Master’s Thesis of Landscape Architecture

A simulation approach to the impact assessment of urban development projects on vegetation ecotone

도시개발사업이 주연부 식생에 미치는 영향 평가에 관한 시뮬레이션 연구

February 2023

Graduate School of Seoul National University
Department of Landscape Architecture and Rural systems Engineering, Landscape Architecture Major

Hyewon Kang

A simulation approach to the impact assessment of urban development projects on vegetation ecotone

Under the Direction of Adviser, Prof. Dong Kun Lee

Submitting a master's thesis of Landscape Architecture

January 2023

Graduate School of Seoul National University
Department of Landscape Architecture and Rural systems Engineering, Landscape Architecture Major

Hyewon Kang

Confirming the master's thesis written by Hyewon Kang
January 2023

Chair ____________________(Seal)

Vice Chair ____________________(Seal)

Examiner ____________________(Seal)

Abstract

Patches are recognized as ecotones as transition zone between adjacent patches that exhibit heterogeneity due to differences in vegetation conditions. Ecotones play an important role in environmental ecology by providing high biodiversity, ecosystem connectivity, and diverse habitat environments. Since South Korea has experienced spatial changes in patches due to rapid industrialization and urbanization, preservation of ecotone depends on confirming the impact of human activities on the natural environment. Therefore, in order to devise sustainable management measures, we tried to monitor the ecotone vegetation dynamics that change due to urbanization and evaluate the extent of impact. This study proposed an impact assessment tool to predict and quantify the range that changes under the influence of urban development projects according to the set peripheral distances (25, 50 m, and 100 m). Normalized Difference Vegetation Index (NDVI) and Vegetation Health Index (VHI) were selected as indices for evaluating the main effects, as well as Landsat and Sentinel-based satellite imagery data were calculated through the Google Earth Engine platform (GEE). Land cover maps provided by the Environmental Spatial Information Service as well as average temperature and precipitation data of the Korea Meteorological Administration were constructed through ArcGIS 10.5. National inventory data of the Environmental Impact Assessment Information Support System (EIASS) were processed and applied as variables. The data analysis method evaluated the vegetation distribution patterns of the research sites using the

Artificial Neural Network (ANN) and Random Forest (RF) machine learning algorithms, and it was set to predict the range of influence on the vegetation index according to the ecotone multiple buffer size. As a result of the analysis, it was confirmed that NDVI was mainly concentrated on high values after the urban development project, while VHI tended to have a high pre-project value, which turned to the opposite trend. This can be interpreted as a significant result of the establish of new urban green spaces in accordance with the provisions of the Act on the Expansion, Management, and Creation of Urban Green Areas for Urban Landscape Planning. As a result of the performance of the machine learning models, the RF model showed the optimal predictive performance in both vegetation indices and along the ecotone distances. The modeled probability heatmap shows significant results at 90% confidence level ($p<10\%$). Moreover, significant results were obtained when comparing the observed and predicted values visualized using the assessment tool. Both NDVI and VHI showed the tendency of the impact of the target site due to urban development to reach a maximum distance of 50 m. This proposal of quantitative evaluation tools is meaningful in that it may emphasize the decisive role of environmental impact assessment in terms of vegetation management by providing information on regional ecological restoration. It is expected that the extent of impact on the vegetation environment by urbanization can be identified to support the project plan while minimizing the loss of vegetation cover.

Keyword: Environmental impact assessment, Land cover change, Vegetation index, Edge effect, Machine learning, Remote sensing
Student Number: 2021-26769

Table of Contents

Chapter 1. Introduction

1.1. Study Background

Patches are defined as two or more habitat type boundaries, and transition zone between adjacent patches that exhibit heterogeneity due to differences in vegetation conditions are recognized as ecotones (Holland, 1988). The edges generally have different species compositions and structures compared to the innermost area; a phenomenon known as the edge effect (Fraver, 1994). The ecotones include much more diverse organisms, conjoining distinct habitats and increase species richness, diversity and abundance in various taxonomic groups, such as vegetations (Łuczaj & Sadowska, 1997). Besides, they serve as a bridge of community to the flow from one group to another as well as function as buffer zones to protect the periphery of ecosystems from natural and potential disasters. Therefore, Ecotone is essential in environmental and ecological research.

Urban patches fragmented into small in size, with edges owing to the anthropogenic disturbances as well as natural changes in the environment (McKinney, 2006). As rapid urbanization, the more accelerated development activities progress, the more extensive natural environment alters over the past decades (MEA, 2005). The acquired patches generated by human-modified landscapes crucially produce multiple peripherals. Previous studies have shown that maintaining numerous fragmented small patches is generally more beneficial for biodiversity conservation than maintaining large patches (Arroyo-Rodríguez et al., 2020). In other words, when significant, the responses to isolated patches per se are positive (Fahrig, 2017). In addition, it is considered by researchers that in terms of environmental heterogeneity, many small patches are easier to manage various soil types than some large patches (Phalan, 2018). Fragmented patches can achieve biodiversity conservation and species coexistence by preventing the spread of competing species, unlike broad single patches (Hernández-Ruedas et al., 2018). Therefore, it is a crucial task to conserve and increase the greenness of isolated small in size patches in ecosystems (Benchimol & Peres, 2013; Carrara et al., 2015; Morante-Filho et al., 2018; Phalan, 2018; Arce−Peña et al., 2019; Galán-Acedo et al., 2019).

Sustainable ecotone conservation depends on identifying the influences of natural environment to human activities. Evaluating edge effects is imperative to enhance ecological understanding and develop conservation strategies and management in landscape ecology. It is considered that research is needed to understand the effect of the ecosystems on the periphery of human−modified patches by the development activities. Hence, monitoring vegetation greenness is needed to minimize vegetation degradation. To monitor the vegetation dynamics of terrestrial ecosystems and to devise sustainable management measures, remote sensing and geographic information system are generally used when evaluating land cover and land cover changes along with related planning tasks.

The methodology discussed in this study is based on the application of remote sensing imagery in vegetation. The relationship between vegetation indices and environmental factors is increasingly important in ecological research (Liu et al., 2019; Peng et al., 2019) and the time-series data have been widely used (Chen, 2021). The vegetation Health Index (VHI) is a widely used remote sensing-based index designed as the weighted sum of two components: the Vegetation Condition Index (VCI) and the Thermal Condition Index (TCI). The first component characterizes moisture conditions and is typically based on information from the visible and near infra-red windows of the electromagnetic spectrum, whereas the latter characterizes the thermal condition and is based on information from the thermal infra-red window. The Normalized Difference Vegetation Index (NDVI) and Land Surface Temperature (LST) or TOA brightness temperature are commonly used to estimate VCI and TCI, respectively. The NDVI, which analyze based on satellite imagery data, is used to determine vegetation distribution and evaluate productivity (Xiao & Moody, 2004; Evans et al., 2006). The NDVI as an indicator has been widely used to evaluate the vegetation growth condition within a pixel basis, reflecting the vegetation growth and coverage status at spatial and temporal scale (Zhi-qiang, & Dennis, 2001; Wan et al., 2004). The VHI was also effective in monitoring vegetation health and stress as well as has been widely used to estimate crop productivity and biomass evaluation (Kogan, 1990). Therefore, it is believed that it can be effectively used to set a homogeneous unit as an index.

1.2. Purpose of Research

The purpose of this study was to evaluate the extent to which urban development projects affect the ecotones along three different magnitude and extent. An impact assessment model was developed that can analyze how vegetation ecology altered and quantitatively evaluate the range of its impact. The NDVI and VHI based Landsat and Sentinel data were extracted in consideration of the spatiotemporal range of the target project, and the analysis conditions were established by applying appropriate environmental variables using national inventory data. An optimal machine learning algorithm comparing the performance of artificial neural networks (ANN) and random forest (RF) models was applied to evaluate the damage and the extent of impact of the ecotone. The simulation approach of this study can be a means to minimize the expected potential side effects by predicting and analyzing environmental impacts that may occur in the development project process in advance.

Chapter 2. Literature Review

Edge effects which caused by patches fragmented, are involved in the effect of species composition, abundance and richness of ecosystem. Despite the importance as the advantages of ecotone, insufficient researchers have studied the influences of urbanized sites and edge effects on the ecosystems of remnant green spaces (Guerra et al., 2017). Silva et al. (2018) evaluated the impact on the edge effect on ferns of two tropical lowland rainforests and montane forests in Mexico by collecting the edaphic parameters through plot sampling. The results detected the factors causing biodiversity loss in ecotones and provide the indication. Aragón et al. (2015) investigate the correlation of edge effects with epiphytes of structural and spatial properties by collecting data via experimental design for each type of edge. On the other hand, Batáry et al. (2014) examined how forest edges and tree diversity affect bird populations, breeding, and survival rates based on field experiments. The results of this study tried to prove the effect of the periphery, and through this, the preservation plan was expected. As a result of the review, most studies emphasized the importance of the edge effects, but only the current status survey proved the validity of the edge effect through the investigation. Data were collected via experiments, and through this, they tried to prove the effect of the periphery. However, crucially, there were significantly few studies on the edge effects related to vegetation cover. Besides, Studies that simulate and predict specific environmental conditions by simulating the ecotone as well as emphasize the importance of monitoring the peripheral effect and suggesting them as an evaluation tool were shown rarely.

Besides, Edge effects are the dominant characteristics of human−modified patches and are the subject that require empirical research. However, the development of methods for effective management is being delayed. As a series of contributions to this, Ewers et al. (2006) developed a statistical model to quantify the variation in response among the edges to measure the extent. As a statistical approach to quantify the strength of edge effects, proposing a model that can determine the size and range of edge effects. It is argued that this could provide an essential management tool for monitoring changes in land cover or changes in edge effects after habitat restoration. On the other hand, Wilson et al. (2014) modeled to test evidence for the edge effects of land cover and association of topography with bird habitats. Variable data was constructed based on breeding surveys, generalized linear models were used for occupancy modelling. Given support for vegetation expansion to support climate change mitigation and policy, this study is timely. Ahmadi et al. (2020) applied variables of forest edges to machine learning algorithms to estimate forest characteristics. The major objectives were to predict the most common variables necessary for sustainable forest management, and to evaluate the indices quantitatively. When reviewed edge effect research conjoined with simulation methods, they were related to research developing quantitative tools, or developing evaluation indicators. However, the studies have investigated terrestrial ecosystem such as community of organisms and interactions of biotic and abiotic components.

Subsequently, to the best of our knowledge, there was also a lack of research on the status of vegetation changes due to changes in land cover, and no research was found to predict future situations that will change according to artificial activities in the future.

Moreover, monitoring can be done accurately with lower costs if the vegetation index is extracted from the remote sensing imagery. This study reviewed previous studies aimed at monitoring performed using machine learning approaches with vegetation indices information obtained from satellite. Park et al. (2016) investigated a meteorological drought index (SPI) and an agricultural drought index based on the various factors such as NDVI and VHI, using machine learning approaches. The approach in this study is applicable to all vegetation areas where remote sensing data is available. Moradi et al. (2022) mapped quantitative vegetation vulnerabilities through machine learning algorithms based on remote sensing data. In addition, a quantification evaluation method was proposed by discovering hydrological influencing variables. Most of the machine learning-based vegetation distribution analysis was dominated by studies aimed at drought monitoring or agricultural yield evaluation. On the other hand, Bao et al. (2021) quantified sensitivity of vegetation covers to climate change. The relationship between the spatial distribution of NDVI sensitivity to climate change was analyzed by constructing an ensemble statistical vegetation model based on meteorological and geographic data. This may be useful in providing a quantitative response to climate vegetation. These results have something in common with this study in that they provide quantitative tools as well as in-depth understanding to provide mitigation strategy guidance to policymakers in prediction and modeling but have different goals for response strategies based on the correlation between land cover and vegetation change.

In order to quantitatively evaluate changes in vegetation affected by development activities, it is essential to investigate changes in vegetation cover for complex factors in addition to topographical and geographical variables. Until now, vegetation zones have evaluated the factors affecting the vegetation indices, and green cover conditions such as NDVI and VHI have been identified. For impact assessment, researchers conducted field measurement campaigns or used statistical analysis. However, considering the peripheral effect of ecotones with biodiversity, the strategy according to vegetation changes is still insufficient, and further research needs to be conducted. In addition, it is necessary to simulate changes in the vegetation ratio according to changes in the natural environment or land cover due to artificial activities and predict changes according to future plans. So as to overcome these limitations, this study proposed simulation indicators using machine learning algorithms. This is not only meaningful in understanding the mechanisms of the vegetation system, but also useful in planning and managing land cover.

Chapter 3. Materials and Methods

3.1. Study flow

The basic approach of the research flow consists of two main parts: data collection including preprocessing and environmental impact assessment (Figure 1). First of all, in collecting the data, the current status of the evaluation report and the impact evaluation report became the basis for the research framework. It was based on ecotone setting according to the existing urban development project, and assorted variables were set as analysis targets. The database was constructed based on the data provided by the government. In addition, NDVI and VHI indices, which are mainly used to identify vegetation distribution and evaluate productivity, were selected to analyze greenness in the surrounding area. Also, we inspected the accuracy of two different machine learning algorithms to evaluate vegetation vitality before and after development projects for ecotones. As a data analysis method, the vegetation distribution characteristics of the research area were analyzed using ANN and RF algorithms, investigating vegetation changes and major factors. Subsequently, the effect of changes in land cover on the periphery was evaluated.

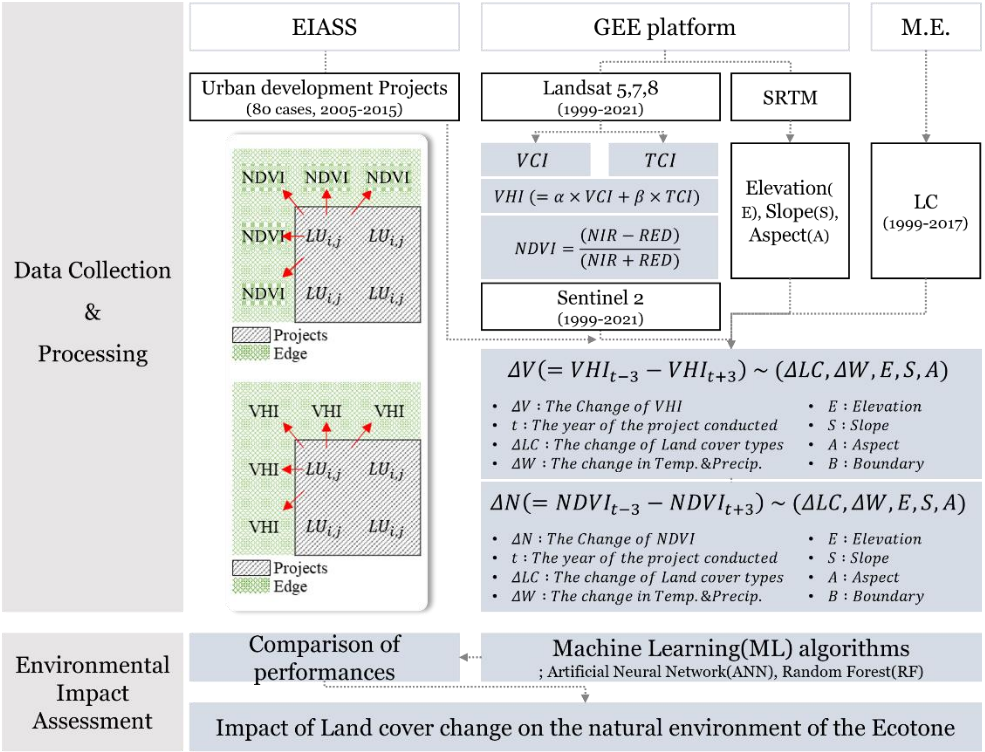

Figure 1 Research flow

3.2. Study scope

This study was aimed at environmental impact assessment projects, focused on urban development in South Korea. The assessment reports were reviewed to investigate the actual condition of environmental impact assessment and to set the scope of the research subject as land cover on terrestrial ecosystem are being converted to non−forest such as agriculture, energy generation and other infrastructure (Curtis et al., 2018). Urban planning in Seoul, the capital of South Korea, has been active since the 1980s (Liao & Pitts, 2006). Likewise, the Ministry of Land, Infrastructure and Transport reported that housing demand exploded in the 1980s and

1990s due to the increase in population density of cities and housing supply also increased, so that citizens experienced frequent land cover changes in the past few decades. By considering the temporal scale, development projects that occurred in the 1990s and 2000s were reviewed (Table 1). Compared to 1950, the number of houses in 2011 increased by about 5.5 times, and the housing penetration rate already exceeded 100% in 2000. In the case of housing site development, the number of projects has declined since the 1990s, but the number has still been high for 2 decades. This is the result of various housing-related development policies such as modern housing construction projects, housing construction projects, and new city construction projects. As the result, it shows that land development, urban development and housing complex development were the main trends in the period. The past changes in Korea's national territory can be summarized as government-led national land development projects, urbanization, and industrialization, and the urban development projects can be seen as a newly rising business keyword of 2000s.

Table 1 Urban development trends in the 1990s and 2000s

Type	Keyword	Number(90s)	Number(00s)	*YoY
Urban planning	Land	153(50%)	123(36%)	-30
	Urban	0	86(25%)	86
	Housing site	0	28(8%)	28
	Sewage treatment	47(15%)	0	-47
	etc	40	91	
	Total	307	346	

*YoY: Year on year

Even though land cover information and plans for the target sites where the project takes place have been established, there is a lack of information and investigation for the possibility of influence on adjacent areas and ecotones. Besides, it was found that the environmental impact assessment was analyzed as a qualitative assessment without scientific basis. Accordingly, research is needed to develop quantitative indicators. From that, the spatial scope of this study was established based on the impact assessment reports provided by the Ministry of Environment. Based on the tendency of the assessment reports in the 1990s and 2000s, urban development projects which was the most dominant, were focused on.

This study aims to detect the changed environment in the 2000s after urban development projects that of 1990s. Therefore, the temporal range was set from 2005 to 2015, five years after the actual project of the impact assessment report occurred, based on the long-term project period. The temporal range was also set to compare before and after environmental changes according to developmental activities, applying of variables such as land cover maps. After classifying the types of urban development, the projects were refined with the keywords of the urban and housing complex.

Land cover before the development projects was confirmed using land cover mapping, along with in the case of urban or agriculture areas, which are difficult to examine the characteristics of vegetation, they were deleted from the study sites. This is a process that can infer the extent of vegetation green space of the target site, and areas without green space before the project cannot produce significant results, so it is excluded from the project list. The green area ratio was calculated for the refined study site. The median value (18.50%) or less for the ratio of green areas to all target sites was deleted, and only development projects including more green areas were considered. The progress on areas containing forests and grassland was extracted using ArcGIS 10.5 geoprocessing clip feature. In addition, classification was conducted according to the green area to evaluate the contribution of the ratio of this green area. Accordingly, classification was conducted to evaluate the contribution of the ratio of the green area ratio (Table 2). Weighted by increasing the median value to a minimum of 10% intervals.

Table 2 Grading according to the green area ratio of the study sites

Percentage of vegetation cover	Grading
18 ~ 27%	1
28 ~ 37%	2
38 ~ 47%	3
48 ~ 57%	4
58 ~ 100%	5

Since all sites with ecotone land cover changes were deleted during the target period, only sites with the same land cover code for 10 years from 2006 to 2015 were included in the analysis. This can only consider the effect of internal changes in the target site as one of the analysis variables, and it can be seen that the effect of the external ecotone itself has been removed. Five main steps were performed to refine the study target site, such as the procedure for the scope of the study (Table 3). The study target period was initially set from 2005 to 2015, but the remaining business after a series of refining processes was actually between 2006 and 2015. We considered that the actual execution and completion of the project would be much later than the reception and completion date stated in the Impact Assessment because construction would not begin immediately after the assessment. To compare and analyze the status before and after urban development projects according to land cover changes, three points of view were used for land cover map data. The land cover before the project was set to 2006, and the land cover map of 2011, which is the median value of the project target period, was used. And after the project, it was set to about 5 years, which is the period required to complete the project after designating the urban development zone and set to 2020. This was applied in the same way as the five years set when setting the 2000s, an era when urbanization occurred and stabilized. In addition, this study found the most suitable setting related to it because the purpose of this study is not to detect changes, but to evaluate the effects of surrounding nature before and after the project occurs. The detailed illustration of the cover shows the land cover classification map code. In conclusion, a total of 80 sites were deployed (Figure 2).

Table 3 Refining process against study site selection

Division	Detail
Keyword	Urban and housing
Period	2005 to 2015
Vegetation cover	310 and 320 for Forest area 410 and 420 for Grassland
Grading	Median value (18.50%) and above
Land cover change	Changes in land cover of 25 m inside the study sites

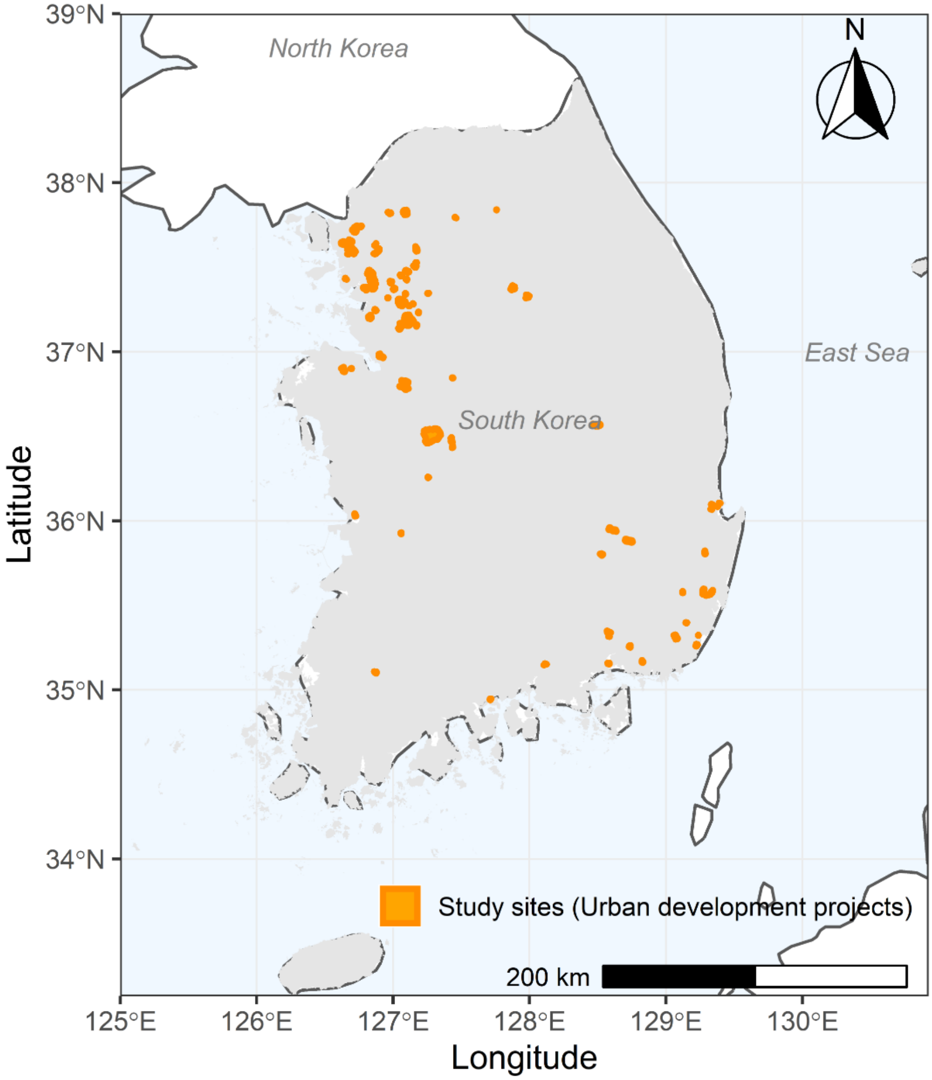

Figure 2 Study sites

The Ecotones conjoin with various habitats and increase species richness and diversity in various taxonomic groups, such as plants (Łuczaj & Sadowska, 1997). The spatial extent of edge effect may be narrow or wide, and edge effect has rarely been quantified (Łuczaj & Sadowska, 1997) among taxonomy. Besides, the scope and impact of the development project periphery may vary depending on the characteristics of the space. As to the characteristics, the spatial range of the ecotone may be contrasting. Hence, in this research, the spatial extent of periphery analyzed from various angles.

The most dominant peripheral distance was a constant multiple (Table 4). The most dominant range setting was to evaluate the effect of forest edge size on the abundance of forest fragments by sampling at four distances: 5, 25, 50 m and 100 m (Rheault et al., 2003; Esseen et al., 1998). On the other hand, Sampaio et al. (2011) established six distances (0, 40, 80, 160, 280 m and 400 m) from the edge and studied composition and diversity of tree community. In the study of Esseen et al. (2019), the edge effects of 25, 50 m and 100 m were estimated then reported that maximum edge effects extended to 50 m at moderately exposed sites. The commonality of the previous research was evaluation of the periphery through the expansion of the ecotone buffer.

Table 4 The effect of edge distances (unit: meters)

Reference	Transect
Alignier et al., 2013	40
Aragón et al., 2015	100
Rheault et al., 2003	5/25/50/100
Kivistö et al., 2000	10/20/50/100
Esseen et al., 1998	5/25/50/100
Batáry et al., 2014	210
Łuczaj et al., 1997	50
Sampaio et al., 2011	0/40/80/160/280/400
Magura et al., 2002	39
Esseen et al., 2019	25/50/100

In addition, the New Zealand Conservation Authority defined that many patches of greenness today are small in size, and environmental conditions limit the range of green patches for up to 60 m, creating an edge effect.

In this study, the distance of multiple 3 of 25, 50, and 100m was set as the edge effect distance based on the 100m boundary where the edge effect was reported the most.

3.3. Sources of Data Collection

The variables used to achieve the evaluation tool are divided into two main variables: the independent variable and the dependent variable. In the machine learning algorithms, a total of 7 environmental factors and vegetation indices are used as independent variables as well as dependent variables, respectively, and variables were set through literature review. The dependent variables, which are the main factors of the analysis, were placed in the analysis including 80 development projects provided by environmental impact assessment along with NDVI and VHI (Table 5). Other independent variables comprise three ecotones according to distance, land cover data, topographic variables, and climate variables (Table 6).

Meteorological effects on NDVI, especially precipitation and temperature, can predict vegetation production changes (Wang, 2003). As weather parameters, data from 1999 to 2001, 2009 to 2009 and 2019 to 2021 were used. Climate data were obtained from the Meteorological Agency (https://data.kma.go.kr). In addition, topographic data such as altitude and slope, which are most frequently used for domestic research GIS analysis, were acquired by Shuttle Radar Topography Mission by using Landsat satellite and obtained through the GEE. Land cover changes due to urban development can be applied by applying land cover maps, which can be obtained from the Ministry of Environment (http://me.go.kr).

Table 5 Dependent data list with source and period

Type	Data	Period	Source
Dependent variable	Normalized Difference Vegetation Index (NDVI) Vegetation Health Index (VHI)	2004–2018	*GEE
	Urban development projects	2006–2015	**EIASS

*GEE: Google Earth Engine Platform (https://earthengine.google.com)
**EIASS: Environmental Impact Assessment Support System (https://www.eiass.go.kr)

Table 6 Independent data list with source and period

Type	Data		Period	Source
Independent variable	Ecotone ; 20, 50 m and 100 m		2006–2015	**EIASS
	Land cover (LC) ; Built-up, Agriculture, Forest, Grassland, Wetland, Barren, Water		2006, 2011, 2020	***ME
	Elevation Slope Aspect	Obtained from SRTM by using Landsat	1999–2018	*GEE
	Meteorological parameters ; Temperature, Precipitation		2004–2021	****KMA

***ME: Ministry of Environment (http://me.go.kr)
****KMA: Korea Meteorological Administration (https://data.kma.go.kr)

Various environmental variables and two difference vegetation indices were applied to analyze the effects of ecotone by distance at urban development project sites. The spatial aspect was supported by applying the three distances of ecotone and topographic variables such as elevation, slope, and aspect. The average temperature and precipitation, which are climate factors affecting the vegetation indices, were controlled. NDVI and VHI were evaluated by assigning a vegetation variable correlation, analyzing the temporal and spatial differences between the performance of vegetation dynamics.

3.4. Satellite Data and Preprocessing

According to the advance of remote sensing technology, research related vegetation monitoring can apply the satellite images as data sources. Satellite imagery was used as a major data source to quantitatively evaluate the possibility that urban development projects may affect vegetation ecology and change the vegetation of ecotone. Two different data sources were used to calculate the NDVI and VHI: the Landsat and Sentinel satellites. Both Landsat and Sentinel series were obtained via the Google Earth Engine Platform (https://earthengine.google.com). GEE is an advantageous cloud computing platform, processing, storing and analyzing of petabyte-scale archives of data on large amounts of remote sensing platforms (Gorelick et al., 2017). In the GEE platform, vegetation index analysis is possible. Recently, GEE has been widely used in time series data processing to extract various phenological indexes (Dong et al., 2016; Mutanga & Kumar, 2019). After creating images of 10 years according to study period with code through GEE, exported to Google Drive. The NDVI and VHI were employed in this study and obtained from the Google Earth Engine Platform(https://earthengine.google.com). NDVI and VHI were acquired by Landsat-5, Landsat 7 ETM+ and Landsat 8 OLI, at a spatial resolution of 30 m to construct, assessing differences in the study sites. Besides, cloud-mask equipped in the GEE can remove the pixels which were covered with clouds for Landsat and Sentinel data production. The 'QA60' band was applied to filter out cloudy pixels. The minimum and maximum NDVI values were chosen from NDVI values using GEE code, ranging from 0 to +1. This is the most widely adopted indicator (Coluzzi et al., 2007; Pignatti et al., 2015; Simoniello et al., 2015). The negative values are generated by factors such as clouds, water, and snow.

Sentinel-2 collects high-resolution multispectral images. Cloud masks are applied to represent clear conditions for Level-2A, which provide atmospheric floor reflection or modification with subpixel multispectral registration. Sentinel-2 launch took place in June 2015. Due to the limitation of temporal range, this research focuses on defined period that the performances achieved by Sentinel-2. Regarding Sentinel satellite data, the identical approaches as Landsat were applied. The NDVI was calculated from surface reflection Sentinel-2 Multi Spectral Instrument data collected at 10 m resolution. To mitigate cloud contamination, the average NDVI of all available Sentinel scenes across Korea was calculated during the period corresponding to the development project (a total of three years, including before and after the project year). The minimum and maximum NDVI values were ranged from -1 to +1. This result will be described in the conclusion.

The NDVI was determined as follows:

$$NDVI = \frac{NIR - RED}{NIR + RED} \qquad (1)$$

The NDVI was calculated for a decade and the minimum and maximum reflectance values of NDVI and LST were extracted to generate VCI and TCI. The VCI is derived from NDVI values (Kogan, 1995). Where NIR is the value of the near infrared band, and RED is the value of the red band, indicating 845-885 nm and 630-680 nm, respectively.

The VHI was determined as follows:

$$VHI = \alpha \times VCI + \beta \times TCI \qquad (2)$$

VHI is the sum of the weighted value of VCI and TCI reflecting temperature and vegetation conditions. From 2006 to 2016, VCI, TCI, and VHI were estimated using NDVI and LST remotely detected in domestic sites.

3.5. Hyperparameter

To quantitatively evaluate the extent to which the affected ecotone vegetation changes in the study area, we applied two machine learning methods to the NDVI, and VHI index maps. The machine learning algorithms used were ANN, and RF. The contribution of development activities to vegetation changes was quantitatively assessed through temporal and spatial based analysis.

RF is an ensemble learning algorithm for classification, regression, and other tasks that operate by constructing multiple decision trees in training. The random forest classifier, the output is the methods expressed by most decision tree models. For regression operations, the average or average prediction of an individual tree is returned. Based on the classification and regression tree (Breiman, 2001), the RF generates numerous independent trees to reach the final decision through two randomized approaches to training sample selection and variable selection at each node of the tree. MLPClassifier analyses were used to show the associations of the vegetation indices (NDVI and VHI) in the development activities with each reference distances. In this context, the ANN model has been applied as a function to model linear and highly nonlinear relationships between input and output datasets. By default, ANNs consist of one input layer, one output layer, and zero or more hidden layers used to solve complex problems. The sklearn package was employed in the program language the Python module for machine learning in this study.

To determine the extent of change on vegetation changes for the periphery of the urban development sites, an impact evaluation tool was devised. After identifying different edge distance, analysis was conducted to assess the relationships among the land cover maps, topography and meteorological variables. Hyperparameter configuration is the process of setting values for each parameter before the learning process. Comparison of model performance are categorized into training and testing. The training set is used to make network learn. The testing set is used to analyze the neural network.

To either enhance the performance and predictive power of models or to make the model faster, hyperparameters are used in random forests. Estimators that mean number of trees the algorithm builds before averaging the predictions was applied. In addition, the minimum number of samples required to split an internal node is controlled with the parameter and the whole dataset is used to build each tree. The classifier was imported and fit the data. The reliability of the ANN in the estimation was evaluated using the coefficient of accuracy (ACC) measuring the fitness between actual and predicted values and Root−Mean−Square Error (RMSE) which means sample standard deviation calculated by measuring the difference between actual and predicted values. The optimizer 'Adam' was chosen to train the models. This minimizes the RMSE between the target values and the results and sets weights and bias in the ANN.

Chapter 4. Results

4.1. Vegetation Dynamics Variations

Negative values represent water features, values close to zero indicate no vegetation, and positive values indicate the presence of green vegetation. Higher values indicate greener, more dense surface vegetation (Weier, 2000). The temporal variation of the ecotone vegetation indices extracted from the remote sensing imagery according to the influence of urban development is displayed (Figure 3). A visual analysis of the NDVI and VHI temporal dynamics indicates two different patterns associated with before and after the development projects, which are characterized by a range of values. While the NDVI confirmed the tendency to be distributed at high values after the development projects, the VHI was distributed at high values before the urban projects. The land cover transition from vegetation regions to urban areas is the main cause of the tendency of declining greenness (Yao et al., 2019), which is the main result of the VHI trend in this study. In order to evaluate the dynamics of vegetation, the time series of VHI for urban development projects was analyzed. Both vegetation indices have respective values before and after the development activities. The corresponding values were set based on the average for three years before and after the completion date of the each of the projects. The NDVI extracted from Landsat ranged from 0 to 0.83, including the three different distances. It showed NDVI values of 0.829 at 25 m, 0.831 which is the maximum value at 50 m, and 0.798 at a distance of 100 m. However, under the same conditions, the Sentinel imagery displayed different results. In the order of short distances, 0.728, 0.742 and 0.75, respectively. This is relatively low compared to the Landsat data. Parallel could be found between the NDVI and the VHI regarding tendency of decreasing value after the urban projects and spatial patterns in terms of the extent concentrated. The maximum value of VHI was 0.79 established 25 m, 0.855 at 50 m, and 0.822 at the foremost outside distance. In common, VHI and NDVI calculated using Landsat imagery presented a tendency to increase in value from 25 m to 50 m and then detracted at 100 m. This may be interpreted that the internal impact of urban development projects reaches up to 50 m.

Significantly, after urban development, distributions concentrated in high NDVI values, turning positive trends. The results occurred especially when land cover modification was made to residential land. Matters necessary for the expansion, management, use, and urban greening of green areas in cities shall be prescribed in accordance with the Urban Parks and Green Areas Act in order to create a pleasant urban environment. It is stipulated that a plan to secure green space must be included in the development plan. In addition, the green space of apartments has become abundant since 1991 when the ratio of green space to land area has been strengthened to 30%. This is a result consistent with previous studies that the greening trend in some cities was caused by construction of urban green space (Zhao et al., 2013; Zhou & Wang, 2011).

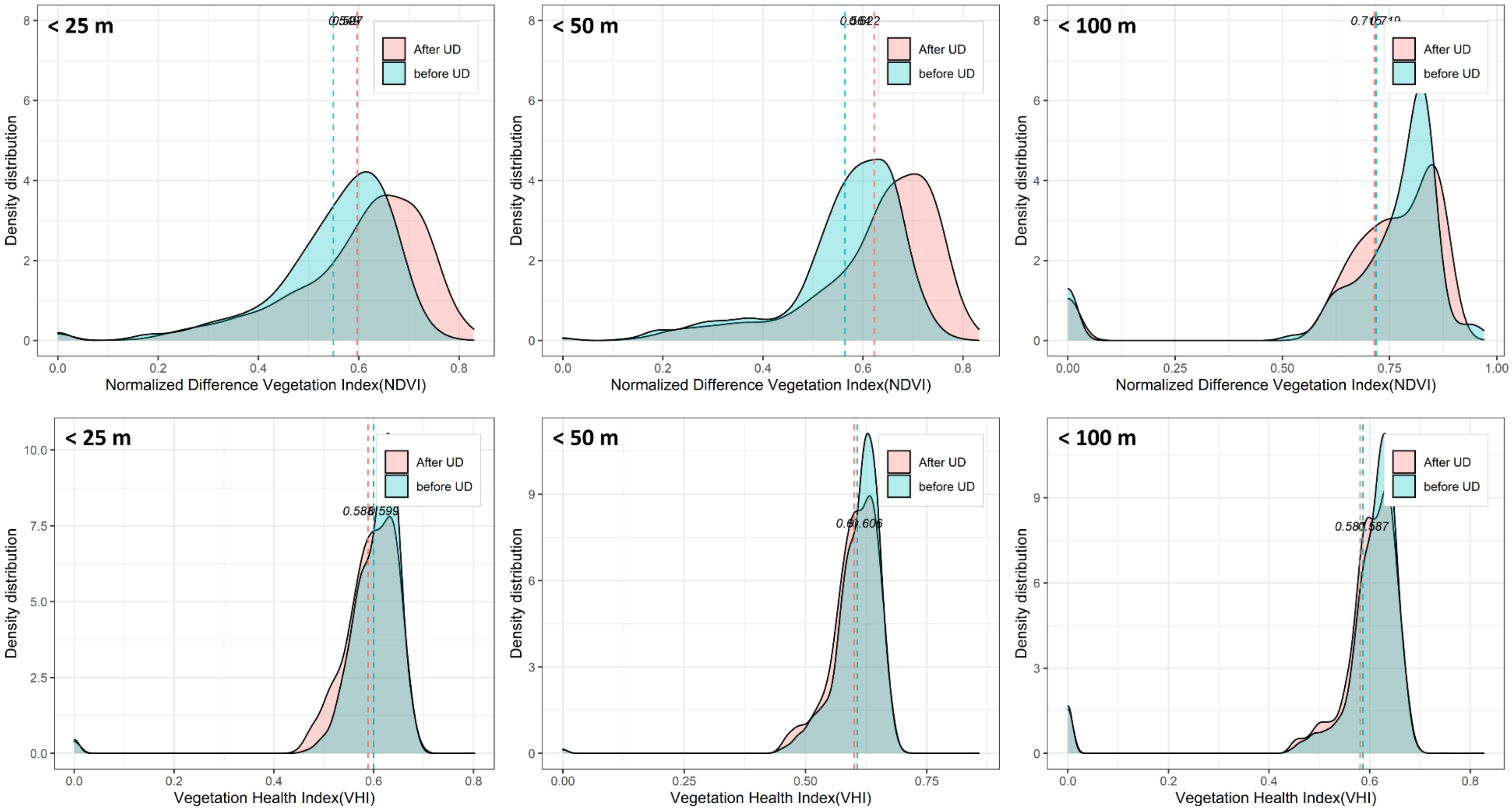

Figure 3 Trend change in NDVI and VHI time series according to the urban development projects.
(The bold value above represents the distance from the edge fragment)

4.2. Comparing the Performance of Machine Learning Algorithms

This study comprised three parts as follows: (1) training and testing of various supervised machine learning models for the diagnosis of vegetation indices changes before and after urban development projects; (2) validation of the best−performing RF and ANN models; and (3) heatmap by land covers using final machine learning model in the study sites to observe the modeling. The major hyperparameters incorporated in the ANN model are as follows: activation had a default value with solver as the default data type, and max_iter, acc, and RMSE were optimized with the hyperparameter values achieved from each iteration as the output. ACC values were expressed in various ways according to the ecotone distance and cross−validation procedure. As can be observed in Table 8, the ACC value reached the maximum value for NDVI, and the optimal hyperparameter value was as follows. In addition, Table 9 shows the optimal values for VHI.

Table 7 Hyperparameter tuning of Artificial Neural Networks for Normalized Difference Vegetation Index

Hyperparameters	25 m	50 m	100 m
*solver	adam	sgd	adam
**max_iter	150	300	100
***acc	0.778	0.789	0.691
****RMSE	0.222	0.211	0.309

Table 8 Hyperparameter tuning of Artificial Neural Networks for Vegetation Health Index

Hyperparameters	25 m	50 m	100 m
*solver	adam	adam	adam
**max_iter	300	250	250
***acc	0.751	0.720	0.723
****RMSE	0.249	0.280	0.277

*solver: Loss function minimization
**max_iter: The maximum number of major iterations for solution
***acc: Accuracy
****RMSE: Root Mean Square Error

Hyperparameters covered in the RF model used in this study are listed in Table 9(Sun et al., 2020). The min_samples_split, max_depths, max_features and min_samples_leaf had a default data type; and n_estimators and bootstrap were optimized, with the hyperparameter values of 20 and 'True' respectively obtained in each iteration as output. Hyperparameters obtained in iteration training were output, which were the identical in entire phase.

Table 9 Hyperparameter tuning of Random Forest for integrated vegetation indices

Hyperparameters	25 m	50 m	100 m
n_estimators		20	
bootstrap		True	

*n_estimators: The number of trees in the forest
**bootstrap: drawing of sample data

The dataset has been divided into 80% training and 20% testing. Initially, 80% of the data were randomly chosen for training, and the remaining 20% for testing. In this study, the Receiver Operating Characteristic (ROC) curve was used to evaluate the performance of each model, which is used as a tool for evaluating the performance of the model in most modeling studies using machine learning (Pham et al., 2016; Shahabi & Hashim, 2015). Here, the high ROC value means that the prediction performance of the model distinguishing the class is high. The cross−validation procedure was reiterated 1000 times. To adopt the optimal algorithm for predicting the extent of vegetation changes, the two distinct models were compared with regard to the accuracy (Table 10). Of the two models compared, RF registered the highest area under the receiver operating characteristic curve of 0.96 (0.81 for ANN), as shown in Figure 4. The results obtained show that the RF model demonstrated the best predictive performance in both vegetation indices and in details of ecotone distance. Hence, the RF model was carried forward entire subsequent analyses. Nevertheless, the validation of the NDVI from Sentinel data presented a considerable difference in performance.

Table 10 Validation of the NDVI and predicted parameters based on the machine learning models

Metric	25 m	50 m	100 m
Artificial Neural Network	0.699	0.751	0.537
Random Forest	0.94	0.933	0.679

Table 11 Validation of the VHI and predicted parameters based on the machine learning models

Metric	25 m	50 m	100 m
Artificial Neural Network	0.707	0.684	0.676
Random Forest	0.921	0.913	0.901

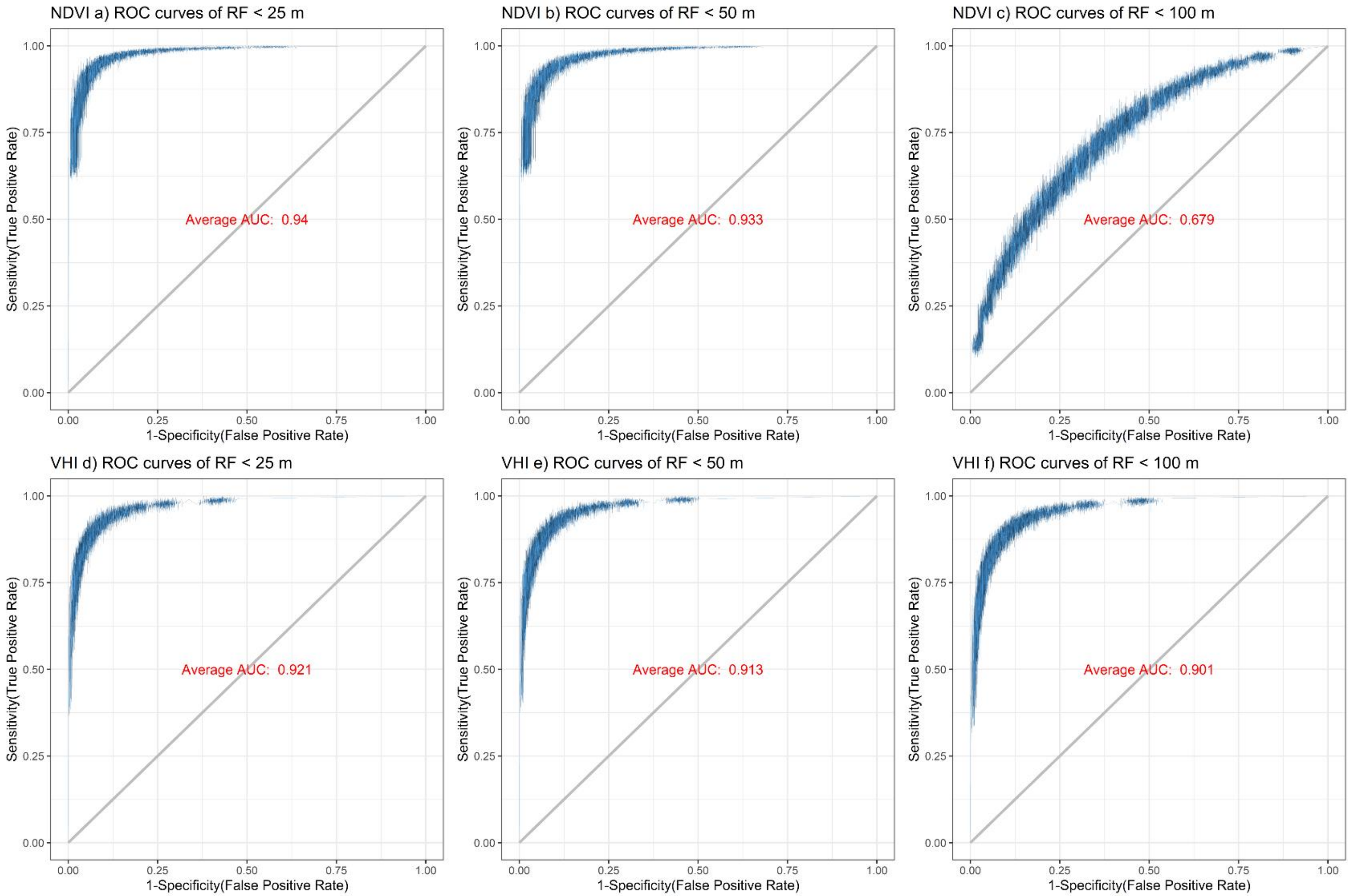

Figure 4 Receiver operating characteristic curves for the Random Forest model

4.3. Relationship between Urban Development and Vegetation Indices

This study proposed an impact assessment model to predict and quantify the extent to which urban development projects affect the edges along three different buffer sizes. The prediction is based on a predefined set of possibilities and is combined with a predictive probability distribution heatmap. Land cover maps, topography, and meteorology were set to detect the influence on the vegetation indices by ecotone multiple buffer sizes. Hence, spatiotemporal pattern for probability of change shown as the heatmap. The observations pattern reflected the data based actual NDVI is shown as the heatmap (Figure 5a). The estimates represent values modeled with the evaluation tool proposed in this study (Figure 5b). The absolute value of the predictive probability for each predictor is given as a percentage. As land cover types and other variables are constrained, the probability of that vegetation indices increases compared to predefined set shows the value of the cell. That is, it means the possibility that the NDVI value of each target site may change positively when the environment changes. If the green area ratio falls sharply due to urban development, it will be close to zero, and if it forms a green area city after development, it will be close to one. In this way, it was set to input values according to specific reference values. Shading represents the intensity of vegetation changes (see legend in the middle upper aspect).

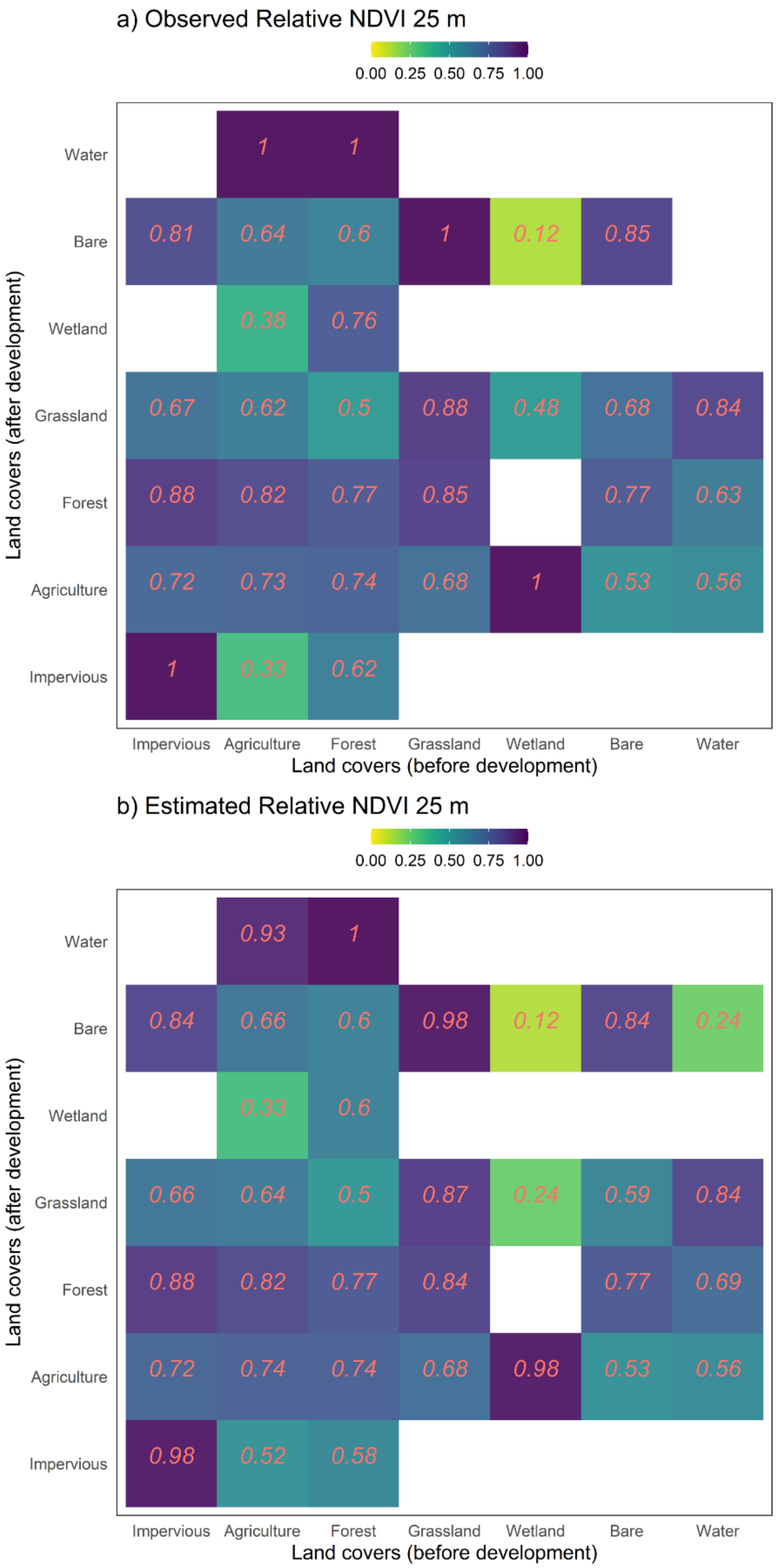

Figure 5 Heat map of the NDVI changes for the 25 m distance buffer

Probability heat map modelled demonstrates significant results at 90% confidence level ($p < 10\%$) even after multiple testing. When land cover changes occur from wetlands to grasslands, the actual data-based heatmap results interpreted that NDVI values would increase with a 48% probability, while the modeling results were considerably low at 24%. In addition, the observed NDVI results showed a 33% probability of increasing values from the predefined set with NDVI when land cover changed from agriculture to impervious, yet the modeled figures were relatively high at 52%. Excluding those cells, significant results were derived when comparing the observed and predicted values (Figure 5).

As can be observed in probability heatmap for the NDVI extracted from Landsat (Figure 8), the further away from the ecotone for 25, 50, and 100 m buffers, the less significant correlation with the inner side of the urban development site. In general, as the distance from the research site increases, the cell value decreases, which means that the probability of being higher than the predefined value is lower. In other words, it was expressed that the influence from internal environmental variables changed. It can be investigated that the color is clearly lighter as it goes up to 100 meters, and the changing probability value decreases. However, it was confirmed that the prediction performance at 100 m was slightly lower than that of 25 and 50 m. In about four cells, there was a larger error between the observed and predicted values. The singularity is that conditions that did not actually exist were also predicted by simulation.

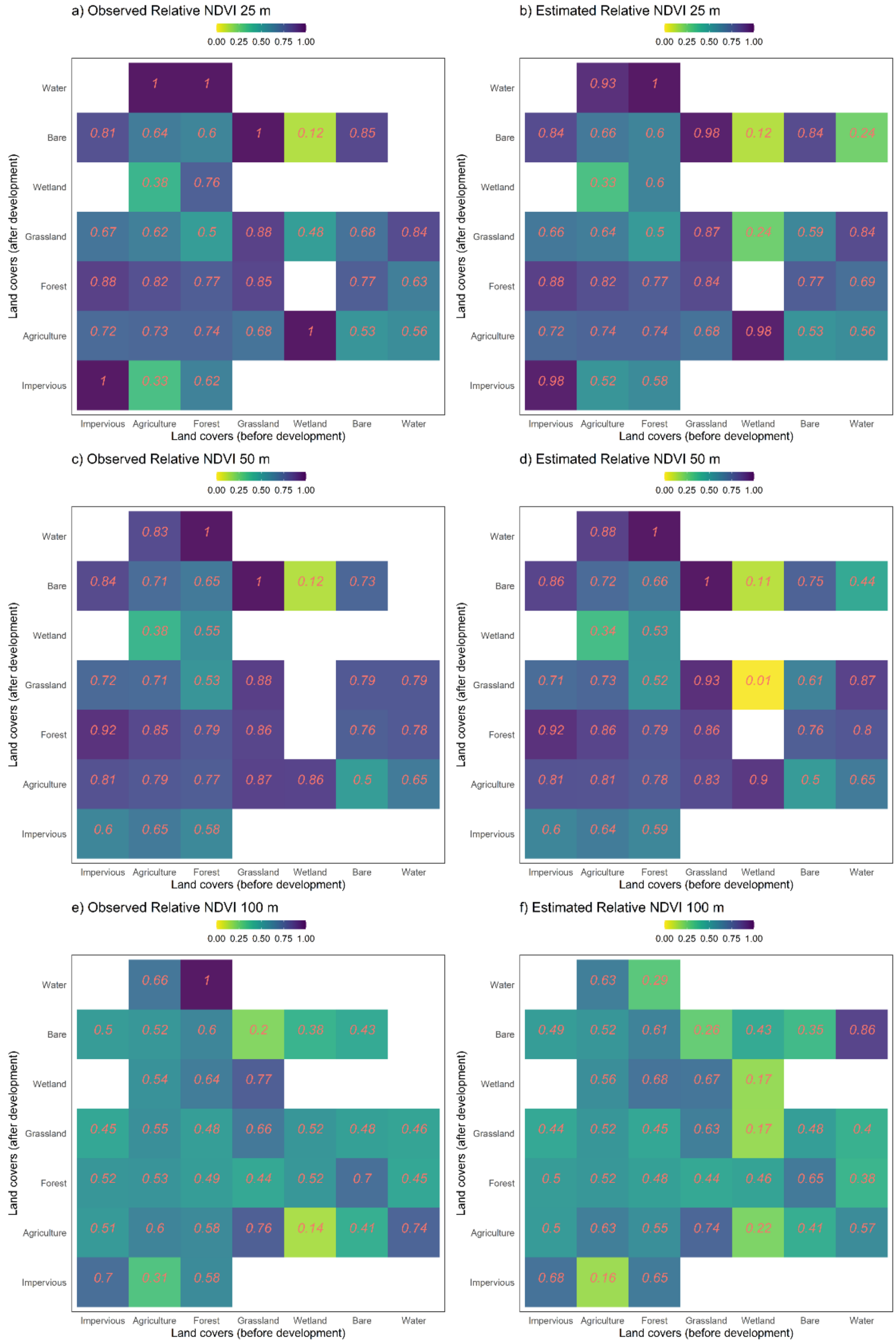

Figure 6 Heat map of the comparison between the NDVI modeling results for 25, 50 m and 100 m buffers

Remote sensing-based vegetation indices have been developed and used to predict vegetation yields in different ecosystem. Previous research has applied VHI for different approaches that may influence ecosystems (Bokusheva et al., 2016; Prasad et al., 2006; Ribeiro et al., 2019; Unganai & Kogan, 1998). The entire range of VHI was analyzed to detect and predict the average extent of vegetation change. The VHI is related to moisture availability and represents vegetation stress (Marengo et al., 2021). As VHI is an index applied to quantifying wetness and dryness, it can be confirmed that the change in land cover type to wetlands is the significant value as can be recognized in Figure 7. This trend also occurs in the change of land cover to agriculture. The VHI and crop yields are highly correlated, especially at the critical stage of crop growth (Prasad et al., 2006; Kogan et al., 2012). It can be seen that when urban development is influenced in wetland and agriculture, the predefined value that converges to approximately 1 is obtained. In addition, most of values were found to have a low effect, and the difference in the extent of influence for the three buffer sizes could not be confirmed significantly.

The estimated probability heatmap presents a step forward in improving the effectiveness of detecting vegetation changes and thus its application prospects. Furthermore, the response characteristics of various vegetation types are identified, deepening understanding on correlations between environmental factors and vegetation indices, which may help decision-makers and authorities to develop better mitigation and adaptation strategies.

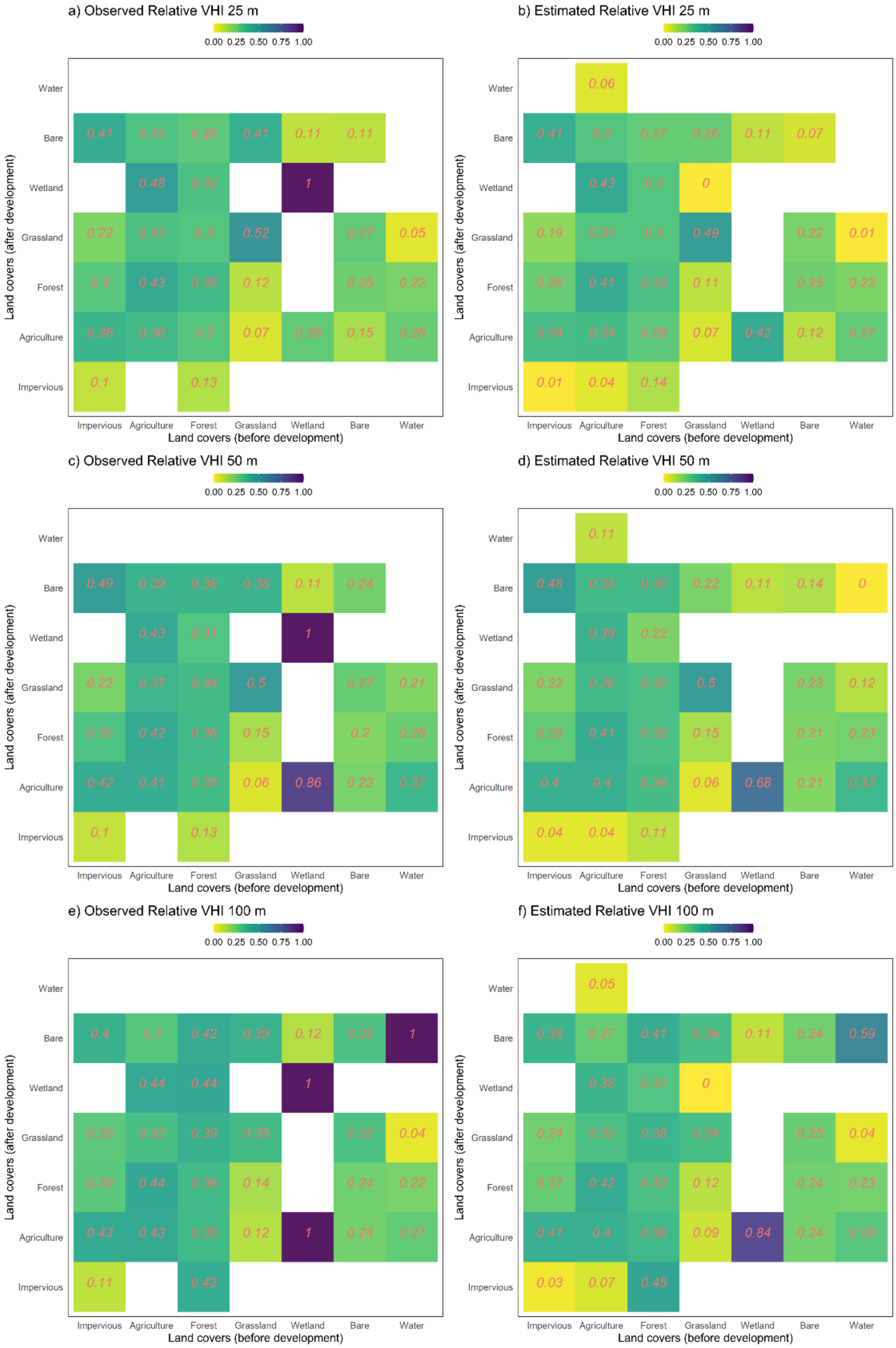

Figure 7 Heat map of the comparison between the VHI modeling results for 25, 50 m and 100 m buffers

Chapter 5. Discussion

5.1. Comparison of Landsat and Sentinel data for estimation of the NDVI

Both Landsat and Sentinel−2 provide high spatial resolution images of 10 m and 30 m respectively. Sentinel−2 data generate high resolution cropping intensity maps to provide high−quality observations. In this study, Landsat data was used as the main source of analysis due to the mismatch between the established temporal range and the satellite imagery to be available, and further comparative analysis was performed using Sentinel−2 data. The Sentinel−2 launched on March 28, 2017, so that data after 2018 was applied as the NDVI value after the urban development project. Hyperparameters applied in the machine learning models are listed (Table 12). Hyperparameters obtained in iteration training were output, were optimized.

Table 12 Hyperparameter tuning of ANN and RF for NDVI from Sentinel−2

Hyperparameters	25 m	50 m	100 m
*solver	adam	adam	adam
*max_iter	100	250	200
*acc	0.858	0.836	0.794
*RMSE	0.142	0.164	0.206
**n_estimators		20	
**bootstrap		True	

* Hyperparameters covered in ANN model
**Hyperparameters covered in RF model

Probability heat map modelled with NDVI from Sentinel data demonstrates significant results at 80% confidence level. Significant results were derived when comparing the observed and predicted values. The NDVI was compared using two different satellite images, Landsat and Sentinel, under the same premise to quantitatively assess the extent of their impact on ecotones over distance (Figure 8). As can be observed, the values of Sentinel based NDVI were clearly lower than the values for Landsat (Figure A 2), and the influence from the inner side of the study site was also found to be less (Figure 8). What is unusual is that the numerical value of the probability heatmap detected the opposite tendency to NDVI from Landsat. The values of Landsat based NDVI was generally higher than Sentinel. The further away from the ecotone for 25, 50, and 100 m buffers, it showed a higher correlation. Due to the nature of the research area composed of relative high vegetation cover, it may be interpreted that the high NDVI value of the surrounding vegetation area was included in the pixel of Landsat, which has a relatively low spatial resolution, resulting in mixed pixel effect. In the case of forest areas, vegetation is distributed over a large dense range, so spectroscopic mixing is alleviated, and Sentinel has a lower value, but there is no significant difference from Landsat data (Chen et al., 2018). Using Landsat data at 30 m affects the variation of the mean NDVI value, which is inferred by the difference in pixel levels. Effective use of images is difficult because Landsat data is temporarily discontinuous due to 16-day revisit cycles, cloud contamination, and hardware problems with sensors. (Cao et al., 2020; Congalton, 2018; Ju & Roy, 2008; Pringle et al., 2009; Santos et al., 2021; Zhu & Woodcock, 2012).

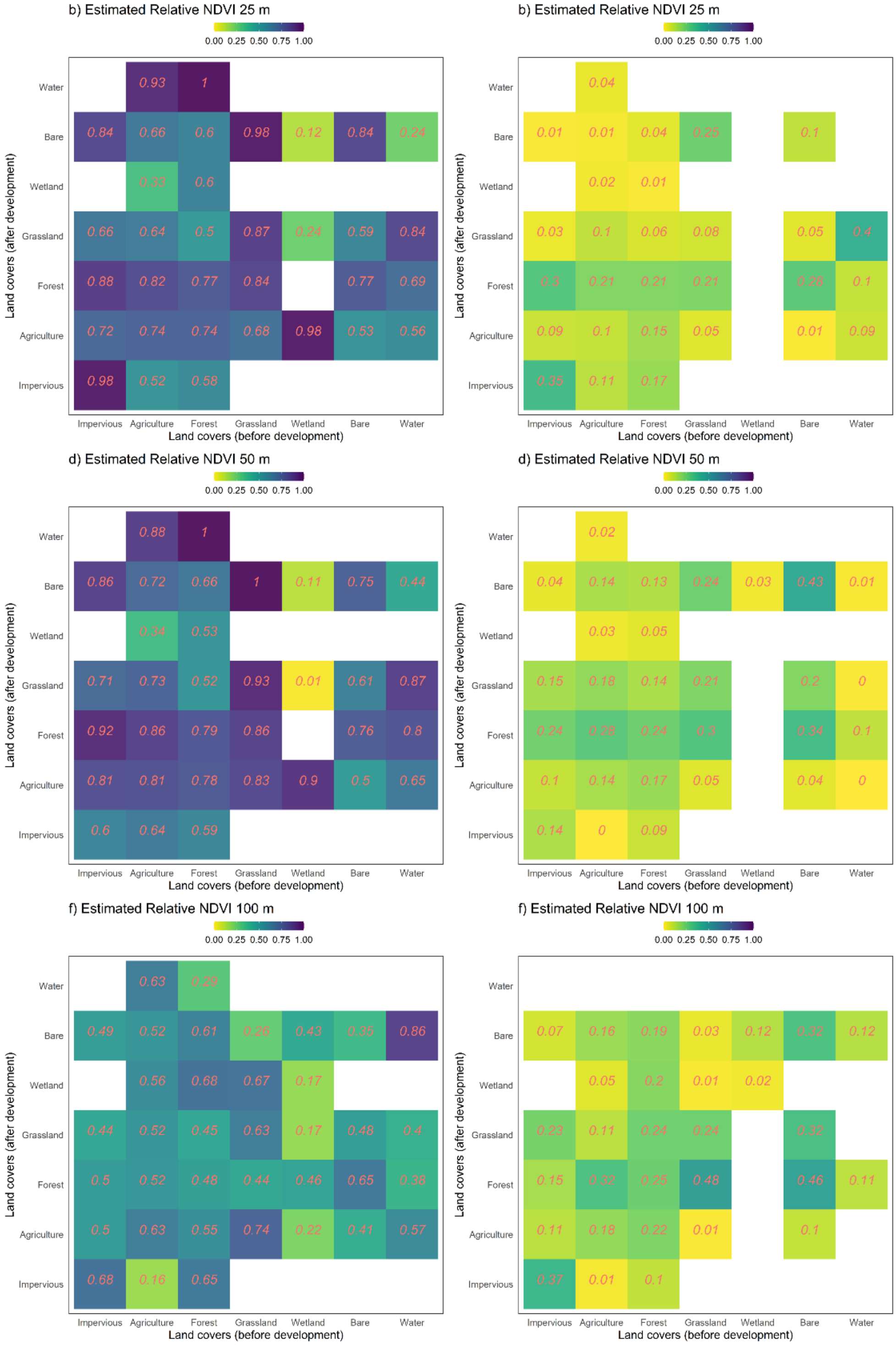

Figure 8 Comparison of estimated results between Sentinel and Landset satellite imagery

5.2. Building a Classifier

In using decision trees for classification, methods are needed to achieve optimal accuracy while avoiding overfitting of training datasets (Ho, 1998). The classifier in this study serves as a criterion for determining the standard of change in NDVI and is one of the most important parts of analysis. The final result is expressed through numerous learning between the value of the vegetation indices which were affected according to the environmental variables and the value by redefined set.

The business year of each target point was extracted, and the value minus the previous year and the following year was set as the reference value for each NDVI and VHI. The probability that the value of each vegetation index changes to a reference value, that is, a value quantifying the degree of influence to change is expressed in the final heat map. Accordingly, a change in the final value may occur according to a change in the setting of the reference value. For example, assuming that the reference value is set to an absolute value, the NDVI is likely to be higher than the reference value according to the environmental variables that change with urban development. These issues limit the design of analytical studies. This procedure cannot rule out the possibility of artificially overestimating or underestimating the mean NDVI value.

5.3. Model Interpretation Strategies

The NDVI based Landsat found a significant correlation in the value of ecotone multiple buffer sizes as the heatmap results, and the NDVI based Sentinel was found to be correlated by changes in the value of NDVI based Sentinel for urban development projects. These differences were statistically further confirmed through the mean and variance differences in NDVI exposure. The findings question the practice of misaligned connections of green space and epidemiological data.

Since the actual construction period is not specified other than the date of completion of the EIASS environmental impact assessment, this study is based on the date of completion of the agreement and does not support the actual time of completion of the construction. Due to data that is partially time−incompatible, the accuracy of the evaluation may have been reduced. In addition, there are inconsistencies between Landsat and Sentinel data availability ranges. In the analysis setting, Landsat data is applied based on urban development projects as there is no limit to the temporal range, yet Sentinel data is set based on satellite images due to the limit of the available range. This is also a factor that is time incompatible and can reduce the accuracy of the results.

Chapter 6. Conclusions

The objective was to analyze the correlation between environmental changes according to the urban projects and vegetation indices, and to quantify the extent of changes in vegetation ecotone along the distances. We interpret the distribution characteristics of vegetation cover using two machine learning models which is ANN and RF models in study sites. In addition, the possibility of ecotone damage and vegetation green change due to urban development were predicted through 10 years of vegetation green change extracted based on two satellite images. The analysis results were found to compare the changes in the actually obtained data-based vegetation index with the prediction results according to the modeling of this simulation model. Our conclusion is that the promotion of urban development projects has a significant effect on peripheral vegetation. It includes both negative and opposite tendencies. The value of NDVI increased after the development act and resulted in greener results, but in the case of VHI, the lower value after urban development prevailed, which supports this result. Land cover change and urbanization induce a decrease in VHI. Land cover changes are a major factor in environmental degradation, and rapid urbanization promotes this situation (Singh et al., 2017). Conversely, urban development increases urban green areas by contributing to the expansion of green areas. Therefore, it is essential to consider various aspects when planning urban development. It should be important to quantify the contribution of climatic and human activities as well as policies to vegetation changes based on a residual analysis. Also, proposing effective measures to improve the services and value provided by ecosystems is fundamental. Our tool is the beginning to advanced knowledge in improving development of conservation strategies, which may help to quantify the edge effect spatially. It is essential to understand the trends of change in Vegetation matters and other relevant parameters to effectively address the issues arising from edge effects, while land cover change related risks can be minimized through initiating proper measures like proper management (Noszczyk et al., 2020).

Bibliography

Ahmadi, K., Kalantar, B., Saeidi, V., Harandi, E. K., Janizadeh, S., & Ueda, N. (2020). Comparison of machine learning methods for mapping the stand characteristics of temperate forests using multi-spectral sentinel-2 data. Remote Sensing, 12(18), 3019.

Alignier, A., & Deconchat, M. (2013). Patterns of forest vegetation responses to edge effect as revealed by a continuous approach. Annals of Forest Science, 70(6), 601-609.

Aragón, G., Abuja, L., Belinchón, R., & Martínez, I. (2015). Edge type determines the intensity of forest edge effect on epiphytic communities. European Journal of Forest Research, 134(3), 443-451.

Arce-Peña, N. P., Arroyo-Rodríguez, V., San-José, M., Jiménez-González, D., Franch-Pardo, I., Andresen, E., & Ávila-Cabadilla, L. D. (2019). Landscape predictors of rodent dynamics in fragmented rainforests. Biodiversity and Conservation, 28(3), 655-669.

Arroyo-Rodríguez, V., Fahrig, L., Tabarelli, M., Watling, J. I., Tischendorf, L., Benchimol, M., ... & Tscharntke, T. (2020). Designing optimal human-modified landscapes for forest biodiversity conservation. Ecology letters, 23(9), 1404-1420.

Assessment, M. E. (2005). Ecosystems and human well-being: wetlands and water. World Resources Institute.

Bao, Z., Zhang, J., Wang, G., Guan, T., Jin, J., Liu, Y., ... & Ma, T. (2021). The sensitivity of vegetation cover to climate change in multiple climatic zones using machine learning algorithms. Ecological indicators, 124, 107443.

Batáry, P., Fronczek, S., Normann, C., Scherber, C., & Tscharntke, T. (2014). How do edge effect and tree species diversity change bird diversity and avian nest survival in Germany's largest deciduous forest?. Forest Ecology and Management, 319, 44-50.

Benchimol, M., & Peres, C. A. (2013). Anthropogenic modulators of species–area relationships in N eotropical primates: a continentalscale analysis of fragmented forest landscapes. Diversity and Distributions, 19(11), 1339-1352.

Bokusheva, R., Kogan, F., Vitkovskaya, I., Conradt, S., & Batyrbayeva, M. (2016). Satellite-based vegetation health indices as a criteria for insuring against drought-related yield losses. Agricultural and Forest Meteorology, 220, 200-206.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Cao, R., Chen, Y., Chen, J., Zhu, X., & Shen, M. (2020). Thick cloud removal in Landsat images based on autoregression of Landsat time-series data. Remote Sensing of Environment, 249, 112001.

Carrara, E., Arroyo-Rodríguez, V., Vega-Rivera, J. H., Schondube, J. E., de Freitas, S. M., & Fahrig, L. (2015). Impact of landscape composition and configuration on forest specialist and generalist bird species in the fragmented Lacandona rainforest, Mexico. Biological Conservation, 184, 117-126.

Chen, X., Wang, D., Chen, J., Wang, C., & Shen, M. (2018). The mixed pixel effect in land surface phenology: A simulation study. Remote Sensing of Environment, 211, 338-344.

Chen, Y., Cao, R., Chen, J., Liu, L., & Matsushita, B. (2021). A practical approach to reconstruct high-quality Landsat NDVI time-series data by gap filling and the Savitzky–Golay filter. ISPRS Journal of Photogrammetry and Remote Sensing, 180, 174-190.

Coluzzi, R., di Donna, I., Lanorte, A., & Lasaponara, R. (2007, October). Mapping forest fuel types by using satellite ASTER data and neural nets. In Remote sensing for agriculture, ecosystems, and hydrology IX (Vol. 6742, pp. 289-296). SPIE.

Congalton, R. G. (2018). Landsat's enduring legacy: pioneering global land observations from space. Photogrammetric Engineering & Remote Sensing, 84(1), 9-10.

Curtis, P. G., Slay, C. M., Harris, N. L., Tyukavina, A., & Hansen, M. C. (2018). Classifying drivers of global forest loss. Science, 361(6407), 1108-1111.

Dong, J., Xiao, X., Menarguez, M. A., Zhang, G., Qin, Y., Thau, D., ... & Moore III, B. (2016). Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. Remote sensing of environment, 185, 142-154.

Esseen, P. A., & Renhorn, K. E. (1998). Edge effects on an epiphytic lichen in fragmented forests. Conservation biology, 12(6), 1307-1317.

Esseen, P. A. (2019). Strong influence of landscape structure on hair lichens in boreal forest canopies. Canadian Journal of Forest Research, 48(8), 994-1003.

Evans, K. L., James, N. A., & Gaston, K. J. (2006). Abundance, species richness and energy availability in the North American avifauna. Global Ecology and Biogeography, 15(4), 372-385.

Ewers, R. M., & Didham, R. K. (2006). Continuous response functions for quantifying the strength of edge effects. Journal of applied ecology, 43(3), 527-536.

Fahrig, L. (2017). Ecological responses to habitat fragmentation per se. Annual Review of Ecology, Evolution, and Systematics, 48(1), 1-23.

Fraver, S. (1994). Vegetation responses along edge-to-interior gradients in the mixed hardwood forests of the Roanoke River Basin, North Carolina. Conservation Biology, 8(3), 822-832.

Galán-Acedo, C., Arroyo-Rodríguez, V., Cudney-Valenzuela, S. J., & Fahrig, L. (2019). A global assessment of primate responses to landscape structure. Biological Reviews, 94(5), 1605-1618.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote sensing of Environment, 202, 18-27.

Guerra, T. N., Araújo, E. L., Sampaio, E. V., & Ferraz, E. M. (2017). Urban or rural areas: which types of surrounding land use induce stronger edge effects on the functional traits of tropical forests plants?. Applied Vegetation Science, 20(4), 538-548.

Hernández-Ruedas, M. A., Arroyo-Rodríguez, V., Morante-Filho, J. C., Meave, J. A., & Martínez-Ramos, M. (2018). Fragmentation and matrix contrast favor understory plants through negative cascading effects on a strong competitor palm. Ecological Applications, 28(6), 1546-1553.

Ho, T. K. (1998). The random subspace method for constructing decision forests. IEEE transactions on pattern analysis and machine intelligence, 20(8), 832-844.

Holland, M. M. (1988). SCOPE/MAB technical consultations on landscape boundaries: report of a SCOPE/MAB workshop on ecotones. Biology International, 17(47), 106.

Ju, J., & Roy, D. P. (2008). The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. Remote Sensing of Environment, 112(3), 1196-1211.

Kivistö, L., & Kuusinen, M. (2000). Edge effects on the epiphytic lichen flora of Picea abies in middle boreal Finland. The Lichenologist, 32(4), 387-398.

Kogan, F. N. (1990). Remote sensing of weather impacts on vegetation in non-homogeneous areas. International Journal of remote sensing, 11(8), 1405-1419.

Kogan, F., Salazar, L., & Roytman, L. (2012). Forecasting crop production using satellite-based vegetation health indices in Kansas, USA. International journal of remote sensing, 33(9), 2798-2814.

Liao, H., & Pitts, A. (2006). A brief historical review of Olympic urbanization. The International Journal of the History of Sport, 23(7), 1232-1252.

Liu, L., Wang, Y., Wang, Z., Li, D., Zhang, Y., Qin, D., & Li, S. (2019). Elevation-dependent decline in vegetation greening rate driven by increasing dryness based on three satellite NDVI datasets on the Tibetan Plateau. Ecological Indicators, 107, 105569.

Łuczaj, Ł., & Sadowska, B. (1997). Edge effect in different groups of organisms: vascular plant, bryophyte and fungi species richness across a forest-grassland border. Folia Geobotanica & Phytotaxonomica, 32(4), 343-353.

Magura, T. (2002). Carabids and forest edge: spatial pattern and edge effect. Forest Ecology and management, 157(1-3), 23-37.

Marengo, J. A., Cunha, A. P., Cuartas, L. A., Deusdará Leal, K. R., Broedel, E., Seluchi, M. E., ... & Bender, F. (2021). Extreme drought in the Brazilian Pantanal in 2019–2020: characterization, causes, and impacts. Frontiers in Water, 3, 639204.

McKinney, M. L. (2006). Urbanization as a major cause of biotic homogenization. Biological conservation, 127(3), 247-260.

Moradi, E., Darabi, H., Alamdarloo, E. H., Karimi, M., & Kløve, B. (2023). Vegetation vulnerability to hydrometeorological stresses in water-scarce areas using machine learning and remote sensing techniques. Ecological Informatics, 73, 101838.

Morante-Filho, J. C., Arroyo-Rodríguez, V., Pessoa, M. D. S., Cazetta, E., & Faria, D. (2018). Direct and cascading effects of landscape structure on tropical forest and non-forest frugivorous birds. Ecological applications, 28(8), 2024-2032.

Mutanga, O., & Kumar, L. (2019). Google earth engine applications. Remote Sensing, 11(5), 591.

Noszczyk, T., Rutkowska, A., & Hernik, J. (2020). Exploring the land use changes in Eastern Poland: Statistics-based modeling. Human and Ecological Risk Assessment: An International Journal, 26(1), 255-282.

Park, S., Im, J., Jang, E., & Rhee, J. (2016). Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions. Agricultural and forest meteorology, 216, 157-169.

Peng, W., Kuang, T., & Tao, S. (2019). Quantifying influences of natural factors on vegetation NDVI changes based on geographical detector in Sichuan, western China. Journal of cleaner production, 233, 353-367.

Phalan, B. T. (2018). What have we learned from the land sparing-sharing model?. Sustainability, 10(6), 1760.

Pham, B. T., Pradhan, B., Bui, D. T., Prakash, I., & Dholakia, M. B. (2016). A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). Environmental Modelling & Software, 84, 240-250.

Pignatti, S., Acito, N., Amato, U., Casa, R., Castaldi, F., Coluzzi, R., ... & Cuomo, V. (2015, July). Environmental products overview of the Italian hyperspectral prisma mission: The SAP4PRISMA project. In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 3997-4000). IEEE.

Prasad, A. K., Chai, L., Singh, R. P., & Kafatos, M. (2006). Crop yield estimation model for Iowa using remote sensing and surface parameters. International Journal of Applied earth observation and geoinformation, 8(1), 26-33.

Pringle, M. J., Schmidt, M., & Muir, J. S. (2009). Geostatistical interpolation of SLC-off Landsat ETM+ images. ISPRS Journal of Photogrammetry and Remote Sensing, 64(6), 654-664.

Rheault, H., Drapeau, P., Bergeron, Y., & Esseen, P. A. (2003). Edge effects on epiphytic lichens in managed black spruce forests of eastern North America. Canadian Journal of Forest Research, 33(1), 23-32.

Ribeiro, A. F., Russo, A., Gouveia, C. M., & Páscoa, P. (2019). Modelling drought-related yield losses in Iberia using remote sensing and multiscalar indices. Theoretical and Applied Climatology, 136(1), 203-220.

Sampaio, A. B., & Scariot, A. (2011). Edge effect on tree diversity, composition and structure in a deciduous dry forest in central Brazil. Revista Árvore, 35(5), 1121-1134.

Santos, L. A., Ferreira, K. R., Camara, G., Picoli, M. C., & Simoes, R. E. (2021). Quality control and class noise reduction of satellite image time series. ISPRS Journal of Photogrammetry and Remote Sensing, 177, 75-88.

Shahabi, H., & Hashim, M. (2015). Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment. Scientific reports, 5(1), 1-15.

Silva, V. L., Mehltreter, K., & Schmitt, J. L. (2018). Ferns as potential ecological indicators of edge effects in two types of Mexican forests. Ecological indicators, 93, 669-676.

Simoniello, T., Coluzzi, R., Imbrenda, V., & Lanfredi, M. (2015). Land cover changes and forest landscape evolution (1985–2009) in a typical Mediterranean agroforestry system (high Agri Valley). Natural Hazards and Earth System Sciences, 15(6), 1201-1214.

Singh, P., Kikon, N., & Verma, P. (2017). Impact of land use change and urbanization on urban heat island in Lucknow city, Central India. A remote sensing based estimate. Sustainable cities and society, 32, 100-114.

Sun, D., Wen, H., Wang, D., & Xu, J. (2020). A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. Geomorphology, 362, 107201.

Unganai, L. S., & Kogan, F. N. (1998). Drought monitoring and corn yield estimation in Southern Africa from AVHRR data. Remote sensing of environment, 63(3), 219-232.

Wan, Z., Wang, P., & Li, X. (2004). Using MODIS land surface temperature and normalized difference vegetation index products for monitoring drought in the southern Great Plains, USA. International journal of remote sensing, 25(1), 61-72.

Wang, J., Rich, P. M., & Price, K. P. (2003). Temporal responses of NDVI to precipitation and temperature in the central Great Plains, USA. International journal of remote sensing, 24(11), 2345-2364.

Weier, J., & Herring, D. (2000). Measuring vegetation (ndvi & evi). NASA Earth Observatory, 20, 2.

Wilson, J. D., Anderson, R., Bailey, S., Chetcuti, J., Cowie, N. R., Hancock, M. H., ... & Thompson, D. B. (2014). Modelling edge effects of mature forest plantations on peatland waders informs landscape-scale conservation. Journal of Applied Ecology, 51(1), 204-213.

Xiao, J., & Moody, A. (2004). Trends in vegetation activity and their climatic correlates: China 1982 to 1998. International Journal of Remote Sensing, 25(24), 5669-5689.

Yao, R., Cao, J., Wang, L., Zhang, W., & Wu, X. (2019). Urbanization effects on vegetation cover in major African cities during 2001-2017. International Journal of Applied Earth Observation and Geoinformation, 75, 44-53.

Zhao, J., Chen, S., Jiang, B., Ren, Y., Wang, H., Vause, J., & Yu, H. (2013). Temporal trend of green space coverage in China and its relationship with urbanization over the last two decades. Science of the Total Environment, 442, 455-465.

Zhi-qiang, G. A. O., & Dennis, O. (2001). The temporal and spatial relationship between NDVI and climatological parameters in Colorado. Journal of Geographical Sciences, 11(4), 411-419.

Zhou, X., & Wang, Y. C. (2011). Spatial–temporal dynamics of urban green space in response to rapid urbanization and greening policies. Landscape and urban planning, 100(3), 268-277.

Zhu, Z., & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. Remote sensing of environment, 118, 83-94.

Abstract in Korean

패치는 식생 조건의 차이로 인해 이질성을 보이는 인접 패치 사이의 전이 영역으로 에코톤으로 인식된다. 에코톤은 높은 생물 다양성, 생태계 연결성 그리고 다양한 서식 환경을 제공함으로써 생태학적으로 중요한 역할을 한다. 우리나라는 급격한 산업화 및 도시화로 패치의 공간적 변화를 경험했기 때문에 에코톤 보존은 인간의 활동이 자연환경에 미치는 영향을 확인하는 데 달려 있다. 따라서, 지속가능한 관리방안을 강구하기 위해 도시화로 인해 변화하는 식생동태를 모니터링하고 그 영향 정도를 평가하고자 하였다. 본 연구는 도시개발사업의 영향을 받아 변화하는 주연부 범위(25, 50 m 및 100 m)를 예측하는 정량적 평가도구를 제안하였다. 영향을 평가하기 위한 지표로는 정규식생지수(NDVI)와 식생건강지수(VHI)가 선정되었으며, 구글어스엔진(GEE) 플랫폼을 통해 Landsat과 Sentinel 기반의 위성영상 데이터를 계산하였다. 환경영향평가정보지원시스템(EIASS)의 국가 인벤토리 데이터를 주요변수로 적용하였으며, ArcGIS 10.5를 통해 환경공간정보서비스의 토지피복도와 기상청 평균기온 및 강수량 데이터세트를 구축하였다. 데이터 분석 방법은 인공신경망(ANN)과 랜덤포레스트(RF) 기계학습 알고리즘을 이용하여 연구대상지의 식생분포 패턴을 분석하였으며, 에코톤 다중 범위에 따라 식생지수에 미치는 영향정도를 예측하도록 설정하였다. 분석 결과, NDVI는 주로 도시개발사업 이후 높은 수치에 집중분포 된 반면, VHI는 사업 전 수치가 높은 경향을 보여 반대의 추세를 보인 것으로 확인되었다. 이는 도시녹지 관리 및 도시경관계획을 위한 「도시공원 및 녹지 등에 관한 법률」의 규정에 따라 새로운 도시녹지를 확충 및 조성에 따른 결과로 해석된다. 기계학습 모델의 성능 비교결과, RF 모델이 식생 지수와

에코톤 거리 모두에서 최적의 예측 성능을 보여주었다. 모델링된 확률 히트맵은 90% 신뢰 수준(p<10%)에서 유의한 결과를 보여주었다. 또한, 평가도구를 사용하여 시각화 된 관측값 및 예측값을 비교했을 때 유의한 결과를 얻을 수 있었다. NDVI와 VHI는 모두 도시개발로 인한 대상지의 영향이 최대 거리 50m에 달하는 경향을 보였다. 이번 정량평가 도구 제안은 지역 생태복원에 대한 정보를 제공함으로써 식생관리 측면에서 환경영향평가의 결정적 역할을 강조할 수 있다는 점에서 의미가 있다. 도시화에 의한 식생환경에 미치는 영향 정도를 파악하여 식생 피복 훼손을 최소화하면서 도시개발을 지원할 수 있을 것으로 기대된다.

키워드: 환경영향평가, 토지 피복 변화, 식생지수, 주연부효과, 기계학습, 원격탐사

학번: 2021-26769

Appendix

Table A 1 Comparison of ANN and RF performance in NDVI from Sentinel-2

Metric	25 m	50 m	100 m
ANN	0.844	0.784	0.749
RF	0.925	0.917	0.89

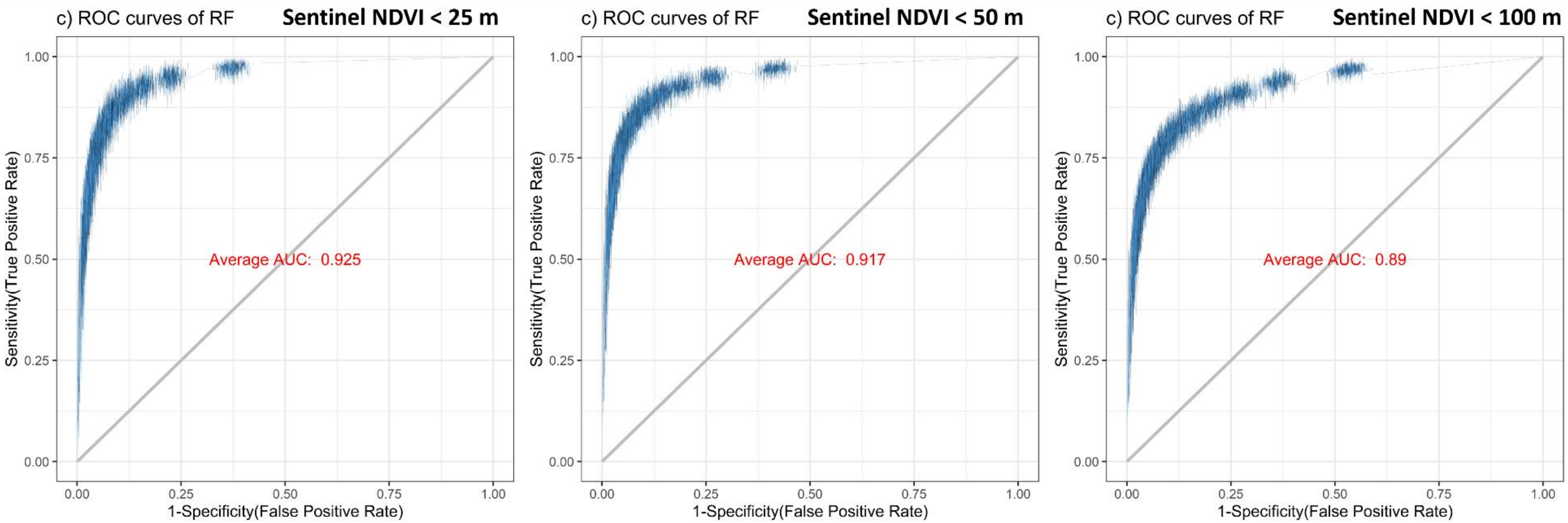

Figure A 1 Receiver operating characteristic curves in NDVI calculated by Sentinel−2 for the Random Forest machine learning model

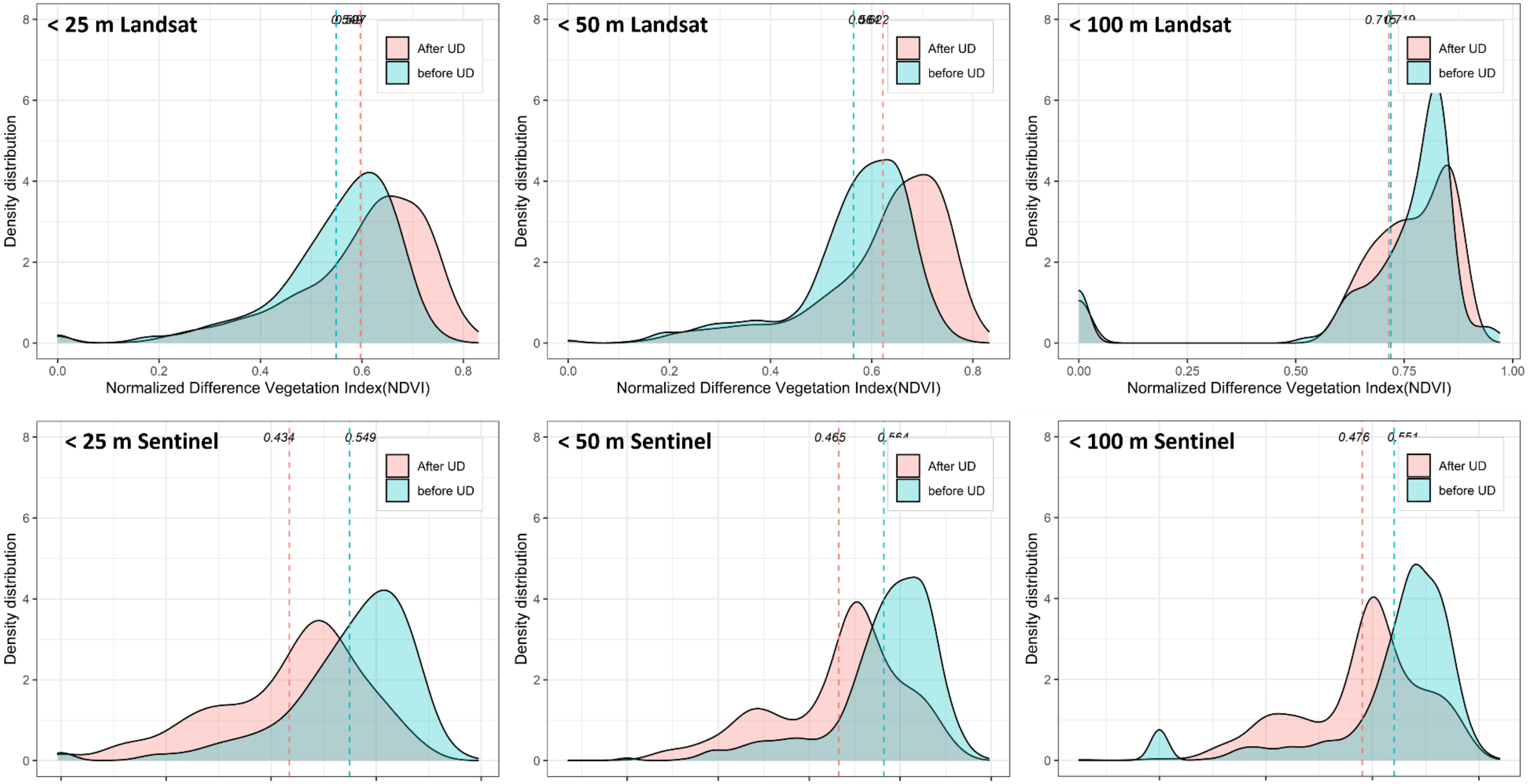

Figure A 2 The variation of the NDVI derived from Sentinel−2 according to the urban development projects