Master's Thesis of Landscape Architecture

Regional Waterlogging Factors Derived by Geographically Weighted Regression and Shapley Additive Explanations

지리 가중 회귀모형 및 섀플리 가법 설명모형에 의한 지역침수 영향요인 분석

August 2022

Seoul National University
Graduate School of Environmental Studies
Landscape Architecture
XIAOLING JIN

Regional Waterlogging Factors Derived by Geographically Weighted Regression and Shapley Additive Explanations

Prof. Youngkeun Song

Submitting a master's thesis of Public Administration
April 2022

Seoul National University
Graduate School of Environmental Studies
Dept. of Landscape Architecture
XIAOLING JIN

Confirming the master's thesis written by
XIAOLING JIN
June 2022

Chair __________________________ (Seal)

Vice Chair ______________________ (Seal)

Examiner _______________________ (Seal)

Regional Waterlogging Factors Derived by Geographically Weighted Regression and Shapley Additive Explanations

서울대학교 환경대학원 환경조경학과
XIAOLING JIN

위 논문은 서울대학교 및 환경대학원 환경조경학과 학위논문
관련 규정에 의거하여 심사위원의 지도과정을 충실히
이수하였음을 확인합니다.

2022년 8월

위 원 장 ______________ (서울대학교 환경대학원 교수)

부위원장 ______________ (서울대학교 농업생명과학대학 교수)

위 원 ______________ (서울대학교 환경대학원 교수)

Abstract

Regional Waterlogging Factors Derived by Geographically Weighted Regression and Shapley Additive Explanations

XIAOLING JIN
Dept. of Landscape Architecture
Graduate school of Environmental Studies
Seoul National University

The landscape is considered as a key component of the ecosystem intervention. Human activities have significantly changed the surface characteristics, such as affected the circulation and flow of natural materials and energy, or weakened the rainwater collection, storage function and runoff drainage capacity of the watershed. These led to waterlogging disasters and increased the risk to the living environment. Therefore, landscape planners and decision-makers need to constantly improve and optimize the landscape pattern to maintain the ecosystem's dynamic balance and reduce waterlogging at the same time. Development of remote sensing technology makes it possible to study large-scale watershed units, meanwhile the experiments on such large-scale sites can be verified by theory. Existing research on verification of theories ignored important interactions within the landscape pattern because the traditional linear regression model (a subfield of supervised learning) such as Geographically Weighted Regression (GWR) could not analyze the relationship between independent variables while analyzing the relationship between independent variables and dependent variables. In recent years, development of interpretable machine learning models in the field of machine learning is making up for this shortcoming. Among them, Shapley Additive Explanations (SHAP) is a representative method which provides an interpretable machine learning model based on game theory. It can not only analyze the relationship between independent variables and dependent variables, but also take into account *correlations between multiple independent variables*, and produce importance ranking according to the contribution degree. Through our extensive and thorough verification and comparative analysis of the two methods, we first find that in the analysis results of GWR, the Shannon Diversity Index (SHDI, one representative landscape metric) is seriously underestimated, while in the results of SHAP, SHDI shows a great impact on waterlogging in any scale of watershed units. At the same time, according to the prediction result of Prediction Mean Squared Error (MSE), although the error value of GWR is small, SHAP is still far more accurate than GWR. Secondly, the water cycle process has characteristics of producing multi-scale geographical watersheds. In order to taking into account the dynamic balance of hydrology, conducting comparative analysis of multi-level watershed-scale units is necessary. Our results show that the use of finer-scale watersheds as the research scale is not necessarily suitable for waterlogging research. In this study, we find that analysis on waterlogging in the Seoul Capital Area (SCA) based on Large-scale watershed units (LSWU) is the most appropriate and accurate. Finally, it is naturally assumed that a threshold for landscape pattern characteristics exists. When the impact on waterlogging reaches this critical point, its role in promoting or alleviating waterlogging will change. Through estimating threshold values of landscape pattern characteristics, the purpose of waterlogging disaster mitigation can be achieved accurately and at a low cost. In summary, this study explores the new analysis method of interactions between landscape patterns and waterlogging, and provid a reference for methods and results of waterlogging control based on landscape ecology.

……………………………………

keywords : Waterlogging, Landscape Patterns, Seoul Capital Area, Geographically Weighted Regression, Shapley Additive Explanations
Student Number : 2020-20489

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Urbanization and Human Intelligence

Urban is an important part of human civilization and development progress. With the acceleration of global urbanization, 66% of the population will live in large urban areas by 2050 [25]. In the process of urbanization, multiple factors keep changing restlessly such as the expansion of roads, the increase of buildings, the increase or decrease of urban green space, and other landscape patterns affections on the ecological environment. These factors then change the spatial characteristics of the urban and thus greatly impact the natural landscape and ecological cycle of urban areas [27, 30, 54, 1]. Therefore, urban planning is important and needs to be constantly adjusted and optimized to adapt to the urbanization development. Due to the inherent complexity of the urban itself, urban planners cannot fully predict how the ecological cycle process would be influenced and disturbed by urban expansion, which has caused several environmental problems that increased the risk to the living environment [2]. Especially in recent years, the impact of urban waterlogging has become more and more significant, and its severity and frequency are increasing all over the world. Therefore, in order to balance the harmony between human and the environment, research on urban ecological environment is thus very urgent and necessary (Fig. 1). Human activities can play a positive role through reflection and consciously promoting constructive change [55]. Nowadays, experts mainly manage and plan the whole landscape through landscape ecology and improve the management of the urban ecological environment [38]. Here, landscape ecology is regarded as the scientific basis of land use development and restoration, landscape planning and management, and is the basis of extensive interdisciplinary human ecology. It is a people-centered research field on the complex interrelationship among society, economy, geography, and culture related to land use. Study the landscape as a whole composed of different and interactive elements, and seek a compromise among these conflicting is needed in order to enrich the human ecological environment. This emphasis on the impact of human beings on land use and the quantitative evaluation of the specific functions of the landscape makes landscape ecology the basis for creating balanced and visionary policy and decision-making tools and helps overcome the tension between modern society and its landscape caused by the increasing needs of industrial development and the potential of natural land [48].

1.2 Landscape and Landscape Ecology

Initially, the definition of "landscape" was limited to the visual perception and aesthetic evaluation of the environment, rather than the ecological environment. However, with the development of land use evaluation ability, "landscape" has began to be considered as a part of the physical environment. Since then, the landscape is considered as a heterogeneous land area composed of interactions among ecosystems [13]. It is a spatial body containing all ecological environments within the human living environment [50, 14]. With the awakening of environmental consciousness dominated by human thinking [61], according to the definition of "landscape ecology", the landscape is not only regarded as the carrier of the ecosystem but also as a control system through land use and the total or partial intervention and control of human intelligence [9]. Among them, the land is the key central point of landscape ecology [72, 48], which means, the evaluation of the overall nature of vegetation is also necessary [12]. The combination of land as the research object and variable with the key variable of human intervention and control [62, 48, 56] helps to clarify the relationship between the landscape structure (the element of spatial pattern) and the ecosystem (the change of ecological process). This is also the main goal of landscape ecology, that is, to clarify the relationship between the internal dynamics and interaction of landscape [58]. In other words, the quantification of spatial heterogeneity is necessary to clarify the relationship between ecological processes and spatial patterns [57, 59]. Refer to Fig. 1 for this process.

1.3 Land Use Land Cover and Landscape Pattern Metrics

The measurement, analysis, and interpretation of spatial patterns through the three main characteristics of landscape structure, function, and change have attracted much attention in landscape ecology [24]. Concretely, structure refers to the spatial relationship between unique ecosystems or elements (i.e., landscape composition and configuration). Function refers to the interaction relationship between spatial features. Change refers to the change of ecological structure and function over time [16, 13]. The landscape pattern metrics (also known as landscape metrics or spatial pattern metrics) are methods to quantify the specific spatial characteristics of patches, patch classification, and patch mosaic near each focus unit or the whole landscape mosaic. It is an effective tool for mapping and quantifying Land Use Land Cover (LULC) characteristics, which is broadly used in landscape ecology research [23, 41, 49, 19, 4]. Based on the compatibility between landscape pattern index and GIS, it is susceptible to landscape definition in resolution, range, and landscape boundary [45]. Therefore, it can objectively and accurately provide the change mode of landscape structure and its configuration information [21]. At the same time, landscape pattern metrics can be used to analyze different landscape levels: (1) patch level (i.e., grassland area), (2) class level (i.e., all grassland areas in a specific area), (3) landscape level (i.e., all landscape types in a specific area) [35, 44]. Therefore, it can measure, analyze, and explain the change of spatial heterogeneity and its impact on ecosystems on different spatial scales (patch, class, or landscape-level) [3, 58, 10, 7, 33, 34, 45, 20, 66]. The ecosystem is the change of ecological process, that is, the change caused by the interaction of landscape structure with time [11]. The landscape pattern metrics can not only help us understand and explain the structure and change of the landscape from different aspects, but also objectively reflect the spatial heterogeneity of the landscape [26, 32]. To this end, we can intervene in the geomorphic system from the outside and then adjust the dynamic balance of the landscape in a relatively fast cycle [51], which has an important contribution to supporting the sustainable development planning and landscape management decision-making of landscape planning [29, 13, 36, 15, 60, 18].

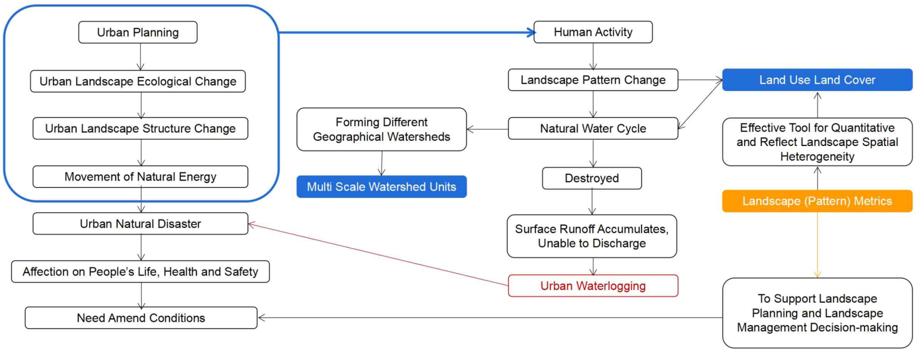

Figure 1. Human activities lead to the change of landscape patterns, which affects and destroys the process of the natural hydrological cycle, resulting in the change of surface characteristics. This change makes the runoff unable to be discharged in time and accumulated on the surface, resulting in urban waterlogging. Therefore, urban planning needs to be constantly adjusted and optimized to adapt to the development of urbanization.

1.4 Natural Water Cycle and Urban Waterlogging

Landscape ecology assumes that the arrangement and combination of spatial patterns in the ecosystem is the response to the ecological environment or ecological processes such as topography, temperature, and natural resource flow. At the same time, the strong changes in LULC caused by the demand for human production activities further strengthen the impact and interference on the spatial combination [21]. It leads to the destruction of the circulation and flow of natural material and energy, disrupts the natural hydrological cycle, and reduces the natural drainage capacity [67]. There are three types of common urban hydrological disasters: urban flood, urban rainstorm, and urban waterlogging. Urban flood refers to the overflow of river water onto normally dry land due to a rainstorm. Urban stormwater refers to rainwater and substances carried by rainwater, which may be leaves or grease on asphalt pavement. Urban waterlogging mainly refers to the phenomenon that human activities lead to LULC change and destroy the natural hydrological cycle. When there is a rainstorm or continuous rainfall, the surface runoff cannot be discharged in time, which gives rise to the accumulation of runoff on the surface [65, 68, 70]. Urban waterlogging is a systematic problem caused by spatial heterogeneity, thus we should examine this phenomenon from the perspective of landscape ecology [37]. The landscape pattern metrics can integrate various characteristics of LULC change, which have an important impact on urban waterlogging [70]. Therefore, landscape pattern metrics are considered to play an important role in rainwater management [31]. Many researchers have focused on the impact of the interaction between landscape patterns and watershed hydrological process on urban waterlogging disasters [64], which ignored important interactions between different landscape metrics.

1.5 Comparison with Previous Studies

Because land is the key center of landscape ecology, the role of its vegetation nature cannot be ignored [12, 72, 48]. In previous studies related to waterlogging, researchers mostly focuse on the impact of the expanded impervious surface on the occurrence of waterlogging events in the process of urban construction [67, 52]. They ignore the nature of "landscape" as a spatial body containing all ecological environments in the human living environment. According to the research object and the key variable of human intervention and control, it is not enough to simply take the impervious property as the research basis. At the same time, some researchers focus on green infrastructure, ignoring gray infrastructure such as roads [70]. Research objects include the administrative unit boundary as the scope benchmark [53], the grid units of different scales [68, 71], and the watershed scale [31]. Among them, the watershed has the function of collecting and storing rainwater, melting snow, and discharging water as runoff. At the same time, the change in land cover and land use within the watershed scale has a great impact on the catchment and subsequent runoff. It is worth noting that the size of the watershed scale plays an important role in this impact [5]. Therefore, the comparative analysis of the watershed scale is also essential, but this is usually missing in previous studies. On the other hand, it should be noted that generally, the study methods for waterlogging can be roughly divided into three categories: (1) The hydrological and hydrodynamic models or software. (2) Qualitative model based on expert knowledge. (3) Multivariate statistical method of machine learning [69], which provides more powerful data analysis tools for conducting regression. With the development of remote sensing technology, it is possible to study large-scale sites. Moreover, the development of theoretical research enables us to verify the reliability of experimental results. In previous studies, researchers have analyzed the relationship between landscape and waterlogging. However, they have neglected the interaction within the landscape structure [31, 53, 64, 67, 68, 70]. In other words, the "function" of one of the three main features of landscape ecology has been neglected (the other two are "structure" and "change") [24]. This is because the traditional linear regression model (an instance of supervised learning) can not analyze the relationship between independent variables at the same time when analyzing the relationship between independent variables and dependent variables [39]. Previous studies usually used traditional supervised learning methods, such as Geographically Weighted Regression (GWR). The influence of specific landscape structures on waterlogging is determined by the absolute value of estimated linear regression model parameters. To adapt to the heterogeneity of landscape space, GWR extends the geographical location-based linear regression model (ordinary least square method). However, due to the limitations of the linear regression model, the correlation between independent variables is intrinsically ignored. In recent years, explainable machine learning models have been widely developed in the field of machine learning, trying to determine the importance of the influence of independent variables on research objectives. This enables us to discover new theoretical explorations. Among them, Shapley Additive Explanations (SHAP) is a new analysis method based on Game Theory, which not only analyzes the relationship between independent variables and dependent variables, but also considers the correlation between multiple independent variables, and ranks the importance according to the contribution degrees. Finally, in order to adjust the ecological balance to mitigate the occurrence of waterlogging events, we also need to calculate threshold values. This is because the existence of landscape structure threshold and the complex feedback response in landscape and geomorphology have played an important role in the process of landscape evolution.

General speaking, in this study, I use the waterlogging area data to calculate the waterlogging density in the watershed unit as the waterlogging degree data. Through the comparison between the supervised learning model and the explainable machine learning model, it is observed whether the landscape pattern structure affecting waterlogging is different when the interaction between landscape patterns is ignored or fully considered. At the same time, considering that we need to objectively examine the ecological process, I further analyze the impact of landscape patterns on waterlogging under different scale watershed units from a macro perspective and select the watershed scale most suitable for landscape planning reference through the comparison of accurate values. Finally, I estimate threshold values of landscape pattern metrics that promote or alleviate waterlogging, which will provide a clearer reference for steak-holders and decision-makers. Therefore, in this study, I conduct research on the following three questions:

Question 1: When we fully consider the interaction between landscape structures, can we reflect the impact of landscape patterns on waterlogging more accurately? Question 2: Considering the geographical watershed division of different scales formed in the process of the water cycle, is it more suitable to be used as a site for waterlogging research if the watershed scale is smaller and finer? Question 3: Is there a threshold for the impact of specific landscape patterns on waterlogging?

1.6 Workflow and Study Area

To study above three research questions, our workflow (Fig. 2) first uses the waterlogging area data and the watershed units of three scales to calculate waterlogging density respectively, and filters out invalid data to obtain the waterlogging density data of three scales as the dependent variable. Then, the 41 kinds of LULC data maps classified with high precision are summarized into 7 kinds (Table 1), and the invalid data are also filtered and only valid data are used. At the same time, in view of the correlation between water cycle and landform, elevation and slope are taken as additional independent variables. Combining these variables with 7 landscape pattern metrics (Table 2), 45 independent variables are obtained in total. After completing the data processing process, I first select a scale watershed unit to compare the analysis results of Geographically Weighted Regression (GWR) and Shapley Additive Explanations (SHAP) methods, and use the Prediction Mean Squared Error (MSE) to compare the error values of the prediction results to decide which method can reflect the impact of landscape pattern on waterlogging more objectively and accurately. Then, I compare the analysis results of the three watershed scales and use the same MSE metric to compare the error values of the prediction results to decide which scale watershed unit is the most suitable for landscape planners as a reference for waterlogging research. Finally, I analyze the thresholds of specific landscape patterns with significant impact of waterlogging in the most suitable analysis results. This process enables us to explore whether new research theories and methods have presumed advantages, provide a more objective analysis scale for the guidance of macro landscape planning, and discover which characteristics of specific landscapes have an impact on waterlogging. Through the calculation of the threshold values, we can optimize and intervene in specific landscape characteristics in limited space and reduce the optimization cost of landscape ecology.

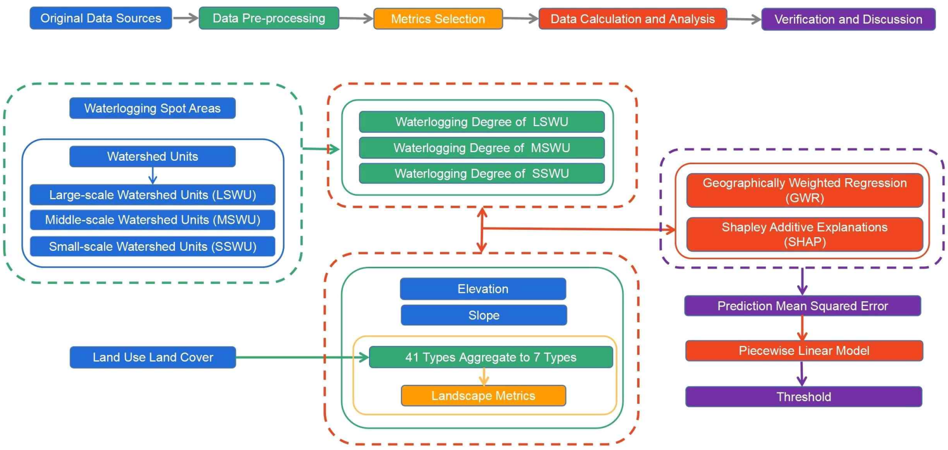

Figure 2. Workflow of this study.

The study area is Seoul Capital Area (SCA), it is the fifth-largest metropolitan area in the world, located in the northwest of the Republic of Korea which includes Seoul Special City, Incheon Metropolitan City, and Gyeonggi-do. The area is 12, 685 km^2 (Fig. 3). According to the Korea National Statistical Office (https://kostat.go.kr/), SCA has more than 26 million residents, accounting for 50.2% population of the Republic of Korea as of the year 2020. Based on the Korea Land and Geospatial Informatix's (The agency is under the Ministry of land, infrastructure and transport of the Republic of Korea) public data (https://www.data.go.kr/data/15048628/fileData.do), the waterlogging degree of SCA is more serious than other regions.

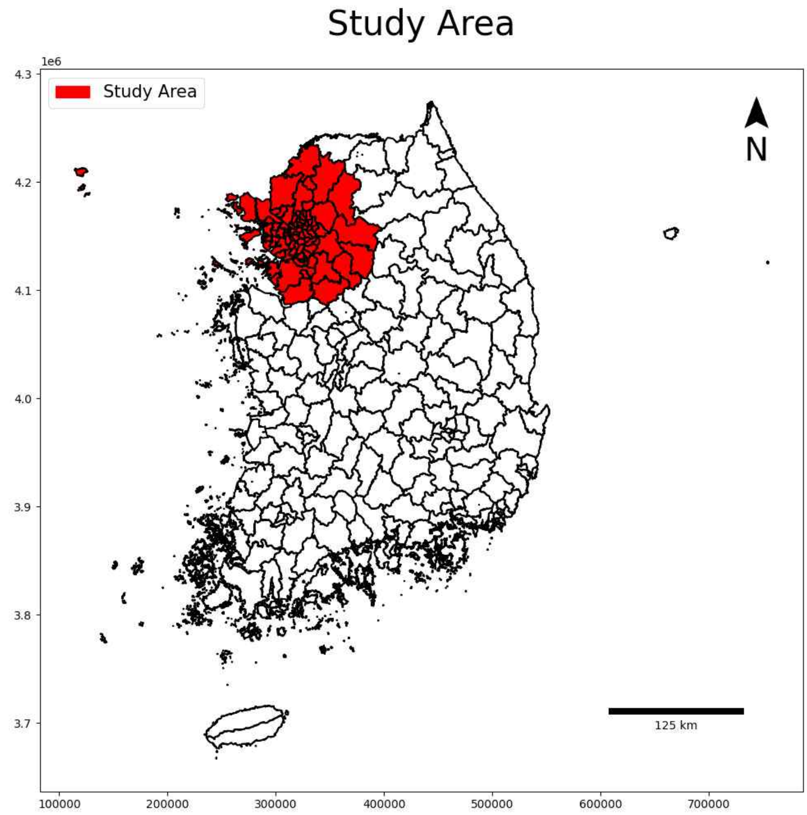

Figure 3. The Seoul Capital Area (SCA) is located in the northwest of the Republic of Korea, including Seoul Special City, Incheon Metropolitan City, and Gyeonggi-do ("do" means Province). The area is 12, 685 km^2, has more than 26 million residents, accounting for 50.2% population of the Republic of Korea as of the year 2020.

Chapter 2. Materials and Methods

2.1 Land Use Land Cover and Landscape Pattern Metrics

In this study, I use the Land Use Land Cover (LULC) map adapted from an original high-resolution satellite map. First of all, I need to process LULC data for later analysis of landscape pattern metrics. The LULC data form a 2020 sub-classified land cover map with 1-meter resolution and a total of 41 categories provided by the Korean Ministry of Environment, which are shapefiles formed as vector data. The data is obtained from aeronautical static satellite images (0.25-meter resolution), Arirang 2 satellite images (1-meter resolution), and Arirang 3 satellite images (0.7-meter resolution). I aggregate 41 categories into 7 categories: Used Area, Transportation Area, Agricultural Area, Forest, Grass, Barren, and Water (Table 1). The Used Area includes Detached Residential Facility, Common Residential Facility, Industrial Facility, Commercial and Business Facility, Commercial and Business Facility Mixed Area, Cultural, Sports, and Recreational Facility, Environmental Basic Facility, Educational and Administrative Facility, and Other Public Facility. The Transportation Area includes Airport, Port, Railroad, Road, and Other Transportation and Communication Facilities. The Agricultural Area includes Cultivated Paddy, Uncultivated Paddy, Cultivated Field, Uncultivated Field, Facility Cultivation Land, Orchard, Pastures and Farm, and Other Cultivation Land. The Forest includes Broad-leaved Forest, Coniferous Forest, and Mixed Forest. The Grass includes Natural Grassland, Golf Course, Cemetery, and Other Grassland. The Barren includes Beach, River Bank, Palisade and Rock, Mining Area, Sports Ground and Other Barren. At last, the Water includes Inland Wetland and Inland Water. Due to the extreme values appearing when analyzing the data containing outside land water, I exclude Coastal Wetland and Sea Water. In addition, considering that elevation and slope are also parts of the landscape structure, and their composition will also affect the process of the water cycle, I add these two kinds of data in addition to the LULC data.

Generally, in previous studies, the landscape pattern metrics were analyzed using the Fragstats software. When using this software for analysis, vector data needs to be converted into raster data such as Tag Image File Format (TIFF), which will reduce the accuracy of results. In order to overcome the loss of heterogeneity that occurs during the aggregation of pixel-based land cover maps, I need alternative solutions that can leverage interdisciplinary strengths [19]. Therefore, in this study, I use Python libraries to directly analyze vector data (shapefile), so the accuracy can be improved.

[Table 1] Aggregate 41 Land Use Land Cover (LULC) categories into 7 categories.

Coarse Classification	Detail Classification
Used Area	Detached Residential Facility
	Common Residential Facility
	Industrial Facility
	Commercial and Business Facility
	Mixed Area
	Cultural, Sports, and Recreational Facility
	Environmental Basic Facility
	Educational and Administrative Facility
	Other Public Facility
Transportation Area	Airport
	Port
	Railroad
	Road
	Other Transportation and Communication Facility
Agricultural Area	Cultivated Paddy
	Uncultivated Paddy
	Cultivated Field
	Uncultivated Field
	Facility Cultivation Land
	Orchard
	Pastures and Farm
	Other Cultivation Land
Forest	Broad-leaved Forest
	Coniferous Forest
	Mixed Forest
Grass	Natural Grassland
	Golf Course
	Cemetery
	Other Grassland
Water	Inland Wetland (Waterfront Vegetation)
	River
	Lake
Barren	Beach
	River Bank
	Palisade and Rock
	Mining Area
	Sports Ground
	Other Barren

As for the selection of landscape pattern metrics, I select those metrics which meet the following two conditions at the same time: (1) The landscape pattern metric is related to hydrological response [68, 53, 71]. (2) The landscape pattern metric can be analyzed by vector data. Therefore, in this study, I select 7 landscape pattern metrics (Table 2). They are Percentage of Landscape (PLAND), Degree of Landscape Division (D), Large Patch Index (LPI), Landscape Shape Index (LSI), Patch Density (PD), Edge Density (ED), and Shannon Diversity Index (SHDI).

The Percentage of Landscape (PLAND) is the ratio of the area of a particular patch to the total area of the regional landscape [66]. It quantifies the proportional abundance of each patch type in the landscape, which is one of the metrics to measure the composition of the landscape [46]. Because of the characteristics of its relative measurement, it is not affected by spatial distribution or configuration. PLAND equals the sum of the areas (square meter) of all patches of the corresponding patch type, divided by the total landscape area (square meter), then multiplied by 100 (convert to percentage). The total landscape area (A) includes any internal background present [8].

P_i =proportion of the landscape occupied by patch type (class).

i. a_{ij} = area (m^2) of patch ij.

A = total landscape (m^2).

$$PLAND = P_i = \frac{\sum_{j=1}^{n} a_{ij}}{A}(100)$$

Jaeger [29] proposed the Degree of Landscape Division (D), which reflects the degree of patch dispersion in the same landscape type. When D equals to 0 it means there is only one patch type [53, 64, 70]. It describes the artificial infiltration of the landscape from a geometric point of view and it is calculated according to the cumulative distribution function of patch size. D represents the probability that two randomly selected places in the investigated landscape are not located in the same undissected area [8]. D equals to the sum of 1 minus the patch area (square meter) divided by the total landscape area (square meter), then summed across all patches of the corresponding patch type. Total landscape area (A) includes any internal background present [46].

$a_{ij} =$ area (m^2) of patch ij.

$A =$ total landscape (m^2).

$$D = \left[1 - \sum_{j=1}^{n} \left(\frac{a_{ij}}{A}\right)^2\right]$$

Largest Patch Index (LPI) is the percentage of the largest patch (also dominance landscape type) in the total landscape area [71, 66, 64]. LPI equals the area (square meter) of the largest patch of the corresponding patch type divided by total landscape area (square meter). Total landscape area (A) includes any internal background present [46].

$a_{ij} =$ area (m^2) of patch ij.

$A =$ total landscape (m^2).

$$LPI = \frac{\max\limits_{j=1}^{n}(a_{ij})}{A}(100)$$

Landscape Shape Index (LSI) is the shape index which measures the shape complexity [71]. LSI provides the total edge (edge density) of the amount of standardization. The exponential formula is equal to .25 (adjustment for raster format) times the sum of the entire landscape boundary and all edge segments (meter) within the landscape boundary involving the corresponding patch type, including some or all of those bordering backgrounds, then divided by the square root of the total landscape area (square meter) [46].

e_{ik}^{*} =total length (m) of edge in landscape between patch types (classes) i and k; includes the entire landscape boundary and some or all background edge segments involving class i.

$A=$ total landscape (m^2).

$$LSI=\frac{.25\sum_{k=1}^{m}e_{ik}^{*}}{\sqrt{A}}$$

Patch Density (PD) reflects the degree of fragmentation of the landscape and the degree of spatial heterogeneity of the landscape [66]. The higher the PD value, the greater the number of patches and the higher the degree of fragmentation [53, 71]. The calculation method is the number of patches of the corresponding patch type divided by the total landscape area (square meter), then multiplied by 10000 square meters (or convert to 100 hectares) [46].

$n_i=$ number of patches in the landscape of patch type (class) i.

$A=$ total landscape (m^2).

$$PD=\frac{n_i}{A}(10{,}000)(100)$$

Edge Density (ED) is the total length of the edges of a particular class per unit, while larger ED value means higher fragmentation [53, 31]. Being calculated according to LSI, ED equals the sum of the lengths (meter) of all edge segments involving the corresponding patch type, divided by the total landscape area (square meter), then multiplied by 10000 square meters (or convert to 100 hectares) [46].

e_{ik} =total length (m) of edge in landscape involving patch type (class) i; includes landscape boundary and background segments involving patch type i.

A = total landscape (m^2).

$$ED = \frac{\sum_{k=1}^{m} e_{ik}}{A}(10{,}000)$$

Different from above selected landscape pattern metrics (PLAND, D, LPI, LSI, PD, ED), the Shannon Diversity Index (SHDI) is not a class metric, but a landscape-level metric to describe the number of landscape elements and the change in their proportions [53]. A larger SHDI value means higher diversity (richness) of the whole landscape [64]. Conversely, when the whole landscape is composed of a single element, its SHDI value becomes 0 [66]. It can be calculated according to PLAND, equals to the sum of subtracting the proportional abundance of each patch type in all patch types multiplied by this proportion, and Pi is based on the total landscape area (A) [46].

P_i = proportion of the landscape occupied by patch type (class) i.

$$SHDI = -\sum_{i=1}^{m}(P_i \ln P_i)$$

[Table 2] Landscape Pattern Metrics.

Landscape Pattern Metrics	Level	Units	Range
Percentage of Landscape (PLAND)	Class	Percentage	$0 < \text{PLAND} \leq 100$
Degree of Landscape Division (D)	Class	Proportion	$0 \leq D \leq 1$
Largest Patch Index (LPI)	Class	Percentage	$0 \leq \text{LPI} \leq 100$
Landscape Shape Index (LSI)	Class	None	$\text{LSI} \geq 1$, without limit
Patch Density (PD)	Class	Number per 100 Hectares	$\text{PD} > 0$
Edge Density (ED)	Class	Meters per Hectare	$\text{ED} \geq 0$, without limit
Shannon Diversity Index (SHDI)	Landscape	Information	$\text{SHDI} \geq 0$. without limit

In addition to landscape pattern metrics, I also add elevation and slope data (Fig. 4-6) as additional independent variables to improve the accuracy of our analysis of landscape elements affecting waterlogging. The elevation and slope data are obtained from the ALOS World 3D - 30m (AW3D30), which is a global digital surface model (DSM) with 30 m resolution provided by the JAXA Earth Observation Research Center taken from 2011. The dataset is based on the DSM dataset (5-meter mesh version) of the World 3D Topographic Data. I achieve the data through Google Earth Engine API.

To summarize, I aggregate LULC into 7 categories and use 7 types of landscape pattern metric tools for measurement. Among them, there are 6 class-level landscape pattern metrics (PLAND, D, LPI, LSI, PD, ED), which can measure each type of LULC. A landscape-level landscape pattern metric (SHDI) measures the whole LULC. At the same time, I also add the data of elevation and slope. Therefore, there are totally 45 independent variables in this study.

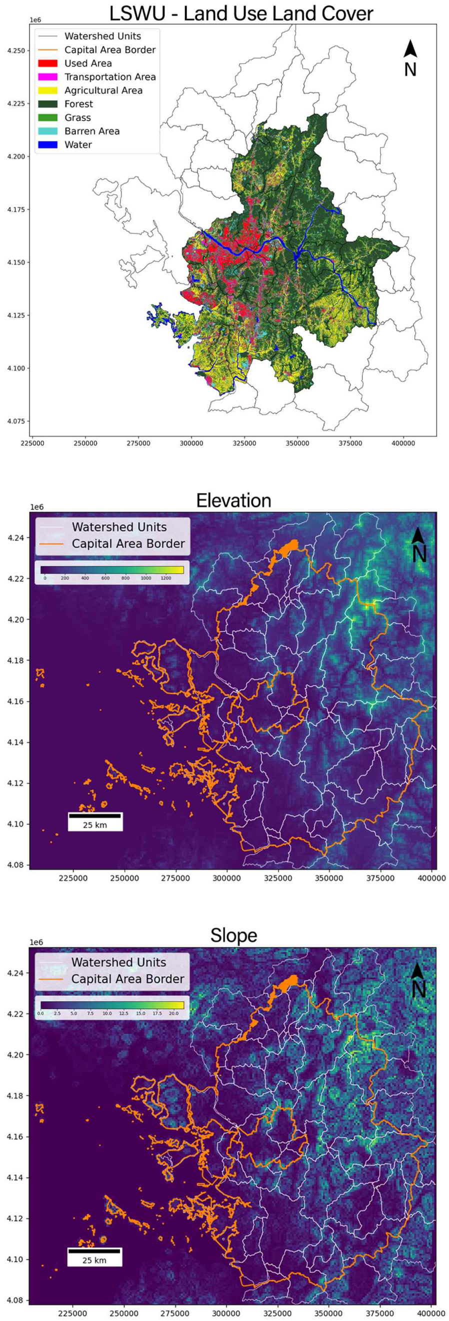

Figure 4. Effective land use land cover (LULC), elevation and slope data map of large-scale watershed units.

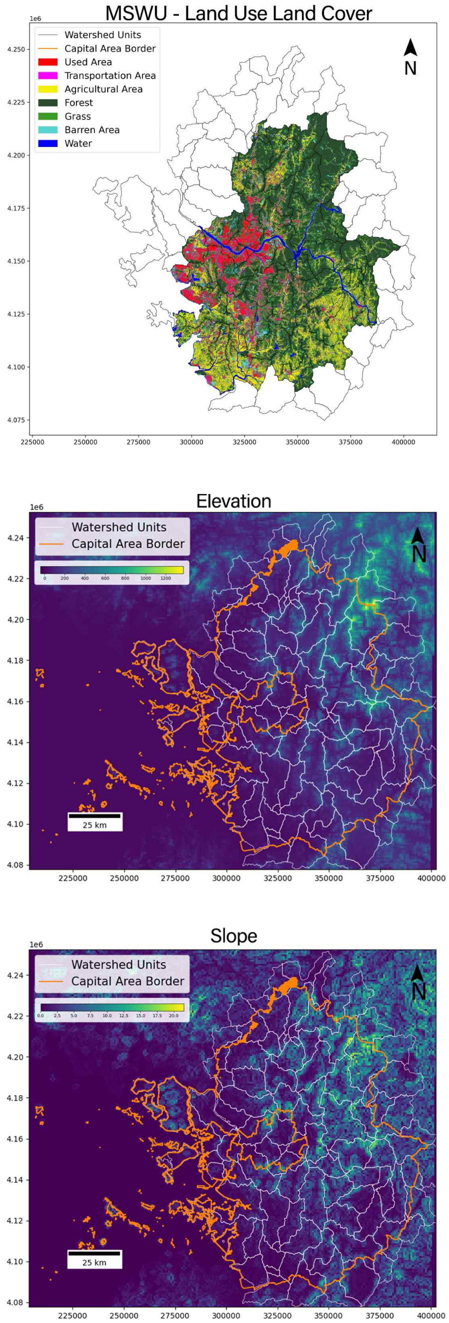

Figure 5. Effective land use land cover (LULC), elevation and slope data map of middle-scale watershed units.

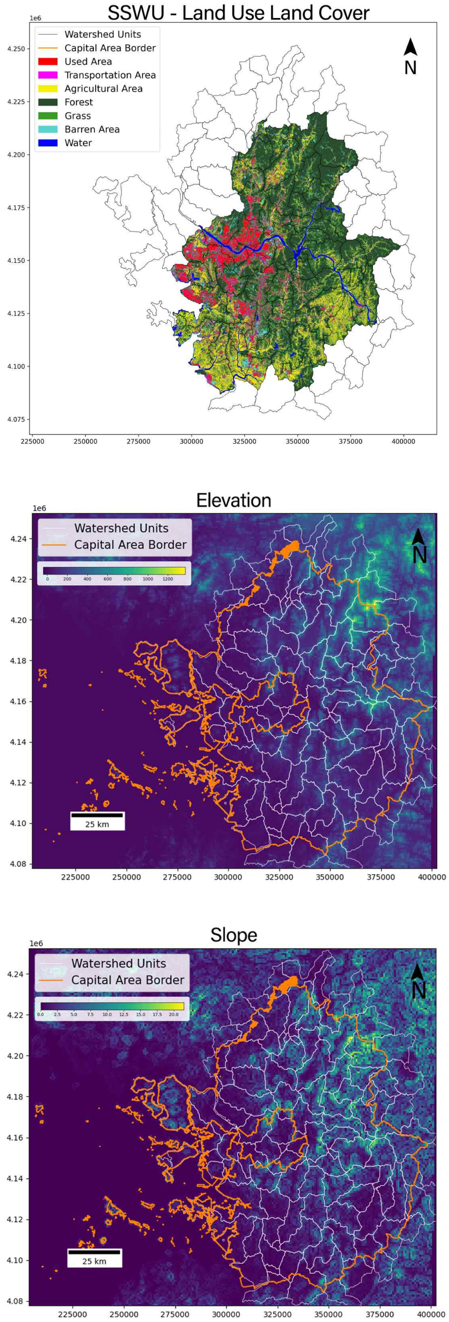

Figure 6. Effective land use land cover (LULC), elevation and slope data map of small-scale watershed units.

2.2 Waterlogging Degree of Watershed Units

Firstly, I assume that the degree of waterlogging can be correctly measured and expressed by the density of waterlogging area in the basin. The waterlogging area data are selected from the waterlogging trace shape files in vector format, which are provided by Korea Land and Geospatial Informatix Corporation in 2020. Then, for scale comparison to be conducted later, I select the data named HydroSHEDS which are developed by the World Wildlife Fund (WWF) conservation science program, U.S. Geological Survey, the International Center for Tropical Agriculture, the nature conservation, and the center for environmental systems research of the University of Kassel, Germany together. HydroSHEDS is a mapping product that provides hydrographic information for regional and global-scale applications in a consistent format. It offers a suite of geo-referenced datasets (vector and raster) at various scales, including river networks, watershed boundaries, drainage directions, and flow accumulations. HydroSHEDS is based on elevation data obtained in 2000 by NASA's Shuttle Radar Topography Mission (SRTM). The resolution is 15 arc second resolution (450m), obtained through Google Earth Engine. Since the smallest scale watershed unit in the Republic of Korea is HYBAS 11, I choose HYBAS 11, HYBAS 10, and HYBAS 9 as the Small-scale Watershed Units (SSWU), Middele-scale Watershed Units (MSWU), and Large-scale Watershed Units (LSWU) for comparative analysis. There are 111 (101, 64) watershed units in SSWU (MSWU, LSWU, respectively) that coincide with the SCA administrative boundary. In order to make the analysis results more accurate, I have removed some parts containing incomplete data. To this end, when selecting each scale watershed unit, I set the following two conditions: (1) The watershed unit should coincide with the SCA administrative boundary more than 90%, and (2) the LULC data should coincide with the watershed unit more than 90% (as the SCA is close the DMZ region, the data of some areas are not public). After screening (Table 3), I end up to actually use 62 (55, 31) watershed units (Table 3) on SSWU (MSWU, LSWU). Finally, the waterlogging area density of each unit under each scale watershed is calculated according to the waterlogging spot area data, and the waterlogging degree is obtained as the dependant variables (Fig. 7-9).

[Table 3] Number of units and effective units at each watershed scale.

Scale of Watershed Units	Number of Watershed Units	Number of Effective Watershed Units
Large-scale Watershed Units (LSWU)	64	31
Middle-scale Watershed Units (MSWU)	101	55
Small-scale Watershed Units (SSWU)	111	62

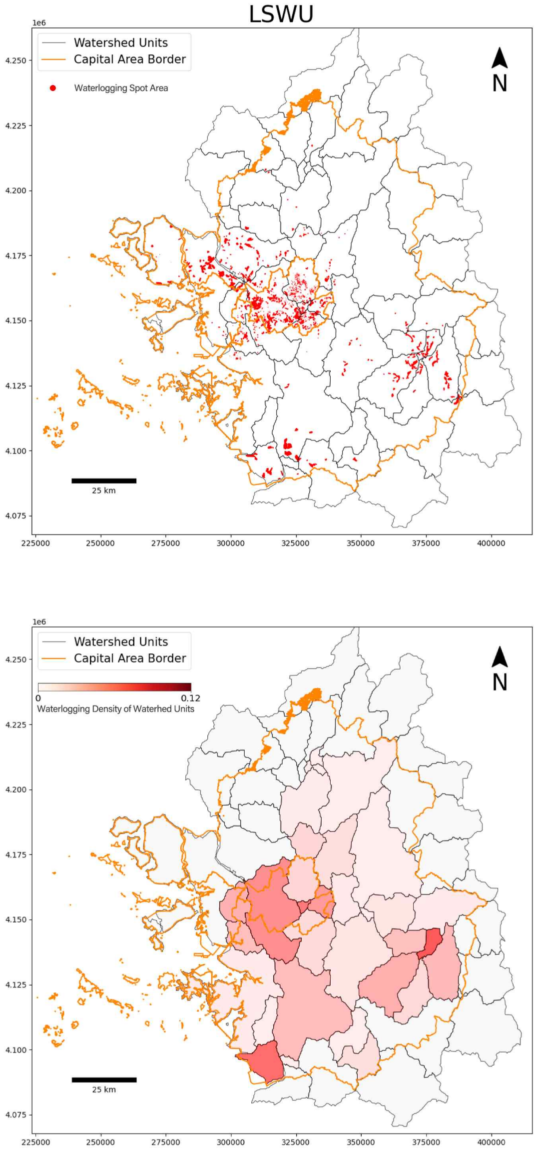

Figure 7. The waterlogging area (red dots) are used to calculate the waterlogging density of large-scale watershed units as the waterlogging degree data (gray areas are invalid watershed units).

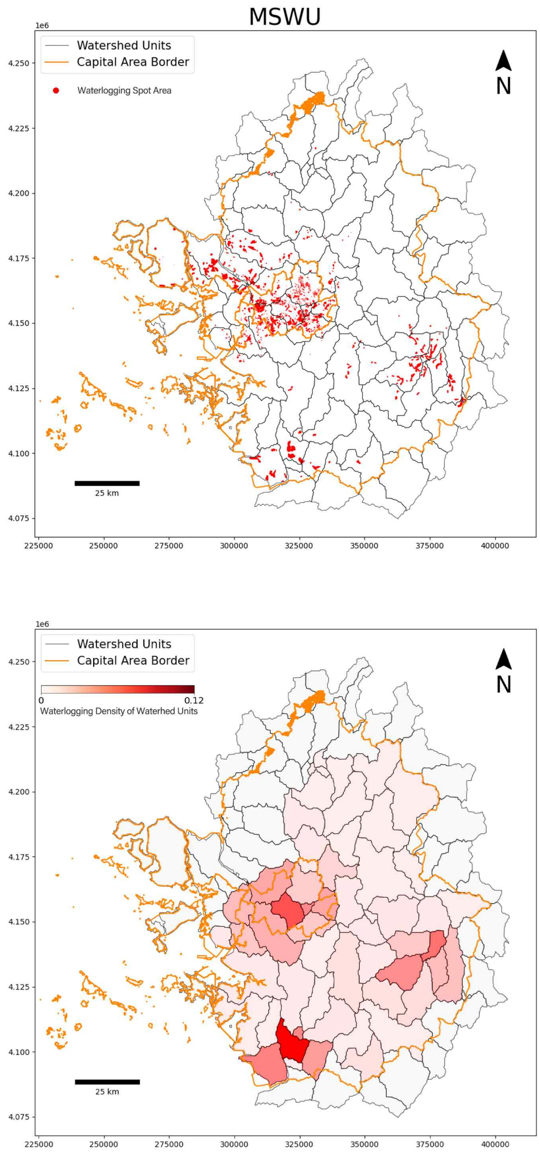

Figure 8. The waterlogging area (red dots) are used to calculate the waterlogging density of middle-scale watershed units as the waterlogging degree data (gray areas are invalid watershed units).

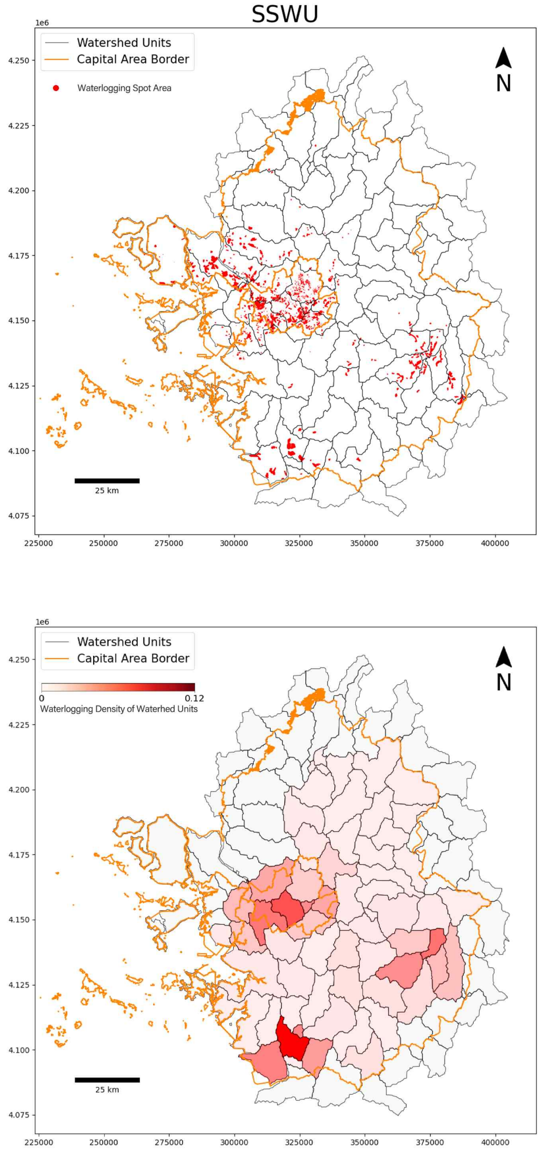

Figure 9. The waterlogging area (red dots) are used to calculate the waterlogging density of small-scale watershed units as the waterlogging degree data (gray areas are invalid watershed units).

2.3 Geographically Weighted Regression (GWR)

The linear regression analysis has long been used by quantitative geographers as a general technique for investigating associations between geographic variables, and the technique is widely used in the research field. However, I would like to note the surprising fact that the technique itself does not consider geographic position when analyzing the relationships between variables [6, 67]. This gap can be addressed by geographically weighted regression (GWR), which is a principled extension of ordinary linear regression that takes into account the geographic information of the variables. GWR was introduced to the geography literature [6] to study the potential for relationships in a regression model to vary in geographical space, or what is termed parametric non-stationarity. GWR is based on the non-parametric technique of locally weighted regression developed in statistics for curve-fitting and smoothing applications, where local regression parameters are estimated using subsets of data relatively close to a model estimation point in variable space. The innovation with GWR is using a subset of data proximate to the model calibration location in geographical space instead of variable space. While the emphasis in traditional locally weighted regression in statistics has been on curve-fitting, that is estimating or predicting the response variable, GWR has been presented as a method to conduct inference on spatially varying relationships, in an attempt to extend the original emphasis on prediction to confirmatory analysis [63].

Specifically, GWR is generally used to describe the spatial variance and explain the effect of the independent variables on the dependent variables [28]. The spatial non-stationarity of the parameters in different spaces is reflected by estimating them for different regions so that the relationship between the variables can change with spatial position [17]. By introducing the spatial positions of the data into the regression coefficients, non-parametric estimation methods can be used to provide local estimates of each geographic position function [40, 47]. The regression relationships are mainly explored and analyzed based on the variation of regression coefficients for each geographic position with space. Therefore many researchers believe that the results of using GWR models to analyze the spatial relationships of different factors are more in line with the objective reality [43, 64].

In the case of this study, GWR essentially repeats ordinary linear regression for each watershed unit, with specific weights which are larger for other watershed units in the neighborhood and smaller for other watershed units in distance. A Gaussian kernel function is usually used for calculating weights and is adopted in our approach. Finally, for the important factor of each feature, I take the absolute value of the mean value of coefficients calculated regarding each watershed unit.

In order to formally define the model for GWR, I first start from a simple linear regression model for each data point (X_i, y_i).

$$y_i = \sum_{j=1}^{d} X_{ij}\beta_j + \epsilon_{i,}$$

where X is the matrix containing independent variables, y is the vector containing dependent variables, d is the dimension of each independent data point, β is the coefficient vector and ε is the noise vector which is assumed to follow a latent normal distribution. In this case, the coefficient vector β can be analytically solved as

$$\hat{\beta} = (X^{\top} X)^{-1} X^{\top} y.$$

I would like to note that this coefficient vector obtained by the traditional linear regression method is agnostic about relative geographic information. The model is calculated in a way that every data point has the same importance to each of other data points. Therefore, in order to address the difference caused by relative geographical positions, GWR introduced a weight matrix which is a diagonal square matrix of size N, the number of data points. This is to say, for each data point $(X_i,\ y_i)$, there is an individual coefficient vector that follows

$$\hat{\beta_i} = (X^{\top} W_i X)^{-1} X^{\top} W_i y,$$

where the $k-th$ dignal element of W_i denotes the distance between X_i and X_k. This distance is further defined using the Gaussian kernel function

$$\exp\left(-\frac{d_{ik}^2}{2h^2}\right),$$

where d_{ik} is the Euclidean distance between the centroids of two watershed units measured in meter, and h is the bandwidth parameter and is set to 10, 000 throughout this thesis. In the end, I achieve N different coefficient vectors, and use the average prediction for new unseen data points.

In our analysis for GWR, I conducted calculation as defined. For the bandwidth of the Gaussian kernel function measuring distance, I used 10, 000 that is derived from the median value of distances, which is a common heuristic method for deciding bandwidth value. Before conducting linear regression, I apply a whitening transformation on all independent variables and dependent variables to improve the stability of GWR. For conducting linear regression, I use the Python library scikit-learn.

2.4 Shapley Additive Explanation (SHAP)

Shapley Additive Explanations (SHAP) is a game-theoretic quantification method for investigating variable importance, which also considers the potential cooperation, namely interactions, among variables. Although having a combinational complexity, significant developments have been recently made to calculate Shapley values in an efficient way [42], which can be easily used off-the-shelf by a Python library and is extensively used in various research areas.

Formally, for a data point X_i SHAP value for the importance of the j-th feature is defined as

$$SHAP_j(f, X_i) = \sum_{z \subset x} \frac{|z|!(d-|z|-1)!}{d!}(f(z) - f(z \setminus j)),$$

where d is the dimension of X_i, $z \subset x$ means all possible combinations of features in X_i represented by $z \in \{0,1\}d, |z|$ is the number of non-zero entries in z and $z \backslash j$ means removing j-th entry from z. My implementation mainly used the Python library xgboost and shap. I used the standard XGBRegressor as the base regressor, with objective set as squared error. Figures of bar chart and beeswarm plot is generated by functions provided by shap.

2.5 Prediction Mean Squared Error (MSE)

Loss functions measure the extent to which model prediction values differ from desired target values. There are many different designs of loss functions for different problem settings, such as classification, regression, ranking, etc. Because regression problem is the focus of this thesis, I adopt to used the famous mean squared error (MSE) as

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(p_i - y_i)^2,$$

where pi denotes the model prediction value for each data points. MSE is first proposed by Gauss and enjoys flourish theoretical and practical advances.

In order to analyze and compare which method between GWR and SHAP is more advanced, thus which can more accurately reflect the landscape pattern analysis that affects waterlogging. I used the Mean Squared Error method for prediction analysis. The analysis method is to randomly select 90 percent of the watershed units in each scale and predict the remaining 10 percent of the results, which is a commonly used analysis accuracy method in machine learning. Therefore, I randomly select 90% watershed units at SSWU, MSWU, and LSWU scales, respectively (55, 49, 27 watershed units, Table 4). Then, I use GWR and SHAP methods to predict the impact of landscape pattern on waterlogging in the remaining 10% watershed units (7, 6, 4 watershed units), and obtain the mean squared error value. Finally, I can distinguish which method and scale watershed are the most accurate for analyzing the impact of landscape patterns on waterlogging.

[Table 4] Number of watershed Units for Prediction Mean Squared Error (MSE).

Scale of Watershed Units	Number of Utilized Watershed Units	Number of Predicted Watershed Units
Large-scale Watershed Units (LSWU)	27	4
Middle-scale Watershed Units (MSWU)	49	6
Small-scale Watershed Units (SSWU)	55	7

2.6 Piecewise Linear Model

Piecewise linear model has a long history in various fields of science, and deep neural network is yet another recent example. In this thesis, I use the piecewise linear models with only two parts, namely its simplest form, to study the threshold of impact.

Formally, it is defined as

$$y = \begin{cases} k_1 x + y_0 - k_1 \tau \; (x \le \tau), \\ k_2 x + y_0 - k_2 \tau \; (x > \tau), \end{cases}$$

where τ is the threshold value for the independent variable and k_1, k_2, y_0 are model parameters.

I implemented the above model in Python using the numpy library and solve it using the scipy library.

Chapter 3. Results

I calculate the correlation among 45 independent variables and 3 dependent variables (corresponding to 3 scales) by using GWR and shake methods to know the impact of landscape patterns on waterlogging and obtain the ranking of the impact degree of 45 independent variables on each dependent variable. The results are sorted by the absolute values of the degree of influence. That is to say, whether the landscape pattern has a positive impact on waterlogging (promoting waterlogging) or a negative impact (alleviating waterlogging), and larger absolute values means greater impact on waterlogging. It should be noted that to accurately analyze which land characteristics and what reasons affect the occurrence of waterlogging, although I have up to 45 independent variables, in order to explain the analysis results more concisely, I will focus on explaining the top 10 independent variables of each analysis result.

3.1 Geographically Weighted Regression (GWR)

By using the Geographically Weighted Regression (GWR) method, I can obtain the absolute values of the impact of different characteristics of specific landscape patterns on waterlogging, the ranking of results (Fig. 10-12), and detailed values (Tables 5-7). According to the ranking of absolute values, I can observe which characteristics of specific landscape patterns have a great impact on waterlogging more intuitively. The detailed values help understand whether the specific characteristics of these specific landscape patterns have a positive impact on waterlogging (promote waterlogging) or a negative impact on waterlogging (alleviate waterlogging).

According to Fig. 10, the top 10 landscape pattern metrics that have the greatest impact on waterlogging in LSWU (Large-scale Watershed Units) are as follows: Grass PLAND (Percentage of Landscape), Transportation Area D (Degree of Landscape Division), Agriculture Area LPI (Largest Patch Index), Agriculture Area LSI (Largest Shape Index), Elevation, Grass D, Water ED (Edge Density), Water LSI, Barren Area D, Forest LPI. The detailed value are as follows (Table 5): Grass PLAND (-3.20×10^{13}), Transportation Area D (-2.79×10^{13}), Agriculture Area LPI (-2.66×10^{13}), Agriculture Area LSI (-2.31×10^{13}), Elevation (-2.24×10^{13}), Grass D (-1.63×10^{13}), Water ED (1.49×10^{13}), Water LSI (1.44×10^{13}), Barren Area D (-1.26×10^{13}), Forest LPI (-1.24×10^{13}).

According to Fig. 11, the top 10 landscape pattern metrics that have the greatest impact on waterlogging in MSWU (Middle-scale Watershed Units) are as follows: Water ED (Edge Density), Agriculture Area ED, Forest PLAND (Percentage of Landscape), Transportation Area LPI (Largest Patch Index), Grass LSI (Largest Shape Index), Water PD (Path Density), Agricultural Area PD, Grass ED, Transportation Area PLAND, Barren Area LSI. The detailed value are as follows (Table 6): Water ED (1.04×10^{13}), Agriculture Area ED (8.51×10^{12}), Forest PLAND (7.72×10^{12}), Transportation Area LPI (6.52×10^{12}), Grass LSI (-5.47×10^{12}), Water PD (-5.34×10^{12}), Agricultural Area PD (-4.86×10^{12}), Grass ED (4.70×10^{12}), Transportation Area PLAND (-4.32×10^{12}), Barren Area LSI (4.12×10^{12}).

According to Fig. 12, the top 10 landscape pattern metrics that have the greatest impact on waterlogging in SSWU (Small-scale Watershed Units) are as follows: Barren Area LPI (Largest Patch Index), Barren Area D (Degree of Landscape Division), Transportation Area ED (Edge Density), Transportation Area LPI, Used Area ED, Agricultural Area LSI (Largest Shape Index), Barren Area LSI, Barren Area ED, Transportation Area LSI, Forest LSI. The detailed value are as follows (Table 7): Barren Area LPI (-6.23×10^{12}), Barren Area D (-5.89×10^{12}), Transportation Area ED (5.27×10^{12}), Transportation Area LPI (-3.79×10^{12}), Used Area ED (-3.76×10^{12}), Agricultural Area LSI (3.51×10^{12}), Barren Area LSI (3.50×10^{12}), Barren Area ED (-3.35×10^{12}), Transportation Area LSI (-3.24×10^{12}), Forest LSI (-2.55×10^{12}).

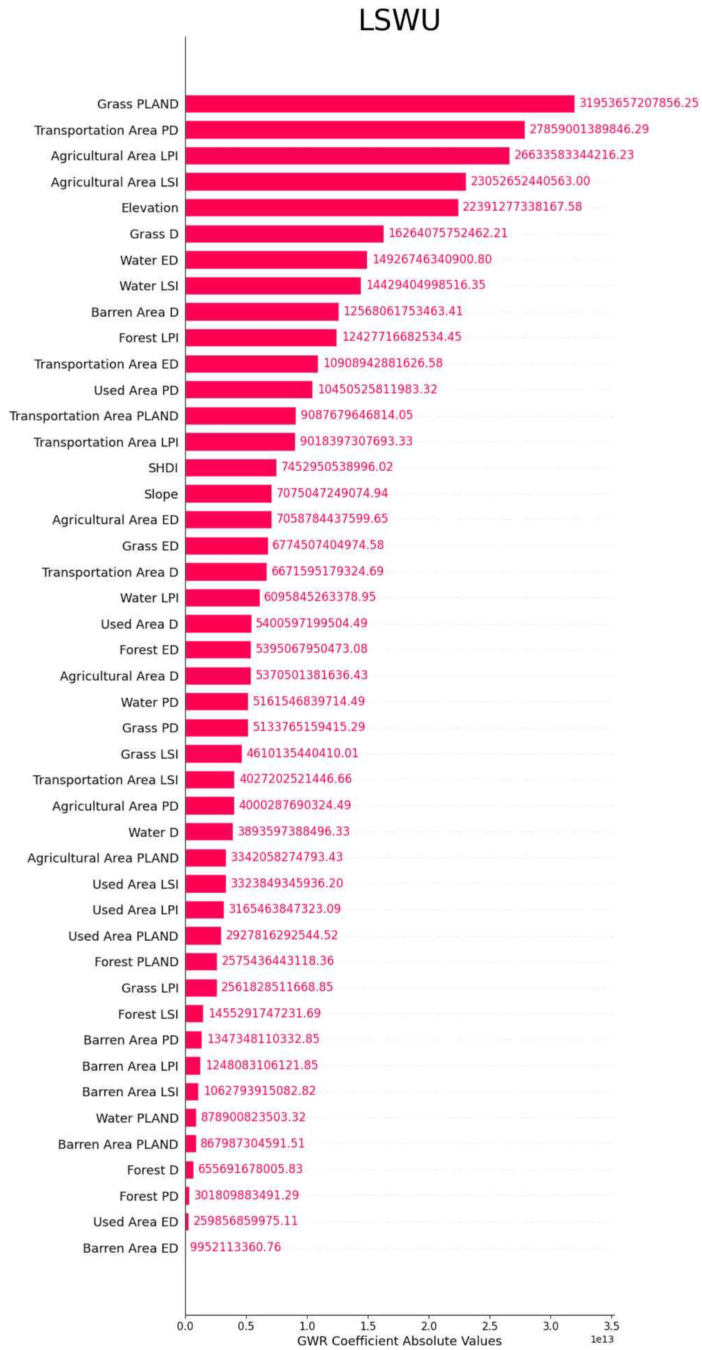

Figure 10. The correlation degree between landscape pattern metrics and large-scale watershed units (LSWU) by geographically weighted regression (GWR).

[Table 5] The values of landscape pattern metrics of Large-scale Watershed Units (LSWU) analyzed by Geographically Weighted Regression (GWR).

Landscape Category	Value
Grass PLAND	-3.20×10^{13}
Transportation Area PD	-2.79×10^{13}
Agricultural Area LPI	-2.66×10^{13}
Agricultural Area LSI	-2.31×10^{13}
Elevation	-2.24×10^{13}
Grass D	-1.63×10^{13}
Water ED	1.49×10^{13}
Water LSI	1.44×10^{13}
Barren Area D	-1.26×10^{13}
Forest LPI	-1.24×10^{13}
Transportation Area ED	-1.09×10^{13}
Used Area PD	1.05×10^{13}
Transportation Area PLAND	9.09×10^{12}
Transportation Area LPI	9.02×10^{12}
SHDI	-7.45×10^{12}
Slope	7.08×10^{12}
Agricultural Area ED	7.06×10^{12}
Grass ED	6.77×10^{12}
Transportation Area D	6.67×10^{12}
Water LPI	-6.10×10^{12}
Used Area D	5.40×10^{12}
Forest ED	-5.40×10^{12}
Agricultural Area D	-5.37×10^{12}
Water PD	-5.16×10^{12}
Grass PD	-5.13×10^{12}
Grass LSI	4.61×10^{12}
Transportation Area LSI	-4.03×10^{12}
Agricultural Area PD	4.00×10^{12}
Water D	3.89×10^{12}
Agricultural Area PLAND	3.34×10^{12}
Used Area LSI	3.32×10^{12}
Used Area LPI	-3.17×10^{12}
Used Area PLAND	-2.93×10^{12}
Forest PLAND	2.58×10^{12}
Grass LPI	2.56×10^{12}
Forest LSI	-1.46×10^{12}
Barren Area PD	1.35×10^{12}
Barren Area LPI	1.25×10^{12}

Barren Area LSI	1.06×10^{12}
Water PLAND	8.79×10^{11}
Barren Area PLAND	8.68×10^{11}
Forest D	-6.56×10^{11}
Forest PD	3.02×10^{11}
Used Area ED	2.60×10^{11}
Barren Area ED	-9.95×10^{9}

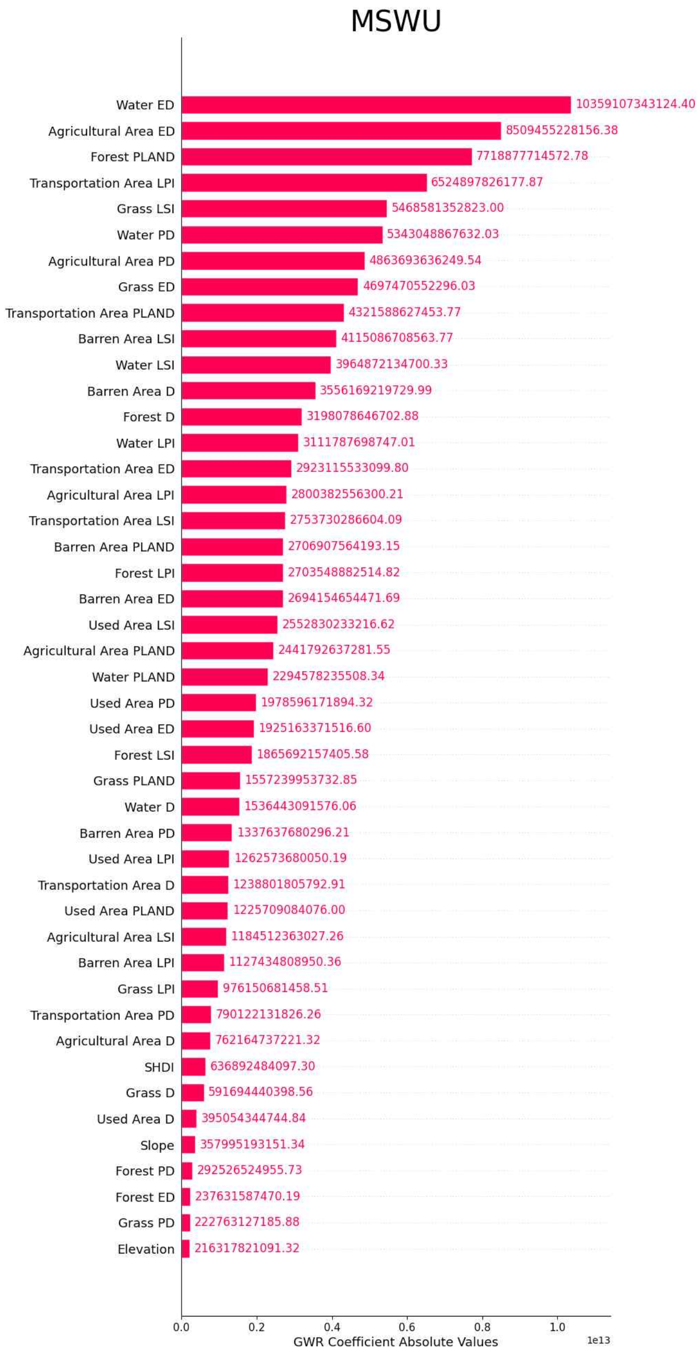

Figure 11. The correlation degree between landscape pattern metrics and middle-scale watershed units (MSWU) by geographically weighted regression (GWR).

[Table 6] The values of landscape pattern metrics of Middle-scale Watershed Units (MSWU) analyzed by Geographically Weighted Regression (GWR).

Landscape Category	Value
Water ED	1.04×10^{13}
Agricultural Area ED	8.51×10^{12}
Forest PLAND	7.72×10^{12}
Transportation Area LPI	6.52×10^{12}
Grass LSI	-5.47×10^{12}
Water PD	-5.34×10^{12}
Agricultural Area PD	-4.86×10^{12}
Grass ED	4.70×10^{12}
Transportation Area PLAND	-4.32×10^{12}
Barren Area LSI	4.12×10^{12}
Water LSI	-3.96×10^{12}
Barren Area D	-3.56×10^{12}
Forest D	-3.20×10^{12}
Water LPI	-3.11×10^{12}
Transportation Area ED	2.92×10^{12}
Agricultural Area LPI	-2.80×10^{12}
Transportation Area LSI	2.75×10^{12}
Barren Area PLAND	-2.71×10^{12}
Forest LPI	-2.70×10^{12}
Barren Area ED	-2.69×10^{12}
Used Area LSI	-2.55×10^{12}
Agricultural Area PLAND	2.44×10^{12}
Water PLAND	2.29×10^{12}
Used Area PD	1.98×10^{12}
Used Area ED	-1.93×10^{12}
Forest LSI	1.87×10^{12}
Grass PLAND	-1.56×10^{12}
Water D	-1.54×10^{12}
Barren Area PD	1.34×10^{12}
Used Area LPI	1.26×10^{12}
Transportation Area D	1.24×10^{12}
Used Area PLAND	1.23×10^{12}
Agricultural Area LSI	-1.18×10^{12}
Barren Area LPI	-1.13×10^{12}
Grass LPI	-9.76×10^{11}
Transportation Area PD	-7.90×10^{11}
Agricultural Area D	7.62×10^{11}
SHDI	6.37×10^{11}

Grass D	-5.92×10^{11}
Used Area D	3.95×10^{11}
Slope	3.58×10^{11}
Forest PD	-2.93×10^{11}
Forest ED	-2.38×10^{11}
Grass PD	2.23×10^{11}
Elevation	2.16×10^{11}

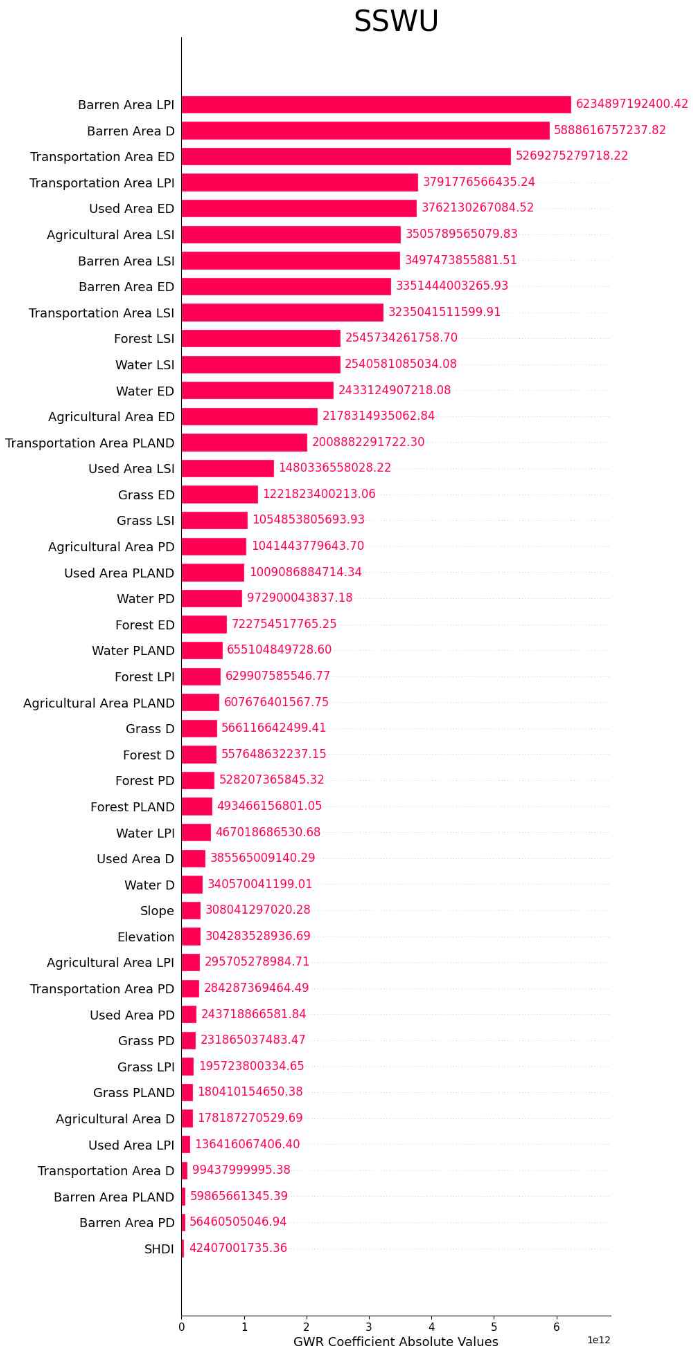

Figure 12. The correlation degree between landscape pattern metrics and small-scale watershed units (SSWU) by geographically weighted regression (GWR).

[Table 7] The values of landscape pattern metrics of Small-scale Watershed Units (SSWU) analyzed by Geographically Weighted Regression (GWR).

Landscape Category	Value
Barren Area LPI	-6.23×10^{12}
Barren Area D	-5.89×10^{12}
Transportation Area ED	5.27×10^{12}
Transportation Area LPI	-3.79×10^{12}
Used Area ED	-3.76×10^{12}
Agricultural Area LSI	3.51×10^{12}
Barren Area LSI	3.50×10^{12}
Barren Area ED	-3.35×10^{12}
Transportation Area LSI	-3.24×10^{12}
Forest LSI	-2.55×10^{12}
Water LSI	-2.54×10^{12}
Water ED	2.43×10^{12}
Agricultural Area ED	-2.18×10^{12}
Transportation Area PLAND	2.01×10^{12}
Used Area LSI	1.48×10^{12}
Grass ED	-1.22×10^{12}
Grass LSI	1.05×10^{12}
Agricultural Area PD	-1.04×10^{12}
Used Area PLAND	1.01×10^{12}
Water PD	9.73×10^{11}
Forest ED	7.23×10^{11}
Water PLAND	-6.55×10^{11}
Forest LPI	-6.30×10^{11}
Agricultural Area PLAND	-6.08×10^{11}
Grass D	5.66×10^{11}
Forest D	5.58×10^{11}
Forest PD	5.28×10^{11}
Forest PLAND	4.93×10^{11}
Water LPI	4.67×10^{11}
Used Area D	3.86×10^{11}
Water D	-3.41×10^{11}
Slope	-3.08×10^{11}
Elevation	-3.04×10^{11}
Agricultural Area LPI	2.96×10^{11}
Transportation Area PD	-2.84×10^{11}
Used Area PD	2.44×10^{11}
Grass PD	2.32×10^{11}
Grass LPI	1.96×10^{11}

Grass PLAND	-1.80×10^{11}
Agricultural Area D	1.78×10^{11}
Used Area LPI	-1.36×10^{11}
Transportation Area D	-9.94×10^{11}
Barren Area PLAND	-5.99×10^{11}
Barren Area PD	5.65×10^{11}
SHDI	-4.24×10^{11}

The results after GWR analysis are shown in the Figures 10-12 and Tables 5-7. To better interpret the results, I classify the top 10 landscape patterns with the highest degree of correlation with the impact of waterlogging according to the landscape category and use the absolute value superposition (based on the 13th power of 10), to sort according to the strength relationship.

In LSWU (Fig. 10), the order of the impact of landscape types on waterlogging is as follows: Agricultural Area (LPI & LSI, 4.97×10^{13}) > Grass (PLAND & D, 4.83×10^{13}) > Water (ED & LSI, 2.93×10^{13}) > Transportation Area (D, 2.79×10^{13}) > Elevation (2.24×10^{13}) > Barren Area (D, 1.26×10^{13}) > Forest (LPI, 1.24×10^{13}). Among them, the landscape properties of the Agricultural Area (LPI & LSI) and the Grass (PLAND & D) are the dominant factors that interfere with the water cycle balance. According to their characteristics, it is helpful to adjust the landscape ecological balance and alleviate waterlogging. As shown in Table 5, both Agricultural Area LPI (-2.66×10^{13}) and Agricultural Area LSI (-2.31×10^{13}) are less than 0 (negative impact), which have an effect on alleviating waterlogging. LPI is the largest patch index, which means that the agricultural area is the dominant landscape type in the watershed unit of this scale. Therefore, it is necessary to expand the agricultural area with the largest proportion of area to achieve the purpose of alleviating waterlogging. The LSI is the landscape shape index, which means that the simpler its edge shape is, the more beneficial it is to alleviate waterlogging. Therefore, in LSWU, we need to find the single patch of agricultural area with the largest one, expand it and simplify the complexity of mosaic edges with other landscape types. And the Grass PLAND (-3.20×10^{13}) and Grass D (-1.63×10^{13}) also play a role in alleviating waterlogging. PLAND is Percentage of Landscape, which means that the higher the proportion of Grass in the watershed unit of this scale, the greater the mitigation effect of waterlogging. D is the landscape division, which reflects the patch dispersion degree of a specific landscape type. Here, the higher the patch dispersion degree of Grass, the more beneficial it is to alleviate waterlogging. Therefore, according to the characteristics of Grass, we need fragmented dispersion increase to achieve our goal.

In MSWU (Fig. 11), the order of the impact of landscape types on waterlogging is as follows: Water (ED & PD, 1.57×10^{13}) > Agricultural Area (ED & PD, 1.34×10^{13}) > Transportation Area (LPI & PLAND, 1.08×10^{13}) > Grass (LSI & ED, 1.02×10^{13}) > Forest (PLAND, 0.77×10^{13}) > Barren Area (LSI, 0.41×10^{13}). Among them, the landscape property of Water (ED & PD, 1.57×10^{13}) has the strongest correlation with waterlogging, but it is similar to that of the subsequent Agricultural Area (ED & PD, 1.34×10^{13}) and Transportation Area (LPI & PLAND, 1.08×10^{13}). Therefore, at this scale, landscape planners need to consider 4 features of three landscape categories at the same time. ED is the edge density, which means the total length of the edge of a specific landscape type. The higher the value, the larger the fragment. PD is patch density, which reflects the degree of fragmentation and spatial heterogeneity of the landscape. The higher the value, the more patches. As shown in Table 6: Water ED (1.04×10^{13}), Water PD (-5.34×10^{12}), Agricultural Area ED (8.51×10^{12}), Agricultural Area PD (-4.86×10^{12}), Transportation Area LPI (6.52×10^{12}), Transportation Area PLAND (-4.32×10^{12}) combined with their values, if we want to achieve the purpose of alleviating waterlogging, we need to reduce the total length of the edge of Water area and Agricultural Area, and increase small water area and agricultural area, such as small-scale constructed wetlands and collective cultivated land in urban areas. At the same time, the Transportation Area, which accounts for the largest proportion, need be reduced, and the overall ratio of Transportation Area cannot be reduced at will because of social demand. This part needs to be compensated by other Agricultural Area adjustments.

In SSWU (Fig. 12), the order of the impact of landscape types on waterlogging is as follows: Barren Area (LPI & D & LSI & ED, 1.90×10^{13}) > Transportation Area (ED & LPI & LSI, 1.90×10^{13}) > Used Area (ED, 0.38×10^{13}) > Agricultural Area (LSI, 0.35×10^{13}) > Forest (LSI, 0.26×10^{13}). In SSWU, the comprehensive impact degree of Barren Area and Transportation Area is the same, and these two landscape types have the highest correlation with waterlogging. Although there are 4 features of the Barren Area, 3 features of the Transportation Area are related to waterlogging. However, because of the low result index of the two landscape types, they are not sensitive to the response of waterlogging. Their respective values are (Table 7): Barren Area LPI (-6.23×10^{12}), Barren Area D (-5.89×10^{12}), Barren Area LSI (3.50×10^{12}), Barren Area ED (-3.35×10^{12}), Transportation Area ED (5.27×10^{12}), Transportation Area LPI (-3.79×10^{12}), and Transportation Area LSI (-3.24×10^{12}). We can observe from the above data that this means that the largest single barren area should be expanded while fragmented barren area being increased. The simpler the mosaic structure of patches and other landscape categories, the better it is. While the edge length of the transportation area has a positive effect on waterlogging, the more complex the patch and shape of the largest single piece of the transportation area are, the greater the mitigation effect on waterlogging.

3.2 Shapley Additive Explanations (SHAP)

By using the Shapley Additive Explanations (SHAP) method, I can obtain the importance ranking of different characteristics of a specific landscape pattern on the impact of waterlogging (Fig. 13, 15, 17), as well as the detailed values. For the detailed values of SHAP (Fig. 14, 16, 18), the x-axis is based on the value of 0. If it is less than 0, it means the landscape type features have a negative impact on the contribution of waterlogging (to alleviate waterlogging). If it is greater than 0, it means the landscape type features have a positive impact on the contribution of waterlogging (to promote waterlogging). Each point on the X-axis represents a single watershed unit, and each watershed unit (point) represents the value of the landscape type feature in the watershed. Y-axis is the order of contribution degree of landscape type characteristics to waterlogging at the whole watershed scale. Considering the different values of each watershed unit (points on the X-axis), in order to comprehensively evaluate the characteristics of each landscape category, I reform the values to obtain Tables 9, 11, and 13.

According to Fig. 13, the top 10 landscape pattern metrics that have the greatest impact on waterlogging in LSWU (Large-scale Watershed Units) are as follows: Transportation Area D (Degree of Landscape Division), SHDI (Shannon Diversity Index), Transportation Area LPI (Largest Patch Index), Barren Area D, Barren Area LPI, Grass ED, Used Area LPI, Forest ED (Edge Density), Grass PLAND (Percentage of Landscape), Transportation Area PD (Path Density). The distribution of positive or negative (based on 0 value) waterlogging impact on watershed units by landscape pattern characteristics is shown in Figure 14. The detailed impact degrees and the number of watershed units affected by negative and positive waterlogging are shown in Table 8: Transportation Area D (2.21×10^{-1}, 24 : 7), SHDI (1.50×10^{-1}, 17 : 14), Transportation Area LPI (1.45×10^{-1}, 26 : 5), Barren Area D (1.43×10^{-1}, 26 : 5), Barren Area LPI (1.40×10^{-1}, 28 : 3), Grass ED (1.39×10^{-1}, 18 : 13), Used Area LPI (1.31×10^{-1}, 24 : 7), Forest ED (1.01×10^{-1}, 28 : 3), Grass PLAND (7.37×10^{-2}, 3 : 28), Transportation Area PD (6.19×10^{-1}, 26 : 5).

According to Fig. 15, the top 10 landscape pattern metrics that have the greatest impact on waterlogging in MSWU (Middle-scale Watershed Units) are as follows: Transportation Area PD (Path Density), Agricultural Area ED (Edge Density), Barren Area LPI (Largest Patch Index), Forest ED, Used Area LPI, Transportation Area D (Degree of Landscape Division), Transportation Area ED, Agricultural Area D, SHDI (Shannon Diversity Index), Agricultural Area LPI. The distribution of positive or negative (based on 0 value) waterlogging impact on watershed units by landscape pattern characteristics is shown in Figure 16. The detailed impact degree and the number of watershed units affected by negative and positive waterlogging are shown in Table 9: Transportation Area PD (2.71×10^{-1}, 44 : 11), Agricultural Area ED (2.35×10^{-1}, 52 : 3), Barren Area LPI (2.06×10^{-1}, 52 : 3), Forest ED (1.87×10^{-1}, 43 : 12), Used Area LPI (1.81×10^{-1}, 46 : 9), Transportation Area D (1.40×10^{-1}, 37 : 18), Transportation Area ED (1.12×10^{-1}, 36 : 19), Agricultural Area D (8.75×10^{-2}, 6 : 48), SHDI (8.37×10^{-2}, 15 : 40), Agricultural Area LPI (7.60×10^{-2}, 48 : 7).

According to Fig. 17, the top 10 landscape pattern metrics that have the greatest impact on waterlogging in SSWU (Small-scale Watershed Units) are as follows: Agriculture Area ED (Edge Density), Transportation Area PD (Path Density), Barren Area LPI (Largest Patch Index), SHDI (Shannon Diversity Index), Used Area PD, Transportation Area D (Degree of Landscape Division), Agricultural Area LPI, Slope, Used Area D, Grass D. The distribution of positive or negative (based on 0 value) waterlogging impact on watershed units by landscape pattern characteristics is shown in Figure 18. The detailed impact degrees and the number of watershed units affected by negative and positive waterlogging are shown in Table 10: Agriculture Area ED (3.13×10^{-1}, 58 : 4), Transportation Area PD (1.92×10^{-1}, 47 : 15), Barren Area LPI (1.80×10^{-1}, 58 : 4), SHDI (1.67×10^{-1}, 41 : 21), Used Area PD (1.06×10^{-1}, 56 : 6), Transportation Area D (9.02×10^{-2}, 58 : 4), Agricultural Area LPI (8.95×10^{-2}, 51 : 11), Slope (8.95×10^{-2}, 55 : 7), Used Area D (8.30×10^{-2}, 19 : 43), Grass D (8.22×10^{-2}, 9 : 53).

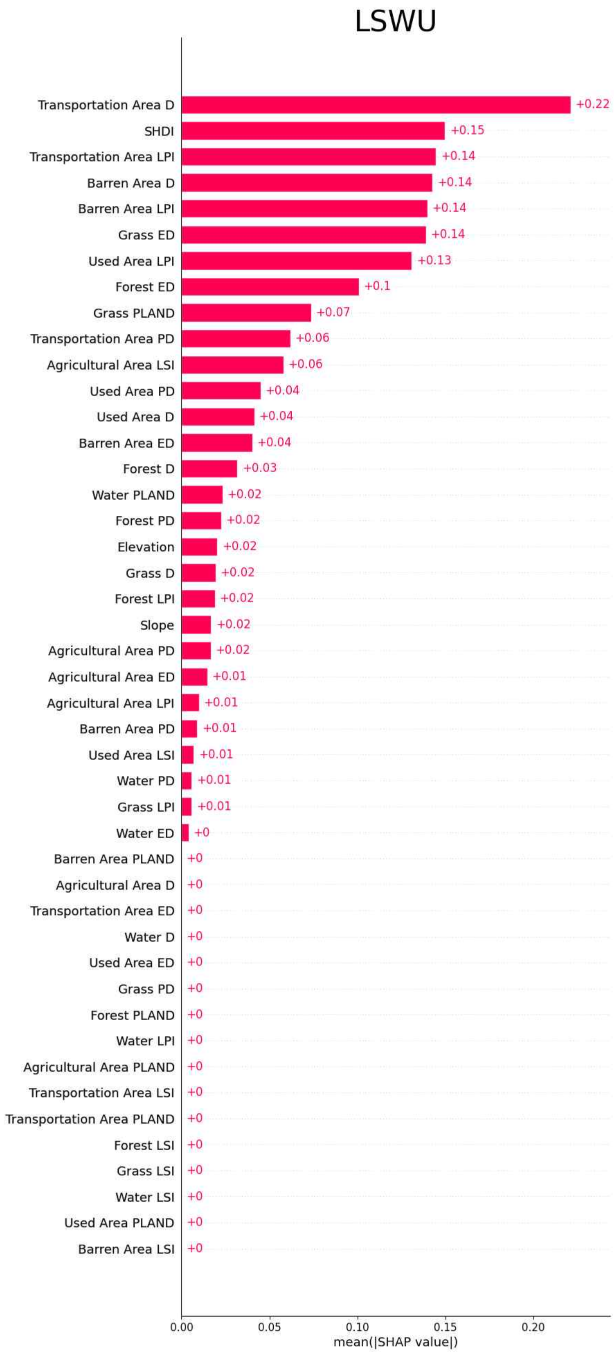

Figure 13. Contribution of landscape pattern metrics to waterlogging in large-scale watershed Units by Shapley Additive Explanations (SHAP).

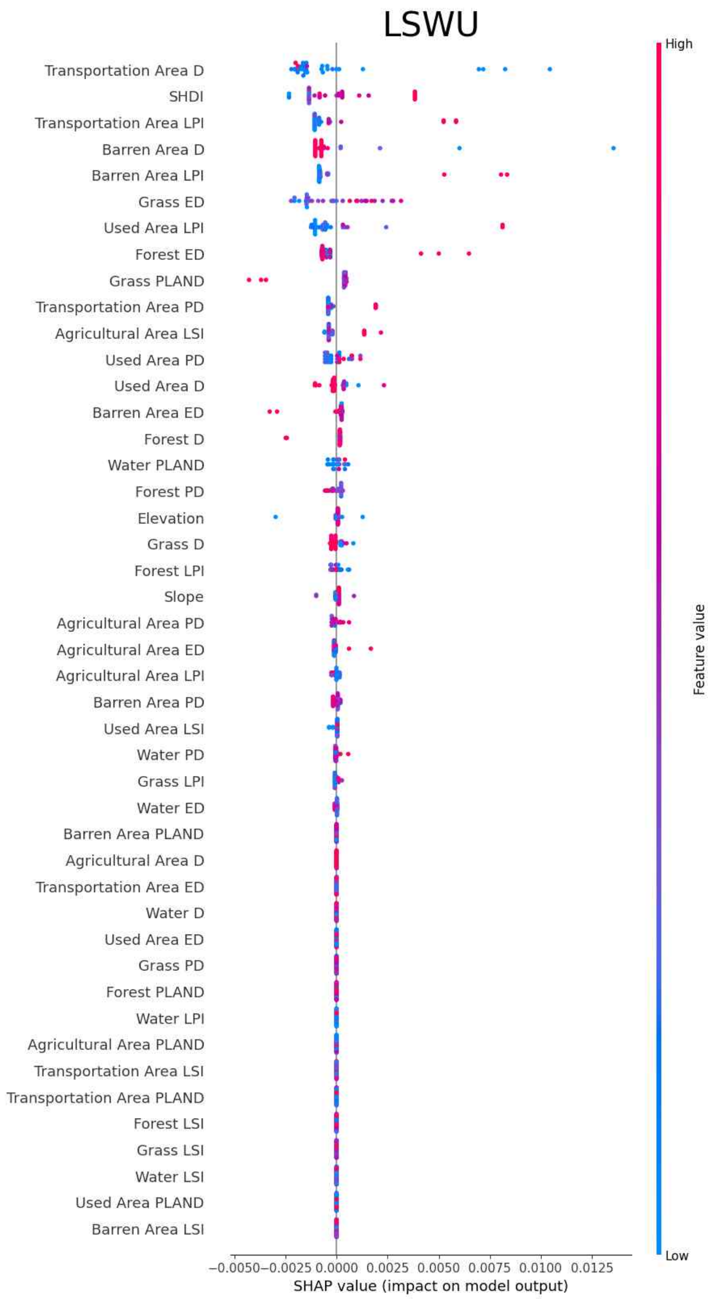

Figure 14. Contribution of landscape pattern metrics to waterlogging in large-scale watershed Units by Shapley Additive Explanations (SHAP).

[Table 8] The values of landscape pattern metrics of Large-scale Watershed Units (LSWU) analyzed by Shapley Additive Explanations (SHAP).

Landscape Category	Value	Number of Watershed Units Less than 0	Number of Watershed Units Greater than 0
Transportation Area D	2.21×10^{-1}	24	7
SHDI	1.50×10^{-1}	17	14
Transportation Area LPI	1.45×10^{-1}	26	5
Barren Area D	1.43×10^{-1}	26	5
Barren Area LPI	1.40×10^{-1}	28	3
Grass ED	1.39×10^{-1}	18	13
Used Area LPI	1.31×10^{-1}	24	7
Forest ED	1.01×10^{-1}	28	3
Grass PLAND	7.37×10^{-2}	3	28
Transportation Area PD	6.19×10^{-2}	26	5
Agricultural Area LSI	5.78×10^{-2}	25	6
Used Area PD	4.49×10^{-2}	16	15
Used Area D	4.15×10^{-2}	21	10
Barren Area ED	4.02×10^{-2}	3	28
Forest D	3.17×10^{-2}	2	29
Water PLAND	2.34×10^{-2}	15	16
Forest PD	2.25×10^{-2}	12	19
Elevation	2.04×10^{-2}	6	25
Grass D	1.95×10^{-2}	22	9
Forest LPI	1.90×10^{-2}	15	16
Slope	1.68×10^{-2}	15	16
Agricultural Area PD	1.66×10^{-2}	19	12
Agricultural Area ED	1.48×10^{-2}	29	2
Agricultural Area LPI	9.95×10^{-3}	15	16
Barren Area PD	9.05×10^{-3}	15	16
Used Area LSI	6.97×10^{-3}	4	27
Water PD	5.84×10^{-3}	28	3
Grass LPI	5.80×10^{-3}	21	10
Water ED	3.97×10^{-3}	9	22
Agricultural Area D	0	0	0
Water D	0	0	0
Water LPI	0	0	0
Transportation Area LSI	0	0	0
Forest LSI	0	0	0

Grass LSI	0	0	0
Water LSI	0	0	0
Barren Area LSI	0	0	0
Grass PD	0	0	0
Used Area PLAND	0	0	0
Transportation Area PLAND	0	0	0
Agricultural Area PLAND	0	0	0
Forest PLAND	0	0	0
Barren Area PLAND	0	0	0
Used Area ED	0	0	0
Transportation Area ED	0	0	0

[Table 9] The top ten landscape pattern characteristics with the highest contribution to the occurrence of waterlogging in Large-scale Watershed Units (LSWU) were obtained by using Shapley Additive Explanations (SHAP). Then by synthesizing the value of each landscape pattern feature, we can confirm whether the impact on waterlogging is positive or negative.

Landscape Category	Sum Value	Promote / Alleviate
Transportation Area D	-1.95×10^{-18}	Alleviate
SHDI	1.30×10^{-18}	Promote
Transportation Area LPI	-8.67×10^{-19}	Alleviate
Barren Area D	1.73×10^{-18}	Promote
Barren Area LPI	1.73×10^{-18}	Promote
Grass ED	1.73×10^{-18}	Promote
Used Area LPI	-6.51×10^{-19}	Alleviate
Forest ED	6.51×10^{-19}	Promote
Grass PLAND	-8.67×10^{-19}	Alleviate
Transportation Area PD	-5.42×10^{-19}	Alleviate

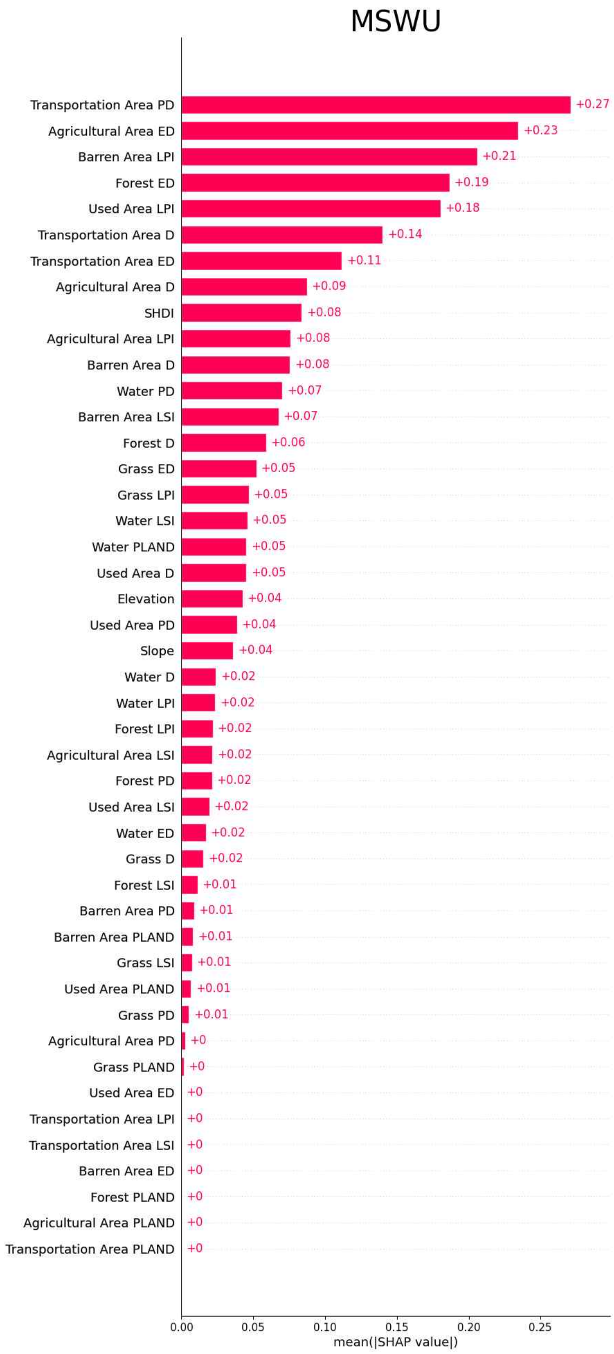

Figure 15. Contribution of landscape pattern metrics to waterlogging in middle-scale watershed Units by Shapley Additive Explanations (SHAP).

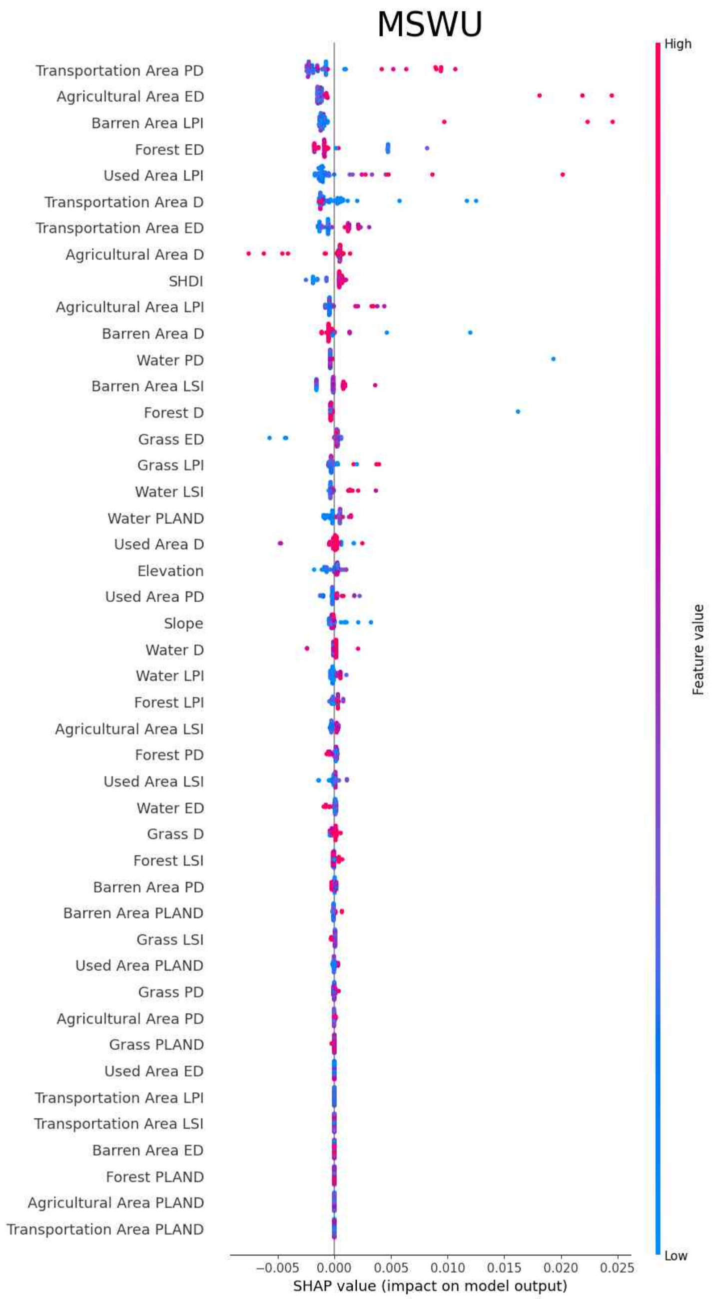

Figure 16. Contribution of landscape pattern metrics to waterlogging in middle-scale watershed Units by Shapley Additive Explanations (SHAP).

[Table 10] The values of landscape pattern metrics of Middle-scale Watershed Units (MSWU) analyzed by Shapley Additive Explanations (SHAP).

Landscape Category	Value	Number of Watershed Units Less than 0	Number of watershed Units Greater than 0
Transportation Area PD	2.71×10^{-1}	44	11
Agricultural Area ED	2.35×10^{-1}	52	3
Barren Area LPI	2.06×10^{-1}	52	3
Forest ED	1.87×10^{-1}	43	12
Used Area LPI	1.81×10^{-1}	46	9
Transportation Area D	1.40×10^{-1}	37	18
Transportation Area ED	1.12×10^{-1}	36	19
Agricultural Area D	8.75×10^{-2}	6	49
SHDI	8.37×10^{-2}	15	40
Agricultural Area LPI	7.60×10^{-2}	48	7
Barren Area D	7.55×10^{-2}	48	7
Water PD	7.04×10^{-2}	54	1
Barren Area LSI	6.76×10^{-2}	36	19
Forest D	5.90×10^{-2}	54	1
Grass ED	5.21×10^{-2}	4	51
Grass LPI	4.70×10^{-2}	44	11
Water LSI	4.60×10^{-2}	48	7
Water PLAND	4.52×10^{-2}	33	22
Used Area D	4.52×10^{-2}	20	35
Elevation	4.27×10^{-2}	17	38
Used Area PD	3.87×10^{-2}	34	21
Slope	3.60×10^{-2}	46	9
Water D	2.40×10^{-2}	25	30
Water LPI	2.36×10^{-2}	40	15
Forest LPI	2.18×10^{-2}	33	22
Agricultural Area LSI	2.16×10^{-2}	27	28
Forest PD	2.15×10^{-2}	15	40
Used Area LSI	1.94×10^{-2}	21	34
Water ED	1.70×10^{-2}	8	47
Grass D	1.53×10^{-2}	23	32
Forest LSI	1.11×10^{-2}	47	8
Barren Area PD	8.70×10^{-3}	19	36
Barren Area PLAND	8.05×10^{-3}	39	16
Grass LSI	7.29×10^{-3}	14	41

Used Area PLAND	6.73×10^{-3}	44	11
Grass PD	5.25×10^{-3}	42	13
Agricultural Area PD	2.72×10^{-3}	38	17
Grass PLAND	1.62×10^{-3}	2	53
Transportation Area LPI	0	0	0
Transportation Area LSI	0	0	0
Transportation Area PLAND	0	0	0
Agricultural Area PLAND	0	0	0
Forest PLAND	0	0	0
Used Area ED	0	0	0
Barren Area ED	0	0	0

[Table 11] The top ten landscape pattern characteristics with the highest contribution to the occurrence of waterlogging in Middle-scale Watershed Units (MSWU) were obtained by using Shapley Additive Explanations (SHAP). Then by synthesizing the value of each landscape pattern feature, we can confirm whether the impact on waterlogging is positive or negative.

Landscape Category	Sum Value	Promote / Alleviate
Transportation Area PD	5.20×10^{-18}	Promote
Agricultural Area ED	4.99×10^{-18}	Promote
Barren Area LPI	-1.52×10^{-18}	Alleviate
Forest ED	1.73×10^{-18}	Promote
Used Area LPI	-1.73×10^{-18}	Alleviate
Transportation Area D	-1.86×10^{-18}	Alleviate
Transportation Area ED	4.34×10^{-19}	Promote
Agricultural Area D	5.42×10^{-20}	Promote
SHDI	8.67×10^{-19}	Promote
Agricultural Area LPI	-4.34×10^{-19}	Alleviate

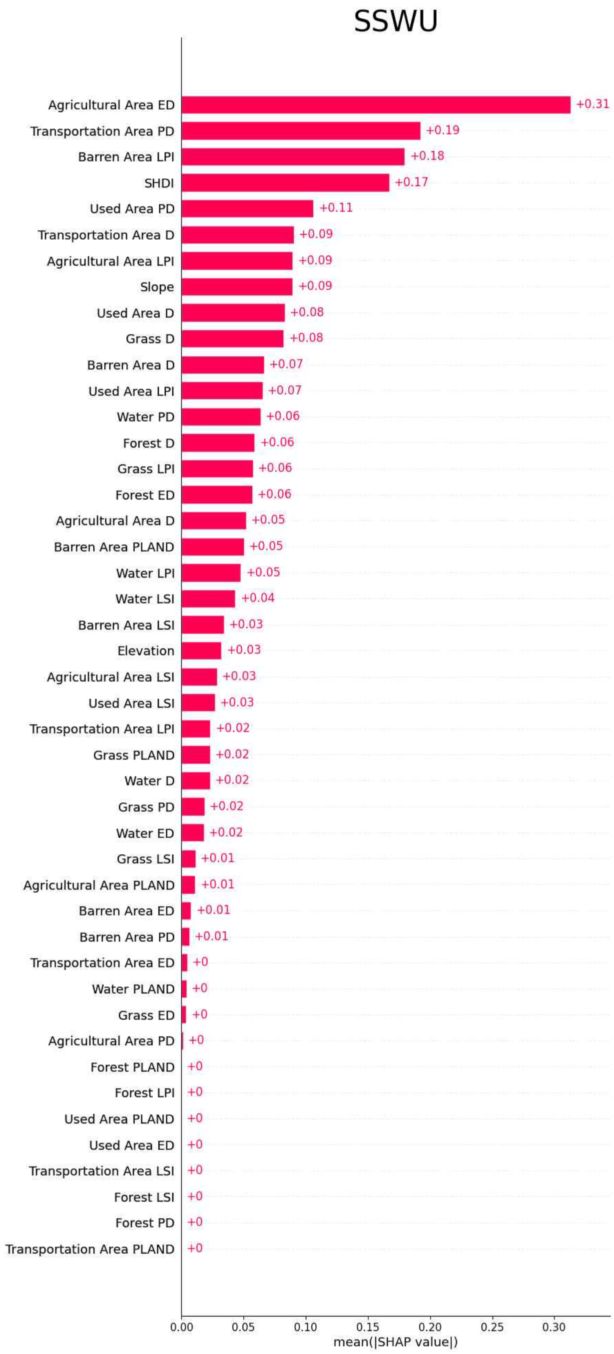

Figure 17. Contribution of landscape pattern metrics to waterlogging in small-scale watershed Units by Shapley Additive Explanations (SHAP).

Figure 18. Contribution of landscape pattern metrics to waterlogging in small-scale watershed Units by Shapley Additive Explanations (SHAP).

[Table 12] The values of landscape pattern metrics of Small-scale Watershed Units (SSWU) analyzed by Shapley Additive Explanations (SHAP).

Landscape Category	Absolute Value	Number of watershed units less than zero	Number of watershed units greater than zero
Agricultural Area ED	3.13×10^{-1}	58	4
Transportation Area PD	1.92×10^{-1}	47	15
Barren Area LPI	1.80×10^{-1}	58	4
SHDI	1.67×10^{-1}	41	21
Used Area PD	1.06×10^{-1}	56	6
Transportation Area D	9.02×10^{-2}	58	4
Agricultural Area LPI	8.95×10^{-2}	51	11
Slope	8.95×10^{-2}	55	7
Used Area D	8.30×10^{-2}	19	43
Grass D	8.22×10^{-2}	9	53
Barren Area D	6.66×10^{-2}	43	19
Used Area LPI	6.54×10^{-2}	50	12
Water PD	6.35×10^{-2}	60	2
Forest D	5.88×10^{-2}	58	4
Grass LPI	5.76×10^{-2}	50	12
Forest ED	5.71×10^{-2}	41	21
Agricultural Area D	5.21×10^{-2}	3	59
Barren Area PLAND	5.04×10^{-2}	43	19
Water LPI	4.78×10^{-2}	39	23
Water LSI	4.33×10^{-2}	51	11
Barren Area LSI	3.42×10^{-2}	35	27
Elevation	3.22×10^{-2}	19	43
Agricultural Area LSI	2.86×10^{-2}	33	29
Used Area LSI	2.69×10^{-2}	40	22
Transportation Area LPI	2.34×10^{-2}	17	45
Grass PLAND	2.29×10^{-2}	6	56
Water D	2.29×10^{-2}	27	35
Grass PD	1.86×10^{-2}	21	41
Water ED	1.80×10^{-2}	5	57
Grass LSI	1.14×10^{-2}	42	20
Agricultural Area PLAND	1.10×10^{-2}	35	27
Barren Area ED	7.73×10^{-3}	36	26
Barren Area PD	6.47×10^{-3}	31	31
Transportation Area ED	4.70×10^{-3}	41	21
Water PLAND	4.26×10^{-3}	42	20

Grass ED	3.89×10^{-3}	22	40
Agricultural Area PD	1.39×10^{-3}	9	53
Forest LPI	0	0	0
Transportation Area LSI	0	0	0
Forest LSI	0	0	0
Forest PD	0	0	0
Used Area PLAND	0	0	0
Transportation Area PLAND	0	0	0
Forest PLAND	0	0	0
Used Area ED	0	0	0

[Table 13] The top ten landscape pattern characteristics with the highest contribution to the occurrence of waterlogging in Small-scale Watershed Units (SSWU) were obtained by using Shapley Additive Explanations (SHAP). Then by synthesizing the value of each landscape pattern feature, we can confirm whether the impact on waterlogging is positive or negative.

Landscape Category	Sum Value	Promote / Alleviate
Agricultural Area ED	-3.04×10^{-18}	Alleviate
Transportation Area PD	-7.81×10^{-18}	Alleviate
Barren Area LPI	-4.88×10^{-18}	Alleviate
SHDI	-3.04×10^{-18}	Alleviate
Used Area PD	2.60×10^{-18}	Promote
Transportation Area D	-1.29×10^{-18}	Alleviate
Agricultural Area LPI	8.67×10^{-19}	Promote
Slope	-1.84×10^{-18}	Alleviate
Used Area D	1.46×10^{-18}	Promote
Grass D	3.31×10^{-18}	Promote

The results of the SHAP analysis are shown in Figures 13-18 and Tables 8, 10, and 12. To better explain the results, I classify the top 10 landscape pattern metrics (Tables 9, 11, 13) with the highest contribution to waterlogging, then confirm the contribution of each landscape category to waterlogging by the sum of absolute values (based on the 18th power of 10), and then at last analyze whether their impact is positive or negative according to the eigenvalue to confirm the impact of each landscape category (Tables 8, 10, 12).

In LSWU (Fig. 14), the contribution of landscape types to the impact of waterlogging is sorted as follows: Barren Area (D & LPI, 3.46×10^{18}), Transportation Area (D & LPI & PD, 3.36×10^{18}), Grass (ED & PLAND, 2.60×10^{18}), SHDI (1.30×10^{18}), Used Area (LPI, 0.65×10^{18}), Forest (ED, 0.65×10^{18}). The contribution values of Barren Area and Transportation Area to waterlogging are similar, followed by Grass, SHDI, Used Area, and Forest. According to Table 9 and the definition of landscape pattern metrics, it is necessary to greatly reduce the barren area in the process of waterlogging control at this watershed-scale unit. Although the more dispersed the transportation area is, the better the alleviation of waterlogging will be. However, because of the actual needs of society, we cannot arbitrarily cut off the transportation area. Therefore, in this part, we can make up for the gap in the contribution of the transportation area by reducing the fine grass and avoiding the excessive dispersion and fragmentation of landscape elements. The impact of the forest is small and can be ignored.

In MSWU (Fig. 16), the contribution of landscape types to the impact of waterlogging is sorted as follows: Transportation Area (PD & ED & D, 7.49×10^{18}), Agricultural Area (ED & D & LPI, 5.47×10^{18}), Forest (ED, 1.73×10^{18}), Used Area (LPI, 1.73×10^{18}), Barren Area (LPI, 1.52×10^{18}), SHDI (0.87×10^{18}). The contribution of Transportation Area and Agricultural Area to waterlogging is much higher than that of Forest, Used Area, Barren Area, and SHDI. According to Table 11 and the definition of landscape metrics, in the process of waterlogging control in this scale watershed, it is necessary to adjust the edge length of the agricultural area to gather the fragmented agricultural areas together to form an agricultural area with a relatively large single area, which is more conducive to the alleviation of waterlogging. The contribution of the forest is relatively small, but since the transportation area cannot be changed at will, it can be adjusted by adding scattered forest areas. At the same time, the maximum single patch area of Used Area and Barren Area shall be further expanded. The contribution of SHDI is relatively low and almost negligible.

In SSWU (Fig. 18), the contribution of landscape types to the impact of waterlogging is sorted as follows: Transportation Area (PD & D, 9.10×10^{18}), Barren Area (LPI, 4.88×10^{18}), Used Area (PD & D, 4.00×10^{18}), Agricultural Area (LPI & ED, 3.91×10^{18}), Grass (D, 3.31×10^{18}), SHDI (3.04×10^{18}), Slope (1.84×10^{18}). Although the contribution of the Transportation Area to waterlogging is absolutely important, as the previous results, the Transportation Area, as the demand of human production activities, cannot be changed at will, the Barren Area, used area, and agricultural area all have high contributions to waterlogging, so we can focus on the intervention of these three landscape types. Of course, the contribution of Grass and SHDI is relatively low, but it can not be ignored. The Slope is a natural landform that should not be changed, and its importance is the lowest, so it can be ignored. According to Table 13 and the definition of landscape metrics, from this result. The Barron area patch with the largest single area should be further expanded to reduce the area proportion of the used area in the basin unit, and it is better to concentrate the Used Area with the scattered area. Reduce and disperse the agricultural area with the largest single area. While focusing on Grass, will increase the landscape diversity within the watershed unit.

3.3 Prediction Mean Square Error (MSE)

According to the analysis results of GWR and SHAP in each scale watershed (Fig. 7-9), the accuracy ranking (starting from the smallest error) are as Table 14: LSWU-SHAP (1.51×10^{-4}) < MSWU-SHAP (1.60×10^{-4}) < SSWU-SHAP (1.90×10^{-4}) < SSWU-GWR (7.45×10^{-2}) < MSWU-GWR (9.20×10^{-2}) < LSWU-GWR (1.64×10^{-1}). That is to say, using the SHAP method to analyze the relationship between landscape patterns and waterlogging in large-scale watershed units, the error is the smallest and the most accurate.

[Table 14] The Results of Prediction Mean Squared Error (MSE).

Scale of Watershed Units	GWR	SHAP
Large-scale Watershed Units (LSWU)	1.64×10^{-1}	1.51×10^{-4}
Middle-scale Watershed Units (MSWU)	9.20×10^{-2}	1.60×10^{-4}
Small-scale Watershed Units (SSWU)	7.45×10^{-2}	1.90×10^{-4}

3.4 Piecewise Linear Model

Through the results of Prediction Mean Squared Error method, we know that the results of LSWU analysis using SHAP method are the most accurate. Therefore, I calculated the threshold of the top 10 landscape pattern characteristics (Fig. 13-14) that show the greatest impact on waterlogging in LSWU as shown as Fig. 19-28. The threshold values are (Table 15): Transportation Area D (3.69×10^{-2}), SHDI (1.65), Transportation Area LPI (1.37×10^{1}), Barren Area D (7.68×10^{-1}), Barren Area LPI (2.36), Grass ED (2.98×10^{1}), Used Area LPI (1.50×10^{-1}), Forest ED (7.19×10^{1}), Grass PLAND (2.64×10^{-1}), Transportation Area PD (2.67).

Combined with Table 9, the results mean that when Barren Area D (Fig. 19) value reache (7.68×10^{-1}), Barren Area LPI (Fig. 20) value reaches (2.36), Grass ED (Fig. 24) value reaches (2.98×10^{1}), SHDI (Fig. 26) value reaches (1.65), Forest ED (Fig. 28) value reaches (7.19×10^{1}). The promoting effect of these landscape features on waterlogging is greatly weakened. On the other hand, when Transportation Area D (Fig. 21) value reaches(3.69×10^{-2}), Transportation Area LPI (Fig. 22) value reaches (1.37×10^{1}), Grass PLAND (Fig. 25) value reaches (2.64×10^{-1}), Used Area LPI (Fig. 27) value reaches (1.50×10^{-1}). The alleviating effect of these landscape features on waterlogging is greatly weakened.

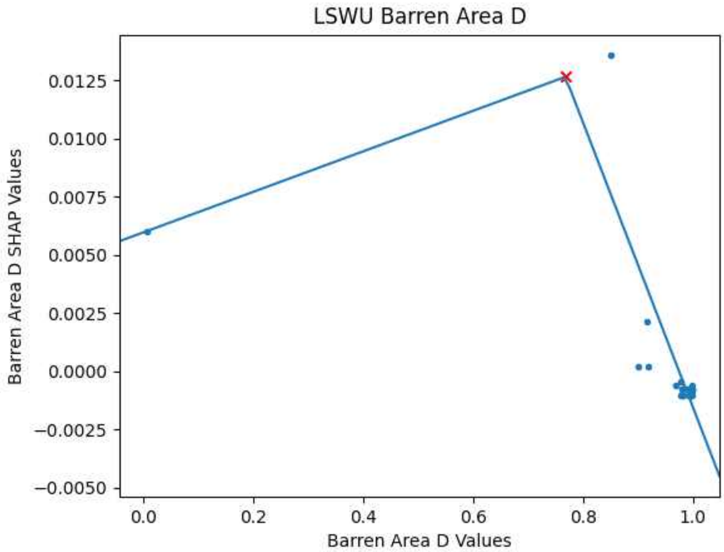

Figure 19. Threshold value of Barren Area D (Landscape Division Index).

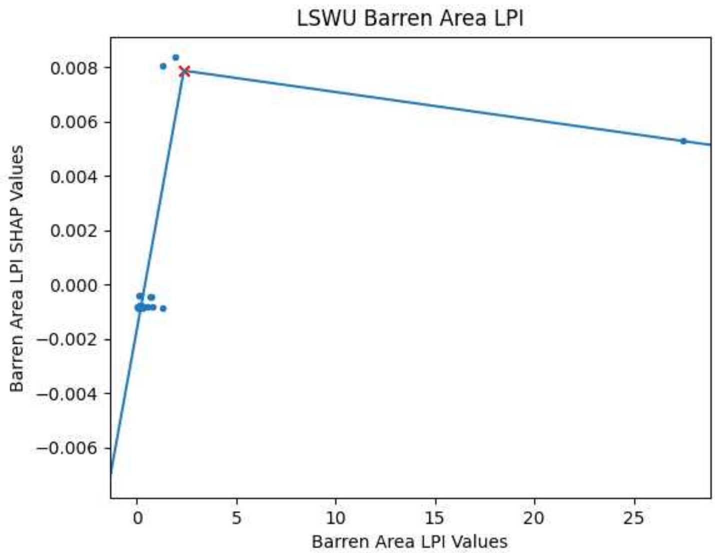

Figure 20. Threshold value of Barren Area LPI (Largest Patch Index).

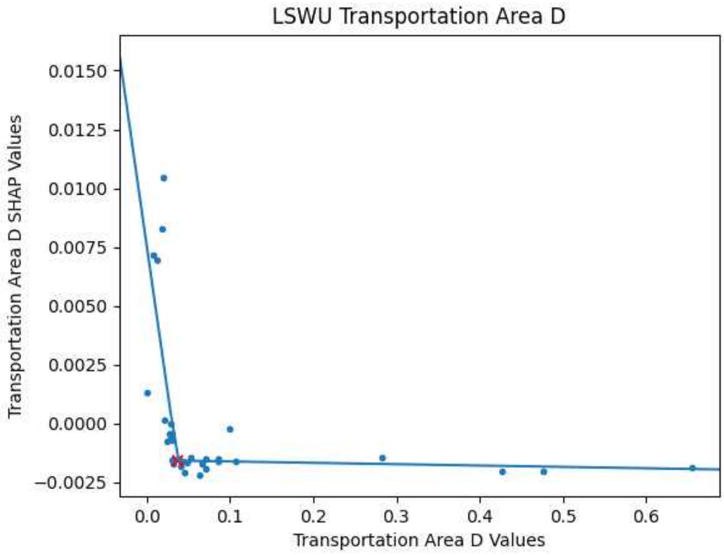

Figure 21. Threshold value of Transportation Area D (Landscape Division Index).

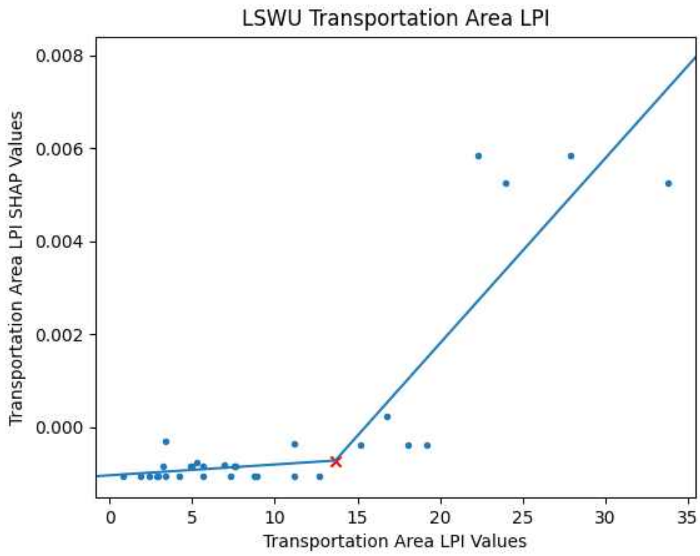

Figure 22. Threshold value of Transportation Area LPI (Largest Patch Index).

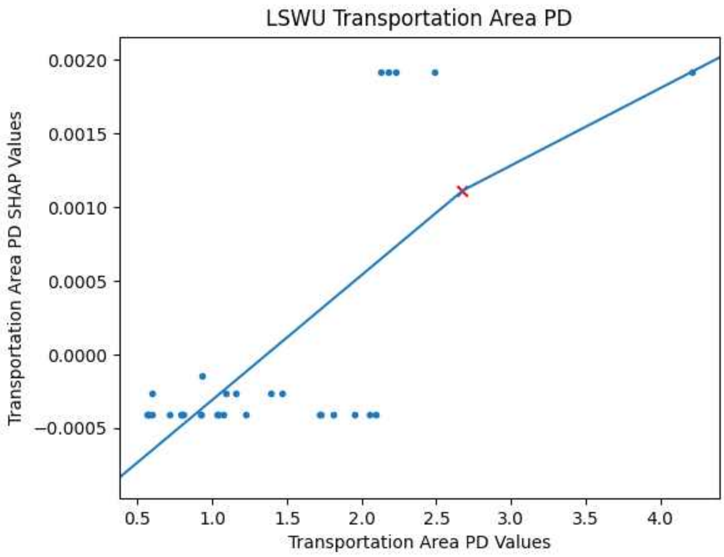

Figure 23. Threshold value of Transportation Area PD (Patch Density).

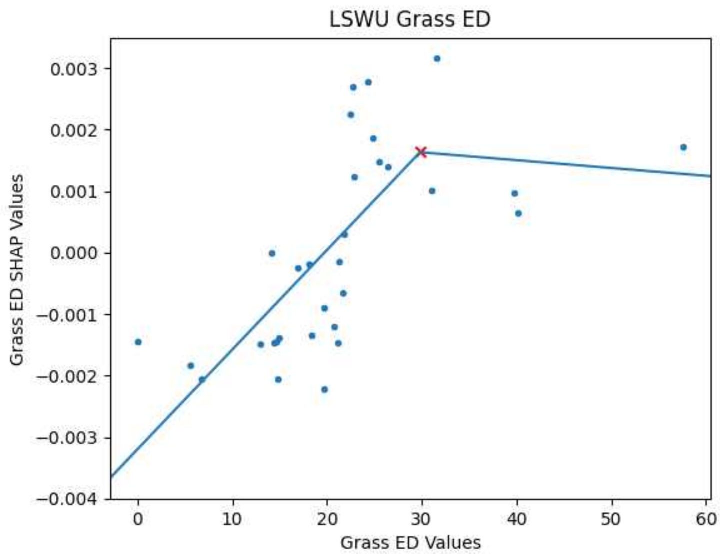

Figure 24. Threshold value of Grass ED (Edge Density).

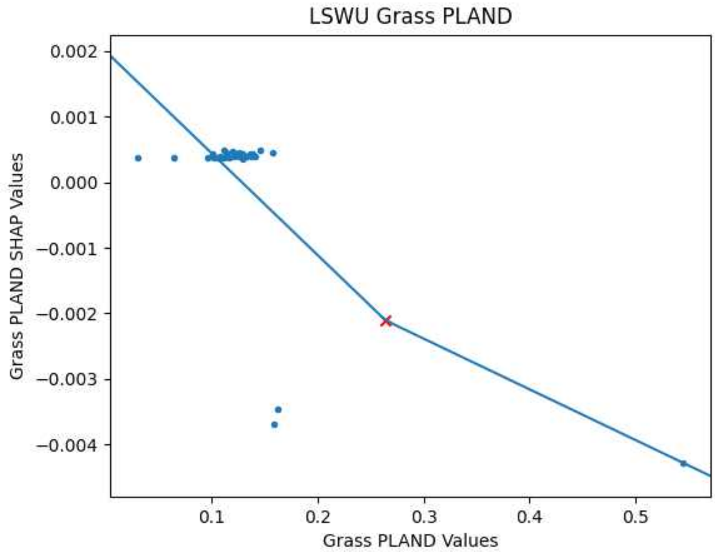

Figure 25. Threshold value of Grass PLAND (Percentage of Landscape).

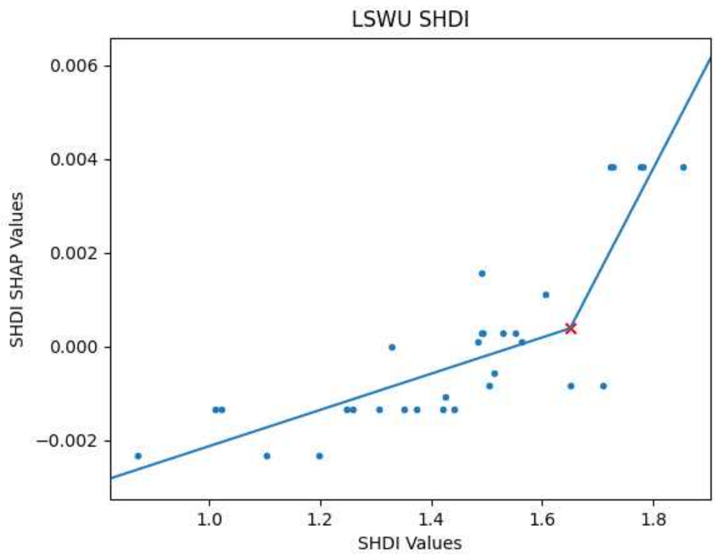

Figure 26. Threshold value of SHDI (Shannon Diversity Index).

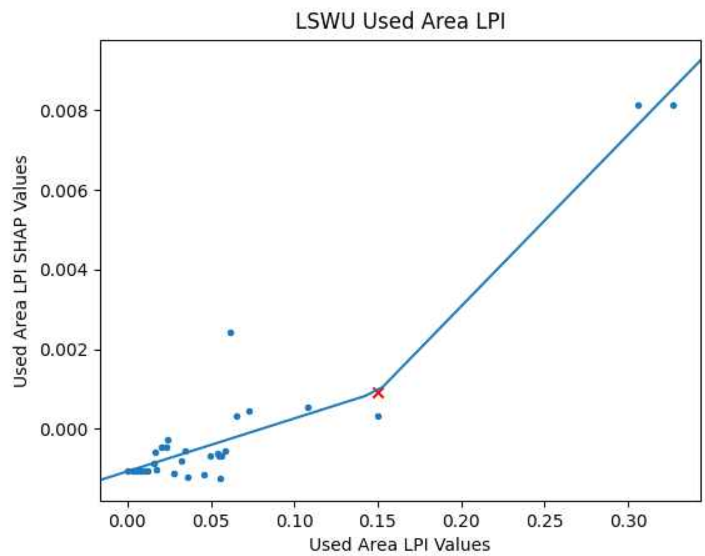

Figure 27. Threshold value of Used Area LPI (Largest Patch Index).

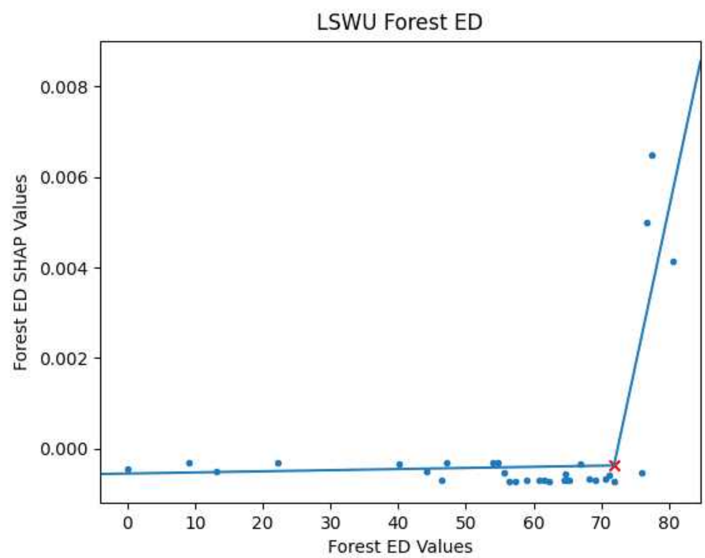

Figure 28. Threshold value of Forest ED (Edge Density).

[Table 15] The threshold value of the top ten landscape pattern characteristics with the highest contribution to waterlogging in LSWU (Large-scale Watershed Units) by using Shapley Additive Explanations (SHAP).

Landscape Category	Value
Transportation Area D	3.69×10^{-2}
SHDI	1.65
Transportation Area LPI	1.37×10^{1}
Barren Area D	7.68×10^{-1}
Barren Area LPI	2.36
Grass ED	2.98×10^{1}
Used Area LPI	1.50×10^{-1}
Forest ED	7.19×10^{1}
Grass PLAND	2.64×10^{-1}
Transportation Area PD	2.67

4. Discussion

4.1 Selection of Data and Tools

High precision landscape pattern analysis in the macro-scale watersheds is helpful for us to objectively examine the role of the landscape ecosystem in waterlogging. If we want to further understand the local situation, we can select the watershed unit which we want to know from the waterlogging degree data, and use the sub-classified landscape types used in this study (There are 41 categories of metadata. In this study, I summarize them into 7 categories). That is to say, the same data can be used to analyze the macro scale, but when necessary, the local scale can be enlarged for analysis. However, it is impossible to do this by neglecting the macro impact and limiting the study to small-scale areas. Python tools can do this freely.

Namely, the Python libraries can directly analyze the shapefile of the fine classified landscape, instead of converting it to Fragstats software for analysis, avoiding the computational loss caused by the conversion of vector data to raster data.

4.2 Supervised Learning and Interpretive Machine Learning

Taking this study as an example, Geographically Weighted Regression (GWR) is an analysis method often used in the study of waterlogging. Many researchers use the results of this theory to try to explain the relationship between landscape patterns and waterlogging. However, in a strict sense, the goal of supervised learning is not to study the correlation, but to reduce the errors in prediction through the support of weight values, so as to provide useful information in the real world. Therefore, the results predicted by the supervised learning method may not reflect the causal relationship in the real world [22, 39]. However, because human beings have rich generalization ability, they hope to interpret the data of supervised learning and observe and infer causality. Also, this method often relies on the verification conclusions of previous studies and requires assumptions on this basis. However, we need to be vigilant that in the study of the natural environment, due to its complexity, there will always be undetected interactions between relevant variables. The use of these methods in this non-stationary environment may invalidate their prediction results [39]. Shapley Additive Explanations (SHAP) is a breakthrough in the field of machine learning in recent years. This method uses game theory to compare and analyze the impact of the combination of all data on the research object and obtains the contribution value [42].

From the results of the Prediction Mean Squared Error (MSE) (Table 14), the error value of GWR is actually very small, but far bigger than that of SHAP. In terms of several landscape types that GWR and SHAP have an impact on waterlogging on LSWU, their common ground is that Barren Area, Transportation Area, and Forest have a strong correlation/contribution to waterlogging. The difference is that according to the results of GWR, Agricultural Area and Elevation are also important, while SHAP gives that Used Area and SHDI (Shannon Diversity Index) are both important. Through Fig. 29, we can observe the difference in the ratio of correlation degree/contribution degree of the elements at these different points (blue is the ratio of correlation degree of GWR and red is the ratio of contribution degree of SHAP). In MSWU comparison, GWR and SHAP common ground are Transportation Area, Agricultural Area, Forest, and Barren Area. The difference is that Water and Grass are also important according to the results of GWR, while the Used Area and SHDI are important according to SHAP. Through Fig. 30, we can observe the difference in the ratio of the correlation degree/contribution degree of the elements at these different points. In SSWU, they still have the common ground of Transportation Area, Barren, and Agricultural Area. The difference is that the Used Area is also considered by GWR to have a strong correlation with waterlogging unlike before. Also, the difference is that according to the results of GWR, the Forest is also very important, while SHAP thinks that Grass and SHDI are both important. Through Fig. 31, we can observe the difference in the ratio of the correlation degree/contribution degree of the elements at these different points.

The above is a description of the diversity of landscape categories summarized from the top 10 landscape categories with the strongest correlation/contribution in the GWR and SHAP results. We can find that both GWR and SHAP believe that the Transportation Area and Barron area play a role in the occurrence of waterlogging, especially since the proportion of Transportation Area is very high, which is also consistent with the relevant research conclusions of the past that impermeable surface has an important impact on waterlogging. However, it is unusual that the SHDI impact index in the results of GWR is very low in any watershed unit. Whether it is the result of separating all landscape features separately (Fig. 10-12) or only based on the analysis of landscape types, the correlation between SHDI and waterlogging is very low, even at the bottom. But, according to the results of SHAP, SHDI plays an important role in any watershed unit (Fig. 13-18), which is one of the important landscape features contributing to waterlogging. This also confirms Lipton et al. [39], namely the response prediction of the supervised learning model to the causal relationship in the real world is likely to be invalid because it does not detect the impact of the interaction between the landscape categories and their characteristics on the real world. Therefore, I stated Finding 1: When we use the interpretive machine model, we can reflect the interaction between landscape patterns and waterlogging more objectively and accurately than supervised learning, because it fully considers the interaction between landscape patterns and the impact of this interaction on waterlogging.

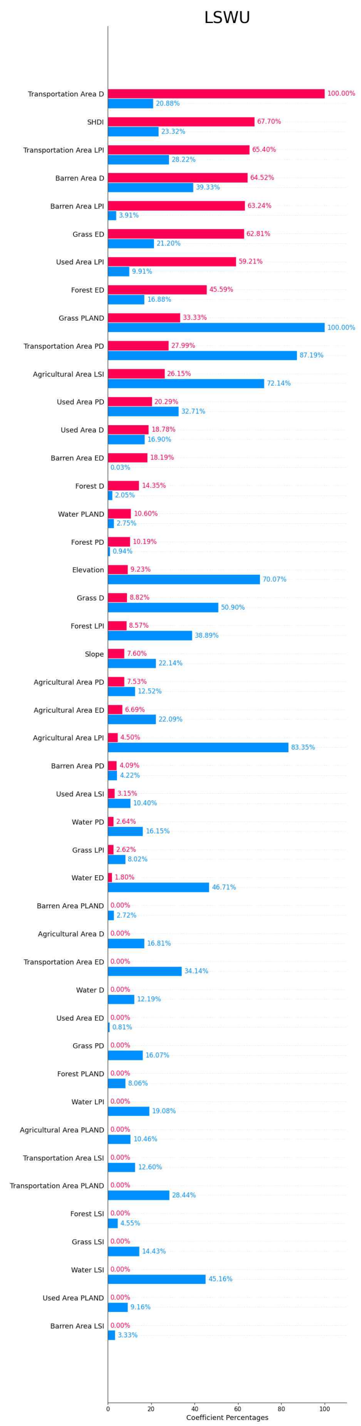

Figure 29. The proportion of the impact of each landscape pattern on waterlogging in the large-scale watershed units (LSWU) obtained by using Geographically Weighted Regression (GWR) and Shapley Additive Explanations (SHAP). Red is SHAP and blue is GWR.

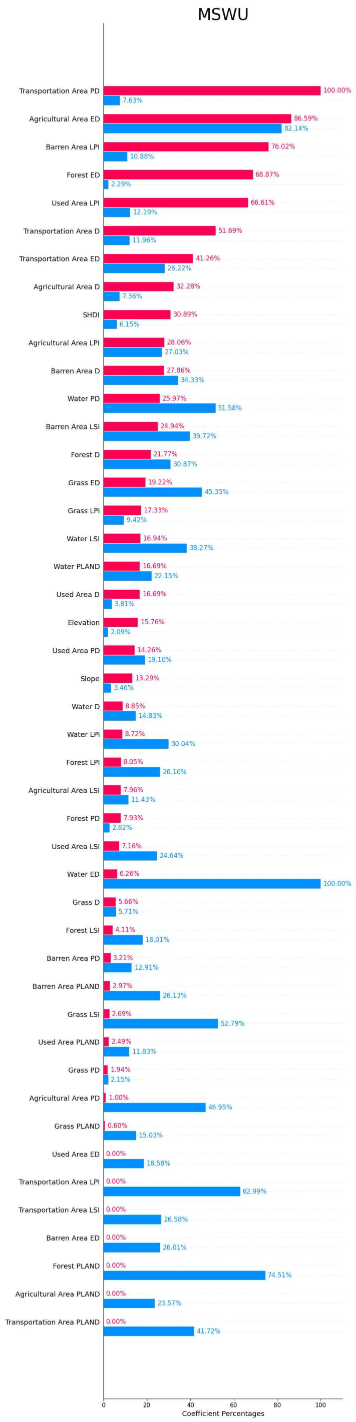

Figure 30. The proportion of the impact of each landscape pattern on waterlogging in the middle-scale watershed units (MSWU) obtained by using Geographically Weighted Regression (GWR) and Shapley Additive Explanations (SHAP). Red is SHAP and blue is GWR.

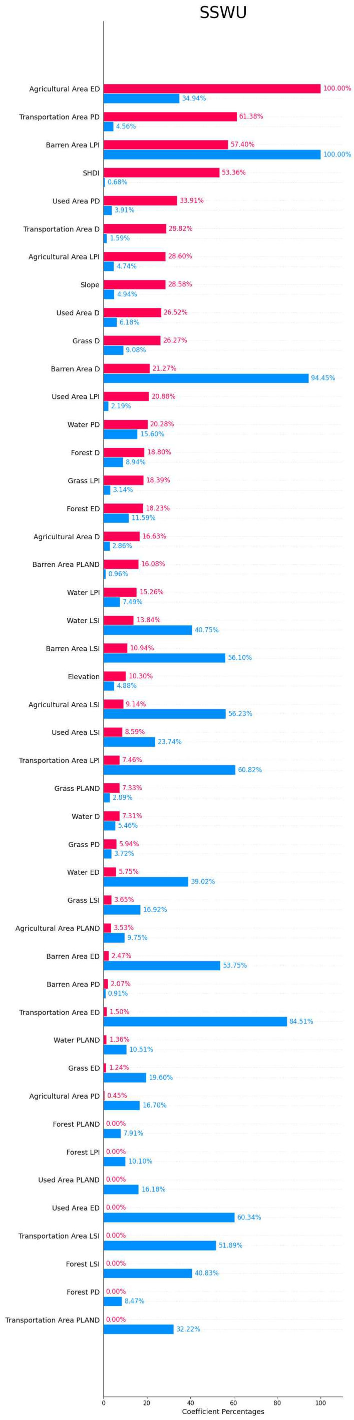

Figure 31. The proportion of the impact of each landscape pattern on waterlogging in the small-scale watershed units (SSWU) obtained by using Geographically Weighted Regression (GWR) and Shapley Additive Explanations (SHAP). Red is SHAP and blue is GWR.

4.3 Landscape Threshold and Hydrological Disaster

According to the results of Prediction Mean Squared Error (MSE, Table 14), it is the most accurate to take Large-scale Watershed Units (LSWU) as the research object of waterlogging, which also conforms to Schumm's viewpoint [51]. This is to say, in large-scale watershed units, the hydrological disasters caused by the change of geomorphic threshold are mainly dominated by climate change. In small-scale watershed units, the hydrological disasters caused by the change of geomorphic threshold are dominated by weather changes. Climate change determines weather change. Therefore, human beings should analyze the abnormal phenomenon of imbalance at a reasonable watershed scale to fundamentally find out the causes of waterlogging while trying to intervene in the ecosystem balance as soon as possible. Therefore, I stated Finding 2: Using a finer-scale watershed as a study site is not necessarily appropriate for waterlogging study.

4.4 Rational Use of Limited Land Resources

Although we can adjust the balance of natural ecology through external intervention in the landscape pattern to achieve the purpose of alleviating waterlogging. However, unlimited expansion of landscape types and specific features to alleviate waterlogging cannot solve the problem. Because nature itself is a complex composition, it is impossible to expand unilaterally and disorderly to alleviate waterlogging, which may cause other problems. At the same time, based on the realistic budget and other realistic concerns, we need to give full play to human wisdom in the limited land resources. Therefore, I stated Finding 3: There are threshold values for landscape features. When the impact on waterlogging reaches this critical point, its role in promoting or alleviating waterlogging will drastically change.

4.5 Limitation and Future Direction

Due to the lack of temporal change data of waterlogging, this study is not able to conduct a spatio-temporal analysis. For future work, if there are more detailed and comprehensive waterlog ging data, combined with the change of geomorphic landscape threshold, different results can be obtained. At the same time, in the process of development, human beings have obtained various materials from nature, and composed and invented many new materials. When these materials reach a certain amount, they will interfere with the dynamic balance of the ecological environment in a short time. Although nature can slowly adjust its balance, the impact of this temporary ecological imbalance on human beings may be huge. Compared with such a complex environment, the current landscape pattern metrics seem a little simple. As people pay more and more attention to their living environment, the formula of landscape ecology theory will developed in near future to better understand and explain the dynamic principle of natural ecology. In the past, it was difficult to verify the experimental results of large scale study area, but with the in-depth development of interpretative machine learning, it can hopefully overturn the results that we put forward a possible misleading but seemingly reasonable explanations.

5. Conclusion

The Seoul Capital Area (SCA) is the fifth-largest metropolitan area in the world, and also the economic hinterland of the Republic of Korea. With the concentration of resources and population, it will continue to attract new local working populations in the future. The developed economy and people's active social production activities have caused drastic changes in the land structure and characteristics of the region, resulting in the disturbance of the natural hydrological cycle, thus causing the waterlogging risk to become higher in SCA. In order to restore the dynamic balance of ecology and prevent potential life and health risks, based on landscape ecology, this study explores the interaction between landscape patterns and waterlogging, provides a reference for landscape planning and decision-makers, and draws the following conclusions:

Firstly, in the study on the impact of landscape patterns on waterlogging, if the interaction between landscape patterns is fully considered, the analysis results are more accurate. Both Geographically Weighted Regression (GWR) and Shapley Additive Explanations (SHAP) show that Barron Area and Transportation Area can promote the occurrence of waterlogging. However, the Shannon diversity index (SHDI) is seriously underestimated in the analysis results of GWR, because the supervised learning method based on regression analysis could not capture the interaction between landscape patterns. On the other hand, interpretive machine learning is an analytical method based on game theory, which can obtain the contribution to waterlogging through any combination of all landscape pattern characteristics. In other words, all possible combinations of elements of landscape patterns were compared. Therefore, in the results of SHAP, SHDI has a high impact on waterlogging in any watershed. At the same time, from the perspective of prediction results, SHAP is also far more accurate than GWR.

Secondly, using a finer-scale watershed as a study site is not necessarily appropriate for waterlogging study. If we want to adjust the landscape from the external intervention to achieve the dynamic balance of hydrology, we need to consider the characteristics of water first. The water cycle process has the characteristics of producing multi-scale geographical watersheds. Therefore, it is necessary to compare and analyze multi-level watershed-scale units. In this study, it is the most appropriate and accurate to study SCA waterlogging based on large-scale watershed units.

Finally, the characteristics of landscape patterns have the existence of a threshold. When the impact on waterlogging reaches a critical point, its role in promoting or alleviating waterlogging will change. Through the threshold of landscape pattern characteristics, the goal of alleviating waterlogging can be achieved accurately and at a low cost.

With the mature development of machine learning, we can interpret the complex natural environment more accurately. This study has made a new exploration of the methods of the interaction between landscape pattern and waterlogging. I wish it can provide a reference for the method and results of waterlogging control based on landscape ecology.

Appendix

Attribution	Full Name of Proper Noun	Abbreviations in this Study
Data	Land Cover Land Use	LULC
	Large-scale Watershed Units	LSWU
	Middle-scale Watershed Units	MSWU
	Small-scale Watershed Units	SSWU
Landscape Pattern Metrics	Percentage of Landscape	PLAND
	Landscape Division Index	D
	Largest Patch Index	LPI
	Landscape Shape Index	LSI
	Patch Density	PD
	Edge Density	ED
	Shannon Diversity Index	SHDI
Study Area	Seoul Capital Area	SCA
Mathematical Methods	Geographically Weighted Regression	GWR
	Shapley Additive Explanations	SHAP
	Prediction Mean Squared Error	MSE

References

[1] F. Aguilera, L. M. Valenzuela, and A. Botequilha-Leit˜ao. Landscape metrics in the analysis of urban land use patterns: A case study in a spanish metropolitan area. Landscape and Urban Planning, 99(3-4):226–238, 2011.

[2] A. H. Anni, S. Cohen, and S. Praskievicz. Sensitivity of urban flood simulations to stormwater infrastructure and soil infiltration. Journal of Hydrology, 588:125028, 2020.

[3] M. Antrop. Landscape change: Plan or chaos? Landscape and urban planning, 41(3-4):155–161, 1998.

[4] A. Arora, M. Pandey, V. N. Mishra, R. Kumar, P. K. Rai, R. Costache, M. Punia, and L. Di. Comparative evaluation of geospatial scenario-based land change simulation models using landscape metrics. Ecological Indicators, 128:107810, 2021.

[5] P. E. Black. Watershed functions 1. JAWRA Journal of the American Water Resources Association, 33(1):1–11, 1997.

[6] C. Brunsdon, S. Fotheringham, and M. Charlton. Geographically weighted regression. Journal of the Royal Statistical Society: Series D (The Statistician), 47(3):431–443, 1998.

[7] B. Burkhard, I. Petrosillo, R. Costanza, et al. Ecosystem services–bridging ecology, economy and social sciences. Ecological complexity, 7(3):257, 2010.

[8] S. A. Cushman and K. McGarigal. Landscape metrics, scales of resolution. In Designing green landscapes, pages 33–51. Springer, 2008.

[9] P. Dansereau. Inscape and landscape. Columbia University Press, 1973.

[10] J. N. DiBari. Evaluation of five landscape-level metrics for measuring the effects of urbanization on landscape structure: the case of tucson, arizona, usa. Landscape and Urban Planning, 79(3-4):308-313, 2007.

[11] W. E. Dramstad, J. D. Olson, and R. T. Forman. Landscape ecology principles in landscape architecture and land-use planning. Number Sirsi) i9781559635141. 1996.

[12] F. E. Egler. Vegetation as an object of study. Philosophy of science, 9(3):245-260, 1942.

[13] A. Farina. Landscape ecology in action kluwer academic. Dordrecht, The Netherlands, 2000.

[14] A. Farina. Principles and methods in landscape ecology: towards a science of the landscape, volume 3. Springer Science & Business Media, 2008.

[15] C. K. Feld, F. de Bello, R. Bugter, U. Grandin, D. Hering, S. Lavorel, O. Mountford, I. Pardo, M. P ̈artel, J. R ̈ombke, et al. Assessing and monitoring ecosystems-indicators, concepts and their linkage to biodiversity and ecosystem services. d4. 1 review paper on ecological indicators. 2007.

[16] R. T. Forman. Some general principles of landscape and regional ecology. Landscape ecology, 10(3):133-142, 1995.

[17] A. S. Fotheringham, C. Brunsdon, and M. Charlton. Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons, 2003.

[18] S. Frank, C. F ̈urst, L. Koschke, and F. Makeschin. A contribution towards a transfer of the ecosystem service concept to landscape planning using landscape metrics. Ecological indicators, 21:30-38, 2012.

[19] A. E. Frazier and P. Kedron. Landscape metrics: past progress and future directions. Current Landscape Ecology Reports, 2(3):63–72, 2017.

[20] S. E. Gergel and M. G. Turner. Learning landscape ecology: a practical guide to concepts and techniques. Springer, 2017.

[21] E. G¨okyer. Understanding landscape structure using landscape metrics. IntechOpen, 2013.

[22] B. Goodman and S. Flaxman. European union regulations on algorithmic decision-making and a "right to explanation". AI magazine, 38(3):50–57, 2017.

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

[24] R. Haines-Young and M. Chopping. Quantifying landscape structure: a review of landscapeindices and their application to forested landscapes. Progress in physical geography, 20(4):418–445, 1996.

[25] M. J. Hammond, A. S. Chen, S. Djordjevi´c, D. Butler, and O. Mark. Urban flood impact assessment: A state-of-the-art review. Urban Water Journal, 12(1):14–29, 2015.

[26] M. Herold, H. Couclelis, and K. C. Clarke. The role of spatial metrics in the analysis and modeling of urban land use change. Computers, environment and urban systems, 29(4):369– 399, 2005.

[27] M. Herold, J. Scepan, and K. C. Clarke. The use of remote sensing and landscape metrics to describe structures and changes in urban land uses. Environment and planning A, 34(8):1443–1458, 2002.

[28] M. Hu, Z. Li, J. Wang, L. Jia, Y. Liao, S. Lai, Y. Guo, D. Zhao, and W. Yang. Determinants of the incidence of hand, foot

and mouth disease in china using geographically weighted regression models. PloS one, 7(6):e38978, 2012.

[29] J. A. Jaeger. Landscape division, splitting index, and effective mesh size: new measures of landscape fragmentation. Landscape ecology, 15(2):115-130, 2000.

[30] W. Ji, J. Ma, R. W. Twibell, and K. Underhill. Characterizing urban sprawl using multi-stage remote sensing images and landscape metrics. Computers, Environment and Urban Systems, 30(6):861-879, 2006.

[31] H. W. Kim and Y. Park. Urban green infrastructure and local flooding: The impact of landscape patterns on peak runoff in four texas msas. Applied geography, 77:72-81, 2016.

[32] J. Kim and C. D. Ellis. Determining the effects of local development regulations on landscape structure: Comparison of the woodlands and north houston, tx. Landscape and Urban Planning, 92(3-4):293-303, 2009.

[33] J. A. Kupfer. Theory in landscape ecology and its relevance to. The SAGE handbook of biogeography, page 57, 2011.

[34] J. A. Kupfer. Landscape ecology and biogeography: rethinking landscape metrics in a postfragstats landscape. Progress in physical geography, 36(3):400-420, 2012.

[35] S. Lang and T. Blaschke. Landschaftsanalyse mit GIS, volume 8347. Ulmer Stuttgart, 2007.

[36] A. B. Leitao and J. Ahern. Applying landscape ecological concepts and metrics in sustainable landscape planning. Landscape and urban planning, 59(2):65-93, 2002.

[37] Y.-P. Lin, N.-M. Hong, P.-J. Wu, C.-F. Wu, and P. H. Verburg. Impacts of land use change scenarios on hydrology and

land use patterns in the wu-tu watershed in northern taiwan. Landscape and urban planning, 80(1-2):111-126, 2007.

[38] D. Lindenmayer, R. J. Hobbs, R. Montague-Drake, J. Alexandra, A. Bennett, M. Burgman, P. Cale, A. Calhoun, V. Cramer, P. Cullen, et al. A checklist for ecological management of landscapes for conservation. Ecology letters, 11(1):78-91, 2008.

[39] Z. C. Lipton. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. Queue, 16(3):31-57, 2018.

[40] B. Lu, M. Charlton, C. Brunsdon, and P. Harris. The minkowski approach for choosing the distance metric in geographically weighted regression. International Journal of Geographical Information Science, 30(2):351-368, 2016.

[41] M. Luck and J. Wu. A gradient analysis of urban landscape pattern: a case study from the phoenix metropolitan region, arizona, usa. Landscape ecology, 17(4):327-339, 2002.

[42] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In Proceedings of the 31st international conference on neural information processing systems, pages 4768-4777, 2017.

[43] J. Luo, P. Du, A. Samat, J. Xia, M. Che, and Z. Xue. Spatiotemporal pattern of pm2. 5 concentrations in mainland china and analysis of its influencing factors using geographically weighted regression. Scientific reports, 7(1):1-14, 2017.

[44] K. McGarigal. Fragstats: Spatial pattern analysis program for categorical maps. computer software program produced by the authors at the university of massachusetts, amherst. http://www. umass. edu/landeco/research/fragstats/fragstats. html, 2002.

[45] K. McGarigal. Landscape pattern metrics based in part on

the article “landscape pattern metrics” by kevin mcgarigal, which appeared in the encyclopedia of environmetrics. Encyclopedia of environmetrics, 2013.

[46] K. McGarigal. Fragstats help. University of Massachusetts: Amherst, MA, USA, page 182, 2015.

[47] J. M. Meik and A. M. Lawing. Considerations and pitfalls in the spatial analysis of water quality data and its association with hydraulic fracturing. In Advances in Chemical Pollution, Environmental Management and Protection, volume 1, pages 227-256. Elsevier, 2017.

[48] Z. Naveh and A. S. Lieberman. Landscape ecology: theory and application. Springer Science & Business Media, 2013.

[49] J. Peng, Y. Wang, Y. Zhang, J. Wu, W. Li, and Y. Li. Evaluating the effectiveness of landscape metrics in quantifying spatial patterns. Ecological Indicators, 10(2):217-223, 2010.

[50] N. A. Rupke. Alexander von Humboldt: a metabiography. University of Chicago Press, 2008.

[51] S. Schumm. Geomorphic thresholds and complex response of drainage systems. Fluvial geomorphology, 6:69-85, 1973.

[52] Z. Shao, H. Fu, D. Li, O. Altan, and T. Cheng. Remote sensing monitoring of multi-scale watersheds impermeability for urban hydrological evaluation. Remote Sensing of Environment, 232:111338, 2019.

[53] M. Su, Y. Zheng, Y. Hao, Q. Chen, S. Chen, Z. Chen, and H. Xie. The influence of landscape pattern on the risk of urban water-logging and flood disaster. Ecological Indicators, 92:133-140, 2018.

[54] J. Tang, L. Wang, and Z. Yao. Analyzing urban sprawl

spatial fragmentation using multitemporal satellite images. GIScience & Remote Sensing, 43(3):218-232, 2006.

[55] P. Teilhard de Chardin. Man’ s place in nature; the human zoological group. 1966.

[56] C. Troll and E. Fischer. Geographic science in germany during the period 1933-1945: A critique and justification. Annals of the Association of American Geographers, 39(2):99-137, 1949.

[57] M. G. Turner. Spatial and temporal analysis of landscape patterns. Landscape ecology, 4(1):21-30, 1990.

[58] M. G. Turner, R. H. Gardner, R. V. O’ neill, and R. V. O’ Neill. Landscape ecology in theory and practice, volume 401. Springer, 2001.

[59] R. E. Turner, N. N. Rabalais, D. Justic, and Q. Dortch. Global patterns of dissolved n, p and si in large rivers. Biogeochemistry, 64(3):297-317, 2003.

[60] E. Uuemaa, M. Antrop, J. Roosaare, R. Marja, and ¨U. Mander. Landscape metrics and indices: an overview of their use in landscape research. Living reviews in landscape research, 3(1):1-28, 2009.

[61] V. I. Vernadsky. The biosphere and the noosphere. American Scientist, 33(1):1-12, 1945.

[62] A. Vink. Development of land use in advancing agriculture. In Land Use in Advancing Agriculture, pages 327-369. Springer, 1975.

[63] D. C. Wheeler and A. P´aez. Geographically weighted regression. In Handbook of applied spatial analysis, pages 461-486. Springer, 2010.

[64] J. Wu, W. Sha, P. Zhang, and Z. Wang. The spatial non-stationary effect of urban landscape pattern on urban waterlogging: a case study of shenzhen city. Scientific reports, 10(1):1-15, 2020.

[65] F. Xue, M. Huang, W. Wang, and L. Zou. Numerical simulation of urban waterlogging based on flood area model. Advances in Meteorology, 2016, 2016.

[66] Y. Yao, S. Zhang, Y. Shi, M. Xu, J. Zhang, Y. Zhang, and J. Zhao. Landscape pattern change of impervious surfaces and its driving forces in shanghai during 1965-2010. Water, 13(14):1956, 2021.

[67] H. Yu, Y. Zhao, Y. Fu, and L. Li. Spatiotemporal variance assessment of urban rainstorm waterlogging affected by impervious surface expansion: A case study of guangzhou, china. Sustainability, 10(10):3761, 2018.

[68] H. Zhang, J. Cheng, Z. Wu, C. Li, J. Qin, and T. Liu. Effects of impervious surface on the spatial distribution of urban waterlogging risk spots at multiple scales in guangzhou, south china. Sustainability, 10(5):1589, 2018.

[69] Q. Zhang, Z. Wu, G. Guo, H. Zhang, and P. Tarolli. Explicit the urban waterlogging spatial variation and its driving factors: The stepwise cluster analysis model and hierarchical partitioning analysis approach. Science of The Total Environment, 763:143041, 2021.

[70] Q. Zhang, Z. Wu, and P. Tarolli. Investigating the role of green infrastructure on urban waterlogging: Evidence from metropolitan coastal cities. Remote Sensing, 13(12):2341, 2021.

[71] Q. Zhang, Z. Wu, H. Zhang, G. Dalla Fontana, and P. Tarolli. Identifying dominant factors of waterlogging events in metropolitan coastal cities: The case study of guangzhou, china.

Journal of Environmental Management, 271:110951, 2020.

[72] I. S. Zonneveld. The land unit—a fundamental concept in landscape ecology, and its applications. Landscape ecology, 3(2):67–86, 1989.

초 록

지리 가중 회귀모형 및 섀플리 가법 설명모형에 의한 지역침수 영향요인 분석

XIAOLING JIN
환경대학원 환경조경학과
환경조경학 전공

경관은 생태계 개입의 핵심 요소로 꼽힌다. 인류의 활동은 지표면의 특징을 크게 변화시키고 있으며, 자연 물질과 에너지의 순환과 흐름에 영향을 주어 유역에 빗물을 모으는 기능과 경류배수의 능력을 약화시켜 침수 재해의 발생을 초래하고 생활환경의 위험을 증가시킨다. 따라서 경관계획가와 정책결정자는 생태계의 동적 균형을 유지하기 위해 경관구조의 최적화를 끊임없이 개선하여 침수를 완화하는 목적을 달성할 필요가 있다. 원격 탐사 기술의 발달로 대규모 유역 단위 연구가 가능해졌으며, 이러한 대규모 현장에서의 실험은 이론으로 검증될 수 있다. 이론 검증에 대한 과거의 연구는 지리 가중 회귀 모델(GWR)와 같은 전통적인 선형 회귀 모델(지도 학습)은 독립변수과 종속변수간의 관계를 분석하면서 독립 변수 간의 관계를 분석할 수 없기 때문에 경관패턴 내의 상호 작용을 무시했다. 최근 머신러닝 분야에서 해석 가능한 머신러닝 모델의 발전이 이러한 단점을 보완하고 있다. 이 중 섀플리 가법 설명모형(SHAP)은 게임 이론에 기반한 해석 가능한 기계 학습 모델의 대표이다. 독립변수와 종속변수의 관계를 분석할 수 있을 뿐 아니라 여러 독립변수의 상관관계를 고려해 기여도에 따른 중요도 순위를 얻을 수 있다. 두 가지 방법의 검증 및 비교 분석을 통해 GWR의 분석결과에서 섀넌 다양성 지수(SHDI)가 심각하게 과소평가된 반면, SHAP 결과에서 SHDI는 모든 규모의 유역 단위에서 침수에 큰 영향을 미친다는 것을 알 수 있다. 또한 예측 평균 제곱 오차(MSE)의 예측 결과에 따르면 GWR의 오차 값은 작지만 SHAP가 GWR보다 훨씬 정확하다. 둘째, 물 순환 과정은 다단계 지리적 유역을 생성하는 특성을 가지고 있다. 수문학의 동적 균형을 실현하기 위해서는 다단계 유역 규모 단위의 비교 분석이 필요하며, 그 결과는 더 미세한 유역을 연구 규모로서 사용하는 것이 반드시 수문 연구에 적합하지 않음을 보여준다. 본 연구에서는 대규모 유역단위(LSWU)를 기반으로 한 수도권(SCA)의 침수 연구가 가장 적절하고 정확하다. 마지막으로 경관패턴 특징은 임계치가 존

재한다. 침수에 대한 영향이 임계점에 도달했을 때, 침수를 촉진하거나 완화하는 작용이 변화한다. 경관패턴 특징의 임계치를 통해 정확하고 저비용으로 침수 재해를 완화하는 목적을 달성할 수 있다. 본 연구는 경관패턴와 침수간의 상호작용 분석방법에 대하여 새로운 탐구를 진행하여 경관생태학에 기초한 침수 완화방법과 결과를 참고로 제공한다.

....

주요어: 침수, 경관 패턴, 수도권, 지리 가중 회귀모형, 섀플리 가법 설명모형
학 번: 2020-20489