonpoint or point pollution sources and modifies the transport and biochemical processes.

This study highlights how LUCC impacts the water quantity and water quality under a changing environment in coastal watersheds. The findings provide new insights into water resource management under global change conditions.

# Zusammenfassung

Studien zu Interaktionen zwischen Gewässern und der Landoberfläche sind Grundlagen des Managements von Wasserressourcen, denn die Eigenschaften der Landoberflächen steuern die Quantität und Qualität der Wasserressourcen. Durch Änderungen der Landnutzung ändern sich auch das Abflussregime und die Verschmutzung der Gewässer. Ein besseres Verständnis der Entwicklung von Quantität und Qualität ist deshalb vor allem in sich schnell ändernden, küstennahen Systemen nötig, weil sie auch wichtige Ökosystemdienstleistungen bereitstellen.

Die Auswirkungen von Landnutzungsänderung auf regionale Wasserressourcen spielen auch eine wichtige Rolle bei der Beurteilung der Sicherheit der Wasserversorgung. In einem sich verändernden hydrologischen System sind die Zusammenhänge zwischen Landnutzung und Wasserhaushalt oft nicht mehr stationär, sondern nicht-linear und skalenabhängig. Das Ziel dieser Arbeit ist deshalb die Identifikation der räumlichen Einflußfaktoren, die Analyse des anthropgenen Einflusses auf hydrologische Extreme und die Analyse der nicht-linearen Beziehungen zwischen Landnutzungsänderungen (LUCC) in küstennahen Ökosystemen.

Die Steuergrößen des LUCC wurden mit einem “machine learning” basierten CA-(Cellular Automata) Markov Modell analysiert, das in dieser Arbeit entwickelt wurde. Die Ergebnisse zeigen, dass die Änderungen in jeder Kategorie gut erfassbar waren, die mittleren AUC (Area Under Curve) Werte bewegten sich zwischen 0.999 und 0.916 für die Trainings- und Testperioden. Wir fanden eine starke Verbindung zwischen der Urbanisierung, Bevölkerungsverteilung und Bruttosozialprodukt. Die Verteilung von Waldflächen wurde vor allem von der Topographie beeinflusst. Mit der Methode konnten die Kontrollgrößen des LUCC identifiziert werden und für das nachhaltige Management der Landnutzung in küstennahen Einzugsgebieten genutzt werden.

Neben LUCC beeinflusst auch die Klimavariabilität das Abflussverhalten. Zum Verständnis der Wechselwirkungen zwischen hydrologischen Extremen, anthropogenen Aktivitäten und der Variabilität des Klimas wurden nicht-stationären Methoden eingesetzt. Die Ergebnisse zeigen, dass erhöhter Abfluss in kürzeren Zeitintervallen besser zu beobachten ist. Anthropogene Aktivitäten verändern hydrologischen Extreme, vor allem in kleinen Einzugsgebieten. Für ein tieferes Verständnis zwischen Klimaänderungen und anthropogenen Aktivitäten wurde der Zusammenhang zwischen regionalen Klimaänderungen und LUCC analysiert. Wir konnten eine Verbindung zwischen Urbanisierung des Minjiang Einzugsgebiets und veränderten klimatologische Extremen identifizieren und damit nachweisen, dass anthropogene Aktivitäten die Variabilität des lokalen Klimas beeinflussen.

Zur vertieften Analyse der Auswirkungen von LUCC auf die Wasserqualität wählten wir

verschiedene küstennahe Regionen mit unterschiedlichen ökonomischen Entwicklungsstadien aus und analysierten sie mit KI-Methoden (interpretable machine learning). Wir fanden heraus, dass die Landnutzung den stärksten Einfluss auf die Wasserqualität hatte, gefolgt von verschiedenen landwirtschaftlichen Aktivitäten. In China spielt die Wasserverschmutzung aus urbanen Regionen immer noch eine wichtige Rolle, durch die Abwasserreinigung nimmt die Rolle in anderen Regionen jedoch ab. Außerdem beeinflusst auch das Klima die Verbindung zwischen LUCC und Wasserqualität, es verstärkt der Export von Nährstoffen und aus Punkt- und Flächenquellen und modifiziert den Transport von Stoffen und andere biogeochemische Prozesse.

In dieser Studie wird analysiert, wie LUCC die Wasserquantität und -qualität von küstennahen Einzugsgebieten bei geänderten klimatischen Rahmenbedingungen beeinflusst. Die Ergebnisse geben neue Einblicke in das Management von Wasserressourcen in Zeiten des Klimawandels.

# Contents

Abstract ..... I

Zusammenfassung ..... III

Chapter 1 General Introduction ..... 1

1.1 Research background ..... 1

1.1.1 Dynamics of water sources ..... 1

1.1.2 Dynamics of land use/cover change ..... 2

1.1.3 Impacts of land use on water quality ..... 3

1.2 The aim of study ..... 4

1.2.1 Research gaps ..... 4

1.2.2 Research questions ..... 5

1.3 Methodology ..... 5

1.4 Thesis structure ..... 6

Chapter 2 A machine learning based CA-Markov model to understand land use changes with multi-scenario simulation in a large watershed in Southeast China ..... 7

Abstract ..... 8

2.1 Introduction ..... 9

2.2 Material and methods ..... 11

2.2.1 Study area ..... 11

2.2.2 Data sources ..... 11

2.2.3 The RF-CA-Markov model ..... 14

2.2.4 Method to evaluate land use change ..... 16

2.2.5 Scenario analysis ..... 16

2.3 Results ..... 17

2.3.1 Patterns of land use change ..... 17

2.3.2 Model training and validation ..... 17

2.3.3 Land use change under different scenarios ..... 20

2.4 Discussion ..... 24

2.4.1 LUCC in the MRW ..... 24

2.4.2 Driving factors for land use change ..... 25

2.4.3 Effects of spatial planning on land use change ..... 26

2.5 Conclusions ..... 27

Chapter 3 A method for detecting the non-stationarity during high flows under global change ..... 28
Abstract ..... 29
3.1 Introduction ..... 30
3.2 Materials and methods ..... 31
3.2.1 Study area ..... 31
3.2.2 Data sources and processing ..... 32
3.2.3 Streamflow reconstruction ..... 32
3.2.4 GEV model ..... 37
3.2.5 Method to distinguish the non-stationary of climate change and human activities 38
3.3 Results ..... 39
3.3.1 Model calibration and validation ..... 39
3.3.2 Patterns of hydrologic extremes ..... 39
3.3.3 Impacts of climate change and human forcing ..... 42
3.4 Discussion ..... 45
3.5 Conclusions ..... 49
Chapter 4 A framework to quantify the effects of urbanization on the regional climate change in a China coastal watershed: A non-stationary perspective ..... 51
Abstract ..... 52
4.1 Introduction ..... 53
4.2 Data and methodology ..... 55
4.2.1 Study area ..... 55
4.2.2 Data sources ..... 56
4.2.3 Urbanization & Climate indices ..... 56
4.2.4 Methods to distinguish the impacts from urbanization ..... 57
4.3 Results ..... 58
4.3.1 Patterns of urbanization ..... 58
4.3.2 Patterns of climate change ..... 60
4.3.3 The impact of urbanization on climate change ..... 63
4.4 Discussion ..... 65
4.4.1 Applicability of the proposed framework ..... 65
4.4.2 Dynamics of climate extremes ..... 65
4.4.3 Impacts of urbanization on local or regional climate extremes ..... 66
4.5 Conclusion ..... 67

Chapter 5 A grid-based interpretable machine learning method to understand the spatial relationships between watershed properties and water quality ..... 68
Abstract ..... 69
5.1 Introduction ..... 70
5.2 Material and methods ..... 72
5.2.1 Study area ..... 72
5.2.2 Data sources ..... 73
5.2.3 Impacts of urbanization on local or regional climate extremes ..... 75
5.3 Results ..... 76
5.3.1 Model training and validation ..... 76
5.3.2 Water quality ..... 78
5.3.3 Watershed properties and water quality ..... 79
5.4 Discussion ..... 84
5.4.1 Interpretable machine learning approach ..... 84
5.4.2 Relationships between watershed properties and water quality ..... 84
5.4.3 Outlook ..... 86
5.5 Conclusions ..... 87
Chapter 6 Use of interpretable machine learning to identify the factors influencing the nonlinear linkage between land use and river water quality in the Chesapeake Bay watershed ..... 88
Abstract ..... 89
6.1 Introduction ..... 90
6.2 Material and methods ..... 91
6.2.1 Study area ..... 91
6.2.2 Land Use and Land Cover ..... 92
6.2.3 Data Source ..... 93
6.2.4 An interpretable machine learning method ..... 94
6.3 Results ..... 94
6.3.1 Nitrogen and Phosphorus changes across land use types and seasons ..... 95
6.3.2 Linkage between riverine nutrients and land use ..... 95
6.3.3 Water quality in the PRB under different scenarios ..... 100
6.4 Discussions ..... 101
6.4.1 Training & validation of the models ..... 101
6.4.2 Nonlinear linkage between land use patterns with nutrient exports ..... 102

6.4.3 Does seasonality matter in the linkage between land use pattern with water quality? ..... 104
6.4.4 Limitations and outlook ..... 105
6.5 Conclusions ..... 105
Chapter 7 General discussion and conclusion ..... 106
7.1 General discussion of research questions ..... 106
7.1.1 What are the major factors influencing the land use/cover change in the coastal watersheds? ..... 106
7.1.2 How does land use/cover change impact regional water resources in the context of climate variability? ..... 107
7.1.3 What is the nonlinear relationship between land use and water quality in coastal watersheds? ..... 108
7.2 General conclusion ..... 108
7.3 Outlook ..... 109
Reference ..... 110
Acknowledgments ..... 127
Appendix ..... 128
Declaration ..... 143

# Chapter 1 General Introduction

## 1.1 Research background

Water resources are the ultimate connectors in global commitment towards a sustainable future (UN WWDR, 2020). Coupled with multiple stressors (e.g. climate change, hydrology processes), human activities may reduce the predictability of water availability and affect water quality. It is not surprising that many rivers across the world suffer from severe water quality deterioration owing to the rapidly increasing human population and economic development (UN WWDR, 2009). With a large population and rapid economic development, coastal watersheds in China have undergone unprecedented urbanization rates, as well as accelerated disturbances from climate change, which may modify regional water resources (Bai et al., 2014; Lu et al., 2019). Therefore, it is essential to quantify and understand the effects of land use change on water resources in coastal watersheds in a sophisticated and clear manner.

### 1.1.1 Dynamics of water sources

Valuing water has become recognized worldwide because over 2 billion people live in areas where water scarcity is becoming more prominent as the world changes (UN WWDR, 2021). Climate change is one of the major factors influencing the hydrological cycle, diminishing water resources availability, and changing spatial distribution worldwide (UN WWDR, 2020). In addition, increases in human activities, such as dam construction, deforestation, and urbanization, have significantly contributed to streamflow regimes alteration and water quality degradation (Zhang et al., 2020a).

Water pollution exacerbates regional water scarcity (Ma et al., 2020) and water pollution is on the rise globally, and eutrophication of surface waters is still a major environmental concern worldwide, especially in developing countries, because of the accelerating development of these areas and high demand for resources (UN WWDR, 2009; Shi et al., 2017). Given that water quality plays a pivotal role in habitat protection, agriculture, industry, and public health (Akasaka et al., 2010; Cuo et al., 2013; Shi et al., 2017; Qu et al., 2022), understanding the dynamics of water quality in watersheds is necessary for sustainable water resources management.

Watershed models serve as powerful tools for understanding watershed processes and are supportive for water resource management (Hörmann et al., 2007; Jackson-Blake et al., 2017; Zhang et al., 2020b). Although some models have clear physical significance, researchers

should consider critical criteria, such as the objectives of the user, data availability, and cost-benefit analysis, when selecting an appropriate model to understand the key processes within watersheds. Overly simplistic models are likely to fail under different field conditions that are beyond the bounds of the calibration data, whereas models with too many parameters may surpass the type of data required, and turn out to be unjustified in practice (Perrin et al., 2001; Paudel and Jwaitz, 2012; Jackson-Blake et al., 2017). Therefore, there is a continuing challenge in constructing or selecting an appropriate model structure that adequately represents the key watershed processes.

### 1.1.2 Dynamics of land use/cover change

Land use/cover change (LUCC) is an important theme regarding the impacts of human activities on the Earth ecosystems and plays a vital role in the evolution of the local or regional environment (Foley et al., 2005; Lambin and Meyfrooidt, 2011; Ning et al., 2018). As a priority region, the dramatic transformations of land use in Asia have been a concern for related studies (Zhao et al., 2006; Song and Deng, 2017; Shfizadeh-Moghadam et al., 2019). Because of the rapid economic development and complex ecosystems, the remarkable land changes, especially urbanization, that have occurred in coastal China during recent decades have been extensively documented and have drawn much attention (Huang et al., 2018; Liu et al., 2021a; Xie et al., 2022). Therefore, it is essential to quantify the patterns of LUCC and identify the major factors that influence LUCC in the coastal areas for watershed management.

Coupling human activities and related environments, spatial and temporal patterns of LUCC can reflect underlying human activities as well as their interactions with nature over time (Lambin and Meyfroidt, 2011; Xie et al., 2022). LUCC was mainly studied from a disciplinary perspective until the 1990s, while new tools and techniques have recently enhanced our ability to monitor and explore changes in land use/cover (Turner et al., 2007; Verburg et al., 2009; Chen and Kirwan, 2022). The recent development of geographic information systems and remote sensing technology has enabled researchers to deal with large amounts of data to understand LUCCs at the spatial and temporal scale (Mundia and Anlay, 2005; Rodriguez-Galiano et al., 2012; Ayalew et al., 2022). Based on these technologies, numerous methods have been proposed for detecting the dynamics of LUCC. Most of the methods were developed for the specific environmental settings (e.g., urbanization, forest retreat), or are too subtle to interpret without the fundamental knowledge of basic metrics of LUCCs. Recently, the intensity analysis, which provides a quantitative framework that can identify the patterns of land use change at different levels, has been widely applied to gain insights into the processes of land use changes in many countries such as Greece, India, and China (Mallinis et al., 2014; Huang et al., 2018; Govind and Ramesh, 2019).

Land use models are useful tools for identifying the driving forces of LUCC and understanding the patterns of land use/cover with specific environmental settings (Verweij et al., 2018; Zhou et al., 2020). Since the 1970s, a few seminal papers have discussed the topic of land use modelling, and boosted by increasing remote sensing data availability, and a number of land use models have been developed in recent decades, such as urban growth, forest landscapes, and agricultural land use models (Wilson, 1971; Batty et al., 1997; Liao et al., 2016; Thompson et al., 2016; Verweij et al., 2018). Thus, various methods have been proposed to simulate land use/cover under different scenarios, such as the multi-agent model, support vector machines, logistic regression, and cellular automata models (Castella and Verburg, 2007; Wagner and Waske, 2016; Thompson et al., 2016; Mustafa et al., 2018). The advantages and disadvantages of these models have been widely discussed in related studies (Perez-Vega et al., 2012; Mustafa et al., 2018; Niya et al., 2020). For example, the Markov chain models were proven to be effective in simulating LUCCs, providing adequate information to support land use management, while it is difficult to identify the spatial variation of the system effectively (Sang et al., 2011; Zhou et al., 2020; da Cunha et al., 2021). Therefore, to take advantage of them and eliminate their defects, researchers have attempted to integrate these models and apply them in land use modelling to achieve a deep understanding of the causes and possible future developments of land-use systems.

### 1.1.3 Impacts of land use on water quality

Nutrient exports in the watersheds are strongly related to increasing anthropogenic influences, such as urbanization, agriculture, industry, and sewage (Brabec et al., 2002; Brett et al., 2005; Huang et al., 2015). According to the Secretary-General of the United Nations, more than 80% of the pollutants entering the seas are contributed by the land-based pollution (UN General Assembly, 2004). As one of the most important hotspots characterized by intensive agricultural activities and dense populations, nutrient pollution in the watershed-coast continuum is one of the most significant environmental problems worldwide (Fohrer and Chicharo, 2011). It alters aquatic ecosystems and affects their capacity to provide essential ecosystem functions (Howarth et al., 2002; Brizzetti et al., 2021). However, compared with the point sources of the nutrients (e.g., industrial, and municipal wastewater discharge), the nonpoint sources of the nutrients, such as fertilizer application during agricultural activities, are more difficult to control in watersheds (Ongley et al., 2010). Therefore, identifying the relationship between land use and nutrient pollutants in watersheds is an imperative step into understanding the dynamics of the water quality in watersheds and in controlling the nutrient export within the catchment-coast continuum.

Several studies have addressed the relationships between land use and water pollution. Generally, human-impacted land has significant positive correlations with water pollution,

whereas the natural land is significantly negatively correlated with nutrients (Lopez et al., 2008; Pratt and Chang, 2012; Huang et al., 2015). Agricultural activities in watersheds may enhance nutrient export. Many studies also point out that riverine nutrients export is highly related to arable land in Europe, China, and the USA (Ahearn et al., 2005; Huang et al., 2015; Kändler et al., 2017) As a result of urbanization, increasing impervious surfaces in the watersheds may modify the streamflow regimes, and additional nutrient exports are usually observed in the urbanized areas (Kaushal et al., 2011). In contrast, the natural lands are usually considered as the net sinks for the pollutants in watersheds because forests can absorb or fix nutrients (Huang et al., 2015; Pacheco et al., 2015; Jachniak et al., 2019). Wetlands can improve downstream water quality by intercepting soil, nutrients, and other pollutants transported from uplands and upstream aquatic ecosystems (Tan et al., 2013).

The relationship between land use and water pollutants may be inconsistent and nonlinear across space (Kaushal et al., 2008; Liu et al., 2016). In addition, climate variability, soil conditions, and topographic factors may modify the relationship between land use and nutrient exports in the watersheds (Lei et al., 2021; Huang et al., 2021). The impacts of land use on water quality may also vary under different climatic conditions (Larned et al., 2004; Julian et al., 2017). The interaction between land use and climate variability may lead to large changes in pollutant concentrations or fluxes over short periods (Kaushal et al., 2008; Vidon et al., 2009; Huang et al., 2021). For example, Kaushal et al. (2008) suggested that the pollutants retained by the impervious surfaces in urbanized areas may amplify the nutrient export the during the flood season. Additionally, the natural land may become a source of nutrients because of anthropogenic activities, too. For example, the excess nitrogen in mature forests may be exported and result in nitrification and nitrate leaching (Pacheco et al., 2015). Thus, more attention should be paid to the relationship between land use composition and water quality in a watershed.

## 1.2 Aim of the study

### 1.2.1 Research gaps

As a result of complex interactions between humans and the physical environment, LUCC has been recognized as an important component of global environmental change (Foley et al., 2005; Huang et al., 2012; Akinemi et al., 2017; Ayalew et al., 2022). Assisted by technologies like remote sensing and geographical information systems, researchers can preliminarily understand the state and trend of LUCC over time and identify its major driving factors (Wilson, 1971; Batty et al., 1997; Liao et al., 2016). With statistical methods like the logistic regression, exploratory regression, and principal component analysis, researchers have identified the major driving factors of the heterogeneity and inertia of land use and employed these factors for land use modelling (Lau and Kam 2005; Feng and Tong, 2017; Lei et al., 2019).

However, it is still difficult to identify the relationship between the driving force and land use patterns.

LUCC can modify the water resources by changing the characteristics of the land surface, infiltration patterns, runoff, groundwater recharge, sediment, water yield, and evapotranspiration (Lee et al., 2009; Cuo et al., 2013; Shrestha et al., 2018). Additionally, land use change may alter the characteristics of nutrient exports by linking water quality to specific land use types. The complexity of processes induced by human activities and other factors in a watershed can exert linear or nonlinear impacts on water resources. Coupled with other factors, the patterns of nutrient exports may be more complex because of nonlinear relationships between land use and water quality under specific environmental settings. Therefore, it remains unclear how LUCCs control watershed water resources in a changing environment.

It is necessary to understand the dynamics of LUCCs and evaluate its potential impact on regional water resources in watersheds caused by intensifying human activities and accelerating economic development. Land use in the coastal areas will change constantly owing to population pressure and limited resources, and one key question is how water resources will be affected under such conditions. Thus, considering the stress caused by rapid LUCC, more attention should be paid to the coupling and coordination between LUCC and regional water resources, especially in coastal watersheds.

### 1.2.2 Research questions

In this thesis we aimed to investigate the effects of LUCC on water resources in a changing environment. The main research questions of this study are as follows:

(1) What are the major factors influencing the land use/cover change in coastal watersheds?

(2) How does land use/cover change impact regional water resources in the context of climate variability?

(3) What is the nonlinear relationship between land use and water quality in coastal watersheds?

## 1.3 Methodology

Focusing on the overall research objective and the key research questions, we developed a machine learning-based CA-Markov model to identify the major driving factors influencing the distribution of land use/cover, and the dynamics of LUCC under different scenarios. The effects of human activities and climate were identified using the non-stationarity theory based on a rainfall-runoff model, remote sensing data, and long-term hydrology records. Using an interpretable machine learning method, the nonlinear relationships between LUCC and water quality were identified in coastal watersheds with different development stages. Figure 1.1 shows the method route followed in this study.

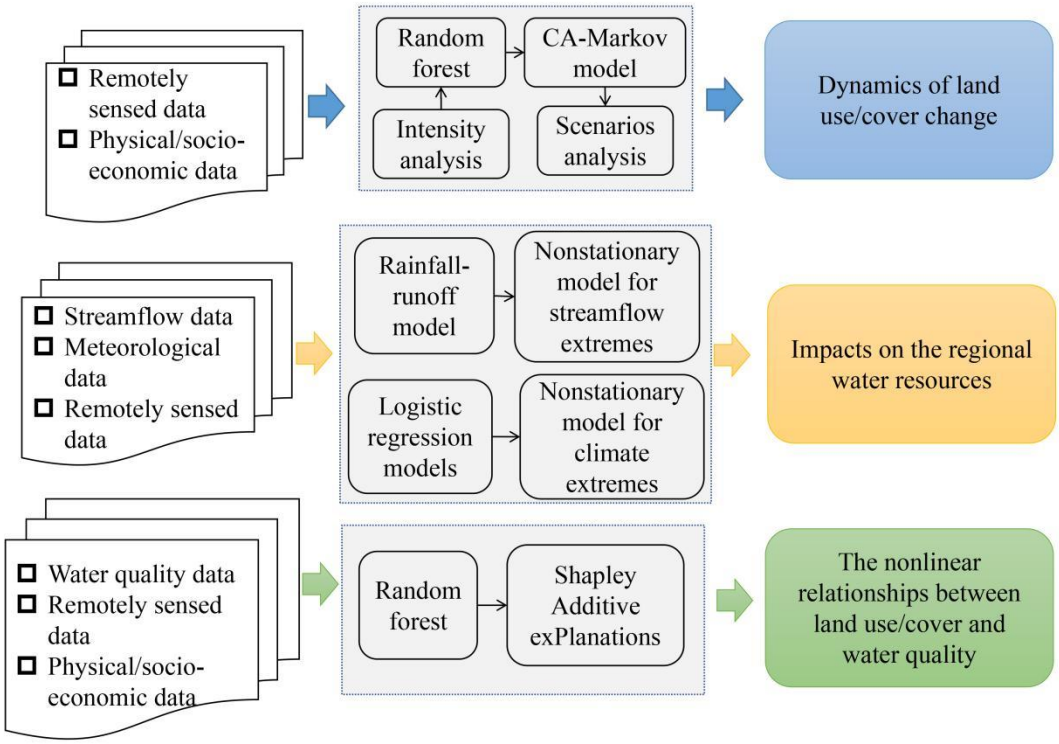


Fig. 1.1 Work flow in this study

## 1.4 Thesis structure

This PhD thesis is organized into seven chapters. Chapter 1 is a general introduction involving the background, research gaps and questions. Chapter 2 identifies the major factors influencing LUCC in coastal watersheds. Chapters 3 and 4 answer the question of how LUCC impacts regional water resources in the context of climate variability. Chapters 5 and 6 investigate the nonlinear relationships between land use and water quality in coastal watersheds. Chapter 7 discusses the main findings of this thesis and draws conclusions.

# Chapter 2 A machine learning based CA-Markov model to understand land use changes with multi-scenario simulation in a large watershed in Southeast China

**Zhenyu Zhang, Georg Hörmann, Jinliang Huang, and Nicola Fohrer**

**To be submitted Remote Sensing on 31th January 2023**

## Abstract

Understanding the land use/cover change (LUCC) in watersheds is essential for sustainable development. To comprehensively evaluate the factors that influence LUCC, we developed and implemented a machine learning-based CA-Markov model to understand the dynamics of LUCC in a coastal watershed located in Southeast China, namely, Minjiang River Watershed (MRW). The proposed method performed well for each category of land use with an average AUC value of 0.999 and 0.916 for the training and testing periods for the suitable images. And the overall accuracy for LUCC was 0.971. The process of urbanization in the MRW was speeding up recently, and the urban area increased by 2.22% of the total area during 2015-2020, and most of that was transferred from woodland and agricultural land. The population and GDP were the major factors influencing the distribution of urbanized land in the MRW. In contrast, the distribution of woodland was highly related to the topographic factors in the MRW. Additionally, the scenario analysis was employed based on the proposed method. The result shows that process of urbanization may be more complex with increasing population and GDP. Land use evolution may be more sustainable with scientific spatial plans which consider facilities for human beings and ecological protection. The proposed method quantified the LUCC with changing environmental settings, which would be a helpful tool for sustainable watershed management.

Key words: Land use/cover change; Machine learning; CA-Markov; Intensity Analysis;

## 2.1 Introduction

Land use/cover change (LUCC), as the most direct indicator depicting the effects of human activities on the natural ecosystem, plays a significant role in the process of global change (Hersperger et al., 2018; Bacau et al., 2022; Wang et al., 2022). As a complex system constructed by the interaction between natural processes and human activities, regional LUCC is influenced by social, political, and natural factors (Wang et al., 2021a; Chen and Kirwan, 2022; Palmate et al., 2022). Therefore, assessing and quantifying LUCC patterns is critical for developing regional environmental management and enhancing the understanding of the interactions between ecosystems and human activities (Marques et al., 2019; Zheng and Hu, 2018; Ayalew et al., 2022).

Modelling of LUCC has recently attracted growing interest, as it is a complex issue involving physical, environmental, and socio-economic factors (Aburas et al., 2017, 2019; Zhou et al., 2020). It has been proven that models are one of the most effective tools for understanding the dynamics of LUCC under various scenarios, grasping the spatio-temporal pattern of LUCC for regional environmental management, and providing a scientific reference for rational decision-making (Bacau et al., 2022; Fu et al., 2018; Yang et al., 2020). For this reason, a number of models have been developed to simulate LULC patterns, such as the logistic regression model, multi-agent models, and the cellular automata-Markov chain (CA-Markov) (Ervinia et al., 2018; Lei et al., 2019; Wang et al., 2021a; Palmate et al., 2022). These models have been widely applied to understand the evolution of land use and play a significant role in spatial planning (Zhou et al., 2020; Bacau et al., 2022). However, it is still difficult to meet the requirements of land use planning and management with these models for the absence of considering impacts from socio-economic and natural factors in these models (Zhou et al., 2020; Wang et al., 2021a). Hereby, the coupled models were developed recently to understand the LUCC in regional or global scale (e.g . Wagner and Fohrer, 2019; Zhou et al., 2020).

Coupled with different modelling techniques, the CA-Markov model has more advantages for simulating changes in complex land use systems by providing improved accuracy and effective simulation (Fu et al., 2018; Palmate et al., 2022; Wang et al., 2022). Typically, the researchers attempted to assign the suitability score for each affecting factor for land use based on experience, literature review, and the fuzzy-logic method, and quantify the impacts of factors on land use by converting original values to a unified scale, namely the degree of suitability. After scoring the suitability of each factor, transition rules were constructed and further applied for land use modelling (Fu et al., 2018). Among them, fuzzy-logic-based methods, such as logistic regression, have been widely used to build suitability data for land use modelling (Fu et al., 2018; Lei et al., 2019; Wagner and Fohrer, 2019). For example, to understand the driving forces for land use, Wang et al. (2022) integrated CA-Markov and logistic regression models

to scientifically simulate the law of land use evolution, and suggested that natural and human factors have the main impact on LULC, while proximity factors have a relatively small impact. Thus, the CA-Markov model is more scientific and practical compared to other models (Arsanjani et al., 2013; Xu et al., 2019; Wang et al., 2021a).

A common problem in existing methods is that the assignment suitability scores may not reflect land use changes for specific regions (Fu et al., 2018; Okwuashi and Ndehedehe, 2021). Although the fuzzy logic-based methods can model uncertainty in real world data with continuous boundaries, these methods may produce the problem of overfitted models because their rules are based on heuristics (Amato et al., 2018; Okwuashi and Ndehedehe, 2021). Meanwhile, these traditional fuzzy-logic based methods may not work well when modelling complex relationships between the potential driving factors and land use for the non-linear behaviour or discard the effects of heterogeneity may not be fully captured, especially for the process of urban development (Xu et al., 2019; Zhou et al., 2020; Viana et al., 2021). Recently, machine learning methods have attracted the attention of geospatial science researchers and have been applied in the remote sensing data analysis (Aburas et al., 2019; Karimi et al., 2019). As a subfield of artificial intelligence, machine learning can successfully overcome the limitations of previous methods, and achieve superior or at least equivalent accuracy outcomes (Ren et al., 2020; Viana et al., 2021). Nowadays, several studies have been carried out to understand land use change using machine learning models in cities or metropolitan regions. For example, coupled with the machine learning method, Zhou et al. (2020) simulated the evolution of urban space in Shanghai from 2015 to 2030 under different scenarios and showed that urban development will be more sustainable under the constraints of ecological and cultivated protection. However, most related studies have focused on the evolution of urbanization; these methods have rarely been introduced to understand the evaluation of multiple land use types at global or regional scales.

Using the integrated CA-Markov Chain model is of significance due to its important role in land use modelling, especially in developing countries (Aburas et al., 2017; Zhou et al.,2020). To date, few studies have combined machine learning models with the CA-Markov model, and limited research has been conducted on the regional scale in China. To close these knowledge gaps, we developed a machine learning-based CA-Markov model with a random forest (RF) algorithm. RF has proven to be a suitable machine-learning algorithm with good interpretability and moderate time complexity (Zhou et al., 2020). To understand the impacts of multiple variables on the evaluation of land use at the regional scale, we applied this model to the largest watershed in Southeast China, namely Minjiang River Watershed (MRW), which is under increasing pressure from continuous population growth, rapid socio-economic development, and limited natural resources. To improve the capability of the CA-Markov model, we selected the MRW as a case to study LUCC in Southeast China, and the aims of this study were (1) to

quantify the land use change in Southeast China, (2) to analyze the evolution of land use under different scenarios, and (3) to understand the role of strategic spatial planning on regional land use change and provide in-depth implications for sustainable watershed management.

## 2.2 Material and methods

### 2.2.1 Study area

The MRW (116°23′–119°43′ E, 25°23′–28°19′N) is the largest watershed in Southeast China, with an area of 60,992 km2. As a watershed located in the subtropical zone, the mean annual temperature and precipitation are 18 ℃ and 1617 mm, respectively, with approximately 70% of the precipitation occurring between April and September (Zhou et al., 2016; Zhang et al., 2019). Based on the official population record, the populations of 2010, 2015, and 2020 were 10.75 million, 11.02 million, and 11.72 million. More than 60% of the population lives in urban areas (Fig. S1). The GDP increased by 200% approximately in the past 10 years, which increased from 417 billion CNY (2010) to 1292 billion CNY (2020) (Fig. S2-1). By 2021, more than 1900 km of high-speed railway and 6000 km of highways were built in Fujian Province, and about 60% of which were built in the MRW. More highways and high-speed railways are planned to be constructed by 2030, which may change the patterns of land use in the MRW (Fig. 2-1).

### 2.2.2 Data sources

Landsat images were obtained from the USGS Earth Explorer (https://earthexplorer.usgs.gov/) for the years 2010, 2015, and 2020 in the MRW. After examining the available images, we realized that the uncertainties caused by the clouds would be more significant than the uncertainties induced by the inconsistent dates of images. Therefore, different dates were used to obtain complete and cloud-free coverage of the study area (Table 2-1). The unsupervised classification for the Landsat images was performed using the Iterative Self Organizing Data Analysis Technique Algorithm (ISODATA) integrated with post-classification enhancement (Yang and Liu, 2005; Huang et al., 2018). Ancillary data, including high-resolution images from Google Earth, GIS data, and information collected during field trips, were employed as reference data for the classification and accuracy assessment. In this study, we classified the land use/cover into 7 categories, namely, Woodland, Grassland, Agriculture, Orchard, Urban, Barren and Water (Table 2-2). The results of the accuracy assessment were shown in Table S2-1 and Table S2-2, and the overall accuracy of the classifications was 0.805±0.024, 0.863±0.021, and 0.864±0.021 for the years 2010, 2015, and 2020, respectively.

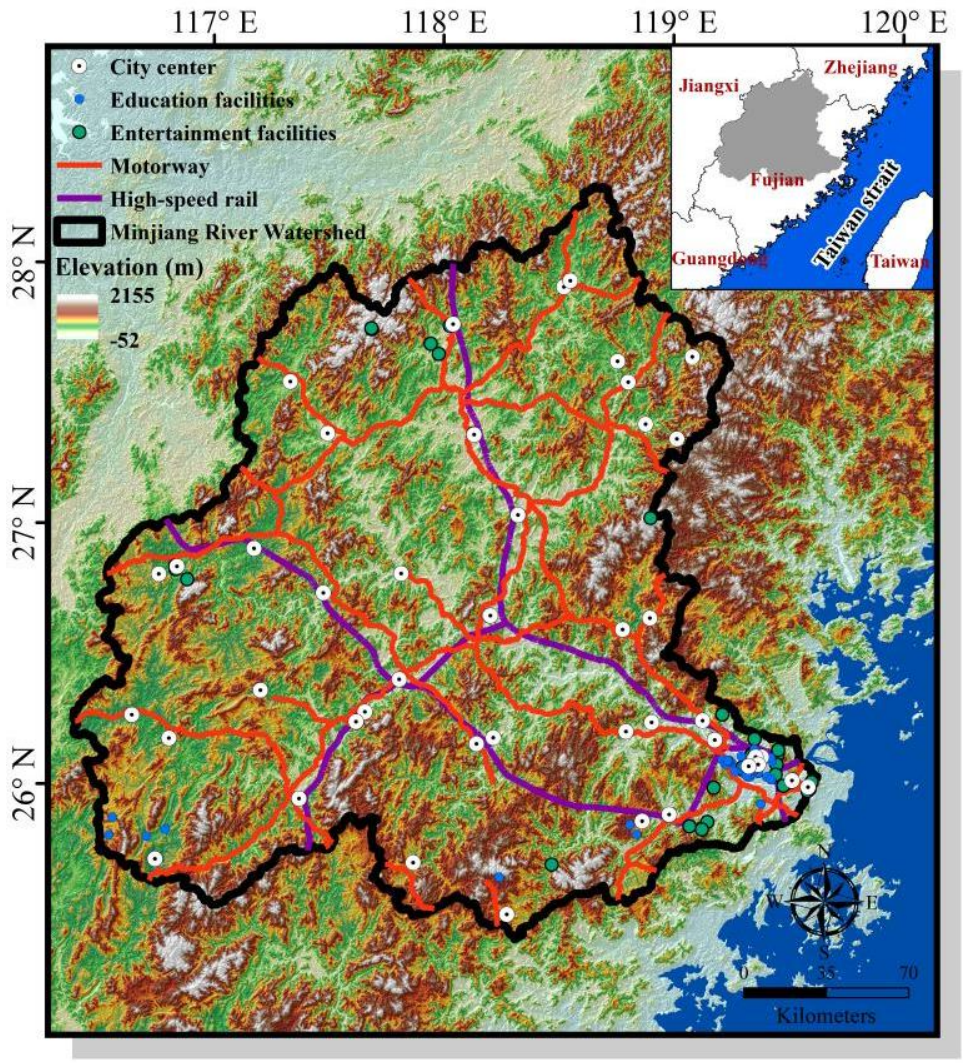


Fig 2-1 Study area

Based on previous studies, the spatial driving factors influencing land use/cover that we employed were listed in Table 2-3. The topographical data, including elevation, slope, and aspect, were estimated using the 30 m SRTM Digital Elevation Database produced by the National Aeronautics & Space of USA (NASA, https//www.nasa.gov/).The vector and Point-of-Interest (POI) data were collected from the OpenStreatMap (OSM, https://www.openstreetmap.org/), including motorway, high-speed rail, city centers, and education facilities in the MRW. The VIIRS Night-time Light (VNL) data were used to identify the socio-economic development in the MRW (Fig. S2-2), which was produced by the Earth Observation Group, Payne Institute for Public Policy, Colorado School of Mines (Elvidge et al., 2017; 2021). The population data were obtained based on China’s population density provided by the WorldPop & Center for International Earth Science Information Network (2018).

Table 2-1 Information of the satellite images

| Year 2010 | | | Year 2015 | | | Year 2020 | | |
|---|---|---|---|---|---|---|---|---|
| Acquisition data | Path/Row | Satellite-senor | Acquisition data | Path/Row | Satellite-senor | Acquisition data | Path/Row | Satellite-senor |
| 10/31/2010 | 119/41 | Landsat 5-TM | 9/27/2015 | 119/41 | Landsat 8-OLI | 10/10/2020 | 119/41 | Landsat 8-OLI |
| 10/31/2010 | 119/42 | Landsat 5-TM | 9/27/2015 | 119/42 | Landsat 8-OLI | 4/17/2020 | 119/42 | Landsat 8-OLI |
| 12/9/2010 | 120/41 | Landsat 5-TM | 2/26/2015 | 120/41 | Landsat 8-OLI | 2/20/2020 | 120/41 | Landsat 8-OLI |
| 12/9/2010 | 120/42 | Landsat 5-TM | 5/13/2015 | 120/42 | Landsat 8-OLI | 2/20/2020 | 120/42 | Landsat 8-OLI |
| 1/14/2010 | 121/41 | Landsat 5-TM | 10/11/2015 | 121/41 | Landsat 8-OLI | 4/15/2020 | 121/41 | Landsat 8-OLI |
| 1/14/2010 | 121/42 | Landsat 5-TM | 2/13/2015 | 121/42 | Landsat 8-OLI | 4/15/2020 | 121/42 | Landsat 8-OLI |

**Note: TM:Thematic Mapper; OLI: Operational Land Imager.**

Table 2-2 Classifications of land use/cover in the MRW

| **Category** | **Description** |
|---|---|
| Woodland | Any significant clustering of dense vegetation, typical with a closed or dense canopy. |
| Grassland | Open areas covered in homogenous grasses with little vegetation. |
| Agriculture | Land used for cultivation, including newly cultivated land, fallow land, swidden land, and rotation plough land. |
| Orchard | Areas for planting perennial woody plants and perennial herb which were used for collecting their fruit, leaves and rhizome. |
| Built-up | Human made structures, road, railway, large homogenous impervious surfaces. |
| Barren | The areas with little vegetation, including exposed rock or soil, desert and sand dunes, dry salt flats/pans, mines. |
| Water | Areas where water was predominantly present throughout the year. |

Table 2-3 Driving factors for land use/cover change modelling

| Category | Driving factor | Year |
|---|---|---|
| Topographical variable | Elevation | |
| | Slope | |
| | Aspect | |
| Proximity variables | Distance to motorway | 2015, 2020 |
| | Distance to high-speed rail | 2015, 2020 |
| | Distance to city center (i.e. cities and counties) | 2020 |
| | Distance to education facilities (i.e. kindergartens, schools, colleges and universities) | 2020 |
| | Distance to entertainment facilities (e.g. public garden) | 2020 |
| Socioeconomic variables | Population density | 2015, 2020 |
| | VIIRS nighttime lights | 2015, 2020 |

## 2.2.3 The RF-CA-Markov model

LUCC models were commonly set up based on the relationship between historical land use and related driving factors (Wu et al., 2019; Feng et al., 2020). In this study, we proposed a method to simulate land use with RF, CA and Markov chain (Fig. 2-2). The relationship between multiple driving factors and each land use type was identified with the RF method, and the cells that may transfer to other types of land use were depicted by the transition

suitability images estimated by the RF models. Previous studies have proven that RF is a powerful machine learning classifier for land use studies because of its non-parametric nature, high classification accuracy, ability to determine variable importance, and capability to avoid over-fitting (Rodriguez-Galiano et al., 2012; Tan et al., 2017; Zhou et al., 2020).

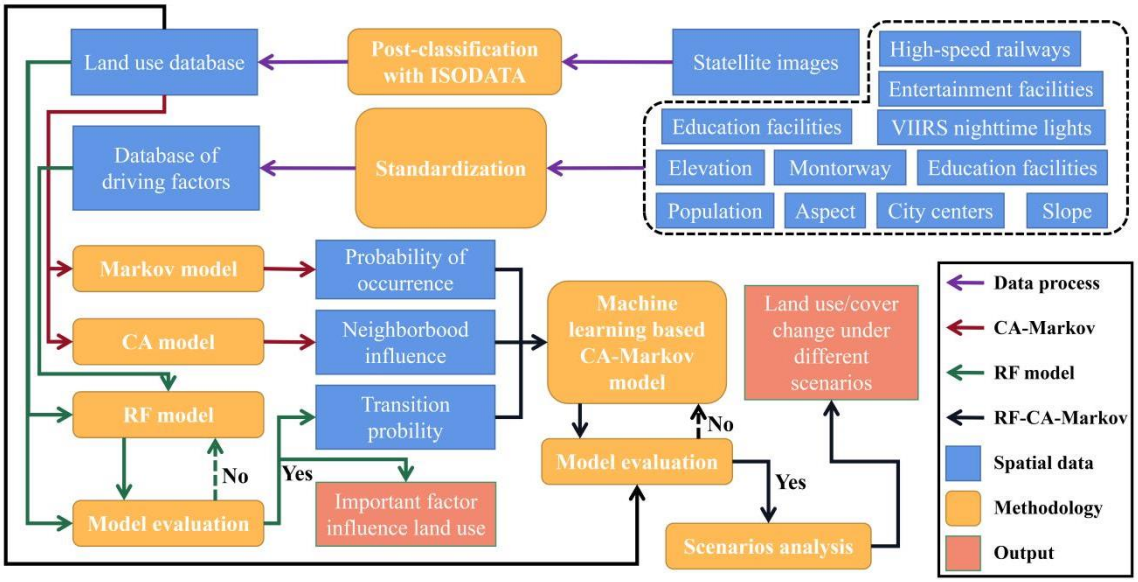


Fig 2-2 The flowchart of the coupling models

The coupled CA with the Markov model can be used to simulate the dynamics of land use. The CA model can describe the changes in spatial positions, while the temporal changes in land use were simulated by the Markov transformation matrix. The CA model is characterized by discreteness in space and the state of land use, which can be used to analyze spatial distribution and neighborhood interactions (Fu et al., 2018; Wang et al., 2021a). The overall complex, self-organizing system was described with a CA model based on individual cell behaviour and interactions between neighbouring cells (Wu & Webster, 1998). The land use changes were simulated with the assumption that areas have a higher tendency to transfer to typical land use when the same type of land use was nearby (Arsanjani et al., 2013). The CA model is composed with four elements, including a cell, cell state, neighbourhood, and transformation rules. The CA model has been widely applied in urbanized land use, such as sprawl (e.g. Zhou et al., 2020; Bacau et al., 2022).

The Markov model is a stochastic model to simulate LUCC in temporal scale to describe how likely one state is to change to another state. Based on the formation of Markov random process for the prediction and optimal control theory, the possibility of one state transferring to another state was described by the probability matrix (Takada et al., 2010). In the land use

analysis, the transition objective was produced by the Markov analysis with the historical land use images (Arsanjani et al., 2013).

## 2.2.4 Method to evaluate land use change

To evaluate the performance of the model, the simulated transition suitability images were validated with the receiver operating characteristic (ROC) curve. The Area Under Curve (AUC) was calculated to evaluate the performance of the model. An AUC value of 1 represents a perfect model performance and value of 0.5 is no better accuracy than chance. Additionally, the user's and producer's accuracy along with figure of merit (FOM) were employed to evaluate the performance of simulated land use maps which calculated values based on the confusion matrix (Table S2-3).

$$Overall\,accuracy = \frac{H + CR}{H + M + FA + CR}$$

$$Producer's\,accuracy(sensitivity) = \frac{H}{H + M}$$

$$User's\,accuracy(precision) = \frac{H}{H + FA}$$

$$FOM = \frac{H}{H + M + FA}$$

Where ***H***, ***M***, ***FA***, and ***CR*** are the number of hits, misses, false alarms, and correct rejections in the confusion matrix.

Intensity analysis for the category level was used to quantify land use change of each category. The change intensity of each category was measured by the percentages of ***Gain*** and ***Loss*** (Pontius et al., 2004). The components of the intensity were estimated by the percentages of ***Quantity, Exchange***, and ***Shift*** for each category in the study area (Pontius and Santacruz, 2014). ***Quantity*** is the quantitative difference of two time points, ***Exchange*** is the location changes between two categories during the period, and ***Shift*** was the location change of more than two categories (Shafizadeh-Moghadam et al., 2019; Pontius 2019).

## 2.2.5 Scenario analysis

To understand the patterns of LUCC under different strategies, we developed three scenarios to explore the dynamics of land use change in MRW under different policies. The scenarios I (Scenarios I) was set up under the assumption that changes of land use will be developed under current environmental setting, and all related factors were same as that in the year of 2020. The scenarios II (Scenarios II) was built with the assumption that the population and socio-economic development will increase or improve based on the expected value. Based on these plans, the Government of China indicated that China's total population will reach 1.45 billion which will increase by 2% compared to the population in 2020, and the GDP will grow at a rate of 5% per year. The scenarios III (Scenarios III) was built with the assumption that the

socio-economic factors will develop as expected, and the traffic, education, and the living facilities in the MRW will be improved continually. More high-speed railways and motorways will be built based on plans of the local government (Fujian Provincial Development and Reform Commission, http://fgw.fujian.gov.cn). The distances to the education and entertainment facilities will be decreased by 20% compared the year of 2020, and human activities will be limited with the protected area in the MRW..

## 2.3 Results

### 2.3.1 Patterns of land use change

The patterns of land use of ten years were presented in the Fig. S2-3, along with the losses and gains of different land use categories during each time interval. The transitions of each category of land use were also calculated in the Table 2-4. Woodland accounted for more than 80% of the MRW area at all time points. Net increased urban area was observed during all time intervals and there were no losses of urban area for these two periods. The urban area increased by 0.84% and 2.22% of the total area in 2010-2015, and 2015-2020. The increasing urban area was mainly transferred from agricultural and woodland areas, and 0.65% and 1.92% of the total area were transferred to the urban area which accounts for more than 75% net increased urban area in these periods.

The intensity of loss and gain and the components of temporal differences in the two periods were quantified in this study (Fig. 2-3). Based on the intensity analysis, the contribution of each category to the overall change was identified. The grassland, agricultural land, orchard, urban, and barren land were always active during the two periods.

### 2.3.2 Model training and validation

The RF models performed well for each land use category, with average AUC values of 0.999 and 0.916 for the training and testing periods, respectively (Fig. 2-4). During the testing period, the AUC values ranged from 0.844 for barren land to 0.981 for the water areas in the MRW. Overall, a high model performance was identified when estimating the transition suitability of woodland, agricultural areas, orchard, urban areas, and water areas, whereas low accuracy was observed for grassland and barren land.

The contribution of each driving factor on the output of RF models was displayed in Fig.2-5. The population, VIIRS nighttime lights, and slope were the most important factors influencing the distribution of urban areas. Population and slope were important factors that influence the distribution of agricultural land. In contrast, natural land use may be influenced by topographical variables. The most important factor that influenced distributions of woodland was slope, meanwhile the top factor that influenced distributions of water was alleviation.

Table 2-4 Land use/cover change during the two periods (% of MRW)

| From Category | Period | Woodland | Grassland | Agriculture | Orchard | Built-up | Barren | Water | Sum | Loss |
|---|---|---|---|---|---|---|---|---|---|---|
| Woodland | 2010-2015 | 84.13 | 0.11 | 0.36 | 0.12 | 0.20 | 0.01 | 0.01 | 84.94 | 0.81 |
| | 2015-2020 | 85.34 | 0.17 | 0.14 | 0.28 | 0.71 | 0.01 | 0.01 | 86.66 | 1.32 |
| Grassland | 2010-2015 | 2.51 | 1.59 | 0.02 | 0.06 | 0.10 | 0.01 | 0.03 | 4.32 | 2.73 |
| | 2015-2020 | 0.15 | 1.28 | 0.09 | 0.13 | 0.06 | | 0.01 | 1.72 | 0.44 |
| Agriculture | 2010-2015 | 0.02 | 0.01 | 3.18 | 0.10 | 0.45 | | | 3.76 | 0.58 |
| | 2015-2020 | 0.02 | 0.07 | 2.16 | 0.15 | 1.21 | | 0.02 | 3.63 | 1.47 |
| Orchard | 2010-2015 | | | 0.05 | 1.80 | 0.09 | 0.01 | | 1.95 | 0.15 |
| | 2015-2020 | | 0.03 | 0.03 | 1.81 | 0.20 | | 0.01 | 2.08 | 0.27 |
| Built-up | 2010-2015 | 0.00 | 0.00 | 0.00 | 0.00 | 3.54 | 0.00 | 0.00 | 3.54 | 0.00 |
| | 2015-2020 | 0.00 | 0.00 | 0.00 | 0.00 | 4.38 | 0.00 | 0.00 | 4.38 | 0.00 |
| Barren | 2010-2015 | | 0.01 | 0.01 | | | 0.10 | | 0.12 | 0.02 |
| | 2015-2020 | 0.02 | 0.02 | 0.02 | 0.02 | 0.01 | 0.04 | | 0.13 | 0.09 |
| Water | 2010-2015 | | | 0.01 | | | | 1.36 | 1.37 | 0.01 |
| | 2015-2020 | 0.03 | 0.01 | | 0.11 | 0.03 | | 1.22 | 1.40 | 0.18 |
| Sum | 2010-2015 | 86.66 | 1.72 | 3.63 | 2.08 | 4.38 | 0.13 | 1.40 | 100 | |
| | 2015-2020 | 85.56 | 1.58 | 2.44 | 2.50 | 6.60 | 0.05 | 1.27 | 100 | |
| Gain | 2010-2015 | 2.53 | 0.13 | 0.45 | 0.28 | 0.84 | 0.03 | 0.04 | | |
| | 2015-2020 | 0.22 | 0.30 | 0.28 | 0.7 | 2.22 | 0.01 | 0.05 | | |

**Note: Blanks indicated less than 0.005% of the total area of the MRW**

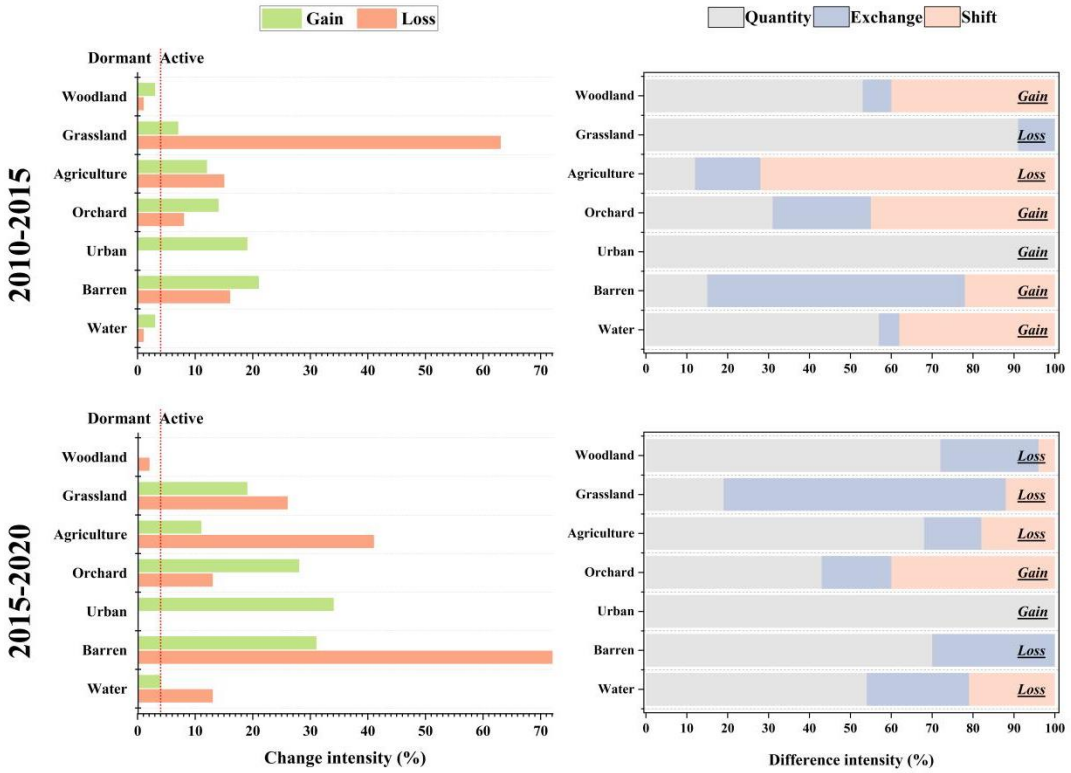


Fig 2-3 Intensity analysis of land use for different periods

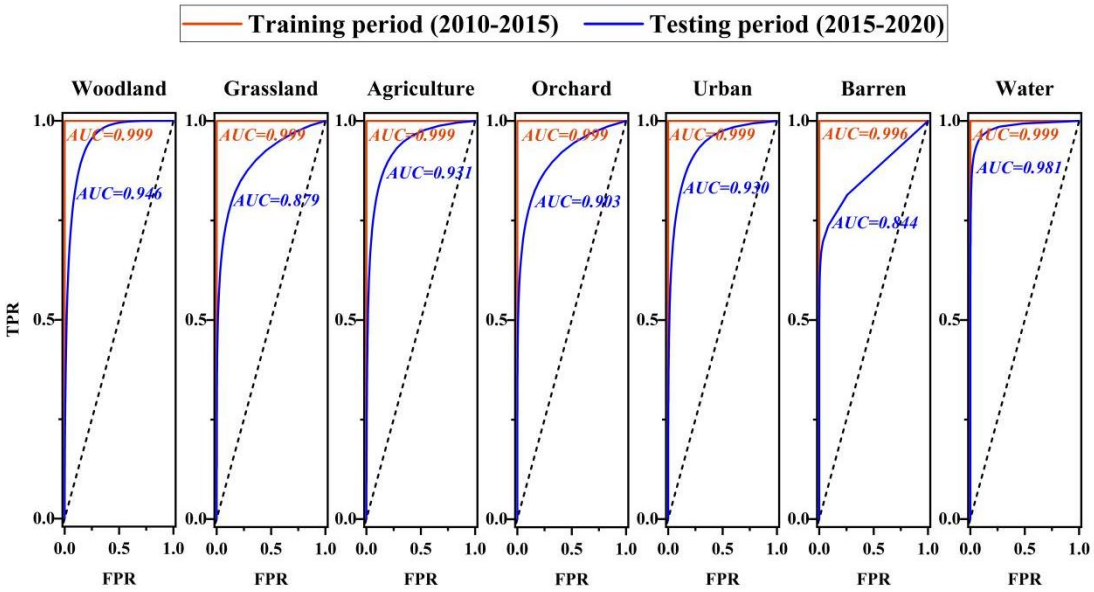


Fig 2-4 Performance of RF model for land use in the MRW

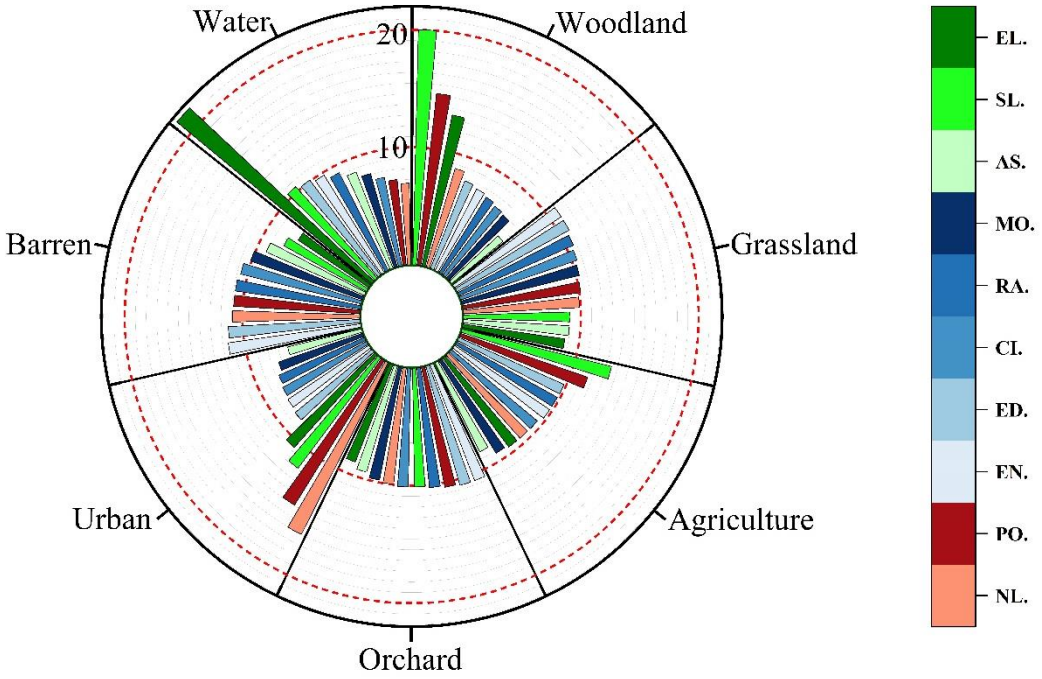


Fig 2-5 Variable importance for determining land use change based on RF models (Note: EL. ,SL., AS., MO., RA., CI., ED., EN., PO., and NL. represented elevation, slope, aspect, distance to motorway, distance to high-speed railway, distance to city center, distance to education facilities, distance to entertainment facilities, population density and VIIRS nighttime lights, respectively)

The land use patterns of the MRW in 2020 were simulated based on the proposed model (Fig. 2-6). The simulated land use was highly consistent with the observed land use. Though the accuracy of barren land (Producer's accuracy=68.30%, User's accuracy=61.70%, FOM=0.60) and grassland (Producer's accuracy=60.98%, User's accuracy=84.59%, FOM=0.64) was lower than other categories of land use, the proposed model still suitable for predict land use change in the MRW. The overall accuracy of the map was 0.971 (Table 2-5).

### 2.3.3 Land use change under different scenarios

Based on the conditions set under the three development scenarios, transition suitability images were simulated using the RF models (Fig. S2-4,S2-5, and S2-6). In Scenario I, the patterns of land use change were identified under the assumption that changes in land use will be developed under the environmental setting of 2020 (Fig. 2-7). The results show that there are no new barren areas, and approximately 81.8% of barren land may be transferred to other types of land use, including urbanized land, orchard, agricultural land, and grassland. And the urban area will increase continually during this process..

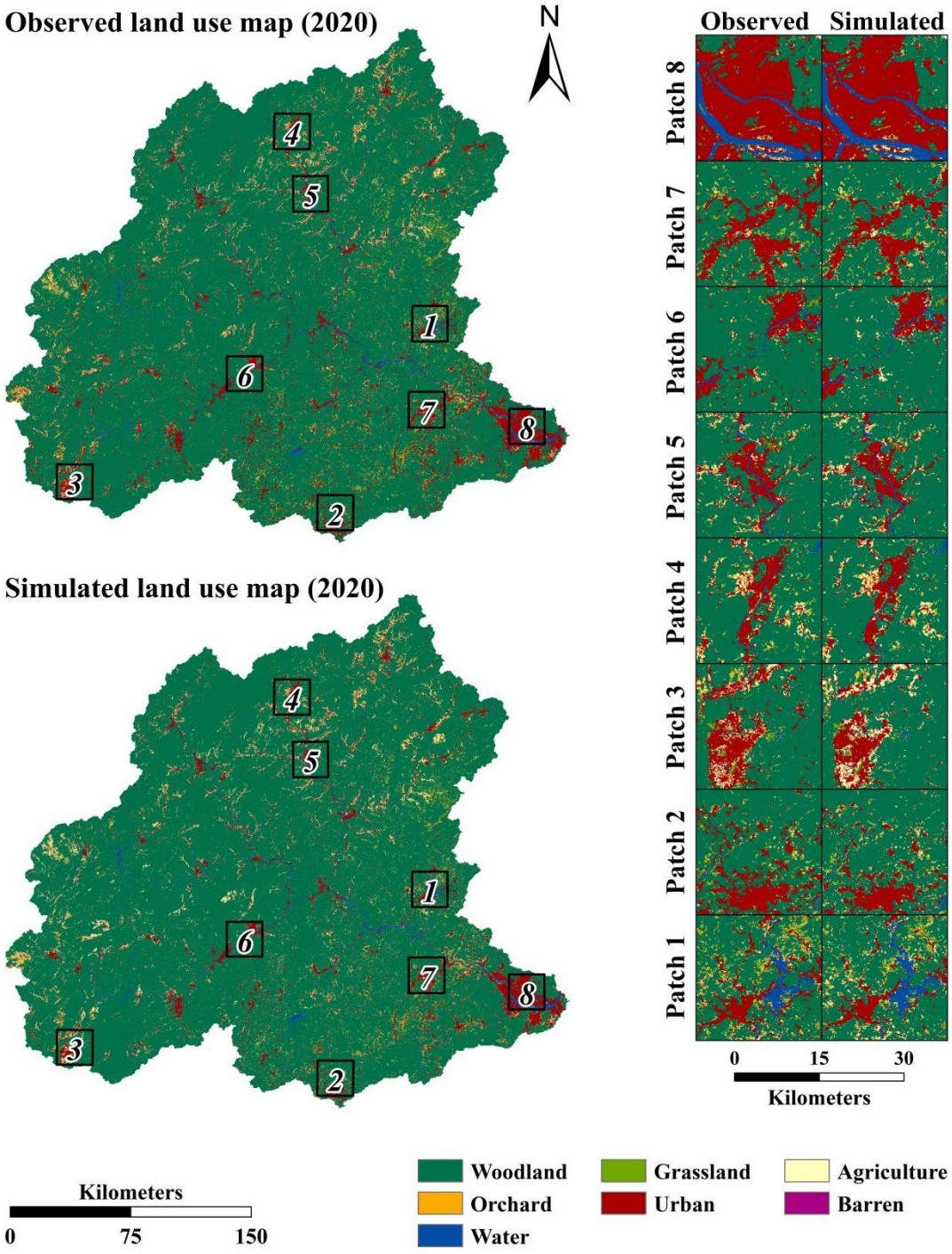


Fig 2-6 Observed and simulated land use patterns in 2020

Table 2-5 Accuracy assessment of the model

| | Woodland | Grassland | Agriculture | Orchard | Built-up | Barren | Water |
|---|---|---|---|---|---|---|---|
| Overall accuracy | 98.58% | 99.46% | 98.73% | 99.18% | 98.53% | 99.91% | 99.78% |
| Producer's accuracy | 99.76% | 60.98% | 96.59% | 75.60% | 78.00% | 68.30% | 96.72% |
| User's accuracy | 98.63% | 84.59% | 66.74% | 92.20% | 99.89% | 61.70% | 87.32% |
| FOM | 0.99 | 0.64 | 0.65 | 0.75 | 0.78 | 0.6 | 0.85 |
| Overall accuracy of map | | | | 0.971 | | | |
| Kappa* | | | | 0.884 | | | |

***Note: Kappa is a metric we do not encourage. We include it here to help user better understand its misleading properties and compare with related studies. For detail discussion of its drawbacks, please see Pontius and Miliones (2011).**

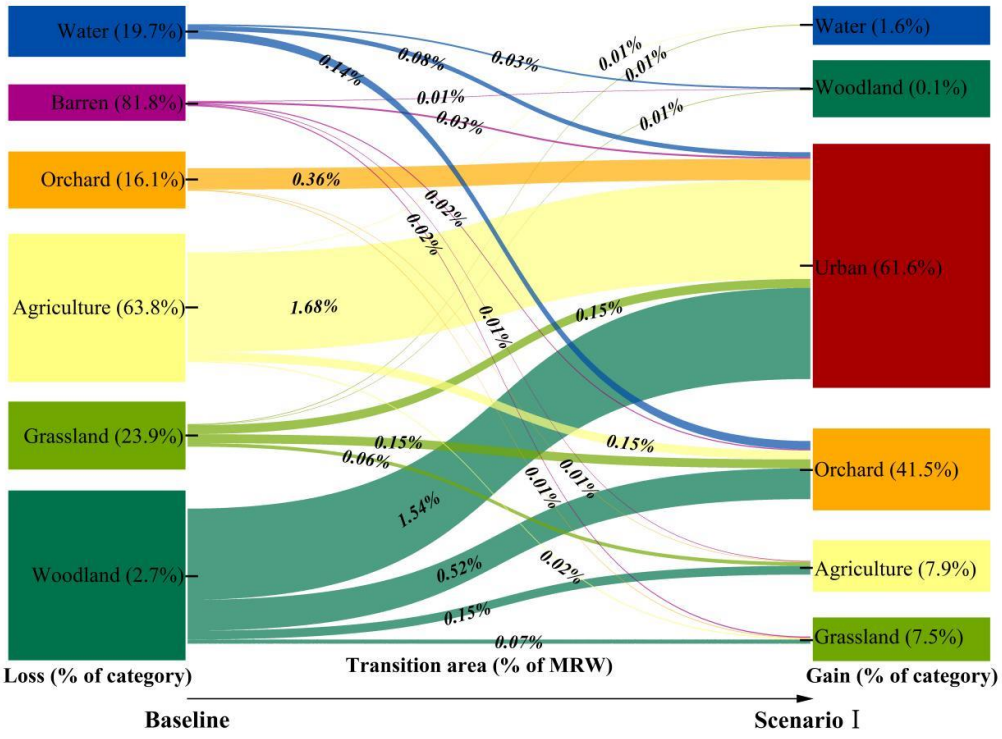


Fig 2-7 Patterns of land use change under Scenarios I

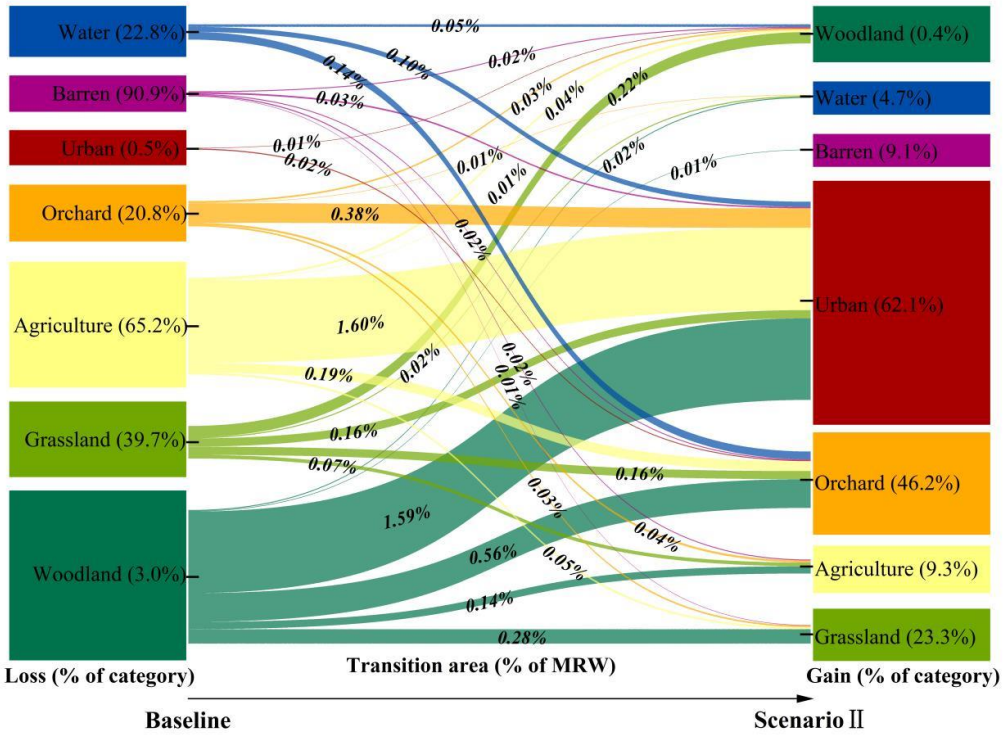


Fig 2-8 Patterns of land use change under Scenarios II

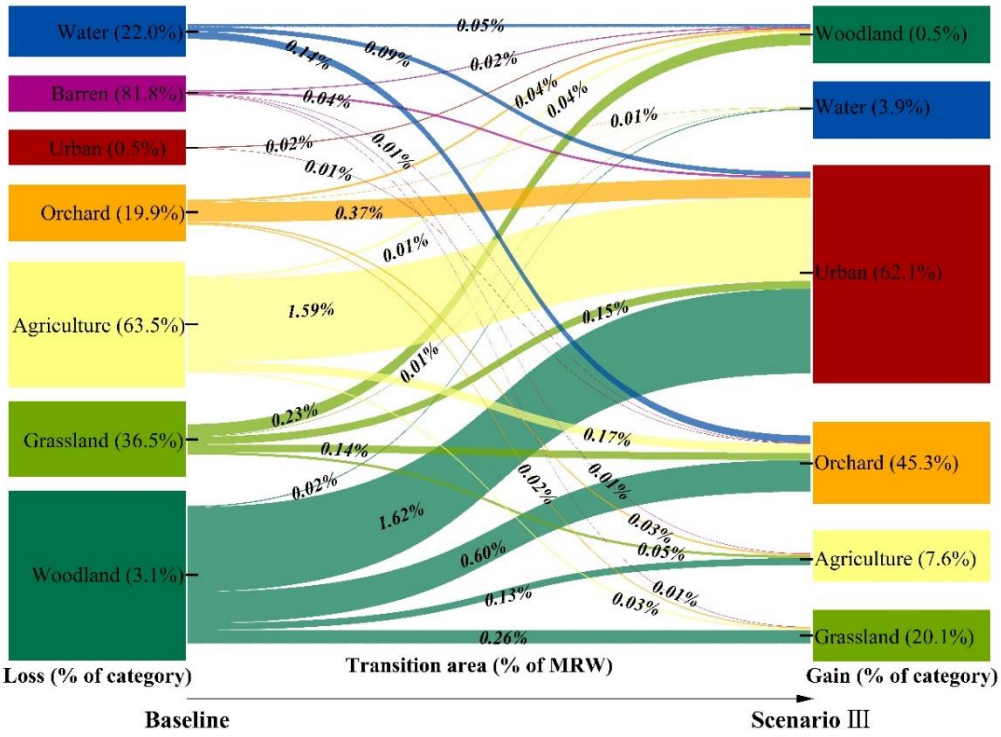


Fig 2-9 Patterns of land use change under Scenarios III

The transformation patterns of land use were also evaluated with Scenario II, which assumes that the social and economic factors will be improved with increasing population and GDP (Fig. 2-8). The results indicated that barren land may be more frequent, and approximately 9.1% of new barren areas were gained during this process. The process of urbanization will be more complex under scenario II, and it is found that approximately 0.5% of urban area may transfer to other types of land use.

Spatial planning of the related facilities may influence land use patterns in the MRW (Fig. 2-9). Compared with the transformation pattern under Scenario II, barren land may be less prominent under the assumption of Scenario III. There was no newly gained barren land under the assumption of Scenario III. Meanwhile, the intensity of agricultural activities may be less pronounced compared to Scenario III. The gained orchard and agricultural land were 45.3% and 7.6%, respectively, under Scenario III (Fig. 2-9). In contrast, approximately 46.2% and 9.3% of orchard and agricultural land may be gained under the assumption of Scenario II. In addition, the process of urbanization may still be complex under this scenario.

## 2.4 Discussion

### 2.4.1 LUCC in the MRW

Chinese cities are undergoing rapid anthropogenic disturbance, and the LUCC in China has drawn much attention recently (Huang et al., 2012; Zhou et al., 2020). The pattern of land use in the MRW during 2010-2020 was analyzed in this study (Table 2-4, Fig. 2-3). The land transformation has accelerated across the two time intervals, which is consistent with the accelerating economic development in Southeast China (Fig. S2-1). This finding is comparable to the results of previous studies in related area (Huang et al., 2012; Zhou et al., 2014). The change in grassland was pronounced at two time intervals, and the change was most prominent during 2015-2020. This result is similar with the trend observed in Asia (Shafizadeh-Moghadam et al., 2019). In this study, an efficient hybrid geospatial explicit approach was developed based on a RF model combined with CA and Markov models to understand patterns of land use in a large watershed located in Southeast China.

Although the CA-Markov model has been widely used in landscape and urban planning, the development of suitability maps for the model has not improved over the past decades (Fu et al., 2018). The RF model has the advantage of exploring the relationship between land use and local factors, especially for nonlinear relationships. The accuracy of the proposed method was relatively high, which was satisfactory for land use modelling. Compared to other land use categories, the accuracy for grassland and barren land was relatively low (Fig. 2-4). The uncertainty from input data, parameters, model structure, processes, and mathematical representation may influence the performance of land use models (Pontius et al., 2006; Ren et al., 2019; Lu et al., 2020). Recently, Lu et al. (2020) suggested that the percentage of land use may also affect the performance of the models. It was confirmed in this study, the total areas of the barren land and grassland in the MRW were less than other categories of land use, and data for training and testing for the RF models may not enough.

### 2.4.2 Driving factors for land use change

The evolution of land use is affected by regional environmental settings, socio-economic development, and related policies (Arsanjani et al., 2013; Lei et al., 2019). Land use models are valuable tools for understanding land use changes and identifying potential outcomes of policies or strategies (Van Vilet et al., 2019; Bacau et al., 2022). The importance of the driving factors influencing land use change was identified using the RF models (Fig. 2-5). It is indicated that the urban sprawl may be significantly affected by socio-economic factors. Population density and VIIRS nighttime lights were proven to be the major factors that influence the distribution of urbanized land in the MRW (Fig. 2-5). These results are consistent with the findings of previous studies (Huang et al., 2018; Lei et al., 2019; Zhou et al., 2020). Meanwhile, the topography was one of the most important factors influencing human-induced land use, and in this study, we also found that slope can highly influence the distributions of urbanized land and agricultural land use in the MRW.

Topographic factors such as elevation and slope gradient can reflect land productivity (Guan et al., 2019). The area with steep slopes may be more difficult to develop and use, and lower economic benefits may be achieved in these lands, which have a lower probability of change. Thus, natural or semi-natural land uses were likely to be found in these areas (Arsanjani et al., 2013; Lei et al., 2019; Guan et al., 2019). For example, Lei et al. (2019) indicated that the probability of fallow, forest, or orchard/garden would increase by 35.7%, 34.7%, and 50.3%, respectively, while the slopes would increase by 1 degree in a rural lowland catchment in Germany. Viana et al. (2021) also proposed that slope can be used as an explanation factor regarding agricultural land use, and different threshold values were found in the Beja district located in southern Portugal. In contrast, the urbanized land uses are usually observed in the area with flat terrain (Zhou et al., 2020; Bacau et al., 2022). In this study, we also found that the distribution of woodland was highly related to slope in the MRW.

### 2.4.3 Effects of spatial planning on the dynamics of land use change

The land use patterns were simulated under different scenarios in this study (Figs. 2-7, 2-8, and 2-9). Overall, the urbanized land may still be dynamic in these scenarios, and most of the newly gained urbanized areas were transferred from the agricultural land and woodland. Similar studies also found that the agricultural land may be replaced by the urbanized land to meet the demand of market in the former socialist countries such as Slovakia and Romania in Europe with changing policies (Pazur and Bolliger, 2017; Bacau et al., 2022). With increasing demands of market, the process of urbanization becomes more complex, some urbanized land may be lost and more of the woodland was transferred to the urbanized land under Scenarios II and III (Figs. 2-8, and 2-9). For the rapid economic development and high population in the highly urbanized area, the urbanized land may be increased in these areas for urban sprawl, while the urbanized land may be reduced in the rural area due to a decreasing population. The similar results were observed in related studies in China. For example, Zhou et al. (2020) found that the urban land area in the Chongming District, the rural area of Shanghai, experienced negative growth during 2015-2020 while the urban area of Shanghai was increasing continuously.

The spatial regional planning plays an important role in land use development (Zhou et al., 2020; Bacau et al., 2020). Comparing the scenarios of rapid economic development and high population (i.e., Scenarios II and III), we found that regional planning may reduce the effect of land degeneration during the process of urbanization. Although more urbanized land may have been gained under these scenarios, increased land abandonment was observed under Scenario II (Figs. 2-8, and 2-9). For the unconstrained development and limited facilities to meet the demands of people, the land use change patterns were more intense under Scenarios II, and more orchard, agricultural land, and barren land were gained compared with the results under

Scenarios III. On the one hand, for the limited facilities, the demands in urbanized area may increase, thus the agricultural and orchard may increase nearby. On the other hand, for unconstrained development in this area, more area may be developed without limitation; thus, more land may be abandoned and more barren land may be gained.

### 2.4.4 Outlook

To understand LUCC with changing environmental settings, the relationship between the driving factors and land use was identified using RF models, and suitable images for each category of land use were simulated. The improved CA-Markov model is more efficient for understanding land use patterns in different scenarios. However, the Markov model was built based on the historical land use images, which may have enhanced the uncertainty of the model. In the next step, we will attempt to improve the Markov model by considering the changing environmental setting to improve the accuracy of the model.

## 2.5 Conclusions

We developed a machine learning-based CA-Markov method to understand LUCC under different scenarios in the MRW. The proposed method exhibited an acceptable performance. Lower accuracy was observed in the grassland and barren land simulations. The land transformation has been accelerating in past decades, which is consistent with accelerating socio-economic development in the MRW. With the RF model, the most important driving factor for land use change was identified. The spatial planning may play a non-negligible role in land use evolution. This study provides an in-depth understanding of LUCC patterns using various strategies.

# Chapter 3 A method for detecting the non-stationarity during high flows under global change

**Zhenyu Zhang, Jinliang Huang, Paul D Wagner, and Nicola Fohrer**

**Science of the Total Environment, 2022, 851:158341**

## Abstract

The sustainability of existing water resources is influenced by extreme streamflow, and climate variability and human activities are generally the major factors controlling these dynamics. However, most of previously proposed methods to determine the effects of these factors have only been developed under the assumption of stationarity. Therefore, to overcome the existing research gap, an innovative method was proposed in this study to analyze and distinguish the effects of climate variability and human activities on extreme streamflow based on the non-stationarity theory. Accordingly, a rainfall-runoff model was developed using long-term hydrological data in the watersheds of Southeast China, which cover >75,000 km2. The model proposed in this study showed an acceptable performance, as indicated by the Nash-Sutcliffe efficiency coefficient (NSE), the Kling-Gupta efficiency (KGE), and percent bias (PBIAS). The NSE, KGE, and |PBIAS| were 0.67–0.75, 0.57–0.74, and 1.22–16.79 during the calibration periods, respectively. And the NSE, KGE, and |PBIAS| were 0.69–0.77, 0.65–0.76, and 0.98–17.51 during the validation periods, respectively. The trends of the extreme streamflow were analyzed for these watersheds at different time scales. The streamflow extremes at short time scales were found to be more sensitive to changing environment than those at longer time scales. The major factor controlling streamflow extremes at short time scales was human activities and climate change may be the dominant factor influencing streamflow extremes at long time scales. The findings of this study could provide useful insights into water management under global change conditions.

Keywords: Rainfall-runoff model; Streamflow; Generalized extreme value; Non-stationarity; Climate change

## 3.1 Introduction

Changes in the frequency and intensity of extreme hydrologic events worldwide threaten human life, infrastructure, and agriculture, thereby adversely impacting society and economy (Milly et al., 2008; Vinnarasi and Dhanya, 2019; UN WWDR, 2020). Hydrologic modeling is a useful tool to estimate the frequency and intensity of hydrologic extremes in the future (Ervinia et al., 2020; Zhang et al., 2020a). However, these models have been developed with stationary parameters. Models are calibrated with present available data, but applied for future projections (Wagner and Fohrer, 2019; Ervinia et al., 2020; Kiesel et al., 2020). It is increasingly recognized that the responses of hydrologic systems to changing environment needs to be better understood (Mentaschi et al., 2016; Pathiraja et al., 2016; Ouarda et al., 2019).

Methods to evaluate extreme events are crucial for assessing the risks that human beings are facing, and for water management (Ishak et al., 2013; Mentaschi et al., 2016). Hydrologic frequency analysis, which estimates the occurrence probability of rare hydrologic events by fitting a probability distribution to the data selected from the maxima in a defined period, is widely applied to design variables to an acceptable exceedance probability (Sraj et al., 2016; Luke et al., 2017; Bracken et al., 2018). Conventionally, most of these analysis methods are developed based on the hypothesis of independency and stationarity of streamflow events, assuming that the probabilistic distribution parameters and exceedance probability of events are static (Huang et al., 2013a; Zhang et al., 2014; Ouarda et al., 2019). Consequently, most hydrologic modeling practices have been performed under this assumption (Mentaschi et al., 2016; Sadegh et al., 2019). However, an assumption that the change in extreme events is stationery has been challenged by the fact that extreme events are changing in the past decades and are likely to change in the future (Sraj et al., 2016; Ceres et al., 2017; Sadegh et al., 2019). Additionally, The stationary models may misestimate flood quantiles relative to the non-stationary models and many studies highly recommend the consideration of a non-stationary framework to enhance the flood management in Europe (Sraj et al., 2016; Hesarkazzazi et al., 2021). Therefore, it is necessary to quantify extreme hydrologic events under the assumption of non-stationary.

Non-stationarity of hydrologic events can be attributed to local anthropogenic impacts, such as land use change and damming, or to the global climate change (Milly et al., 2008; Zhang et al., 2015; Ouarda et al., 2019; Cui et al., 2020). Related studies indicated that extreme precipitation events are projected to be more frequent and intense, exceeding known historical records, and thereby providing strong evidences that extreme hydrologic events are closely related to global climate change (Luke et al., 2017; Sarhadi et al., 2018; Sun et al., 2018). The distribution of hydrologic extremes is likely to be modified in the context of climate change (Nasri et al., 2017; Li et al., 2018a; Carney et al., 2020). Additionally, human activities have

increasingly influenced the streamflow regime and extreme hydrologic events (Zhang et al., 2015; Zhang et al., 2020a). Human land use has a strong impact on water balance components and it has been identified as a major contributing factor to non-stationarity in rainfall-runoff relationships (Wagner et al., 2016; Deb et al., 2019; Wagner et al., 2019; Aghsaei et al., 2020).

Furthermore, the interaction between local human activities and global climate change made this non-stationarity more complex (Li et al., 2018b; Ervinia et al., 2020). For example, in the developing regions, urban expansion may be the major factors to modify the change patterns of the local climate. A study in the Yangtze River Delta found that the recurrence levels of the 1-day and 5-day maximum precipitation increased by 25.9 % and 59.1 % for highly urbanized areas, 34.2 % and 36.9 % for slightly urbanized areas, and 30.7 % and 61.5 % for rural stations, respectively (Lu et al., 2019). In contrast, the climate change may increased the risk of flood or high flow events in the developed regions. Some studies indicated that the increasing magnitude of extreme precipitation may increase the non-stationarity and lead to more uncertainties of flooding. (Tromel and Schonwiese, 2007; Sraj et al., 2016; Huo et al., 2021). Sraj et al. (2016) found that a 10 % increase in annual precipitation, the 10-year flood increases by 8 % in a watershed of Slovenia. Although significant progress has been made in understanding hydrologic extremes, the role of local or global forcing on the frequency and intensity of high flows remains uncertain.

The coastal watersheds in southeast China had been suffering from increased frequencies of climate events (e.g., typhoons) and drastic human activities (e.g. accelerated land use changes) (Huang et al., 2013a; Huang et al., 2013b; Zhang et al., 2020a). To understand the vulnerability induced by extreme hydrology events, the non-stationary patterns of high flows was identified in this study. To achieve this goal, a rainfall-runoff hydrology model coupled with the generalized extreme value (GEV) model were developed. The objectives of this study were: (1) to evaluate the occurrence of non-stationary patterns in high flows, (2) to identify the impacts of climate change and human activities on high flows, and (3) to provide evidences for management of water-related extremes. This study provides new insights into the non-stationary of high flows in the context of human activities and climate changes.

## 3.2 Materials and methods

### 3.2.1 Study area

Minjiang River Watershed and Jiulong River Watershed are the two largest watersheds in Southeast China located west of the Taiwan Strait, in the subtropical zone with a subtropical monsoon climate (Fig 3-1). The Minjiang River Watershed covers 60,992 $km^2$ and the annual average precipitation is 1617 mm, of which 70 % occurs between April and September. The average annual discharge of the Minjiang River Watershed is 1980 $m^3/s$ (Zhang et al., 2019; Huang et al., 2020). The Jiulong River Watershed is the second largest river in the Fujian

Province and it covers an area of approximately 14,700 $km^2$. The average discharge of the Jiulong River Watershed is 380 $m^3/s$ (Huang et al., 2013a; Zhang et al., 2015). These watersheds supply water to >35 million residents, households, and industries, and provide resources for agricultural irrigation (Zhang et al., 2015; Zhang et al., 2019; Huang et al., 2020). As a result of population increase and economic growth, the areas of urban increased significantly after 1980s (Huang et al., 2010; Huang et al., 2012a). Simultaneously, the demands of water resource increased. For example, hundreds of dams or reservoirs were build in the in the Jiulong River Watershed (Zhang et al., 2015).

### 3.2.2 Data sources and processing

The daily river discharge data of nine hydrological stations in two major watersheds in the Fujian Province (i.e., Jiulong River Watershed and Minjiang River Watershed; Fig. 3-1 and Table 3-1) were obtained from the Hydrology and Water Resources Bureau. Meteorological data with 0.5 °×0.5 °spatial resolution from 1961 to 2012, including mean daily air temperature, maximum daily air temperature, minimum daily air temperature, and daily precipitation, were obtained from the China Meteorological Administration (http://data.cma.cn). The potential evapotranspiration was estimated based on the Hargreaves method (Hargreaves et al., 1985; Hargreaves and Allen, 2003). The climate variability trends of these watersheds are shown in Fig 3-2, Fig 3-3, and Fig 3-4.

### 3.2.3 Streamflow reconstruction

A conceptual rainfall-runoff hydrology model was developed in this study based on Jackson-Blake et al. (2017). The model was calibrated for the baseline (i.e., streamflow generated under the environmental setting of the 1980s), and the baseline streamflow (i.e., streamlfow re-constructed based on the environmental setting of baseline) after the 1980s were reconstructed by only changing the meteorological inputs of the model. The surface runoff, soil water, and groundwater were considered as input variables in this model (Fig S3-1), and they were estimated as follows:

$$Q_s = f \cdot P \tag{1}$$

where $\boldsymbol{Q_s}$ is the surface runoff, $\boldsymbol{f}$ is the proportion of precipitation that exceed the infiltration, and $\boldsymbol{P}$ is the precipitation.

$$Q_{soil} = (V_{soil} - fs) \cdot \frac{fl}{T_s} \quad (if \ V_{soil} - fs) \tag{2}$$

$$\frac{dV_{soil}}{dt} = (1-f) \cdot P - alpha \cdot PET \cdot (1 - e^{\mu V_{soil}}) - Q_{soil} \tag{3}$$

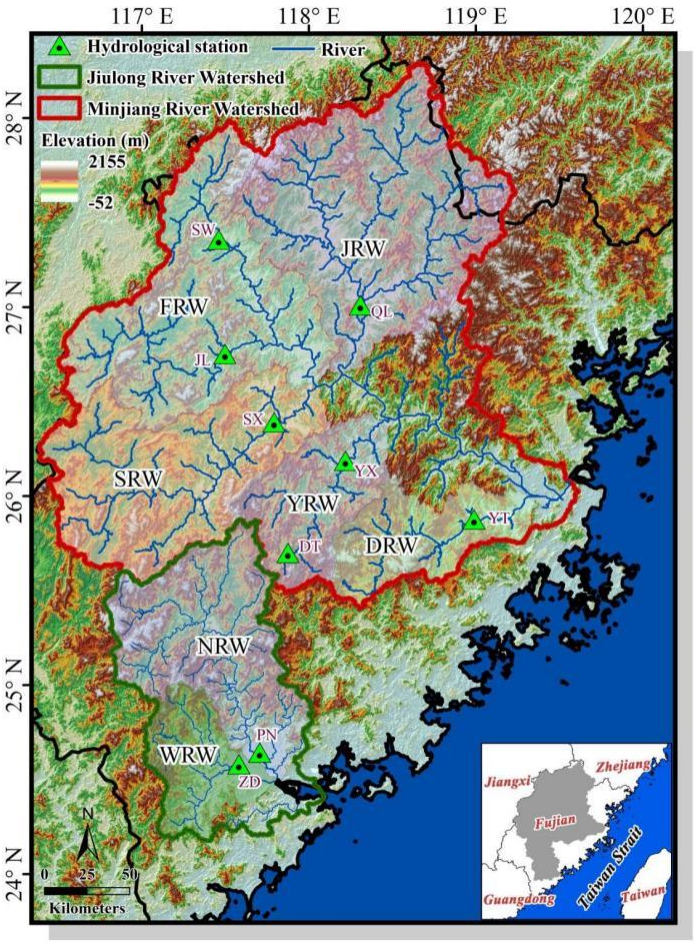


Fig 3-1 Map of the study area

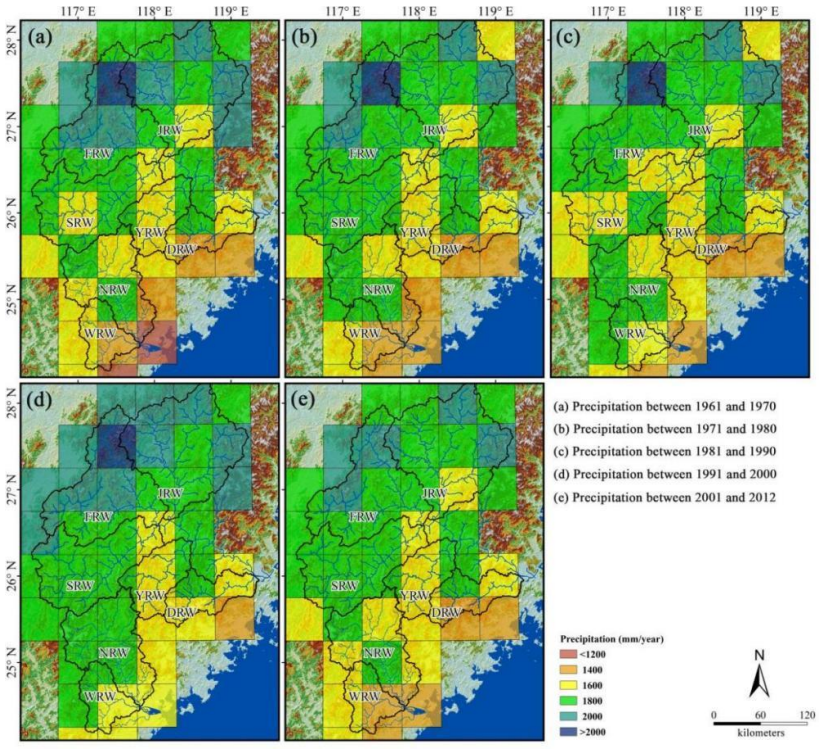


Fig 3-2 Precipitation in the study area

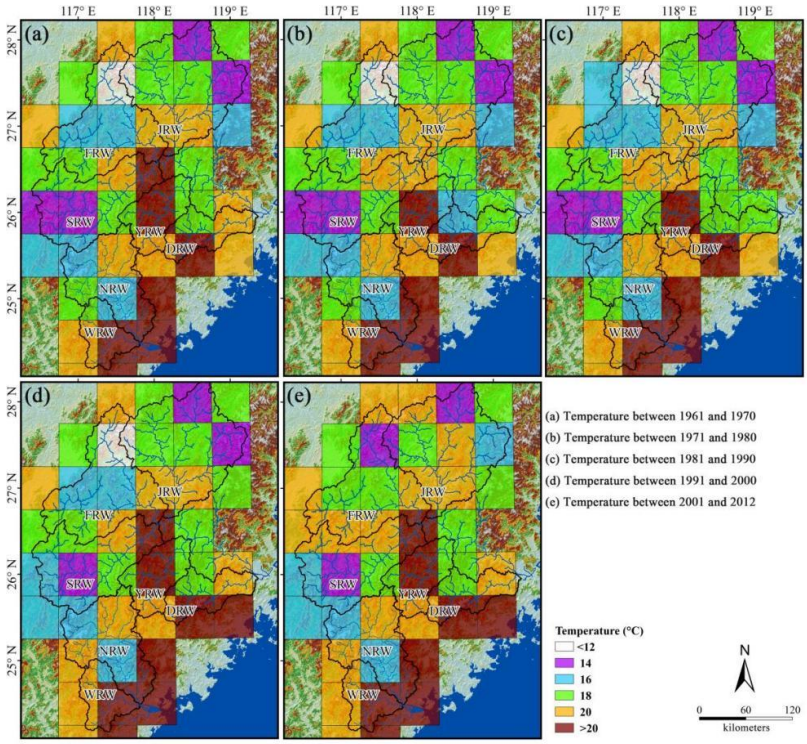


Fig 3-3 Temperature in the study area

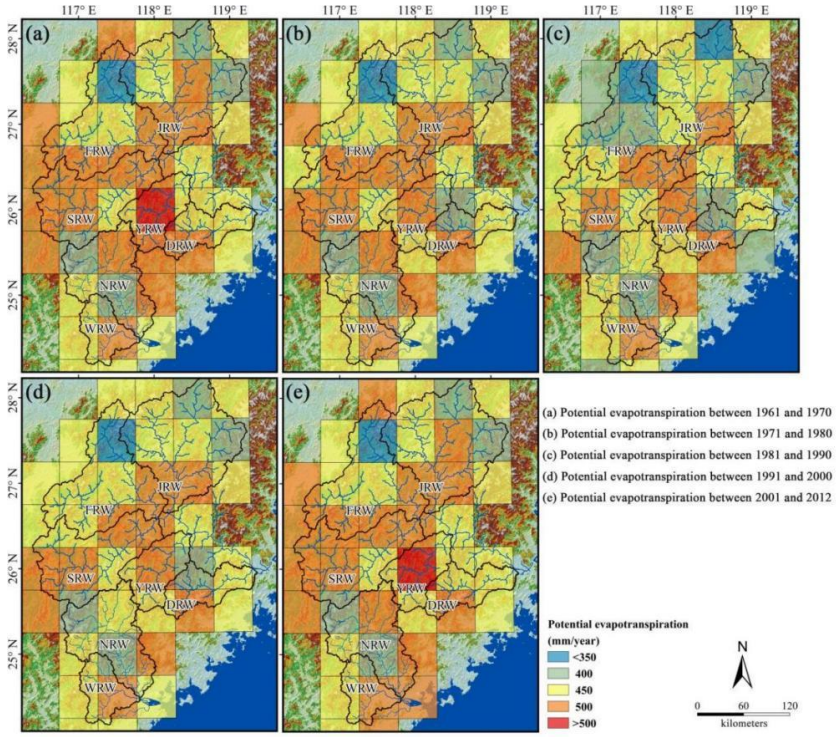


Fig 3-4 Potential evapotranspiration in the study area

Table 3-1 Description of the hydrology stations

| Name | Hydrology station | Watershed size ($km^2$) | Time period | Length (m) |
|---|---|---|---|---|
| West River Watershed (WRW) | Zhengdian (ZD) | 3491.37 | 1961-2012 | 239430 |
| North River Watershed (NRW) | Punan (PN) | 8458.72 | 1961-2012 | 509131 |
| Shaxi River Watershed (SRW) | Shaxian (SX) | 9918.13 | 1965-2012 | 549756 |
| Jianxi River Watershed (JRW) | Qilijie (QL) | 14816.17 | 1961, 1963-1968, 1970-2012 | 647260 |
| Futunxi River Watershed (FRW) | Jiangle (JL) | 5995.75 | 1961, 1965-1966, 1970-2012 | 531577 |
| | Shaowu (SW) | 2712.34 | 1961, 1963-1967, 1969-1972, 1976-2000, 2003-2012 | 205356 |
| Youxi River Watershed (YRW) | Datian (DT) | 370.94 | 1962-1977 | 13865 |
| | Youxi (YX) | 4075.33 | 1984, 1990-2012 | 321146 |
| Dazhangxi River Watershed (DRW) | Youtai (YT) | 3948.38 | 1961, 1963-2012 | 375344 |

where $Q_{soil}$ is the soil water flow, $V_{soil}$ is the soil water volume, ***fs*** is the soil field capacity, ***fl*** is the threshold between soil and groundwater flow, ***Ts*** is the soil water time constant, ***PET*** is the potential evapotranspiration, ***alpha*** is the correction factor for the potential evapotranspiration, and $\boldsymbol{\mu}$ is a parameter that determines the shape of the curve that link the relationship between evapotranspiration and soil water content (Fenicia et al., 2011). .

$$\frac{dV_g}{dt} = beta \cdot Q_{soil} - Q_g \tag{4}$$

$$Q_g = \frac{V_g}{T_g} \quad \text{or} \quad \frac{dQ_g}{dV_g} = \frac{1}{T_g} \tag{5}$$

$$\frac{dQ_g}{dt} = \frac{dV_g}{dt} \cdot \frac{dQ_g}{dV_g} = (beta \cdot Q_{soil} - Q_g) \cdot \frac{1}{T_g} \tag{6}$$

where $V_g$ is the groundwater volume, ***beta*** is the baseflow index, $T_g$ is the groundwater time constant, and $Q_g$ is the groundwater flow.

We argue that physical changes in a watershed transform the underlying processes, which are manifested in the temporal variation of parameters in this conceptual model. The comparison method was employed to understand the non-stationary of streamflow in two periods (i.e. baseline period and impacted period ) with the assumption that the human activities were inactive in the baseline period and were active in the impacted period. Thus, the parameters for conceptual model would not change too much during this period. The model was calibrated and validated using the daily streamflow data by comparing the simulation results with field measurements. The periods for calibration and validation of the model were 1961-1969 and 1970-1977, respectively. The Nash-Sutcliffe efficiency coefficient (NSE) (Nash and Sutcliffe, 1970), percent bias (PBIAS), and Kling-Gupta efficiency (KGE, Gupta et al., 2009) were used to evaluate the performance of the model. PBIAS is a statistical metric providing an estimate of over prediction (PBIAS>0) or under prediction (PBIAS<0) of the model. They were calculated as follows:

$$NSE = \frac{\sum_{i=1}^{n} (Q_{obs} - Q_{cal})^2}{\sum_{i=1}^{n} (Q_{obs} - Q_{\text{ave}})^2} \tag{7}$$

$$PBIAS = 100 * \frac{\sum_{i=1}^{n} (Q_{cal} - Q_{obs})}{\sum_{i=1}^{n} Q_{obs}} \tag{8}$$

where $Q_{obs}$, $Q_{cal}$, $Q_{ave}$, and $n$ are the measured data, simulated data, the average, the number of days.

$$KGE = 1 - \sqrt{(r-1)^2 + (a-1)^2 + (b-1)^2} \quad (9)$$

where $r$, $a$, and $b$ are the linear correlation between observations and simulations, the measure of the flow variability error (i.e. the ratio between the standard deviation of simulated flow and the standard deviation of observed flow), and the bias term respectively (i.e. the ratio between the mean simulated and mean observed flow). Knoben et al. (2019) reported that the KGE scores between -0.41 and 1.0 can be considered as satisfactory model performance. The index of agreement (d) (Willmott 1981) was employed to evaluate the degree of model prediction error. The agreement value of 1 indicates a perfect match, and 0 indicates no agreement at all.

### 3.2.4 GEV model

The GEV distribution is one of the most important extreme distribution functions applied in extreme theory, and has been widely used worldwide (Ishak et al., 2013; Zhang et al., 2014; Zhang et al., 2018; Ragno et al., 2019). The cumulative distribution function for the non-stationary GEV model is defined as follows (Towler et al., 2010):

$$\mathrm{F}\left(x;\mu(\theta),\sigma,\xi\right) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu(\theta)}{\sigma}\right)\right]^{-1/\xi}\right\}, 1+\xi\left(\frac{x-\mu(\theta)}{\sigma}\right) > 0 \quad (10)$$

where $x$ is the annual extreme value; $\mu(\theta)$ is the location parameter in the non-stationary GEV model; $\theta$ is the covariate of the linear function of the location parameter; and $\sigma$ and $\xi$ are the scale and shape parameters of the GEV model, respectively. The employed GEV method is based on a study by Ragno et al. (2019), named Process-informed Nonstationary Extreme Value Analysis (ProNEVA), in which the non-stationarity component can be defined by the temporal explanatory variable. Using the ProNEVA approach, the distribution parameters were estimated by the Bayesian inference approach and the posterior distribution function of the parameters was sampled with a newly-developed hybrid evolution Markov Chain Monte Carlo approach (Ragno et al., 2019).

The precipitation in Southeast China has increased in the past, and may continue to increase in the future (Huang et al., 2013a; Zhang et al., 2015; Ervinia et al., 2020). Moreover, the intense human activities (including land use change) after the 1980 may have modified the streamflow regime in this region (Huang et al., 2012; 2013a; 2018). In addition, patterns of hydrological may be changed at different scales (Tu et al., 2017; Sun et al., 2018). Therefore, in this study, the 3-day maximum streamflow (annual maxima of 3-day means; Sm3day), 7-day maximum streamflow (Sm7day), 30-day maximum streamflow (Sm30day), and 90-day maximum streamflow (Sm90day) were used as the annual extreme values for GEV models.

These indices of the flow statistics were calculated according to previous studies (Zhang et al., 2015; Zhang et al., 2020a).

### 3.2.5 Method to distinguish the non-stationary of climate change and human activities

Theoretically, the hydrologic extremes can be extrapolated with probability density function or cumulative distribution function with observed maxima or minima. According to the assumption of the non-stationary theory, the extrapolated extreme can vary temporally and the return level can be estimated annually. This is also referred to as the effective return level (Cheng et al., 2014). The parameters in the GEV distribution can change temporally, thus facilitating the efficient detection of extreme trends within this framework (Bennett et al., 2015; Zhang et al., 2018; Steirou et al., 2019). In this study, we developed GEV models under two conditions, including the baseline streamlfow (i.e., streamflow re-constructed with the environmental setting of the baseline, which was influenced by the changed climate conditions in 1978-2012) and impacted streamflow (i.e., observed streamflow, which was influenced by the changed climate conditions and human activities). Distribution curves for the effective return levels for the two scenarios were then constructed. The slopes of the return period at the specific scale were further estimated to quantify the impacts from the related factors (Lu et al., 2019) (i.e., α and β; Fig 3-5). The increased slope (i.e., positive value) indicated that the magnitude of the streamflow extreme may increase in the future, and the decreased slope (i.e., negative value) indicated that the magnitude of the streamflow extreme may decrease in the future.

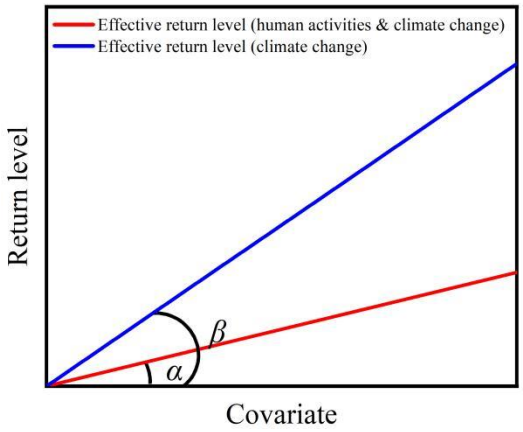


Fig 3-5 Method to identify the non-stationarity of hydrologic extremes associated with climate and human forcing

Then, the contribution of climate change ($\boldsymbol{E_c}$) and human activities (including the coupled effect of climate change and human activities) ($\boldsymbol{E_h}$) can be estimated as follows:

$$E_c = \frac{\beta}{\alpha} \cdot 100\% \tag{11}$$

$$E_h = \frac{\alpha - \beta}{\alpha} \cdot 100\% \tag{12}$$

## 3.3 Results

### 3.3.1 Model calibration and validation

The calibration and validation results are shown in Fig 3-6. The simulated discharge fitted well with the observed values, with an NSE over 0.65, |PBIAS| less than 20, and KGE over 0.60 except for the DT station (Table 3-2), suggesting that the models reasonably reflected the characteristics of the rainfall-runoff processes in these watersheds during the baseline period, and thus can be used to evaluate future changes with the environmental setting in this period. Additionally, the simulated flow duration curves fitted well with the observed curves (d=0.967-0.989), which indicated that the models can perform well for the maxima and minima streamflow (Fig 3-6).

### 3.3.2 Patterns of hydrologic extremes

The time series of maximum streamflow at different scales for the period from 1978 to 2012 are illustrated in Fig 3-7. The alteration of the red to blue from bottom to top indicated that the maximum streamflow decreased with a longer time scale. Moreover, high variabilities were observed in the bottom area of the time series, such as for Sm1day and Sm3day, which indicated that these indicators may be more sensitive to the changing environment. In contrast, low variabilities were observed for Sm90day and Sm30day, and this might suggest that these indicators were less sensitive to the changing environment. In addition, similar patterns of the annual maxima streamflow were observed in adjacent hydrology stations, such as the ZD and PN stations which are located in the Jiulong watershed.

According to the non-stationary GEV analysis, the return level curves of the annual maxima streamflow for the different scales (Fig 3-8) and the observed data fitted well with the maximum likelihood return level curve. All the stations exhibited similar return level patterns (Fig 3-8).

Table 3-2 Nash-Sutcliffe efficiency coefficient (NSE), Kling-Gupta efficiency (KGE), and percent bias (PBIAS) of the rainfall-off model in streamflow simulation

| Station | Calibration | | | Validation | | |
|---|---|---|---|---|---|---|
| | NSE | KGE | PBIAS | NSE | KGE | PBIAS |
| ZD | 0.67 | 0.72 | 1.92 | 0.71 | 0.72 | 1.83 |
| PN | 0.72 | 0.73 | 11.26 | 0.74 | 0.76 | 9.42 |
| SX | 0.72 | 0.57 | 16.61 | 0.77 | 0.74 | 14.89 |
| QL | 0.71 | 0.63 | -6.81 | 0.75 | 0.65 | -7.46 |
| JL | 0.75 | 0.63 | 16.79 | 0.69 | 0.68 | 17.51 |
| SW | 0.67 | 0.74 | 1.22 | 0.72 | 0.70 | 0.98 |
| DT | 0.50 | 0.58 | 25.86 | 0.52 | 0.62 | 18.36 |
| YT | 0.70 | 0.70 | 8.99 | 0.74 | 0.74 | 2.49 |

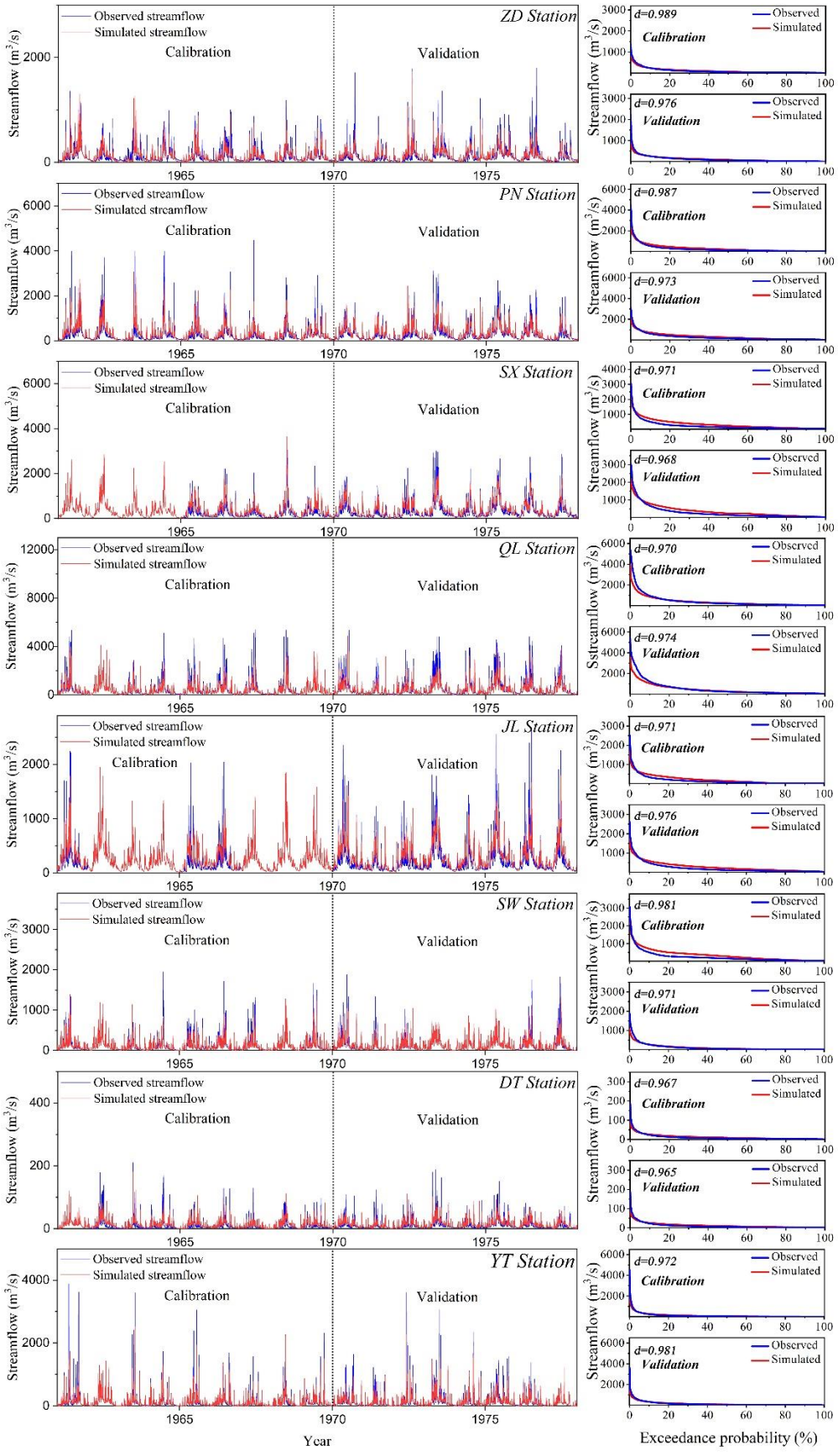


Fig 3-6 Daily streamflow calibration and validation for the rainfall-runoff hydrology model

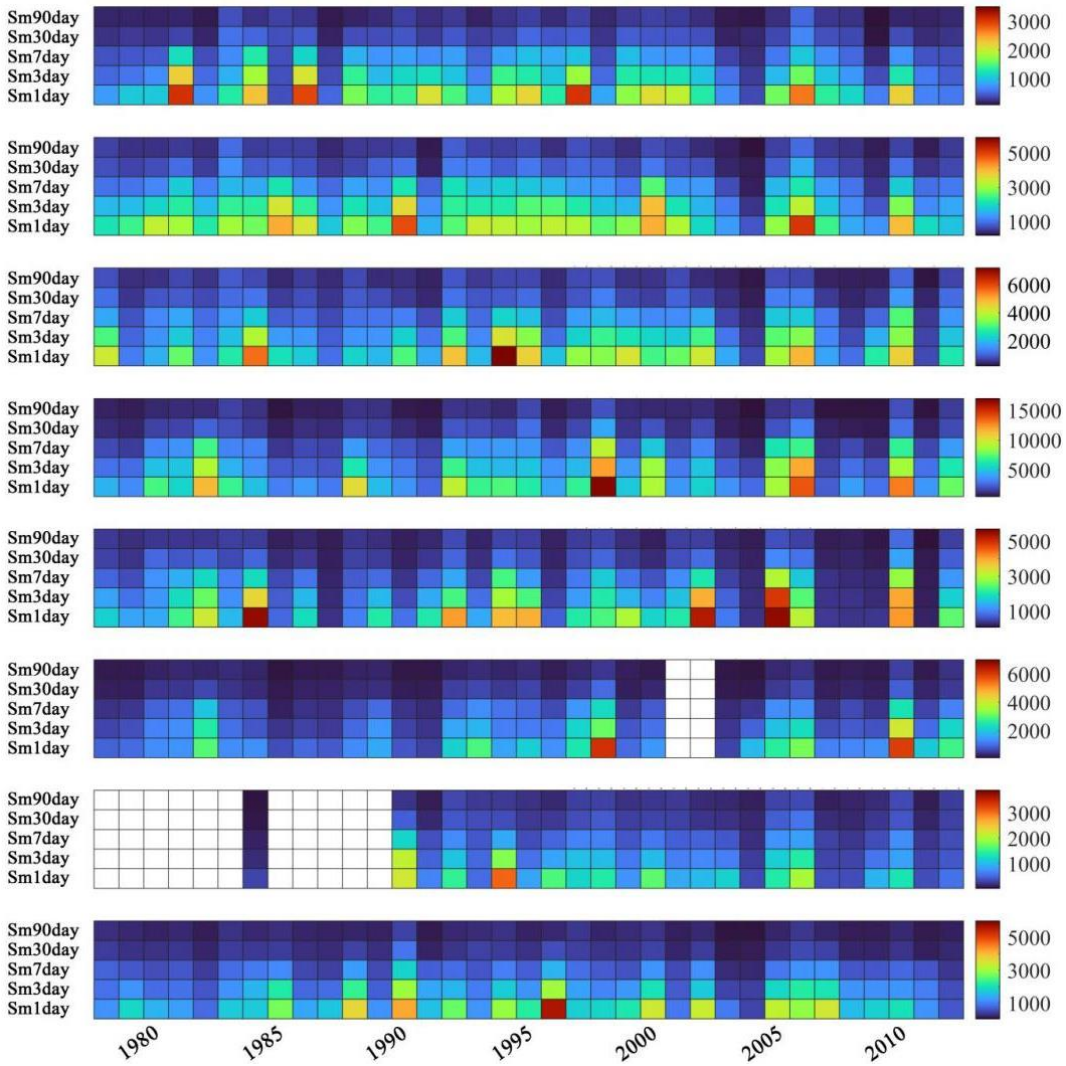


Fig 3-7 Annual maxima streamflow of the study area

(Note: from up to down: ZD, PN, SX, QL, JL, SW, YX, and YT)

## 3.3.3 Impacts of climate change and human forcing

The non-stationary GEV models of the hydrology stations were set up for the different streamflow scenarios, and accordingly, the effects of human activities and climate change were identified (Fig 3-9). Except for the SW and YX stations, the return levels of the annual maxima streamflow might be reduced in the future because of human activities. Climate variability evidently had limited impacts on the annual maxima streamflow in the YT station which might be due to the small watershed area and low variability of climate change (Table 3-1, Fig 3-2, Fig 3-3, and Fig 3-4).

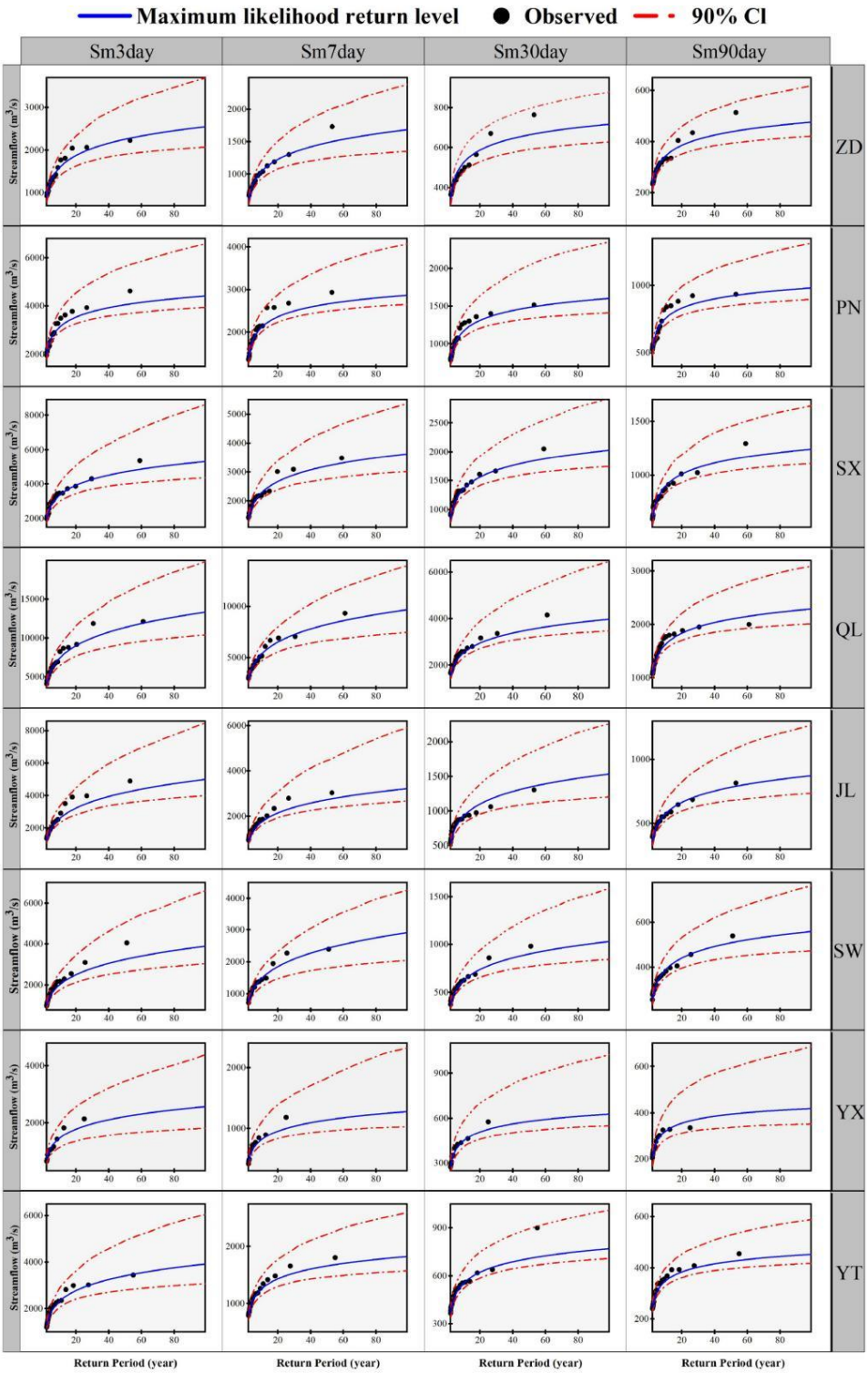


Fig 3-8 Return level of the annual maxima streamflow

Furthermore, the quantified effects of human activities and climate change are presented in Table 3-3. Human activities might alter the patterns of annual maxima streamflow in terms of the Sm3day and Sm7day for the small watersheds (i.e., ZD, SW, and YT stations). The $E_h$ of Sm3day and Sm7day at the ZD, SW, and YT stations, with watershed sizes less than 3500

km$^2$, were 177-208% and 114-214%, respectively. Except for the YT station, climate change was the major factor that influenced the patterns of the annual maxima for the different stations (Table 3-3).

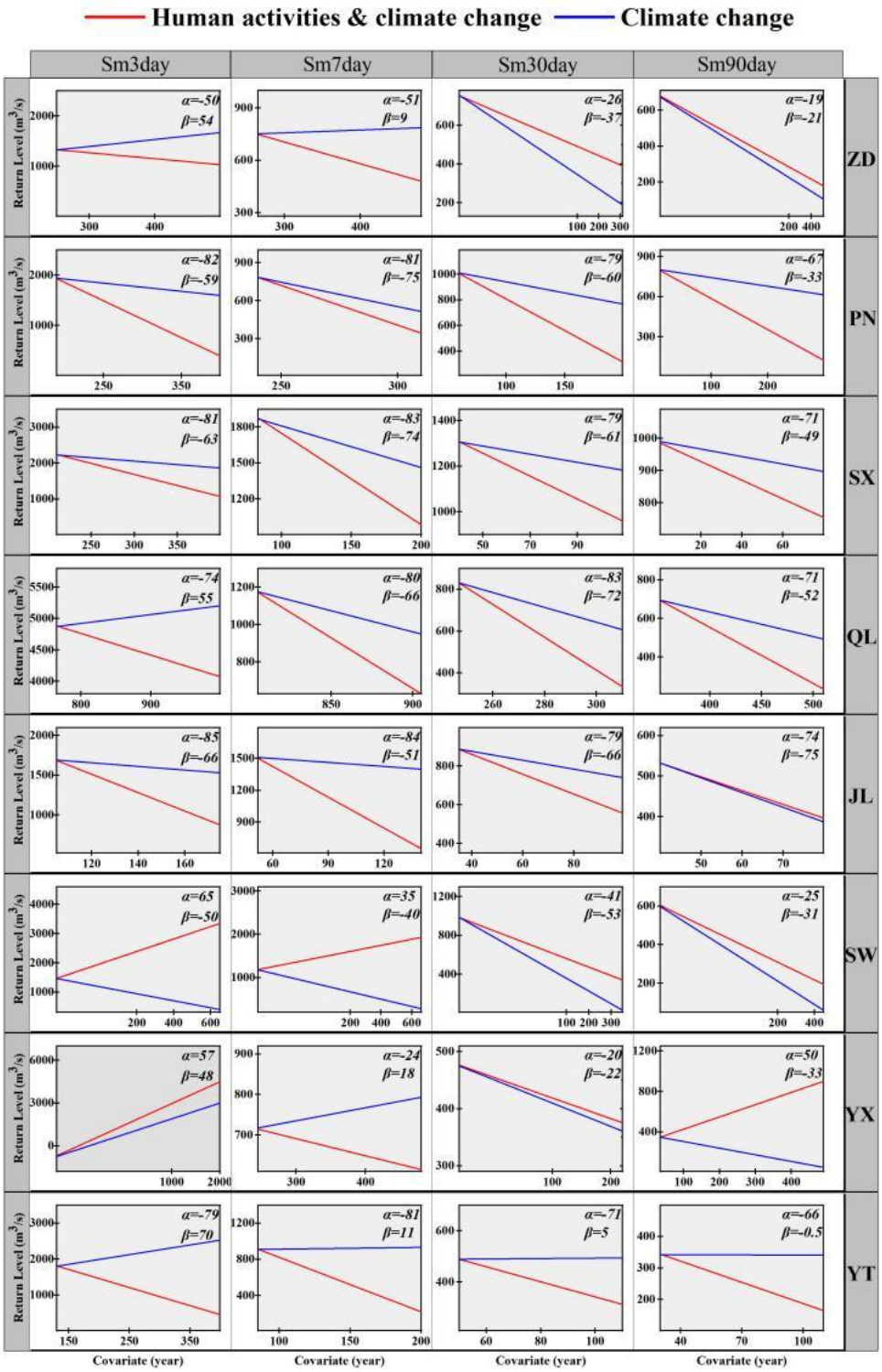


Fig 3-9 Time series of the effective return levels for the annual maxima streamflow based on the non-stationary generalized extreme value (GEV) model (10-year return level)

Table 3-3 Effects of human activities and climate change (%)

| Station | Index | $E_h$ | $E_c$ | Station | Index | $E_h$ | $E_c$ |
|---|---|---|---|---|---|---|---|
| ZD | Sm3day | 208 | -108 | JL | Sm3day | 22 | 78 |
| | Sm7day | 118 | -18 | | Sm7day | 39 | 61 |
| | Sm30day | -42 | 142 | | Sm30day | 16 | 84 |
| | Sm90day | -11 | 111 | | Sm90day | -1 | 101 |
| PN | Sm3day | 28 | 72 | SW | Sm3day | 177 | -77 |
| | Sm7day | 7 | 93 | | Sm7day | 214 | -114 |
| | Sm30day | 24 | 76 | | Sm30day | -29 | 129 |
| | Sm90day | 51 | 49 | | Sm90day | -24 | 124 |
| SX | Sm3day | 22 | 78 | YX | Sm3day | 16 | 84 |
| | Sm7day | 11 | 89 | | Sm7day | 175 | -75 |
| | Sm30day | 23 | 77 | | Sm30day | -10 | 110 |
| | Sm90day | 31 | 69 | | Sm90day | 166 | -66 |
| QL | Sm3day | 174 | -74 | YT | Sm3day | 189 | -89 |
| | Sm7day | 18 | 83 | | Sm7day | 114 | -14 |
| | Sm30day | 13 | 87 | | Sm30day | 107 | -7 |
| | Sm90day | 27 | 73 | | Sm90day | 99 | 1 |

Note: $E_h$ and $E_c$ are the effect of human activities and climate variability, respectively.

## 3.4 Discussion

### 3.4.1 Applicability of the rainfall-runoff model

This study assumed that the marginal distribution of streamflow regimes follows a GEV distribution, where the distribution parameters can vary temporally. However, the streamflow regimes may be controlled by climatic factors (including precipitation, temperature, and weather generator parameters) and human-induced factors (including land use/cover change and dam construction) in the watersheds (Huang et al., 2013a; Zhang et al., 2015; Steirou et al., 2019; Aghsaei et al., 2020). Hydrological models are useful tools for understanding the impacts of climate variability on watersheds (Deb et al., 2019; Ervinia et al., 2020; Zhang et al., 2020a). To identify the streamflow regimes controlled by climate change, the streamflow of the study area was reconstructed, which eliminated the interference of human activities after the 1980s, by only changing the meteorological inputs of the watersheds. The model structure employed in the study was found to be a viable tool to simulate the rainfall-runoff process for the large-scale watershed.

Previous studies in the study area employed models with more complex structures (Huang et al., 2013a; Zhang et al., 2019; Zhang et al., 2020a). Although the assumptions of the rainfall-runoff model proposed in this study were simpler than the popular process-based models, that is, Soil & Water Assessment Tool (SWAT) and Hydrological Simulation Program-FORTRAN (HSPF), the performance of the proposed model was comparable to those models in previous studies (Table S3-1). For example, the NSE for the daily streamflow simulation of SWAT in the NRW during the calibration and validation periods were 0.64 and 0.60 while those were 0.72 and 0.74 respectively, using the rainfall-runoff model in this study (Table 3-2, and Table S3-1).

The development of watershed models may be highly complex under certain environmental settings (Jackson-Blake et al., 2017). In this study, this conclusion for watershed modeling was further confirmed by comparing the results of this model with the other models employed in previous studies in Southeast China (Table S3-1). Three major runoff generation processes were considered in this study, and both the models performed similarly during calibration and validation in the large watersheds. Although the process-based models, including SWAT and HSPF, better represent watershed processes, such as land use and soil condition, these might also intensify the uncertainty in the structure of the model due to insufficient or parsimonious inputs of models for the large watersheds. Compared to the performance on small watersheds, the model exhibited better performance on the large watershed because the effects of some processes (e.g., land use changes and dam construction) can be neglected in such cases (Zhang et al., 2020a).

Compared with the combination methods for estimating potential evapotranspiration (e.g., Penman-Monteith method), the temperature-based methods have been widely applied in water energy balance analysis because of its low data requirements and better correlation with temperature, which would be beneficial for future projections (Weiland et al., 2012; Seong et al., 2018). Some researchers have attempted to adapt these methods to estimate daily potential evapotranspiration. Aguilar and Polo (2011) indicated that the Hargreaves method can be used to study the water and energy balance on a daily scale after correction. For example, Weiland et al. (2012) proposed a revised Hargreaves method to estimate daily global reference potential evapotranspiration. In this study, the potential evapotranspiration was further corrected to simulate the soil runoff, and the models were acceptable (Table 3-2). Some models have employed the similar methods to correct potential evapotranspiration to satisfy requirements (e.g., Jung et al., 2016).

Obtaining representative meteorological data for watershed-scale hydrological modeling can be difficult and time-consuming (Fuka et al., 2014). Grid-based meteorological data are becoming an important source for driving hydrological models because of their high temporal resolution and reliability (Wang et al., 2021b; Tan et al., 2021). Fuka et al. (2014) also indicated

that grid-based data such as CFSR data (~0.5 ° resolution) can be as good as the traditional weather gauging stations in small watersheds (22.1-1200 $km^2$). In this study, we used grid-based meteorological data as the driving force for the rainfall-runoff model to understand streamflow regime in Southeast China. It required less time due to the less data acquired to establish the model. Although the performance of this were lightly poorer in the small sub-watersheds, for example the NSE of simulated streamflow during the calibration and validation periods were 0.50 and 0.52 at DT station (390 $km^2$) (Table 3-2 and Table S3-1). In this study, we also compared the performance of the rainfall-runoff model with different driving forces (Fig S3-2), the results shown that the performance of grid-based meteorological data can be as good as the local stations. Compared to the performance of related studies (e.g. Fuka et al., 2014), our model is acceptable.

### 3.4.2 Hydrologic extremes in Southeast China

Streamflow regimes are generally controlled by climate variability such as precipitation, temperature, evapotranspiration, soil moisture (Huang et al., 2013a; Zhang et al., 2015), and they may be effective indicators that identify regional change (Wang et al., 2013; Huang et al., 2014; Zhang et al., 2018). Moreover, changes in precipitation amplify changes in streamflow, as observed for the past decades across the Asian Pacific region (Chen et al., 2007; Huang et al., 2014; Zhang et al., 2018). However, this trend would be altered due to intensive human activities. For example, a different trend was observed in Southeast China. Owing to global change, an amplified runoff response to rainfall was observed at the eastern Taiwan Strait, China (Huang et al., 2014). However, this trend was rarely observed in the Jiulong River Watershed, which is located at the western Taiwan Strait China (Huang et al., 2013a).

In this study, these conclusions were further confirmed by analyzing the patterns of streamflow extremes across eight hydrologic stations in Southeast China. These differed from each other even though they exhibited similar patterns of climate variability since the past decades (Figs 3-2, 3-3, 3-4, and 3-7). For example, the streamflow extremes of West River Watershed (WRW; ZD station) were evidently more sensitive to the climate variability than the DRW (YT station), even though they exhibited similar climate variability and watershed characteristics (Table 3-1). In addition, a study in the Pearl River, which is near the present study area, indicated that the response of extreme low flows to climate change would vary with different time scales and the annual minima with a short time scale may be more sensitive to climate change based on the observed streamflow in the past decades (Tu et al., 2017). A similar result was obtained in this study, as compared to shorter time scales, a lower variability was observed in the streamflow extreme for the annual maxima streamflow (e.g., 1-day maximum streamflow; Fig 3-7). Thus, the amplified streamflow according to precipitation may be more evident for shorter time scales.

The joint return periods for streamflow extreme is one of the most important factors that should be considered in the field of water management (Bennett et al., 2015; Vinnarasi and Dhanya, 2019). However, the stationary model may underestimate streamflow quantiles under climate change or overestimate streamflow quantiles because of human activities relative to the non-stationary models (Zhang et al., 2015; Sraj et al., 2016; Liu et al., 2021b). The hydrological return period must be carefully reformulated and extended under non-stationary conditions due to the time-varying aspect of climate extremes as well as the changed characteristics of watersheds that may be induced by human activities (Sraj et al., 2016; Lu et al., 2019). The return levels of the annual maxima streamflow were also evaluated in this study based on the non-stationary theory, and all the stations exhibited similar patterns in terms of the return level and the uncertainty, which might increase in the high return period (e.g., 100-year scale). Thus, more attention should be paid to the non-stationarity of the high return period level of the streamflow extreme, when considering a hydraulic project in the watershed.

### 3.4.3 Impacts of climate change and human forcing

During the last decades, numerous theories and conceptions have been proposed to evaluate the impacts of human activities and climate change, including hydrological indicators (Richter et al., 1996), the elasticity of rainfall-runoff (Sankarasubramanian et al., 2001), ecohydrological analysis methods (Tomer and Schilling, 2009), and other related methods (Zhang et al., 2015; Huang et al., 2014; Zhang et al., 2020a). However, most methods were developed based on the stationary theory, which assumed that the return level of annual streamflow extremes is constant and a simple relationship exists between the return period and the return level of streamflow extremes. However, the non-stationarity of signals is a common problem including hydrological research (Bennett et al., 2015; Mentaschi et al., 2016; Bracken et al., 2018). To our knowledge, this study was the first to identify the impacts of human activities and climate variability on the streamflow based on the non-stationarity theory in Southeast China. Compared with the previous methods, this study not only distinguished the impacts of human activities and climate variability on the watershed processes but also evaluated these impacts on different time scales.

The maximum flow at different periods can be used to understand the magnitude and duration of annual extreme flow conditions (Richter et al., 1996). The maximum flow at different scales may respond differently to the effects of climate variability and human activities. For example, in Japan, due to monsoons, persistent heavy rainfall over several days frequently occurs during the rainy season, and 5-day maximum flow have proven to be a useful index to represent the potential damage caused by persistent heavy rain (Hu et al., 2017). Sraj et al. (2016) suggested that streamflow regimes with 10-year return levels may be more sensitive to precipitation. In this study, the impacts of climate variability and human activities were also

determined in terms of the 10-year return period in Southeast China (Fig 3-9 and Table 3-3). Human activities and climate change in Southeast China evidently reduced the return levels of extreme high flow based on the 10-year return period.

Human activities potentially amplify floods in the watershed (Sun et al., 2018; Zhang et al., 2018). Extreme streamflow at low time scales (e.g. Sm3day and Sm7day) may be more sensitive to human activities, particularly in small-scale watersheds (Table 3-3). Accordingly, amplified floods have been observed more frequently in high urbanized areas with intensified human activities and small discharge areas (Fang et al., 2022). The observed $E_h$ values of Sm3day and Sm7day at stations where the discharge areas are less than 3500 $km^2$ were 177-208% and 114-214%, respectively (Table 3-3). Moreover, human activities may directly decrease streamflow owing to the increasing demand for water consumption (Zhang et al., 2015; Wang et al., 2020). Mahmoodi et al. (2021) found that the Sm90day may decrease 35% with increasing demand for water resources and climate change in a dry area. The dam construction may decrease the magnitude of 90-day max flow while climate variability may increase the magnitude of 90-day max flow (Zhang et al., 2020a). The values of $E_c$ for the Sm30day and Sm90day in the stations, except for the YT, were high compared that those of $E_h$ (Table 3-3). Notably, human activities were always the dominant factor controlling the streamflow extreme in the YT station, which might be attributed to the intensive water intake in this watershed. Similar results were also observed in North China (Wang et al., 2020).

### 3.4.4 Limitations and outlook

The performance of the rainfall-runoff model proposed in this study might be lower for the small watersheds than the large ones. To reduce the complexity, the effect of the evapotranspiration on surface flow was neglected in the model, which may overestimate the surface flow and underestimate the soil flow. Moreover, the different forces of human activities (e.g., land use/cover change, water intake, and hydropower dam operation) might have a different effect on the regional hydrological processes. The effects of climate and human activities may be coupled. Therefore, the rainfall-runoff model in this study should be improved, and the effects of the different forces of human activities on streamflow regimes could be distinguished and quantified in future studies.

## 3.5 Conclusions

This study presents an innovative approach to distinguish the effects of climate change and human activities on extreme streamflow based on the non-stationarity theory. A rainfall-runoff model was adopted to evaluate the streamflow regimes under the natural conditions. The proposed approach facilitates the detection of the non-stationarity of high flow driven by climate change and human activities at different time scales. The method proposed in this study could be valuable for understanding the potential impacts of climate change and human

activities on the watershed system. For the watersheds in Southeast China, human activities may be the major factor that controls the streamflow extremes for short time scales. Climate change would be the dominant factor that influences streamflow extremes at long time scales. Furthermore, the water intake in the watershed might be an important factor that influences its streamflow extremes.

# Chapter 4 A framework to quantify the effects of urbanization on the regional climate change in a China coastal watershed: A non-stationary perspective

**Zhenyu Zhang, Georg Hörmann, Jinliang Huang, and Nicola Fohrer**

**Submitted to Sustainable Cities & Society on 7th September 2022**

## Abstract

Urbanization has profoundly influenced the local environment and climate conditions. Based on non-stationary theory, a grid-to-grid framework was proposed in this study to quantify the effects of urbanization on the regional climate change in the Minjiang River Watershed (MRW), the largest watershed in Southeast China. The urbanization in the MRW was inactive during 1980-2000, and cities in the MRW have continued to expand at an increasing rate since 2000. These changes may modify climate extremes. A significant increasing trend was observed for the 3-day maximum precipitation (Rx3day) during 1980-2000 in some grid cells, while a decreasing trend was detected from 2001 to 2020 in the same area. The 30-day maximum PET (Ex30day) and 90-day maximum PET (Ex90day) decreased slightly from 1980 to 2000, but the values of most grid cells increased from 2001 to 2020. The patterns of 90-day maximum precipitation (Rx90day), 7-day maximum temperature (Tx7day), and 30-day maximum temperature (Tx30day) may be more sensitive to the urban sprawl. And the extreme temperatures may be amplified in the highly urbanized areas. The proposed method quantified the impacts of urbanization on local or regional climate and could be useful for understanding the relationship between human activities and climate change.

Key words: Urbanization; Regional climate change; Non-stationary;

## 4.1 Introduction

The past several decades have seen a widespread urban expansion as well as a changed climate (Song et al., 2019; Li et al., 2020a; Chen et al., 2021). Land use/cover change, as one of the most important human-induced change, have been proved by many studies that it may exert a great influence on local environment (Grimm et al., 2008; Fenner et al., 2019; Song et al., 2021). And there is a broad agreement that urbanization may influence local or regional climate extremes (Sun et al., 2016; Gu et al., 2019). As a result of an exponential population growth and accelerated economic development, the increased impervious surface area caused by urbanization may alter the surface energy balance and the regional climate (Yang et al., 2013; Li et al., 2020a; Song et al., 2021). However, the response of regional climate to urbanization may be different for varying climate parameters and the urbanization may modify the patterns of climate extremes in the local or regional scales (Vittal et al., 2016; Lu et al., 2019). It is therefore crucial to identify the impacts of urbanization on regional climate change.

Previous studies suggested that urbanization may influence the urban boundary layer and modify the local climate including regional climate extremes by increasing the impervious surface areas (Grimm et al., 2008; Gu et al., 2019; Li et al., 2019; Wang et al., 2021c). For example, several studies found that urbanization may amplify precipitation in cities, and increasing frequencies and amounts of short-duration intense rainfall was observed in the urbanized areas (Yang et al., 2017; Lu et al., 2019). Meanwhile, some studies also found a negative influence of urbanization on local precipitation (Wang et al., 2012). Many studies also concluded that urbanization has a significant influence on extreme temperatures. As a result of the urban heat island induced by the effects of urbanization, higher air temperature was usually observed in urban areas (Grimm et al., 2008; Mishra et al., 2015; Li et al., 2020b). However, different trends were observed in different regions. For example, several studies indicated that unexpected heat waves may be induced as the result of urbanization in some cities of Japan, USA, and China (Nonomura et al., 2009; Mishra et al., 2015; Li et al., 2020b). The relationship between urbanization and regional climate change is very complex and varies spatially, which is still an open question (Scott et al., 2018; Fenner et al., 2019); climate extremes may be amplified or reduced by urbanization (Li et al., 2014; Gu et al., 2019; Li et al., 2020b). To contribute to this discussion, we analyzed the impacts of urbanization on regional climate extremes for cities within the largest watershed in Southeast China.

Extreme value theory has been widely used to analyze meteorological and hydrological extremes (Vasilliades et al., 2015; Tian et al., 2021). Due to the changes in climate patterns and human activities, the traditional assumption of stationarity might underestimate the extreme events. Many studies suggested that the temporal evolution of climate extremes would follow the patterns of non-stationary (Milly et al., 2008; Vu and Mishra, 2019). For water resources

management, the non-stationary frequency analysis was more valuable than climate models and the presence of any trend can have a considerable effect on the interpretation of observations (Chen et al., 2017: Su and Chen, 2019). The non-stationary patterns of climate extremes may be controlled by many factors and many studies suggested that regional factors induced by human activities are more responsible for these patterns than other factors (Liu et al., 2022). The non-stationary generalized extreme value (GEV) has been widely used to understand the characteristics of climate extremes through fitting annual series of maxima (Lu et al., 2019; Gu et al., 2019). For this reason, many studies developed non-stationary models for climate extreme based on GEV theory for water resource management including trend analysis, recurrence levels analysis, frequency analysis and engineering design. For example, Siliverstovs et al. (2010) assessed the likelihood of heat waves in summer with the GEV distribution. Su and Chen (2019) proposed a framework to identify major influences on extreme precipitation in the Pearl River Basin and suggested that extreme precipitation is not only affected by climate factors but also by human activities (e.g., urbanization). However, most studies use climate indices (e.g. ENSO) to understand the non-stationary patterns of climate extremes, but this may not be appropriate for modelling. The impacts of human activities, especially urbanization should be fully understood.

Since the 1980s, cities in China began to sprawl and urbanization has accelerated in the 2000s in most of China (Huang et al., 2012; Lu et al., 2019: Song et al., 2021). Recently, studies tried to identify the impacts of urbanization on regional climate using the difference of time series between urban and rural stations. For example, after comparing the patterns of climate extreme in urban, suburban and rural areas, Lu et al. (2019) found that urbanization increases the magnitudes and recurrence levels of extreme precipitation for different return periods. The estimated recurrence levels of the annual maximum 1-day and 5-day precipitation were increased by 25.9% and 59.1% for urbanized stations. Gu et al. (2019) also proved that urbanization may induce non-stationarity in precipitation extremes at local or regional scales in China, and significant non-stationarity may tend to occur in urbanized areas compared with rural or urbanized areas for the precipitation extremes. The classification of stations (e.g. urban, suburban or rural) is usually based on administrative boundaries in most studies. However, this method may influence the results of the studies if regions have a different developing level or a complex shape of the administrative unit. In addition, most of these studies were conducted based on limited station data. Meanwhile, most non-stationary modeling studies focus primarily on changes in extremes over time (Ragno et al., 2019). Few studies have explored large-scales areas.

To extend the existing studies, we analyzed the spatial and temporal patterns of urbanization in Southeast China as well as the climate extremes during the past 40 years. To achieve this goal, a framework will be proposed based on the GEV method. The objectives of

this study are: (1) to identify the spatial-temporal variability of urbanization; (2) to understand the patterns dynamics of climate extremes in Southeast China; and (3) to analyze the impacts of urbanization on regional climate changes.

## 4.2 Data and methodology

### 4.2.1 Study area

The Minjiang River Watershed (MRW) is the largest watershed in Southeast China with a area of 60992 km$^2$ (116°23′–119°43′ E, 25°23′–28°19′N) (Fig 4-1) with a typical subtropical monsoon climate with about 70% precipitation between April and September (Zhou et al., 2016; Zhang et al., 2019). The annual temperature and precipitation are 18 ℃ and 1617 mm, respectively (Zhou et al., 2016; Wang et al., 2017). The watershed provides water for residential, industrial and agricultural activities of its more than 12 million residents which account for approximately half of the population in Fujian Province (Zhou et al., 2016; Zhang et al., 2019). Similar to other regions in Coastal China, MRW is experiencing rapid urbanization since the beginning of the reform and opening-up period. The impervious areas in the MRW expand with an increasing rate since 2000.

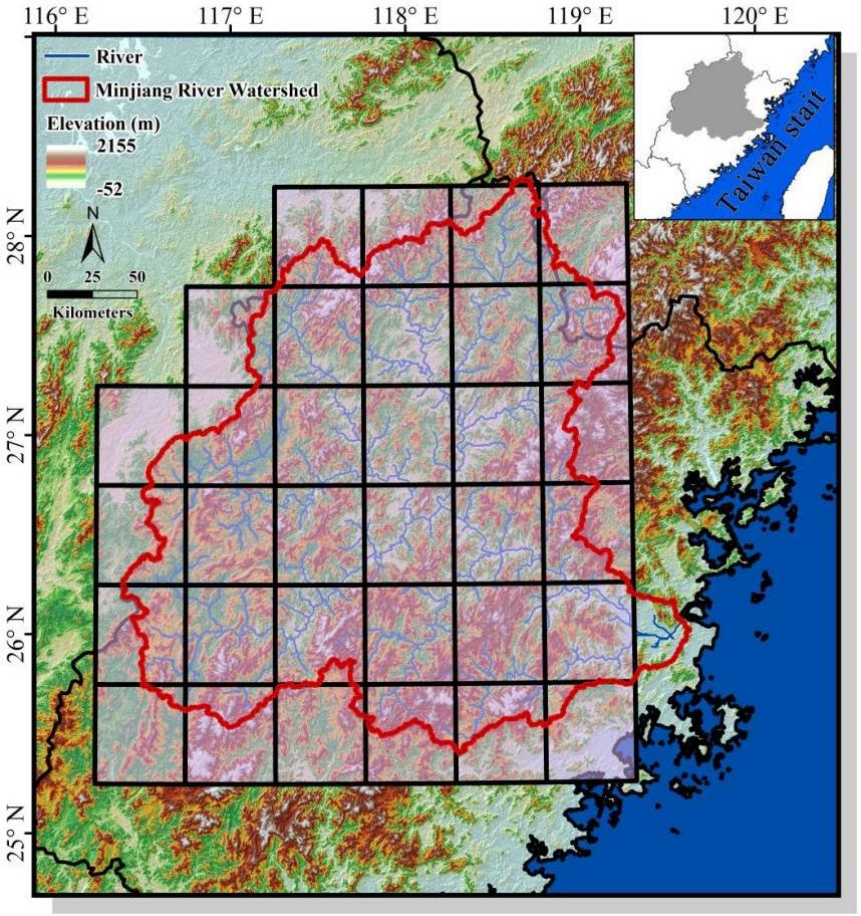


Fig 4-1 Study area

### 4.2.2 Data sources

The land use data for the MRW of the years 1980, 1990, 1995, 2000, 2005, 2010, 2015, 2018, and 2020 were provided by the Data Center for Resources & Environmental Science, Chinese Academy of Sciences (RESDC) (http://www.resdc.cn). The land use information was interpreted manually based on TM digital images. The accuracy of the six classes of land use (i.e. Cropland, Woodland, Grassland, Water body, Built-up land, and Unused land) was above 94.3% while the overall accuracy of the 25 subclasses was above 91.2% (Liu et al., 2014). The meteorological data, including mean daily air temperature, maximum daily air temperature, minimum daily air temperature, and daily precipitation was obtained from the China Meteorological Administration (http://data.cma.cn) with 0.5 °×0.5 °spatial resolution during 1980-2020 and potential evapotranspiration was estimated with the Hargreaves method (Hargreaves et al., 1985; Hargreaves and Allen, 2003).

### 4.2.3 Urbanization & Climate indices

Impervious surfaces have been used as an important indicator of urbanization for decades (Brabec et al., 2002; Claggett et al., 2013; Yang et al., 2021). The impervious surface coefficients (ISC) were applied in this study to estimate the percent of impervious surface area for each grid cell.

$$ISG = \frac{\sum A_i \cdot ISC_i}{A}$$

Where ***ISG*** is the percent of imperviousness for each analysis grid cell, $\boldsymbol{A_i}$ is the area of the particular land use, $\boldsymbol{ISC_i}$ is the impervious surface coefficient for the specific land use category (Table 4-1), and ***A*** is the area of the analysis grid cell.

Table 4-1 Impervious surface coefficients (ISC) for major land use category

| Land use | Description | ISC (%) |
|---|---|---|
| Open space | Natural areas | 2 |
| Agriculture | Farming area | 4 |
| Urban area | Community, neighborhood, office in urban area | 83 |
| Suburban area | Community, neighborhood, office in suburban area | 69 |
| Other built-up | Publicly owned, buildings in rural area | 50 |

Note: ISC was estimated based on Washburn et al. (2010)

The extreme indices were calculated for precipitation, temperature and potential evapotranspiration (PET). The details of the indices are listed in Table 4-2.

Table 4-2 Indices of climate extremes

| Index | Indicator name | Definition |
|---|---|---|
| Rx1day | 1-day maximum precipitation | Annual maximum precipitation of 1 day mean |
| Rx3day | 3-day maximum precipitation | Annual maximum precipitation of 3 day means |
| Rx7day | 7-day maximum precipitation | Annual maximum precipitation of 7 day means |
| Rx30day | 30-day maximum precipitation | Annual maximum precipitation of 30 day means |
| Rx90day | 90-day maximum precipitation | Annual maximum precipitation of 90 day means |
| Tx1day | 1-day maximum temperature | Annual maximum temperature of 1 day mean |
| Tx3day | 3-day maximum temperature | Annual maximum temperature of 3 day means |
| Tx7day | 7-day maximum temperature | Annual maximum temperature of 7 day means |
| Tx30day | 30-day maximum temperature | Annual maximum temperature of 30 day means |
| Tx90day | 90-day maximum temperature | Annual maximum temperature of 90 day means |
| Ex1day | 1-day maximum PET | Annual maximum PET of 1 day mean |
| Ex3day | 3-day maximum PET | Annual maximum PET of 3 day means |
| Ex7day | 7-day maximum PET | Annual maximum PET of 7 day means |
| Ex30day | 30-day maximum PET | Annual maximum PET of 30 day means |
| Ex90day | 90-day maximum PET | Annual maximum PET of 90 day means |

### 4.2.4 Methods to distinguish the impacts from urbanization

Logistic regression models were used in this study to simulate urbanization (i.e. ISC) from 1980 to 2020. In order to understand the relationship between climate extremes and urbanization, the process-informed non-stationary extreme value analysis proposed by Ragno et al. (2019) was used. With this method, the non-stationary GEV method was used to estimate return periods of climatic extremes with urbanization which was developed on a hybrid evolution Markov Chain Monte Carlo approach for numerical parameters estimation and uncertainty assessment.

In order to understand the impacts of an increasing rate of urbanization, the non-stationary GEV models were built for different periods (i.e. 1980-2000 and 2001-2020). Based on the GEV models for the different periods, the relationship between magnitude of climate extremes

and the levels of urbanization at 10-year return levels. Based on previous study (i.e. Lu et al., 2019), the impacts of the urbanization can be quantified as follow:

$$Alteration = f'(ISC) - g'(ISC)$$

Where ***Alteration*** is the index to quantify the impact of urbanization on the regional climate extremes; ***f(ISC)*** is the function to estimate the relationship between climate extremes and urbanization in the pre-impacted period (i.e. 1980-2000) and the ***g(ISC)*** is the function to estimate the relationship between climate extremes and urbanization in the post-impacted period (i.e. 2001-2020)

The nonparametric Mann-Kendall trend (M-K) test (Mann, 1945; Kendall, 1975; Alan et al., 2003) was used to detect the trends of climate extremes. For the Z-score calculate for M-K trend test, Z-score more than 0 means increasing trend and less than 0 means decreasing trend. If the Z-score more than 2.58 or less than -2.58, the significance level is better than 1%. If the Z-score more than 1.96 or less than -1.96, the significance level is better than 5%. Z-score more than 1.65 or less than -1.65, the significance level is better than 10%.The Mann-Whitney test was used to test the difference between two groups.

## 4.3 Results

### 4.3.1 Patterns of urbanization

Urbanization from 1980 to 2020 was estimated based on the land use change analysis (Figs S4-1, S4-2, 4-2, and Table S4-1). From 1980 until 2000 the urbanization was quite slow and the suburban area increased from 3 grid cells in 1980 to 6 grid cells in 2000. From 2000 onwards, the trend increased from 6 grid cells in 2000 to 15 grid cells in 2020. Urban grids also increased 1 grid cell in 2000 to 6 grid cells in 2020.

The patterns of the urbanization in the MRW were quantified based on the process of the urbanization in the two periods (i.e. 1980-2000 and 2001-2020). And three patterns of land use was identified as follows, Group 1: the grid cell that was inactive during two periods; Group 2: the grid cell that was inactive in 1980-2000 and transferred to suburbanized grid cell in 2001-2020; Group 3: the grid cell that was inactive in 1980-2000 and transferred to urbanized grid cell in 2001-2020 (Fig 4-3).

Logistic regression models were employed to further understand the process of urbanization (Fig S4-3). The Reduced Chi-Sqr and Adj. $R^2$ were calculated for the logistic regression models in the MRW (Fig 4-4). The observed data matched well with the simulated data. The value of Reduced Chi-Sqr was less than 0.2 and the value of Adj. $R^2$ was more than 0.90. The models simulate urbanization quite well.

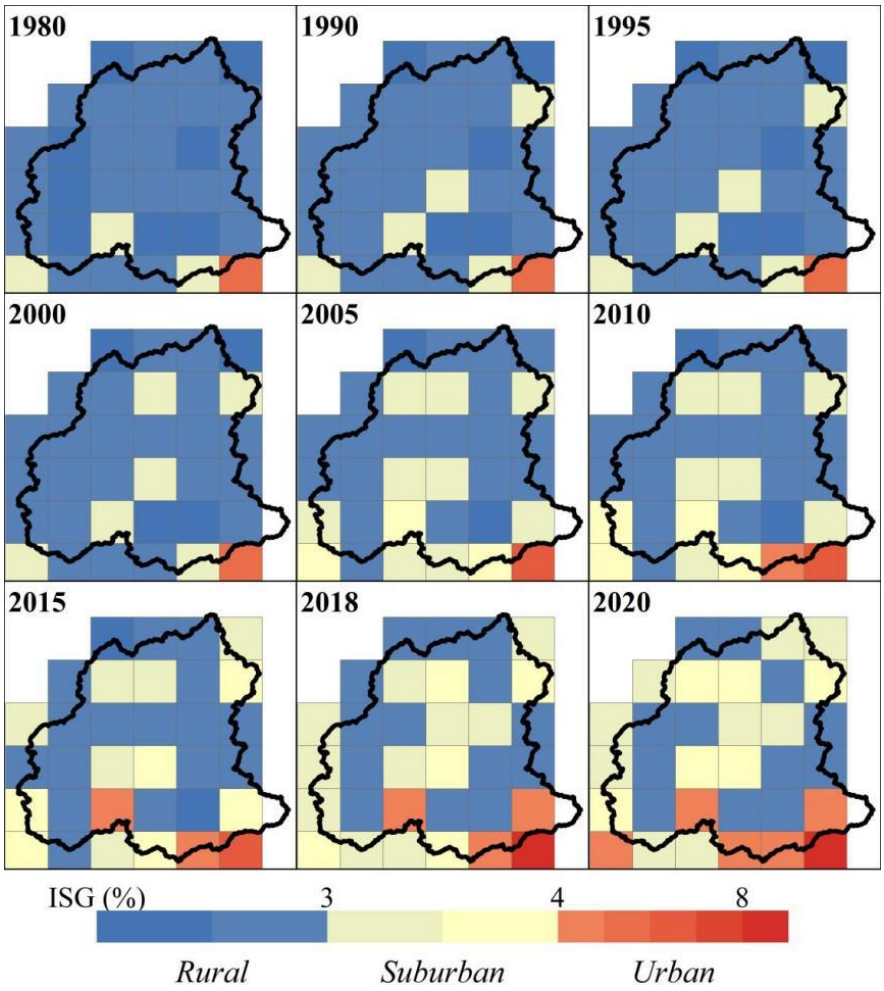


Fig 4-2 Process of urbanization in the MRW

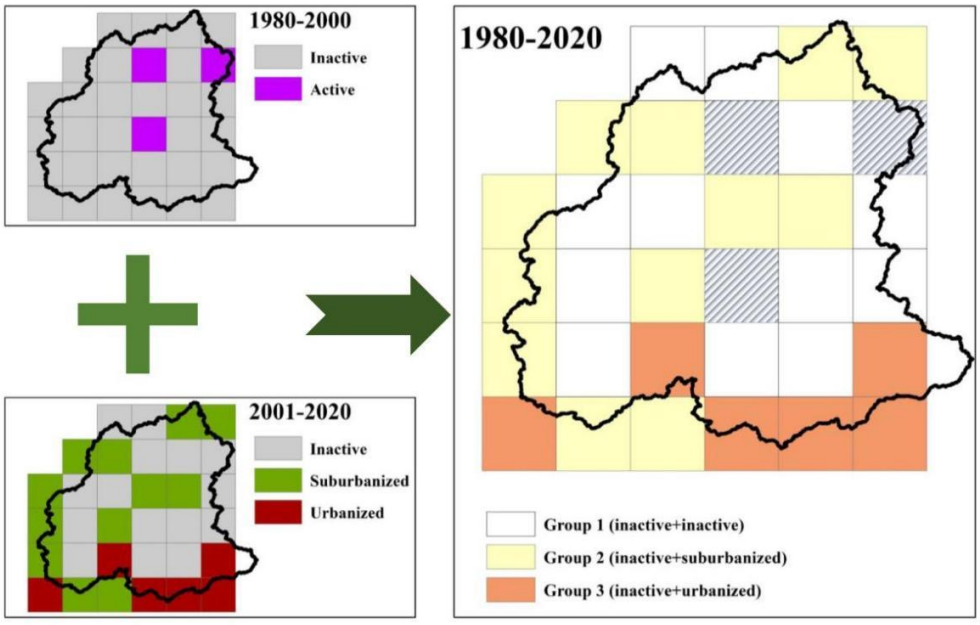


Fig 4-3 Land use patterns in the MRW

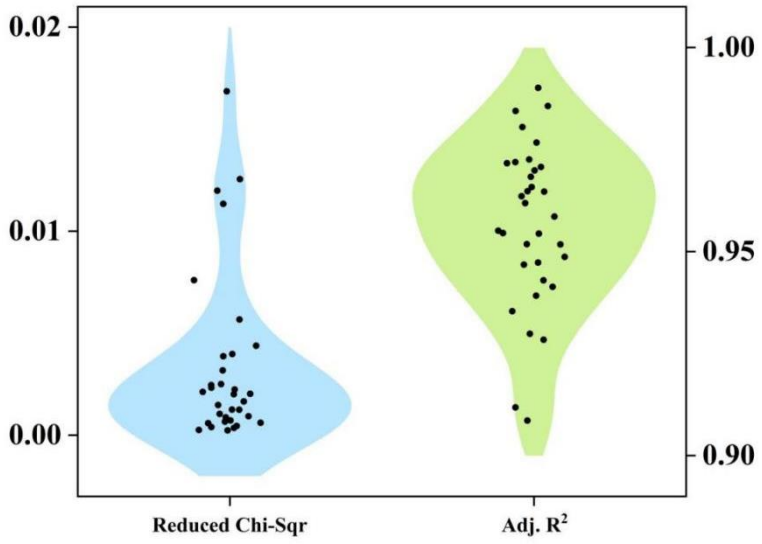


Fig 4-4 Reduced Chi-Sqr and Adj. $R^2$ for the urbanization simulation

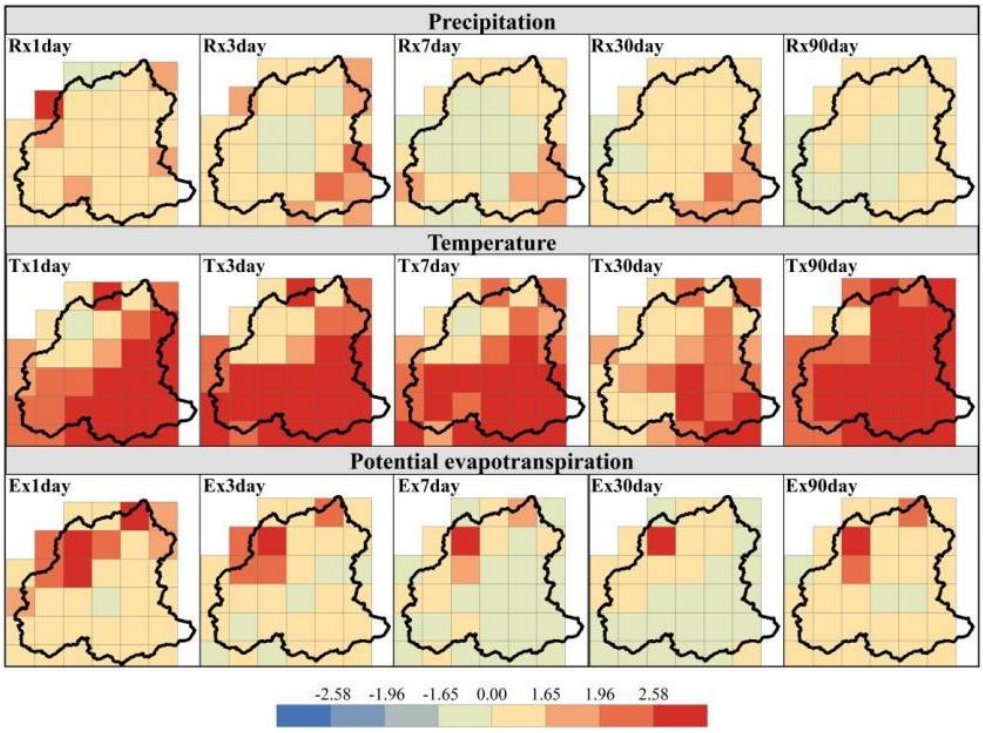


Fig 4-5 Z-score of M-K trend test in the MRW from 1980 to 2020

## 4.3.2 Patterns of climate change

Based on the M-K trend test, trends of climate extremes in terms of precipitation, temperature and PET from 1980 to 2020 were detected (Fig 4-5).The indices of extreme temperature in the Southeast MWR where the level of urbanization was high increased more significantly, especially for the Tx1day, Tx3day, Tx7day and Tx90day. The indices of extreme

PET in the Northeast MRW where the level of urbanization was relatively low, decreased more significantly than other indices including Ex1day and Ex3day.

To further understand the interaction between urbanization and climate change, the trend of the climate variables was also estimated with an M-K test for different periods, i.e. from 1980 to 2000 and from 2001 to 2020 (Fig 4-6). A significantly increasing trend was observed for the R3day and Tx90day from 1980 to 2000 in some grid cells and a decreasing trend from 2001 to 2020 in the same area. Ex30day and Ex90day decreased slightly from 1980 to 2000, but the values of most of the grid cells increased from 2001 to 2020. This shows that urbanization may influences trend of climate variability in the MRW.

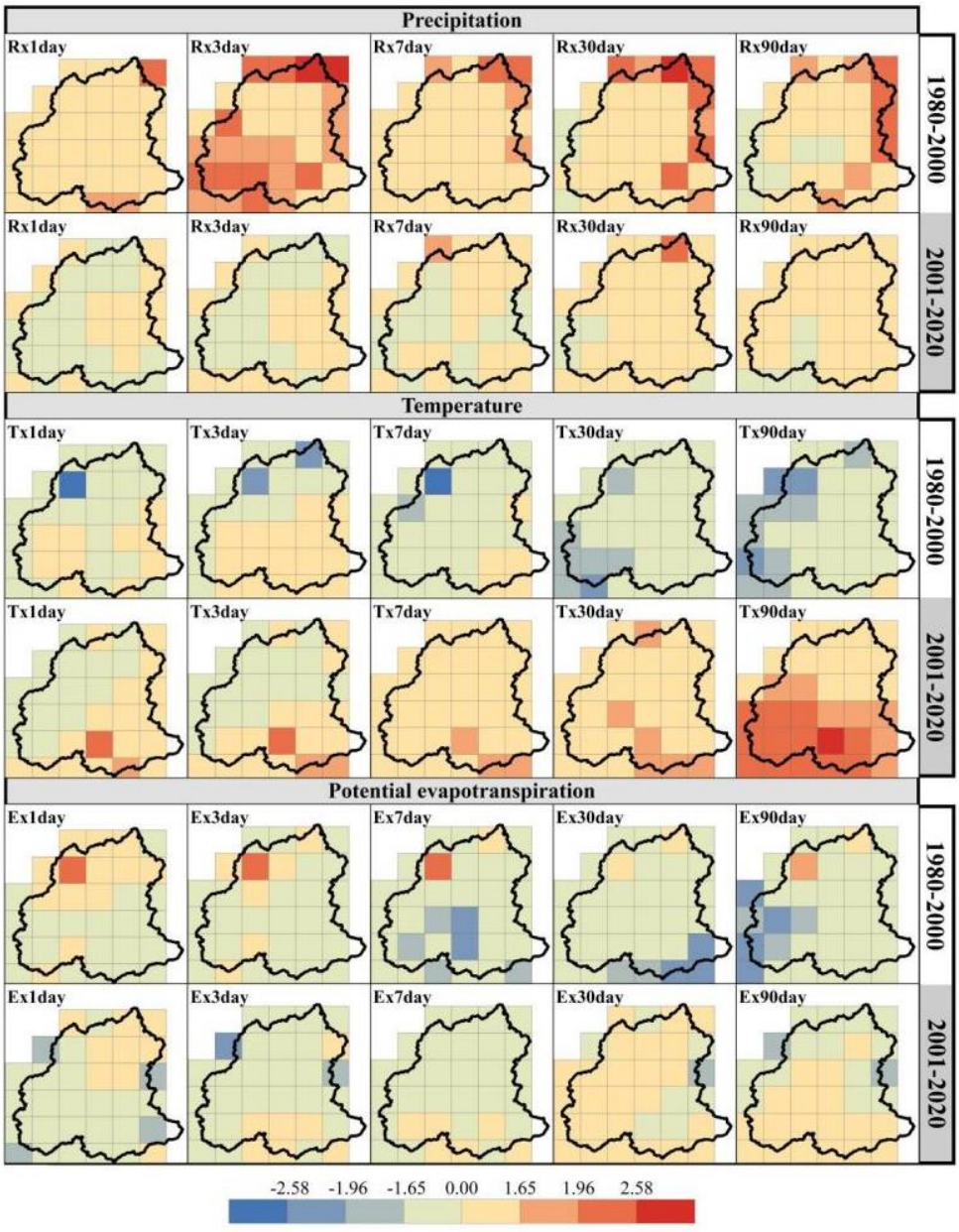


Fig 4-6 Z-score of M-K trend test in the MRW in different periods

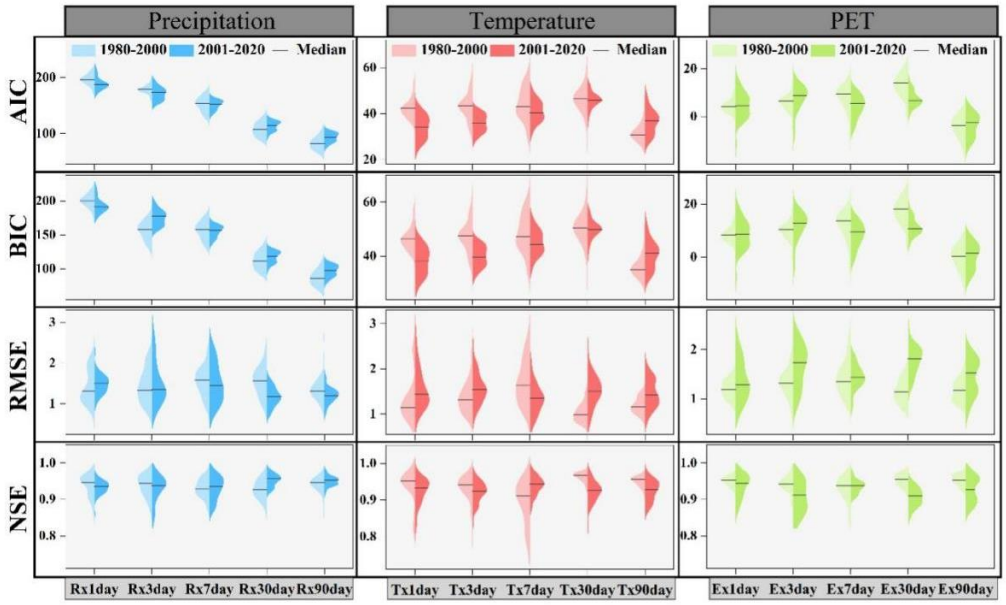


Fig 4-7 Evaluation of the non-stationary GEV models

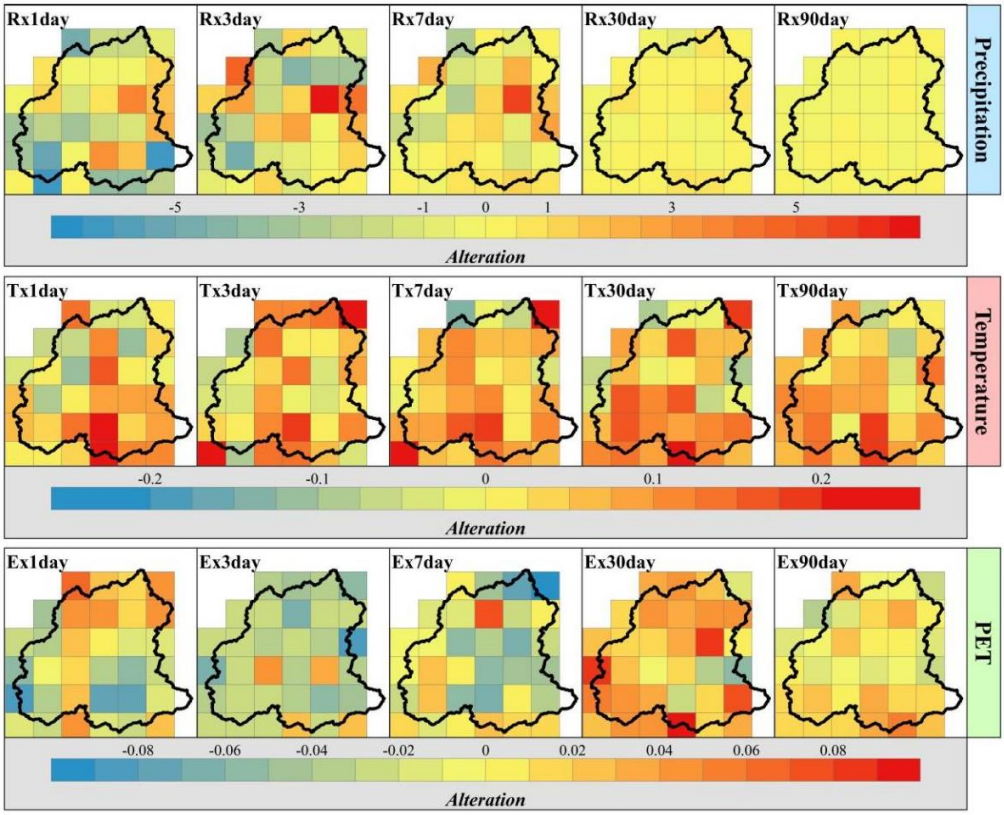


Fig 4-8 Effects of urbanization on climate extremes

### 4.3.3 The impact of urbanization on climate change

In order to understand the non-stationary patterns of climate extremes influenced by urbanization, the non-stationary GEV models were built for each grid cell in different periods. The results show that the models can be used to estimate non-stationary patterns of climate extremes (Fig 4-7).

Based on the models above, the values of a 10-year return period were estimated. The degree of change was estimated by comparing the difference between 1980-2000 and 2001-2020 (Fig 4-8).

The effects of urbanization were estimated by comparing the differences among the groups classified (Table 4-3). The climate extremes may be modified in the high urbanized areas (inactive+urbanized). In contrast, the patterns of climate extremes in the suburbanized areas (inactive+suburbanized) were similar with the inactive areas (inactive+inactive).

The temperature extremes were more sensitive to urbanization (Figs 4-9, and 4-10), especially in the high urbanized areas (Fig 4-9). And the amplified temperature extremes were observed with increasing urbanization, especially for the Tx7day and Tx30day (Table 4-3, and Fig 4-9).

Table 4-3 Result of Mann-Whitney test among different group (Z-values)

| | Category | Indices | Inactive+suburbanized | Inactive+urbanized |
|---|---|---|---|---|
| Inactive+inactive | Precipitation | Rx1day | 0.61 | 1.45 |
| | | Rx3day | -0.78 | -0.33 |
| | | Rx7day | -0.72 | -1.08 |
| | | Rx30day | -0.09 | -1.08 |
| | | Rx90day | -1.18 | **-2.01*** |
| | Temperature | Tx1day | 0.20 | -1.92 |
| | | Tx3day | 0.00 | -1.73 |
| | | Tx7day | -0.09 | **-2.20*** |
| | | Tx30day | -0.61 | **-2.01*** |
| | | Tx90day | -0.12 | -0.52 |
| | PET | Ex1day | -0.49 | -0.89 |
| | | Ex3day | -0.14 | -1.45 |
| | | Ex7day | 0.26 | -0.98 |
| | | Ex30day | -1.70 | -1.64 |
| | | Ex90day | 1.15 | -1.45 |

****p<0.05***

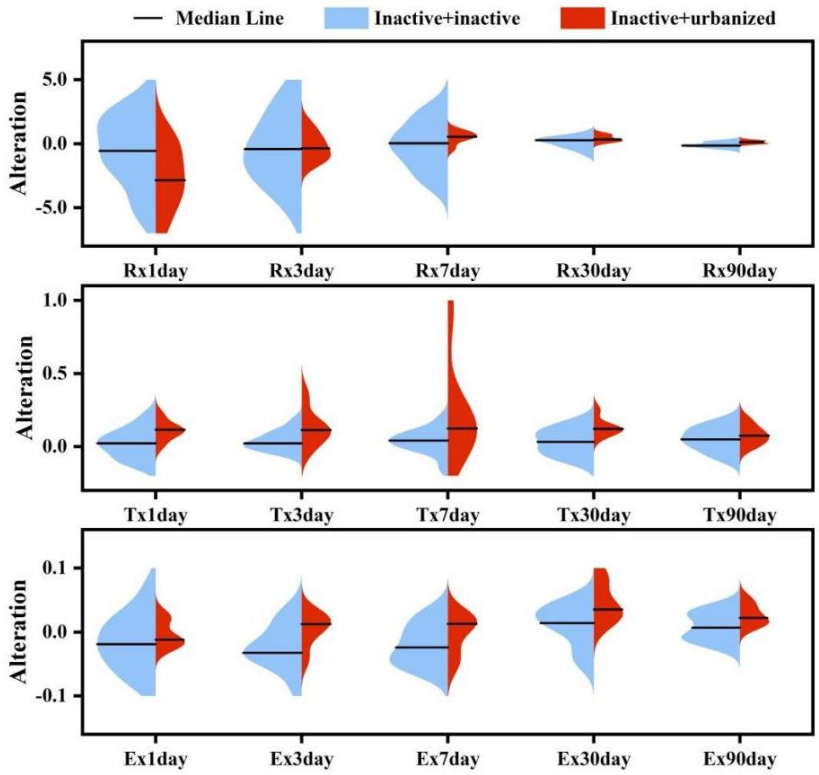


Fig 4-9 The effects of urbanization on climate extremes (Inactive+inactive V.S. Inactive+urbanized)

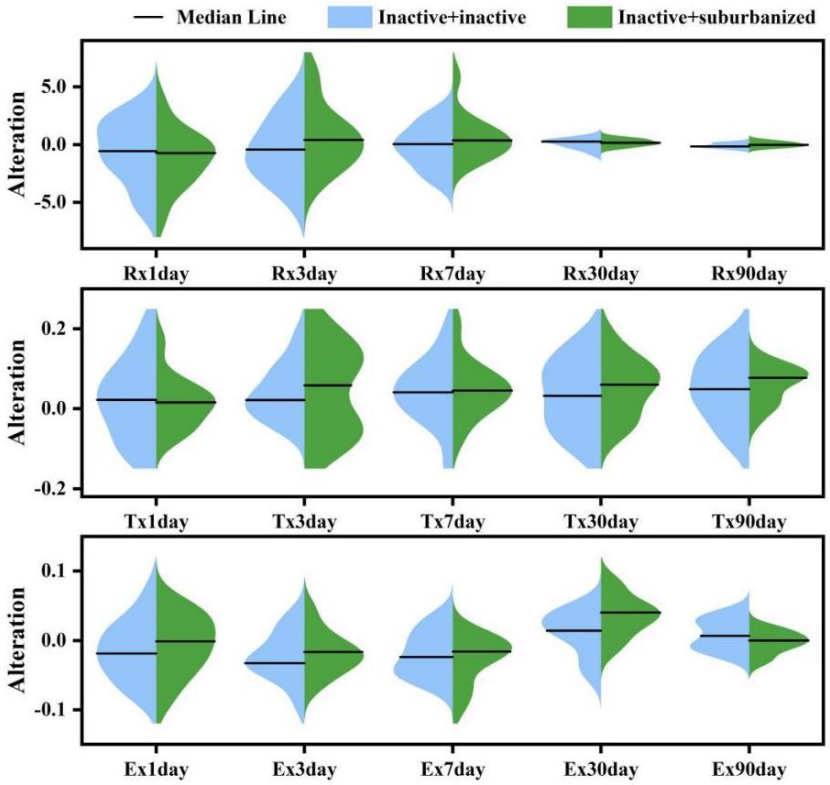


Fig 4-10 The effects of urbanization on climate extremes (Inactive+inactive V.S. Inactive+suburbanized)

## 4.4 Discussion

### 4.4.1 Applicability of the proposed framework

Recent studies have confirmed the strong relationship between urbanization and climate extremes (Yang et al., 2017;Song et al., 2021; Pimonsree et al., 2022). They found that the patterns of regional climate extremes follow a non-stationary distribution as the results of urbanization and developed methods to understand the relationship between urbanization and regional climate change. For example, Lu et al (2019) analyzed this relationship by comparing the differences between the non-stationary patterns of urban stations, suburban stations, and rural stations. With the GAMLSS (Generalized Additive Model for Location, Scale and Shape) framework, Gu et al. (2019) found that urbanization may induce non-stationary behavior in precipitation extremes. In this study, we further extend existing studies and propose a new framework to quantify the impacts of urbanization on regional climate extremes.

Instead of using time covariates non-stationary models which may not cover the physical relations, we introduced an index of urbanization as covariates to build grid based non-stationary models for climate extremes. To understand urbanization in the MRW, logistic models were used with Reduced Chi-Sqr and Adj. $R^2$ as performance measures. Compared with previous studies, the non-stationary GEV models are more useful to quantify the patterns of climate extremes. The different metrics used to evaluate the models included AIC, BIC, RMSE, and NSE. Though the NSE and RMSE were similar in the three types of models (i.e. GEV models for Precipitation, Temperature and PET, respectively), the Observed AIC and BIC were more effective in these models. The calculated AIC for the precipitation, temperature, and PET were 50-250, 20-60, and <20 respectively (Fig 4-7). The observed BIC for the precipitation, temperature, and PET were 50-300, 10-70, and <30 respectively (Fig 4-7). The value of AIC and BIC may be more effective to evaluate the uncertainty of the non-stationary GEV models. Similar results were found in related studies. Panagoulia et al. (2014) showed that both AIC and BIC are always effective to detect non-stationary behavior, and BIC may be more effective to evaluate the models except in very small samples. Kim et al. (2017) also suggested that the AIC may be better than other metrics for non-stationary data with a relatively small sample size.

### 4.4.2 Dynamics of climate extremes

The patterns of regional climate have changed remarkably in the past decades as a consequence of global climate change and intensified human activities (Gu et al., 2019; Lu et al., 2019; Li et al., 2021). As a consequence of global climate change, precipitation can become more extreme in some regions (Donat et al., 2016). For example, Van den Besselaar et al. (2013) observed a general increase of extreme precipitation in Northern Europe in autumn, winter, and spring. In China, both annual maximum daily precipitation and heavy precipitation days

increase in southern, but decrease in northern China (Chen et al., 2011). Huang et al. (2013a) also found that extreme precipitation seems to be intensified in the past decades in Southeast China. In this study, an upward trend of extreme precipitation was also found (Fig 4-5). Slightly increasing trends were observed in more than 80% of the grid cells for Rx1day, Rx3day, and Rx30day. Significantly increasing trends were found in grid cells located in the southeast for Rx3day, Rx7day, and Rx30day. This suggests that the coastal area may be more sensitive to climate change and exhibit more frequent extreme climate events. Additionally, the extreme temperatures increased significantly including Tx1day, Tx3day, Tx7day, Tx30day, and Tx90day (Fig 4-5).

Changes in climate extremes would follow similar patterns in all grid cells if climate change was the only factor controlling the climate patterns. However, the observed patterns for the two periods (i.e. 1980-2000 and 2001-2020) were totally different from each other (Fig 4-6). It is therefore highly probable that the patterns of climate extremes are controlled by urbanization. China has experienced rapid urbanization since its reform process (Lu et al., 2019; Li et al., 2020a). The changes of land use in the MRW were similar to other areas in China. In this study, we divided the data set into two periods, a period from 1980 until 2000 and from 2001 until 2020. In the first period, the values of ISG in each grid cell increased slowly, only a few rural cells were changed into suburban cells. The number of urban cells did not change. In contrast, in the period of 2001-2020, the number of urban grid cells increased from 1 (2000) to 6 cells (2020). Urbanization was speeding up since 2000. Similar of land use and land cover changes were found in other areas in China. For example, Huang et al. (2012) found a more intense land use and cover change since 2000 in the Jiulong River Watershed in Southeast China. Lu et al. (2019) reported an increasing urbanization in the Yangtze River Delta since 2000.

### 4.4.3 Impacts of urbanization on local or regional climate extremes

Chinese cities have been experiencing unprecedented growth for over three decades and the resulting urbanization has a remarkable impact on the hydrological cycle at the local and regional scale (Huang et al., 2013a; Gu et al., 2019; Lu et al., 2019). Urbanization may modify the relationship between regional climate extremes and global change and urbanization may contribute differently to local or regional climate extremes (Lu et al., 2019; Gu et al., 2019). The response to urbanization may vary with climate indices. Though an increasing trend of observed temperature was found in the past decades in the MRW (Fig 4-5), the opposite trends were identified for different periods (Fig 4-6). During the period from 1980 to 2000 with low urbanization, the observed extreme temperatures decreased but increased in the period of 2001-2020 with a high urbanization rate. Most of the temperature extremes in the same period decreased including Tx7day, Tx30day, and Tx90day. 12 grids of Tx90day showed a

significantly decreasing trend in this period. Similarly, the different trends of precipitation extremes were also observed in these two periods, respectively. Rx1day, Rx3day, Rx7day, and Rx30day increased in more than 80% of the grid cells in period of 1980-2000, and most of these indices decreased in the period of 2001-2020 (Fig 4-6). Especially for Rx3day, 24 grids increased significantly from 1980 to 2000.

Additionally, the speed of the urbanization may influence regional climate extremes. Gu et al. (2019) also proved that the transition patterns of land use may play an important role in local or regional climate patterns. The patterns of Rx90day, Tx7day, and Tx30day may be more sensitive to the urban sprawl (Table 4-3). There is no doubt that increasing urbanization increases the area of impervious surface and urban heat islands (Mishra et al., 2015; Zhou et al., 2019; Li et al., 2020b). Among the indices of climate extremes, the observed temperature and precipitation were more sensitive to urbanization then PET in this study.

Urbanization contributes to the climate extremes and may introduce non-stationarity of local or regional climate data sets (Mishra et al., 2015; Lu et al., 2019; Gu et al., 2019). The non-stationary GEV plays an important role to understand the patterns of climate extremes (Lu et al., 2019; Vu and Mishra, 2019). In order to quantify the non-stationarity of climate extremes, Lu et al. (2019) developed a method to estimate the non-stationary impacts of urbanization in terms of rate and applied it in the Yangtze River Delta. We also estimated the non-stationary patterns of climate extremes based on the 10-year return period with the non-stationary GEV models (Figs 4-9 and 4-10). The non-stationary patterns of climate extreme in the long-time scale may be more sensitive to the process of urbanization. The Rx90day, Tx7day, and Tx30day changed much in the MRW (Table 4-3). Meanwhile, the urbanization may amplify the temperature extremes in the watershed (Fig 4-9).

## 4.5 Conclusion

This study presents a grid-based framework to identify the impacts of urbanization on regional climate extremes from a non-stationary perspective . The urbanization in the MRW may d the trend of the local climate extremes. The increase of urban area may change the non-stationary patterns of local or regional climate by modifying the rate of the expected return period. The proposed method quantifies the impacts of urbanization on local or regional climate and would be helpful for understanding the relationship between human activities and climate change.

# Chapter 5 A grid-based interpretable machine learning method to understand the spatial relationships between watershed properties and water quality

**Zhenyu Zhang, Georg Hörmann, Jinliang Huang, and Nicola Fohrer**

**Submitted to Ecological Indicators on 1th March 2023**

## Abstract

Understanding the spatial relationship between watershed properties and water quality is essential for watershed management. However, it remains challenging to identify such a relationship because of its nonlinearity. To understand the temporal and spatial effects of the driving of water quality, we developed a grid-baed interpretable machine learning method that integrated the random forest regression (RFR) model with the Shapley Additive exPlanations (SHAP) method to explore the relationship between water quality (e.i. nitrogen, phosphorus, and chemical oxygen demand) and watershed properties in the Minjiang River Watershed (MRW). A grid-based model was developed with 1837 input features for streams and reservoirs, respectively, based on land use, population, fertilizer, and Noah land surface model in the MRW. The simulated data fitted well with the observed values. Compared to the water quality in the stream, the water quality in the reservoirs may be more sensitive to the change environmental settings. The reservoirs may amplify the effects of climate variability on water quality. The soil moisture may modify the water quality, especially the dryness of the top soil may control the pollutants that entry into the watershed. The effects of the urbanization may modify the distributions of important feature that control the regional water quality. The point source pollution per inhabitant may be reduced with increasing urbanization. This study provides an in-depth understanding of the relationship between water quality and watershed properties, temporally and spatially.

Keywords: Land-use; water quality; nonlinearity; machine learning

## 5.1 Introduction

Climate change, topography, and anthropogenic activities can strongly affect regional ecosystems on different spatio-temporal scales (Collins et al., 2019; Lei et al., 2021). Good water quality is one of the most critical factors to support the regional ecosystem as well as the livelihood of human beings. However, the increasing export of agrochemicals and nutrients threatens water quality globally, and an increasing in flow of waste makes these systems vulnerable to the adverse effects of global and regional environmental changes (Liang et al., 2020; Ahmed and Lin, 2021). To quantify effects of global and regional changes on the watersheds system, the relationships between water quality and watershed properties should be fully understood.

Water quality is essentially controlled by land use/cover change as well as other watershed properties such hydrology conditions, soil conditions and anthropogenic activities (Nielsen et al., 2012; Huang et al., 2014; Liu et al., 2021c). Although these relationships have been extensively investigated, there is no generally accepted theory (Liu et al., 2021c). Theoretically, land use change determines the water quality of watersheds (Baker 2003; Roberts and Prince 2010; Liang et al., 2020). Excessive fertilizer application in agriculture has been regarded as a major non-point source of nutrients in rivers (Haidary et al., 2013; Lee et al., 2014; Strehmel et al., 2016). Urbanization may amplify nutrient exports by increasing impervious surface area (Kaushal et al., 2014; Huang et al., 2021). Natural land cover such as wetlands or forests may reduce nutrient exports by reducing soil erosion, filtering pollutants, and assimilating nutrients (Huang et al., 2015). Unfortunately, the results of these studies were not consistent. One of the most important reasons for the inconsistent results was that most studies ignored the nonlinear response of water quality to these factors (Liu et al., 2021c). To analyze water quality of a watershed, the contributions of different properties must be considered separately.

In the past decades, many efforts have been made to understand the relationships between water quality and multiple properties (e.g. land use/cover change, soil conditions, and streaflow regimes) of watersheds (Baker, 2003; Kaushal et al., 2014; Collins et al., 2019). These relationships are commonly analyzed using conventional statistical methods such as ordinary least squares regression, redundancy analysis, and correlation analysis. Using multiple linear regression, Huang et al. (2013c) found that urbanization may be one of the most important factors that influence water quality in a watershed located in Southeast China. Using stepwise multiple linear regression and redundancy analyses, Lei et al. (2021) found that water quality is affected by soil properties, land use composition (areal shares of arable or pasture land, respectively with slopes >2%, forest, and urban) and configuration in a North-German catchment. Additionally, geographically weighted regression (GWR) was used to identify spatial relationships (Tu 2011; Huang et al., 2015; Wang et al., 2021d). Using this method, Tu

(2011) found that percentages of commercial and industrial lands have stronger positive relationships with the concentrations of water pollutants in less-urbanized areas than in highly-urbanized areas in the watersheds around the metropolitan Boston USA which indicated that the inverse relationship may be found between land use change and water quality. However, these methods were developed with the assumption that the relationships between water quality and catchment properties are linear. Therefore, a new method should be developed to understand the nonlinearity and threshold effects of catchment properties on water quality.

Recently, a nonlinear relationship between water quality and catchment properties has been observed in some studies (Zhou et al., 2016; Zhang et al., 2019; Liu et al., 2021c). After investigating the nonlinear relationships between water quality and ecological responses in streams with varying laud use, Amairo et al. (2019) assumed that nonlinearity may be common in watershed systems and found it occurred in more than 50% of the cases compared with other relations. This highlights the potential complexity of nonlinear relationships across the watershed systems. For example, Liu et al. (2021c) indicated that the limited capability of self-purification in the polluted water body may induce a nonlinear result. The hydrology conditions may also modify the relationships across the watersheds. Zhang et al. (2019) found that reservoirs may increase the concentration of ammonia nitrogen ($NH_4^+$-N) in the downstream for the modified streamflow regimes. Although the traditional nonlinear regression method such as power, exponential, quadratic and segmented regressions can be used to detected the potential nonlinear relationship across the watersheds, it is still difficult to explain the development of water quality because the relationships across the watershed may be modified at different stages. Recently, several machine learning algorithms have been proved to outperform traditional methods when it comes to explain complex relationship across the watersheds, especially for the nonlinear cases (Ahmed et al., 2021; Adedeji et al., 2022).

Assessment and modelling of water quality is needed to understand the potential threats to watersheds, because it describes temporal and spatial changes in watersheds (Kaushal et al., 2014; Gallo et al., 2015; Wang et al., 2021e). To depict the nonlinear variation of water quality with changing multiple properties and project the water quality in the watersheds, the machine learning models were employed in a few studies, recently. For example, Zhou et al. (2016) used self-organizing maps to identify the relationships between water quality and land use with a focus on point source pollution. However, the results of these models are difficult to shed mechanistic insights to the underlying physical processes (Lipton, 2018; Wang et al., 2021e). Thus, the interpretable machine learning approaches have recently been developed to gain a deeper understanding of the relationship between water quality and different properties. For example, Wang et al. (2021e) coupled the SHapey Additive exPlanations (SHAP) method with random forest regression (RFR) to simulate the distribution of water quality in the Taihu Lake Basin. They identified the driving forces and found that the total nitrogen (TN) was mainly

affected by agricultural non-point sources, while the permanganate index (CODMn) and total phosphorus (TP) were affected by agricultural and domestic sources. Wang et al., (2021d) analyzed the spatial relationship between water quality and different properties with the GWR method and interpretable machine learning modelling. They found that urban development influences stream water quality and that high-density urban development was more effective in reducing nutrient exports than current sprawl development. However, there is still a lack of robust methods to link water quality and different watershed properties at different temporal and spatial scales.

Many rivers across the globe are experiencing declining water quality and over 45% of rivers in China (Vorosmarty., 2010; Lintern et al., 2018). Owing to intensified anthropogenic activities, water quality seems to have degraded in Southeast China (Huang et al., 2015; Zhou et al., 2016; Zhang et al., 2019). The Minjinag River Watershed (MRW), the largest watershed in Southeast China, is an important source of water for more than 13 million residents for their residential, industrial, and agricultural activities and plays an important role in the regional economy and ecology (Huang et al., 2020). Previous studies in the MRW primarily explored the nonlinear relationships between water quality and watershed properties as a function of the heterogeneity of urbanization, landscape characteristics, and hydrological conditions (Zhou et al., 2016; Zhang et al., 2019; Huang et al., 2020). For example, Zhou et al. (2016) assumed that additional point source pollution in the MRW may weaken the relationship between water quality and land use. Zhang et al. (2019) also indicated that reservoirs in the MRW modify the hydrodynamic characteristics, which may change water quality. The objectives of this study are to: (1) develop an interpretable machine learning approach for water quality modelling in temporal and spatial scales; (2) to identify the most important watershed properties that control water quality (e.g. nitrogen, phosphorus, and chemical oxygen demand); (3) to understand how the temporal and spatial heterogeneity of watershed properties controls water quality and finally (4) derive consequences for regional water resource management..

## 5.2 Material and methods

### 5.2.1 Study area

The MRW is the largest watershed in Southeast China, with an area of 60992 $km^2$ (116°23′–119°43′ E, 25°23′–28°19′N), located west of the Taiwan Strait (Fig 5-1). It lies in a subtropical zone with subtropical monsoon climate, with about 70% precipitation between April and September, and only 2-3% of precipitation occurring during November and December (Zhang et al., 2019; Huang et al., 2020). The annual mean temperature and precipitation in this watershed were 18 ℃ and 1617 mm, respectively (Zhou et al., 2016; Wang et al., 2017). As the seventh highest watershed in China in terms of runoff, more than 60 billion $m^3$ runoff was produced each year which flow into the East China Sea (Zhang et al., 2019). As

one of the most developed regions in China, more than 15 million residents use it as water source for households, industries, and agricultural irrigation. With increasing demand, several dams or reservoirs were built in the MRW, which modified streamflow and nutrient budget (Zhang et al., 2019).

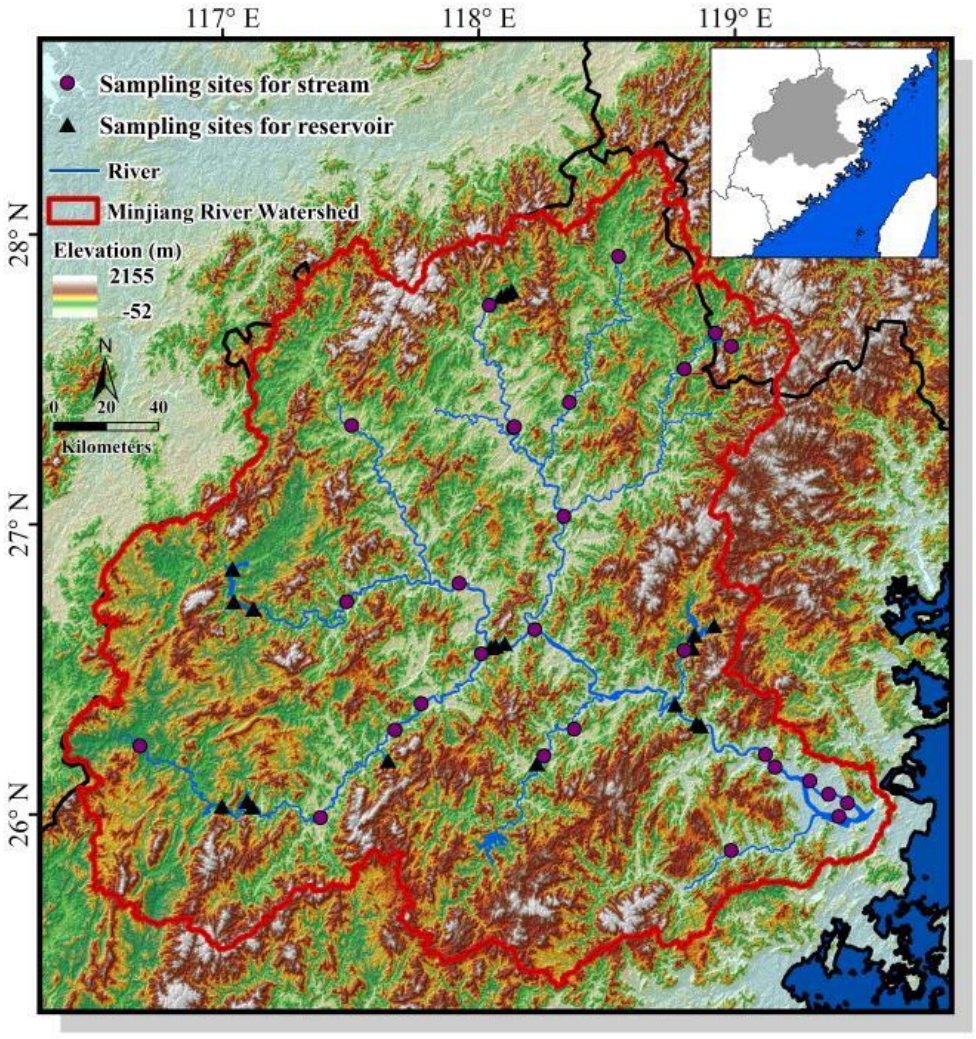


Fig 5-1 Study area

## 5.2.2 Data sources

Water quality data from 2000 to 2014 were collected at 32 sampling sites in the stream and 20 sampling sites in the reservoirs by the Hydrology and Water Resources Survey Bureau of Fujian Province based on the Environmental Quality Standards for Surface Water (GB3838-2002) (Fig 5-1). The three water quality indicators TP, $NH_4^+$-N, and $COD_{Mn}$ were monitored routinely and reflected the changes of water quality in the MRW. TP and $NH_4^+$-N are indicators of nutrient pollution, $COD_{Mn}$ can be used as an indicator of pollution with oxygen-demanding organic matter.

Table 5-1 selected features to simulate the water quality in the MRW

| Features | Description | TP | $NH_4^+$-N | COD |
|---|---|---|---|---|
| ***Climatic variable*** | | | | |
| PotEvap_tavg | Potential evaporation rate (W $m^{-2}$) | √ | √ | √ |
| Rainf_f_tavg | Rain precipitation rate (kg $m^{-2}$ $s^{-1}$) | √ | √ | √ |
| ***Hydrology*** | | | | |
| Qs_acc | Surface runoff (kg $m^{-2}$) | √ | √ | √ |
| Qsb_acc | Baseflow-groundwater runoff (kg $m^{-2}$) | √ | √ | √ |
| ***Soil property*** | | | | |
| SoilMoi0_10 | Soil moisture in 0-10 cm (kg $m^{-2}$) | √ | √ | √ |
| SoilMoi10_40 | Soil moisture in 10-40 cm (kg $m^{-2}$) | √ | √ | √ |
| SoilMoi40_100 | Soil moisture in 40-100 cm (kg $m^{-2}$) | √ | √ | √ |
| SoilMoi100_200 | Soil moisture in 100-200 cm (kg $m^{-2}$) | √ | √ | √ |
| SoilTM0_10 | Soil temperature in 0-10 cm (K) | √ | √ | √ |
| SoilTM10_40 | Soil temperature in 10-40 cm (K) | √ | √ | √ |
| SoilTM40_100 | Soil temperature in 40-100 cm (K) | √ | √ | √ |
| SoilTM100_200 | Soil temperature in 100-200 cm (K) | √ | √ | √ |
| ***Anthropogenic variables*** | | | | |
| Population | Population | | | |
| P_app | P fertilizer applied (g $m^{-2}$ cropland $yr^{-1}$) | √ | | |
| N_app | N fertilizer applied (g $m^{-2}$ cropland $yr^{-1}$) | | √ | |
| Fer_app | Application of fertilizer (g $m^{-2}$ cropland $yr^{-1}$) | | | √ |
| ***Land use/cover*** | | | | |
| Agriculture | Percentage of agricultural land (%) | √ | √ | √ |
| Nature | Percentage of natural land (%) | √ | √ | √ |
| Urban | Percentage of urbanized land (%) | √ | √ | √ |

To understand the relationship between water quality and related properties (e.g. variables of climate, hydrology, soil properties, anthropogenic factors, and land use/cover), a database was set up to develop the machine learning models based on the location of sampling sites in the MRW (Table 5-1) and the sources of the data were shown in the Table 5-2. The land use information was interpreted manually based on TM digital images and the accuracy of the six classes of land use (i.e. Cropland, Woodland, Grassland, Water body, Built-up land, and Unused land) was above 94.3% while the overall accuracy of the 25 subclasses was above 91.2% (Liu et al., 2014). The land use/cover in the MRW was further classified into 3 categories: nature, agriculture, and urban.

Table 5-2 Sources of the data for the database

| ***Indicator*** | ***Sources*** |
|---|---|
| PotEvap_tavg,Rainf_f_tavg,Qs_acc,Qsb_acc,SoilMoi0_10,SoilMoi10_40,SoilMoi40_100,SoilMoi100_200,SoilTM0_10,SoilTM10_40,SoilTM40_100,SoilTM100_200 | Noah land surface model (Noah-LSM, Kumar et al., 2019) based on the NASA GLobal Land Data Assimilation System (GLDAS) |
| Population | WorldPop & Center for International Earth Science Information Network (2018) |
| P_app,N_app,Fer_app | Spatial dataset provided by Lu and Tian (2017) |
| Agriculture,Nature,Urban | Land use data in the year of 2000, 2005, 2010, and 2015 provided by the Data Center for Resources & Environmental Science, Chinese Academy of Sciences (RESDC) |

### 5.2.3 Impacts of urbanization on local or regional climate extremes

We propose a grid-based interpretable machine learning approach to simulate water quality and identify the properties that control water quality with spatial data of 0.25-degree resolution. Each property was further classified into 108 sub-features according to location of the central point of the grid. In other words, 1836 sub-features (108×17) were used as input features for modelling. Additionally, the locations of the sampling sites were considered. Thus, 1837 input features (i.e. 1836 sub-features for watershed feature and 1 feature for location of sampling sites) were used for the final models. Specifically, to understand the impacts of different streamflow regimes on water quality, the machine learning models for streams and reservoirs were set up, respectively.

For the machine learning models, we used the RFR method, which is a suitable method because it detects nonlinearity between independent variables and is robust to outliers (Wang et al., 2021e). As an ensemble learning method, it is based on a large number of individual decision trees. Each tree was independent during the modelling process because it was built with a unique bootstrap sample of the training dataset (Breiman 2001). The models were verified with the observed data and the key hyperparameters were identified with ***k***-fold cross-validation. As a resampling method to tune the hyperparameters of the models, the ***k***-fold cross-validation was usually used to evaluate models with limited samples. The ***k*** in this method means the number of groups. In this study, we used 10-fold cross-validation to identify the number of trees (n_estimators), the minimum number of samples required to split a node (min_sample_split), the minimum number of samples required to be at a leaf node (min_samples_leaf), the number of features to consider when looking for the best split (max_features), and the maximum depth of the tree (max_depth) for the water quality models. Based on related studies, 80% and 20% of entire dataset were resampled randomly and further

split into the training and testing dataset (Wang et al., 2021e). The mean absolute error (MAE), mean squared error (MSE), root mean square error (RMSE), $R^2$ coefficient of determination ($R^2$), percent bias (PBIAS), and Kling-Gupta efficiency (KGE) (Gupta et al., 2009) were used to evaluate the performance of the model.

The SHAP analysis was used to identify the major watershed properties that control the water quality at different scales. The SHAP analysis, which is based on game theory, can calculate the contribution of each property to the output of the models (Strumbelj and Kononenko, 2014; Lundberg and Lee, 2017). The contribution of each variable to the model output is allocated based on the marginal contribution (Lundberg and Lee, 2017). To identify the spatial distribution of watershed properties, Spearman's test of rank correlation was used to estimate the correlation between SHAP values of input properties and indicators of topography or land use including slope, elevation, topographic wetness index (TWI), and percentage of agricultural or urbanized land use.

## 5.3 Results

### 5.3.1 Model training and validation

We computed RFR models of the water quality parameters (i.e. $NH_4^+$-N, TP, and $COD_{Mn}$) for stream and reservoir, respectively (Table 5-2). The simulated data fitted well with the observed values based on the MAE, MSE, RMSE, $R^2$, PBIAS, and KGE.

Table 5-3 Performance of models during the training and testing periods

| | | Stream | | | Reservoir | | |
|---|---|---|---|---|---|---|---|
| | | $NH_4^+$-N | TP | $COD_{Mn}$ | $NH_4^+$-N | TP | $COD_{Mn}$ |
| Training period | MAE | 0.158 | 0.023 | 0.394 | 0.055 | 0.014 | 0.157 |
| | MSE | 0.094 | 0.003 | 0.520 | 0.008 | 0.001 | 0.051 |
| | RMSE | 0.307 | 0.050 | 0.721 | 0.088 | 0.032 | 0.226 |
| | $R^2$ | 0.619 | 0.709 | 0.691 | 0.811 | 0.674 | 0.885 |
| | PBIAS | -0.541 | -0.622 | -0.236 | 0.436 | -0.674 | 0.280 |
| | KGE | 0.616 | 0.577 | 0.544 | 0.779 | 0.562 | 0.770 |
| Testing period | MAE | 0.159 | 0.023 | 0.390 | 0.060 | 0.013 | 0.170 |
| | MSE | 0.088 | 0.002 | 0.321 | 0.008 | 0.001 | 0.054 |
| | RMSE | 0.296 | 0.039 | 0.567 | 0.089 | 0.026 | 0.232 |
| | $R^2$ | 0.724 | 0.591 | 0.585 | 0.828 | 0.603 | 0.861 |
| | PBIAS | 3.494 | 5.120 | 1.509 | 0.794 | 4.750 | -0.400 |
| | KGE | 0.762 | 0.773 | 0.598 | 0.738 | 0.620 | 0.723 |

The performance of the reservoir models was better than that of the water quality models for the stream (Table 5-3, Fig 5-2), particularly for the $COD_{Mn}$ models. The extreme values may

not be well predicted by the models for the stream and the high concentrations may be underestimated by the models (Fig 5-2).

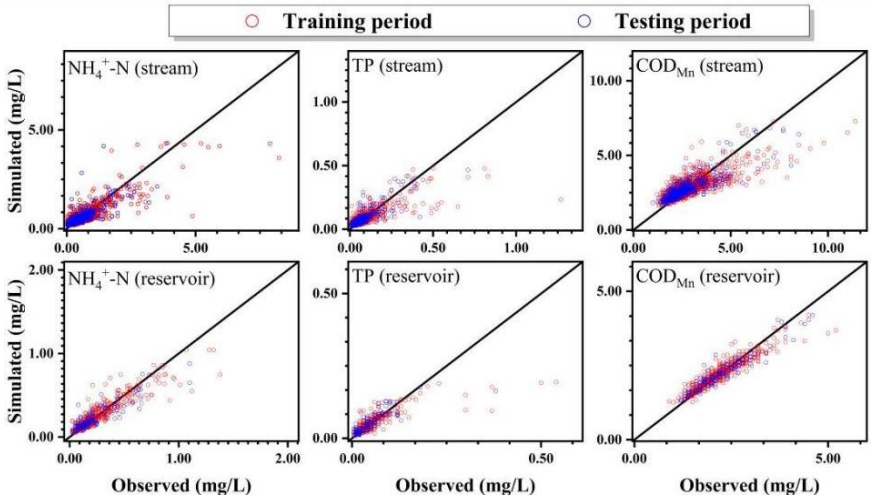


Fig 5-2 The training and testing results of RFR models

We also tested the spatial accuracy of the models. Based on the estimated local PBIAS of the models, the $COD_{Mn}$ models achieved the highest accuracy, followed by the TP and $NH_4^+$-N models (Fig 5-3). For the water quality models of stream, the sites of low accuracy were found in the outlet of the MRW which may be the induced by interference from the invaded sea water from tidal action. The low accuracy of sites from the reservoir models were found in downstream of the sub-watershed where multiple sub-watersheds meet.

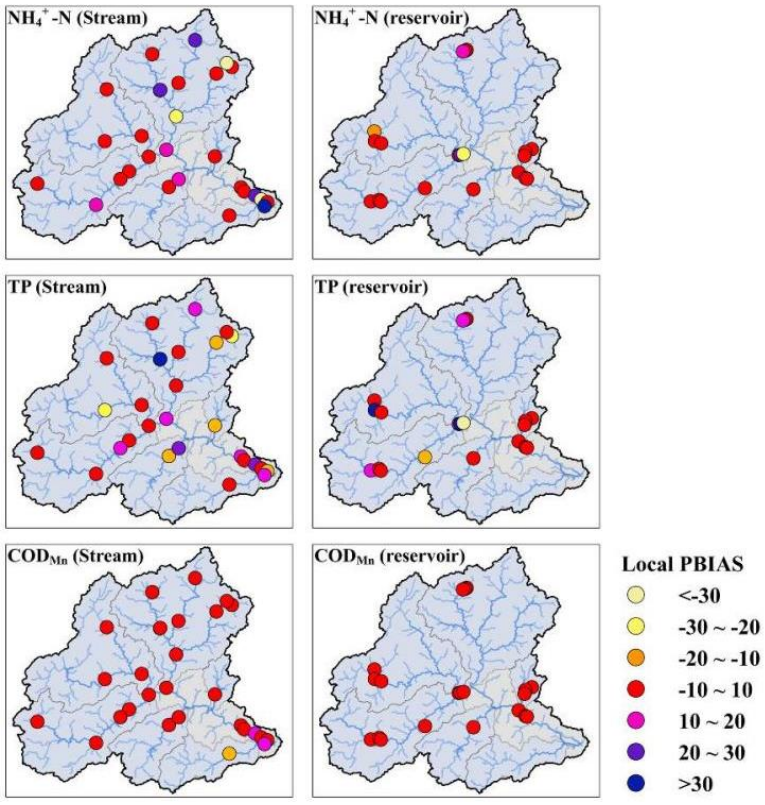


Fig 5-3 Spatial performance of the models during the training and testing period in the MRW

## 5.3.2 Water quality

Fig 5-4 shows an overview of the water quality in the streams and reservoirs. The variation and absolute values were higher in the streams than in the reservoirs. Meanwhile, the water quality in the streams was poorer than that in the reservoirs. And the median values of the three indicators in the streams were higher than that in the reservoirs (Fig. 5-4)

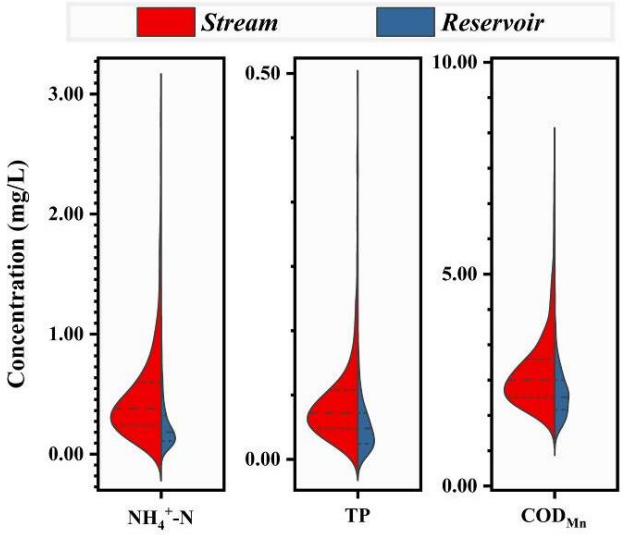


Fig 5-4 Overview of the water quality in the MRW

The inter-annual variation of water quality is shown in Fig 5-5. The highest variations were observed in $NH_4^+$-N, followed by $COD_{Mn}$ and TP. There were no obvious trends.

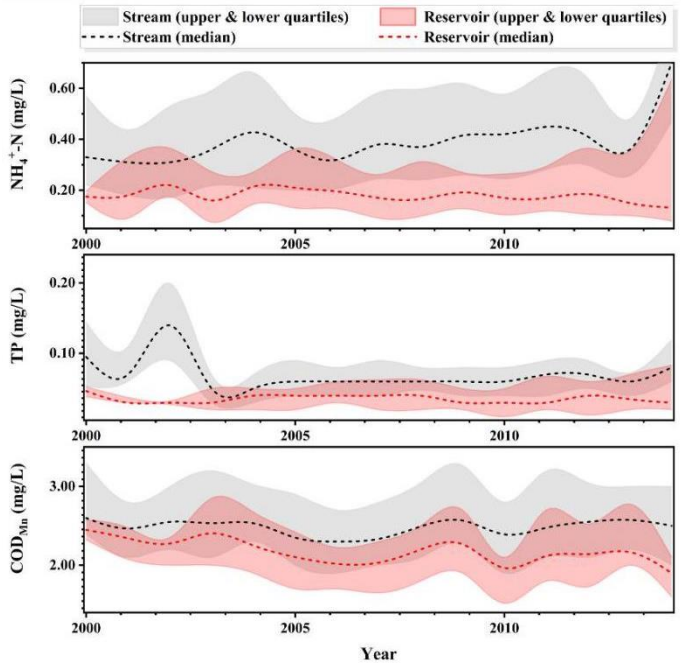


Fig 5-5 Inter-annual variations of water quality during 2000-2014

The seasonal variation of water quality is shown in Fig 5-6. The patterns are different for streams and reservoirs. High concentrations of $NH_4^+$-N were found from April to October in

the reservoirs, and lower concentrations were found in the streams. Similar trends were also observed for TP. Compared with $NH_4^+$-N and TP, the seasonal patterns of $COD_{Mn}$ in the streams and reservoirs in the MRW.

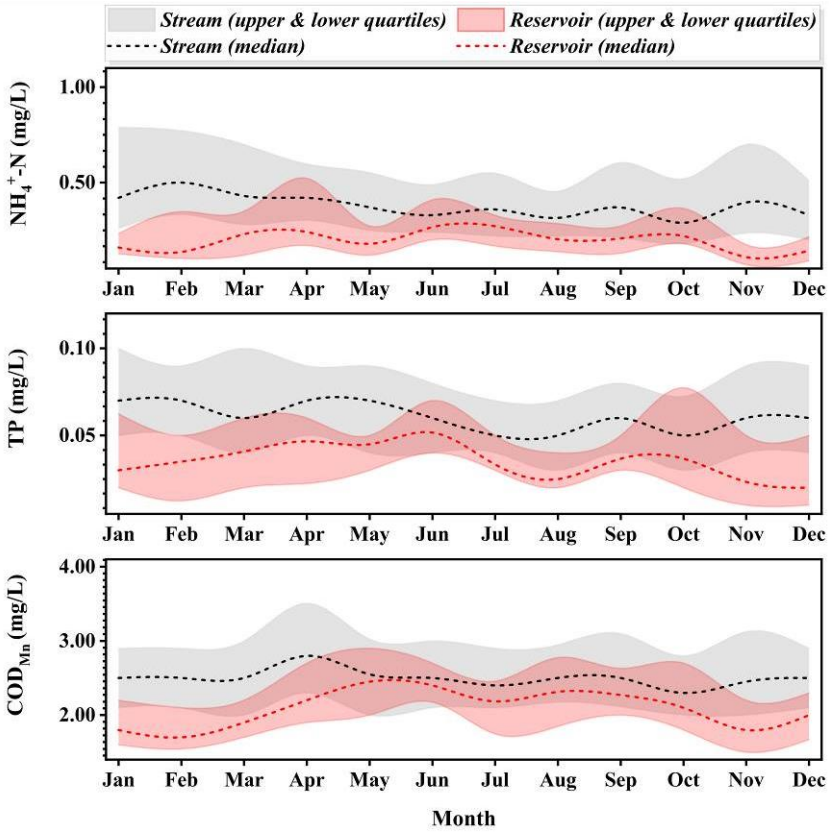


Fig 5-6 Seasonal variation of water quality during 2000-2014

## 5.3.3 Watershed properties and water quality

The features driving the deterioration of the water quality were identified with the SHAP analysis and the SHAP values were calculated for each input variable of the RFR models (Fig 5-7). A higher SHAP value means a stronger driving force. A positive SHAP value means a positive influence on nutrient concentrations in the related model. As shown in Fig 5-7, population, soil conditions (including soil moisture and soil temperature) were the major factors that controlled water quality. The soil moisture at the soil surface may play a different role in nutrient export than soil moisture in the deep soil. Increasing SoilMoi0_10 increases the concentrations of $NH_4^+$-N and TP in the reservoirs. However, increasing SoilMoi100_200 may reduce the concentrations $NH_4^+$-N and TP in the streams (Fig 5-7).

The concentrations of NH4+-N and TP in the streams are more sensitive to surface runoff than to the baseflow-groundwater runoff, In the reservoirs baseflow-groundwater runoff is more important. The increasing surface runoff and baseflow-groundwater increases concentrations of $NH_4^+$-N and TP in the reservoirs, but decreases concentrations of $NH_4^+$-N and TP in the streams (Fig. 5-7) Different trends were observed for COD. CODMn is more sensitive to

surface runoff than baseflow-groundwater. The increasing storm surface runoff increases CODMn in the reservoirs (Fig. 5-7).

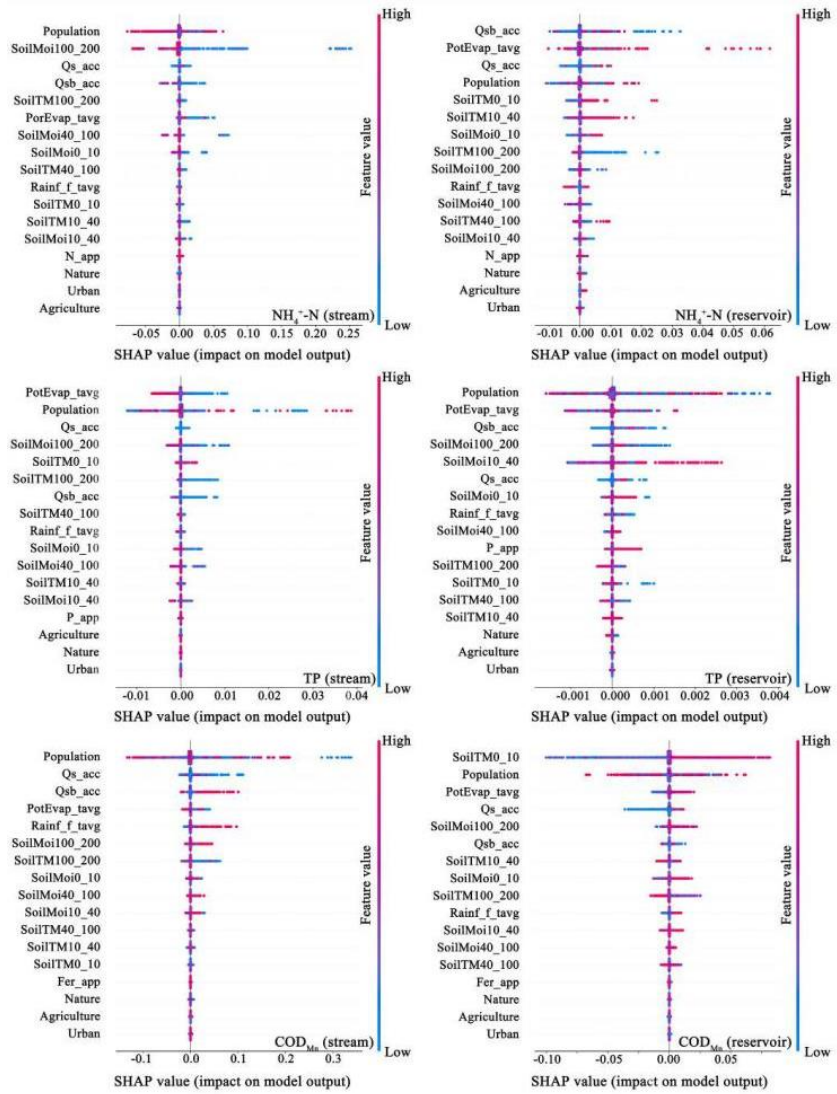


Fig 5-7 SHAP results of variable importance based on the RFR models

The spatial distributions of the mean absolute SHAP values are shown in Figs 5-8, 5-9, 5-10, 5-11, 5-12, and 5-13. The spatial distribution of the SHAP values of land use/cover (Fig 5-13) and fertilizer (Fig 5-12) were low compared other variables and have similar contribution to water quality. Other variables have a much higher influence like e.g. the climate and hydrological variables (Figs 5-8, and 5-9). The soil variables show a similar distribution with high SHAP values in the southeast (Figs 5-10, and 5-11). The highest number of grid cells with a SHAP value were found for population variable. The grids cells located in the northeast have a bigger influence on regional water quality (Fig 5-12). Additionally, the distributions of TP were different from other two models.

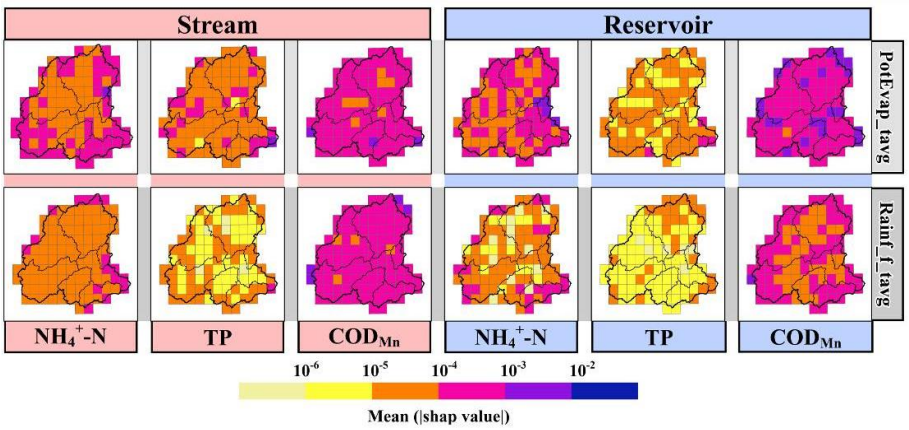


Fig 5-8 Spatial distribution of the SHAP results for climate variables

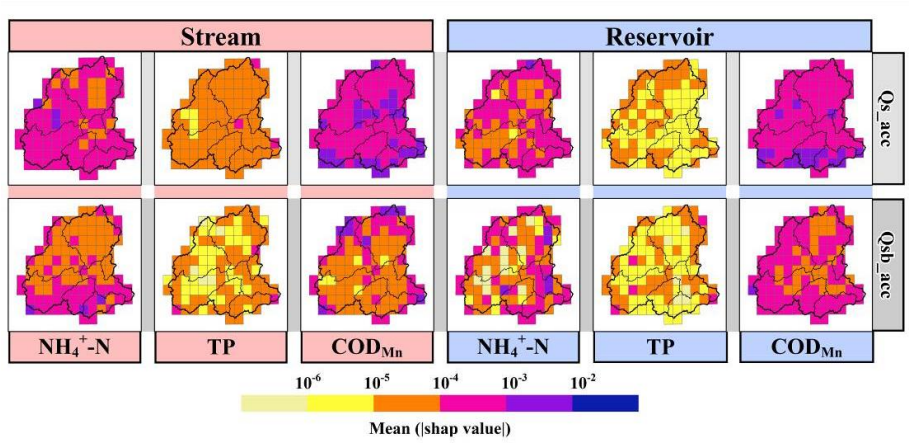


Fig 5-9 Spatial distributions of the SHAP results for the hydrology in the MRW

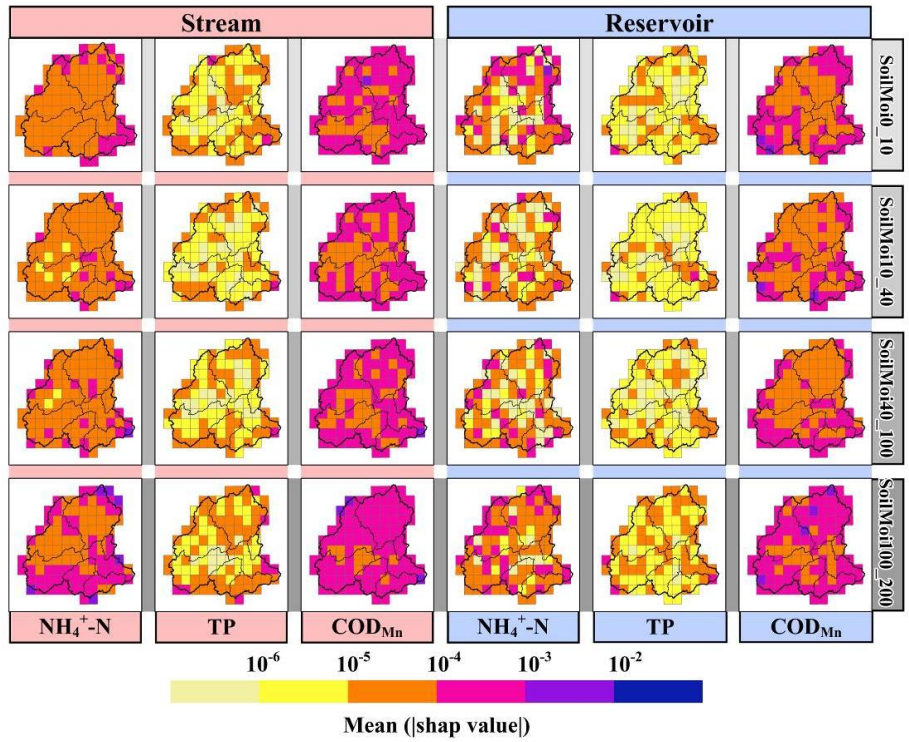


Fig 5-10 Spatial distributions of the SHAP results for the soil moisture in the MRW

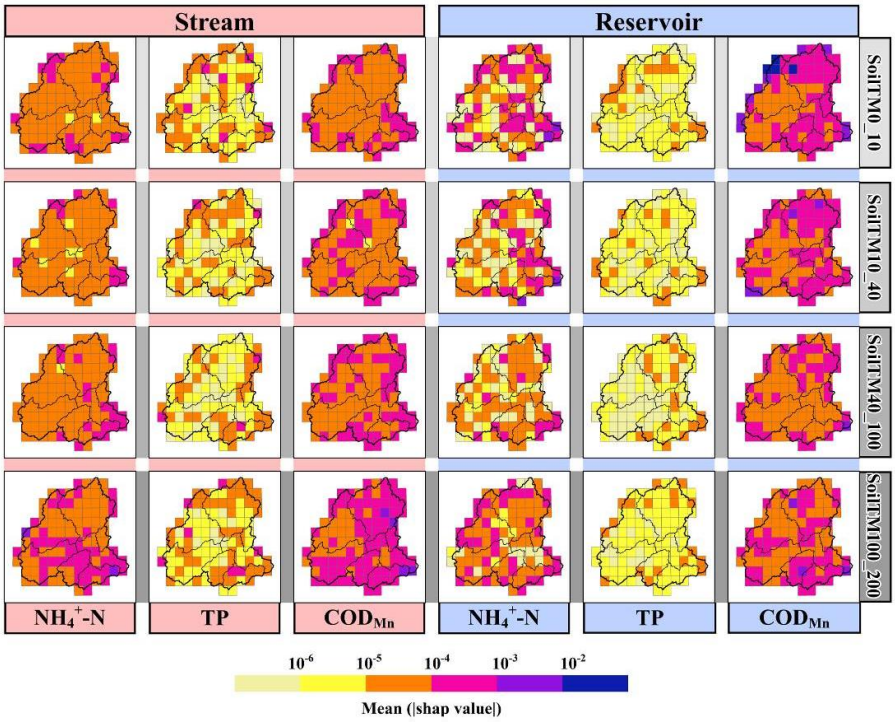


Fig 5-11 Spatial distributions of the SHAP results for the soil temperature in the MRW

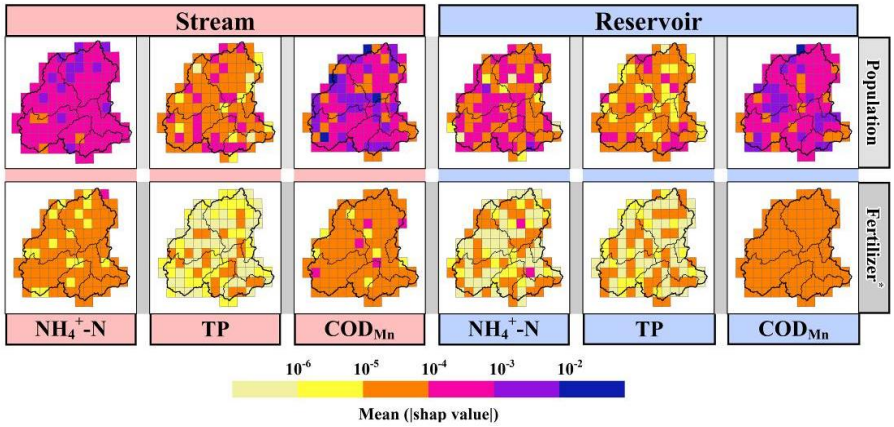


Fig 5-12 Spatial distributions of the SHAP results for the anthropogenic varibales in the MRW

(*Note: Fertilizer for $NH_4^+$-N, TP, and $COD_{Mn}$ were used P_app, N_app, and Fer_app, respectively)

To understand the spatial variation of the variables of the water quality models, Spearman correlation analysis was used to identify the relationships between the variables with high spatial variation and land use or topography indicators (Fig 5-14). Of all watershed properties, urbanization has the most significant influence on the spatial heterogeneity of water quality. Significant relationships also existed between the SHAP values of population and the percentage of urban areas in the MRW. The Spearman correlation coefficients for NH4+-N, TP and CODMn suggest that the population in the less urbanized areas may contribute more than

in highly urbanized areas. Compared to the stream models, the effects of urbanization on the reservoirs contribute more to the spatial heterogeneity of the watershed variables, especially, for the CODMn models. In the reservoirs, the effect of urbanization is higher.

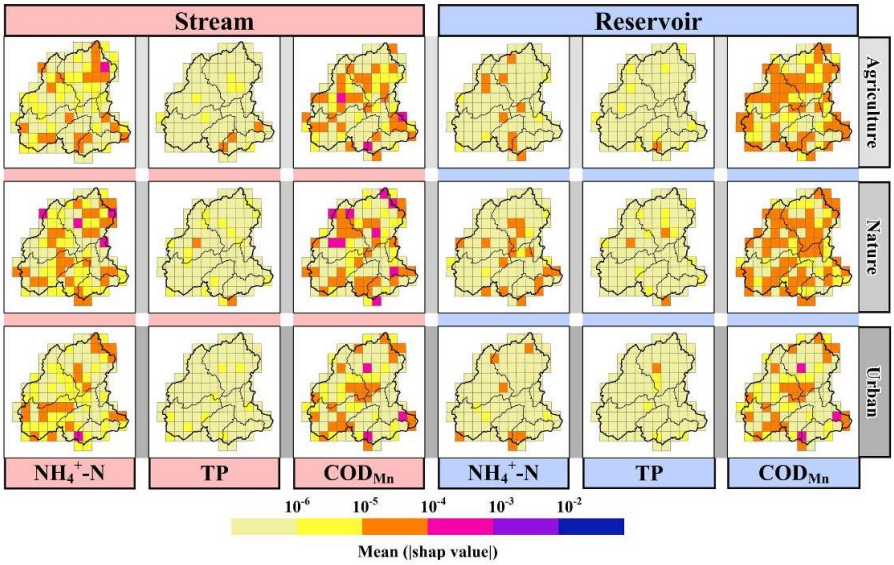


Fig 5-13 Spatial distributions of the SHAP results for the land use/cover in the MRW

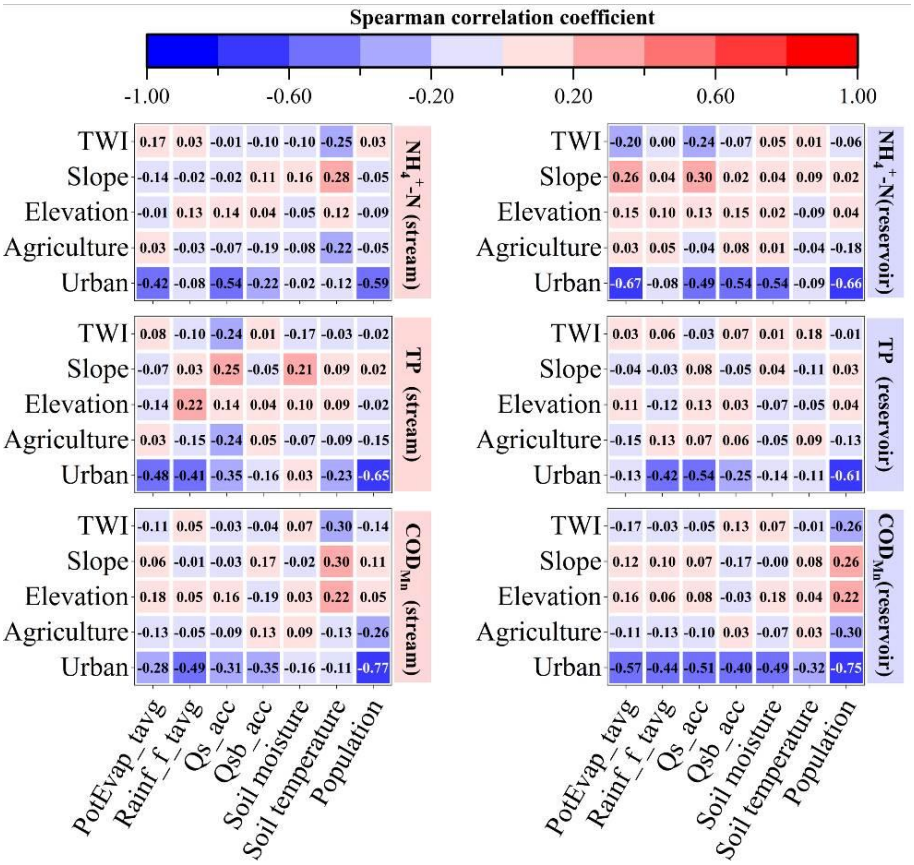


Fig 5-14 Factors influence the spatial variations of the SHAP value of variables for water quality

## 5.4 Discussion

### 5.4.1 Interpretable machine learning approach

Statistical models like multiple linear regression and geographically weighted regression are effective tools to understand the relationship between watershed properties and water quality (Huang et al., 2013c; 2015). However, these methods do not cover the spatial settings. To address these limitations, we propose a grid-based interpretable machine learning method. The watershed properties of each grid cell are used as independent drivers for the models. The accuracy of the models was estimated on global and local scales to provide a full picture of water quality modelling for the entire watershed (Figs 5-2, and 5-3). The estimated global |PBIAS| of the models during training and testing periods were 0.236-0.674 and 0.400-5.120, respectively (Table 5-2) was satisfying.

Higher accuracy of the water quality models was found in the reservoir compared to the stream (Table 5-2). The water quality in the reservoir may be less sensitive than that in the stream, and extreme values can be predicted in the models for the reservoir. The high values of water quality parameters may be underestimated for these models of stream water quality (Fig 5-2). The estimated accuracy of sites in the mouth of the MRW was relatively low because of saltwater intrusion (Fig 5-3). In reservoirs, the models for $COD_{Mn}$ achieved the highest accuracy. In contrast, higher accuracy was found in stream water quality models for nutrients (i.e. NH4+-N, TP). This suggested that the reservoir may also modified the relationships between watershed properties and water quality.

### 5.4.2 Relationships between watershed properties and water quality

Several studies have described seasonal variation of water quality as a result of agricultural activities, climate variability, biochemical processes and so on (Kaushal et al., 2014; Greaver et al., 2016; Duan et al, 2021; Lei et al., 2021). For example, Lei et al (2021) observed lower water quality in winter compared to summer in a German lowland catchment. In Southeast China, low water quality occurred more frequently in the wet season in summer compared to the dry season in winter (Zhang et al., 2019; Huang et al., 2021). The climate is also an important driving factor of water quality. The German catchment is located in a high-latitude marine climatic zone with moderate precipitation, whereas the watersheds in Southeast China are located in a subtropical monsoon climate with approximately 70% precipitation during summer. High precipitation may transport more pollutants to water bodies after fertilizer application. Additionally, biogeochemical transformations may also decrease with increased runoff (Gallo et al, 2015; Greaver et al., 2016). High concentrations of $NH_4^+$-N, TP, and $COD_{Mn}$ were observed from May to August in the reservoirs, and the observed seasonal variability of water quality in streams was low compared to the reservoirs (Fig 5-6). This indicates that

reservoirs may amplify the effects of climate variability on water quality. A previous study in Southeast China also found that reservoirs or dams may enhance evapotranspiration in the watershed (Zhang et al., 2020a).

Relationships between land use and water quality have been clarified in previous study, and they suggested that human-impact land use may be negatively correlated to water quality and natural land use may be positively correlated to water quality (Tu 2009; Lee et al., 2013; Lei et al., 2021). In this study, high SHAP values were observed with low feature values of natural land use (Nature) for the NH4+-N and TP models in the reservoirs which indicated that the high percent of natural land use would reduce the concentrations of nutrients in the reservoirs. In contrast, this relationships between land use and water quality may be weak in the stream compared to that in the reservoirs (Fig 5-7). Though the population and urbanized land use are highly correlated, they play different roles for the water quality. The importance of population (Population) was ranked high in the all models (Fig 5-7), while the percentage of urbanized land use (Urban) is not as important as the population. It indicated the effects of population on water quality may be nonlinearity.

The spatial distribution of the population influences water quality in the MRW (Figs 5-12, and 5-14). Negative correlations were also found between the SHAP values of population and percentage of the urbanized land use (Fig 5-14), suggesting that point source pollution per inhabitant may be reduced with increasing urbanization. During urbanization, more sewage treatment plants are built and therefore point sources are reduced. Regional population and the fraction of residential land have a significant influence on water quality (Wang et al., 2021e). Similar results were also found in previous studies. Tu (2009) found in Boston that the percentages of commercial and industrial land have stronger positive relationships with the concentration of water pollutants in less urbanized areas than in highly-urbanized areas. Wang et al. (2021e) found that population density is positively correlated with water quality parameters where the population density is less than 3000 people/km$^2$. An increasing population in rural areas may lower water quality. Urbanization has the most important influence on the spatial distribution of water quality parameters. Based on Spearman correlation analysis, more properties were correlated between water quality and urban land use in reservoirs than in stream models (Fig 5-14). Increasing urbanization increases the area of impervious surface and modifies the regional climate and streamflow (Mishra et al., 2015; Li et al., 2020b). Urbanization in the MRW may reduce the effects of climate, hydrology and soil properties on water quality, especially for the organic matter in the reservoir. (Fig 5-14).

Soils act as primary sink and source of terrestrial contaminants, influence water quality through subsurface and soil water flows (Huang et al., 2017; Liu et al., 2017; Wang et al., 2021d) and biochemical processes in the change the patterns of pollutant export (Huang and Hall, 2017; Lei et al., 2021, Xie et al., 2023). Coupled with the regional environmental setting, the soil

property may influence the magnitudes of nutrient exports by changing the patterns of the sediment in the watersheds (Drewry et al., 2006; Lintern et al., 2018). For this reason, the nutrients export in the reservoirs may be more sensitive to the increasing runoff in the watersheds for the highly sediment exports in the storm events (Fig. 5-7). Besides, the processes in the soil modify the mobilization and delivery processes of the pollutants in the watersheds (Huang and Hall, 2017; Lintern et al., 2018; Xie et al., 2023). Huang and Hall (2017) found that increasing soil moisture modifies process mineralization leads to a higher release of organic matter. In this study we also found that soil moisture in 100-200 cm (SoilMoil100_200) was the important feature influence the CODMn models for streams (ranked 6/17) and reservoirs (ranked 5/17). However, there is a different relationship between the nutrient concentrations and soil conditions in streams and reservoirs. Increasing temperature and moisture in the bottom soil (i.e. SoilTM100_200, SoilMoil100_200) decreases the concentrations of NH4+-N and TP in the stream while increasing temperature and moisture in the topsoil (i.e. SoilTM0_10, SoilMoil0_10) increases the concentrations of $NH_4^+$-N and TP in reservoirs (Fig. 7). The delivery of the pollutants from the catchment to receiving waters via subsurface flow pathways is strongly affected by soil hydrological properties (Lintern et al., 2018). The moisture of the soil may change the connectivity to the stream, and the high connectivity means more pollutants may entry into the streams with runoff. The studies in the USA and Sweden indicated that the greater pollutants may be found in the well-draining soil where the soil is more wet (Arheimer and Liden, 2000; Franklin et al., 2013). In this study, we also found that the nutrients exports were also highly related to the soil moisture in the watershed (Fig. 5-7).

Streamflow is a holistic driver regulating material and energy flows in a catchment. Baseflow-groundwater runoff and surface runoff were identified as the key factors for the transport of nonpoint pollutants (He and He, 2008; Zhu et al., 2019). Compared to the nutrients in the stream, baseflow-ground water runoff (Qs_acc) and storm surface runoff (Qsb_acc) may transport more pollutants from non-point sources to reservoirs. A previous study in the MRW also indicated that storm events may significantly enhance nutrient exports in reservoirs compared to streams (Zhang et al., 2019). Additionally, we further confirmed that storm surface runoff contributes more to the nutrient export in the reservoir than baseflow-ground water runoff. A higher rank of storm surface runoff for the $NH_4^+$-N (ranked 1/17) and TP (ranked 3/17) models in reservoirs was observed (Fig 5-7). In contrast, baseflow-ground water runoff may play a more important role for the transport of nutrients than the storm surface runoff (e.g. $NH_4^+$-N (ranked 3/17) and TP (ranked 3/17)). Compared to the nutrient exports, the COD in the MRW may be more sensitive to baseflow-ground water runoff and storm surface runoff. Increasing baseflow-ground water runoff and storm surface runoff enhance exports of organic matter in the watershed (Fig 5-7).

### 5.4.3 Outlook

In this study we analyzed the spatial distributions of the properties which influence the water quality in our watershed with a grid-based interpretable machine learning approach. This method can shed light on how to improve water quality in the watershed by considering the spatial and temporal variability of watershed properties. This method can also be used to illustrate the dependency of biogeochemical variables in response to changes in land use, management, and policy. The increase of pollution with increasing population may be reduced with increasing level of urbanization level in the MRW, and the other impacts from urbanization were also evaluated with the proposed method. The method proposed in this study can be used to evaluate the source of the pollutants and help to understand the export of nutrients and organic matter at different scales.

## 5.5 Conclusions

We developed a grid-based interpretable machine learning method to understand the spatial relationships between watershed properties and water quality in the MRW. The proposed method exhibited acceptable performance. The water quality in the reservoir may be less sensitive than that in the stream, and extreme values can be predicted in the models for the reservoir. Higher accuracy of the water quality models was found in the reservoir compared to the stream. The climate change, soil conditions, population and hydrological conditions were identified as the most important features influence the water quality in the MRW. The nonlinear relationships between water quality and watershed properties were further clarified in this study. The reservoirs may amplify the effects of climate variability on water quality. The effects of the urbanization may modify the distributions of important feature that control the regional water quality. The point source pollution per inhabitant may be reduced with increasing urbanization. This study provides an in-depth understanding of the relationship between water quality and watershed properties, temporally and spatially.

# Chapter 6 Use of interpretable machine learning to identify the factors influencing the nonlinear linkage between land use and river water quality in the Chesapeake Bay watershed

**Zhenyu Zhang, Jinliang Huang, Shuiwang Duan, Yaling Huang, Juntao Cai, and Jing Bian**

**Ecological Indicators, 2022, 140:108977**

## Abstract

Understanding the relationship between land use and water quality is essential for effective watershed management. However, it remains challenging to identify such a relationship owing to its nonlinearity. We developed an interpretable machine learning method that integrated the random forest regression (RFR) model with the Shapley Additive exPlanations (SHAP) method to explore the relationship between water quality and land use in the Potomac River Basin (PRB), the second largest tributary entering Chesapeake Bay from 2006 to 2019. The water quality of the 26 sub-watersheds, classified into five types (natural, forested, agricultural, mixed, and urbanized), was investigated using statistical methods and scenario analysis. The results showed that the models employed were effective in predicting the water quality. The mean absolute error (MAE), root mean square error (RMSE), percent bias (PBIAS), $R^2$ coefficient of determination ($R^2$), and Kling-Gupta efficiency (KGE) were 0.011-0.159 mg/L, 0.019-0.219 mg/L, -0.14-0.64%, 0.79-0.99, and 0.69-0.98, respectively, during the training period, which were 0.010-0.201 mg/L, 0.017-0.292 mg/L, -1.87-0.41%, 0.82-0.99, and 0.80-0.97, respectively, during the testing period. The threshold effects of land use patterns were obvious for water quality indicators with high concentrations (i.e., TN and $NH_4^+$-N). In contrast, the water quality at low concentrations (i.e., TP and $NO_3^-$-N) may be more sensitive to wetland or barren land with changing climate. Agricultural activities and urbanization could be the dominant factors determining nutrient export to the PRB. Meanwhile, the typical 'sink' for the nutrient such as wetland may change into the 'source' for different nutrient. This study provides an in-depth understanding of how riverine nutrient export responds to the land use gradient in the Chesapeake Bay watershed.

Keywords: Land-use; water quality; nonlinearity; threshold effect; Potomac River basin

## 6.1 Introduction

Global cycles of nitrogen (N) and phosphorus (P) are altered by human activities, which have contributed to increased coastal eutrophication (Baker et al., 2003; Duan et al., 2012; Huang et al., 2015). Land use within a watershed is important for the water quality of rivers, lakes, estuaries, and coastal waters (Huang et al., 2013c; Hur et al., 2014; Duan et al., 2021). Therefore, exploring the linkage between land use and water quality, particularly in coastal watersheds, is critical for developing watershed management strategies and controlling land-based pollution in coastal bays (Huang et al., 2013c; Mulkey et al., 2017).

Previous studies have suggested that land use change may alter the characteristics of N and P exports by linking water quality to specific land use types (Huang and Klemas 2012; Lacher et al., 2019; Ahmadisharaf et al., 2020). Natural vegetation, such as forests, can filter pollutant particles and reduce water pollution (Lei et al., 2021).In contrast, human-impacted land use may increase the risk of water quality degradation through additional sources of non-point or point pollution (Zhang et al., 2020b; Duan et al., 2021). Jordan et al. (2018) found that TP and dissolved phosphate increased with a high percentage of cropland. Duan et al. (2021) found that non-point agricultural sources made a significant contribution to N export in streams.

However, land use is not the only factor controlling regional nutrient export because the relationship between land use and riverine nutrients may be modified by regional climate variability (Huang and Klemas, 2012; Zhang et al., 2020b). Land use change, coupled with climate variability, can significantly affect nutrient export (Duan et al., 2012; Huang et al., 2021). Huang and Klemas (2012) indicated that the percentage of built-up land is a good predictor of downstream water quality, whereas the linkage between water quality and landscape variables during wet years is stronger than that during dry years. Natural vegetation may also be an important source of pollution under specific conditions such as nutrient export from forest soils during storm events (Lee et al., 2013; Duan et al., 2021).

Recently, the threshold effects of land use patterns have been identified (Huang et al., 2015; Liu et al., 2021c). The patterns of nutrient export may be modified by changing vegetation cover (Mulkey et al., 2017; Jordan et al., 2018; Lacher et al., 2019). For example, using geographically weighted regression, Huang et al. (2015) found that the built-up may be more associated with pollution in watersheds with less urbanized areas. Therefore, it is necessary to understand the nonlinear relationship between land use and water quality.

Conceptually, riverine nutrients are strongly controlled by land use, and many effects have been made in the past decades to identify this relationship. Conventional statistical methods, such as stepwise regression, redundancy analysis, and correlation analysis, were commonly applied in this study. For example, using nonlinear regression (i.e. power regression, exponential regression, quadratic regression, and segmented regression), Liu et al. (2021c)

analyzed the nonlinearity and threshold effects of landscape patterns on water quality in a rapidly urbanized watershed. However, it may be difficult to identify the relationship between multiple factors and water quality. Compared with conventional statistical models, machine learning models may be more effective for modelling water quality with multiple factors. For example, Zhou et al. (2016) demonstrated how point source pollution weakens the land use-water quality correlation using a self-organizing map-based approach. However, the interpretations of machine learning models were limited to quantifying the impact of related factors on riverine nutrients (Wang et al., 2021d; 2021e). Recently, Shapley Additive exPlanations (SHAP) was developed to interpret machine learning models with improved computational performance and consistency with human intuition (Lundberg and Lee, 2017). The SHAP method has been proved to be a powerful interpreter for machine learning models (Li et al., 2021; Wang et al., 2021d).

Chesapeake Bay, a large temperate estuary in the mid-Atlantic region of the United States, has been proven to be a nutrient over-enrichment area that has produced significant symptoms of anthropogenic eutrophication (Duan et al., 2012; Jordan et al., 2018). In the past decades, many efforts have been made to improve water quality in Chesapeake Bay (Mulkey et al., 2017; Ator et al., 2020). Understanding the linkage between nutrient exports and land use patterns is helpful for watershed management and updating the knowledge for watershed modelling (Baker, 2003; Hernandez Cordero et al., 2020). In this study, we hypothesized that the impact of climate variability can be quantified by water temperature and precipitation, the impact of topography can be qualified by slope, the threshold effects of land use can be quantified by types of watershed, and the effects of climate variability with human interference can be quantified by sampling time. With these external factors, the nonlinear relationship between land-use gradient and water quality in the PRB was analyzed with an interpretable machine learning method by answering the following two questions: (1) how does the stream water quality respond to distinct land-use gradients in the Potomac River Basin? (2) what is the temporal variation in land use-water quality linkages?

## 6.2 Material and methods

### 6.2.1 Study area

Chesapeake Bay, the largest estuary in the United States, has experienced increased eutrophication over the past century (Mulkey et al., 2017, Yu et al., 2020). The Potomac River Basin (PRB) is the second largest tributary entering Chesapeake Bay in terms of water quantity, and the fourth largest river on the eastern coast of the United States (Belval and Sprague, 1999; Guardian et al., 2021) (Fig 6-1). It covers approximately 37,990 $km^2$, with an average flow of 306 $m^3/s$. More than five million residents rely on the PRB as their source of water for residential, industrial, and agricultural activities (Tanir et al., 2021). In addition, an

average of approximately 1,840,000 $m^3$ of water per day is withdrawn daily in the Washington area for water supply. Land use is distributed unevenly across the basin, with predominantly forested areas in the west, dense agriculture in the central area, and major population centers in the east. This pattern of land use follows a general transition in topography from steep mountains in the west to gently rolling terrain in the east (Miller et al., 1997).

The PRB lies in seven geological provinces, including the Appalachian Plateau, Valley and Ridge, Great Valley, Blue Ridge, Piedmont, and Triassic Lowlands in the upper PRB, and Coastal Plain Basin in the lower of the PRB. The river meanders from Fairfax Stone, West Virginia to Point Lookout, Maryland, which is located in parts of the Commonwealth of Virginia, Maryland, West Virginia and Pennsylvania state, and the District of Columbia (Blomquist et al., 1996, Bricker et al., 2014, Tanir et al., 2021). The PRB experiences a moderate climate pattern marked by cold winters and warm, humid summers, with a mean annual precipitation of 700–1200 mm/yr and monthly precipitation reaching its highest magnitude in June (Sridhar et al., 2019).

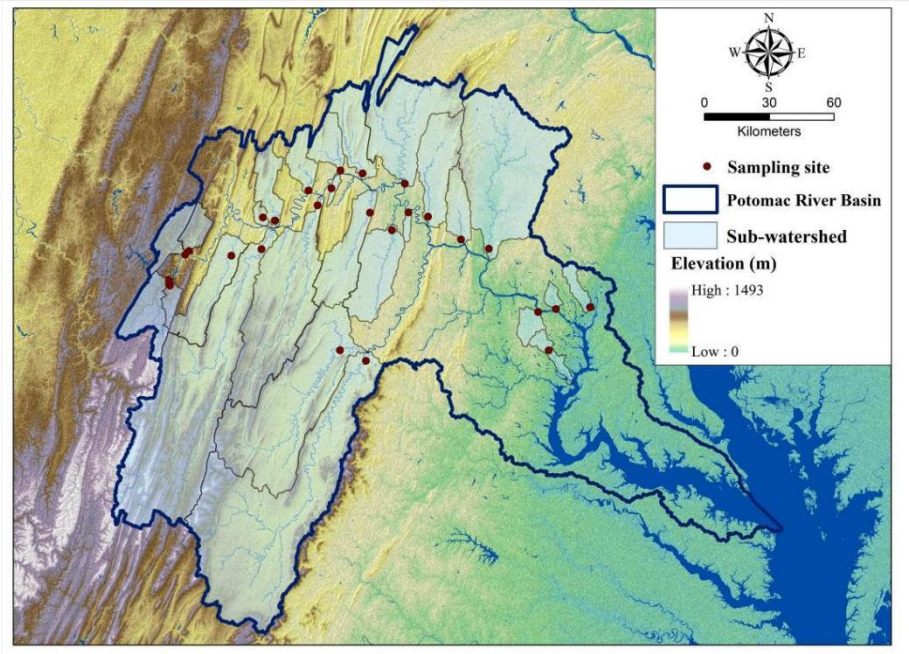


Fig 6-1 Sampling sites in the Potomac River Basin during 2006-2019 (Note: watershed boundary for 26 stations in derived from NHD, USGS)

## 6.2.2 Land Use and Land Cover

The land-use data in 2006, 2011, 2016, and 2019 within the PRB were considered in this study based on the National Land Cover Database (NLCD) (Fig S6-1). The land use categories were classified into eight classes based on the NLCD land cover legend for Level I: water, urban, barren, forest, shrubland, herbaceous, agriculture, and wetland. The estimated overall

accuracy of the NLCD2016 was 90.6% (Wickham et al., 2021).

To examine the threshold effects of land use patterns, the sub-watersheds in the PRB were classified as natural, forested, agricultural, mixed, and urbanized based on land use composition. Natural and forested watersheds were identified with forest that more than 80% and 50%-80%, respectively. Meanwhile, the agricultural watershed was identified with agriculture that more 50%, and the urban watershed was identified with urban that more than 50%. The remaining watersheds were classified as mixed.

Land use patterns in the PRB were shown in Fig S6-1. Overall, by 2019, the largest land-cover class in the PRB was forest (52.55%), followed by agriculture (25.21%), and urban (14.67%). Minor proportions of the watershed were covered by water (4.04%), wetland (2.05%), herbaceous (0.72%), shrubland (0.59%), and barren (0.17%). During the period of 2006–2019, the shrubland increased markedly by 23.43%. The urban and agriculture increased slightly by 0.43% and 0.75%, respectively (Table 6-1). Meanwhile, a decrease of 9.53% in barren was observed, followed by forest (decreased by 0.29%). The water and wetland remained almost unchanged during this period, decreasing by only 0.06% and 0.01%, respectively.

Table 6-1 Overview of land use change during period of 2006-2019 (%)

| Land use | Unchanged | Lost | Gained | Net gain/loss |
|---|---|---|---|---|
| Water | 99.91 | 0.09 | 0.03 | -0.06 |
| Urban | 100 | 0.00 | 0.43 | 0.43 |
| Barren | 86.39 | 13.61 | 4.08 | -9.53 |
| Forest | 99.24 | 0.76 | 0.47 | -0.29 |
| Shrubland | 58.46 | 41.55 | 64.98 | 23.43 |
| Herbaceous | 63.39 | 36.62 | 40.98 | 4.36 |
| Agriculture | 99.82 | 0.18 | 0.93 | 0.75 |
| Wetland | 99.90 | 0.10 | 0.09 | -0.01 |

## 6.2.3 Data Source

In this study, water quality data during 2006.1-2019.12 were obtained from Chesapeake Bay Program (www.chesapeakebay.net/data). Four representative parameters, including total phosphorus (TP), total nitrogen (TN), ammonia nitrogen ($NH_4^+$-N), and nitrate nitrogen ($NO_3^-$-N), were employed in this study. The total number of samples for TN, TP, $NH_4^+$-N, and $NO_3^-$-N was 5360, 4249, 3996, and 4067, respectively (Fig S6-2). Water temperature data were obtained from the Chesapeake Bay Program. Precipitation data were obtained from the CRU TS which was produced by the UK's National Centre for Atmospheric Science at the University of East Anglia's Climatic Research Unit (Harris et al., 2020). Based on the climate variability of the PRB, the seasonal cycle for sampling was divided into two periods (i.e. the dry and wet seasons). To understand the impact of topography, the slopes were estimated based on the

SRTM 90m Digital Elevation Database from the National Aeronautics & Space Administration (NASA, https://www.nasa.gov/). Additionally, the density of point source pollution in each sub-watershed was estimated based on the point source information provided by the Chesapeake Bay Program.

### 6.2.4 An interpretable machine learning method

In this study, an interpretable machine method was employed to understand the relationship between land use and water quality in the PRB based on Random Forest regression (RFR) and the corresponding SHAP. The nonlinear relationships between land use and water quality were quantified using the trained RFR models. As an ensemble learning method, a large number of individual decision trees were constructed in RFR, and each independent tree was planted using a unique bootstrap sample of the training dataset (Breiman 2001). To reflect the relationship between land use and riverine nutrients, the data in year of 2006, 2011, 2016 and 2019 were selected as the basic data for modelling. The dataset was split randomly into a training dataset and a testing dataset, which accounted for 80% and 20% of the entire dataset, respectively, based on related studies (Wang et al., 2021e).

In this study, the k-fold cross-validation was employed to tune the hyperparameters of the models. The k-fold cross-validation is a resampling method used to evaluate models with a limited data sample, and the k refers to the number of groups split. Using 10-fold cross-validation, the training dataset was split into 10 sub-datasets. The models were trained with training folds and validated with the testing fold (Fig S6-3). The number of trees (n_estimators), the minimum number of samples required to split a node (min_sample_split), the minimum number of samples required to be at a leaf node ( min_samples_leaf), the number of features to consider when looking for the best split (max_features), and the maximum depth of the tree (max_depth) were identified in this study. The models were evaluated by the mean absolute error (MAE), root mean square error (RMSE), percent bias (PBIAS, Gupta et al., 1999), $R^2$ coefficient of determination ($R^2$), and Kling-Gupta efficiency (KGE, Gupta et al., 2009). In this study, R2>0.60, KGE≥0.60, and PBIAS≤±25% were used as the satisfactory thresholds for assessing the model performance based on previous studies (Moriasi et al., 2007, 2015; Odusanya et al., 2019).

To understand the factors that influence water quality in the PRB, the SHAP analysis was employed in this study. As a game-theoretic approach, SHAP can interpret machine learning models by estimating the feature importance (Strumbelj and Kononenko, 2014; Lundberg and Lee, 2017). The contribution of each feature to the model output was allocated based on the marginal contribution (Lundberg and Lee, 2017).

## 6.3 Results

### 6.3.1 Nitrogen and Phosphorus changes across land use types and seasons

The spatial and seasonal variations of water quality among the five types of watersheds were shown in Fig 6-2. The highest concentrations of $NO_3^-$-N and TN were observed in the mixed watersheds, followed by the agricultural and urbanized watersheds. Whereas, the concentrations of $NH_4^+$-N and TP were higher in the urbanized watersheds than that in the agricultural watersheds (Table S6-2). Specifically, the concentrations of $NH_4^+$-N (0.002-1.527 mg/L) and TP (0.003-3.006 mg/L) were relatively lower than those of $NO_3^-$-N (0.001-5.467 mg/L) and TN (0.100-7.164 mg/L). TN and $NO_3^-$-N in the dry season were higher than those in the wet season in all types of watersheds (Fig 6-2). There was no distinguishing difference between the two seasons in the concentrations of $NH_4^+$-N and TP (Fig 6-2).

The inter-annual variation of water quality in the PRB was shown in Fig 6-3. A degeneration trend in water quality was observed during the period of 2006-2009 and which seemed to have improved after 2010s in the PRB. However, concentrations of TN and TP have increased in recent years.

### 6.3.2 Linkage between riverine nutrients and land use

The RFR models were employed to understand the relationship between water quality and land use in the PRB considering land use, intensity of human activities (i.e., type of watershed), climate variability (i.e., water temperature, precipitation), topography, point source, and climate variability with human interference (i.e., time of sampling). The 10-fold cross-validation method was employed to tune the hyperparameters of the models in the PRB (Fig 6-4). All of the models fit well during the two periods, which indicated that the land use, season, topography, point source, and type of watershed could explain a high proportion of variability of the water quality in the PRB.

The performance of the models was listed in Table 6-2. Compared to the performances of TN and $NO_3^-$-N, the performances of $NH_4^+$-N and TP were relatively poor. The KGE of the $NH_4^+$-N during the training and testing period were 0.78 and 0.80, respectively. The calculated KGE for the TP were 0.69 and 0.85, respectively. In contrast, the KGE of TN and $NO_3^-$-N were 0.93-0.98 and 0.95-0.97, respectively, during the two periods. This suggested that other factors may have influenced the concentration of $NH_4^+$-N and TP in the PRB.

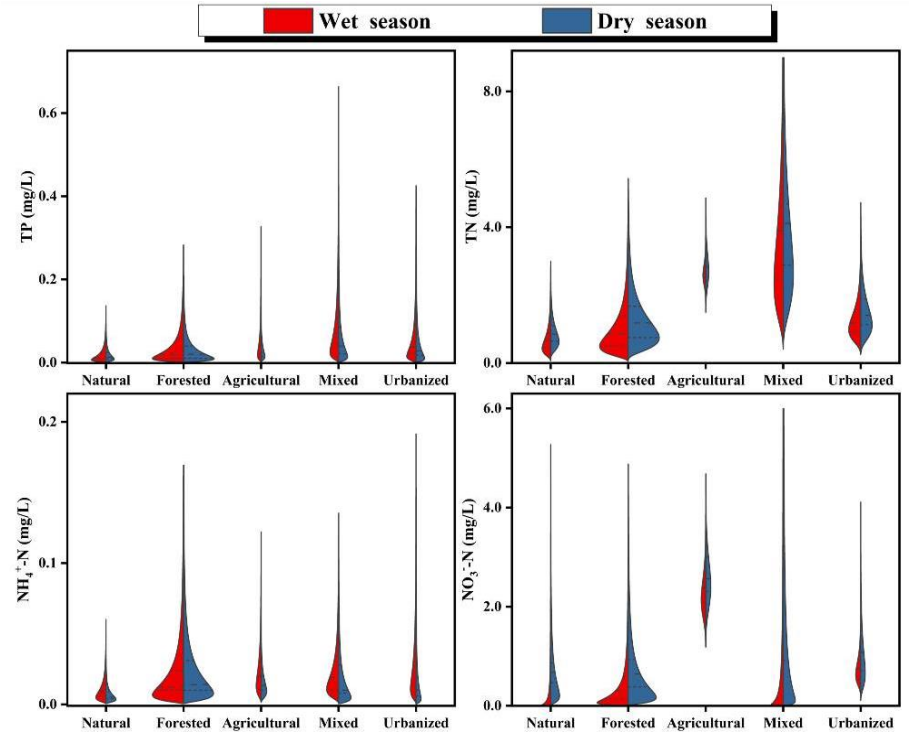


Fig 6-2 Spatial and seasonal variations among five types of watersheds in the PRB

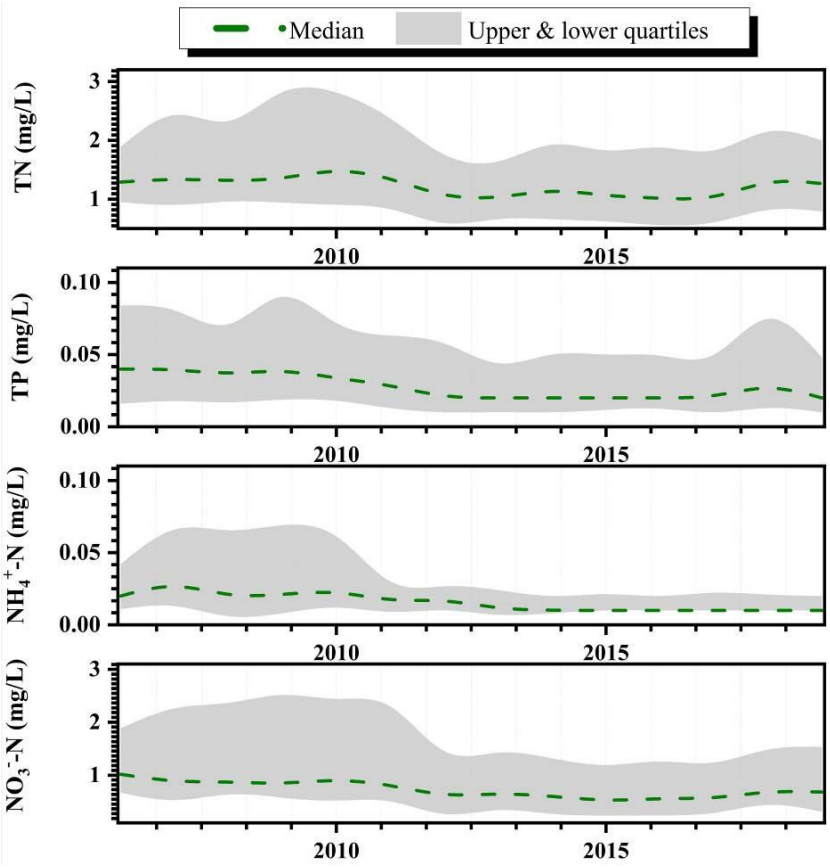


Fig 6-3 Inter-annual variation in water quality during 2006-2019

The features that controlled the water quality in the PRB were identified through the SHAP analysis based on the RFR models (Fig 6-5). A higher SHAP value indicated a stronger driving force. If the SHAP value is positive, this feature is a positive driving force for increasing the concentration of water quality parameters. Point source pollution may play a more important role in the TP model (rank 3/14) than that in the other models (TN model (rank 9/14), $NH_4^+$-N model (rank 12/14), $NO_3^-$-N model (rank 12/14)). Agriculture, forest, type of watershed, and slope were the top factors that controlled the exports of TN and $NO_3^-$-N in the PRB. Consistently, the feature importance for the type of watershed was low in the other two models (i.e., TP and $NH_4^+$-N). Meanwhile, wetland and barren may be one of the most important factors influencing the migration process of TP or $NH_4^+$-N. Increasing SHAP values of TP and $NH_4^+$-N were observed with increasing feature values of wetland or barren. In contrast, increasing wetland may reduce TN. This indicated that human activities were still the major factors controlling the exports of TN and $NO_3^-$-N in the PRB, and the threshold effects of land use patterns on water quality may be found in the PRB. Climate variability may play a more important role in the TP and $NH_4^+$-N models than the TN and $NO_3^-$-N models. Increasing SHAP values of TN and $NO_3^-$-N were observed with increasing feature values of temperature, whereas opposite trends were observed in the models of TP or $NH_4^+$-N.

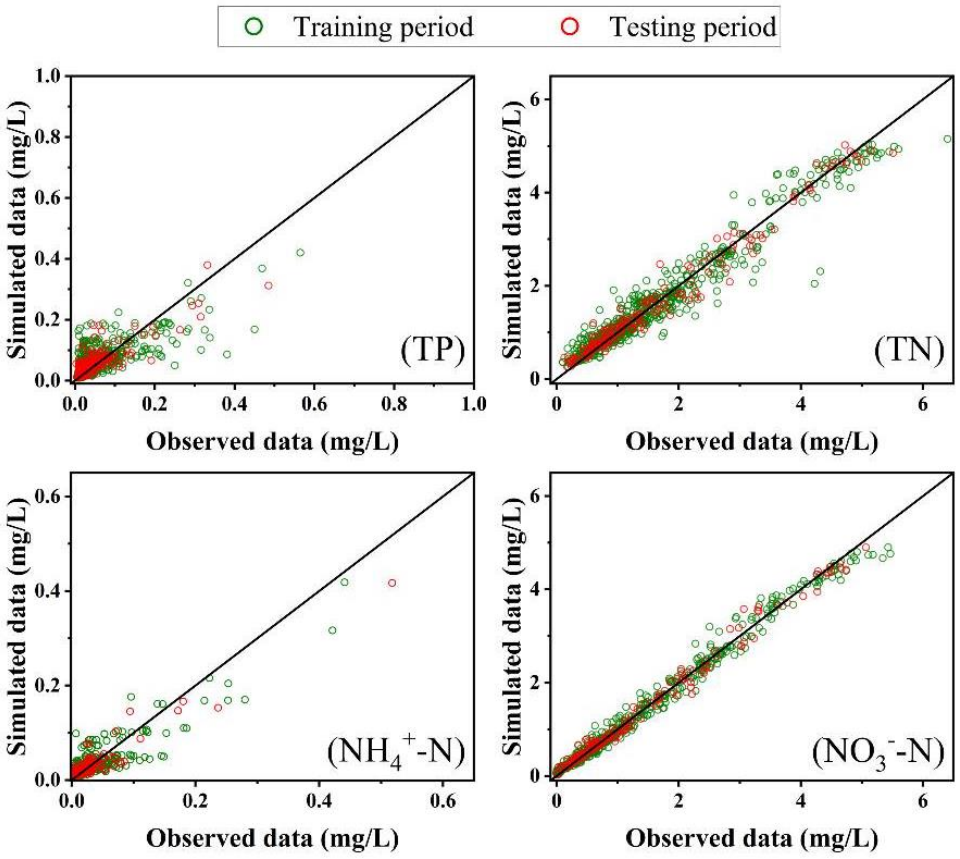


Fig 6-4 The training and testing results of RFR models in the PRB

Table 6-2 Performance of models during the training and testing periods

| | Training period | | | | | Testing period | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MAE (mg/L) | RMSE (mg/L) | $R^2$ | PBIAS (%) | KGE | MAE (mg/L) | RMSE (mg/L) | $R^2$ | PBIAS (%) | KGE |
| TP | 0.034 | 0.068 | 0.94 | 0.64 | 0.69 | 0.032 | 0.095 | 0.82 | -1.87 | 0.85 |
| TN | 0.159 | 0.219 | 0.97 | -0.14 | 0.93 | 0.201 | 0.292 | 0.94 | 0.41 | 0.95 |
| $NH_4^+$-N | 0.011 | 0.019 | 0.79 | 0.61 | 0.78 | 0.010 | 0.017 | 0.91 | -1.67 | 0.80 |
| $NO_3^-$-N | 0.094 | 0.137 | 0.99 | 0.13 | 0.98 | 0.098 | 0.139 | 0.99 | 0.26 | 0.97 |

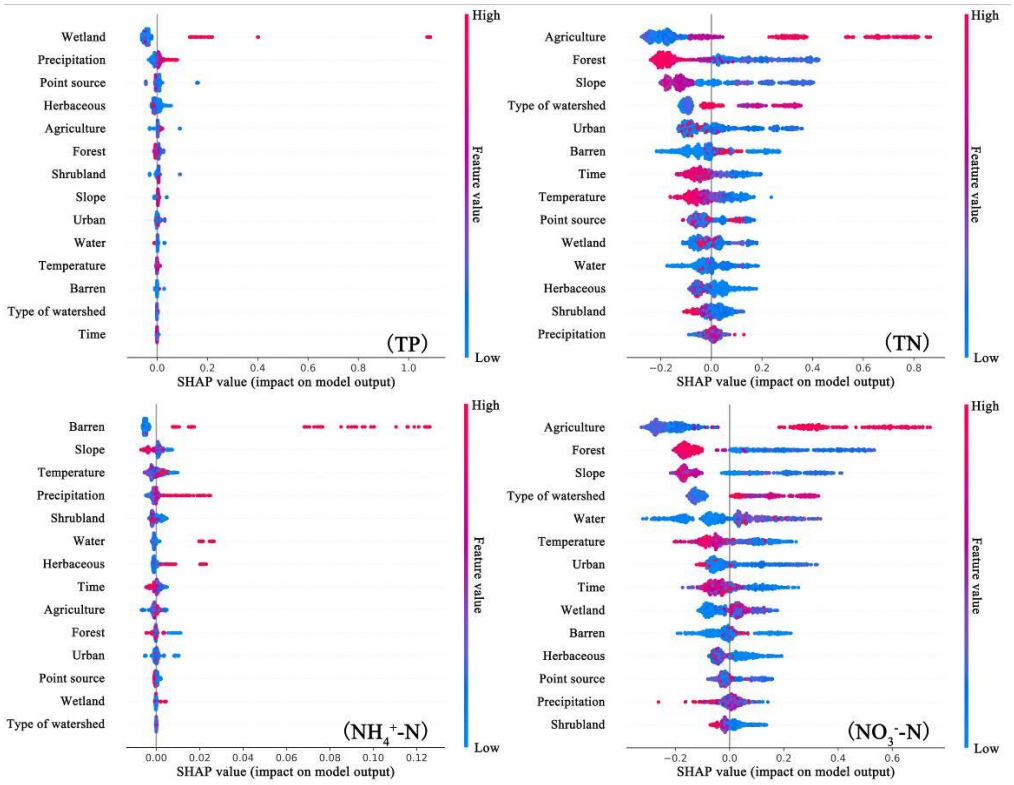


Fig 6-5 SHAP results of feature importance based on the RFR models

(Note: the relative importance of features were ranked from up to down)

.

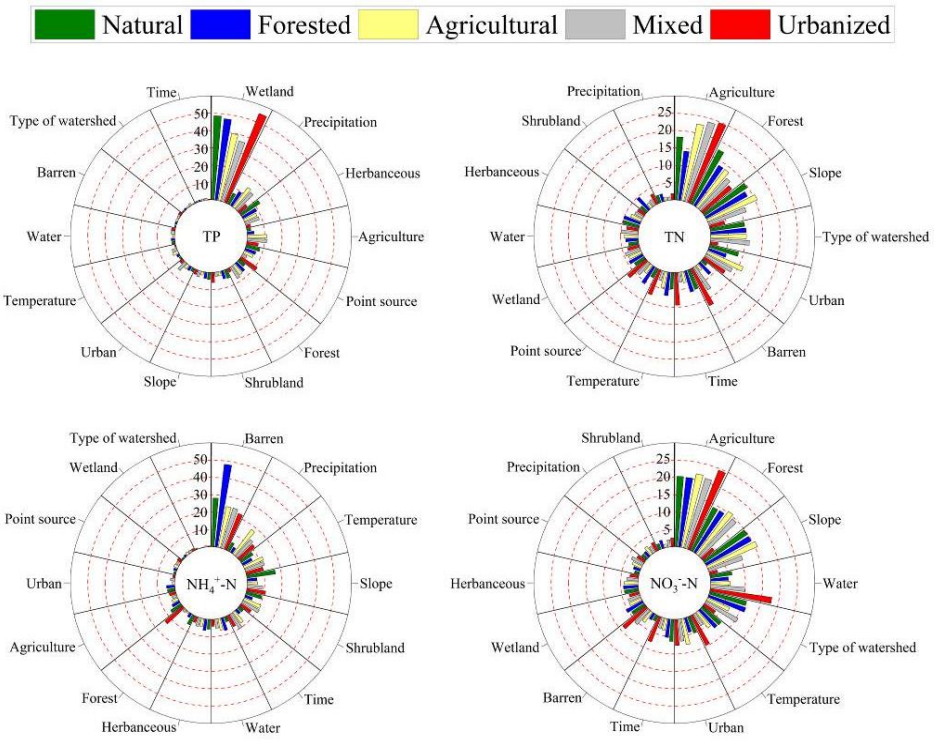


Fig 6-6 The features relative importance ranking in the PRB (%)

The relative contribution of each feature to the models was shown in Fig 6-6. The threshold effect was observed with the modified composition of land use. Although the wetland (36%-55%) was the major factor influencing TP in the PRB, the contribution of this factor changed with increasing human-impact land use. The contribution of wetland may decrease when forest transferred to the agriculture, and a decreasing trend was observed from the natural watershed (49%) to the agricultural watershed (40%). Meanwhile, the observed contribution of wetland to the TP model in PRB decreased from 40% (agricultural watershed, agriculture >50%) to 36% (mixed watershed, agriculture≤50%, urban≤50%; forest≤50%), and increased from 36% (mixed watershed, agriculture≤50%, urban≤50%; forest≤50%) to 55% (urbanized watershed, urban >50%). Similar results were found in other models as well.

### 6.3.3 Water quality in the PRB under different scenarios

Most of the forest, agricultural land, and urbanized land likely gained from shrubland, herbaceous or shrubland, and herbaceous or barren, respectively, in the PRB (Table 6-1). Three scenarios were set up based on the land use in 2019, Scenario I, all of the shrubland transfer to the forest; Scenario II, all of the herbaceous and shrubland transfer to the agricultural land; and Scenario III, all of the herbaceous and barren transfer to the urbanized land. As shown in Fig 6-7, TP and $NH_4^+$-N were more sensitive to land use changes. Undoubtedly, increasing agricultural land (Scenario II) and urbanized land (Scenario III) may increase the concentration of pollutants in the PRB. Although increasing forest may reduce the export of TP, TN, and $NO_3^-$-N in the PRB, shrubland may play a more important role than forest. The observed $NH_4^+$-N increased when the shrubland transfer to the forest (Scenario I).

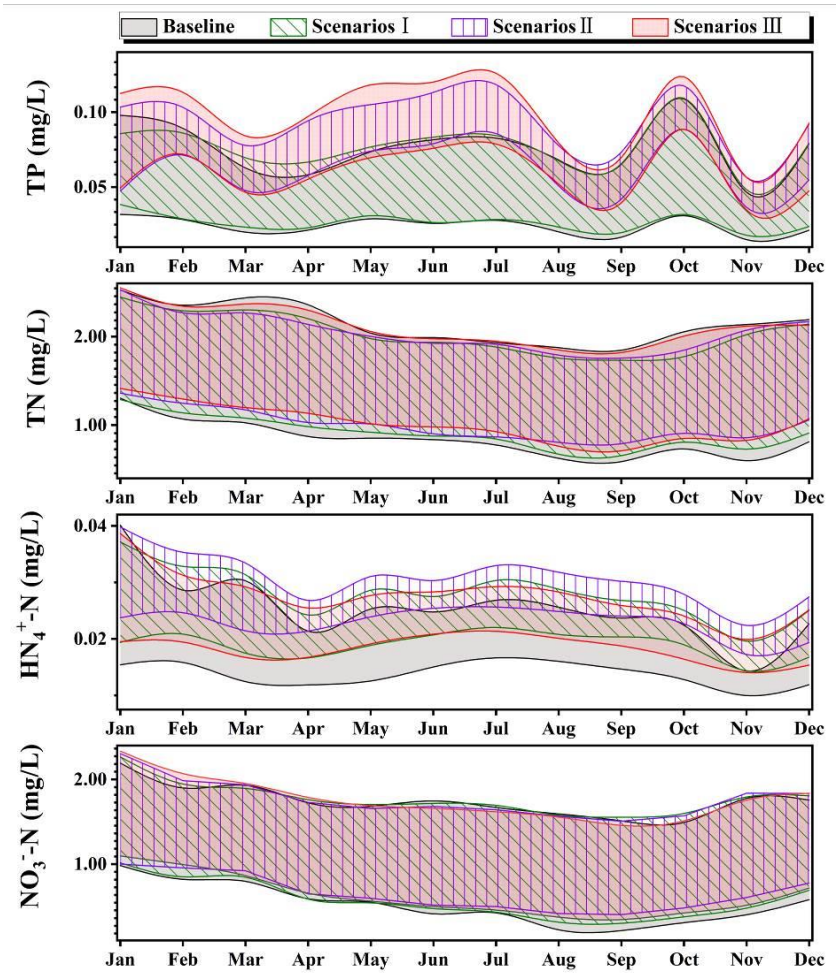


Fig 6-7 Water qualify in the PRB under different scenarios (upper & lower quartiles)

## 6.4 Discussions

### 6.4.1 Training & validation of the models

Understanding the linkages between land use and water quality is commonly recognized as an imperative step to forecast riverine nutrients (Galbraith et al., 2007; Huang et al., 2012; Zhou et al., 2016). Data-driven models are convenient for providing quality assessments and for identifying the relationship between land use and water quality. Our previous studies detected global and local variations in the relationship between land use and water quality using multiple linear regression and geographically weighted regression (Huang et al., 2012; 2015). However, this relationship may weaken or become nonlinearity because of interference from climate variability or human activities (Zhou et al., 2016; Huang et al., 2021; Liu et al., 2021c). To fill this gap, we developed an interpretable machine learning method that integrated the RFR model with the SHAP method to explore the relationship between water quality and land use in the PRB. Compared to the multiple linear regression models used in previous studies (e.g. Huang et al., 2013c), the estimated accuracy of the models in this study was relatively high.

The PBIAS of the models during the training and testing period were -0.14-0.64 and -1.84-0.41, respectively, indicating satisfactory model generalizability.

The relationship between land use and riverine pollutants has been widely explored with multiple linear regression models. And the accuracy of these models may be influenced by the size of the data (Chen et al., 2020; Zhang et al., 2020b; Wang et al., 2021e). With limited data, the RFR models can provide satisfactory performance for scenario analysis and water management. Although the performances of TP and $NH_4^+$-N models were slightly poorer than other models (Table 6-2), there were no significant differences between the observed and simulated baseline scenarios (Fig S6-4). Meanwhile, compared to the performance of other studies (e.g. Wang et al., 2021e), our models were acceptable. We believed that the trained models can be used to assess the water quality in the PRB.

The TN and $NO_3^-$-N models achieved the highest accuracy, followed by the TP and $NH_4^+$-N models (Table 6-2). Consistently, the highest concentrations of water quality indicators were found in TN and $NO_3^-$-N (1-3 mg/L, Fig 6-3), followed by TP and $NH_4^+$-N (0-0.1 mg/L, Fig 6-3). This means that other external factors may have influenced the migration processes of $NH_4^+$-N and TP in the PRB. Theoretically, $NH_4^+$-N is more unstable than other riverine nutrients and can be oxidized into $NO_3^-$-N via biochemical processes. Climate variability may modify this process in the watershed, and in this study we also found that the climate variability (i.e. temperature (rank 3/14) and precipitation (rank 4/14)) was one of the most important factor that influence concentration of $NH_4^+$-N in the PRB (Fig 6-5). In contrast, riverine P has poor mobility and soil erosion may modify the migration processes of the TP. Correspondingly, the precipitation (rank 2/14) and point source (rank 3/14) were the important factors influencing the TP in the PRB. This suggested that extenal factors (e.g. climate variability, point source) may alter the relationship between land use and water quality. A similar result was also found in a previous study in the Minjiang River Watershed (Zhou et al., 2016).

### 6.4.2 Nonlinear linkage between land use patterns with nutrient exports

In this study, RFR models were employed to identify the relationship between land use and water quality in the PRB. The percentages of agriculture, forest, urban, and type of watershed were the top factors that controlled the exports of TN and $NO_3^-$-N in the PRB. Accelerated urbanization is a major contributor to water quality degradation because of increasing point source pollution. (Yu et al., 2013; Huang et al., 2015; Zhang et al., 2020b). However, the high percentage of developed land may not indicate high nutrient export in the PRB. As shown in Fig 6-5, low SHAP values were observed with high feature values of urbanized land. The percentage of urban may not be an important factor controlling $NH_4^+$-N export in the PRB (Fig 6-5). Unlike the developing world, a high rate of domestic wastewater treatment is usually found in urban areas. Most studies have shown that nutrient export is poorly

correlated with specific land uses. (Brett et al., 2005; Lee et al., 2013). In contrast, pollutants from stormwater may be difficult to control in the PRB. For example, the deposited N in the barren may become the major source of $NH_4^+$-N in the PRB when the point source pollution is well-controlled (Fig 6-5).

Rather than urbanization, agricultural activities in the PRB may be the main factor determining nutrient export to the watershed (i.e., TN and $NO_3^-$-N). In this study, we found that agricultural land may be the most significant factor influencing the water quality in the watershed. Similarly, previous studies in Chesapeake Bay suggested that agricultural activities may contribute to more than half of the nutrient exports to that area. (Duan et al., 2012; Mulkey et al., 2017; Jordan et al., 2018; Duan et al. 2021). As there are larger areas of agricultural land use in the PRB, non-point sources may be more difficult to control for each type of watershed. It can be inferred that non-point sources from agricultural activities continue to be the primary source of pollutants in the PRB.

The relationship between land use and water quality has been widely examined in previous studies. After investigating the relationships between water quality and ecological responses in streams with varying laud uses, Amaio et al. (2019) calculated that nonlinearity may be common in watershed systems, occurring in half of their study compared with linearity cases that occurred in less than 14%. The linear relationship between land use and water quality was observed in this study as well. The contributions of forest and slope to riverine nutrients were relatively unchanged with changing type of watershed. For example, the impacts of forest on $NO_3^-$-N were similar for natural (14%), forested (14%), agricultural (15%), and mixed (15%) watersheds (Fig 6-6). However, compared with linear relationships, nonlinear relationships were more common in this study.

The nonlinear relationships between land use patterns and water quality have been clarified in previous studies. For example, climate variability may amply nutrients in urbanized watersheds (Kaushal et al., 2014; Huang et al., 2021). Pollutants may be more sensitive to the percentage of built-up areas in low-urbanized watersheds than highly urbanized watersheds. The threshold effects of land use patterns were examined in this study, and these effects were observed in the TN (rank 4/14) and $NO_3^-$-N (rank 4/14) models, where the initial concentrations of nutrients were high (Figs 6-3, and 6-5). In contrast, the threshold effects of land use patterns were not obvious in the remaining models, including the TP (rank 13/14) and $NH_4^+$-N (rank 14/14) models, where the initial concentrations of nutrients were low (Figs 6-3, and 6-5).

Many studies have indicated that wetlands or forests are the sinks of nutrients, which means that a negative relationship must be established between wetlands and water quality (Galbraith and Burns, 2007; Wojciechowska, 2017). However, in organic-rich wetlands, nutrient release can occur under denitrification or redox conditions. Increasing SHAP values were observed with increasing feature values of wetland (i.e., TP, and $NH_4^+$-N ), suggesting

that wetlands may be a source of nutrient export in the PRB (Fig 6-5)**.** Natural land use, such as forests and wetlands, can be a source of nutrients because of the modified hydrologic regimes and biological processes. Thus, natural land use may become nutrient sources owing to their insufficient capacity for specific nutrients.(Lee et al., 2013; Zhang et al., 2020b).

### 6.4.3 Does seasonality matter in the linkage between land use pattern with water quality?

Various land use patterns have controlled nutrient export as the changing seasons (i.e., dry and wet seasons) in the PRB. Seasonal variations in TN, TP, $NO_3^-$-N, and $NH_4^+$-N across five types of watersheds were displayed in Fig 6-2. Generally, TN and $NO_3^-$-N in the dry season were higher than those in the wet season. Besides, the TP during the wet season was much higher than that during the dry season, especially in natural watersheds. High levels of phosphorus in soils can contribute to elevated levels of TP during high-flow events (Huang et al., 2015).

The seasonality of the water quality may be controlled by human activities and climate variability. For example, fertilizer application in spring and fall may enhance nutrients in watersheds. Climate variability, such as changes in precipitation and temperature, may influence biochemical processes and alter nutrients in the watershed. Therefore, the sampling time (i.e. Time), precipitation, and water temperature (Temperature) were used as the seasonal indicators. Water temperature and precipitation can be used to depict climate variability in the PRB. The effects of human activities and climate variability can be depicted by sampling time. The sampling time may be a less important feature that controls water quality in the PRB (Fig 6-5), and the SHAP value of that for the TP model was close to zero. The export of TN, $NO_3^-$-N, and $NH_4^+$-N may change seasonally. This process may be complex with different land use. Different biochemical processes can be observed in different seasons (Hur et al., 2014; Ahmadisharaf et al., 2020).

Biochemical processes may be influenced by the initial nutrient concentrations. As shown in Fig 6-5, decreasing SHAP values were observed in the TN and $NO_3^-$-N models with increasing feature values of temperature. An opposite relationship was observed in the other two models. Similar results were also found in TP models in the Texas Gulf Region (Wang et al., 2021b). Additionally, with the continued disturbance of human activities, the linkages between land use and water quality would change. For example, precipitation in forested watersheds may dilute nutrients in winter, whereas in agricultural-dominated watersheds, high levels of nutrients due to precipitation and runoff. (Pratt et a al. 2012). The patterns of seasonal nutrients may change with interference from human activities, and the relationship between the feature value of time and SHAP value in the TP model changed more than the temperature.

### 6.4.4 Limitations and outlook

In this study we analyzed the nonlinear relationship between land use and riverine water quality using an interpretable machine learning method. The method proposed in this study can be used to evaluate the impacts of land use policy on riverine nutrient. However, the spatial variability of watershed properties may increase the uncertainty of the results. In the next step, we will try to understand the impacts of the spatial and temporal variability of watershed properties on watershed water quality.

## 6.5 Conclusions

We developed an interpretable machine learning method to explore the threshold effects of land use patterns on water quality and its nonlinearity in the Potomac River Basin. The proposed models exhibited an acceptable performance. The threshold effects of land use patterns were obvious for water quality indicators with high concentrations. In contrast, the water quality at low concentrations may be more sensitive to wetland or barren land with changing water temperatures. The typical “sink” for the nutrient such as forest or wetland may change into the “source” for the nutrient under specific environmental setting. This study provides an in-depth understanding of how riverine nutrient export responds to the land use gradient in Chesapeake Bay watershed.

# Chapter 7 General discussion and conclusion

The main objective of this thesis is to understand how LUCC impacts water resources in coastal watersheds in a changing environment. Based on the remotely sensed data and a machine learning-based CA-Markov model, we gained insights into the major driving factors of LUCC in coastal watersheds. Using machine learning methods and hybrid watershed modelling, we explored how LUCC/urbanization interacts with climate change and its links with water resources from a nonstationary perspective. Furthermore, we investigated the nonlinear relationships between LUCC and water quality in two coastal watersheds with different development stages. This chapter addresses the research questions proposed in the first chapter and summarizes them from a general perspective.

## 7.1 General discussion of research questions

### 7.1.1 What are the major factors influencing the land use/cover change in coastal watersheds?

LUCC in coastal China has drawn much attention for its population pressure and rapid socioeconomic development (Huang et al., 2012; Zhou et al., 2020). Compared with the developed countries such as the USA, the LUCC in the coastal China has been more active in the past decades (Figs. 2-3 and 6-2). The signals of each land use category were analyzed using intensity analysis. The urbanization rate in the Minjiang River Watershed has been increasing recently, and the built-up area has increased significantly in recent years, most of which has been transferred from woodland and agricultural land. We developed a machine learning-based CA-Markov (Cellular automata and Markov Chain) model to gain insights into the driving forces of the LUCCs in the coastal watersheds of China. Population and economic development were the major factors that controlled the distributions of the built-up area in the MRW (Fig. 2-5), which indicated that urban sprawl in the coastal areas may be largely affected by socioeconomic factors. We concluded that urbanization may be more complex with increasing population and GDP. In this study, we also found that the distribution of woodlands was highly related to slope. To achieve the best cost-effectiveness, human beings prefer to develop in a place that is less steep, and the human-impacted land uses are more likely to be found in the flat areas (Lei et al., 2019; Viana et al., 2021; Bacau et al., 2022).

We proposed a machine learning-based CA-Markov model to simulate LUCC under different scenarios (Fig. 2-2). The proposed method can help improve the performance of the CA-Markov model and increase the accuracy of the results (Fig. 2-6 and Table 2-5). Similar to other studies (Pazur and Bolliger, 2017; Bacau et al., 2022), most of the newly developed built-

up areas were transferred from agricultural land and woodland in all scenarios (Figs. 2-7, 2-8, and 2-9). Similar to related studies (Pazur and Bolliger, 2017; Bacau et al., 2022), we also found that land use policy may significantly influence the distribution of land use in the MRW. Given that the economy may continue to develop rapidly, and a higher percentage of the population will reside in these areas, a reasonable spatial plan may play an important role in landscape ecology in the watershed by mitigating the negative effects of land degeneration during urbanization (Figs. 2-8 and 2-9).

### 7.1.2 How does LUCC impact regional water resources in the context of climate variability?

The theoretical basis of water resources management is strongly questioned, because traditionally, the stationarity assumption of hydrology faces a challenge from the changing environment. The impacts of climate variability and human activities on the high flow regime were analyzed using a nonstationary framework based on the hydrology model developed in this study. It is well known that the streamflow regimes are the results of precipitation, temperature, evapotranspiration, and soil moisture. Changes in precipitation may amplify the changes in streamflow that have been observed in the past decades in the Asian Pacific region (Chen et al., 2007; Huang et al., 2014; Zhang et al., 2018). These conclusions were further confirmed at a study site located in southeast China. After comparing the patterns of streamflow extremes at different time scales, we found that the streamflow extremes at shorter time scales may be more sensitive to the climate variability in this region. The maximum flow at different periods can be used to understand the magnitude and duration of extreme annual flow conditions (Richter et al., 1996). The impacts of climate variability and human activity on high flows during different periods were also determined in this study. Similar to other studies (Sun et al., 2018; Zhang et al., 2018), we found that human activities may amplify floods in the watersheds, especially small watersheds. Additionally, compared with high flow at long time scales, the high flows at short time scales may be more sensitive to the human activities in the watersheds.

The effects of climate and human activity may be coupled. To further understand these effects, the relationship between LUCC and regional climate change was analyzed in this study based on a grid-to-grid framework. This study further confirms the results of previous studies (Lu et al., 2019; Song et al., 2021; Pimonsree et al., 2022), which concluded that patterns of regional climate extremes follow a nonstationary distribution as the results of urbanization, and the different patterns of climate extremes were also observed in different periods (Fig. 4-6). Additionally, the patterns of the climate extremes were analyzed under different patterns of land use. Low intensity human activities may not make a big difference in regional climate variability, and the patterns of climate extremes in the suburbanized areas were similar to those

in the natural areas (Table 4-3 and Fig. 4-10). In contrast, the patterns of precipitation and temperature may be modified in the highly urbanized areas.

In summary, streamflow dynamics are consistent with human activity and climate variability. Human activities with low intensity contribute little to local climate variability, whereas high-intensity human activities may modify the local climate, and these coupled effects may enhance the modification of regional water resources (Fenner et al., 2019; Govind et al., 2019; Li et al., 2020a).

### 7.1.3 What is the nonlinear relationship between land use and water quality in coastal watersheds?

Although a nonlinear relationship between land use and water quality has been identified in previous studies (Mulkey et al., 2017; Amaio et al., 2019; Liu et al. 2019c), this relationship is still difficult to understand using conventional statistical methods. To further understand this relationship in the coastal watersheds at different stages of development, we proposed an interpretable machine learning method and applied it to typical coastal watersheds in China and the USA (Chapters 5 and Chapter 6). The estimated accuracy of this method was relatively high in the two study areas, indicating a satisfactory generalized model.

As the largest developing country in the world, the watersheds in China still suffer from high pollution levels. Pollutants in the watershed were more sensitive to climate variability and hydrological processes (Fig. 5-7). Additionally, human activities may increase its complexity, and the reservoirs built in the watershed may amplify the effects of climate variability on water quality. Urban point source pollution is still an important factor influencing the water quality in coastal China, although the point source pollution per inhabitant may be reduced with increasing urbanization. In contrast, urbanization and point source pollution in the USA did not significantly contribute to the water quality in the coastal watershed (Fig. 6-6). The high percentage of developed land do not indicate high nutrient export in the Potomac River watershed. Rather than urbanization, agricultural activities in the USA may be the main factor determining nutrient exports to the watersheds. Climate variability is still an important factor that influences the relationship between land use and water quality in the Potomac River Basin by modifying the migration and biochemical processes of nutrients （Fig. 6-7）.

## 7.2 General conclusion

Water resource conditions are highly influenced by human activities. As one of the most important indicators that reflects the intensity of human activities, LUCC has drawn much attention in recent decades. Thus, it is necessary to understand the LUCC patterns in watersheds and identify their impacts on the local water resources. We also analyzed the impacts of the human activities on the streamflow regime as well as the regional climate changes. Furthermore,

the nonlinear relationship between land use and water quality was identified in this study. The major findings of this study are as follows:

1. Spatial variation in land use was highly related to the driving factors, and population and local economic development may be the major factors influencing urbanization processes in the coastal watersheds. Land use policies may play an important role in land use evolution.
2. Streamflow extremes are highly impacted by the human activities and climate variability, and the human activities may be the major factor controlling streamflow extremes at short time scales.
3. The coupled effects of climate variability and human activities were identified by analyzing the relationship between urbanization and climate patterns in the studied watersheds, and the patterns of precipitation and temperature may be modified in highly urbanized areas.
4. A nonlinear relationship between land use and water quality has been widely observed, especially in highly polluted watersheds.

## 7.3 Outlook

Understanding the catchment hydrology, nutrient exports, and their future changes is a prerequisite for sustainable water management. These components are strongly influenced by the land use dynamics in watersheds. In this study, we analyzed the patterns of LUCC and identified their impacts on the conditions of water resources; however, there are still limitations that require further investigation. For example, the transitions of land categories were estimated based on past LUCCs, and the nonlinear processes of LUCC were not fully considered in these models. This study focused on distinguishing the individual effects of human activities and climate variability, but the coupled effects of related factors remain unclear. To reduce the complexity, a rainfall-runoff model was proposed, and the performance of the model was lower in the small watersheds than in large watersheds. Thus, more efforts should be made to improve the model to understand the complex processes in the watersheds. In this study, we proposed a method that can shed light on how to improve water quality in watersheds from the perspective of LUCCs; however, the effects of different landscapes were not fully discussed in this study. Therefore, more data must be collected to improve the models and methods used in this study. Future research could focus on modelling catchment processes, including soil processes, river processes and other key processes in catchment systems. In addition, long time series sampling data with high spatial resolution are needed for further analysis.

**Reference**

Aburas MM, Ho YM, Ramli MF, Ash'aari ZH. Improving the capability of an integrated CA-Markov model to simulate spatio-temporal urban growth trends using an analytical hierarchy process and frequency ratio. International Journal of Applied Earth Observation & Geoinformation. 2017, 59:65-78

Aburas MM, Ahamad MSS, Omar NQ. Spatio-temporal simulation and prediction of land-use change using conventional and machine learning models: a review. Environmental Monitoring & Assessment. 2019, 191: 205

Adedeji IC, Ahmadisharaf E, Sun Y. Predicting in-stream water quality constituents at the watershed scale using machine learning. Journal of Contaminant Hydrology. 2022, 251: 104078

Aghsaei H, Mobarghaee Dinan N, Moridi A, Asadolahi Z, Delavar M, Fohrer N, Wagner PD. Effects of dynamic land use/land cover change on water resources and sediment yield in the Anzali wetland catchment, Gilan, Iran. Science of the Total Environment, 2020, 712: 136449.

Aguilar C, Polo MJ. Generating reference evapotranspiration surfaces from the Hargreaves equation at watershed scales. Hydrology and Earth System Sciences. 2011, 15:2495-2508

Ahmed MH, Lin L. Dissolved oxygen concentration predictions for running waters with different land use land cover using a quantile regression forest machine learning technique. Journal of Hydrology. 2021, 597: 126213

Ahearn DS, Sheibley RW, Dahlgren RA, Anderson M, Johnson J, Tate KW. Land use and land cover influence on water quality in the last free-flowing river draining the western Sierra Nevada, California. Journal of Hydrology.2005, 313:234-247

Ahmadisharaf E, Lacher IL, Fergus C, Benham BL, Akre T, Kline KS. Projecting land use change impacts on nutrients, sediment and runoff in multiple spatial scales: Business-as-usual vs. Stakeholder-informed scenarios. Journal of Cleaner Production. 2020, 257: 120466

Akasaka M, Takamure N, Mitsuhashi H, Kadono Y. Effects of land use on aquatic macrophyte diversity and water quality of ponds. Freshwater Biology. 2010, 55: 909-922

Akinemi FO, Pontius RG, Bralmoh AK. Land change dynamics: insights from intensity analysis applied to an African emerging city. Journal of Spatial Science. 2017,62: 69-83

Alan DZ, Justin S, Edwin PM, Bart N, Eric FW, Dennis PL. Detection of intensification in global- and continental-scale hydrological cycles: temporal scale of evaluation. Journal of climate. 2003, 16:535-547

Amario SC, Rearick DC, Fraching C, Kembel SW, Porter-Goff E, Spooner DE, Williams CJ, Wilson HF, Xenopoulos MA. The prevalence of nonlinearity and detection of ecological breakpoints across a land use gradient in streams. Scientific Reports. 2019, 9:3878

Amato F, Tonini M, Murgante B, Kanevski M. Fuzzy definition of rural urban interface: an application based on land use change scenarios in Portugal. Environmental Modelling & Software. 2018, 104:171-187

Arheimer B, Liden R. Nitrogne and phosphorus concentrations from agricultural catchments---influence of spatial and temporal variables. Journal of Hydrology. 2000, 227:140-159

Arsanjani JJ, Helbich M, Kainz W, Boloorani AD. Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. International Journal of Applied Earth Observation & Geoinformation. 2013, 21: 265-275

Ator SW, Blomquist JD, Webber JS, Chanat JG. Factors driving nutrient trends in streams of the Chesapeake Bay watershed. Journal of Environmental Quality. 2020, 49:812-834

Ayalew AD, Wagner PD, Sahlu D, Fohrer N. Land use change and climate dynamics in the Rift Valley Lake Basin, Ethiopia. Environmental Monitoring and Assessment 2022, 194: 791

Bacau S, Domingo D, Palka G, Pellissier L, Kienast F. Integrating strategic planning intentions into land-change simulations: designing and assessing scenarios for Bucharest. Sustainable Cities & Society. 2022, 76: 103116

Baker A. Land use and water quality. Hydrological Processes. 2003. 17:2499-2501

Bai X, SHi P, Liu Y. Realizing China's urban dream. Nature. 2014, 509:158-160

Batty M, Couclelis H, Eichen M. Urban systems as cellular automata. Environment & Planning B: Planning & design. 1997, 24:159-164

Bennett KE, Cannon AJ, Hinzman L. Historical trends and extremes in boreal Alaska river basin. Journal of Hydrology. 2015,527:590-607

Belval DL, Sprague LA. Monitoring nutrients in the major rivers draining to Chesapeake Bay. Water. Resour. Invest. Rep. 1999: 99-4238

Brabec E, Schulte S, Richards PL. Impervious surfaces and water quality: a review of current literature and its implications for watershed planning. Journal of Planning Literature. 2002, 16(4): 499-514

Bracken C, Holman KD, Rajagopalan B, Moradkhani H. A Bayesian hierarchical approach to multivariate nonstationary hydrologic frequency analysis. Water Resources Research. 2018,54: 243-255

Breiman L, Random forests. Machine Learning. 2001, 45:5-32

Brett MT, Arhonditsi GB, Mueller SE, Hartley DM, Frodge JD, Funke DE. Non-point-source impacts on stream nutrient concentration along a forest to urban gradient. Environmental Management, 2005, 35(3): 330-342

Bricker SB, Rice KC, Bricker OP. From headwaters to coast: influence of human activities on water quality of the Potomac River Estuary. Aquatic Geochemistry. 2014, 20: 291-323

Brizzetti B, Vigiak O, Udias A, Aloe A, Zanni M, Bouraoui F, Pistocchi A, Dorati C, Friedland R, De Roo A, Benitez Sanz C, Leip A, Bielza M. How EU polices could reduce nutrient pollution in European inland and coastal waters. Global Environmental Change. 2021, 69: 102281

Blomquist JD, Fisher GT, Denis JM, Brakebill JW, Werkheiser WH. Water-quality assessment of the Potomac River Basin; basin description and analysis of available nutrient data, 1970-1990. US Geological Survey Water-Resources Investigations Report 95-4221, 1996

Carney M, Azencott R, Nicol M. Nonstationarity of summer temperature extremes in Texas. International Journal of Climatology. 2020, 40: 620-640

Castella J, Verburg PH. Combination of process-oriented and pattern-oriented models of land-use change in a mountain area of Vietnam. Ecological Modelling. 2007, 202:410-420

Ceres RL, Forest CE, Keller K. Understanding the detectability of potential changes to the 100-year peak storm surge. Climatic Change. 2017,145: 221-235

Chen Y, Li W, Xu C, Hao X. Effects of climate change on water resources in Tarim River Basin, Northwest China. Journal of Environmental Sciences. 2007, 19:221-235

Chen Y, Chen X, Ren G. Variation of extreme precipitation over large river basins in China. Advances in Climate Change Research. 2011, 6:108-114

Chen PC, Wang YH, You GJY, Wei CC. Comparison of methods for non-stationary hydrologic frequency analysis: Case study using annual maximum daily precipitation in Taiwan. Journal of Hydrology. 2017, 545:197-211

Chen K, Chen H, Zhou C, Huang Y, Qi X, Shen R, Liu F, Zuo M, Zou X, Wang Y, Chen D, Chen X, Deng Y, Ren H. Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. Water Research. 2020, 171(15): 115454

Chen Y, Lara MJ, Jones BM, Frost GV, Hu F. Thermokarst acceleration in Arctic tundra driven by climate change and fire disturbance. One Earth. 2021, 4: 1718-1729

Chen Y, Kirwan ML. A phenology- and trend- based approach for accurate mapping of sea-level driven coastal forest retreat. Remote Sensing of Environment. 2022, 281:113229

Cheng L, AghaKouchak A, Gilleland E, Katz RW. Non-stationary extreme value analysis in a changing climate.Climatic Change, 2014,127(2):353-369

Claggett PR, Irani FM, Thompson RL. Estimating the extent of impervious surfaces and turf grass across large regions. Journal of the American Water Resources. 2013, 49(5): 1057-1077

Collins SM, Yuan S, Tan PN, Oliver SK, Lapierre JF, Cheruvelil KS. Fergus CE, Skaff NK, Stachelek J, Wagner T, Soranno PA. Winter precipitation and summer temperature predict lake water quality at macroscales. Water Resources Research. 2019, 55: 2708-2721

Cui Z, Huang JC, Tian F, Gao J, Wang X, Li J. Quantifying the impacts of climate change and land use on hydrological processes: A comparison between mountain and lowland agriculture watersheds. Hyrological Processes, 2020, DOI:10.1002/hyp.13950

Cuo L, Zhang Y, Gao Y, Hao Z, Cairang L. The impacts of climate change and land cover/use transition on the hydrology in the upper Yellow River Basin, China. Journal of Hydrology. 2013,502:37-52

da Cunha ER, Santos CAG, da Silva RM, Bacani M, Pott A. Future scenarios based on a CA-Markov land use and land cover simulation model for a tropical humid basin in the Cerrado/Atlantic forest ecotone of Brazil. Land Use Policy. 2021. 101:105141

Deb P, Kiem AS, Willgoose G. A linked surface water-groundwater modelling approach to more realistically simulate rainfall runoff non-stationarity in semi-arid regions. Journal of Hydrology. 2019, 575: 273-291

Donat MG, Lowry AL, Alexander LV, O'Gorman PA, Maher N. More extreme precipitation in the world's dry and wet regions. Nature Climate Change. 2016, 6:508-514

Drewry JJ, Newham LTH, Greene RSB, Jakeman AJ, Croke BFW. A review of nitrogen and phosphorus export to waterways: context for catchment modelling. Marine & Freshwater Research. 2006, 57: 757-774

Duan S, Kaushal SS, Groffman PM, Band LE, Belt KT. Phosphorus export across an urban to rural gradient in the Chesapeake Bay watershed. Journal of Geophysical Research. 2012, 117: G01025

Duan S, Kaushal SS, ROsenfeldt EJ, Huang J, Murthy S. Changes in concentrations and source of nitrogen along the Potomac River with watershed land use. Applied Geochemistry. 2021, 131:105006

Ervinia A, Huang J, Zhang Z. Nitrogen sources, processes and associated impacts of climate and land-use changes: insights from the INCA-N model. Marine Pollution Bulletin. 2020, 159:111502

Elvidge CD, Baugh K, Zhizhin M, Hsu FC, Ghosh T. VIIRS night-time lights. International Journal of Remote Sensing. 2017, 38: 5860-5879

Elvidge CD, Zhizhin M, Ghosh T, Hsu FC. Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019. Remote Sensing. 2021,13(5): 922

Fang L, Huang JL, Cai JT, Nitivattananon V.Hybrid approach for flood susceptibility assessment in a flood-prone mountainous catchment. Journal of Hydrology. 2022, 612: 128091

Feng Y, Tong X. Using exploratory regression to identify optimal driving factors for cellular automaton modelling of land use change. Environmental Monitoring & Assessment. 2017, 189:515

Feng Y, Lei Z, Tong X, Gao C, Chen S, Wang J, Wang S. Spatially-explicit modeling and intensity analysis of China's land use change 2000-2050. Journal of Environmental Management. 2020,263:110407

Fenner D, Holtmann A, Meier F, Langer I, Scherer D. Constrsting changes of urban heat island intensity during hot weather episodes. Environmental Research Letters. 2019,14(12):124013

Fenicai F, Kavetski D, Savenije H. Elements of a flexible approach for conceptual hydrological modeling: 1. motivation and theoretical development. Water Resources Research. 2011, 47: W11510

Fohrer N, Chicharo. Interaction of river basins and coastal waters-an integrated ecohydrological view. Treatise on Estuarine and Coastal Science, McLusky D and Wolanski E, Eds. 2011: 109-150

Foley JA, DeFries R, Asner GP, Barford C, Bonan G, Carpenter SR, Chapin FS, Coe MT, Daily GC, Gibbs HK, Helkowski JH, Holloway T. Howard EA, Kucharik CJ, Monfreda C, Patz JA, Prentice IC, Ramankutty N, Snyder PK. Global consequence of land use. Science. 2005, 309:570-574

Fu X, Wang X, Yang Y. Deriving suitability factors for CA-Markov land use simulation model based on local historical data. Journal of Environmental Management. 2018, 206: 10-1

Fuka DR, Walter MT, MacAlister C, Degaetano AT, Steenhuis TS, Easton ZM. Using the climate forecast system reanalysis as weather input data for watershed models. Hydrological Processes, 2014, 28:5613-5623

Galbraith LM, Burns GW. Linking land-use, water body type and water quality in southern New Zealand. Landscape Ecology. 2007, 22:231-241

Gallo EL, Meixner T, Aoubid H, Lohse KA, Brooks PD. Combined impact of catchment size, land cover, and precipitation on streamflow and total dissolved nitrogen: A global comparative analysis. Global Biogeochemical Cycles. 2015, 29(7):1109-1121

Govind NR, Ramesh H. The impact of spatiotemporal patterns of land use land cover and land surface temperature on an urban cool island: a case study of Bengaluru. Environmental Monitoring & Assessment. 2019, 191:283

Greaver TL, Clark CM, Compton JE, Vallano D, Talhelm AF, Weaver CP, Band LE, Barson JS, Davidson EA, Tague CI, Felker-Quinn E, Lynch JA, Herrick JD, Liu I, Goodale CI, Novak KJ, Haeuber RA. Key ecoligcal responses to nitrogen are altered by climate change. Nature Climate Change. 2016, 6:836-843

Grimm NB, Faeth SH, Golubiewski NE, Redman CL, Wu J, Bai X, Briggs JM. Global change and the ecology of cities. Science. 2008:756-760

Gu X, Zhang Q, Li J, Singh VP, Sun P. Impact of urbanization on nonstationarity of annual and seasonal precipitation extremes in China. Journal of Hydrology. 2019, 575:638-655

Guan D, Zhao Z, Tan J. Dynamic simulation of land use change based on logistic-CA-Markov and WLC-CA-Markov models: a case study in three gorges reservoir area of Chongqing, China. Environmental Science & Pollution Research. 2019, 26:20669-20688

Guardian MG, He P, Bermudez A, Duan S, Kaushal SS, Rosenfeldt E, Aga DS. Optimized suspect screening approach for comprehensive assessment of the impact of best management practices in reducing micropollutants transport in the Potomac River watershed. Water Research X. 2021, 11:100088

Gupta HV, Sorooshian S, Yapo PO. Status of automatic calibration for hydrologic models: comparison with multilevel expert calibration. Journal of Hydrologic Engineering. 1999, 4(2):135-143

Gupta HV, Kling H, Yilmaz KK, Martinez GF. Decomposition of the mean squared error and NSE performance criteria: Implications of improving hydrological modelling. Journal of Hydrology. 2009, 377: 80-91

Haidary A, Amiri BJ, Adamowski J, Fohrer N, Nakane K. Assessing the impacts of four land use types on the water quality of wetlands in Japan. Water Resources Management. 2013, 27: 2217-2229

Harris I, Osborn TJ, Jones P, Lister D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. Scientific Data. 2020, 7:109

Hargreaves GL, Hargreaves GH, Riley JP. Irrigation water requirements for Senegal river basin. Journal of Irrigation & Drainage Engineering. 1985, 111:265-275

Hargreaves GH, Allen RG. History and evaluation of hargreaves evapotranspiration equation. Journal of Irrigation & Drainage Engineering. 2003, 129(1):53-63

He L, He Z. Water quality prediction of marine recreational beaches receiving watershed baseflow and stormwater runoff in southern California, USA. Water Research. 2008,42:2563-2573

Hernanadez Codero AL, Tango PJ, Batiuk RA. Development of a multimetric water quality indicator for tracking progress towards the achievement of Chesapeake Bay water quality standards. Environmental Monitoring & Assessment. 2020, 192: 94

Hersperger AM, Olivira E, Pagliarin S, Palka G, Verburg P, Bolliger J, Gradinaru S. Urban land-use change: the role of strategic spatial planning. Global Environment Change. 2018, 51:32-42

Hesarkazzazi S, Arabazadeh R, Hajibabaei M, Rauch W, Kjeldsen TR, Prosdocimi I, Castellarin A, Sitzenfrei R. Stationary vs non-stationary modelling of flood frequency distribution across northwest England. Hydrological Sciences Journal. 2021, 66(4):729-744

Hu M, Sayama T, Duan W, Takara K, He B, Luo P. Assessment of hydrological extremes in the Kamo River Basin, Japan. Hydrological Science Journal. 2017, 62:1255-1265

Huang J, Lin J, Tu Z. Detecting spatiotemporal change of land use and landscape pattern in a coastal gulf region, southeast of China. Environment, Development & Sustainability. 2010, 12:35-48

Huang J, Klemas V. Using remote sensing of land cover change in coastal watershed to predict downstream water quality. Journal of Coastal Research. 2012, 28: 930-944

Huang J, Pontius RG, Li Q, Zhang Y. Use of intensity analysis to link patterns with processes of land change from 1986 to 2007 in a coastal watershed of southeast China. Applied Geography. 2012, 34:371-384

Huang J, Zhang Z, Feng Y, Hong H. Hydrologic Response to Climate change and human activities in a Subtropical Coastal Watershed of Southeast China. Regional Environment Change, 2013a, 13:1195-1210

Huang J, Zhou P, Zhou Z, Huang Y. Assessing the influence of land use and land cover datasets with different points in time and levels of detail on watershed modeling in the North River Watershed, China. International Journal of Environmental Research & Public Health. 2013b, 10(1):144-157

Huang J, Li Q, Pontius RG, Klemas V, Hong H. Detecting the dynamic linkage between landscape characteristics and water quality in a subtropical coastal watershed, southeast China. Environmental Management.2013c, 51:32-44

Huang JC, Lee TY, Lee JY. Observed magnified runoff response to rainfall intensification under global warming. Environmental Research Letters, 2014,9(3): 034008

Huang J, Huang Y, Pontius RG, Zhang Z. Geographically weighted regression to measure spatial variations in correlations between water pollution versus land use in a coastal watershed. Ocean & Coastal Management. 2015, 103:14-24

Huang W, Hall S. Elevated moisture simulates carbon loss from mineral soils by releasing protected organic matter. Nature Communications. 2017, 8:1774

Huang B, Huang J, Pontius RG, Tu Z. Comparison of intensity analysis and land use dynamic degree to measure land changes outside versus inside the coastal zone of Longhai, China. Ecological Indicators. 2018, 89:336-347

Huang Y, Tian M, Jin F, Chen M, Liu Z, He S, Li F, Yang L, Fang C, Mu J. Coupled effects of urbanization level and dam on microplastics in surface waters in a watershed of Southeast China. Marine Pollution Bulletin, 2020, 154:11089

Huang Y, Huang J, Ervinia A, Duan S, Kaushal SS. Land use and climate variability amplifies watershed nitrogen exports in coastal China. Ocean & Coastal Management. 2021, 207: 104428

Huo R, Li L, Chen H, Xu C, Chen J, Guo S. Extreme precipitation changes in Europe from the last millennium to the end of the twenty-first century. Journal of Climate. 2021, 34(2):567-588

Hur J, Nguyen HM, Lee BM. Influence of upstream land use on dissolved organic matter and trihalomethane formation potential in watersheds for two different seasons. Environmental Science and Pollution Research, 2014, 21:7489-7500.

Hörmann G, Zhang X, Fohrer N. Comparison of a simple and a spatially distributed hydrologic model for the simulation of a lowland catchment in Northern Germany. Ecological Modelling. 2007, 209: 21-28

Howarth RW, Sharpley A, Walker D. Sources of nutrient pollution to coastal waters in the United States: implications for achieving coastal water quality goals. Estuaries. 2002,

Ishak EH, Rahman A, Westra S, Sharma A, Kuczera G. Evaluating the non-stationarity of Australian annual maximum flood. Journal of Hydrology.2013,494:134-145

Jachniak E, Jagus A, Mlyniuk A, Nycz B. The quality problems of the dammed water in the mountain forest catchment. Journal of Ecological Engineering. 2019, 20:165-171

Jackson-Blake LA, Sample JE, Wade Aj, Hellwell RC, Skeffington RA. Are our dynamic water quality models too complex? A comparison of a new parsimonious phosphorus model, SimplyP, and INCA-P. Water Resources Research. 2017, 53: 5382

Jordan TE, Weller DE, Pelc Carey E. Effects of local watershed land use on water quality in Mid-Atlantic Coastal Bays and Subestuaries of the Chesapeake Bay. Estuaries & Coasts. 2018, 41: S38-S53

Julian JP, de Beurs KM, Owsley B, Davies-Colley RJ, Ausseil AE. River water quality changes in new Zealand over 26 years: response to land use intensity. Hydrology & Earth System Sciences. 2017, 21:1149-1171

Jung CG, Lee DR, Moon JW.Comparison of the Penman-Monteith method and regional calibration of the Hargreaves equation for actual evapotranspiration using SWAT-simulated results in the Seolma-cheon basin. Hydrology Science Journal. 2016, 61:793-800

Kändler M, Blechinger K, Seidler C, Pavlu V, Sanda M, Dostal T, Krase J, Vitvar T, Stich M. Impact of land use on water quality in the upper Nisa catchment in the Czech Republic and in Germany. Science of the Total Environment. 2017, 586:1316-1325

Karimi F, Sultana S, Babakan AS, Suthaharan S. An enhanced support vector machine model for urban expansion prediction. Computers, Environment & Urban Systems. 2019 (75):61-75

Kaushal SS, Groffman PM, Band LE, Shields CA, Morgan RP, Paimer MA, Belt KT, Swan CM, Findlay SEG, Fisher GT. Interaction between urbanization and climate variability amplifies watershed nitrate export in Maryland. Environmental Science & Technology. 2008, 42:5872-5878

Kaushal SS, Groffman PM, Band LE, Elliott EM, Shield CA, Kendall CK. Tracking nonpoint source nitrogen pollution in human-impacted watershed. Environmental Science & Technology. 2011, 45: 8225-8232

Kaushal SS, Mayer PM, Vidon PG, Smith RM, Penniono MJ, Newcomer TA, Duan SW, Welty C, Belt KT, Land use and climate variability amplify carbon, nutrient, and contaminant pulses: a review with management implications. Journal of the American Water Resources Association. 2014, 50:585-614

Kendall MG. Rank correlation methods. Charles Griffin, London, UK. 1975

Kim NW, Chung M, Won YS, Arnold JG. Development and application of the integrated SWAT-MODFLOW model. Journal of Hydrology. 2008, 356: 1-16

Kim H, Ki S, SHin H, Heo JH. Appropriate model selection methods for nonstationary gerneralized extreme value models. Journal of Hydrology. 2017, 547:557-574

Kiesel J, Stanzel P, Kling H, Fohrer N, Jahnig SC, Pechivanidis I. Sreamflow-based evaluation of climate model sub-selection methods. Climatic Change, 2020, 163: 1267-1285

Knoben WJM, Freer JE, Woods RA. Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. Hydrology & Earth System Sciences. 2019,23: 4323-4331

Kumar SV, Jasinki M, Mocko DM, Rodell M, Borak J, Li B, Beaudoing HK, Peters-Lidard CD. NCA-LDAS land analysis: development and performance of a mutisensor, multivariate land data assimilation system for the national climate assessment. Journal of Hydrometeorology. 2019,20(8):1571-1593

Lacher IL, Ahmadisharaf E, Fergus C, Akre T, Mcshea WJ, Benham BL, Kline KS. Scale-dependent impacts of urban and agricultural land use on nutrient, sediment, and runoff. Science of the Total Environment. 2019, 652: 611-622

Lambin EF, Meyfroidt P. Global land use change, economic globalization, and the looming land scarctiy. Proceedings of the National Academy of Sciences. 2011,108:3465-3472

Larned ST, Scarsbrook MR, Snelder TH, Norton NJ, Biggs BJF.Water quality in low-elevation streams and rivers of New Zealand: recent state and trends in contrasting land-cover classes. New Zealand Journal of Marine and Freshwater Research. 2004, 38: 347-366

Lau KH, Kam BH. A celluar automata model for urban land-use simulation. Environment & Planning B: Planning & Design. 2005, 32:247-263

Lu M, Xu Y, Shan N, Wang Q, Yuan J, Wang J. Effect of urbanisation on extreme precipitation based on nonstationary models in the Yangtze River Delta metropolitan region. Science of the Total Environment. 2019, 673: 64-73

Lee SW, Heang SJ, Lee SB, Hwang HS, Sung HC. Landscape ecological approach to the relationships of land use patterns in watersheds to water quality characteristics. Landscape & Urban Planning. 2009, 92:80-89

Lee TY, Huang JC, Kao SJ, Tung CP. Temporal variation of nitrate and phosphate transport in headwater catchments: the hydrological controls and land use alteration. Biogeoscience. 2013, 10:2617-2632

Lee TY, Shih YT, Huang JC, Kao SJ, Shiah FK, Liu KK. Speciation and dynamics of dissolved inorganic nitrogen export in the Danshui River, Taiwan. Biogeoscience. 2014, 11:5307-5321

Lei C, Wagner PD, Foherer N. Identifying the most important spatially distributed variables for explaining land use patterns in a rural lowland catchment in Germany. Journal of Geographical Science. 2019, 29(11):1788-1806

Lei C, Wagner PD, Fohrer N. Effects of land cover, topography, and soil on stream water quality at multiple spatial and seasonal scales in a German lowland catchment. Ecological Indicators, 2021, 120:106940

Li Q, Huang J, Jiang Z, Zhou L, Chu P, Hu K. Detection of urbanization signals in extreme winter minimum temperature changes over Northern China. Climatic Change. 2014, 122: 595-608

Li W, Jiang Z, Zhang X, Li L. On the emergence of anthropogenic signal extreme precipitation change over China. Geophysical Research Letters. 2018a,45: 9179-9185

Li X, Cheng G, Lin H, Cai X, Fang M, Ge Y, Hu X, Chen M, Li W. Watershed system model: the essential to model complex human-nature system at the river basin scale. Journal of Geophysical Research: Atmospheres, 2018b, 123:3019-3034

Li P, Zuo D, Xu Z, Gao X, Peng D, Kan G, Sun W, Pang B, Yang H. Impact of urbanization on variability of annual and flood season precipitation in a typical city of North China. Hydrology Research. 2020a, 51(5): 1150-1169

Li L, Zha Y, Wang R. Relationship of surface urban heat island with air temperature and precipitation in global large cities. Ecological Indicators. 2020b, 117: 106683

Li H, Qin C, He W, Sun F, Du P. Improved predictive performance of cyanobacterical blooms using a hybrid statistical and deep-learning method. Environmental Research Letters. 2021, 16: 124045

Liang K, Jiang Y, Qi J, Fuller K, Nyiraneza J, Meng F. Characterizing the impacts of land use on nitrate load and water yield in an agriculture watershed in Atlantic Canada. Science of the Total Environment. 2020, 729: 138793

Liao J, Tang L, Shao G, Su X, Chen D, Xu T. Incorporation of extended neighborhood mechanisms and its impact on urban land-use cellular automata simulations. Environmental Modelling & Software. 2016, 75:163-175

Lin JY, Huang JL, Prell C, Bryan BA, Changes in supply and demand mediate the effects of land-us change on freshwater ecosystem service flows. Science of the Total Environment, 2021a, 7639(1):143012

Lin JY, Huang JL, Hadjikakou M, Huang YL, Li K, Bryan BA. Reframing water-related ecosystem service flows. Ecosystem Service. 2021b, 50: 101306

Lintern A, Webb JA, Ryu D, Liu S, Bende-Michl U, Waters D, Leahy P, Wilson P, Western AW. Key factors influencing differences in stream water quality across space. WIREs Water. 2018, 5: e1260

Liu J, Kuang W, Zhang Z, Xu X, Qin Y, Ning J, Zhou W, Zhang S, Li R, Yan C, Wu S, Shi X, Jiang N, Yu D, Pan X, Chi W. Spatiaotemporal characteristics, patterns and causes of land-use changes in China since the late 1980s. Journal of Geographical Sciences, 2014,24(2):195-210

Liu R, Xu F, Zhang P, Yu W, Men C. Identifying non-point source critical source areas based on multi-factors. Journal of Hydrology. 2016, 533: 379-388

Liu J, Zhang X, Wu B, Pan G, Xu J, Wu S. Spatial scale and seasonal dependence of land use impacts on riverine water quality in the Huai River basin, China. Environmental Science & Pollution Research. 2017, 24(26): 20995-21010

Liu C, Yang M, Hou Y, Xue X. Ecosystem service multifunctionality assessment and coupling coordination analysis with land use and land cover change in China's coastal zones. Science of The Total Environment. 2021a, 797:149033

Liu M, Ma X, Yin Y, Zhang Z, Yin J, Ullah I, Arshad M. Non-stationary frequency analysis of extreme streamflow disturbance in a typical ecological function reserve of China under a changing climate. Ecohydrology. 2021b, 14: e2323

Liu J. Xu J, Zhang X, Liang Z, Rao K. Nonlinearity and threshold effects of landscape pattern on water quality in a rapidly urbanized headwater watershed in China. Ecolgical Indicators, 2021c, 124:107389

Liu H, Zhou L, Xia J, Chen T, Wang F. Impact assessment of climate and urbanization on the nonstationarity of extreme precipitation: a case study in an urban agglomeration in the middle reaches of the Yangtze rive. Sustainable Cities & Society. 2022, 85: 104038

Lu C, Tian H. Global nitrogen and phosphorus fertilizer use for agriculture production in the past half century: shifted hot spots and nutrient imbalance. Earth System Science Data. 2017, 9(1):181-192

Lu M, Xu Y, Shan N, Wang Q, Yuan J, Wang J. Effect of urbanization on extreme precipitation based on nonstationary models in the Yangtze River Delta metropolitan region. Science of the Total Environment. 2019,672:64-73

Luke A, Vrugt JA, AghaKuchak A, Mattheew R, Sanders BF. Predicting nonstationary flood frequencies: evidence supports an updated stationarity thesis in the United States. Water Resources Research. 2017, 53:5469-5494

Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems. 2017. pp. 4766-4775

Lopez RD, Nash MS, Heggem DT, Ebert DW. Watershed vulnerability predictions for the Ozarks using landscape models. Journal of Environmental Quality. 2008, 37:1769-1780

Ma T, Sun S, Fu G, Hall J, Ni Y, He L, Yi J, Zhao N, Du Y, Pei T, Cheng W, Song C, Fang C, Zhou C. Pollution exacerbates China's water scarcity and its regioanal inequality. Nature Communications. 2020, 11:650

Mahmoodi N, Kiesel J, Wagner PD, Fohrer N. Spatially distributed impacts of climate change and groundwater demand on the water resources in a wadi system. Hydrology & Earth System Sciences, 2021.25: 5065-5081.

Mallinis G, Koutsias N, Arianoutsou M. Monitoring land use/land cover transformations from 1945 to 2007 in two peri-urban mountainous areas of Athens metropolitan area, Greece. Science of the Total Environment. 2014, 490:262-278

Mann HB. Nonparametric tests against trend. Econometrica. 1945, 13:245-259

Marques A, Martins IS, Kastner T, Plutzar C, Theurl MC, Eisenmenger N, Huijbregts MAJ, Wood R, Stadler K, Bruckner M, Canelas J, Hibers JP, Tukker A, Erb K, Pereira HM. Increasing impacts of land use on biodiversity and carbon sequestration driven by population and economic growth. Nature Ecology & Evolution. 2019, 3:628-637

Mentaschi L,Vousdoukas M, Voukouvalas E, Aartini L, Feyen L, Besio G, Alfieri L. The transformed-stationary approach: a generic and simplified methodology for non-stationary extreme value analysis. Hydrology & Earth System Sciences. 2016, 20:3527-3547

Miller CV, Denis JM, Ator SW, Brakebill JW. Nutrients in streams during baseflow in selected environmental setting of the Potomac River Basin. Journal of the American Water Resource Association, 1997, 33(6): 1155-1171

Milly PCD, Betancourt J, Falkenmark M, Hirsch RM, Kundzewicz ZW, Lettenmaier DP, Stouffer RJ. Climate change. Stationarity is dead: whither water management? Science. 2008, 319: 572-574

Mishra V, Ganguly AR, Nijssen B, Lettenmaier DP. Changes in observed climate extremes in global urban areas. Environmental Research Letters. 2015, 10(2): 024005

Moriasi DN, Arnold JG, van Liew MW, Binger RL, Harmel RD, Veith TL. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE. 2007, 50(3): 885-900

Mulkey AS, Coale FJ, Vadas PA, Shenk GW, Bhatt GX. Revised method and outcomes for estimating soil phosphorus losses from agricultural land in the Chesapeake Bay Watershed Model. Journal of Environmental Quality. 2017, 46: 1388-1394

Mundia CN, Anlya M. Analysis of land use/cover changes and urban expansion of Nairobi city using remote sensing and GIS. International Journal of Remote Sensing. 2005,26: 2831-2849

Mustafa A, Rienow A, Saadi I, Cools M, Teller J. Comparing support vector machines with logistic regression for calibrating cellular automata land use change models. European Journal of Remote Sensing. 2018, 51:391-401

Nash JE, Sutcliffe JV. River flow forecasting through conceptual models part I - A discussion of principles. Journal of Hydrology. 1970, 10(3): 282-290

Nasri B, Bouezmarni T, St-Hilaire A, Ouarda TBMJ. Non-stationary hydrologic frequency analysis using B-spline quantile regression. Journal of Hydrology. 2017,554: 532-544

Nielsen A, Trolle D, Sondergaad M, Lauridsen TL, Bjerring R, Olesen JE, Jeppesen E. Watershed land use effects on lake water quality in Denmark. Ecological Application. 2012, 22: 1187-1200

Ning J, Liu J, Kuang W, Xu X, Zhang S, Yang C, Li R, Wu S, Hu Y, Du G, Chi W, Pan T, Ning J. Spatiotemporal patterns and characteristics of land-use change in China during 2010-2015. Journal of Geographical Sciences. 2018, 28: 547-562

Niya AK, Huang J, Kazemzadeh-Zow A, Karimi H, Keshtkar H, Naimi B. Comparison of three hybrid models to simulate land use changes: a case study in Qeshm Island, Iran. Environmental Monitoring & Assessment. 2020, 192: 302

Nonomura A, Kitahara M, Masuda T. Impact of land use and land cover changes on the ambient temperature in a middle scale city, Takamatsu, in Southwest Japan. Journal of Environmental Management. 2009, 90: 3297-3304

Odusanya AE, Mehdi B, Schurz C, Oke AO, Awokola OS, Awomeso JA, Adejuwon JO, Schulz K. Multi-site calibration and validation of SWAT with satellite-based evapotranspiration in a data-sparse catchment in southwestern Nigeria. Hydrology & Earth System Science. 2019, 23:1113-1144.

Ongley ED, Zhang X, Yu T. Current status of agricultural and rural non-point source pollution assesment in China. Environmental Pollution. 2010, 158: 1159-1168

Okwuashi O, Ndehedehe CE. Integrating machine learning with Markov chain and cellular automata models for modelling urban land use change. Remote Sensing Applications: Society and Environment. 2021, 21:100461

Ouarda T, Yousef L, Charron C. Non-stationary intensity-duration-frequency curves integrating information concerning teleconnections and climate change. International Journal of Climatology. 2019, 39: 2306-2323

Pacheco FAL, Santos RMB, Sanches Fernandes LF, Pereira MG, Cortes RMV. Controls and forecasts of nitrate yields in forested watersheds: A view over mainland Portugal. Science of the Total Environment. 2015, 537: 421-440

Pratt B, Chang H. Effects of land cover, topography, and built structure on seasonal water quality at multiple spatial scales. Journal of Hazardous Materials, 2012, 209-210: 48-58

Ragno E, AghaKouchak A, Cheng LY, Sadegh M. A generalized framework for process-informed nonstationary extreme value analysis. Advances in Water Resources. 2019, 130: 270-282

Palmate SS, Wagner PD, Fohrer N, Pandey A. Assessment of uncertainties in modelling land use change with an integrated Cellular Automata-Markov chain model. Environmental Modeling & Assessment. 2022, 27:275-293

Panagoulia D, Econnomou P, Caroni C. Stationary and nonstationary generalized extreme value modelling of extreme precipitation over a mountainous area under climate change. Environmetrics. 2014, 25:29-43

Paudel R, Jawitz JW. Does increased model complexity improve description of phosphorus dynamics in a large treatment wetland? Ecological Engineering. 2012, 42: 283-294

Pathiraja S, Marshall L, Sharma A, Moradkhani H. Detecting non-stationary hydrologic model parameters in a paired catchment system using data assimilation. Advances in Water Resources. 2016, 94: 103-119

Pazur R, Bolliger J. Land changes in Slovakia: past processes and future directions. Applied Geography. 2017, 85:163-175

Perez-Vega A, Mas J, Ligmann-Zielinska A. Comparing two approaches to land/cover change modeling and their implication for the assessment of biodiversity loss in a deciduous tropical forest. Environmental Modelling & Software. 2012, 29: 11-23

Perrin C, Michel C, Andreassian. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429. Journal of Hydrology. 2001, 242:275-301

Pimonsree S, Limsakui A, Kammuang A, Kachenchart B, Kamlangkla C. Urbanization-induced changes in extreme climate indices in Thailand during 1970-2019. Atmospheric Research. 2022, 265: 105882

Pontius RG, Shusas E, McEachern M. Detecting important categorical land changes while accounting for persistence. Agriculture, Ecosystems & Environment. 2004, 101:251-268

Pontius RG, Versluis AJ, Malizia NR. Visualizing certainty of extrapolations from models of land change. Landscape Ecology. 2006, 21:1151-1161

Pontius RG, Millones M. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. International Journal of Remote Sensing. 2011, 32: 4407-4429

Pontius RG, Santacruz A. Quantity, exchange, and shift components of difference in a square contingency table. International Journal of Remote Sensing. 2014, 35(21):7543-7554

Pontius RG. Component intensities to relate difference by category with difference overall. International Journal of Applied Earth Observation & Geoinformation. 2019,77:94-99

Pratt B, Chang H. Effects of land cover, topography, and built structure on seasonal water quality at multiple spatial scales. Journal of Hazardous Materials, 2012,209-210:48-58

Qu Y, Wu N, Guse B, Fohrer N. Distinct indicators of land use and hydrology characterize different aspects of riverine phytoplankton communities. Science of the Total Environment. 2022, 851: 158209

Ren Y, Lu Y, Comber A, Fu B, Harris P, Wu LH.Spatially explicit simulation of land use/land cover changes: current coverage and future prospects. Earth-Science Reviews. 2019, 190: 398-415

Ren X, Mi Z, Georgopoulos PG. Comparison of machine learning and land use regression for fine scale spatiotemporal estimation of ambient air pollution: modeling ozone concentration across the contiguous United States. Environment International. 2020, 142:105827

Ragno E, AghaKouchak A, Cheng L, Sadegh M. A Generalized Framework for processinformed Nonstationary Extreme Value Analysis, Advances in Water Resources, 2019,130, 270-282

Richter BD, Baumgartner JV. Powell J, Braun DP. A method for assessing hydrologic alteration within ecosystems. Conservation Biology, 1996,10(4):1163-1174

Roberts AD, Prince SD. Effects of urban and non-urban land cover on nitrogen and phosphorus runoff to Chesapeake Bay. Ecological Indicators. 2010, 10(2):459-474

Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry & Remote Sensing. 2012, 67: 93-104

Sadegh M, Aghakouchak A, Flores A, Mallakpour I, Nikoo MR. A multi-model nonstationary rainfall-runoff modeing framework: analysis and toolbox. Water Resources Management. 2019,33: 3011-3024

Sang L, Zhang C, Yang J, Zhu D, Yun W. Simulation of land use spatial pattern of towns and villages based on CA-Markov model. Mathematical & Computer Modelling. 2011, 54: 938-943

Sankarasubramanian A, Vogel RM, Limbrunner JF. Climate elasticity of streamflow in the United States. Water Resource Research. 2001,37:1771-1781

Sarhadi A, Ausin MC, Wiper MP, Touma D, Diffenhaugh NS. Multidimensional risk in a nonstationary climate: joint probability of increasingly severe warm and dry conditions. Science Advances. 2018, 4:eaau3487

Scott AA, Waugh DW, Zaitchik BF. Reduced urban heat island intensity under warmer conditions. Environmental Research Letters. 2018, 13(6): 064003

Seong C, Sridhar V, Billah MM. Implications of potential evapotranspiration methods for streamflow estimations under changing climatic conditions. International Journal of Climatology. 2018, 38 :896-914

Shafizadeh-Moghadam H, Minaei M, Feng Y, Pontius RG. Globeland30 maps fours times larger gross than net land change from 2000 to 2010 in Asia. International Journal of Applied Earth Observation & Geoinformation. 2019, 78:240-248

Shi P, Zhang Y, Li Z, Li P, Xu G. Influence of land use and land cover patterns on seasonal water quality at multi-spatial scales. CATENA. 2017, 151: 182-190

Shrestha S, Bhatta B, Shrestha M, Shrestha PK. Integrated assessment of the climate and landuse change impact on hydrology and water quality in the Songkhram River Basin, Thailand. Science of the Total Environment. 2018, 643:1610-1622

Siliverstovs B, Otsch R, Kemfert C, Jaeger CC, Haas A, Kremers H. Climate change and Modelling of extreme temperatures in Switzerland. Stochastic Environmental Research & Risk Assessment. 2010. 24: 311-326

Strumbelj E, Kononenko I. Explaining prediction models and individual predictions with feature contributions. Knowledge & Information Systems. 2014,41: 647-665

Sridhar V, Ali SA, Lakshmi V. Assessment and validation of total water storage in the Chesapeake Bay using GRACE. Journal of Hydrology: Regional Studies. 2019, 24: 100607

Song W, Deng X. Land-use/land-cover change and ecosystem service provision in China. Science of the Total Environment. 2017, 576:705-719

Song X, Zhang J, Zou X, Zhang C, AghaKouchak A, Kong F. Changes in precipitation extremes in the Beijing metropolitan area during 1960-2012. Atmospheric Research. 2019, 222: 134-153

Song X, Mo Y, Xuan Y, Wang Q, Wu W, Zhang J, Zou X. Impacts of urbanization on precipitation patterns in the greater Beijing-Tianjin-Hebei metropolitan region in northern China. Environmental Research Letters. 2021, 16:014042

Sraj M, Viglione A, Parajka J, Bloschl G. The influence of non-stationary in extreme hydrological events on flood frequency estimation. Journal of Hydrology & Hydromechanics. 2016,64(4): 426-437

Steirou E, Gerlitz L, Apel H, Sun X, Merz B. Climate influences on flood probabilities arcoss Europe. Hydrology & Earth System Sciences. 2019,23:1305-1322

Strehmel A, Schmalz B, Fohrer N. Evaluation of land use, land management and soil conservation strategies to reduce non-point source pollution loads in the Three Gorges Region, China. Environmental Management. 2016, 58:906-921

Su C, Chen X. Covariates for nonstationary modeling of extreme precipitation in the Pearl River Basin, China. Atmospheric Research. 2019, 229: 224-239

Sun Y, Zhang X, Ren G, Zwiers FW, Hu T. Contribution of urbanization to warming in China. Nature Climate Change. 2016, 6: 706-710

Sun P, Wen Q, Zhang Q, Singh V, Sun Y, Li J. Nonstationarity-bassed evaluation of flood risk in the Huai River basin, China. Journal of Hydrology. 2018,567:393-404

Takada T, Miyamoto A, Hasegawa SF. Derivation of a yearly transition probability matrix for land-use dynamics and its application. Landscape Ecology. 2010, 25:561-572

Tan Y, Li J, Chen J, Gu B. Hong J. The sinks of dissolved inorganic nitrogen in surface water of wetland mesocosms. Ecological Engineering. 2013, 52: 125-129

Tan M, Liu K, Liu L, Zhu Y, Wang D. Spatialization of population in the Pearl River Delta in 30m grids using random forest model. Progress in Geography. 2017, 36(10): 1304-1312

Tan M, Gassman PW, Liang J, Maywood JM. A review of alternative climate products for SWAT modelling: sources, assessment and future directions. Science of the Total Environment. 2021, 795:148915

Tanir T, de Lima AS, Coelho GA, Uzun S, Cassalho F, Ferreira CM. Assessing the spatiotemporal socioeconomic flood vulnerability of agricultural communities in the Potomac River Watershed. Natural Hazards. 2021, 108: 225-251

Thompson JR, Simons-Legaard E, Legaard K, Domingo JB. A LANDIS-II extension for incorporating land use and other disturbances. Environmental Modelling & Software. 2016, 75: 202-205

Tian Q, Li Z, Sun X. Frequency analysis of precipitation extremes under a changing climate: a case study in Heihe River basin, China. Journal of Water & Climate Change.2021, 12(3):772-786

Tomer MD, Schilling KE. 2009. A simple approach to distinguish land-use and climate-change effects on watershed hydrology. Journal of Hydrology, 376:24-33

Towler EB, Rajagopalan B, Summers RS, Yates D. An approach for probabilitistic forecasting of seasonal turbidity threshold exceedance. Water Resource Research. 2010, 46:W06511

Tromel S, Schonwiese CD. Probability change of extreme precipitation observed from 1901 to 2000 in Germany. Theoretical & Applied Climatology. 2007, 87:29-39

Tu J. Combined impact of climate and land use changes on streamflow and water quality in eastern Masseachusetts, USA. Journal of Hydrology, 2009, 379: 268-283

Tu J. Spatially varying relationships between land use and water quality across an urbanization gradient explored by geographically weighted regression. Applied Geography. 2011, 31:376-392

Tu X, Du X, Singh V, Chen X, Du Y, Li K. Joint risk of interbasin water transfer and impact of the window size of sampling low flows under environmental change. Journal of hydrology, 2017, 554: 1-11

Turner BL, Lambin EF, Reenberg A. Land change science special feature: the emergence of land change science for global environmental change and sustainability. Proceeding of the National Academy of Sciences. 2007, 104:20666-20671

UN General Assembly Oceans and the Law of the Sea, Report of the Secretary-General of 18 August 2004: A/59/62/Add.1

UN WWDR (United Nations World Water Developent Report), Water in a change world. 2009

UN WWDR (United Nations World Water Development Report). Water and climate change. 2020

UN WWDR (United Nations World Water Development Report). Valuing Water. 2021

Van den Besselaar, Tank AMGK, Buishand TA. Trend in European precipitation extremes over 1951-2010. International Journal of Climatology. 2013. 33: 2682-2689

Van Vliet J, Verburg PH, Gradinaru SR, Hersperger AM. Beyond the urban-rural dichotomy: Towards a more nuanced analysis of changes in built-up land. Computer, Environment & Urban Systems. 2019, 74:41-49

Vasilliades L, Galiatsatou P, Loukas A. Nonstationary frequency analysis of annual maximun rainfall using climate covariates. Water Resource Management. 2015, 29:339-358

Verburg PH, van de Steeg J, Veldkamp A, Willemen L. From land cover change to function dynamics: a major challenge to improve land characterization. Journal of Environmental Management. 2009, 90: 1327-1335

Verweij P, Cormont A, van Eupen M, Janssen S, te Roller J, de Winter W, Perez-Soba M, Staritsky IG. Improving the applicability and transparency of land use change modelling: the iCLUE model. Environmental Modelling & Software. 2018, 108:81-90

Viana CM, Santos M, Freire D, Abrantes P, Rocha J. Evaluation of the factors explaining the use of agriculture land: A machine learning and model-agnostic approach. Ecological Indicators. 2021, 131: 108200

Vidon P, Hubbard LE, Soyeux E. Seaaonal solute dynamics across land uses during storms in glaciated landscape of the US Midwest. Journal of Hydrology. 2009,376：34-47

Vinnarasi R, Dhanya CT. Briging realism into a dynamic copula-based non-stationary in intensity-duration model. Advances in Water Resources. 2019,130: 325-338

Vittal H, Ghosh S, Karmakar S, Pathak A, Murtugudde R. Lack of dependence of Indian summer monsoon rainfall extremes on temperature: an observation evidence. Scientific Reports. 2016, 6:31039

Vorosmarty CJ, Mclntyre PB, Gessner MO, Dudgeon D, .Global threats to human water security and reiver biodiveristy. Nature. 2010, 467:555-561

Vu TM, Mishra AK. Nonstationary frequency analysis of the recent extreme precipitation events in the United States. Journal of Hydrology. 2019, 575:999-1010

Wagner PD, Wake B. Importance of spatially distributed hydrologic variables for land use change modeling. Environmental Modelling & Software. 2016, 83:245-254

Wagner PD, Bhallamudi SM, Narasimhan B, Kantakumar LN, Sudheer KP, Kumar S, Schneider K, Fiener P. Dynamic integration of land use changes in a hydrologic assessment of a rapidly developing Indian catchment. Science of the Total Environment. 2016, 539: 153-164

Wagner PD, Fohrer N. Gaining prediction accuracy in land use modeling by integrating modeled hydrologic variables. Environmental Modelling & Software. 2019,115: 155-163

Wagner PD, Bhallamudi SM, Narasimban B, Kumar S, Fohrer N, Fiener P. Comparing the effects of dynamic versus static representations of land use change in hydrologic impact assessments. Environmental Modelling & Software. 2019, 122: 103987

Washburn B, Yancey K, Mendoza J. User Guide for the California impervious Surface Coefficients. Ecotoxicology Program, Integrated Risk Assessment Branch, Office of Environmental Health Hazard Assessment, California Environmental Protection Agency. 2010

Wang J, Feng J, Yan Z, Hu Y, Jia G. Nested high-resolution modeling of the impact of urbanization on regional climate in three vast urban agglomerations in China. Journal of Geophysical Research: Atmospheres. 2012, 117:D21103

Wang G, Zhang J, Xuang Y, Liu F, Jin J, Bao Z, He R, Liu C, Liu Y, Yan X. Simulating the impact of climate change on runoff in a typical river catchment of the Loess Plateau, China. Journal of Hydrometeorology, 2013,14(5):1553-1561

Wang Y, Chen X, Chen Y, Liu M, Gao L. Flood/drought event identification using an effective indicator based on the correlations between multiple time scales of the Standardized Precipitation Index and river discharge. Theoretical & Applied Climatology. 2017, 128: 159-168

Wang Y, Duan L, Liu T, Li J, Feng P. A non-stationary standardized streamflow index for hydrological drought using climate and human-induced indices as covariates. Science of the Total Environment. 2020, 699:134278

Wang Q, Guan Q, Lin J, Luo H, Tan Z, Ma Y. Simulating land use/land cover change in an arid region with the coupling models. Ecological Indicators. 2021a, 122: 107231

Wang Y, Yang G, Gu X, He X, Gao Y, Tian L, Liao N. Application of SWAT model with CMADS data for hydrological simulation in western China. Journal of Water & Climate Change. 2021b, 12:1154-1167

Wang J, Chen F, Doan Q, Xu Y. Exploring the effects of urbanization on hourly extreme rainfall over Yangtze River Delta of China. Urban Climate. 2021c, 36:100781

Wang F, Wang Y, Zhang K, Hu M, Weng Q, Zhang H. Spatial heterogeneity modeling of water quality based on random forest regression and model interpretation. Environmental Research. 2021d, 202: 111660

Wang R, Kim JH, Li MH. Predicting stream water quality under different urban development pattern scenarios with an interpretable machine learning approach. Science of the Total Environment. 2021e, 761:144057

Wang Q, Wang H, Chang R, Zeng H, Bai X. Dynamic simulation patterns and spatiotemporal analysis of land-use/land-cover changes in the Wuhan metropolitan area, China. Ecological Modelling. 2022, 464: 109850

Weiland F, Tisseuil C, Durr HH, Mrac M, Van Beek LP. Selecting the optimal method to calculate daily global reference potential evaporation from CFSR reanalysis data for application in a hydrological model study. Hydrology & Earth System Sciences. 2012, 16:983-1000

Wickham J, Stehman SV, Sorenson DG, Gass L, Dewitz JA. Thematic accuracy assessment of the NLCD 2016 land cover for the conterminous United States. Remote Sensing of Environment. 2021, 257: 112357

Willmott CJ. On the validation models. Physical Geography. 1981, 2:184-194

Wilson AG. A family of spatial interaction models, and associated developments. Environment & Planning A. 1971, 3:1-32

Wojciechowska E, Gajewska M, Ostojski A. Reliability of nitrogen removal processes in multistage treatment wetlands receiving high-streength wastewater. Ecological Engineering. 2017, 365-371

WorldPop, Center for International Earth Science Information Network (CIESIN), Columbia University. Global high resolution population denominators project. 2018. https://dx.doi.org/10.5258/SOTON/WP00660

Wu F, Webster CJ. Simulation of land development through the integration of cellular automata and multicriteria evaluation. Environment & Planning B: Planning & Design. 1998, 25(1): 103-126

Wu H, Li Z, Clarke KC, Shi WZ, Fang LC, Lin AQ, Zhou J. Examining the sensitivity of spatial scale in cellular automata Markov chain simulation of land use change. International Journal of Geographical Information Science. 2019, 33:1040-1061

Xie Z, Liu J, Huang J, Chen Z, Lu X. Linking land cover change with landscape pattern dynamics induced by damming in a small watershed. Remote Sensing. 2022, 14:3580

Xie Z, Zhang Y, Zhang Z, Huang J. Nitrate removal mechanism in riparian groundwater in an internsified agricultural catchment. Agricultural Water Management. 2023, 280:108223

Xu T, Gao J, Coco G. Simulation of urban expansion via integrating artificial neural network with Markov chain-cellular automata. International Journal of Geographical Information Science. 2019, 33(10): 1960-1983

Yang X, Liu Z. Using satellite imagery and GIS for land-use and land-cover change mapping in an estuarine watershed. International Journal of Remote Sensing. 2005, 26: 5275-5296

Yang L, Villarini G, Smith J, Tian F, Hu H. Changes in seasonal maximum daily precipitation in China over the period 1961-2006. International Journal of Climatology. 2013, 33: 1646-1657

Yang X, Zu Z, Liu W, Liu L. Spatiotemporal characteristics of extreme precipitation at multiple timescales over Northeast China during 1961-2014. Journal of Water & Climate Change. 2017, 8(3):535-556

Yang Y, Bao W, Liu Y. Scenario simulation of land system change in the Beijing-Tianjin-Hebei region. Land Use Policy. 2020, 96: 104677

Yang Q, Huang X, Yang J, Liu Y. The relationship between land surface temperature and artificial impervious surface fraction in 682 global cities: spatiotemporal variations and drivers. Environmental Research Letters. 2021, 16: 024032

Yu D, Shi P, Liu Y, Xun B. Detecting land use-water quality relationships form the viewpoint of ecological restoration in an urban area. Ecological Engineering. 2013, 53: 205-216

Yu X, Shen J, Du J. A machine-learning-based model for water quality in coastal waters, taking dissolved oxygen and hypoxia in Chesapeake Bay as an example. Water Resources Research. 2020, 56: e2020WR027227

Zhang Q, Gu X, Singh VP, Xiao M, Xu C.Stationarity of annual flood peaks during 1951-2010 in the Pearl River basin, China. Journal of Hydrology.2014, 519, 3263-3274.

Zhang Z, Huang J, Huang Y, Hong H. Streamflow variability response to climate change and cascade dams development in a coastal China watershed. Estuaine, Coastal & Shelf Science. 2015, 166: 209-217

Zhang Q, Gu X, Singh VP, Shi P, Sun P. More frequent flooding? Changes in flood frequency in Pearl River basin, China, since 1951 and over the past 1000 years. Hydrology and Earth System Sciences. 2018,22: 2637-2653

Zhang Z, Huang J, Zhou M, Huang Y, Lu Y. A coupled modeling approach for water management in river-reservoir system. International Journal of Environmental Research & Public Health. 2019, 16:2949

Zhang Z, Liu J, Huang J. Hydrologic impacts of cascade dams in small head water watershed under climate variability. Journal of hydrology. 2020a, 590: 125426

Zhang Z, Huang J, Xiao C, Huang JC. A simulation-based method to develop strategies for nitrogen pollution control in a creek watershed with sparse data. Environmental Science & Pollution Research. 2020b, 27: 38849-38860

Zhao S, Peng C, Jing H, Tian D, Lei X, Zhou X. Land use change in Asia and the ecological consequences. Ecological Research. 2006, 21:890-896

Zheng F, Hu Y. Assessing temporal-spatial land use simulation effects with CLUE-S and Markov-CA models in Beijing. Environmental Science & Pollution Research. 2018, 25: 32231-32245

Zhou P, Huang JL, Pontius RG, Hong HS. Land classification and change intensity analysis in a coastal watershed of southeast China. Sensors. 2014, 14:11640-11658

Zhou P, Huang J, Pontius RG, Hong H. New Insight into the Correlations between Land Use and Water Quality in a Coastal Watershed of China: Does Point Source Pollution Weaken It? Science of the Total Environment, 2016, 543: 591-600

Zhou Q, Leng G, Su J, Ren Y. Comparison of urbanization and climate change impacts on urban flood volumes: Importance of urban planning and drainage adaption. Science of the Total Environment. 2019: 24-33

Zhou L, Dang X, Sun Q, Wang S. Multi-scenarios simulation of urban land change in Shanghai by random forest and CA-Markov model. Sustainable Cities & Society. 2020, 55:102045

Zhu Y, Chen L, Wei G, Li S, Shen Z. Uncertainty assessment in baseflow nonpoint source pollution prediction: the impacts of hydrographic separation methods, data sources and baseflow period assumptions. Journal of Hydrology. 2019,574:915-925

# Acknowledgments

The time I spent at the Department of Hydrology and Water Resources Management is a unique experience that will continue to encourage, inspire, and motivate me to fight on in my life. Thanks for the positive atmosphere in the department for international students to live and learn.

There is no one more important than Prof. Dr. Nicola Fohrer, who accepted me and gave me a chance to study in her research group. Thank you for always trusting me, which encourage me to explore my ideas during my research. I would like to express my deepest gratitude to you for her patient guidance, timely feedback, and uplifting advice that have helped me to meet every milestone during my research.

Many thanks to Dr. Paul Wagner for your patience and dedication to help me start this challenging yet exciting journey in Kiel and for your generous support and advice on my research. I cannot thank enough for Georg Hörmann, who is an excellent mentor and friend to me. Not only because of his notes and scripts for my research but also because of his generous and affable way to help me understand various cultures, which is so much fun to me.

I also want to thank members of the Department of Hydrology and Water Resources Management: Dr. Nariman Mahmoodi, Dr. Daniel Rosado, Dr. Lishani Nisansala Wijewardene, Dr. Tibebe Belete Tigabu, Dr. Yueming Qu, Dr. Chaogui Lei, Dr. Sun Xiuming, Mr. Ayenew Desalegn Ayalew, Mr. Lukas Paul Loose, Mr. Sven Grantz, Ms. Maryam Rhezanezhad, Ms. Henrike Traute Risch, Ms. Anne-Kathrin Wendell, and Ms. Kristin Peters for their friendship and company.

I forward my gratitude to my friend and supervisor, Prof. Dr. Jinliang Huang from Xiamen University, who led me to the area of watershed science. Thanks for always being supportive to me over the past 15 years. I would like to thank Prof. Jiacong Huang, who kindly recommended me to join this nice group in Kiel, Germany. I also would like to acknowledge the support from China Scholarship Council.

Finally, I would like to forward my heartfelt thanks to my family.

# Appendix

## Supplementary materials: Chapter 2

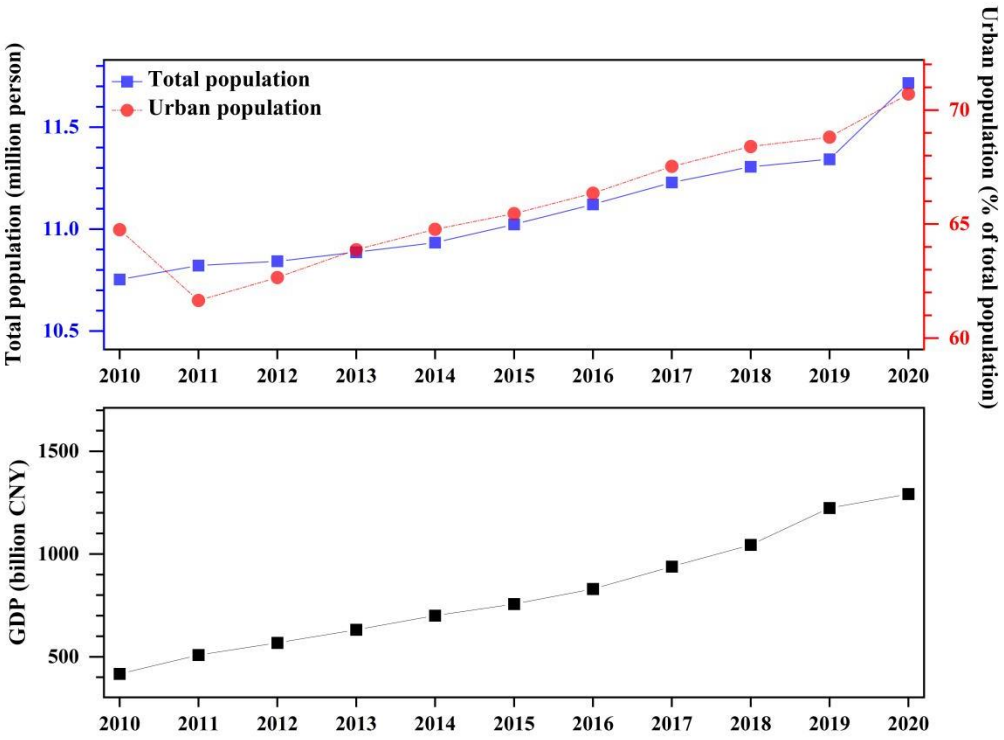


Fig. S2-1 Population and GDP in the MRW during 2010-2020

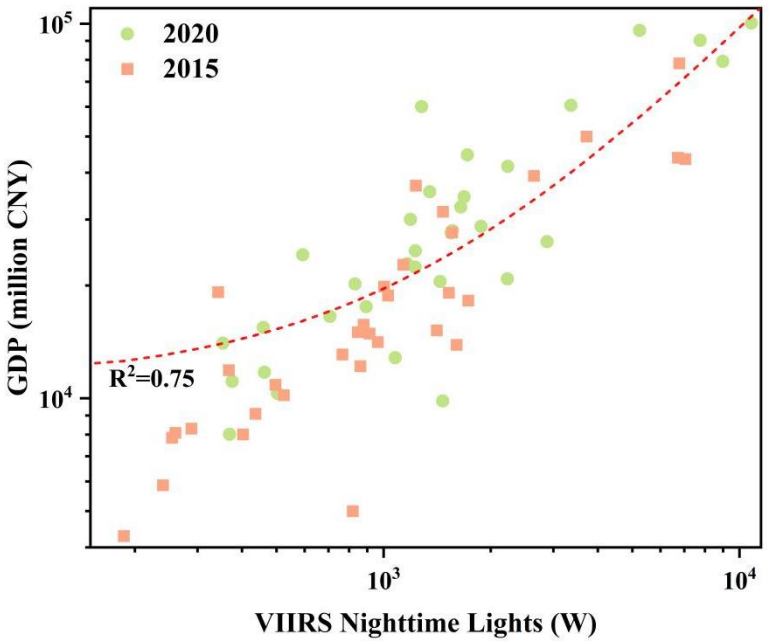


Fig S2-2 Relationship between GDP and VIIRS Nighttime Lights

The unit of VIIRS Nighttime Lights data provided by the Earth Observation Group, Payne Institute for Public Policy, Colorado School of Mines (https://payneinstitute.mines.edu/eog/) is $nW/cm^2/sr$. However, this unit do not allow us to compare the counties or districts in the Minjiang River Watershed. Thus, we have to convert it to other unit allowing us to estimate the information of nighttime lights in these counties and districts.

The resolution of the VIIRS Nighttime Lights data is 15 arc second (~500m at the Equator), and the area of each pixel is

$$\left(500m \cdot \frac{1cm}{0.01m}\right)^2 = 2.5 \times 10^9 cm$$

The sr is the unit for the solid angles, thus we convert it to a unit for the flat surface

$$\frac{1nW}{(cm^2/sr)} \cdot \frac{2.5 \times 10^9 cm^2}{2\pi \cdot sr} = 3.981 \times 10^8 nW = 0.3981W$$

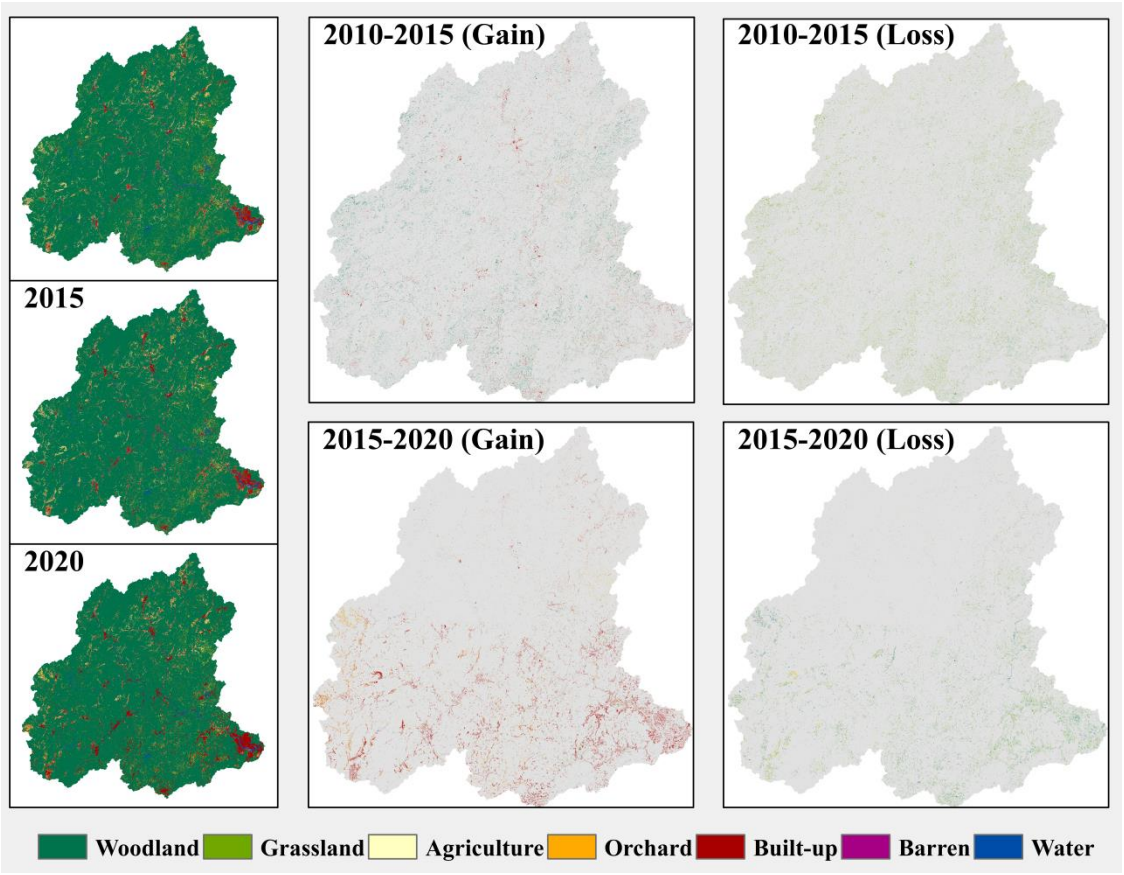


Fig S2-3 Maps of land use categories at three time point and two time intervals

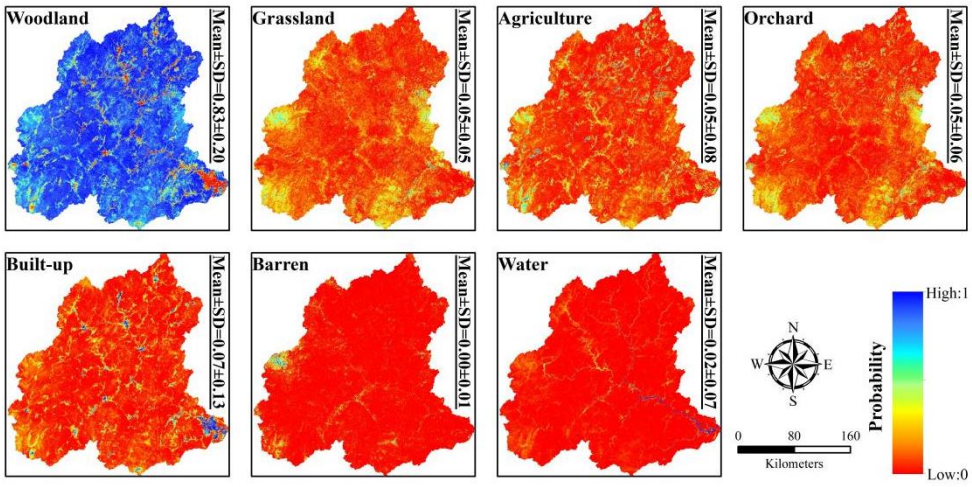


Fig S2-4 Transition suitability images for Scenarios I

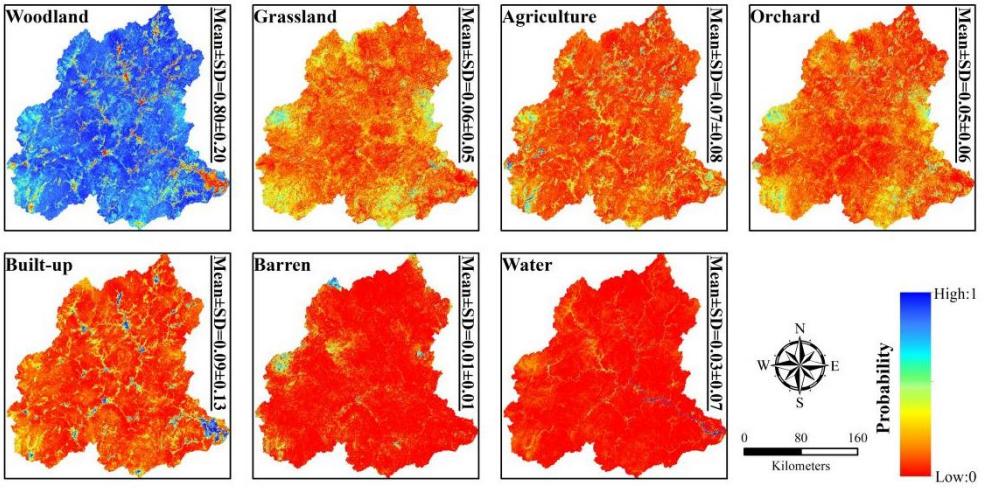


Fig S2-5 Transition suitability images for Scenarios II

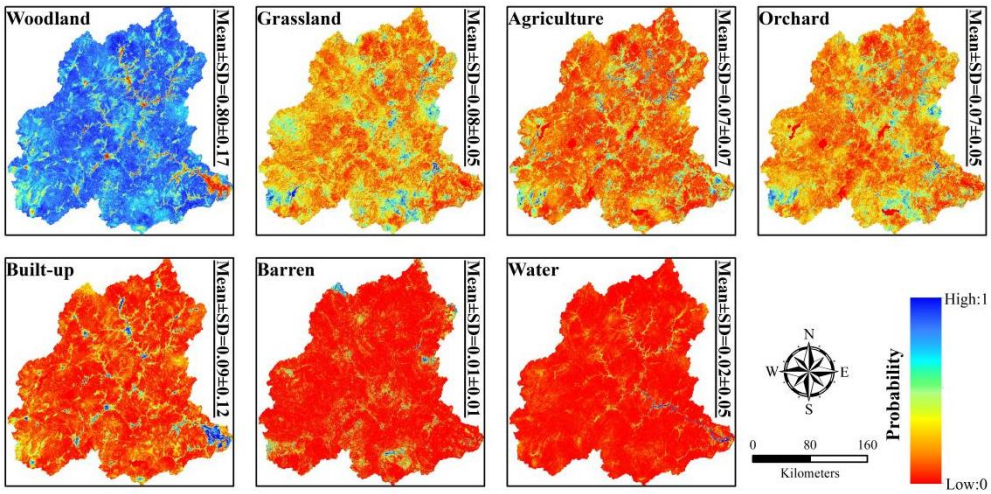


Fig S2-6 Transition suitability images for Scenarios III

Table S2-1 result of accuracy assessment for each class

| Class | User's Accuracy | | | Producer's Accuracy | | |
|---|---|---|---|---|---|---|
| | 2010 | 2015 | 2020 | 2010 | 2015 | 2020 |
| Agriculture | 0.88±0.05 | 0.95±0.04 | 0.81±0.06 | 0.84±0.05 | 0.82±0.05 | 0.72±0.06 |
| Built-up | 0.81±0.06 | 0.84±0.06 | 0.90±0.05 | 0.94±0.04 | 0.92±0.04 | 0.98±0.03 |
| Orchard | 0.76±0.07 | 0.84±0.06 | 0.85±0.06 | 0.79±0.06 | 0.90±0.05 | 0.82±0.05 |
| Water | 0.90±0.05 | 0.95±0.04 | 0.93±0.04 | 0.92±0.04 | 0.92±0.04 | 0.98±0.02 |
| Forest | 0.74±0.07 | 0.94±0.04 | 0.97±0.03 | 0.81±0.06 | 0.88±0.05 | 0.93±0.04 |
| Bareland | 0.72±0.07 | 0.69±0.08 | 0.71±0.07 | 0.66±0.06 | 0.84±0.06 | 0.89±0.05 |
| Grassland | 0.84±0.06 | 0.83±0.06 | 0.89±0.05 | 0.73±0.06 | 0.78±0.06 | 0.79±0.05 |

Table S2-2 Overall of accuracy assessment

| Metrics | 2010 | 2015 | 2020 |
|---|---|---|---|
| Overall accuracy | 0.805±0.024 | 0.863±0.021 | 0.864±0.021 |
| Allocation disagreement | 0.155 | 0.090 | 0.088 |
| -Shift | 0.073 | 0.037 | 0.031 |
| -Exchange | 0.081 | 0.054 | 0.058 |
| Quantity disagreement | 0.041 | 0.047 | 0.048 |

Table S2-3 Variables for the confusion matrix

| | Observed (Yes) | Observed (No) |
|---|---|---|
| Simulated (Yes) | Hits (H) | False Alarms (FA) |
| Simulated (No) | Misses (M) | Correct Rejections (CR) |

**Supplementary materials: Chapter 3**

Table S3-1 Streamflow simulation with other models in Southeast China

| Model | Study area | Discharge area ($km^2$) | Monthly streamflow | | Daily streamflow | | Citation |
|---|---|---|---|---|---|---|---|
| | | | Cal* | Val** | Cal | Val | |
| SWAT*** | NRW, Jiulong River Watershed | 8459 | 0.86 | 0.86 | 0.64 | 0.6 | Huang et al., 2013b |
| | L-S, Jiulong River Watershed | 132 | 0.92 | 0.77 | 0.89 | 0.73 | Zhang et al., 2020 |
| HSPF**** | SRW, Minjiang River Watershed | 9918 | - | - | 0.78 | 0.75 | Zhang et al., 2019 |
| | JRW, Minjiang River Watershed | 14816 | - | - | 0.78 | 0.74 | Zhang et al., 2019 |
| | Aojiang River Watershed | 2676 | 0.67 | 0.61 | - | - | Lin et al., 2021a |
| | NRW, Jiulong River Watershed | 8459 | 0.79 | 0.65 | - | - | Lin et at., 2021b |
| | WRW, Jiulong River Watershed | 3491 | 0.65 | 0.65 | - | - | Lin et al., 2021b |

Note: *Cal: Nash-Sutcliffe efficiency coefficient (NSE) of the calibration period; **Val: NSE of the validation period; ***SWAT: Soil and Water Assessment Tool; ****HSPF: Hydrological Simulation Program - FORTRAN

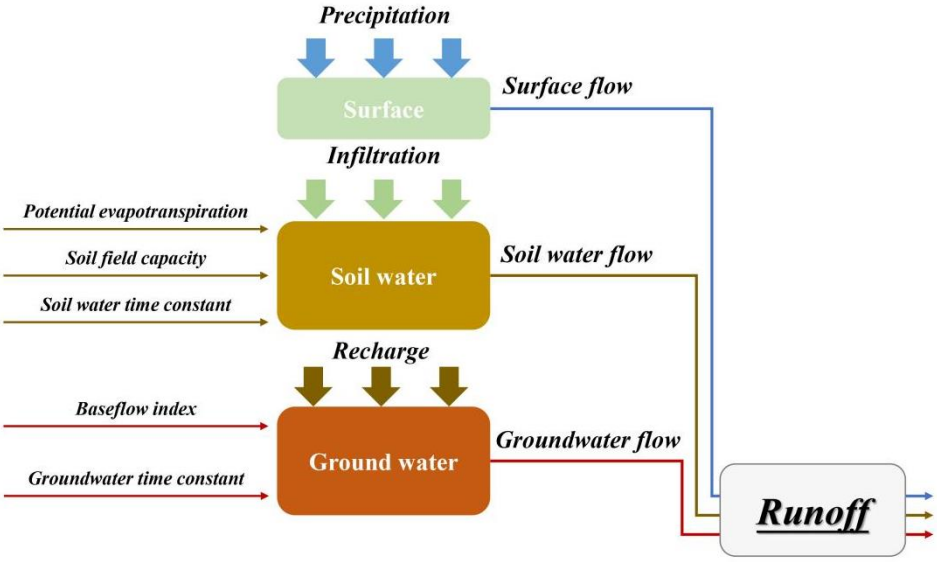


Fig S3-1 The structure of rainfall-runoff model

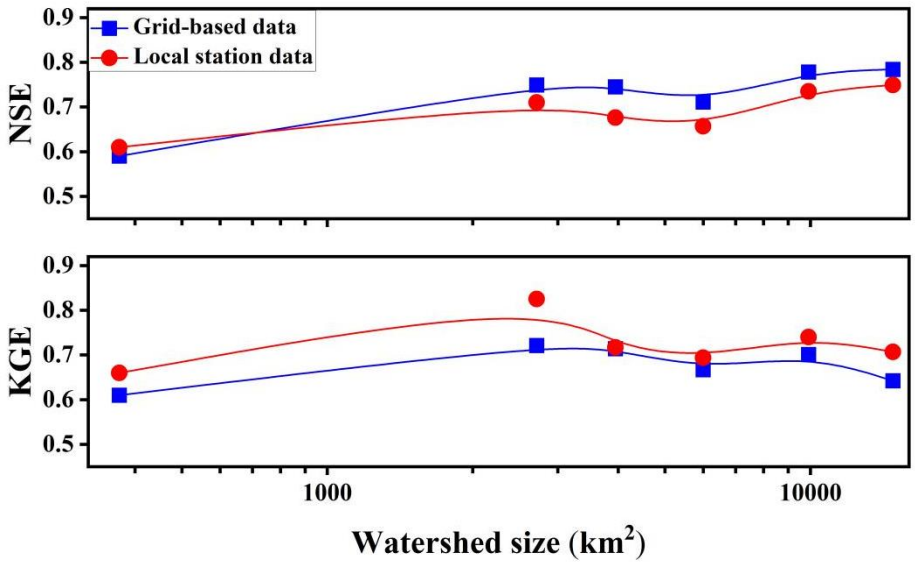


Fig S3-2 Performance of the model with different sources of meteorological data during baseline period

## Supplementary materials: Chapter 4

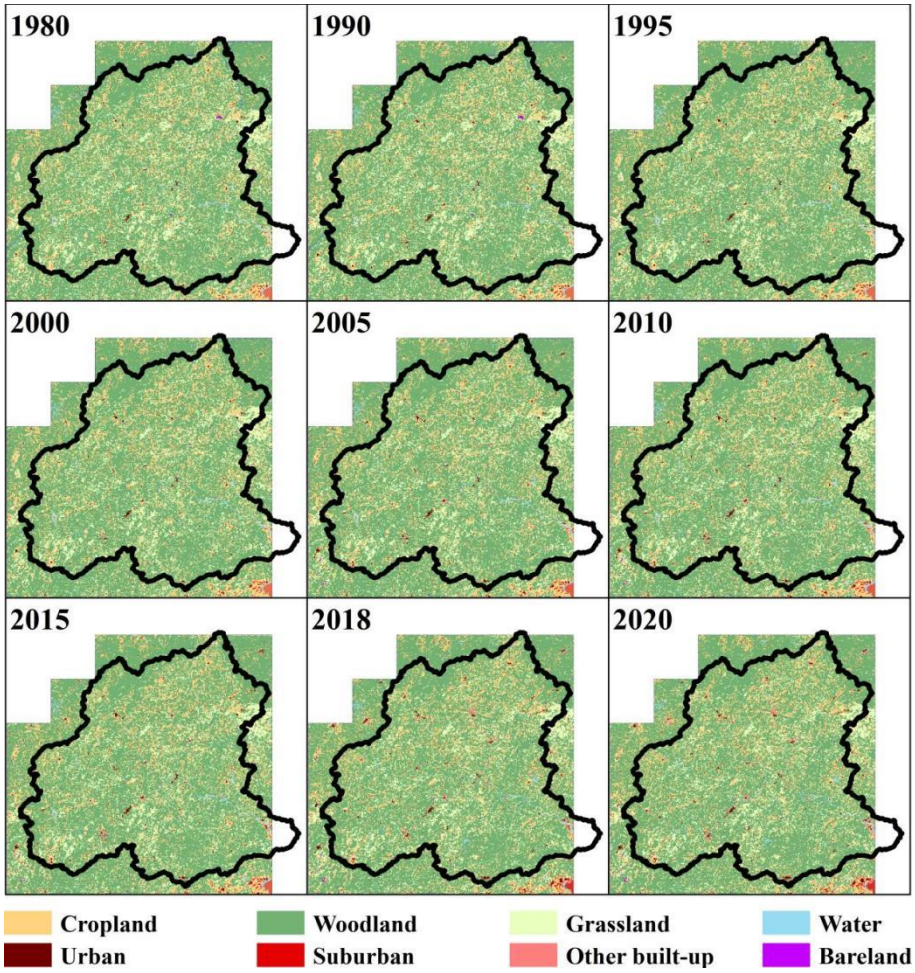


Fig S4-1 Land use patterns in the MRW

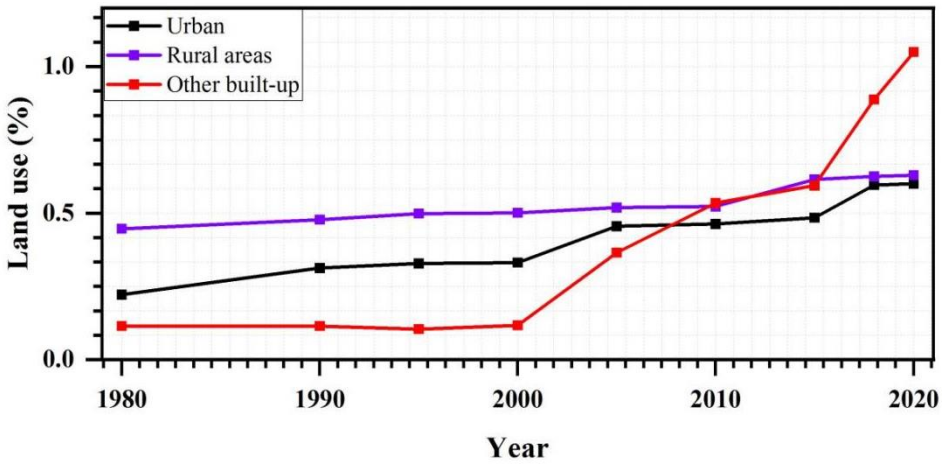


Fig S4-2 Overview of urbanization process in the MRW

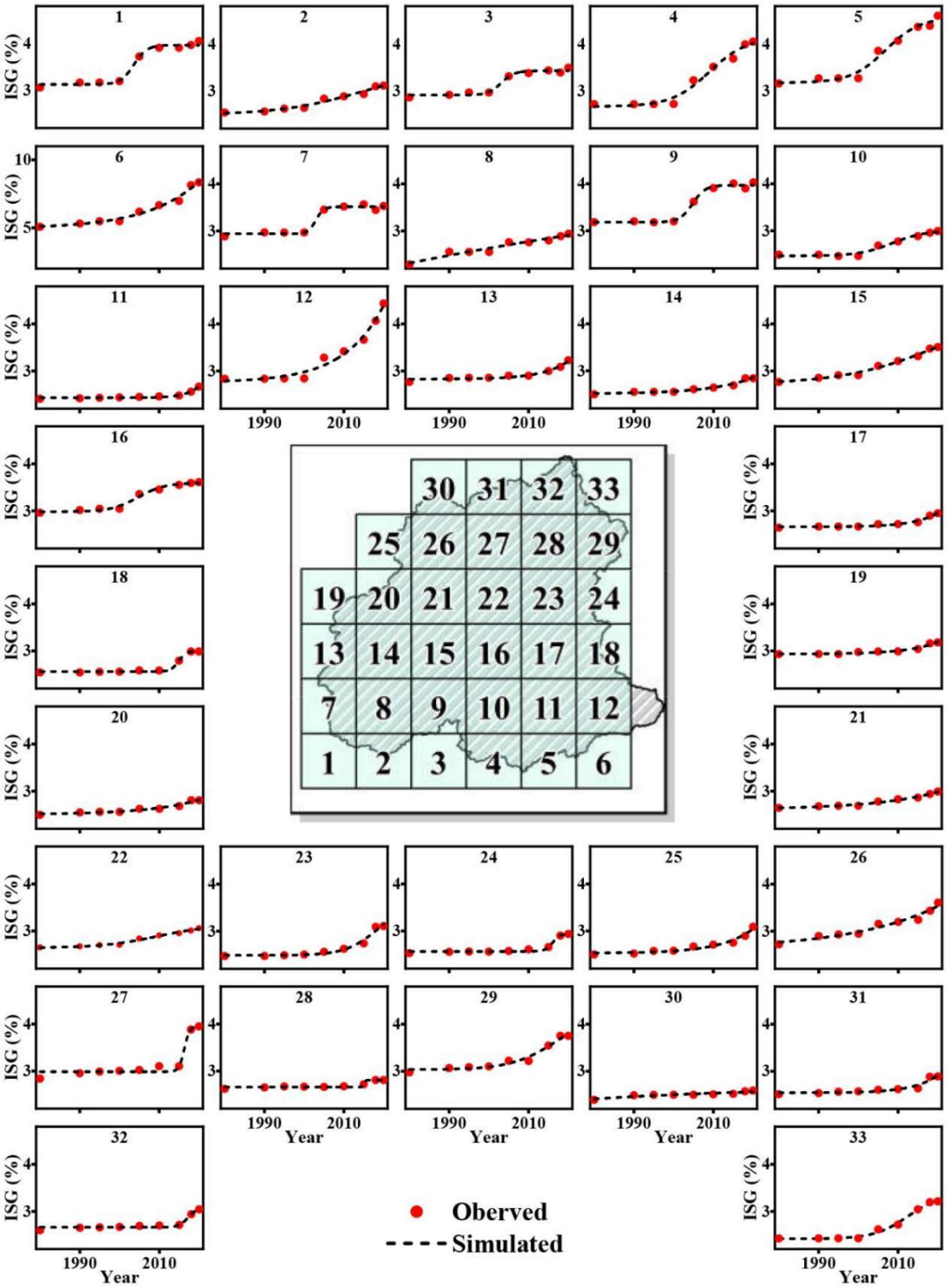


Fig S4-3 Logistic models for urbanization in the MRW

Table S4-1 Transfer matrix for different periods in terms of urbanization (grid cells)

| From | To | 1980-1990 | 1990-2000 | 2000-2010 | 2010-2020 |
|---|---|---|---|---|---|
| Rural | Rural | 27 | 26 | 20 | 12 |
| | Subrban | 2 | 1 | 6 | 8 |
| | Urban | 0 | 0 | 0 | 0 |
| Suburban | Rural | 0 | 0 | 0 | 0 |
| | Subrban | 3 | 5 | 5 | 7 |
| | Urban | 0 | 0 | 1 | 4 |
| Urban | Rural | 0 | 0 | 0 | 0 |
| | Subrban | 0 | 0 | 0 | 0 |
| | Urban | 1 | 1 | 1 | 2 |

**Supplementary materials: Chapter 6**

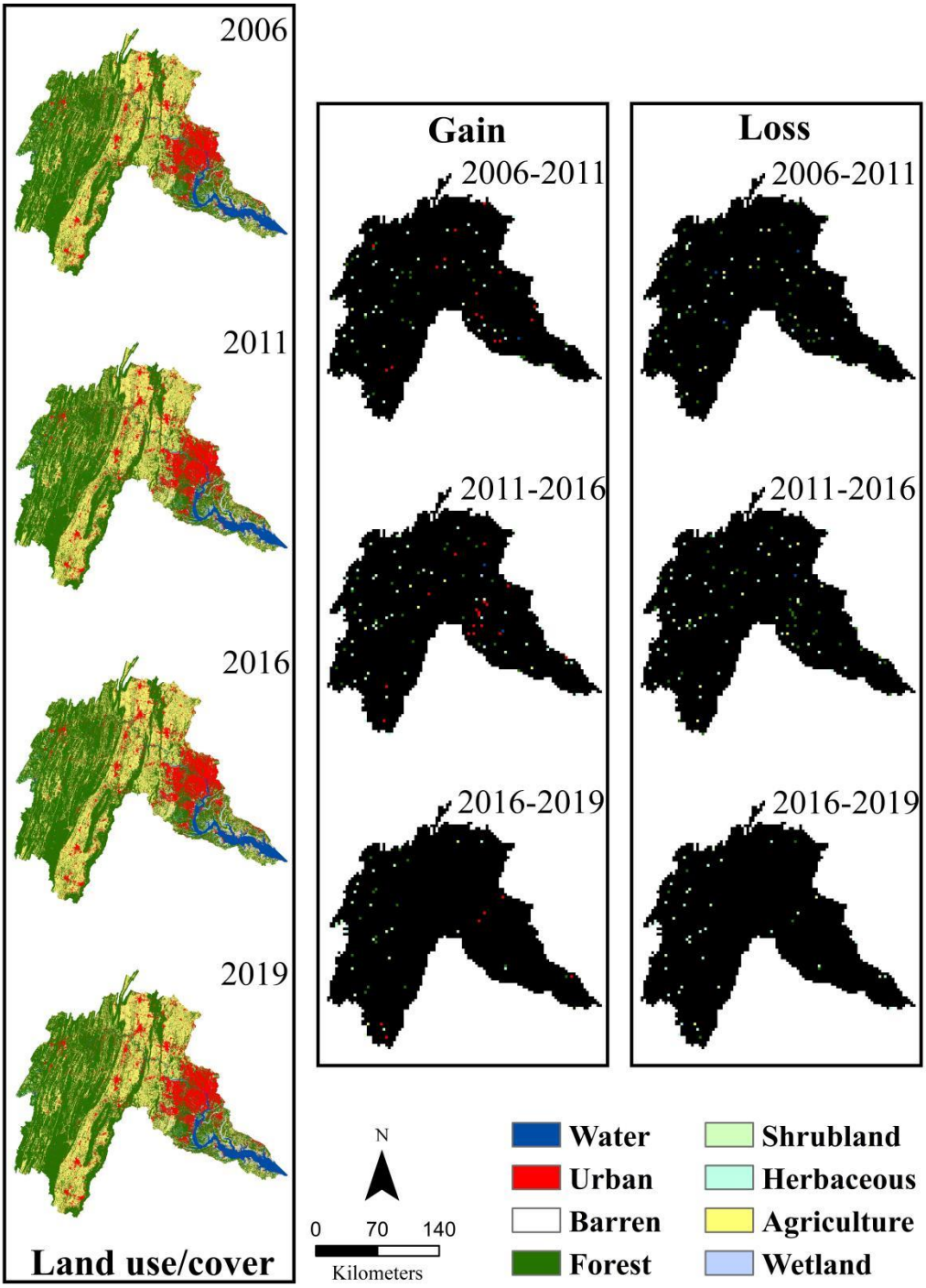


Fig. S6-1 Maps of land use/cover, loss and gains in the PRB

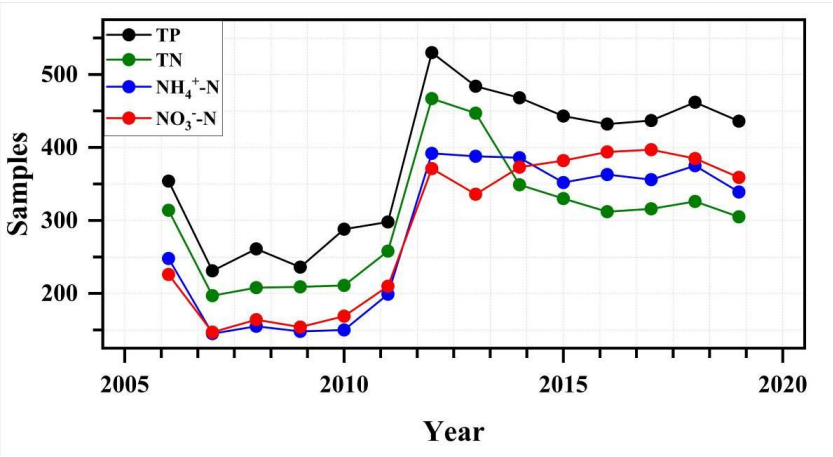


Fig S6-2 Overview of water quality sampling during 2006-2019

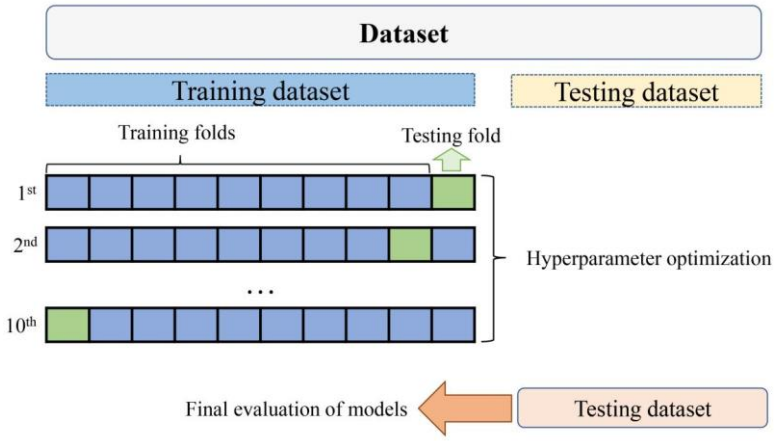


Fig S6-3 10-fold cross-validation

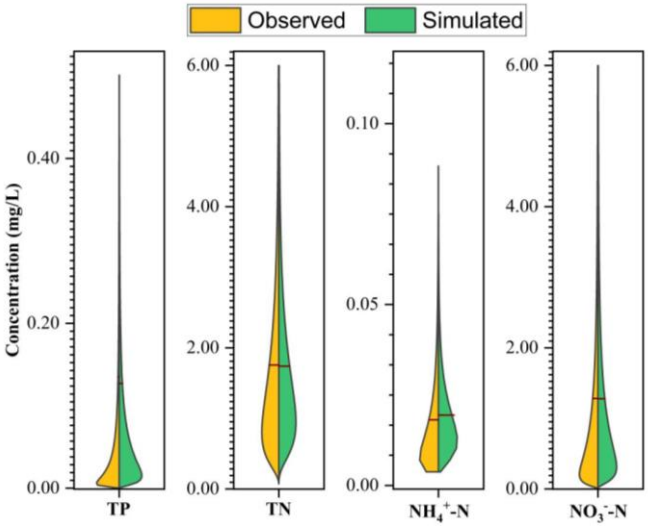


Fig S6-4 Result of observed and simulated data in the baseline

Table S6-1 land use transfer matrix for different periods in the PRB (%)

| To | From | 2006-2011 | 2011-2016 | 2016-2019 | To | From | 2006-2011 | 2011-2016 | 2016-2019 |
|---|---|---|---|---|---|---|---|---|---|
| Water | Urban | 0 | 0 | 0 | Shrubland | Water | 0 | 0 | 0 |
| | Barren | 0.57 | 0.7 | 0.99 | | Urban | 0 | 0 | 0 |
| | Forest | 0.02 | 0.01 | 0 | | Barren | 0.11 | 0.05 | 12.1 |
| | Shrubland | 0.1 | 0.03 | 0 | | Forest | 0.32 | 0.13 | 0.14 |
| | Herbaceous | 0.24 | 0.13 | 0.11 | | Herbaceous | 15.54 | 25.32 | 31.5 |
| | Agriculture | 0.01 | 0 | 0 | | Agriculture | 0.32 | 0.02 | 0 |
| | Wetland | 0.47 | 0.14 | 0.05 | | Wetland | 0.01 | 0 | 0 |
| Urban | Water | 0.03 | 0.01 | 0.02 | Herbaceous | Water | 0.06 | 0.07 | 0.03 |
| | Barren | 2.04 | 0.81 | 0.84 | | Urban | 0 | 0 | 0 |
| | Forest | 0.26 | 0.13 | 0.22 | | Barren | 0.58 | 0.35 | 41.95 |
| | Shrubland | 1.36 | 0.32 | 0.70 | | Forest | 1.13 | 0.64 | 0.71 |
| | Herbaceous | 6.93 | 1.6 | 3.45 | | Shrubland | 3.19 | 5.89 | 2.11 |
| | Agriculture | 0.68 | 0.31 | 0.43 | | Agriculture | 0.06 | 0.05 | 0.01 |
| | Wetland | 0.27 | 0.11 | 0.2 | | Wetland | 0.02 | 0.01 | 0.01 |
| Barren | Water | 0.04 | 0.04 | 0.04 | Agriculture | Water | 0.01 | 0.01 | 0 |
| | Urban | 0 | 0 | 0 | | Urban | 0 | 0 | 0 |
| | Forest | 0 | 0 | 0.04 | | Barren | 0.20 | 0.09 | 0.76 |
| | Shrubland | 0.02 | 0.02 | 0.15 | | Forest | 0.16 | 0.06 | 0.02 |
| | Herbaceous | 0.21 | 0.14 | 1.80 | | Shrubland | 2.77 | 1.68 | 0.39 |
| | Agriculture | 0 | 0.01 | 0.04 | | Herbaceous | 1.27 | 1.64 | 5.14 |
| | Wetland | 0 | 0 | 0 | | Wetland | 0.21 | 0.23 | 0.03 |

Table S6-1 land use transfer matrix for different periods in the PRB (%) (continued)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Water | 0.3 | 0.1 | 0.04 | | Water | 0.21 | 0.18 | 0.45 |
| | Urban | 0 | 0 | 0 | | Urban | 0 | 0 | 0 |
| | Barren | 1.13 | 2.86 | 0.65 | | Barren | 0.03 | 0.03 | 0.03 |
| Forest | Shrubland | 55.91 | 44.44 | 65.10 | Wetland | Forest | 0.01 | 0 | 0 |
| | Herbaceous | 11.56 | 14.16 | 11.53 | | Shrubland | 0.02 | 0.04 | 0.05 |
| | Agriculture | 0.27 | 0.13 | 0.03 | | Herbaceous | 0.08 | 0.08 | 0.02 |
| | Wetland | 0.13 | 0.08 | 0 | | Agriculture | 0.03 | 0.01 | 0.01 |

Table S6-2 Descriptive statistics of pollutant concentration (mg/L)

| | TP | | | | |
|---|---|---|---|---|---|
| | Nature | Semi-nature | Agriculture | Mix | Urban |
| Min | 0.004 | 0.003 | 0.010 | 0.005 | 0.003 |
| Max | 3.006 | 0.870 | 0.263 | 1.268 | 0.716 |
| Mean | 0.040 | 0.049 | 0.047 | 0.098 | 0.079 |
| Standard deviation | 0.202 | 0.086 | 0.046 | 0.143 | 0.115 |
| | TN | | | | |
| Min | 0.102 | 0.100 | 1.693 | 0.251 | 0.220 |
| Max | 3.911 | 7.164 | 3.566 | 6.403 | 8.850 |
| Mean | 0.787 | 1.153 | 2.733 | 3.662 | 1.382 |
| Standard deviation | 0.414 | 0.719 | 0.408 | 1.310 | 0.664 |
| | $NH_4^+$-N | | | | |
| Min | 0.002 | 0.002 | 0.004 | 0.002 | 0.003 |
| Max | 0.055 | 1.527 | 0.149 | 0.186 | 0.253 |
| Mean | 0.010 | 0.081 | 0.022 | 0.025 | 0.033 |
| Standard deviation | 0.009 | 0.214 | 0.021 | 0.028 | 0.044 |
| | $NO_3^-$-N | | | | |
| Min | 0.001 | 0.018 | 1.199 | 0.019 | 0.174 |
| Max | 1.587 | 2.361 | 3.377 | 5.467 | 1.937 |
| Mean | 0.567 | 0.562 | 2.409 | 2.418 | 0.894 |
| Standard deviation | 0.302 | 0.458 | 0.455 | 1.729 | 0.402 |

# Declaration

I, Zhneyu Zhang, hereby declare that the dissertation submitted, entitled “Data-driven Analysis of Potential Impacts of Land-use/cover Change on Water Resources in Coastal Watersheds: Perspectives from Non-stationarity and Nonlinearity” was written independently by me. The content and design of this thesis, apart from the supervisor’s guidance, is my own work. The thesis has not been submitted either partially or wholly as a part of a doctoral degree to another examining body and is my first and only doctoral procedure. Chapter 3 and 6 of the thesis have been published in peer-reviewed journals and chapter 2, 4, 5 is currently under review. This work has been prepared respecting the Rules of Good Scientific Practice of the German Research Foundation. I have not been deprived of an academic degree.

Kiel, March-2023

Zhenyu Zhang