博 士 論 文

Research on Spatial Downscale Temperature Prediction by using

Machine Learning and its Application in Urban Heat Island

機械学習を用いた空間ダウンスケールの気温予測法及び

都市ヒートアイランドへの応用に関する研究

北九州市立大学国際環境工学研究科

2021 年 6 月

王　鋭

WANG　RUI

Doctoral Thesis

Research on Spatial Downscale Temperature Prediction by using

Machine Learning and its Application in Urban Heat Island

June 2021

Rui WANG

The University of Kitakyushu

Faculty of Environmental Engineering

Department of Architecture

Gao Laboratory

Preface

This study focused on the air temperature spatial downscaling method based on machine learning algorithms. Due to the contradiction between the temporal and spatial resolution of some meteorological images, or the inability to provide high-resolution images, five metropolitan areas in Japan were selected as case studies to analyze the feasibility of the downscaling method. Then based on the downscaled high-resolution air temperature data, this study analyzed and compared the atmosphere urban heat islands of each metropolitan area. Through comparison with surface urban heat island, the characteristics of two types of urban heat island were indicated.

ACKNOWLEDGEMENTS

RESEARCH ON SPATIAL DOWNSCALE TEMPERATURE PREDICTION BY USING MACHINE LEARNING AND ITS APPLICATION IN URBAN HEAT ISLAND

ABSTRACT

With the rapid development of urban around the world, the urban environment has been greatly damaged, the most notable of which is the growth of urban heat island. Urban heat islands are not only harmful to the health of residents, but also to society development. Various metropolitan areas in Japan are also facing the problem of urban heat island.

This study mainly applied the spatial downscaling method of air temperature based on machine learning algorithm to study the urban heat island in five typical metropolitan areas in Japan. The downscaling method used in this study is a statistical downscaling method. The main principle is the regression invariance of air temperature and underlying surface characteristics at different scales, that is, the regression with low resolution was established first, and then the regression model was used to predict high-resolution air temperature in combination with high-resolution underlying surface characteristics. Based on the downscaled high-resolution air temperature, urban heat island intensity and urban heat island ratio index were applied to evaluate and compare the atmosphere urban heat islands of each metropolitan area. Moreover, the surface urban heat island based on land surface temperature was compared with the atmosphere urban heat island, and the similarities and differentiates were indicated.

In Chapter One, the research background and significance of urban heat island and the application of temperature downscaling in urban heat island research are demonstrated. In addition, the importance of urban heat island research and temperature downscaling research were analyzed, and previous research was reviewed. Then, the purpose of this study was purposed.

In Chapter Two, this is the core of this study, the air temperature spatial downscaling method was introduced. Firstly, it was the basic principle of statistical downscaling method. Secondly, three different types of downscaling methods are analyzed and compared. The results showed that the random forest model based on machine algorithms was the most suitable for temperature spatial downscaling in highly heterogeneous regions. Thirdly, five machine learning based downscaling models were evaluated and compared to select the optimal downscaling model, namely, extra trees model. Fourthly, in order to prevent the model from overfitting and underfitting, it is necessary to tune the parameters of the extra trees model. Finally, after using the downscaling model to predict

the high-resolution air temperature, residuals were fitted on the downscaling results to obtain the final downscaled high-resolution air temperature.

In Chapter, maximum likelihood classification method was developed to identify the urban structure which was the independent variables in the air temperature downscaling model. Based on the remote sensing data and Google Earth images, urban was divided into 17 types, including 10 built types and 7 land cover types. After accuracy test, we found that this method can be used to identity urban structure. Moreover, ENVI-met, a climate simulation tools, was applied to simulate the micro-climate of some typical urban structure. The results showed that urban structure had a great impact on air temperature, which provided theoretical support for the air temperature downscaling model.

In Chapter Four, Kanto Major Metropolitan Area was selected as a case study to verify whether the downscaling mentioned proposed in Chapter Two was feasible and efficient. K-fold cross-validation method and measured air temperature data obtained from meteorological station were used to test the precision. It had been verified that the air temperature downscaling method has high accuracy and can be applied to the subsequent air temperature downscaling in various metropolitan areas.

In Chapter Five, five metropolitan areas of different climatic zones in Japan were selected as the research object. The climate characteristics of each metropolitan areas were introduced. Based on the above urban structure identification method and air temperature downscaling method. The urban structure which is necessary for the downscaling model was identified. Then, combined with the digital evaluation model which indicate altitude, the monthly average air temperature in January and August and annual average air temperature in 2010 were downscaled from 1 kilometer to 250 meters for the next analysis and comparative study of atmosphere urban heat islands in each metropolitan area.

In Chapter Six, based on the air temperature spatial distribution in each metropolitan area obtained in the previous, the urban heat island intensity and urban heat island ratio index were introduced to evaluate and compare atmosphere urban heat island. Densely inhabit districts were defined as urban areas, and other regions were referred to suburb area, the average air temperature difference of the two regions was the urban heat island. Meanwhile, according to the normalized air temperature, urban area was divided into seven temperature zones, from low temperature zone to extremely high temperature zone. High temperature zone and extremely high temperature zone were defined as urban heat island zone, then the urban heat island ratio index was calculated. According to these two indicators, the atmosphere urban heat islands of each metropolitan area were compared horizontally, and then the atmosphere urban heat island in winter (January) and summer (August) were also contrasted.

In Chapter Seven, previous scholars indicated that there were some great differences between

atmosphere urban heat island and surface urban heat island. Limited by the weather conditions of remote sensing data, land surface temperature of only Kanto Major Metropolitan Area and Kinki Major Metropolitan Area could be retrieved for the surface urban heat island analysis. According to the normalized air temperature and normalized land surface temperature, we calculated the normalized urban heat island of the two metropolitan areas separately and compared atmosphere and surface urban heat island. In addition, four landscape pattern metrics were introduced to evaluate the impact of different urban structure on atmosphere urban heat island and surface urban heat island.

In Chapter Eight, the whole summary of each chapter has been presented.

CONTENTS

Preface

Acknowledgement

Abstract

Chapter 1. Research Background and Purpose of the Study

Chapter 2. Downscale Air Temperature Prediction based on Machine Learning Model

Chapter 3. Urban Structure Identification and its Implication of Urban Climate

Chapter 4. Case Study and Precision Evaluation

Chapter 5. Downscale Air Temperature Prediction of Metropolitan Area by Extra Trees Model

Chapter 6. Atmosphere Urban Heat Island Analysis of Metropolitan Area based on High-resolution Air Temperature

Chapter 7. Comparative Study of Atmosphere Urban Heat Island and Surface Urban Heat Island

Chapter 8. Conclusion

CONTENTS OF FIGURES

CONTENTS OF TABLES

Chapter 1. Research Background and Purpose of the Study

1.1. Introduction

The process of urbanization is advancing rapidly around the world, and the connection with global climate change is getting closer. Urbanization refers to the population shift from rural to urban areas, the decrease in the proportion of people living in rural areas[1,2]. Today, about 55% of the world's population lives in urban areas, and this proportion is expected to increase to 68% by 2050 [3]. Urbanization has spread rapidly throughout the western world. Since the 1950s, it has also begun to occupy an important place in the developing world. At the beginning of the 20th century, only 15% of the world's population lived in cities [4]. By 1950, the urban population has reached 751 million, accounting for about 29.6% of the total population. It is a stage of rapid urban population growth after 1950. According to the UN, the year 2007 witnessed the turning point when more than 50% of the world population were living in cities, for the first time in human history. And in 2018, the world's urban population has increased to 4.2 billion. According to Figure 1.1, despite the relatively low level of urbanization, Asia still accounts for 54% of the world's urban population, followed by Europe and Africa with 13% each. After entering the 21st century, the urbanization of Asia and Africa is obviously faster than that of other continents. Although the area of the city is not large, the urban population accounts for more than half of the total population. Therefore, the urban climate problem directly affects the residents' life and production activities, and also affects the health of the urban population.
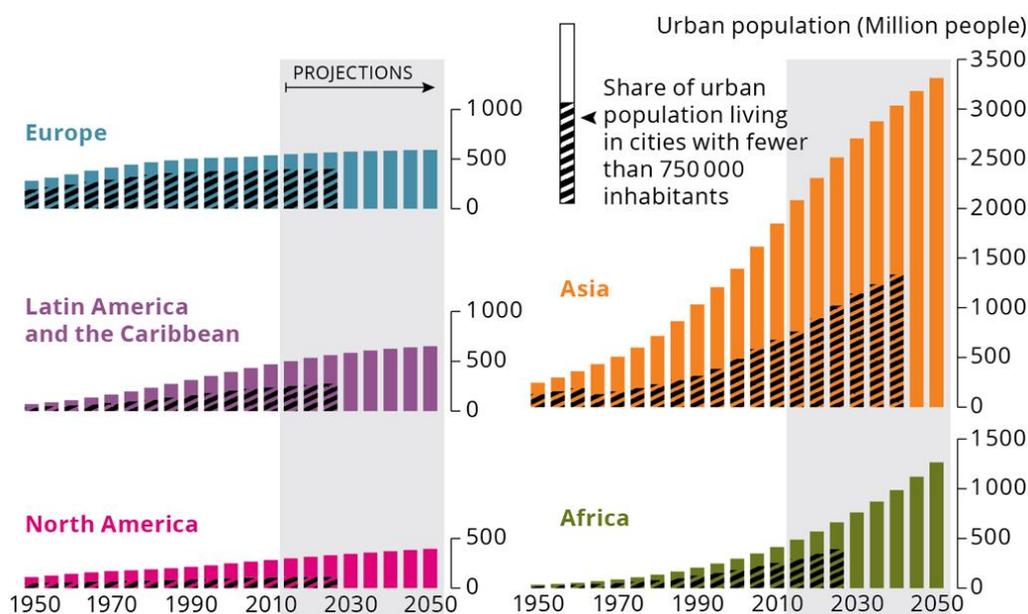
Figure 1.1. Urban trends by world regions.
(Source: European Environment Agency)

With the development of urbanization, there are many urban phenomena occurred. This includes population issues, environmental issues, social issues, and many other aspects[5]. Rapid urbanization

has led to drastic changes in land use and land cover, resulting in a decrease in latent heat flux and an increase in sensible heat flux, which has a significant impact on the climate [6]. Most notably, it causes global warming. Urbanization has led to a continuous increase in greenhouse gas production (mainly CO_2). Figure 1.2 shows the global and Japanese air temperature trends. The blue line represents the air temperature variation trend around the world, and the orange line represents the trend in Japan. It is obvious that the air temperature has risen by about 1.6 ℃ globally. And in Japan, the air temperature has risen by about 1.5 ℃. Generally speaking, the trends of air temperature in Japan and all the world are almost equal. A major manifestation of global warming is the intensification of urban heat island.
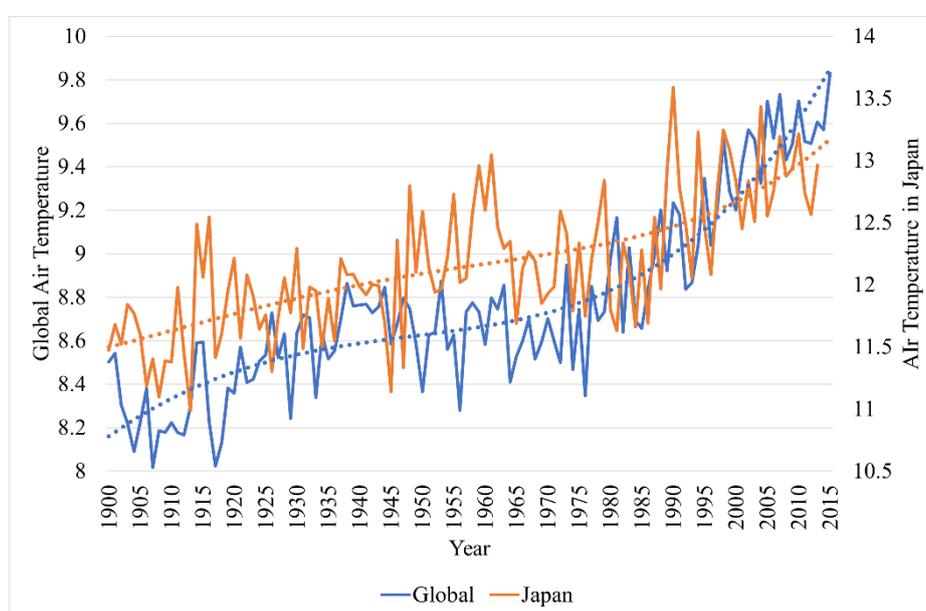
Figure 1.2. Air temperature trends around world and in Japan from 1900 to 2015.
(Source: World Bank Open Data)

Changes in urban climate have brought many ecological problems, for example, the common "heat island", "dry island", "wet island", etc. Among them, urban heat island is one of the most significant urbanization climate effects, and it is widespread in almost all urbans in the world. Urbanization has caused the natural ground to be replaced by the artificial impermeable surface, which affects the material circulation and energy exchange process on the surface, leading to the continuous deterioration of the environment. The larger the urban size and population density in the city, the more prominent the impact of the side effect of urban heat islands, including high temperature disasters and extreme weather. As the main and common climatic feature of all urbans, the urban heat island affects urban water bodies, urban soils, and urban residents' health.

For the research of urban heat island, there are mainly three methods: traditional observation, numerical simulation, and remote sensing inversion. Faced with different urban heat islands,

(including atmosphere urban heat island and surface heat island), the data sources used are not the same. Remote sensing inversion is mostly applied in the study ode surface urban heat islands. However, it is tedious to obtain the air temperature distribution required in the study of atmosphere urban heat island. Japan National Land Numerical Information provided the air temperature distribution throughout Japan with a spatial resolution of only 1 kilometer. It is not enough to support high-resolution atmosphere urban heat island research. Temperature downscaling based on the regression of urban structure and urban climate can effectively improve the resolution of air temperature distribution data, facilitate the analysis and research of atmosphere urban heat island, and provide effective responses to the government and policymakers.

1.2. Research Background

1.2.1 Introduction of Urban Heat Island

Urban heat island refers to a climatic phenomenon in which the temperature in the urban area is higher than that in suburbs area due to the changes in the underlying surface structure and nature and the population density and other factors [7–10]. The concept of urban heat island was first proposed by Howard in 1833 in response to the phenomenon that the temperature in central London is higher than that in the suburbs.

Currently, the causes of urban heat islands are known and mainly concentrated in the following three aspects (Figure 1.3):

- urban morphology causes: high-density buildings cause short and long wave radiation to be trapped in the urban [11], and the high roughness buildings also reduces the wind speed, thereby preventing heat dissipation [12–14].
- urban surface causes: a high percentage of impervious surface reduces water storage, so it cannot be cooled by evapotranspiration like vegetation [15]. Meanwhile, artificial urban materials (eg. asphalt) will absorb short and long wave radiation which contributes to ground and air overheating [16].
- human activity causes: inhabitants and appliances release anthropogenic heat, including the traffic [17,18]
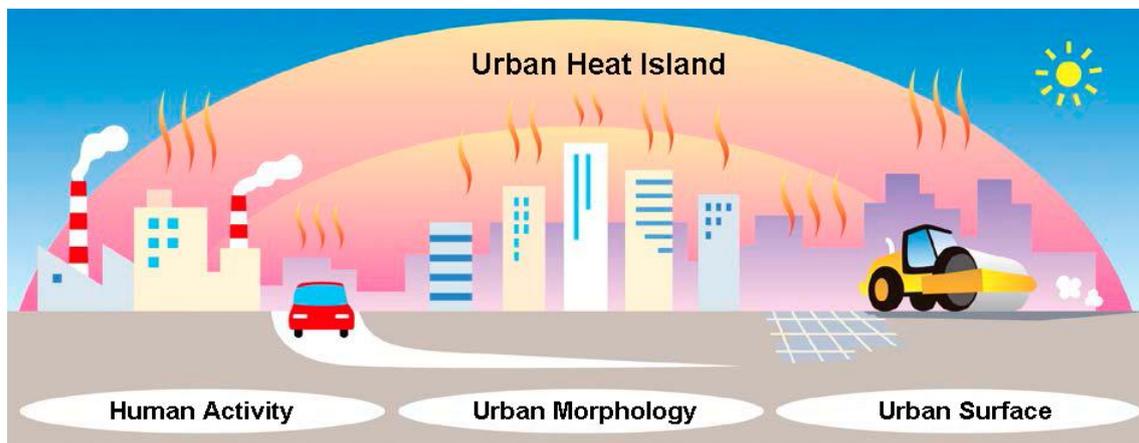
Figure 1.3. The causes of urban heat islands.

(Source: Heat Island Countermeasure Guidelines, 2012 Edition)

According to the Intergovernmental Panel on Climate Change (IPCC) results, UHI may become a serious issue in the next decades. With the increasing severity of UHI, the impact on the urban ecological environment is also multifaceted. Since the appropriate temperature is an important

ecological factor for biological growth and activities, UHI have changed biological physiological activities, phenology, and their distribution. Previous studies have shown that the increase in urban temperature will advance physiological time, including germination, flowering, etc., and delay the falling leaves time [19]. UHI will not only endanger the normal growth of vegetation, but also directly affect the health of urban residents. They can potentially increase the magnitude and duration of heat waves within cities. Related studies have found that the mortality rate during the heat wave increases exponentially with the maximum, and UHI exacerbates this effect [20]. Increasing UHI can cause heat stroke, physical exhaustion, syncope and cramps [21]. Not only health, but heat can also affect behavior. It can make people more irritable and aggressive. Meanwhile, UHI promote the formation of clouds and accelerate their movement, thereby increasing rainfall in urban areas and changing urban precipitation characteristics and hydrological conditions [22,23], and forming "urban rain islands". The existence of UHI makes air exchange between the urban center and suburbs form a closed circulation, which causes the concentration of urban air pollutants, including volatile organic compounds , carbon monoxide , nitrogen oxides , and particulate matter [24]. The warm airflow in urban contains a large amount of dust and smog. During the movement of the airflow, it may land on the urban and its surrounding area, or gather in the sky about the urban, and combine with the water vapor in atmosphere to form "urban pollution islands" and "urban fog islands". These air pollutants and fogs reduce the visibility and increase the urban pollution level to cause harm to the human respiratory system. UHI not only have a negative impact on the quality of the urban environment and residents' health, but also bring a great economic burden. The continuous high temperature will increase the energy consumption of urban industrial and commercial electricity, residential electricity, etc., resulting in electricity tensions.

According to the difference of research objects, UHI are divided into atmosphere urban heat island (AUHI) and surface urban heat island (SUHI) [25,26]. Early research on UHI was mainly proposed for AUHI, which refers to the difference between the air temperature in urban area and that in the suburbs. It usually depends on fixed or mobile meteorological observation stations to obtain discrete air temperature data values, both temporal and spatial resolution are not reach standards. It was not until 1972 that Rao first proposed the use of thermal infrared remote sensing to study UHI, making a new stage in UHI research [27]. The temperature obtained by thermal infrared remote sensing is the radiant temperature of the surface, which can cover a large area of the earth's surface, with good spatial continuity and high spatial resolution [25,28]. With the use of thermal infrared remote sensing in the study of UHI, the meaning of urban heat island has also changed slightly, and SUHI represented by land surface temperature.

1.2.2 The Application of Temperature Downscaling Prediction in UHI Research

Image fusion, spatial sharpening [29–31], downscaling [32,33], and disaggregation [34] are some terms that describe the methods that improve the spatial resolution of input data based on auxiliary data with higher spatial resolution [32,35]. Downscaling methods mainly include two categories, namely dynamical downscaling and statistical downscaling [36].

Dynamic downscaling refers to the use of a regional climate model driven by global climate model output or reanalysis data to generate regional climate information [37–40]. This method has been widely used in the prediction of climate change [41,42]. In contrast, statistical downscaling is a method based on the invariance of the regression at different scales. In the past 20 years, various statistical downscaling methods have been developed, which have a range of application in climate and climate change research. They are mainly divided into three types: analogs, weather typing, and regression [43]. Among the three methods, the regression method is relatively easy to implement, and the calculation requirements are small, so the regression method is the widely used method.

Each downscaling method has the advantages and disadvantages in application. The statistical downscaling method has low requirements for calculation and can be easily applied, however, it requires long-term, high-quality surface observation to establish a reliable statistical regression between independent variables and dependent variables. In contrast, dynamic downscaling methods involves the use of a regional numerical model that contains the complete physical set, and this method is based on physics, and computationally expensive. Previous studies [44–46] found that when downscaling the current climate, the results of the two methods are very similar, but they are very different in predicting climate. Since this research mainly analyzed and compared the urban heat island based on the current climate data. And the prediction of future climate is not involved. Therefore, this study used statistical downscaling method for the following research.

Temperature downscaling prediction was originally generated to meet the needs of agricultural remote sensing monitoring, because crops have a short growth cycle and require a high temporal resolution land surface temperature for early identification of crop, drought monitoring, etc. Nowadays, the research on downscaling in urban heat islands mainly focuses on land surface temperature, that is, the research on temperature downscaling is mainly on surface urban heat islands. Surface urban heat island based on land surface temperature is a phenomenon with high temporal and spatial variability. In this case, land surface temperature data obtained from satellite thermal infrared remote sensing with high temporal and spatial resolution is required. However, due to the technical limitations, the thermal sensor in satellite have a trade-off between spatial and temporal resolution [47]. The higher spatial resolution, the lower the temporal resolution, and vice versa. The characteristic of thermal sensor on satellite is shown in Table 1-1. For example, Landsat 4-5 TM, Landsat 7 ETM+, and Landsat 8 OLI/TIRS can provide high spatial resolution, however, the use in urban climate research is limited due to the limited availability of image data at night and low temporal resolution. On the other hand,

remote sensing data with low spatial resolution such as Moderate Resolution Imaging Spectroradiometer (MODIS) can provide land surface temperature with high temporal resolution and provide land surface temporal data of the observation area twice a day.

Table 1-1 The characteristic of thermal sensor on satellite

Sensor–Satellite	Spatial Resolution (m)	Wavelength (μm)	Temporal Resolution
AVHRR-NOAA*	1100	Band 4: 10.3-11.3, Band 5: 11.5-12.5	Twice a day
MODIS-Terra*	1000	Band 31-36: 10.78-14.39	Twice a day
TM-Landsat 5**	120 (resampled to 30)	Band 6: 10.40-12.50	16 days
ETM+-Landsat 7**	60 (resampled to 30)	Band 6: 10.40-12.50	16 days
TIRS-Landsat 8**	100 (resampled to 30)	Band 10: 10.60-11.19 Band 11: 11.50-12.51	16 days
ASTER–Terra**	90	Band 10-14:8.125-11.65	16 days
AATSAR–Envisat	1000	11μm band, 12 μm band	35 days
ATSAR – ERS	1000	11μm band, 12 μm band	35 days

Note: * represents the remote sensing data for land surface temperature with high temporal and low spatial resolution; ** represents the remote sensing data for land surface temperature with low temporal and high spatial resolution, the spatial resolution of TM on Landsat 5, ETM+ on Landsat 7, and TIRS on Landsat 8 can be resampled to 30 m.

Land surface temperature downscaling has been widely applied in the surface urban heat island research, however there are still few previous studies on temperature spatial downscaling in atmosphere urban heat island. The main reason is that air temperature images are difficult to obtain. It is not like land surface temperature that can be obtained on a large scale through satellite remote

sensing.

Japan National Land Numerical Information provides monthly and annual average weather data in 2010m, including air temperature and precipitation, etc. based on the meteorological station and meteorological observation points [48]. With the help of downscaling in the application of land surface temperature, we can also downscale the air temperature through the same method.

1.3. Review of Previous Study

1.3.1.   Study on Urban Heat Island

In 1833, Howard first proposed the urban heat island in the scientific magazine "The climate of London", pointing out that the air temperature in the central of London was higher than that in the surrounding countryside [7]. Subsequently, researchers from various countries have carried out more and more studies on urban heat islands.

Previous studies of urban heat islands were mainly based on the long-term observation of meteorological station and meteorological observation points. After statistics and analysis, the characteristic and evolution of urban heat islands were summarized. Based on the measured meteorological data, Kazimierz and Krzysztof [49] conducted research on Lodz, Poland, and found that the urban heat island is closely related to wind speed, and the urban heat island intensity is the highest at night on a clear and windless day. And Morris et al. [50] concluded that the urban heat island is influenced by cloud cover and wind speed through the research on Melbourne, Australia. The research of Gedzelman et al. [51] and Kim et al. [52] found that there are great difference in the urban heat island structure between inland urban and coastal urban. The urban heat island intensity in coastal urban is weak, and the time is delayed. The analysis shows that sea-land breeze circulation is the main reason.

In the past 20 years, scholars have mainly used satellite remote sensing data in urban land surface temperature for surface urban heat island research. Carnahan and Larson [53] used Landsat TM thermal infrared band data to study the land surface temperature characteristics of urban areas and found that urban areas showed lower land surface temperature than rural areas. Gallo et al. [54] used satellite remote sensing data to study Seattle, the United States, and found that the normalized difference vegetation index (NDVI) is inversely proportional to the land surface temperature. Mackey et al. [55] took Chicago, the United States, as an example, and used Landsat remote sensing images to study the effect of urban scale on the reduction of urban heat island. Camilloni & Barrucand [56] took Buenos Aires, the capital of Argentina, as an example. Based on long-term land surface temperature retrieved from thermal infrared remote sensing data, they studied the annual and seasonal changes of the urban heat island.

The urban heat island research based on the remote sensing data can reflect the spatial distribution of land surface temperature, but the air temperature cannot be obtained from remote sensing to study the spatial distribution of atmosphere urban heat island. The resolution of the air temperature spatial distribution provided by Japan National Land Numerical Information is only 1 km. Therefore, in order to study the atmosphere urban heat island more accurately, we spatially downscaled the air temperature data to obtain the high-resolution spatial distribution of air temperature for high-resolution atmosphere urban heat island research.

1.3.2.　　　Study on Temperature Downscaling

The statistical downscaling method refers to the scale invariance based on the statistical regression between air temperature and some explanatory variables, that is, an algorithm that uses high-resolution urban structure and altitude data to enhance the spatial resolution of low-resolution air temperature images. This algorithm is widely used to enhance the spatial resolution of low-resolution thermal infrared temperature images. Based on the assumption that the relationship between land surface temperature and normalized difference vegetation index is constant under different spatial resolutions, Kustas et al. [57] proposed the DisTrad algorithm that uses the relationship between land surface temperature and NDVI to achieve sub-pixel decomposition of thermal infrared images. Agam et al. [29] proposed the TsHARP algorithm on the basis of the DisTrad algorithm. Through the comparison of four vegetation index – land surface temperature relationship models, they proved that the unary linear regression between vegetation index and the land surface temperature is optimal for the land surface temperature downscaling. Hutengs and Vohland [58] establish a random forest model based on machine learning algorithm for a complex landscape in the Eastern Mediterranean, the Jordan River Region to downscale land surface temperature.

Most of the existing statistical downscaling methods are applied to land surface temperature. We used the same principle to downscale the air temperature data and verify accuracy.

1.4. Research Content and Purpose

1.4.1.   Research Content

This study focused on the use of downloaded high-resolution air temperature data to analyze the atmosphere urban heat island. Based on the previous study and theory analysis, using a statistical downscaling method based on machine learning algorithms, combined with low-resolution air temperature obtained from Japan National Land Numerical Information and high-resolution urban structure identified from remote sensing data and digital elevation model, we obtained high-resolution air temperature distribution through downscaling and used these data to study the atmosphere urban heat islands of metropolitan areas in various climatic zones in Japan. The research flow is shown in Figure 1.4.

Previous Study	CHAP 1 Research Background and Purpose of the Study	
Research Method	CHAP 2 Downscale Air Temperature Prediction based on Machine Learning Model	
Precision Analysis	CHAP 3 Urban Structure Identification and its Implication of Urban Climate	CHAP 4 Case Study and Precision Evaluation
Urban Heat Island of Metropolitan Area in Japan	CHAP 5 Downscale Air Temperature Prediction of Metropolitan Area by Extra Trees Model	
	CHAP 6 Atmosphere Urban Heat Island Analysis of Metropolitan Area based on High-resolution Air Temperature	
	CHAP 7 Comparative Study of Atmosphere Urban Heat Island and Surface Urban Heat Island	
Conclusion	CHAP 8 Conclusion	

Figure 1.4. Research flow

● Previous Study

In Chapter One, we introduce the research background and significance of urban heat island and the application of temperature spatial downscaling. In addition, we reviewed the previous studies of urban heat island and temperature spatial downscaling method. Finally, we elaborated on the purpose of this study.

● Research Method

In Chapter Two, we mainly introduced the spatial downscaling method. Firstly, we explained that the basic principle of the statistical downscaling method is that the regression between air temperature and underlying surface characteristics remain unchanged under different scales. Next, we compared

and selected the different types of downscaling models, that mainly include two parts. The first part is to select a random forest model based on machine learning algorithms from three types of downscaling models. The second part is to select the optimal model from machine learning models and determine the independent variables of the downscaling model. Then, we introduced the parameter tuning method of downscaling model to obtain the optimal model and use this model to predict high-resolution air temperature. Finally, we fitted the residual of the predicted air temperature and get the final downscaling result.

- Data Preprocessing and Precision Analysis

Chapter Three mainly includes the identification of urban structure and the study of the influence of urban structure on air temperature. We used the maximum likelihood classification method to divide the urban area into 17 types combined with Landsat series remote sensing data and Google Earth satellite images and use the error matrix to verify the accuracy. Then, ENVI-met was applied to simulate the micro-climate of several typical urban structure to analyze the impact of urban structure on air temperature, so as to provide a strong theoretical basis for the downscaling method mentioned in Chapter Two.

In Chapter Four, Kanto Major Metropolitan Area was selected as the case study to verify the feasibility of the above downscaling method. In this chapter, we use two verification methods. The first is to cross-validate the predicted low-resolution air temperature with the original air temperature. Secondly, we obtained the air temperature data of all meteorological stations and meteorological observation points in the study area, and then compared them with the downscaling high-resolution air temperature data to verify the accuracy of the downscaling method.

- Urban Heat Island of Metropolitan Area in Japan

In Chapter Five, five metropolitan areas in each climatic zone in Japan were selected as the research object. And we selected the monthly average air temperature in January and August and annual average air temperature in 2010 as the data source. According to the downscaling method, the low-resolution air temperature of metropolitan areas was downscaled in this chapter for the further atmosphere urban heat island analysis.

In Chapter Six, based on the downscaled high-resolution air temperature in Chapter Five, we calculated the urban heat island intensity and urban heat island ratio index for the urban heat island evaluation of each metropolitan area to compare and analysis the atmosphere urban heat island of each metropolitan area.

In Chapter Seven, in order to compare the surface urban heat island and the atmosphere urban heat island, we retrieved the land surface temperature of Kanto Major Metropolitan Area and Kinki Major Metropolitan Area. Then, urban heat islands were normalized to the range between 0 and 1. Meanwhile,

we introduced two landscape pattern metrics and used multiple linear regression model to indicate the effect of urban structure on urban heat island.

- Conclusion

Chapter Nine presented the whole summary of each chapter.

1.4.2.   Research Purpose

Nowadays, many previous research on urban heat island and temperature spatial downscaling method were carried out. However, for the research on atmosphere urban heat island, the current data source and method have the following problems.

- Reason for atmosphere urban heat island research

Most of the research on urban heat island focuses on surface urban heat island. Due to the shortage of air temperature distribution data, there are few studies on atmosphere urban heat island. According to the Web of Science [59], we counted the frequency of atmosphere and surface urban heat island in previous studies during 1965 and 2021. And the results as shown in Figure 1.5 indicate that the study of atmosphere urban heat island is significantly less than that of surface urban hear island since 1965, and is only less than one-fifth of that on surface urban heat island.
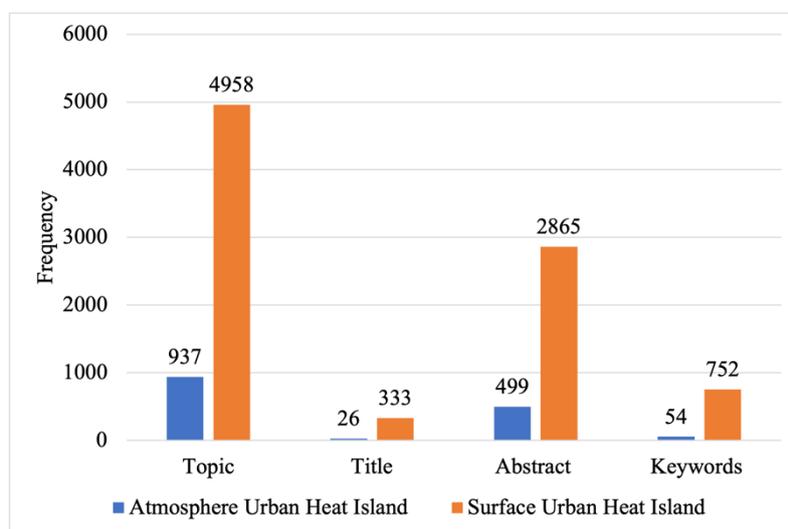
Figure 1.5. The frequency of atmosphere and surface urban heat island in previous studies

Previous studies have also shown that there is a great difference between the atmosphere and surface urban heat island due to the different formation mechanisms. Therefore, we mainly studied the atmosphere urban heat islands of metropolitan areas in different climatic zones in Japan, and selected

Kanto and Kinki Major Metropolitan Areas to compare atmosphere and surface urban heat islands.

- Difficulty in obtaining data

With the development of remote sensing technology, the acquisition of land surface temperature has become more convenient and efficient, which has greatly promoted the surface urban heat island research. In contrast, the accuracy and efficiency of the estimation of air temperature data based on remote sensing data is low, therefore, the obtaining of air temperature is mainly based on a large number of meteorological stations and meteorological observation points. The lack of data has led to less research on atmosphere urban heat island based on air temperature.

However, in recent years, some government agencies and scientific research institution in various countries have increased research on meteorology, so they have also provided air temperature distribution data in some regions based on meteorological data. In order to study the atmosphere urban heat islands of metropolitan areas in different climatic zones in Japan, we obtained air temperature distribution data from the National Land Numerical Information.

- The resolution of air temperature is low

The resolution of air temperature obtained from National Land Numerical Information is only 1 km. In order to study atmosphere urban heat island, we need to downscale the air temperature. In this study, the resolution was downscaled from 1 km to 250 m.

Previous studies show that the statistical downscaling method for land surface temperature is very mature, which mainly predicted the high-resolution land surface temperature based on the regression model of the land surface temperature and underlying surface characteristics. Scholars used the principle to develop a variety of models for land surface temperature downscaling, including DisTrad model established by the regression between land surface temperature and NDVI, the TsHARP model based on the further optimization from DisTrad model, HUTS model and other commonly used temperature downscaling methods.

This study also used the principle of invariance of the regression model to establish an air temperature spatial downscaling model to improve the air temperature spatial resolution for the more detailed atmosphere urban heat island research.

- Reason to use machine learning models

The first land surface temperature downscaling method was mainly used in the agricultural field. The downscaling accuracy for highly heterogeneous areas is low and the research area is small, therefore, only traditional statistical downscaling method, such as TsHARP model, can be applied for research, which will be proved in Section 2.2.

The research objects of this study are five metropolitan areas in Japan with high spatial

heterogeneity, and the area of the research area is huge. Taking the Kanto Major Metropolitan Area as an example, the number of low-resolution data used to establish the model is about 13 thousand, and the number of high-resolution independent variables for prediction is about 217 thousand. Obviously, the traditional regression formula cannot accurately describe the relationship between the independent variables and dependent variables. Therefore, it is necessary to introduce machine learning models for spatial downscaling of air temperature. The machine learning model can represent the regression between air temperature and several underlying surface characteristics more accurately, thereby greatly improving the accuracy and efficiency of downscaling.

- Selection of independent variables

Previous studies generally set remote sensing bands or some high-order land surface parameters as independent variables to establish models and predict high-resolution land surface temperature. The research object of this study is air temperature, meanwhile, the past research results show that urban structure has a great influence on air temperature, which has been proved in Chapter 3. Therefore, we used urban structure as an independent variable of the air temperature downscaling model and quantified it. Then, according to lapse rate, the air temperature drops by about 0.6 °C every 100 m rises on average. It indicates that altitude can determine air temperature. Thus, we also add DEM, which represents altitude, as the independent variable. In summary, we set DEM and urban structure as independent variables.

Reference

1.   Varenyam, A.; Abhijit, M. *Ecological Wisdom Inspired Restoration Engineering*; 2019;

2.   Allender, S.; Foster, C.; Hutchinson, L.; Arambepola, C. Quantification of urbanization in relation to chronic diseases in developing countries: A systematic review. *J. Urban Heal.* 2008, *85*, 938–951.

3.   Desa, U. 2018 Revision of World Urbanization Prospects. *May* **2018**, *16*, 2018.

4.   Buckley, R.M.; Patricia, C.A. *Urbanization and Growth*; The World Bank, 2008;

5.   Redman, C.L.; Jones, N.S. The environmental, social, and health dimensions of urban expansion. *Popul. Environ.* **2005**, *26*, 505–520, doi:10.1007/s11111-005-0010-1.

6.   Kalnay, E.; Cai, M. Impact of urbanization and land-use change on climate. *Nature* **2003**, *423*, 528–531, doi:10.1038/nature01675.

7.   Howard, L. *The climate of London: deduced from meteorological observations made in the metropolis and at various places around it*; Harvey and Darton, J. and A. Arch, Longman, Hatchard, S. Highley [and] R. Hunter, 1833; Vol. 3;.

8.   Oke, T.R. *Boundary layer climates*; Routledge, 2002; ISBN 1134951337.

9.   Oke, T.R.; Maxwell, G.B. Urban heat island dynamics in Montreal and Vancouver. *Atmos. Environ.* **1975**, *9*, 191–200, doi:10.1016/0004-6981(75)90067-0.

10.   Oke, T.R. The energetic basis of the urban heat island. *Q. J. R. Meteorol. Soc.* **1982**, *108*, 1–24, doi:10.1002/qj.49710845502.

11.   Bernabé, A.; Musy, M.; Andrieu, H.; Calmet, I. Radiative properties of the urban fabric derived from surface form analysis: A simplified solar balance model. *Sol. Energy* **2015**, *122*, 156–168, doi:10.1016/j.solener.2015.08.031.

12.   Britter, R.E.; Hanna, S.R. Flow and dispersion in urban areas. *Annu. Rev. Fluid Mech.* **2003**, *35*, 469–496.

13.   Mirzaei, P.A.; Haghighat, F. Approaches to study Urban Heat Island - Abilities and limitations. *Build. Environ.* **2010**, *45*, 2192–2201, doi:10.1016/j.buildenv.2010.04.001.

14.   Magli, S.; Lodi, C.; Contini, F.M.; Muscio, A.; Tartarini, P. Dynamic analysis of the heat released by tertiary buildings and the effects of urban heat island mitigation strategies. *Energy Build.* **2016**, *114*, 164–172, doi:10.1016/j.enbuild.2015.05.037.

15.   Asaeda, T.; Ca, V.T. The subsurface transport of heat and moisture and its effect on the environment: A numerical model. *Boundary-Layer Meteorol.* **1993**, *65*, 159–179, doi:10.1007/BF00708822.

16.   Santamouris, M. Cooling the cities - A review of reflective and green roof mitigation technologies to fight heat island and improve comfort in urban environments. *Sol. Energy* **2014**, *103*, 682–703, doi:10.1016/j.solener.2012.07.003.

17.   Sailor, D.J. A review of methods for estimating anthropogenic heat and moisture emissions in the

urban environment. *Int. J. Climatol.* **2011**, *31*, 189–199, doi:10.1002/joc.2106.

18. De Munck, C.; Pigeon, G.; Masson, V.; Meunier, F.; Bousquet, P.; Tréméac, B.; Merchat, M.; Poeuf, P.; Marchadier, C. How much can air conditioning increase air temperatures for a city like Paris, France? *Int. J. Climatol.* **2013**, *33*, 210–227, doi:10.1002/joc.3415.

19. Sukopp, H. Urban Ecology — Scientific and Practical Aspects. In *Urban Ecology*; Springer Berlin Heidelberg, 1998; pp. 3–16.

20. Buechley, R.W.; Van Bruggen, J.; Truppi, L.E. Heat island = death island? *Environ. Res.* **1972**, *5*, 85–92, doi:10.1016/0013-9351(72)90022-9.

21. Kovats, R.S.; Hajat, S. Heat Stress and Public Health: A Critical Review. *Annu. Rev. Public Health* **2008**, *29*, 41–55, doi:10.1146/annurev.publhealth.29.020907.090843.

22. van Heerwaarden, C.C.; de Arellano, J.V.G. Relative humidity as an indicator for cloud formation over heterogeneous land surfaces. *J. Atmos. Sci.* **2008**, *65*, 3263–3277, doi:10.1175/2008JAS2591.1.

23. Shepherd, J.M.; Pierce, H.; Negri, A.J. Rainfall modification by major urban areas: Observations from spaceborne rain radar on the TRMM satellite. *J. Appl. Meteorol.* **2002**, *41*, 689–701, doi:10.1175/1520-0450(2002)041<0689:RMBMUA>2.0.CO;2.

24. Koppe, C.; Sari Kovats, R.; Menne, B.; Jendritzky, G.; Wetterdienst, D.; Organization, W.H. Heat-waves: risks and responses. **2004**.

25. Voogt, J.A.; Oke, T.R. Thermal remote sensing of urban climates. *Remote Sens. Environ.* **2003**, *86*, 370–384, doi:10.1016/S0034-4257(03)00079-8.

26. Yuan, F.; Bauer, M.E. Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in Landsat imagery. *Remote Sens. Environ.* **2007**, *106*, 375–386, doi:10.1016/j.rse.2006.09.003.

27. Rao, P.K. Remote sensing of urban "heat islands" from an environmental satellite. *Bull. Am. Meteorol. Soc.* **1972**, *53*, 647–648.

28. Arnfield, A.J. Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island. *Int. J. Climatol.* **2003**, *23*, 1–26, doi:10.1002/joc.859.

29. Agam, N.; Kustas, W.P.; Anderson, M.C.; Li, F.; Neale, C.M.U. A vegetation index based technique for spatial sharpening of thermal imagery. *Remote Sens. Environ.* **2007**, *107*, 545–558, doi:10.1016/j.rse.2006.10.006.

30. Jeganathan, C.; Hamm, N.A.S.; Mukherjee, S.; Atkinson, P.M.; Raju, P.L.N.; Dadhwal, V.K. Evaluating a thermal image sharpening model over a mixed agricultural landscape in India. *Int. J. Appl. earth Obs. Geoinf.* **2011**, *13*, 178–191.

31. Zhan, W.; Chen, Y.; Zhou, J.; Li, J.; Liu, W. Sharpening thermal imageries: A generalized theoretical framework from an assimilation perspective. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 773–789, doi:10.1109/TGRS.2010.2060342.

32. Pardo-Igúzquiza, E.; Chica-Olmo, M.; Atkinson, P.M. Downscaling cokriging for image sharpening. *Remote Sens. Environ.* **2006**, *102*, 86–98, doi:10.1016/j.rse.2006.02.014.

33. Liu, D.; Pu, R. Downscaling thermal infrared radiance for subpixel land surface temperature retrieval. *Sensors* **2008**, *8*, 2695–2706, doi:10.3390/s8042695.

34. Merlin, O.; Duchemin, B.; Hagolle, O.; Jacob, F.; Coudert, B.; Chehbouni, G.; Dedieu, G.; Garatuza, J.; Kerr, Y. Disaggregation of MODIS surface temperature over an agricultural area using a time series of Formosat-2 images. *Remote Sens. Environ.* **2010**, *114*, 2500–2512, doi:10.1016/j.rse.2010.05.025.

35. Zakšek, K.; Oštir, K. Downscaling land surface temperature for urban heat island diurnal cycle analysis. *Remote Sens. Environ.* **2012**, *117*, 114–124, doi:10.1016/j.rse.2011.05.027.

36. Sylwia, T.; Emilie, S. *A Review of Downscaling Methods for Climate Change Projections*; 2014;

37. Rockel, B. The Regional Downscaling Approach: a Brief History and Recent Advances. *Curr. Clim. Chang. Reports* 2015, *1*, 22–29.

38. Giorgi, F. Simulation of regional climate using a limited area model nested in a general circulation model. *J. Clim.* **1990**, *3*, 941–963.

39. Mass, C.F.; Ovens, D.; Westrick, K.; Colle, B.A. Does increasing horizontal resolution produce more skillful forecasts?: The Results of Two Years of real-Time Numerical Weather Prediction over the Pacific Northwest. *Bull. Am. Meteorol. Soc.* **2002**, *83*, 407–430.

40. Wang, Y.; Leung, L.R.; McGregor, J.L.; Lee, D.K.; Wang, W.C.; Ding, Y.; Kimura, F. Regional climate modeling: Progress, challenges, and prospects. *J. Meteorol. Soc. Japan* 2004, *82*, 1599–1628.

41. Sato, T.; Xue, Y. Validating a regional climate model's downscaling ability for East Asian summer monsoonal interannual variability. *Clim. Dyn.* **2013**, *41*, 2411–2426, doi:10.1007/s00382-012-1616-5.

42. Wang, D.; Menz, C.; Simon, T.; Simmer, C.; Ohlwein, C. Regional dynamical downscaling with CCLM over East Asia. *Meteorol. Atmos. Phys.* **2013**, *121*, 39–53, doi:10.1007/s00703-013-0250-z.

43. Gutiérrez, J.M.; San-Martín, D.; Brands, S.; Manzanas, R.; Herrera, S. Reassessing statistical downscaling techniques for their robust application under climate change conditions. *J. Clim.* **2013**, *26*, 171–188.

44. Flaounas, E.; Drobinski, P.; Vrac, M.; Bastin, S.; Lebeaupin-Brossier, C.; Stéfanon, M.; Borga, M.; Calvet, J.-C. Precipitation and temperature space–time variability and extremes in the Mediterranean region: evaluation of dynamical and statistical downscaling methods. *Clim. Dyn.* **2013**, *40*, 2687–2705.

45. Gutmann, E.D.; Rasmussen, R.M.; Liu, C.; Ikeda, K.; Gochis, D.J.; Clark, M.P.; Dudhia, J.; Thompson, G. A comparison of statistical and dynamical downscaling of winter precipitation

over complex terrain. *J. Clim.* **2012**, *25*, 262–281.

46.   Haylock, M.R.; Cawley, G.C.; Harpham, C.; Wilby, R.L.; Goodess, C.M. Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios. *Int. J. Climatol. A J. R. Meteorol. Soc.* **2006**, *26*, 1397–1415.

47.   Stathopoulou, M.; Cartalis, C. Downscaling AVHRR land surface temperatures for improved surface urban heat island intensity estimation. *Remote Sens. Environ.* **2009**, *113*, 2592–2605.

48.   National Land Numerical Information Available online: https://nlftp.mlit.go.jp/ksj/ (accessed on Jan 31, 2021).

49.   Klysik, K.; Fortuniak, K. Temporal and spatial characteristics of the urban heat island of Lodz, Poland. In Proceedings of the Atmospheric Environment; Elsevier Science Ltd, 1999; Vol. 33, pp. 3885–3895.

50.   Morris, C.J.G.; Simmonds, I.; Plummer, N. Quantification of the influences of wind and cloud on the nocturnal urban heat island of a large city. *J. Appl. Meteorol. Climatol.* **2001**, *40*, 169–182.

51.   Gedzelman, S.D.; Austin, S.; Cermak, R.; Stefano, N.; Partridge, S.; Quesenberry, S.; Robinson, D.A. Mesoscale aspects of the Urban Heat Island around New York City. *Theor. Appl. Climatol.* **2003**, *75*, 29–42, doi:10.1007/s00704-002-0724-2.

52.   Kim, Y.H.; Baik, J.J. Daily maximum urban heat island intensity in large cities of Korea. *Theor. Appl. Climatol.* **2004**, *79*, 151–164, doi:10.1007/s00704-004-0070-7.

53.   Carnahan, W.H.; Larson, R.C. An analysis of an urban heat sink. *Remote Sens. Environ.* **1990**, *33*, 65–71, doi:10.1016/0034-4257(90)90056-R.

54.   Gallo, K.P.; McNab, A.L.; Karl, T.R.; Brown, J.F.; Hood, J.J.; Tarpley, J.D. The use of a vegetation index for assessment of the urban heat island effect. *Int. J. Remote Sens.* **1993**, *14*, 2223–2230, doi:10.1080/01431169308954031.

55.   Mackey, C.W.; Lee, X.; Smith, R.B. Remotely sensing the cooling effects of city scale efforts to reduce urban heat island. *Build. Environ.* **2012**, *49*, 348–358, doi:10.1016/j.buildenv.2011.08.004.

56.   Camilloni, I.; Barrucand, M. Temporal variability of the Buenos Aires, Argentina, urban heat island. *Theor. Appl. Climatol.* **2012**, *107*, 47–58, doi:10.1007/s00704-011-0459-z.

57.   Kustas, W.P.; Norman, J.M.; Anderson, M.C.; French, A.N. Estimating subpixel surface temperatures and energy fluxes from the vegetation index-radiometric temperature relationship. *Remote Sens. Environ.* **2003**, *85*, 429–440, doi:10.1016/S0034-4257(03)00036-1.

58.   Hutengs, C.; Vohland, M. Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sens. Environ.* **2016**, *178*, 127–141, doi:10.1016/j.rse.2016.03.006.

59.   Web of Science Available online: https://clarivate.com/products/web-of-science/ (accessed on Jun 22, 2021).

Chapter 2. Downscale Air Temperature Prediction based on Machine Learning Model

2.1. The Basic Principles of Temperature Downscaling

The fundamental nature of temperature downscaling is to improve the spatial resolution of low-resolution temperature products by using high-resolution auxiliary parameters. And the principle was that the quantitative correlation between temperature and urban structure remained unchanged with different resolution; that is to say, the downscaling regression model between temperature and independent variables with low resolution, including DEM and urban structure, could still be suitable to the high-resolution model. It could be expressed as the following equation:

$$T_{HR} = f(P_{HR}) + \Delta T$$

(2-1)

$$\Delta T = T_{LR} - f(P_{LR})$$

(2-2)

where T_{HR} and T_{LR} represent high-resolution and low-resolution temperature, respectively; f represents the downscaling regression model between temperature and independent variables with both high and low resolution; P_{HR} and P_{LR} represent the high-resolution and low-resolution independent variables, including DEM and urban structure; and ΔT represents the residual.

2.2. Selection of Commonly used Downscale Model and Surface Urban Heat Island Analysis: A Case Study of Hangzhou

There are many commonly used downscaling models. In this section, we selected a square urban area with a side length of 25 km in Hangzhou, China as a research object, and used three different models to downscale land surface temperature (LST) obtained from Moderate Resolution Imaging Spectroradiometer (MODIS). The main objectives of this section were (1) to estimate the accuracy of the downscaled LST in a heterogeneous urban landscape (Hangzhou, China) based on three different models; (2) to assess the seasonal variation of the results during 2013 and 2014; (3) to confirm the change of the surface urban heat island (SUHI) of Hangzhou across four seasons; (4) to verify the feasibility of the optimal downscaling model combined with LST retrieved at a resolution of 30 m.

LST is an important parameter reflecting the interaction between surface and atmosphere at the regional and global scales [1]. LST is also a natural indicator closely related to human production and life. It can characterize the urban thermal environment [2,3] and is widely used in urban heat island analysis [4], soil moisture estimation [5], surface flux estimation [6] and other fields. Therefore, obtaining measures of LST is an important research objective in the fields of climate, ecology, hydrology, soil and urban studies. However, due to the restrictions of imaging conditions, existing remote sensing products have a contradiction between temporal resolution and spatial resolution. A single dataset cannot satisfy LST spatiotemporal monitoring and application research [7]. For example, the Landsat 8 Thermal InfraRed Sensor (TIRS) band has a spatial resolution of 100 m and can be resampled to 30 m to match multispectral bands. However, it has a long revisit period of about 16 days, and is greatly affected by weather [8,9]. On the contrary, MODIS, with the resolution of 1 km can obtain images four times per day. Thus, the fusion of multi-source remote sensing data based on their respective resolution advantages to obtain images with both high spatial resolution and high temporal resolution is a popular research topic in LST inversion and application.

LST downscaling, which create a composite of remote sensing images information with various spatial resolution, involves lowering the detail of high-resolution data to that of low-resolution data. Scholars have proposed a variety of downscaling methods, mainly divided into thermal sharpening (TSP) and temperature unmixing (TUM) [10]. The TSP method can improve the spatial resolution of thermal infrared band images, and the TUM method can obtain the LST information of different components in the same pixel. Kustas et al. [11] proposed a DisTrad method, which constructed a linear regression between LST and the normalized difference vegetation index (NDVI). This method achieved downscaling of LST from the kilometer level to the hundred-meter level. Based on the DisTrad method, Agam et al. [12] suggested the thermal sharpen (TsHARP) method, which used the NDVI as the regression kernel. Essa et al. [13] calculated the correlation between LST and remote sensing of various land use and land cover types, and then improved the DisTrad method based on this information. Weng et al. [14] further considered the LST trend and landscape heterogeneity, and

implemented the spatial-temporal fusion of LST based on radiance, proposing the spatial-temporal adaptive data fusion algorithm for temperature mapping (SAFAT) method, and successfully verifying the approach in Los Angeles, California.

The simple single-factor and multi-factor regression methods mentioned above cannot completely summarize the complex relationships between different scale factors and LST. Hutengs et al. [15] used the random forest (RF) model to downscale MODIS products from 1000 m to 250 m for the vegetation coverage area around the Jordan Valley. However, in this section, the land cover type in the study area was mainly vegetation, and mostly comprised a single type. Extension of the RF model to urban areas with complex underlying types needs further study. Generally, the most popular downscaling methods apply the NDVI, which, however, cannot solely explain the variation in LST in urban areas with complex surface types. Bonafoni et al. [16] proposed a traditional downscaling method combining both built-up and vegetation spectral indices that was demonstrated in Milan, Italy.

In validation processing, Govil et al. [17] used 30-m retrieved LST to validate 30-m downscaled LST of a humid tropical city. Hua et al. [18] verified a downscaling model based on retrieved LST and determined that the downscaling effects of various land cover types are different. Hutengs et al. [15] used a 240-m Enhanced Thematic Mapper Plus (ETM+)/LST map as a direct reference to evaluate downscaling results. Standard LST products inversed from ETM+/LST based on a mono-window algorithm were introduced to confirm the accuracy of the downscaling method in the research of Zhan et al. [19] Combined with the previous research on LST downscaling, most of these studies validated downscaling methods based on special time nodes, in which the scan time between several remote sensing products was the same or similar to existing high-resolution LST product correlations.

2.2.1.   Downscaling Models and Accuracy Evaluation

We defined a rectangular area in Hangzhou with a side length of 25 km as the study area. Landsat 8 OLI/TIRS with a resolution of 30 m and MODIS/LST products with a resolution of 1 km were used as the original data for this section. These data comprised a digital elevation model (DEM), which only represents height information without any further definition about the surface [20], the normalized difference vegetation index (NDVI) and the normalized difference built-up index (NDBI), calculated from Landsat 8 OLI [21] and other bands in Landsat 8 OLI, as independent variables. The dependent variable was pre-processed MODIS/LST products. The objective was to achieve LST downscaling from 1 km to 100 m to analyze SUHI during the day and night in four seasons, based on three different models, MLR, TsHARP and RF models. MLR and RF models are multivariate models with several independent variables, while the TsHARP model has only one independent variable. From another perspective, the RF model is a nonlinear regression model, the MLR model is a linear regression model, and the TsHARP model includes both linear regression and nonlinear regression models. The coefficient of determination (R^2) and root mean square error (RMSE) were used to

evaluate the accuracy of the downscaling models. According to high-resolution LST data, we analyzed the SUHI of Hangzhou during day and night throughout the year. Finally, combined with the retrieved LST computed from Landsat 8 TIRS with a resolution of 30 m, the downscaling results showed little error, that is, the RF model is a feasible method to downscale LST in highly heterogeneous areas.

● Downscaling Models

The thermal sharpen (TsHARP) model employs NDVI in a regression model to sharpen LST. It assumes that the relationship between LST and NDVI is the same at all scales [22]. Correlations between LST and NDVI are established [22], caused by shadows and evapotranspiration which make vegetation surface cooler than bare soil [23]. The building process is shown in Figure 2.1.
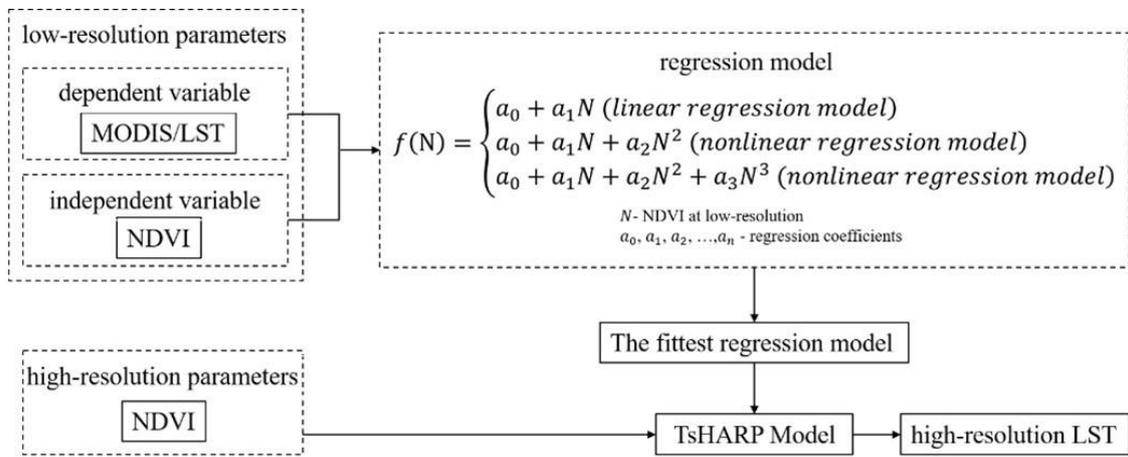
Figure 2.1. Building process of the thermal sharpen (TsHARP) model.

The key to the TsHARP model is determination of the most appropriate relationship between LST and NDVI through regression analysis. In this section, three regression models—linear regression, nonlinear binary curve regression and nonlinear ternary curve regression model—were used to fit the scatter distribution of LST and NDVI at a scale of 1 km. The fitting function is as shown in the following equation. From these three regression models, by comparing the R^2 and RMSE, the best-fitting regression model can be used to predict the LST distribution at a scale of 100 m.

$$f(N) = \begin{cases} a_0 + a_1 N \text{ (linear regression model)} \\ a_0 + a_1 N + a_2 N^2 \text{ (nonlinear regression model)} \\ a_0 + a_1 N + a_2 N^2 + a_3 N^3 \text{ (nonlinear regression model)} \end{cases}$$

(2-3)

where a_0, a_1, a_2 and a_3 represent regression coefficients, and N represents NDVI.

The multiple linear regression (MLR) model, shown in Figure 2.2, is based on multiple linear regression [24]. In downscaling low-resolution remote sensing products, additional high-resolution remote sensing information needs to be introduced to achieve downscaling conversion.
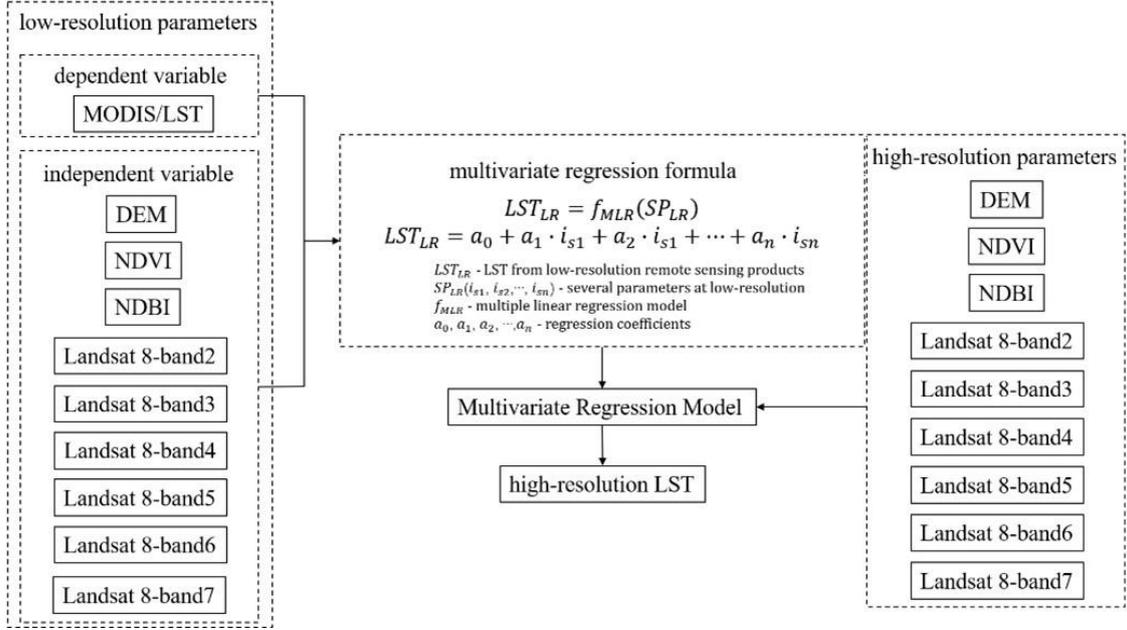
Figure 2.2. Building process of the multiple linear regression (MLR) model.

In the following building process, low-resolution parameters comprised two parts, namely, dependent and independent variables. LST was set as the dependent variable, and independent variables consisted of DEM, NDVI, NDBI, and band 2 to band 7 of Landsat 8. According to the low-resolution variables, we used the least squares method to build a MLR model, as shown in Equations (2-4) and (2-5) [25]. The LST at high-resolution was estimated based on the p corresponding independent variable and multiple regression models.

$$LST_{LR} = f_{MLR}(SP_{LR})$$

(2-4)

$$LST_{LR} = a_0 + a_1 \cdot i_{s1} + a_2 \cdot i_{s1} + \cdots + a_n \cdot i_{sn} LST_{LR} = f_{MLR}(SP_{LR})$$

(2-5)

where LST_{LR} is the LST from low-resolution remote sensing products, $SP_{LR}(i_{s1}, i_{s2}, \cdots, i_{sn})$ are several parameters which are DEM, NDVI, NDBI and Landsat 8 OLI band 2 to band 7 at low-

resolution, f_{MLR} is the multiple linear regression model, a_0, a_1, a_2, $\cdots$, a_n are regression coefficients.

The random forest (RF) model is a machine learning model which prevents overfitting, and was proposed by Breimans [26] in 2001. The term "random forest" was derived from the random decision forest proposed by Tin Kam Ho [27] in 1995. RF is a non-linear statistical ensemble method [28]. It uses bootstrap resampling technology to merge multiple samples extracted from the original training samples to generate a new series of training samples, then creates decision trees based on these training samples and establishes an RF model [15]. The RF model is not sensitive to multicollinearity, which can effectively prevent overfitting during the downscaling process [18]. The current research used Python 3.8 and the scikit-learn third-party open-source machine learning algorithm library, which is one of the most popular machine learning libraries [29].

Figure 2.3 shows the building process of the RF model. The training samples were remote sensing images with low resolution (1 km), and selection of dependent and independent variable was the same as that in the MLR model. In order to verify the accuracy of the models, we divided the sample into training samples and test samples according to a 6:4 ratio. The RF model was created by *n* decision trees generated by training samples. In the process of creating the model, several parameters needed to be adjusted, namely: n_estimators, bootstrap, and oob_score of the RF framework parameters and max_features, max_depth, min_samples_leaf, min_samples_split, max_leaf_nodes, min_impurity_decrease, criterion and min_samples_leaf of the RF decision tree parameters [30]. Among these, n_estimators, max_depth, and max_features were the three that most affect the downscaling result. In order to prevent the model from underfitting, we tuned these three parameters for fitting to achieve the optimal model. Then, we used the previously divided training samples for cross-validation based on the cross_val_score module in the scikit-learn libraries to determine the feasibility of the model.
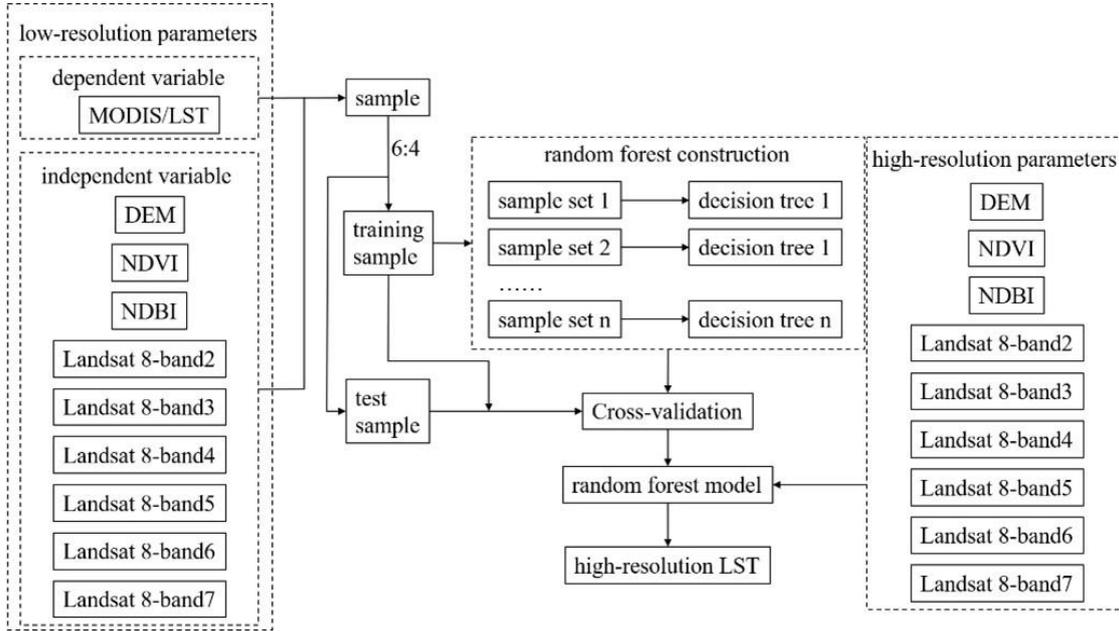
Figure 2.3. Building process of the random forest (RF) model.

● Accuracy Evaluation and Fit Residual

Traditional quantitative evaluation usually uses one evaluation indicator. In order to compare the accuracy of the three downscaling models for each day and night during four seasons more objectively, this section used two evaluation indicators for comprehensive evaluation and analysis, R^2 (coefficient of determination) and RMSE (root mean square error).

R^2 (coefficient of determination) is an important statistic to reflect the model fit. In statistics, it is used to measure the proportion of dependent variables that can be explained by independent variables to determine the explanatory power of the regression model [31]. R^2 takes values between 0 and 1 with no units. It is the most commonly used index to evaluate the pros or cons of regression models. The larger the value of R^2 (closer to 1), the better the regression model is fitted.

RMSE (root mean square error) is a commonly used measure of the similarity between two vectors in n-dimensional space [32]. RMSE can test the consistency of real images and simulation images, and thus be used to judge the effect of different downscaling models. The RMSE calculation is shown in Equation (2-6). Larger errors have a disproportionately greater effect on RMSE. Consequently, RMSE is sensitive to outliers [32]. RMSE is non-negative. A lower RMSE means higher consistency between simulation images and real images.

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (LST_{LR} - LST_{LRS})^2}$$

(2-6)

where the $RMSE$ represents root mean square error, LST_{LR} represents the low-resolution real images to reflect LST, LST_{LRS} represents the low-resolution simulation images to reflect LST, and n is the total number of pixels in the low-resolution real images or simulation images.

In the process of establishing a correlation model at a low resolution, a residual exists between the real and simulation images. In order to improve the accuracy of the simulation of high-resolution LST images, this section fitted the residual to the simulation of high-resolution images. The flow chart is shown in Figure 2.4. A spline was used to interpolate adjacent cells to downscale the LST residual. The last step was to fit the high-resolution residual to the high-resolution simulation images, finally resulting in high-resolution land surface temperature images.
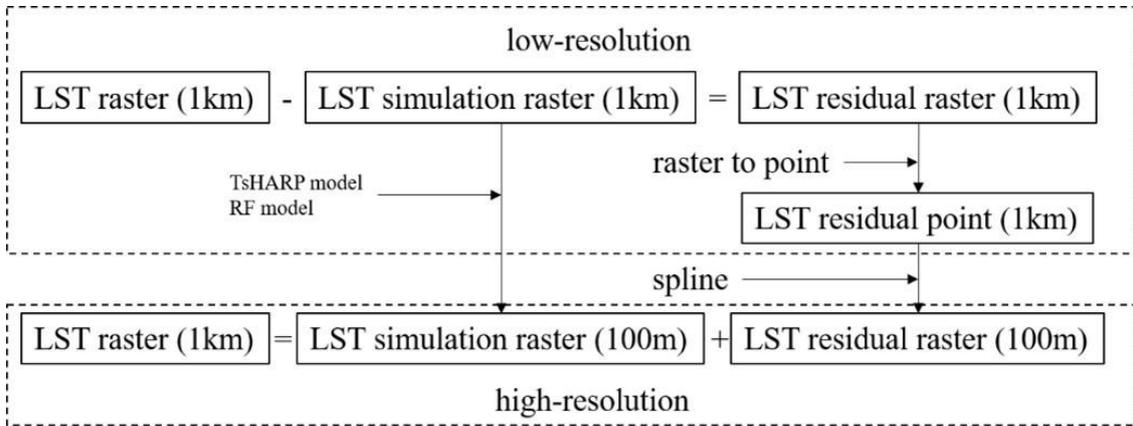
Figure 2.4. Fit residual flow chart.

●   Downscaling Result Validation based on Retrieved LST

To provide further confirmation of the approach, this section verified the downscaling accuracy using retrieved LST values from Landsat 8 TIRS with similar time and weather conditions as those of MODIS/LST. The single-channel algorithm proposed by Giannini et al. [33] and Dissanayake et al. [34] for LST retrieval of Landsat 8 TIRS has high accuracy and sensitivity.

Firstly, the proportion of vegetation was calculated using Equation (2-7) [35]:

$$P_v = \left(\frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \right)^2$$

(2-7)

where P_v represents the proportion of vegetation; $NDVI$ represents the normalized difference vegetation index (explained in Equation (2-10)); $NDVI_{min}$ and $NDVI_{max}$ represent the minimum and maximum value of NDVI, respectively.

Secondly, land surface emissivity was computed using Equation (2-8) [36]:

$$E = 0.004P_v + 0.986$$

(2-8)

where E represents land surface emissivity; P_v represents the proportion of vegetation.

Finally, LST corrected for spectral emissivity was computed using Equation (2-9):

$$LST = \frac{T_b}{1 + \left(\frac{\lambda \cdot T_b}{\rho}\right) \cdot \ln(E)}$$

(2-9)

where LST represents land surface temperature; T_b represents the at-satellite brightness temperature [37]; λ represents the band 10 wavelength in Landsat 8 TIRS (10.8 μm); ρ is 1.438×10^{-2} mK; E represents land surface emissivity.

Due to the difference between Terra satellite and Landsat 8 orbits, the revisit period of MOD11A2 is 8 days and that of Landsat 8 is 16 days, meaning that the images cannot be obtained in the same day. In addition, there is also an error of several minutes in the scanning time. In order to solve the contradiction in temporal resolution, we introduced meteorological conditions, including maximum and minimum air temperature, relative humidity, wind speed [38], and solar radiation [39] on the basis of selecting two adjacent dates as much as possible. Finally, we selected a set of downscaling LSTs and retrieved LSTs with the closest time and the most similar meteorological conditions to verify the downscaling method accuracy.

2.2.2. Case Study

● Study Area

The study area is located in the center of Hangzhou, as shown in Figure 2.5. This section selected a square urban area with a side length of 25 km. Hangzhou is located in the central and southern areas of the Yangtze River Delta. Hangzhou's climate is humid subtropical with four distinct seasons.

The study area includes various land use and land cover. Qiantang River runs through this area. West Lake is located on the west side. To the southwest of West Lake is a forest area. The urban area is mainly concentrated in areas to the north, east, and northeast of West Lake. As an important part of the urban agglomeration in the Yangtze River Delta, Hangzhou developed with rapid urbanization from the end of the previous century. Due to urban expansion and population growth, the urban structure has changed significantly. This has also led to climate change in this area, particularly in terms of SUHI. Previous research shows that SUHI is a significant contributor to regional warming in this area [40].
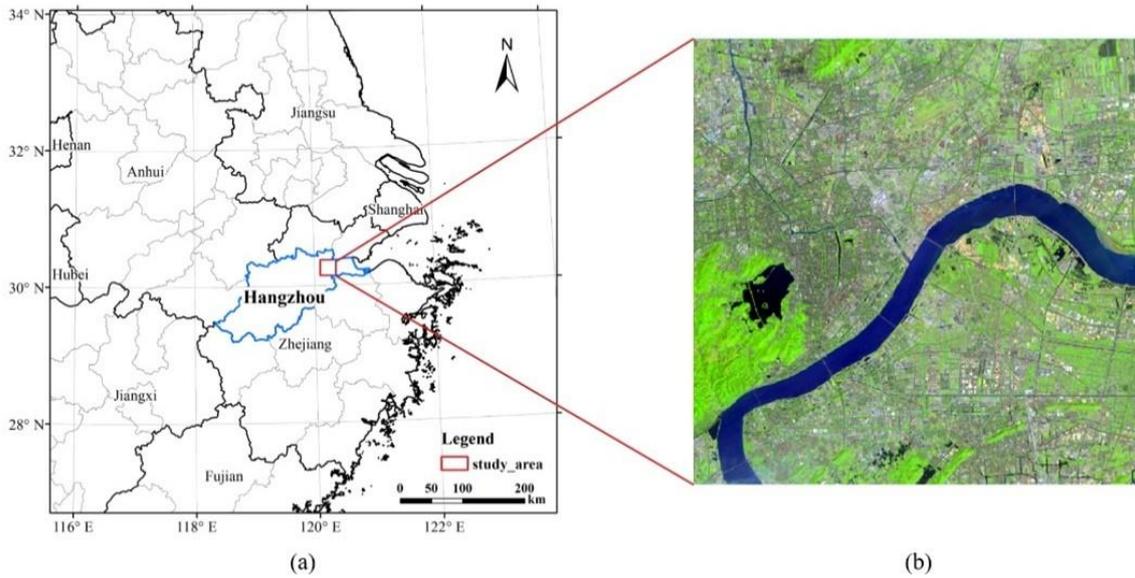
Figure 2.5. Study area (a) Location of the study area in the Yangtze River Delta; (b) Landsat 8 composite of study area.

- Data Source and Preprocessing

In this section, MODIS/LST products and Landsat 8 OLI were obtained from summer 2013 to spring 2014. Landsat 8 OLI provides seasonal coverage of global land images with nine bands. These remote sensing images use the World Reference System (WRS) to enable users to search for images from any part of the world using path and row numbers [41]. In Landsat 8 OLI, the coastal aerosol band (band 1) focuses on aerosols research in coastal areas; the panchromatic band (band 8) produces black and white images with a resolution of 15 m used to enhance and improve resolution; and the cirrus band (band 9) is designed for clouds, particularly for cirrus clouds [42]. These three bands were not useful for the downscaling of this section. In contrast to these bands, the visible blue band (band 2), green band (band 3) and red band (band 4) can help identify various land uses and land covers; the near infrared band (band 5) provides vegetation indexes, such as NDVI, which allow measurement of plant health in combination with other bands; and the shortwave infrared bands (bands 6 and 7) are particularly useful for distinguishing wet from dry earth, and for geology. Thus, we only selected bands 2 to 7 from Landsat 8 OLI as the data source. Landsat 8 OLI is greatly influenced by clouds and weather. Accordingly, several sunny days without any clouds above the study area were chosen April 14, July 19, November 8 in 2013 and January 27 in 2014. We selected Landsat 8 Level-1 Data Products after system radiation correction and geometric correction [37]. The WRS path and row were 119 and 039, respectively. The Landsat data we chose are shown in Table 2-1. The additional parameters at high-resolution extracted from Landsat 8 OLI were pre-processed according to Equations (2-10) and (2-11) [43,44]. Meanwhile, DEM data with a resolution of 30 m, which reflect the altitude situation, were also used in the RF model and the MLR model as independent variables. In order to downscale

from 1 km to 100 m, this section resampled these parameters at a scale of 100 m and 1 km.

$$\text{NDVI} = \frac{NIR - RED}{NIR + RED}$$

(2-10)

$$\text{NDBI} = \frac{SWIR_1 - NIR}{SWIR_1 + NIR}$$

(2-11)

where RED, NIR and $SWIR_1$ represent band 4, band 5, and band 6 in Landsat 8 OLI, respectively [37].

Table 2-1. The dates and remote sensing images IDs of data source.

Landsat 8 OLI/TIRS		MOD11A2/LST	
Date	Landsat scene ID	Date	MODIS/LST ID
2013.04.14	LC81190392013104LGN02	2013.04.07	A2013097.h28v05.006.2016156021756
			A2013097.h28v06.006.2016156021753
2013.07.19	LC81190392013200LGN01	2013.07.20	A2013201.h28v05.006.2016166200144
			A2013201.h28v06.006.2016166200148
2013.11.08	LC81190392013312LGN02	2013.11.09	A2013313.h28v05.006.2016173161718
			A2013313.h28v06.006.2016173161720
2014.01.27	LC81190392014027LGN01	2014.01.17	A2014017.h28v05.006.2016197155044
			A2014017.h28v06.006.2016197155043

MOD11A2/LST products with a resolution of 1 km were selected as the low-resolution LST data with a temporal resolution of 8 days, including day data (10:30 a.m.) and night data (10:30 p.m.). MOD11A2 products were retrieved based on the split channel algorithm [45]. The path and row were 28/05 and 28/06, respectively. The imaging dates, shown in Table 2-1, were April 7, July 20, November 9 in 2013 and January 17 in 2014, similar to the Landsat 8 OLI dates. Then, the MODIS Tools called MRT were used for reprocessing tasks, such as creating a mosaic and resampling.

2.2.3.   Model Selection and Precision Analysis

Based on SPSS and Python, we constructed correlation models for statistical low-resolution data. Downscaling based on the TsHARP and MLR models was run on SPSS, and the RF model was run using Python.

After statistical calculation, the TsHARP model with the best fitting was the third of three equations, the unary cubic model. Table 2-2 shows the regression coefficients of the TsHARP model during day and night in four seasons. According to theses regression coefficients, we formed the corresponding downscaling models and thereby predicted low-resolution LST. Among the regression coefficients, a_0 has high significance for the models regardless of the seasons or whether day or night; a_2 and a_3 have low significance, especially in autumn and winter daytime. Figure 2.6 shows the scatter plot of predicted LST data based on the TsHARP models versus MODIS LST data. The x-axis represents the MODIS/LST product values, which are the true LST (1 km level); the y-axis represents the predicted LST from the TsHARP model (1 km level).

Table 2-2. The regression coefficients of the TsHARP models during day and night in four seasons.

season	day or night	Regression coefficients			
		a_0	a_1	a_2	a_3
spring	day	22.672***	48.924***	-290.510***	373.146***
	night	12.887***	5.591**	-110.425***	239.119***
summer	day	37.672***	48.924***	-290.510***	373.146***
	night	28.351***	15.334***	-151.171***	273.468***
autumn	day	19.222***	12.474*	35.560	-591.507**
	night	14.234***	26.034***	77.402**	110.314
winter	day	9.952***	13.535***	-288.013**	764.451*
	night	3.323***	-26.812***	-280.854***	3898.314***

Note: * - $p < 0.1$; ** - $p < 0.05$; *** - $p < 0.01$
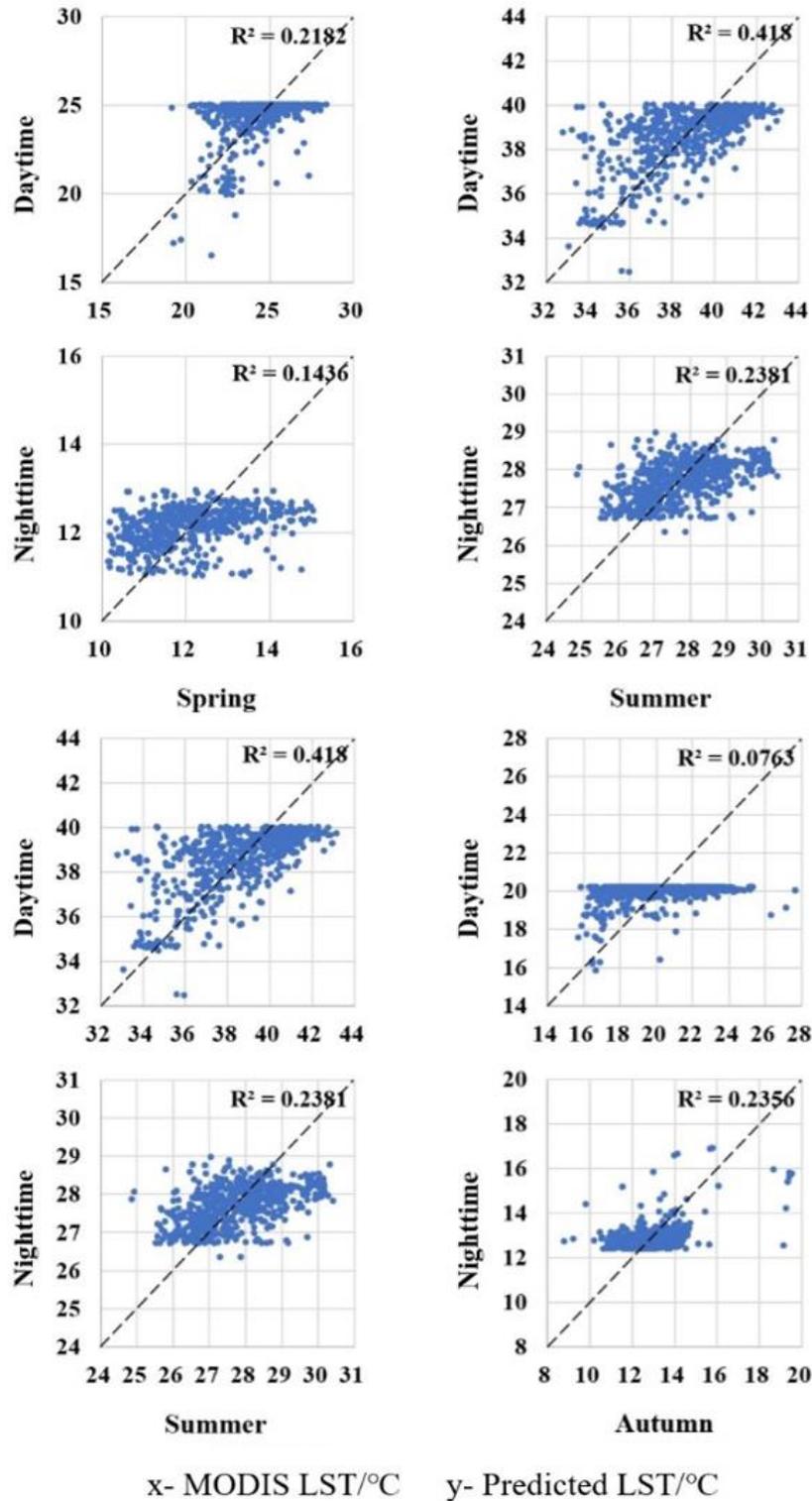
Figure 2.6. Scatter diagrams of predicted land surface temperature (LST) data based on the TsHARP models (y-axis) versus MODIS LST data (x-axis). (The black dashed lines represent the 1:1 line, that is, the predicated LST and MODIS LST of the scattered points falling on this line are equal).

Scatter points were not distributed near the 1:1 line, which meant that this model was poor and could not be used in downscaling research in this section area. The TsHARP model is based on the correlation between LST and NDVI. Due to shadows and transpiration, the vegetation surface is usually cooler than that of other landscapes[46]. This theoretical basis had considerable errors because of the strong spatial heterogeneity of the study area [47–49], and the predicted LST had an obvious boundary value. Therefore, the TsHARP model was not suitable for this section.

In order to solve the problem of the sharp drop in the correlation between LST and NDVI due to spatial heterogeneity, we introduced more independent variables to build the MLR model based on NDVI, including DEM, NDBI, and Landsat B2 to B7. According to the independent and dependent variables, we calculated the regression coefficients (a_0, a_1, a_2…a_n) using the least squares method. Table 2-3 shows the regression coefficients of the MLR models. Overall, CT, DEM, NDVI, B2, B3, B4, and B7 have high significance, compared with other variables. The significance of NDBI during the day is generally higher than that at night. Scatter diagrams comparisons of MODIS LST and predicted LST (Figure 2.7) show that the predictive capabilities of the MLR model are improved compared with the TsHARP model. However, since the MLR model is a linear model, which cannot easily characterize the complex nonlinear regression between LST and independent variables, there are a large number of outliers. Thus, the MLR model is not a perfect downscaling model in this section.

Table 2-3. The regression coefficients of the MLR models during day and night in four seasons.

season	day or night	Regression coefficients									
		CT	DEM	NDVI	NDBI	B2	B3	B4	B5	B6	B7
spring	day	18.1***	0.008*	26.8*	-29.1*	0.004***	-0.008***	0.005***	-0.001	0.001*	-0.002***
	night	16.2***	0.012***	26.9**	-19.6	0.003***	-0.006***	0.004***	-0.001	0.001	-0.002***
summer	day	16.9***	-0.009*	45.7***	-90.9***	0.006***	-0.005***	0.004***	0.001*	-0.002**	-0.002***
	night	27.4***	0.004*	18.5***	-23.7**	0.003***	-0.004***	0.003***	-0.001***	0.001****	-0.002***
autumn	day	-4.47	-0.017**	107***	71.3**	-0.001	0.002	0.005**	-0.007***	0.001	0.003***
	night	34.7***	0.010***	-89.0***	24.0*	0.001	-0.006***	-0.001	0.002**	0.002**	-0.001*
winter	day	42.1***	0.009***	-106***	-4.90***	0.003***	-0.016***	0.005***	0.005**	-0.001	0.001*
	night	2.53	0.019***	-26.2	-49.6***	0.006***	-0.008***	0.003***	0.002	0.000	-0.003***

Note: CT - Constant Term; * - $p<0.1$; ** - $p<0.05$; *** - $p<0.01$
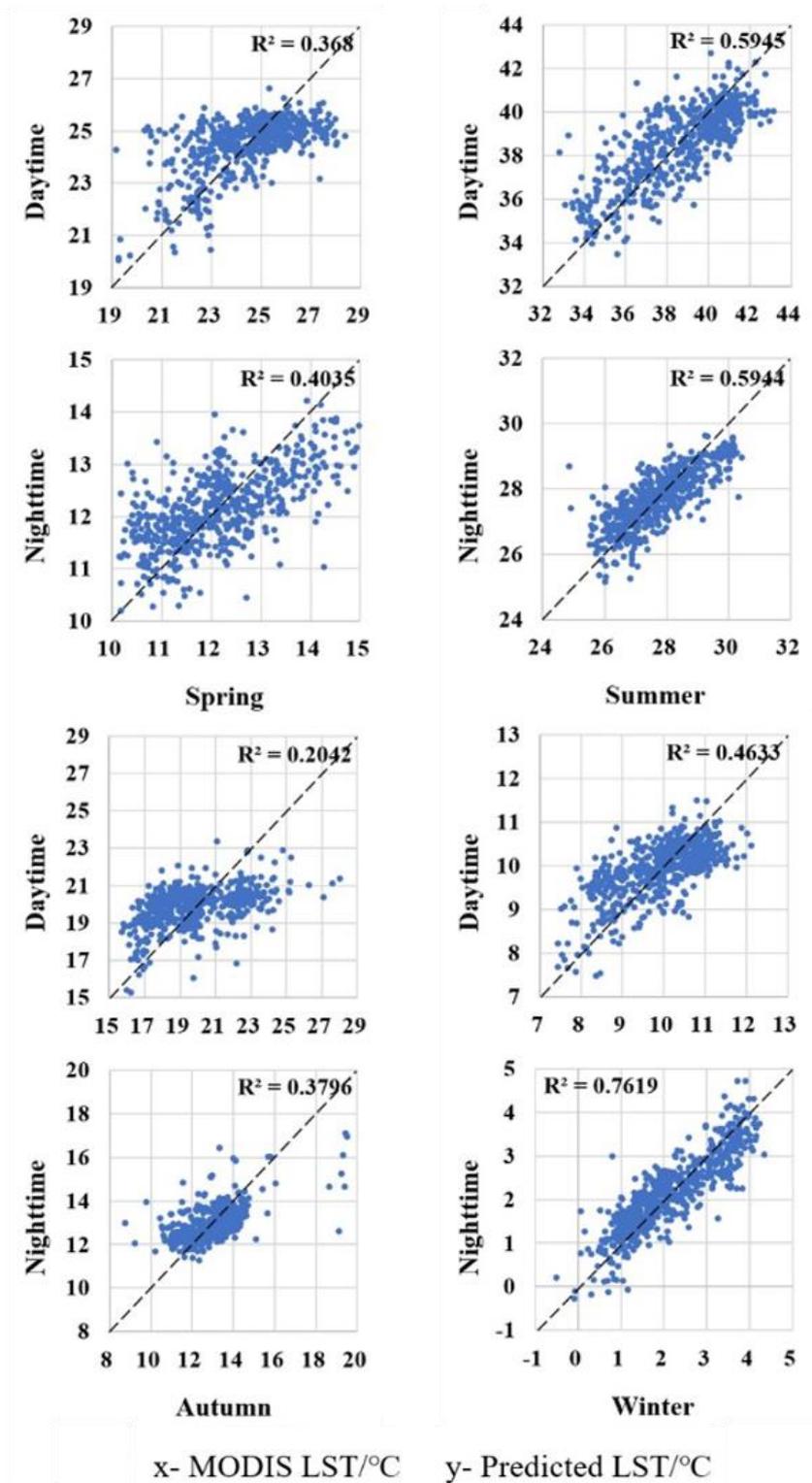
Figure 2.7. Scatter diagrams of predicted LST data based on the MLR models (y-axis) versus MODIS LST data (x-axis). (The black dashed lines represent the 1:1 line, that is, the predicated LST and MODIS LST of the scattered points falling on this line are equal).

The simple single-factor and multi-factor regression models cannot completely summarize the complex relationship between different factors and LST. Under the premise that the physical mechanism is still unclear, a better choice is to build a downscaling model with the help of machine learning methods. Compared with some other machine learning methods, such as artificial neural networks and support vector machines, the RF model has the advantages of low computation needs and a large number of samples, which are appropriate for downscaling research. The training process of the RF model mainly comprises the process of adjusting hyperparameters, which is generally called parameter tuning. Various parameters combinations will have different predicted results. Therefore, there is no single set of parameters that can optimize the various models. Optimization requires continuous training and adjustment to achieve the optimal combination for a certain type of problem [50]. We tuned parameters according to the importance of the three most significant parameters, which are n_estimators, max_depth, and max_features. Due to the small number of samples in this section, the division depth was not constrained, that is, max_depth was set to "None". Thus, this section only tuned n_estimators and max_features. Other parameters were set to default values. Figure 2.8 shows the changes of the model's obb score, that is R^2 when tuning n_estimators and max_features in three parts: Figure 2.8(a) represents the changes with n_estimators ranging from 1 to 200; Figure 2.8(b) represents the partial enlarged detail with n_estimators ranging from 1 to 40; and Figure 2.8(c) represents the range from 30 to 200.
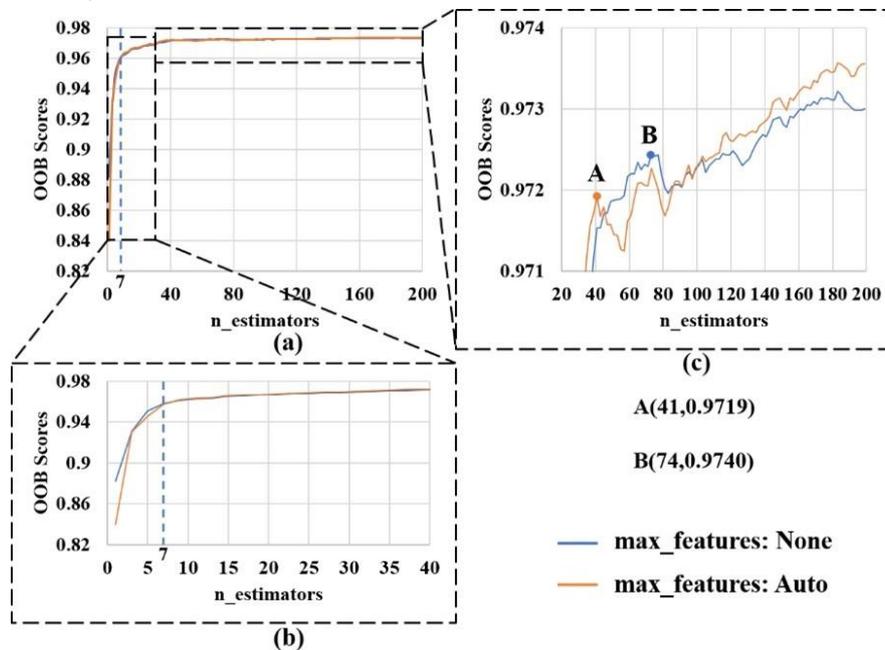
Figure 2.8. The changes of model Out Of Bag (OOB) scores (y-axis) when tuning n_estimators (x-axis) and max_features (blue line and orange line):(a) the changes with n_estimators ranging from 1 to 200; (b) the partial enlarged detail with n_estimators ranging from 1 to 40; (c) the partial enlarged detail with n_estimators ranging from 30 to 200.

In order to prevent underfitting of the RF model, in the tuning process we increased n_estimators to improve the model's fitting ability; when the Out Of Bag (OOB) scores did not significantly improve for the first time, the value of n_estimators was optimal (Point A and B in Figure 2.8). Meanwhile, we adjusted max_features, and set it to None (blue line in Figure 2.8, max_features are the square root of the sample features) and Auto (orange line in Figure 2.8, max_features are the sample features).

Combining the predicted results of the above two models, the fitting degree of winter night was the highest. Thus, we took winter night as an example of parameter tuning. When all parameters were set to default values, the OOB score was 0.9621. When n_estimators increased from 1 to 7, OOB scores rose rapidly, then tended to be flat. According to the Figure 2.8(c), the orange line reached the first maximum OOB scores (OOB score = 0.9719), Point A, when n_estimators was 41. When n_estimators was 71, the first maximum OOB score of the blue line was 0.9740, Point B. Consequently, Point B was the best parameter combination as shown in Table 2-4. After parameter tuning, we obtained the optimal combination corresponding to an OOB score of 0.9740, which was 0.0119 higher than the original OOB score. Then, we combined the training samples and test samples to perform a total of 10 cross-validation on the optimal model based on the cross_val_score module to verify whether the model was good fitting. The 10-fold cross-validation results are shown in Figure 2.9. The mean of accuracy was about 0.9625. The fourth cross-validation had the highest accuracy, of about 0.9834, and the lowest was the ninth, of about 0.9396. The mean squared error (MSE) of the training sample was about 0.025 °C and that of test sample was about 0.053 °C. The MSE of test samples was slightly higher than that of training samples, indicating that the model was not overfitting. Overall, the cross-validation results meet the requirements, that is, the optimal RF model could be used in the subsequent downscaling research.

Table 2-4. RF model parameter list and main optimal combination of this model.

Parameter Name	Parameter Description	Ranges	Optimal Value
n_estimators	The number of trees in the forest.	1,3,5,7…199	71
max_depth	The maximum depth of the tree.	None,1,2…100	None
max_features	The number of features to consider when looking for the best split	None, Auto	None
oob_score	Whether to use out-of-bag samples to estimate the generalization accuracy.	True, False	True
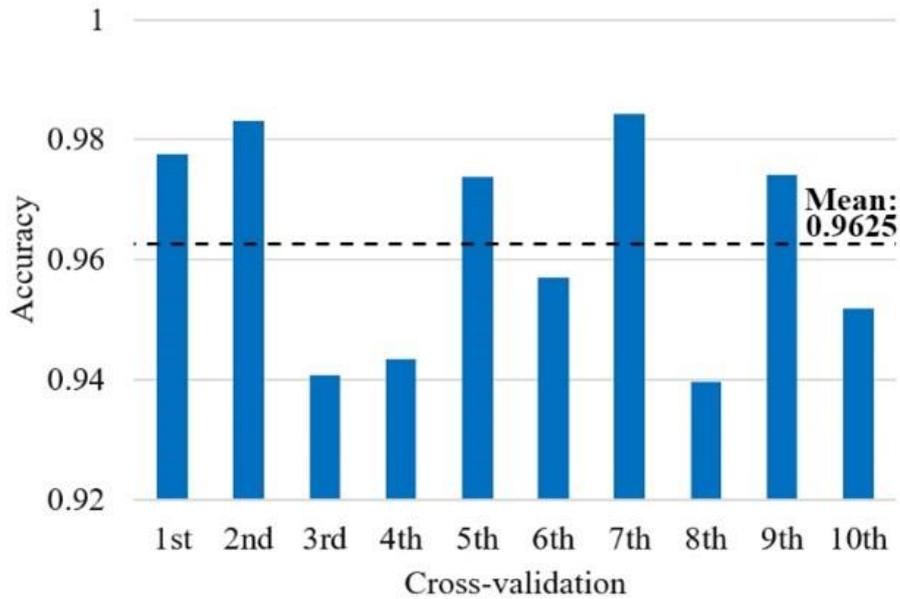
Figure 2.9. Cross-validation results of the optimal model (x-axis represents the cross-validation times; y-axis represents the model accuracy; the black dashed line represents the mean of cross-validation results).

Through the above method, we tuned the parameters of the RF models for different dates so that their accuracy complied with the requirement based on cross validation, we then predicted low-resolution LST based on these RF models and compared results with MODIS/LST products to verify model accuracy. The scatter diagrams are shown in Figure 2.10. The scatterplot comparisons of MODIS LST and predicted LST show the improved predictive capabilities of the RF model in comparison to the TsHARP and MLR models, with almost all scatter points clustered around the 1:1 line and fewer outliers. Compared with the earlier two models, the RF model is more suitable for downscaling in this highly heterogeneous research region. Furthermore, the error histograms (Figure 2.11) show that the prediction errors of the RF model approximately obeyed the normal distribution; the peak value appeared around 0 ℃, and values gradually decreased on both sides. Peaks at night were generally higher than those during the daytime. The prediction errors in summer and winter were significantly less than those in spring and autumn, especially during daytime. The daytime errors in autumn were more discrete than those during other seasons. Compared to the minimum value, the value of winter night was closest to 0 ℃, respectively, -0.7 ℃ and 0.5 ℃. However, 1% of values were less than -2 ℃ and 0.2% of values were more than 2 ℃.
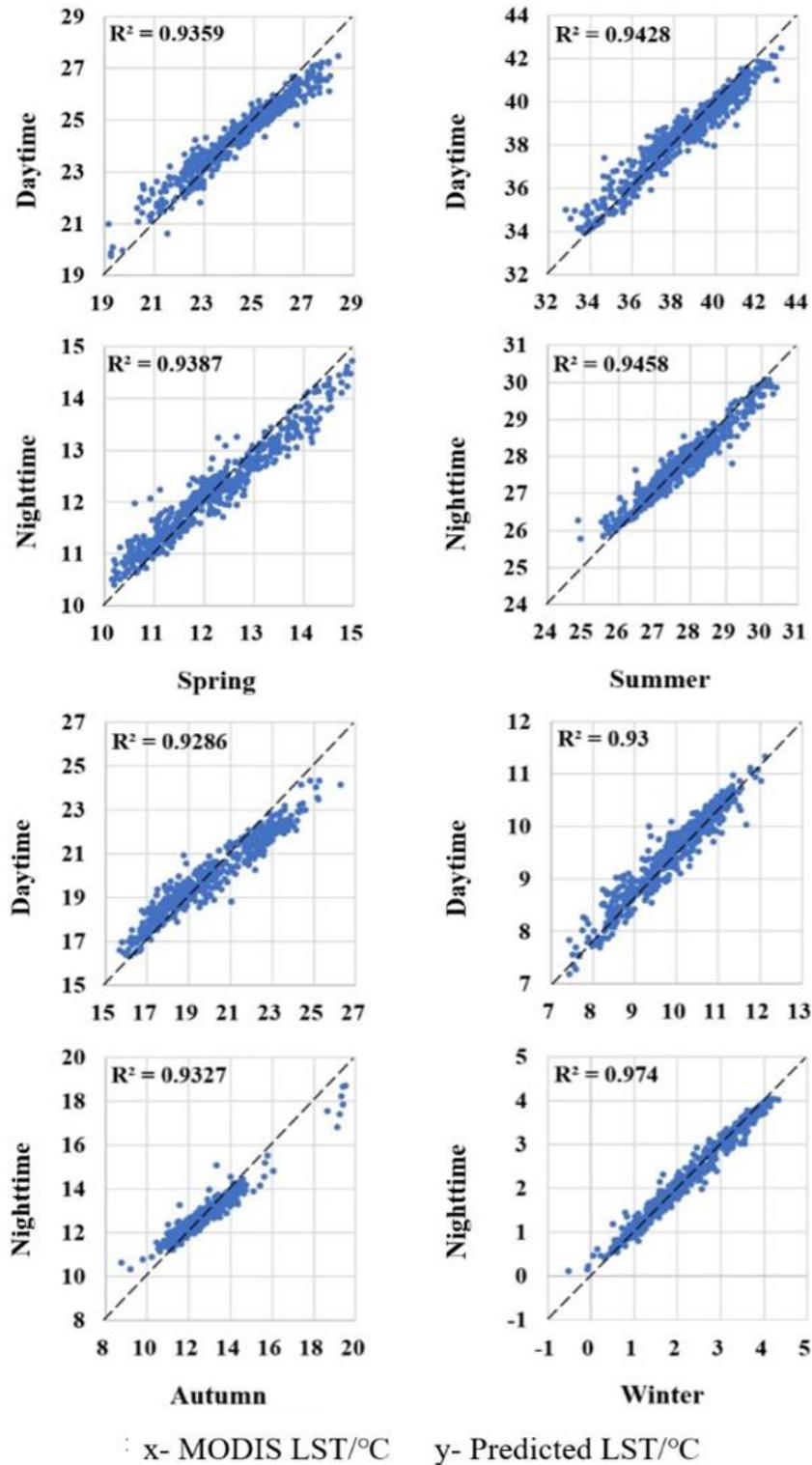
Figure 2.10. Scatter diagrams of predicted LST data based on the RF models (y-axis) versus MODIS LST data (x-axis). (The black dashed lines represent the 1:1 line, that is, the predicated LST and MODIS LST of the scattered points falling on this line are equal).
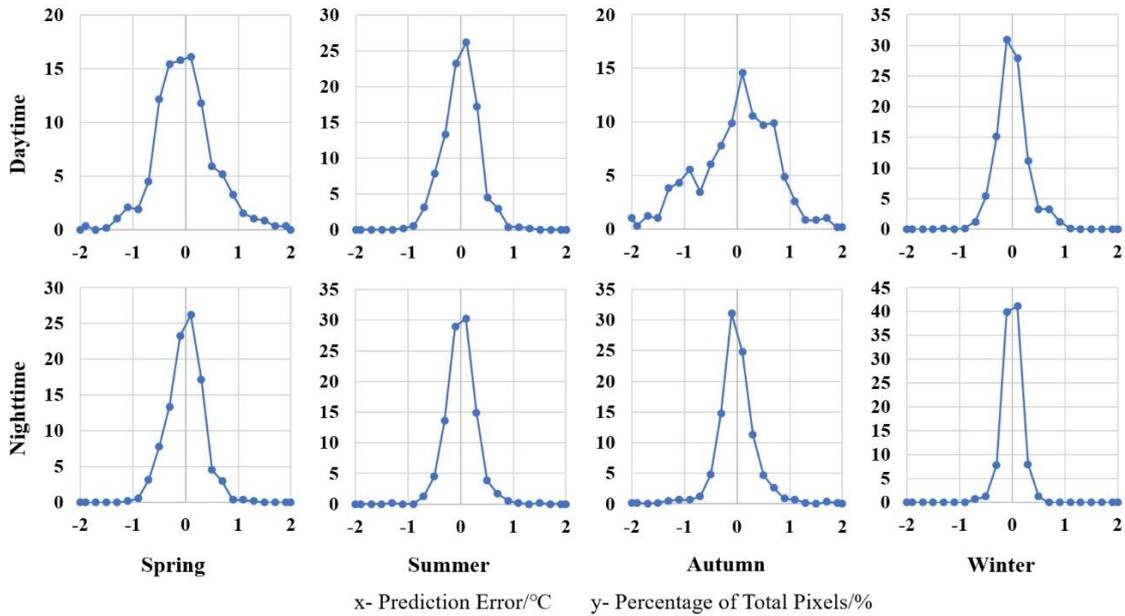
Figure 2.11. Error histogram of RF models. (x-axis represents the prediction error/°C; y-axis represents percentage of total pixels of study area/%).

To sum marize, R^2 and RMSE were calculated for MODIS/LST products and predicted LST, as shown in Figure 2.12. The histogram shows R^2, and line chart shows RMSE, of various models. Blue, orange, and gray represent the TsHARP, MLR, and RF models, respectively. The results clearly show that the RF model was more suitable for this section than the TsHARP and MLR models. This is evident in the significantly higher R^2 and lower RMSE of the RF model compared to the other two models. For the RF model, the prediction effect at night was better than that during the daytime, and that at winter night was the best, with R^2 of 0.9740 and RMSE of 0.1678. The worst effect was for autumn daytime: R^2 was 0.9286 and RMSE was 0.7556. However, even the worst RF model performed better than the other two models. From a seasonal perspective, R^2 values in summer and winter were higher than those in spring and autumn. From low to high, RMSE values were winter, spring, summer, and autumn. By comparison with the single-factor TsHARP model, the prediction improvement of the MLR model with more independent variables was limited due to the simple linear regression. The application of the RF model greatly enhanced the model's predictive capabilities, because, under the premise of multiple-factors, machine learning could perform complex nonlinear regression. According to the above model selection and precision analysis, we only selected the RF model to undertaken downscaling of MODIS/LST products from 1 km to 100 m.
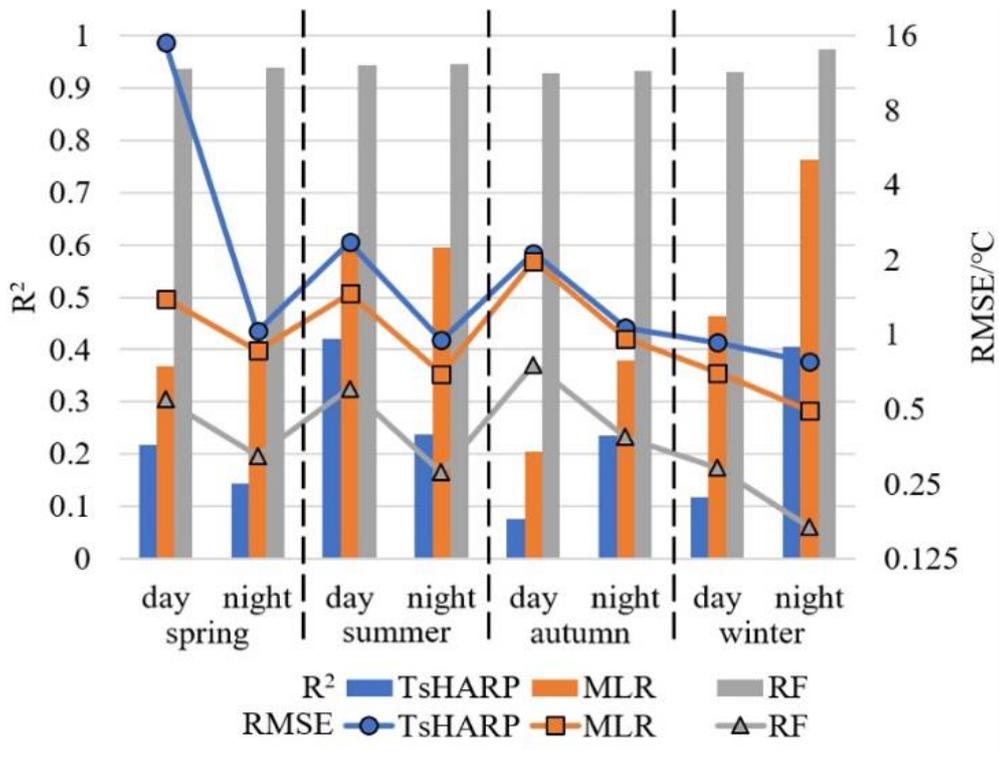
Figure 2.12. Prediction precision for TsHARP, MLR, and RF models of low-resolution LST for the entire research region. (Histogram represents R^2; line chart represents RMSE).

2.2.4.   Downscaling Results and Surface Urban Heat Island

During the construction of nonlinear regression models, the RF model was able to provide feature importance based on randomized variable selection. The importance scores were presented in mean squared error (MSE). The larger the MSE of an independent variable, the more important that variable is to a model [51]. Figure 2.13 shows the independent variable importance scores from all research data; the x-axis represents MSE and the y-axis represents the independent variables. Among the independent variables, b2 to b7 represent Band 2 to Band 7 from Landsat 8 OLI. During daytime, the importance scores of various factors were relatively balanced, and the difference between high and low score factors was large at night, especially at winter night. The b6 score reached 43.2%; in comparison, the highest score in daytime was 30%, for NDBI in winter. At night, b5 and b6 were the most important, with high scores. Meanwhile, DEM played a major role in the RF models at night. Contrary to nighttime, DEM scores in daytime were not large. According to the Environmental Lapse Rate [52], high-altitude areas usually received more solar radiation with more xeric and warmer conditions, particularly north-facing slopes. In the current research, different orientations resulted in a significant difference of LST. Therefore, the effect of DEM is weak when receiving solar radiation [53]. NDBI during daytime played a key role in the RF models, with the highest scores in summer, autumn and winter, and the second highest in spring. This meant that buildings heated by solar

radiation had a significant influence on LST. By contrast, at night without solar radiation. NDBI had lower importance scores than DEM.
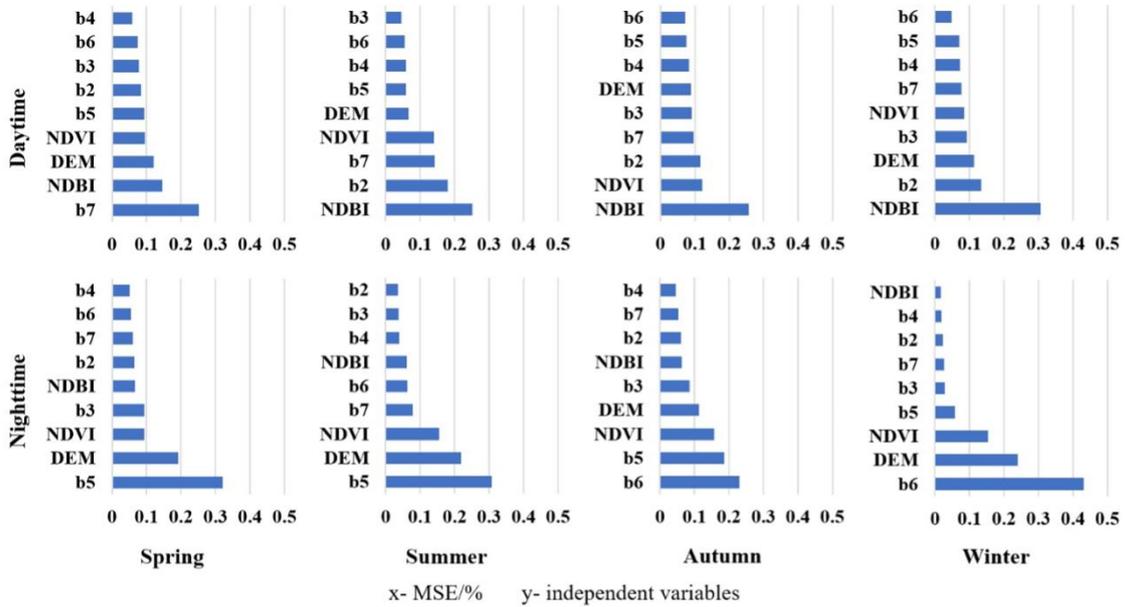
Figure 2.13. The independent variable importance scores across all research date for RF model. (x-axis: MSE/%, represents importance scores; y-axis: independent variables).

The intention of a LST downscaling model is to overcome the contradiction between the spatial-temporal resolutions of various remote sensing images to obtain LST products with high spatial-temporal resolution. According to the RF models constructed in Section 2.2.3, we downscaled MODIS/LST from 1 km to 100 m, that is, independent variables with a scale of 100 m were used to predict the corresponding LST. In order to improve prediction accuracy and reduce errors, we used a spline method to fit residuals, as introduced in Section 2.2.1. Due to some restrictions of the study area, MODIS/LST products could not provide high-resolution LST of Qiantang River, which crosses the urban area, mostly resulting in a lack of water surface temperature data. The data were unable to provide enough training samples for the RF model. Therefore, the downscaling accuracy based on the RF model of water surfaces will be greatly reduced. The average annual sediment discharge was 6.68 million tons [54]. The tidal bore is one of the symbolic features of the Qiantang River. The effect of tidal bores causes abrupt changes of the riverbed, thus changing the land cover, which in turn influences LST [55]. Hence, we eliminated the LST downscaling of large areas of water, such as Qiantang River and West Lake.
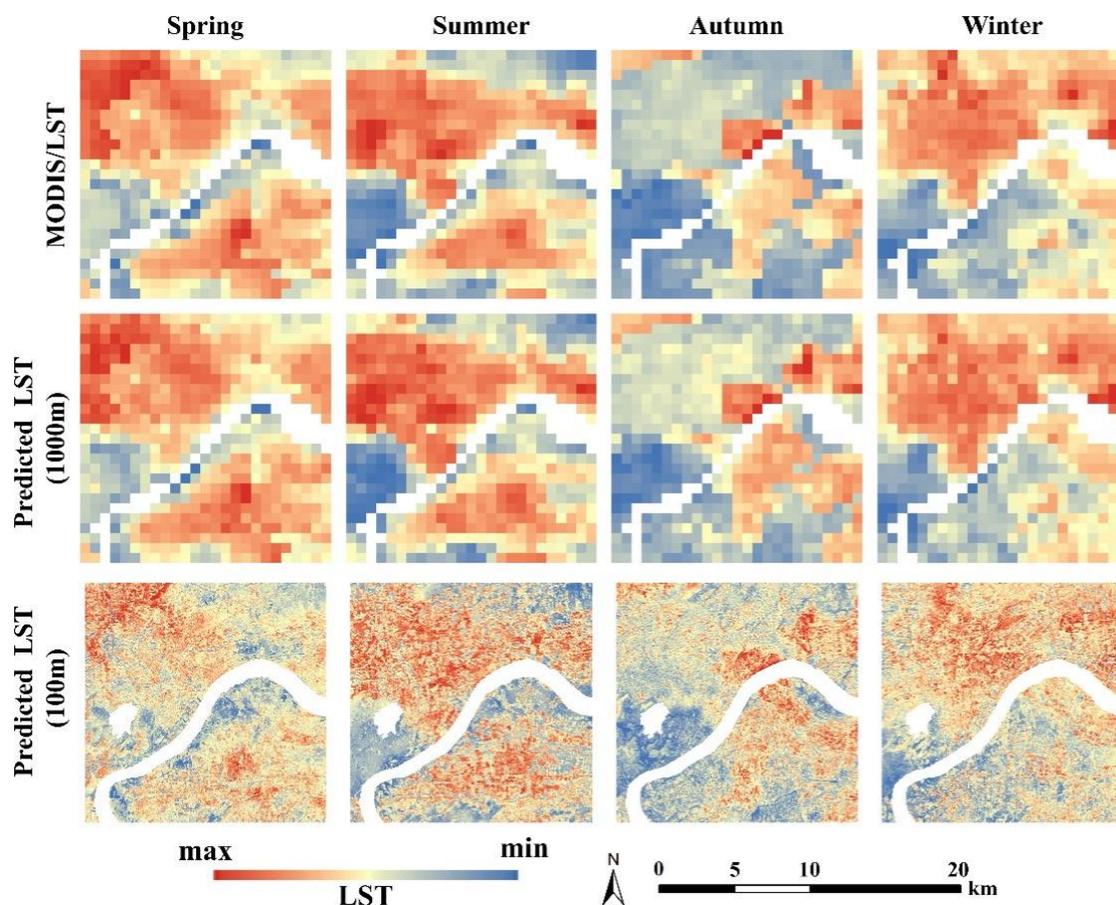
Figure 2.14. Predicted and downscaling results during daytime at different scales in various seasons from MODIS/LST products

Figures 2.14 and 2.15 show LST distribution during day and night, respectively, including the MODIS/LST products, predicted LST with a scale of 1km, and predicted LST with a resolution of 100 m; the latter two represent the downscaling results. LST changes from blue to red. Blue regions represent low temperature areas, and red regions represent high temperature areas. Comparing MODIS/LST products and predicted LST with a scale of 1 km, the similarity of each pair of images is extremely high, whether during the day or at night, which also shows that the RF model is suitable for future downscaling research in this section area. The following prediction results, with a scale of 100 m, are the downscaling results after the fitting of residuals. The distribution of downscaled LST is basically consistent with the distribution of MODIS/LST products.
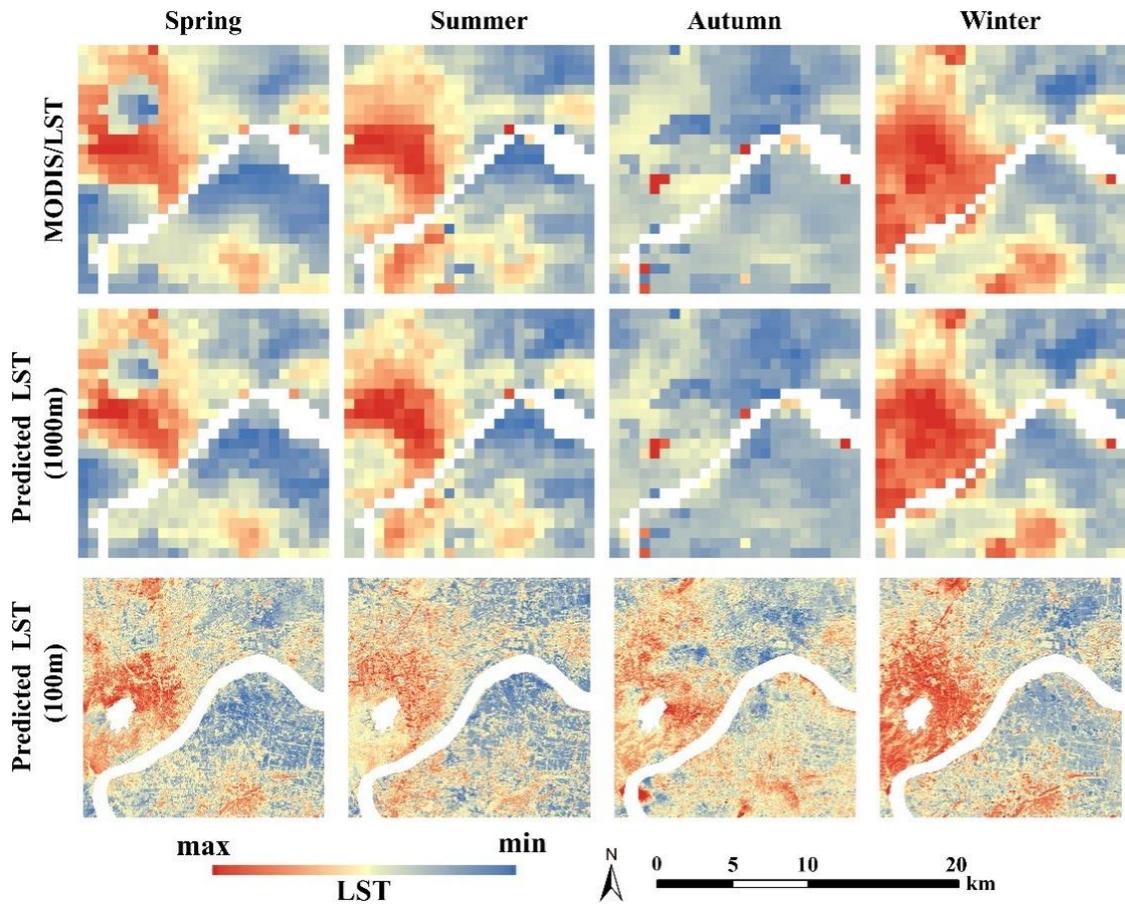
Figure 2.15. Predicted and downscaling results at night at different scale in various seasons from MODIS/LST products.

With the exception of autumn, the SUHI distribution during the day is similar across the seasons. On the north bank of the river and the northeast side of the lake, a large number of high temperature areas are generally distributed. The LST of the southwest area of the lake is lower than that of the other areas. Compared with spring and summer, there are more high temperature areas in the northeast corner of the region and the north bank of the river. However, the low temperature areas in the southern part of the river are more than those in spring and summer. In contrast to spring, summer, and winter, the high temperature areas in autumn are obviously fewer. Red areas mainly appear on the north bank of the river and near the east side of the research region. Similar to the other three seasons, the low temperature areas are also located at the southwest of the lake.

At night, high temperature areas are mainly concentrated around the lake in low-resolution MODIS/LST products. However, SUHI in autumn is obviously different from other seasons. Several small areas with extremely high temperatures can be found in MODIS/LST products. Through downscaling and excluding large water areas including West Lake and Qiantang River, the

corresponding high-resolution LST shows that the SUHI distributions at night are similar in all seasons. The blue areas are located in the northeast of the study area and the south bank of river. The dissimilarity in the four seasons lies in the relative differences of LST in high temperature areas. The red area is larger in winter than in other seasons.

From the downscaling results, we found that there are significantly more high temperature areas during the day than at night. This means that the distribution range of SUHI is wider during the day than at night. During the daytime, SUHI spreads throughout the research region but at night SUHI shrinks towards the West Lake and the south of the study area. Two obvious SUHI area are distributed on both sides of Qiantang River. The LST of SUHI on the north bank is higher than that on the south bank. In daytime, the urban center, which is to the northeast of West Lake, is not clearly the warmest area; however, at night the urban center became the core zone of SUHI irrespective of the season. Ranges of LST during daytime (Figure 2.14) and at night (Figure 12.5) are shown in Table 2-5. Comparing MODIS/LST and predicted LST with a scale of 1 km, we found that the average LSTs were almost equal, with differences less than 0.02 ℃. However, the ranges were smaller. Comparing MODIS/LST and predicted LST with a scale of 100 m, the mean LST difference showed improvement compared with the former, but within the allowable range. Furthermore, the LST ranges were close to those of MODIS/LST. The predicted LST (100 m) difference was largest in autumn during both day and night, that is, SUHI in autumn is the most serious. The difference in winter was the smallest. Generally speaking, differences during the day were always greater than those at night, and differences ranged of 2 ℃ and 5 ℃, except in winter. In winter, the difference at night was 0.39 ℃ higher than that during the day. From the mean LST throughout the year, LST rose sharply to reach 38.42 ℃ in the daytime and 27.56 ℃ at night from spring to summer, then gradually decreased to 9.91 ℃ during the day and 1.98 ℃ at night in winter.

Table 2-5. Ranges of LST during daytime (Figure 2.14) and at night (Figure 2.15) of all research dates.

Season	day or night	MODIS LST/°C			Predicted LST (1000m)/°C			Predicted LST (100m)/°C			
		max	mean	min	max	mean	min	max	mean	min	Δ
spring	day	28.37	24.49	19.15	27.48	24.48	19.75	27.97	24.34	20.49	7.48
	night	15.05	12.15	10.15	14.81	12.15	10.40	15.10	12.08	9.72	5.38
summer	day	43.15	38.51	32.81	42.50	38.52	33.92	42.82	38.42	33.18	9.64
	night	30.41	27.69	24.85	30.09	27.70	25.78	30.43	27.56	25.32	5.11
autumn	day	28.03	19.88	15.69	25.43	19.86	16.33	26.74	19.91	15.96	10.78
	night	19.51	12.89	8.77	18.71	12.87	10.35	16.20	12.75	10.03	6.17
winter	day	12.13	9.89	7.43	11.72	9.89	7.59	11.94	9.91	7.49	4.45
	night	4.33	2.13	-0.51	4.03	2.13	0.12	4.32	1.98	-0.52	4.84

Note: Δ - The predicted LST difference with a resolution of 100m between the maximum and minimum

2.2.5.   Validation Results comparing Downscaling LST and Retrieved LST

Due to satellite orbit restrictions, we could not obtain MODIS/LST products and Landsat 8 OLI/TIRS to retrieve LST with high-resolution with the same scan time. We referred to historical meteorological dates from Greenhouse Data [56] as shown in Table 2-6. According to the date, we preliminarily excluded spring and winter, because the acquisition dates of Landsat 8 OLI/TIRS and MOD11A2 were too far apart. Compared with spring and winter, the dates in summer and autumn were adjacent. The scan times [8,57] of the two remote sensing types was similar, concentrated around 10:30 a.m., and only 2 or 3 minutes apart. Then, we organized and analyzed the obtained meteorological data. The smallest difference was in summer (with an asterisk in Table 2-6). The maximum and minimum air temperatures differed by only 1 °C; the difference in relative humidity was 1%; the wind speed difference was 0.1 m/s; and the solar radiation difference was about 0.4 MJ/m2. Compared with other sets of data, we selected the Landsat 8 TIRS on July 19, 2013 to retrieve LST, and combined with MOD11A2 on July 20 to verify downscaling accuracy.

Table 2-6. RF model parameter list and the optimal combination of this model.

Season	Spring		Summer*		Autumn		Winter	
RS type	LC08	MOD	LC08	MOD	LC08	MOD	LC08	MOD
Date	04.14	04.07	07.19	07.20	11.08	11.09	01.27	01.17
Scan Time (a.m.)	10:33	10:30	10:33	10:30	10:33	10:30	10:32	10:30
AT_{max}/°C	28	17	37	36	25	28	11	12
AT_{min}/°C	15	5	27	28	13	16	2	0
RH/%	33	39	49	50	69	67	35	34
Wind Speed/m/s	4.0	1.6	2.7	2.8	1.8	2.0	1.9	1.2
SR/MJ/m^2	31.6	34.6	35.6	36.0	12.6	13.4	11.1	10.9

Note: LC08 – Landsat 8 OLI/TIRS; MOD – MOD11A2; AT – air temperature; RH - relative humidity; SR – Solar Radiation; * - the season selected for validation

Using the single channel algorithm, the retrieved LST (that is, the real LST) with a resolution of 30 m is shown in Figure 2.16(b). Figure 2.16(a) shows the downscaled LST with a resolution of 100 m. The differences in the LST ranges of the two figures is small: the real LST ranges from 34.89 °C to 41.62 °C, while the other ranges from 33.18 °C to 42.82 °C. According to the LST distribution, the high-temperature areas (red areas) and low-temperature areas (blue areas) are basically similar. From the comparison of the downscaled and real LST, we present the error distribution histogram as shown in Figure 2.16(c). The x-axis represents the error between the two types of LST, and the y-axis represents the number of pixels. The error is approximately normally distributed. The peak value of the error is around 0.3 °C, the mean of errors is about 0.2617 °C, the median is about 0.3 °C, and the standard deviation is about 1.56 °C. Since the scan times and scan methods of Landsat 8 TIRS and MODIS/LST are different, as mentioned above, and the meteorological conditions at the two time were not exactly the same, a few errors between the downscaled and real LST. This small number of errors meets the research requirement, that is, the downscaled LST based on the RF model meets the accuracy requirement and the RF model can be used in downscaling research.
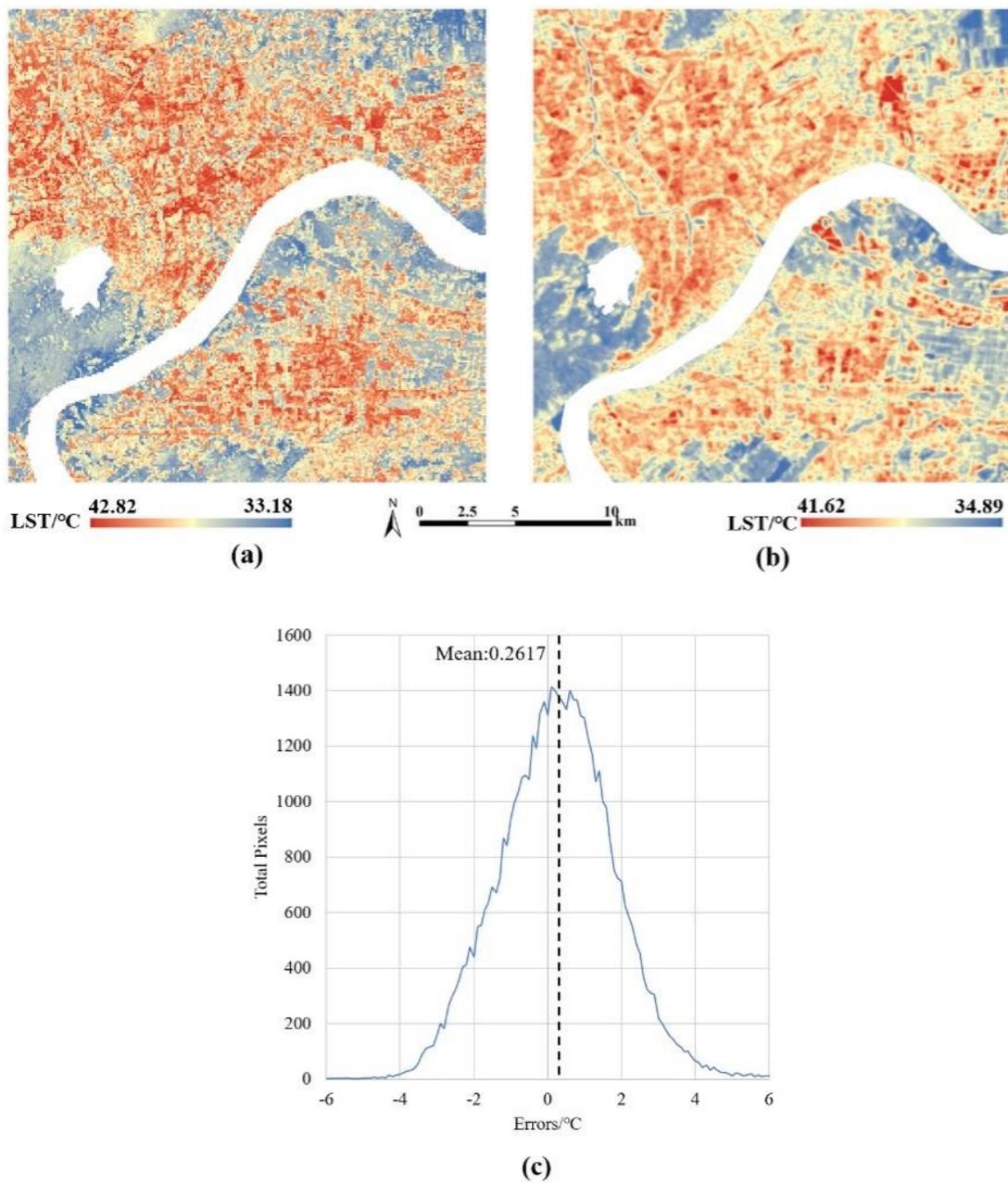
Figure 2.16. Validation results during summer daytime: (a) downscaled LST with a resolution of 100 m; (b) retrieved LST with a resolution of 30 m; (c) error distribution histogram between downscaling LST and real LST. (the black dashed line represents the mean value).

c

2.2.6.   Summary

This section used three different models (TsHARP, MLR and RF models) to downscale MODIS/LST products from 1 km to 100 m based on Landsat 8 OLI and a DEM with high-resolution data, and selected the highly heterogeneous Hangzhou urban area as the research region. Of the three model types examined, the TsHARP model was a single-factor regression model that favors nonlinearity based on the correlation between LST and NDVI. The MLR model was a multi-factor linear regression model, which introduced more independent variables compared to the TsHARP model, including DEM, NDBI, and Landsat 8 OLI Band 2 to Band 7. The RF model was used as a multi-factor nonlinear regression model based on machine learning to predict LST. Then, we used R^2 and RMSE to evaluate the prediction effect of these three models. According to the evaluation comparison, the suitable model—that is, the RF model—was selected for the subsequent downscaling study. After parameter tuning, we built the optimal RF model to downscale the LST products obtained from MODIS for four seasons during day and night, and analyzed SUHI based on the high-resolution LST. Finally, we selected similar retrieved LST based on Landsat 8 TIRS to verify the feasibility of the RF model.

However, the choice of independent variables in this section was flawed. This section selected DEM, NDVI, NDBI, and Landsat 8 OLI Band 2 to Band 7, thus including only two topography derived variables, NDVI and NDBI. In other studies, Hamid and Mohsen [58] selected ratio vegetation index (RVI), difference vegetation index (DVI), renormalized difference vegetation index (RDVI), NDVI, soil adjusted vegetation index (SAVI) and modified soil-adjusted vegetation index (MSAVI), while Wei Z. et al. [59] chose NDVI, enhanced vegetation index (EVI), normalized difference vegetation index (NDWI), leaf area index (LAI), surface albedo (ALB), surface elevation (ELV) and surface slope (SLP). In precision evaluation, we were unable retrieve high-resolution LST based on Landsat 8 TIRS for the same periods to verify downscaling accuracy because the Terra Satellite, which provides MODIS/LST products, and Landsat 8 OLI/TIRS are not synchronized. We were only able to select a few high-resolution LST data at particular times to verify the downscaling accuracy.

We concluded that the proposed RF model downscaling method, based on the multi-factor nonlinear regression of LST and DEM, NDVI, NDBI, and Landsat 8 OLI Band 2 to Band 7, was proven to be effective and flexible in downscaling the LST spatial resolution from 1 km to 100 m for various seasons in the research region. Compared to the downscaling methods based on the TsHARP model with single-factor nonlinear regression and the MLR model with multi-factor linear regression, both statistics and visual analysis supported this conclusion. According to the prediction precision, the RF model effects in winter and summer were slightly better than those in spring and autumn, and those at night were better than those during the day. Combined with high-resolution LST, we concluded that SUHI was spread throughout the city across a large area, with the exception of the hills to the southwest of West Lake. At night, SUHI shrank sharply in the urban center around West Lake and the low

2-31

temperature area increased. According to the LST difference across the four seasons, we found that SUHI was the most obvious in autumn and the weakest in winter. Finally, the error distribution histogram between the downscaled and real LST supported the conclusion that the RF model can be applied to downscaling research in highly heterogeneous regions.

2.3. Air Temperature Downscaling for Atmosphere Urban Heat Island Analysis

UHIs refers to the phenomenon that the air temperature in the urban center is almost always higher than that in the surrounding areas due to the continuous urbanization process, the closer to the urban center, the higher the air temperature [3]. Rapid urbanization has led to urban population explosion. More than 50% population lives in urban around the world, however the proportion of urban areas is only less than 3% in the global land area. Urban have provided the main driving force of global climate change [60].

UHIs, as an emergent property of a metropolitan area [61], are distinguish between surface UHI (SUHI) and atmosphere UHI (AUHI) [62]. Although this two UHI types are formed in different ways, they may be related to the exchange of energy between the land surface and the atmosphere [63]. SUHI can be directly monitored by infrared thermography mounted on satellite and airplanes. Remote sensing can provide regular and fixed-site observation of land surface temperature (LST), which greatly promoted the SUHI research [64–66]. AUHI is mainly detected surface air temperature above the city. It is the earliest research on UHI. Howard reported the first study on UHI in 1818 [67]. AUHI is usually conducted by simultaneous measuring the air temperature at meteorological station in urban areas and suburban areas. As a result, AUHI can only analyze the air temperature of the observation points where the meteorological stations are located. However, the fixed-point measured air temperature date cannot reveal the temperature of entire urban and surrounding areas like SUHI based on remote sensing [68]. This chapter considered the latter, this is AUHI. In order to analyze the climate issue closely related to national life, the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) of Japan released detail grid data of the 2010 climate (including air temperature and precipitation) with a spatial resolution of only 1 km throughout Japan in 2012. Take the Kanto Major Metropolitan Area as an example, it has a large population and dense street blocks. The low-resolution climate raster data cannot accurately describe the climate conditions. In order to meet the local application of air temperature data, we downscaled the air temperature data from 1 km to 250 m.

According to Section 2.2, the comparative study found that the machine learning model has higher accuracy than the traditional TsHARP model and MLR model. Against this background, the main purpose is to indicate that the AT obtained from MLIT can be downscaled from 1 km to 250 m, which makes it a detailed data for analyzing the high-resolution AUHI of metropolitan areas in Japan. DEM and urban structure which identified by maximum likelihood classification method from remote sensing were selected as independent variables, and the 2010 annual average air temperature data as dependent variable. We established the regression relationship between dependent and resampled independent variables at low resolution by using extra trees (ET) model and used K-fold cross-validation (CV) to check the precision of the downscaling model to confirm whether it was overfitting or underfitting. Using tested model, we predicted the air temperature result with a spatial resolution of 250 m. Finally, we fitted residuals of the high-resolution predicted air temperature based on the low-

resolution original air temperature and the predicted air temperature to obtain the downscaled air temperature result.

2.4. Comparison and Selection of Machine Learning Model

The downscaling results in Section 2.3 show that the RF model which belongs to machine learning model, has higher accuracy that the traditional TsHARP and MLR models. Machine learning is the study of computer algorithms that improve automatically through experience and by the use of data. It builds a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so [69]. Machine learning is mainly divided into supervised learning and unsupervised learning. The difference between them is whether the target of the training samples is labeled by manual. This research mainly uses supervised learning.

Supervised learning learns a function from a given training samples When new data inputs, it can predict the result based on this function. The training samples requirements for supervised learning include input and output, which can also be said to be features and targets.

In this section, we selected five supervised learning models for screening and comparison, including support vector regression model, K-nearest neighbors model, decision trees model, random forest model, and extra trees model, so as to select the optimal model for the next downscaling. The results are as follows.

2.4.1.   Support Vector Regression Model

Support vector is supervised learning models with associated learning algorithms that analyze data for classification and regression analysis. This section mainly used the regression version of the support vector machine algorithms. The regression version was proposed in 1996 by Harris Drucker et al. [70]. The support vector regression model is not only suitable for linear but also for nonlinear regression. But the traditional support vector machine algorithm is only appropriate for dichotomous classification models. Although this problem can be solved by a combination of multiple support vector regression, the accuracy will be affected.

Figure 2.17 shows the scatter plot of real air temperature versus predicted air temperature based on support vector regression model. The dotted line is the 1:1 line, color bar represents the frequency of scatter. The R^2 is about 0.738. The results show that support vector regression model has a poor prediction accuracy for the area whose air temperatures are below 11 ℃, and there is a clear dividing line around 11 ℃, which is obviously very unfavorable for the air temperature downscaling. So the support vector regression model is not suitable for air temperature downscaling.
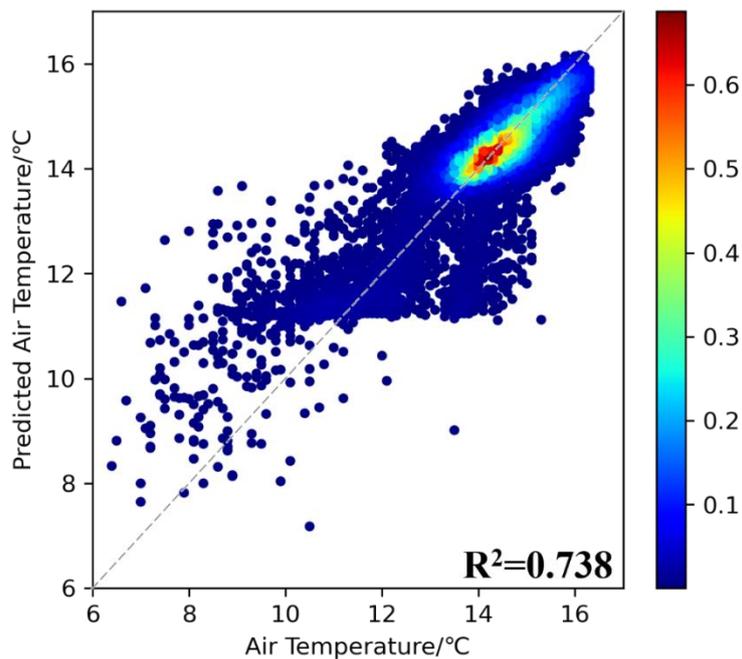
Figure 2.17. Scatter plot of real air temperature (x-axis) versus predicted air temperature (y-axis) based on support vector regression model

2.4.2. K-nearest Neighbors Model

The k -nearest neighbors algorithm is a non-parametric classification method first developed by Evelyn Fix and Joseph Hodges in 1951 [71], and later expanded by Thomas Cover [72]. The basic principle of k-nearest neighbors is very simple. It is to calculate the K data points closest to the target point, and then predict this target point. In k -nearest neighbors regression, the k -nearest neighbors algorithm is used for estimating continuous variables. One such algorithm uses a weighted average of the k nearest neighbors, weighted by the inverse of their distance.

Figure 2.18 shows the scatter plot of real air temperature versus predicted air temperature predicted by k-nearest neighbors model. Compared to the prediction result by support vector regression model, the prediction accuracy of this model is much higher, R^2 is about 0.792. And there is no obvious data dividing line in the whole range, but there are still many abnormal points, so k-nearest neighbors model cannot be used for air temperature downscaling.
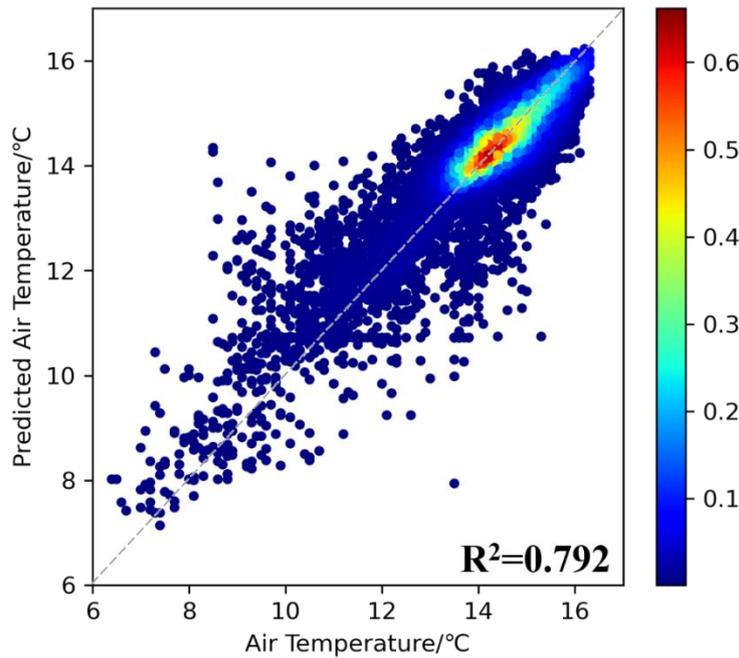
Figure 2.18. Scatter plot of real air temperature (x-axis) versus predicted air temperature (y-axis) based on k-nearest neighbors model

2.4.3.   Decision Trees Model

Decision trees are a parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. Decision trees are simple to understand and interpret. But they are often relatively inaccurate. Many other predictors perform better with similar data. This can be remedied by replacing a single decision tree with a random forest of decision trees, but a random forest is not as easy to interpret as a single decision tree. And They are unstable, meaning that a small change in the data can lead to a large change in the structure of the optimal decision tree.

Figure 2.19 shows scatter plot of real air temperature versus predicted air temperature based on decision trees. As mentioned above, it is relatively inaccurate. Although, the R^2 reached 0.847, which is much higher than support vector regression model and k-nearest neighbors model, we found that there are many points on the 1:1 line, which indicates that this model is overfitting seriously.

In addition, a line composed of many scattered points located near 11 °C. And there were more outliers in the mid-air temperature area. Therefore, this model cannot be used for air temperature downscaling either.
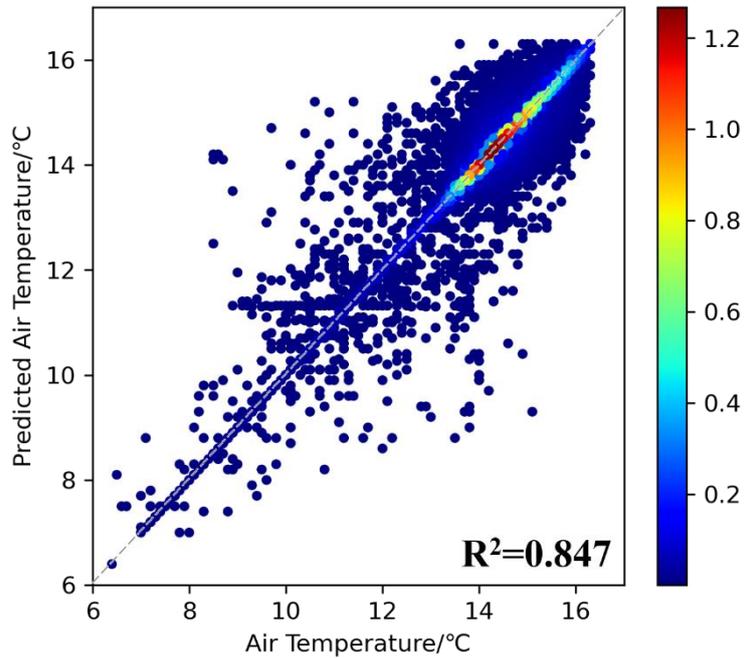
Figure 2.19. Scatter plot of real air temperature (x-axis) versus predicted air temperature (y-axis) based on decision trees model

2.4.4.   Random Forest Model

Random forest model has been verified in Section 2.2 that it can be used to downscale the land surface temperature. In order to further compare with other machine learning models and select the most suitable model for air temperature downscaling, we set the default values as each parameter to predict air temperature on the study area.

The comparison result is shown in Figure 2.20. the scatter plot indicates that the prediction accuracy of the random forest model is significantly higher than the above three machine learning models. The R^2 is about 0.882, and the random forest model is not overfitting obviously. In summary, random forest model can provide technology for the next air temperature downscaling. But in order to improve the prediction accuracy, we also verified the extra trees model.
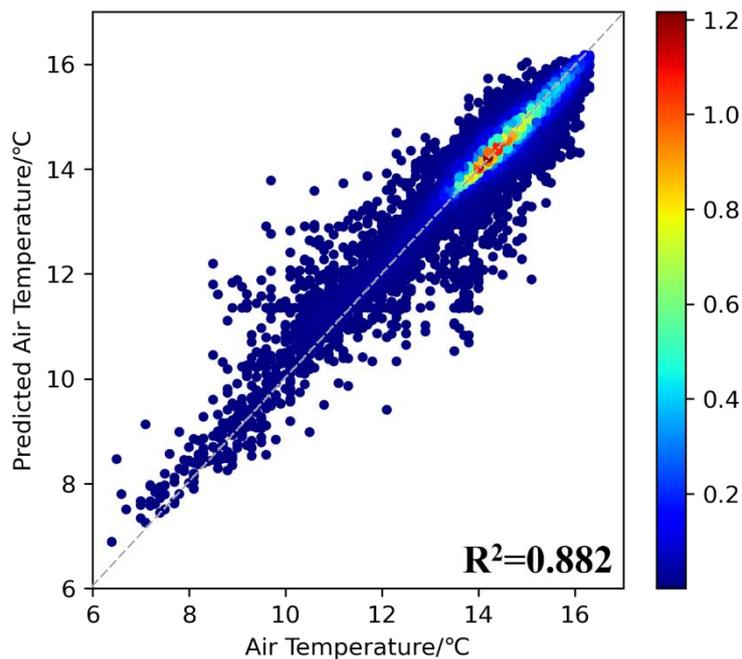
Figure 2.20. Scatter plot of real air temperature (x-axis) versus predicted air temperature (y-axis) based on random forest model

2.4.5.   Extra Trees Model

Random forest model adds a further step of randomization yield extra trees model. Similarly to ordinary random forest models, they are an integration of single decision trees. First, each decision tree is trained using the whole training sample, and second, the top-down splitting in the tree learner is randomized. Instead of computing the locally optimal cut-point for each feature under consideration, a random cut-point is selected. This value is selected from a uniform distribution within the feature's empirical range. Then, of all the randomly generated splits, the split that yields the highest score is chosen to split the node. Similar to ordinary random forests, the number of randomly selected features to be considered at each node can be specified [73].

Comparing to the prediction results of random forest model, R2 of extra trees model is about 0.912, which is significantly higher than the prediction accuracy of the random forest model. And extra trees model is not overfitting. Therefore, we selected the extra trees model which is one of the most popular machine learning model as the further air temperature downscaling model.
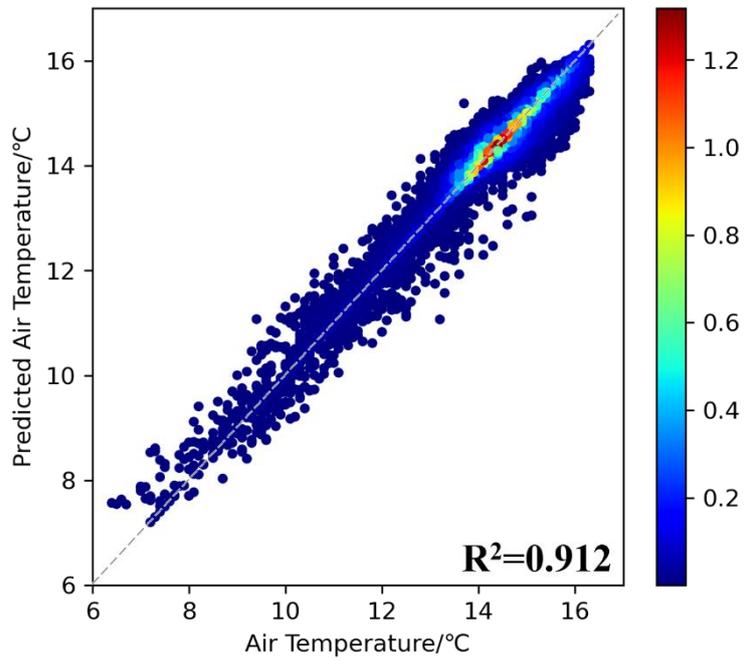
Figure 2.20. Scatter plot of real air temperature (x-axis) versus predicted air temperature (y-axis) based on random forest model

2.5. Extra Trees Model for Air Temperature Downscaling

Combining the previous sections and some other previous studies, we found that compared with other downscaling methods, the downscaling accuracy of the extra trees model is significantly higher than that of other models. [15,74]. This section selected ET model as the downscaling model. ET model, also known as extremely randomized trees, is a relatively recent machine learning model used to perform classification and regression tasks [75], which is an extension of random forest model [76]. This model belongs to the class of ensemble learning methods based on decision trees model[77], and unlikely to be overfitting [73]. Decision tree is a basic regression model, the structure is shown in Figure 2.22. Decision tree is a typical binary tree structure, the value of internal node characteristics is "Yes" or "No". The tree is to divide the feature space into several units, and each divided unit has a specific output, that is the "c" in the decision tree structure. According to the input variables, the feature space is divided. The process of division is the process of building a tree. Each time it is divided, one more node is added. When the division is stopped, the final output of each unit is determined, thereby establishing the entire decision tree. ET model is an algorithm that integrates multiple trees through ensemble learning. The structure of ET model is shown in Figure 2.23. The model randomly sampling samples from the input variables based on the bootstrap algorithm, and then building N corresponding decision trees based on these samples. In the process of building a decision tree, different from the RF model, the feature of each node of each decision tree in RF model is fixed, and only the input sample of each decision tree is random. However, in ET model, the feature of each node of each decision tree are also random. Then, vote according to the generated decision tree to select an optimal tree as the regression model.
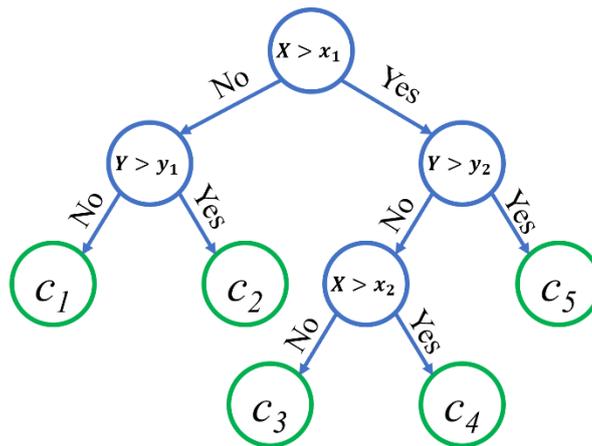
Figure 2.22. The structure of a single decision tree.
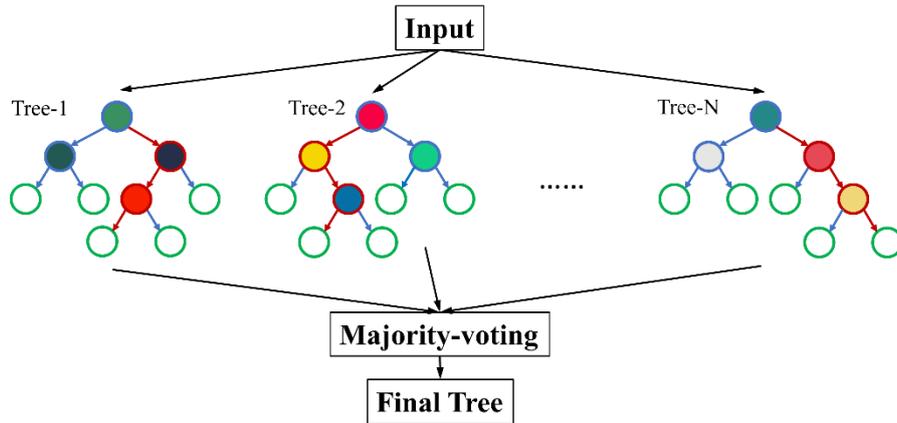
Figure 2.23. The structure of extra trees model.

ET model has the similar principle as random forest model. It uses bootstrap resampling technology to extract from initial samples based on the randomly selected best feature to establish a new series of training samples, and then create decision trees and establishes an ET model. Compared with the random forest model, ET model uses the whole samples to train each regression trees [76].

In the following establishing process (shown in Figure 2.24), low resolution parameters at a spatial resolution of 1 km included two types: dependent variables (AT from NLNI) and independent variables (DEM, and urban structure). In order to test the model precision, we divided the sample into ten folds equally based on K-fold CV [78], and randomly extract one fold as the test sample, and the other nine fold (train sample) were used to train ET model. We used scikit-learn open-source machine learning libraries based on Python 3.8, which is the most famous machine learning libraries [29], to train and test the ET model. In the establishing process, several parameters needed to be tuned, including two parts: ET framework parameters, namely n_estimators, oob_score, and criterion; and ET decision trees parameters, namely max_features, max_depth, min_samples_split, min_samples_leaf, min_weight_fraction_leaf, max_leaf_nodes, and min_impurity_split [79]. Among them, n_estimators, max_features, max_depth, and min_samples_leaf are the four most important parameters that influence the downscaling accuracy. For preventing overfitting and underfitting in ET model, we adjusted these four parameters for the good fitting state to establish the ideal model according to model score, train and test mean square error (MSE). Finally, based on previous training and test samples, K-fold cross validation was used to determine the feasibility of the ET model.
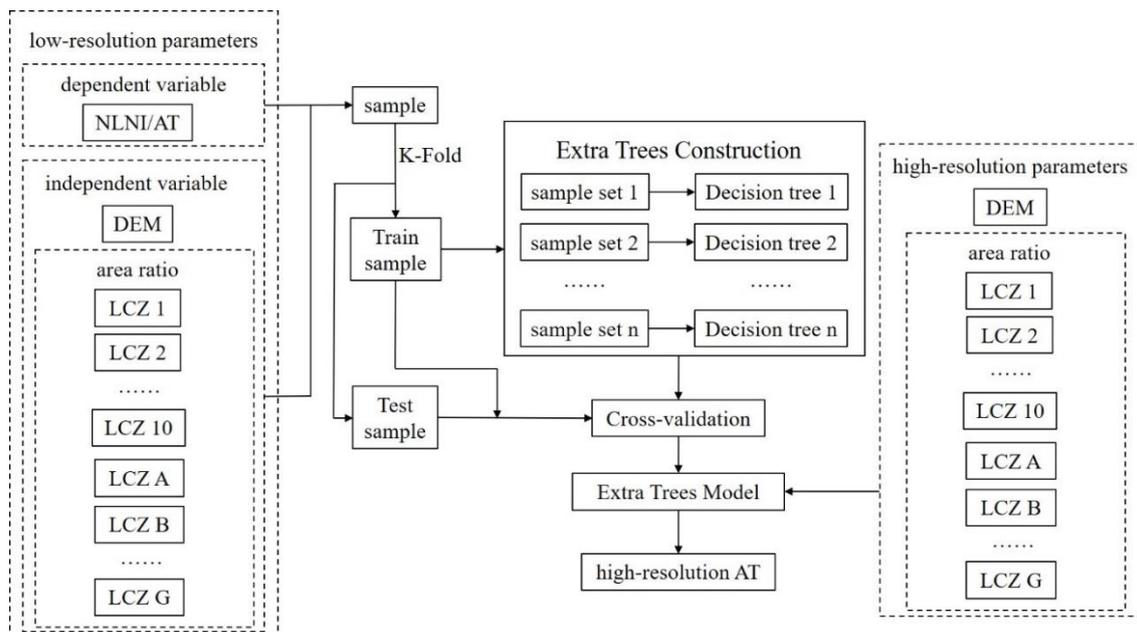
Figure 2.24. Establishing process of the extra trees model.

2.6. Residual Fitting

In this section, we used the same method as introduced in Section 2.2.1 to fit the residuals. In the establishing process for a low-resolution ET model, there were residuals between the real image and the predicted image. We fitted residual to high-resolution predicted results in order improve downscaling accuracy. Residuals were downscaled by spline interpolator [80]. Figure 2.25 shows the flow chart of residual fitting. The last step was to fit the residual with high resolution to the predicted results with high resolution, finally resulting in AT data with high resolution.
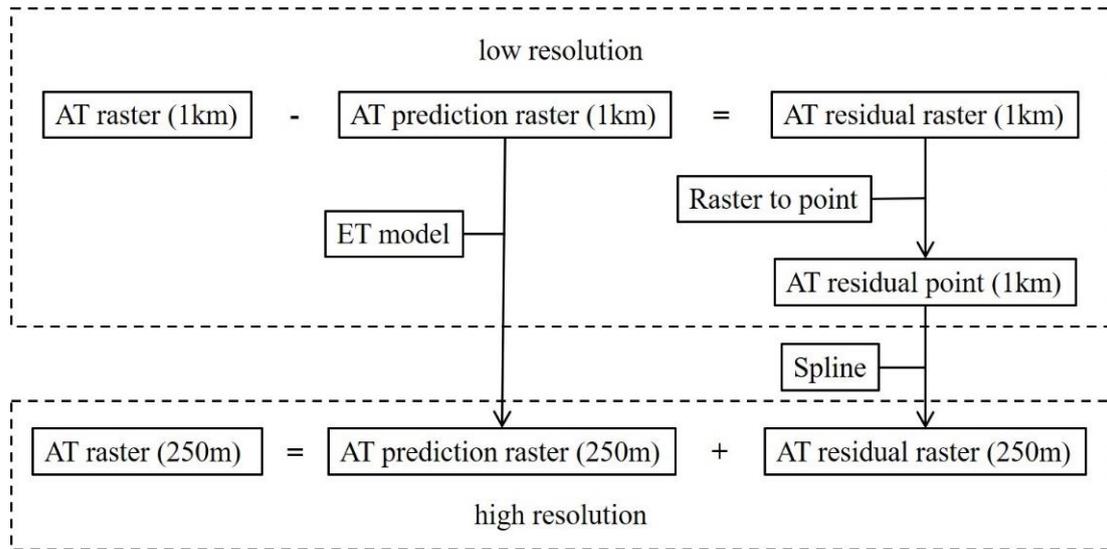
Figure 2.25. Residual fitting flow chart in air temperature downscaling

2.7. Summary

This chapter is the core of this study. In this chapter, we introduced the detailed downscaling method. The principle of temperature downscaling is that under different scales, the regression between the underlying surface characteristics and temperature remains unchanged. Firstly, we selected Hangzhou, China as a case study to compare three downscaling models, including thermal sharpen model, multiple linear regression model, and random forest model. The result shows that the random forest model has the highest downscaling accuracy. Next, according to the above results, we selected five models based on machine learning algorithm, and selected the optimal model for air temperature downscaling. The result shows that extra trees model is the most suitable compared with other downscaling models based on machine learning algorithm. Then, we introduced the parameters tuning method to establish the optimal model. Finally, through the residual fitting, we can obtain the high-resolution air temperature.

Reference

1. Weng, Q.; Lu, D. A sub-pixel analysis of urbanization effect on land surface temperature and its interplay with impervious surface and vegetation coverage in Indianapolis, United States. *Int. J. Appl. earth Obs. Geoinf.* **2008**, *10*, 68–83.

2. Stathopoulou, M.; Cartalis, C. Downscaling AVHRR land surface temperatures for improved surface urban heat island intensity estimation. *Remote Sens. Environ.* **2009**, *113*, 2592–2605.

3. Zakšek, K.; Oštir, K. Downscaling land surface temperature for urban heat island diurnal cycle analysis. *Remote Sens. Environ.* **2012**, *117*, 114–124, doi:10.1016/j.rse.2011.05.027.

4. Kustas, W.; Anderson, M. Advances in thermal infrared remote sensing for land surface modeling. *Agric. For. Meteorol.* **2009**, *149*, 2071–2081.

5. Chunqiao, S.; Songcai, Y.O.U.; Gaohuan, L.I.U.; Linghong, K.E.; Xinke, Z. The spatial pattern of soil moisture in Northern Tibet based on TVDI method. *Prog. Geogr.* **2011**, *30*, 569–576.

6. Anderson, M.C.; Allen, R.G.; Morse, A.; Kustas, W.P. Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources. *Remote Sens. Environ.* **2012**, *122*, 50–65.

7. Ha, W.; Gowda, P.H.; Howell, T.A. A review of downscaling methods for remote sensing-based irrigation management: Part I. *Irrig. Sci.* **2013**, *31*, 831–850.

8. Survey, U.S.G. No Title Available online: https://earthexplorer.usgs.gov/.

9. Quan, J.; Zhan, W.; Chen, Y.; Liu, W. Downscaling remotely sensed land surface temperatures: A comparison of typical methods. *Yaogan Xuebao- J. Remote Sens.* **2013**, *17*, 361–387.

10. Zhan, W.; Chen, Y.; Zhou, J.; Wang, J.; Liu, W.; Voogt, J.; Zhu, X.; Quan, J.; Li, J. Disaggregation of remotely sensed land surface temperature: Literature survey, taxonomy, issues, and caveats. *Remote Sens. Environ.* **2013**, *131*, 119–139.

11. Kustas, W.P.; Norman, J.M.; Anderson, M.C.; French, A.N. Estimating subpixel surface temperatures and energy fluxes from the vegetation index-radiometric temperature relationship. *Remote Sens. Environ.* **2003**, *85*, 429–440, doi:10.1016/S0034-4257(03)00036-1.

12. Agam, N.; Kustas, W.P.; Anderson, M.C.; Li, F.; Neale, C.M.U. A vegetation index based technique for spatial sharpening of thermal imagery. *Remote Sens. Environ.* **2007**, *107*, 545–558, doi:10.1016/j.rse.2006.10.006.

13. Essa, W.; Verbeiren, B.; van der Kwast, J.; Van de Voorde, T.; Batelaan, O. Evaluation of the DisTrad thermal sharpening methodology for urban areas. *Int. J. Appl. earth Obs. Geoinf.* **2012**, *19*, 163–172.

14. Weng, Q.; Fu, P.; Gao, F. Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data. *Remote Sens. Environ.* **2014**, *145*, 55–67.

15. Hutengs, C.; Vohland, M. Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sens. Environ.* **2016**, *178*, 127–141, doi:10.1016/j.rse.2016.03.006.

16.   Bonafoni, S. Downscaling of Landsat and MODIS land surface temperature over the heterogeneous urban area of Milan. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 2019–2027.

17.   Govil, H.; Guha, S.; Dey, A.; Gill, N. Seasonal evaluation of downscaled land surface temperature: A case study in a humid tropical city. *Heliyon* **2019**, *5*, e01923.

18.   Junwei, H.U.A.; Shanyou, Z.H.U.; Guixin, Z. Downscaling land surface temperature based on random forest algorithm. *Remote Sens. L. Resour.* **2018**, *30*, 78–86.

19.   Zhan, W.; Chen, Y.; Wang, J.; Zhou, J.; Quan, J.; Liu, W.; Li, J. Downscaling land surface temperatures with multi-spectral and multi-resolution images. *Int. J. Appl. earth Obs. Geoinf.* **2012**, *18*, 23–36.

20.   Peckham, R.J.; Gyozo, J. Development and Applications in a Policy Support Environment Series: Lecture Notes in Geoinformation and Cartography. *Heidelberg, Ger.* **2007**.

21.   Macarof, P.; Statescu, F. Comparasion of ndbi and ndvi as indicators of surface urban heat island effect in landsat 8 imagery: A case study of iasi. *Present Environ. Sustain. Dev.* **2017**, *11*, 141–150.

22.   Sattari, F.; Hashim, M.; Pour, A.B. Thermal sharpening of land surface temperature maps based on the impervious surface index with the TsHARP method to ASTER satellite data: A case study from the metropolitan Kuala Lumpur, Malaysia. *Measurement* **2018**, *125*, 262–278.

23.   Petropoulos, G.; Carlson, T.N.; Wooster, M.J.; Islam, Sjp. A review of Ts/VI remote sensing based methods for the retrieval of land surface energy fluxes and soil surface moisture. *Prog. Phys. Geogr.* **2009**, *33*, 224–250.

24.   Wilby, R.L.; Dawson, C.W.; Barrow, E.M. SDSM—a decision support tool for the assessment of regional climate change impacts. *Environ. Model. Softw.* **2002**, *17*, 145–157.

25.   Wang, Y.; Xie, D.; Li, Y. Downscaling remotely sensed land surface temperature over urban areas using trend surface of spectral index. *J. Remote Sens.* **2014**, *18*, 1169–1181.

26.   Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.

27.   Ho, T.K. Random decision forests. In Proceedings of the Proceedings of 3rd international conference on document analysis and recognition; IEEE, 1995; Vol. 1, pp. 278–282.

28.   Hais, M.; Kučera, T. The influence of topography on the forest surface temperature retrieved from Landsat TM, ETM+ and ASTER thermal channels. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 585–591.

29.   Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

30.   Matthew, W. Bias of the Random Forest out-of-bag (OOB) error for certain input parameters. *Open J. Stat.* **2011**, *2011*.

31.   Draper, N.R.; Smith, H. *Applied regression analysis*; John Wiley & Sons, 1998; Vol. 326; ISBN 0471170828.

32.   Willmott, C.J.; Matsuura, K. On the use of dimensioned measures of error to evaluate the performance of spatial interpolators. *Int. J. Geogr. Inf. Sci.* **2006**, *20*, 89–102.

33.   Giannini, M.B.; Belfiore, O.R.; Parente, C.; Santamaria, R. Land Surface Temperature from Landsat 5 TM images: comparison of different methods using airborne thermal data. *J. Eng. Sci. Technol. Rev.* **2015**, *8*.

34.   Dissanayake, D.; Morimoto, T.; Murayama, Y.; Ranagalage, M. Impact of landscape structure on the variation of land surface temperature in sub-saharan region: A case study of Addis Ababa using Landsat data (1986–2016). *Sustainability* **2019**, *11*, 2257.

35.   Carlson, T.N.; Ripley, D.A. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sens. Environ.* **1997**, *62*, 241–252.

36.   Artis, D.A.; Carnahan, W.H. Survey of emissivity variability in thermography of urban areas. *Remote Sens. Environ.* **1982**, *12*, 313–329.

37.   Zanter, K. Landsat 8 (L8) data users handbook. *Landsat Sci. Off. Website* **2016**.

38.   Mutiibwa, D.; Strachan, S.; Albright, T. Land surface temperature and surface air temperature in complex terrain. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 4762–4774.

39.   Sheng, J.; Wilson, J.P.; Lee, S. Comparison of land surface temperature (LST) modeled with a spatially-distributed solar radiation model (SRAD) and remote sensing data. *Environ. Model. Softw.* **2009**, *24*, 436–443.

40.   Sang, Y.-F. Spatial and temporal variability of daily temperature in the Yangtze River Delta, China. *Atmos. Res.* **2012**, *112*, 12–24.

41.   Olmos-Trujillo, E.; González-Trinidad, J.; Júnez-Ferreira, H.; Pacheco-Guerrero, A.; Bautista-Capetillo, C.; Avila-Sandoval, C.; Galván-Tejada, E. Spatio-Temporal Response of Vegetation Indices to Rainfall and Temperature in A Semiarid Region. *Sustainability* **2020**, *12*, 1939.

42.   Loyd, C. Putting Landsat 8's Bands to Work. *NASA, USA* **2013**.

43.   Zha, Y.; Gao, J.; Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* **2003**, *24*, 583–594.

44.   Purevdorj, T.S.; Tateishi, R.; Ishiyama, T.; Honda, Y. Relationships between percent vegetation cover and vegetation indices. *Int. J. Remote Sens.* **1998**, *19*, 3519–3535.

45.   Wan, Z.; Dozier, J. A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 892–905.

46.   Sandholt, I.; Rasmussen, K.; Andersen, J. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sens. Environ.* **2002**, *79*, 213–224.

47.   Inamdar, A.K.; French, A. Disaggregation of GOES land surface temperatures using surface

emissivity. *Geophys. Res. Lett.* **2009**, *36*.

48.   Jeganathan, C.; Hamm, N.A.S.; Mukherjee, S.; Atkinson, P.M.; Raju, P.L.N.; Dadhwal, V.K. Evaluating a thermal image sharpening model over a mixed agricultural landscape in India. *Int. J. Appl. earth Obs. Geoinf.* **2011**, *13*, 178–191.

49.   Chen, X.; Li, W.; Chen, J.; Rao, Y.; Yamaguchi, Y. A combination of TsHARP and thin plate spline interpolation for spatial sharpening of thermal imagery. *Remote Sens.* **2014**, *6*, 2845–2863.

50.   Probst, P.; Boulesteix, A.-L. To Tune or Not to Tune the Number of Trees in Random Forest. *J. Mach. Learn. Res.* **2017**, *18*, 6673–6690.

51.   Zhao, W.; Sánchez, N.; Lu, H.; Li, A. A spatial downscaling approach for the SMAP passive surface soil moisture product using random forest regression. *J. Hydrol.* **2018**, *563*, 1009–1024, doi:10.1016/j.jhydrol.2018.06.081.

52.   Jingchao, J.; Junzhi, L.; Chengzhi, Q.; Yamin, M.; A-Xing, Z. Near-surface air temperature lapse rates and seasonal and type differences in China. *Prog. Geogr.* **2016**, *35*, 1538–1548, doi:10.18306/dlkxjz.2016.12.010.

53.   Måren, I.E.; Karki, S.; Prajapati, C.; Yadav, R.K.; Shrestha, B.B. Facing north or south: Does slope aspect impact forest stand characteristics and soil properties in a semiarid trans-Himalayan valley? *J. Arid Environ.* **2015**, *121*, 112–123, doi:10.1016/j.jaridenv.2015.06.004.

54.   Jiyu, C.; Cangzi, L.; Chongle, Z.; Walker, H.J. Geomorphological development and sedimentation in Qiantang Estuary and Hangzhou Bay. *J. Coast. Res.* **1990**, 559–572.

55.   Hu, P.; Han, J.; Li, W.; Sun, Z.; He, Z. Numerical investigation of a sandbar formation and evolution in a tide-dominated estuary using a hydro-morphodynamic model. *Coast. Eng. J.* **2018**, *60*, 466–483.

56.   Greenhouse Data Available online: http://data.sheshiyuanyi.com/WeatherData/ (accessed on Apr 29, 2021).

57.   MODIS Available online: https://modis.gsfc.nasa.gov/about/specifications.php (accessed on Apr 29, 2021).

58.   Ebrahimy, H.; Azadbakht, M. Downscaling MODIS land surface temperature over a heterogeneous area: An investigation of machine learning techniques, feature selection, and impacts of mixed pixels. *Comput. Geosci.* **2019**, *124*, 93–102, doi:10.1016/j.cageo.2019.01.004.

59.   Zhao, W.; Wu, H.; Yin, G.; Duan, S.B. Normalization of the temporal effect on the MODIS land surface temperature product using random forest regression. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 109–118, doi:10.1016/j.isprsjprs.2019.04.008.

60.   Grimm, N.B.; Faeth, S.H.; Golubiewski, N.E.; Redman, C.L.; Wu, J.; Bai, X.; Briggs, J.M. Global change and the ecology of cities. *Science (80-.).* 2008, *319*, 756–760.

61.   Oke, T.R. The energetic basis of the urban heat island. *Q. J. R. Meteorol. Soc.* **1982**, *108*, 1–24, doi:10.1002/qj.49710845502.

62.   Clinton, N.; Gong, P. MODIS detected surface urban heat islands and sinks: Global locations and controls. *Remote Sens. Environ.* **2013**, *134*, 294–304, doi:10.1016/j.rse.2013.03.008.

63.   Huang, F.; Zhan, W.; Wang, Z.H.; Voogt, J.; Hu, L.; Quan, J.; Liu, C.; Zhang, N.; Lai, J. Satellite identification of atmospheric-surface-subsurface urban heat islands under clear sky. *Remote Sens. Environ.* **2020**, *250*, 112039, doi:10.1016/j.rse.2020.112039.

64.   Chakraborty, T.; Lee, X. A simplified urban-extent algorithm to characterize surface urban heat islands on a global scale and examine vegetation control on their spatiotemporal variability. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *74*, 269–280, doi:10.1016/j.jag.2018.09.015.

65.   Zhou, D.; Xiao, J.; Bonafoni, S.; Berger, C.; Deilami, K.; Zhou, Y.; Frolking, S.; Yao, R.; Qiao, Z.; Sobrino, J. Satellite Remote Sensing of Surface Urban Heat Islands: Progress, Challenges, and Perspectives. *Remote Sens.* **2018**, *11*, 48, doi:10.3390/rs11010048.

66.   Peng, S.; Piao, S.; Ciais, P.; Friedlingstein, P.; Ottle, C.; Bréon, F.M.; Nan, H.; Zhou, L.; Myneni, R.B. Surface urban heat island across 419 global big cities. *Environ. Sci. Technol.* **2012**, *46*, 696–703, doi:10.1021/es2030438.

67.   Howard, L. *The climate of London: deduced from meteorological observations made in the metropolis and at various places around it*; Harvey and Darton, J. and A. Arch, Longman, Hatchard, S. Highley [and] R. Hunter, 1833; Vol. 3;.

68.   Zhou, J.; Zhang, X.; Zhan, W.; Göttsche, F.M.; Liu, S.; Olesen, F.S.; Hu, W.; Dai, F. A Thermal Sampling Depth Correction Method for Land Surface Temperature Estimation From Satellite Passive Microwave Observation Over Barren Land. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4743–4756, doi:10.1109/TGRS.2017.2698828.

69.   Koza, J.R.; Bennett, F.H.; Andre, D.; Keane, M.A. Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. In *Artificial Intelligence in Design '96*; Springer Netherlands, 1996; pp. 151–170.

70.   Drucker, H.; Burges, C.J.C.; Kaufman, L.; Smola, A.; Vapnik, V. Support vector regression machines. *Adv. Neural Inf. Process. Syst.* **1997**, *9*, 155–161.

71.   Fix, E.; Hodges, J.L. Nonparametric discrimination: Consistency properties. *Randolph Field, Texas, Proj.* **1951**, 21–49.

72.   Altman, N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **1992**, *46*, 175–185, doi:10.1080/00031305.1992.10475879.

73.   Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42, doi:10.1007/s10994-006-6226-1.

74.   Mukherjee, S.; Joshi, P.K.; Garg, R.D. A comparison of different regression models for downscaling Landsat and MODIS land surface temperature images over heterogeneous landscape. *Adv. Sp. Res.* **2014**, *54*, 655–669, doi:10.1016/j.asr.2014.04.013.

75.   Gall, J.; Yao, A.; Razavi, N.; Van Gool, L.; Lempitsky, V. Hough forests for object detection,

tracking, and action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2188–2202, doi:10.1109/TPAMI.2011.70.

76. Ahmad, M.W.; Reynolds, J.; Rezgui, Y. Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *J. Clean. Prod.* **2018**, *203*, 810–821, doi:10.1016/j.jclepro.2018.08.207.

77. John, V.; Liu, Z.; Guo, C.; Mita, S.; Kidono, K. Real-time lane estimation Using Deep features and extra trees regression. In Proceedings of the Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer Verlag, 2016; Vol. 9431, pp. 721–733.

78. Xiong, Z.; Cui, Y.; Liu, Z.; Zhao, Y.; Hu, M.; Hu, J. Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation. *Comput. Mater. Sci.* **2020**, *171*, 109203, doi:10.1016/j.commatsci.2019.109203.

79. sklearn.ensemble.ExtraTreesRegressor — scikit-learn 0.24.1 documentation Available online: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesRegressor.html (accessed on Jan 31, 2021).

80. Quan, J.; Zhan, W.; Ma, T.; Du, Y.; Guo, Z.; Qin, B. An integrated model for generating hourly Landsat-like land surface temperatures over heterogeneous landscapes. *Remote Sens. Environ.* **2018**, *206*, 403–423, doi:10.1016/j.rse.2017.12.003.

Chapter 3. Urban Structure Identification and its Implication of Urban Climate

3.1. Introduction

Rapid urbanization has greatly changed urban structures, thereby impacting the urban climate and associated phenomena, such as, urban heat stress in particular [1–3]. Built-up and urban areas are characterized by less vegetation and a large number of impervious surfaces, which leads to low evaporation heat flux; high urban surface roughness, which reduces wind speed; and increase radiation absorption due to urban structures. These factors explain the primary reasons that urban climates differ from rural climates [4]. With the transformation of natural landscapes into impermeable surfaces, changes in land use and land cover have a profound impact on urban structure and urban heat stress [5]. During the past few decades, a large number of research methods have been developed to study urban structure. Scholars began to study the internal spatial structure of urban area and explore models related to these spaces before the 1950s [6]. As the world entered the post-modernization era in the 21st century, scholars focused more on the spatial structure of urban land use and land cover; For example, Inostroza et al [7]vv. studied the scale and spatial form of urban land from a static perspective, and Burgalassi and Luzzati [8] researched the efficiency of urban spatial structures. Currently, single/multi-center forms are one of the most important dimensions, representing a current hotspot and future development direction for urban spatial structure research [9,10]. The urban spatial structure is used to characterize areas with similar building types, open spaces, and land use [11]. Combined with research on urban heat islands (UHIs), Stewart et al. [12] proposed a new low-scale urban structure model called the local climate zone (LCZ) classification. This system could better characterize the urban structure, especially with respect to buildings of various heights and densities. The LCZ model divided the urban structure into 17 categories, including ten building types and seven land cover types.

The extraction of urban structures has become an important link in urban ecology and urban climate research. Traditional urban structure classification is mainly based on manual interpretation of the urban functional zoning map. For example, Wilson et al. [13] used extremely detailed urban functional zoning maps for evaluating urban vegetation and the urban thermal environment. Scholars have also begun to use high-resolution remote sensing data to distinguish urban structures. Puissant and Weber [14] analyzed the effect of high-resolution remote sensing data on urban structure monitoring and management and defined four types of urban structures that can be classified from remote sensing data: urban structures with continuous residential buildings, sparse urban structures with discontinuous residential buildings, low-rise urban structures with discontinuous residential buildings, and high-rise urban structures with continuous residential buildings. Herold et al. [15] characterized various types of urban structures and used high-resolution remote sensing data to realize the classification. Banzhaf and Hofer [11] defined the spatial distribution of land use and land cover in different urban structures based on grid units to construct a large-scale urban structure classification. Similarly, Montanges et al. [16] used multi-source data (land cover, building height, remote sensing data, etc.) to express the characteristics of urban structures, and used multiple classifiers to experiment with the urban structure functional area classification. High-resolution

remote sensing data provide a basis for realizing the urban structure classification, promoting the application of remote sensing technology in urban ecological environments, and planning management directly. This study combined the LCZ classification system and high-resolution remote sensing data to identify urban structures for further urban climate research.

Urban expansion has intensified UHIs [17,18]. The global climate change caused by UHIs has a huge impact on urban residents' health, causing human discomfort and even mortality [19,20]. Those who attend outdoor activities need a comfortable outdoor thermal environment [21]. Meanwhile, an uncomfortable outdoor thermal environment may affect social activities and even economic activities in an urban area [22]. Owing to the large amount of building materials, especially concrete and asphalt, huge quantities of solar radiation are stored and then released into the urban environment. This phenomenon, called the "canyon effect", combined with anthropogenic heat from air conditioning, vehicles, etc. [23], are all contributing to the UHI intensification [24]. Urban design strategies, such as urban structure adjustment, can be exploited to mitigate this phenomenon and the consequent negative impact on human health [25]. Urban morphology, building density, and building plot ratio affect the air temperature distribution, thereby influencing the urban thermal environment [26,27].

Urban structure planning requires a reliable method to ensure urban thermal comfort from the perspective of the urban microclimate [28], with the main goal of verifying whether the adopted structure is effective for optimizing the urban thermal environment [29]. In the past two decades, the evaluation of outdoor thermal comfort through numerical simulation modeling or field measurements has become popular. Although field measurements can provide highly accurate data, they are expensive and time-consuming tasks that can only show the thermal environment at designated locations and times [30]. Moreover, the urban microclimate is a complex system based on various parameters, involving countless natural and urban processes. Natural parameters such as air temperature, humidity, wind speed, and solar radiation are very sensitive to the 3D urban microclimate model [31]. Owing to the various processes that lead to different thermal environments, the most feasible method for urban microclimate simulation is numerical modeling [32]. Given the advantages of numerical modeling and the increasing computing power advances in personal computers, some simulation tools have emerged to help assess the impact of various urban structures on microclimate and human thermal comfort [33–36].

This study selected ENVI-met as the simulation software. ENVI-met can be considered one of the most commonly used software options for determining the impact of architectural design and urban planning on the microclimate [37,38]. Although there are several microclimate simulation tools, such as RayMan [35], Townscope [39] and SOLWEIG [40], ENVI-met is particularly popular because of its extremely high spatiotemporal resolution, advanced 3D interactive interface and modeling technology, and ability to adjust air temperature and humidity [31]. Through the overall analysis of this software, the air temperature, mean radiation temperature (MRT), relative humidity, wind speed, solar radiation, and other parameters can be

obtained as well as higher-order predicted mean vote (PMV) and physiological equivalent temperature (PET) values [29].

Several previous studies have shown that numerical simulation modeling can evaluate the impact of urban structure on the microclimate and outdoor thermal comfort. Ketterer and Matzarakis [41] studied the impact of urban canyons on urban thermal stress, and found that the heat stress can be reduced with a northwest–southeast orientation and an aspect ratio of at least 1.5. Krüger et al. [37] observed and evaluated the relationship between urban structure and urban microclimate changes, and Ali-Toudert and Mayer [42] used ENVI-met to analyze the relationship between outdoor thermal comfort and urban street design. They found that streets in different directions will have a moderate impact on the air temperature and have a strong impact on the heat generated by humans: the greater the openness of the urban canyon to the sky, the higher the heat stress during the daytime. Middel et al. [43] studied the outdoor thermal environment of the three main residential landscape types in different LCZs in the Phoenix metropolitan area in the United States, emphasizing how the dense urban structure determines the urban cool island effect. Johansson et al. [44] considered a typical summer in São Paulo, Brazil and analyzed the changes in the microclimate in six different types of urban structures. Wang et al. [45] compared the MRTs of different urban densities in Toronto, Canada based on ENVI-met simulation modeling to reflect the outdoor thermal environment; they evaluated the urban microclimate based on urban density, and the results showed that the duration of direct sunlight and MRT will greatly affect the thermal comfort of the city. After sunset, the MRT will drop and maintain balance with the air temperature.

In the past, most of the urban thermal environment research based on ENVI-met used MRT [45,46], PET [47,48] and PMV [49]. In contrast, this study used the discomfort index of the wet-bulb globe temperature (WBGT) to measure the thermal comfort. WBGT is the most commonly used heat stress index [50]. WBGT mainly depends on three factors: the wet-bulb temperature, black-globe temperature and dry-bulb temperature (see Equation (3-7) in section 3.2.4). The WBGT can also be calculated based on empirical formulas (see Equation (3-8) in section 3.2.4) [51].

This study intends to introduce a classification system that can be used as a classification standard for urban structure identification. The study area is divided into 17 types of LCZs through the maximum likelihood classification (MLC) method. The accuracy of the results was validated using a confusion matrix, and the Kanto major metropolitan area (MMA) was chosen as the study area in this chapter to verify the feasibility of this identification. Combined with the WBGT, ENVI-met was used to simulate and evaluate the urban microclimate changes of some quintessential LCZs under three different urban structure conditions. Figure 3.1 demonstrates the structure of this study. First, the LCZ classification system and research methods are presented. Second, the results are shown in a case study. Third, idealized models were created to simulate the microclimate of each LCZ. Fourth, the WBGT was calculated to evaluate the urban microclimate. The discussion and conclusion summarize and close out the research.
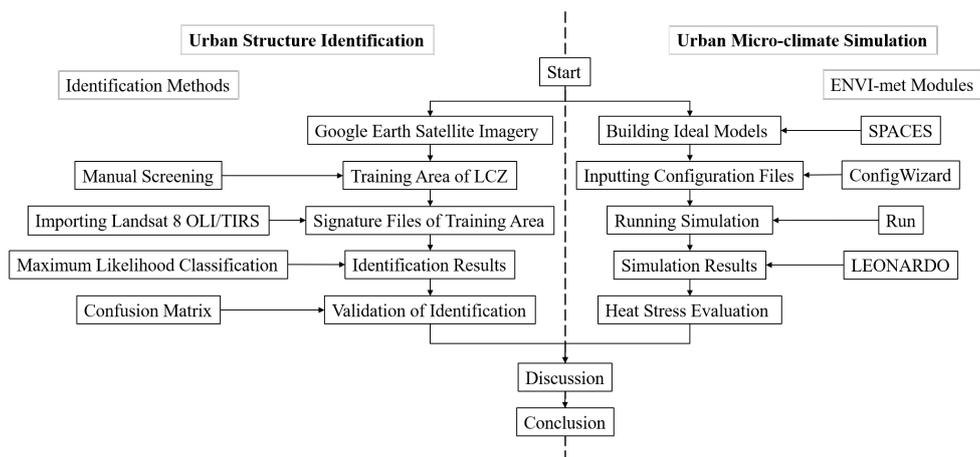
Figure 3.1. A flow chart of this research process, including identification and simulation.

3.2. Research methods

3.2.1.   Study area and data source

   The Kanto MMA as the most populous metropolitan area in the world, consists of the Tokyo Metropolis and the prefectures of Kanagawa, Chiba, Saitama, Ibaraki, Tochigi, and Gunma. The mainland portion of Kanto MMA lies northwest of Tokyo Bay and measures about 170 km east to west and 160 km north to south. It has an area of about 13,500 km^2 with a population of about 3.81 million in 2016 [52].

   The Kanto MMA has a subtropical monsoon climate with an annual average temperature of 15.6 °C. The four seasons are distinct, and precipitation is abundant. There is substantial precipitation in the summer, which is impacted by the southeastern monsoon, and there is little snowfall in the winter [53].

   The study area is divided into a densely inhabited district (DID) and surrounding area (SA), as shown in Figure 3.2a. Figure 3.2b shows the Google Earth imagery of the study area. DID is a statistically defined area set in the Japan National Census in which the population density is over 4,000 people/km^2, and the sum of the adjacent unit area population is more than 5,000. In 2010, the population of the Kanto MMA DID was about 1.11 million, with a total population of 2.97 million. Each LCZ was represented in the study area.

(a)
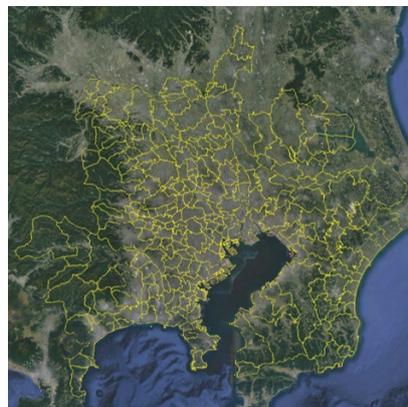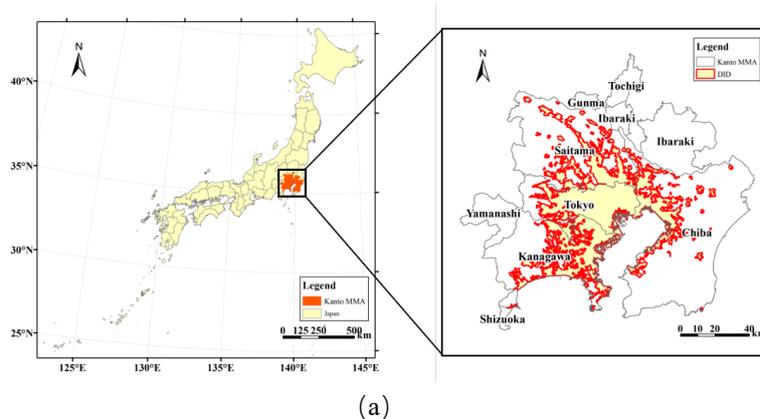
(b)

Figure 3.2. (a) Location of the study area in Japan. (b) Google Earth imagery of the study area.

Remote sensing data are greatly influenced by clouds and weather. Accordingly, a sunny day without any cloud cover was chosen: November 5, 2019. Multiple remote sensing images needed to be merged given the size of the study area. Table 3-1 shows the research data and some basic information, including the Landsat scene identifier, The worldwide reference system (WRS) path and row, date, start and stop time, land and scene cloud cover, and sun elevation [54].

Table 3-1. Basic information of study area remote sensing data

Location	Kanto Major Metropolitan Area
Landsat Scene Identifier	LC81070352019309LGN00/LC81070362019309LGN00
WRS Path, WRS Row	107, 035/107, 036
Date	2019-11-05
Start Time	10:15:56/10:16:20
Stop Time	10:16:28/10:16:52
Land Cloud Cover	0.64/2.06
Scene Cloud Cover	0.49/6.85
Sun Elevation	36.06°/37.31°

3.2.2.   Urban structure classification

Urban structures are mainly classified by manual sampling methods, classification methods based on geographic information systems (GIS) and classification method based on remote sensing. The manual sampling method, which requires field measurements, is the most basic method. However, it is time-consuming, laborious, and unsuitable for large-scale urban structure classification. A classification method based on GIS is commonly used currently (e.g., Lelovics et al. (Hungary) [55] and Zheng et al. (China) [56]). This method can achieve high accuracy using high-precision GIS data for urban planning and architectural form, but it is difficult to obtain urban GIS data in some regions which limits the application of this method. The classification method based on remote sensing is automatically divided into different types of LCZs by analyzing the spatial and spectra remote sensing information. Gamba et al. [57], for example, developed an LCZ classification method based on high-resolution remote sensing data and applied it to Heraklion, Greece. Although these approaches improve the classification accuracy, high-resolution remote sensing images also increase the

acquisition cost. Overall, combining the classification methods based on remote sensing and Landsat series images can improve identification accuracy and solve the shortcomings of high-resolution remote sensing's acquisition cost to classify urban structures based on the LCZ classification system.

Figure 3.3 shows the detailed steps of urban structure identification, including the four main steps. The first step is pre-processing Landsat 8 OLI imagery, in which the raw data consist of two series of Landsat 8 data points with various WRS Raw data. We corrected and merged these data to obtain a series of remote sensing images that can cover the entire study area as much as possible and then composited and extracted the imagery by research region. Second, we used ArcGIS 10.5 to select the training area in Google Earth imagery. Table 3-2 [58] below shows the training area of the Google Earth imagery and Landsat 8 OLI imagery of some of the typical LCZs. We selected 250 training areas for later urban structure identification. Among them, there are 10 training areas for each type of LCZ (from LCZ-1 1 to LCZ- 6) and 15 training areas (for LCZ-7 to LCZ-10 and LCZ-A to LCZ-G. The next step is to import selected training areas into remote sensing data to create signature file [59]. Then, urban structures are identified using MLC based on the signature files.
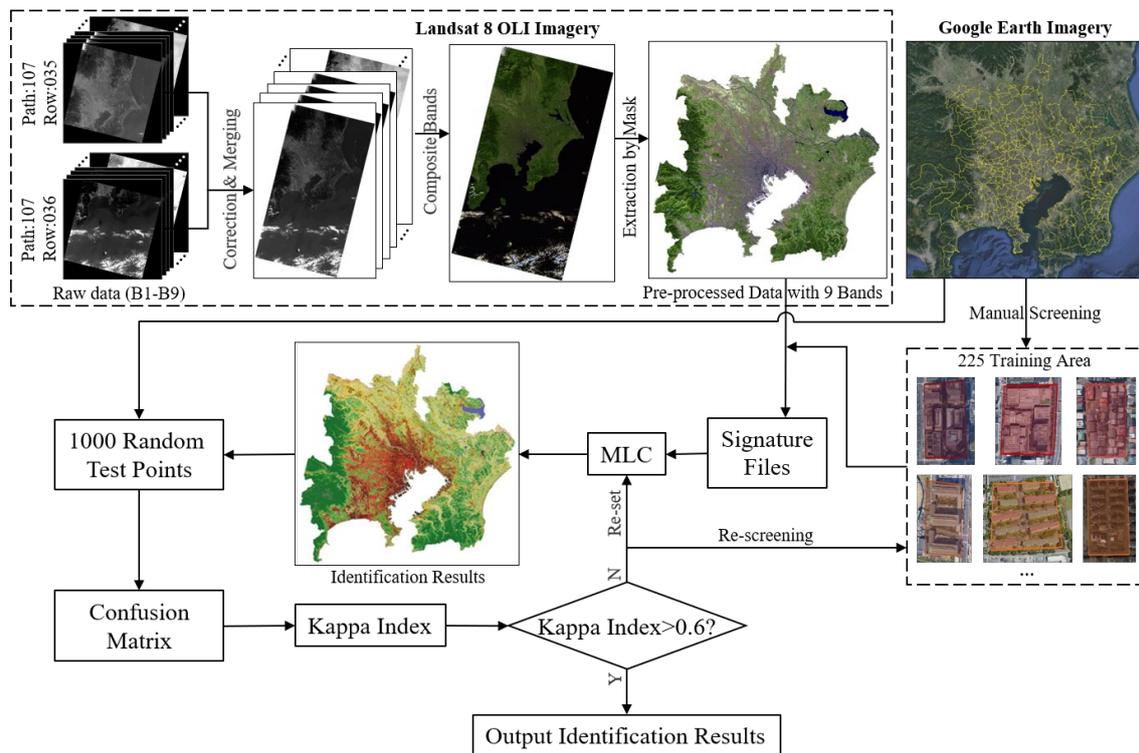
Figure 3.3. Detailed steps of urban structure identification

Table 3-2. The training area of Google Earth Imagery and Landsat 8 OLI/TIRS of some typical Local Climate Zones. (The data to create the sketch comes from)

LCZ ID	LCZ-1	LCZ-2	LCZ-3	LCZ-4	LCZ-5
Sketch					
Google Earth Imagery					
Landsat 8 OLI Imagery					
Street view					

LCZ ID	LCZ-A	LCZ-B	LCZ-C	LCZ-E	LCZ-F
Sketch					
Google Earth Imagery					
Landsat 8 OLI Imagery					
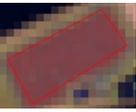
MLC assumes that all the distribution functions are normal distributions, selects training areas, and sets up a set of nonlinear discriminant functions to calculate the attribution probability of each area to be classified by statistical methods based on Bayes' theorem [60]. This classification method is a nonlinear classification, and its failure probability is the smallest; it is also a widely used and

sophisticated supervised classification method [61]. In Figure 3.4, according to Bayes' theorem, the posterior probability that Type x_i belongs to Area A or Area B is represented by the following equation:

$$P(XA) = P(A) \times P(X|A) = P(X) \times P(A|X)$$

(3-1)

where $P(A)$ represents the prior probability of Type a; $P(X|A)$ represents the prior conditional probability of Type x_i appearing in Area A (known); $P(X)$ represents the prior probability of Type x_i; and $P(A|X)$ represents the posterior probability of Type x_i belonging to Area A. $P(A)$, $P(X|A)$, and $P(X)$ were obtained from signature files created in the previous step.
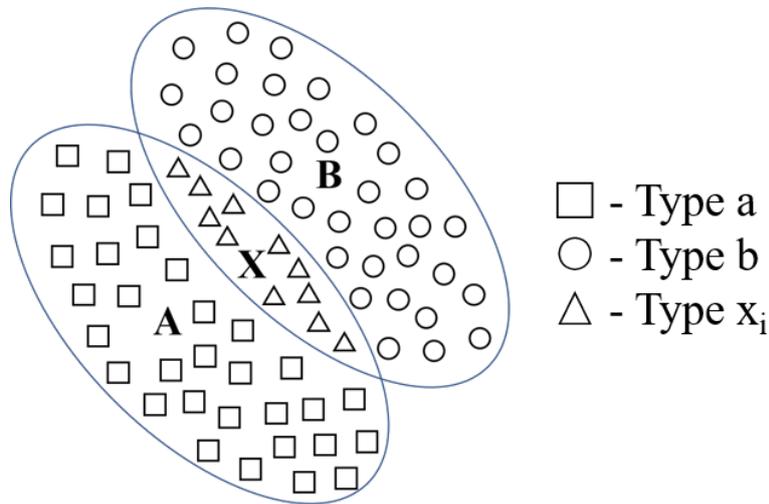
Figure 3-4. Two-dimensional space image classification (Area A represents the area consisting of Type a; Area B represents the area consisting of Type b; Area X represents the area consisting of Type x_i. Among them, Area A and B are known areas, Type x_i in Area X is the type need to be identified)

The posterior probability that Type x_i belongs to Area B, $P(B|X)$, could be calculated using the same method. When $P(A|X) > P(B|X)$, then Type x_i belongs to Area A; otherwise, Type x_i belongs to Area B. When the area type covers two types, the largest posterior probability is taken as the identification result.

Through the above three steps, we identified the urban structure. To validate the accuracy of the MLC, we randomly selected 1000 test points in the research region [62]. By combining them with the Google Earth imagery, we can determine whether the LCZs that are classified by MLC conform to reality.

This validation method is called a confusion matrix or error matrix. It is a standard format for accuracy evaluation, represented by a matrix with *k* rows and *k* columns [63], as shown below:

	A	B	...	K	n_{i+}
A	N_{11}	N_{12}	...	N_{1k}	N_{1+}
B	N_{21}	N_{22}	...	N_{2k}	N_{2+}
...	...	...	...	...	...
K	N_{k1}	N_{k2}	...	N_{kk}	N_{k+}
n_{+j}	N_{+1}	N_{+2}	...	N_{+k}	n

where *n* is the total number of samples in the remote sensing classification accuracy evaluation, and *k* is the total number of classified categories. N_{ij} is the number of samples that are classified as *i* in remote sensing and belong to *j* in the reference category. Row data indicate the LCZ type of test points in the identification results, and the column data are from the Google Earth images.

The specific evaluation index includes the producer's accuracy (PA), the user's accuracy (UA), overall accuracy, and Kappa index. Based on the confusion matrix, the PA and UA can be calculated using Equations (3-2) and (3-3):

$$PA_j = \frac{N_{jj}}{N_{+j}} = \frac{N_{jj}}{\sum N_{ij}} \ (j = 1 \sim k)$$

(3-2)

$$UA_i = \frac{N_{ii}}{N_{i+}} = \frac{N_{ii}}{\sum N_{ij}} \ (i = 1 \sim k)$$

(3-3)

According to PA_j and UA_i, the overall accuracy and Kappa index (k) can be obtained by Equations (3-4) and (3-5):

$$OA = \frac{\sum N_{ii}}{n} \ (i = 1 \sim k) = \frac{\sum N_{jj}}{n} \ (j = 1 \sim k)$$

(3-4)

$$k = \frac{p_0 - p_e}{1 - p_e}$$

(3-5)

where p_0 is the overall accuracy calculated by Equation (3-4), and p_e is calculated by the following equation:

$$p_e = \frac{\sum(N_{i+} \times N_{+j})}{n \times n}$$

(3-6)

Furthermore, the overall accuracy and Kappa index were obtained [64]. The result of Kappa is between -1 and 1, but Kappa usually falls between 0 and 1, and it can be divided into five groups to represent different levels of consistency: 0.0–0.20 is slight, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial, and 0.81–1.0 is almost perfect [65,66].

3.2.3.   Microclimate simulation

In the simulation six typical LCZs were selected to verify different local climates in different LCZs. This study selected ENVI-met as the simulation software, LCZ-1 to LCZ-5 as typical LCZs, and an additional model with mid-building height and mid-building density.

ENVI-met is suitable for mesoscale models, especially for microclimate simulations of urban block sizes. It includes four modules: modelling (SPACES), programming (ConfigWizard), simulation (Run), and results (LEONARDO) [67]. It is widely used in the analysis of urban microclimate and thermal environments [68,69].

For the present research, the selected LCZs were divided into two categories: compact buildings and open buildings. These two categories include high-rise buildings, mid-rise buildings, and low-rise buildings. To observe the climate differences of various LCZs, three qualifications were set: the same building density, building height, and plot ratio. Figure 3-5 shows a typical simulation model [70]. According to the demand, the grid size of the simulation space is determined as $100 \times 100 \times 35$, and the size of the grid is dx = 1, dy = 1, dz = 2 [71]. Table 3-3 shows the characteristics of the various models.
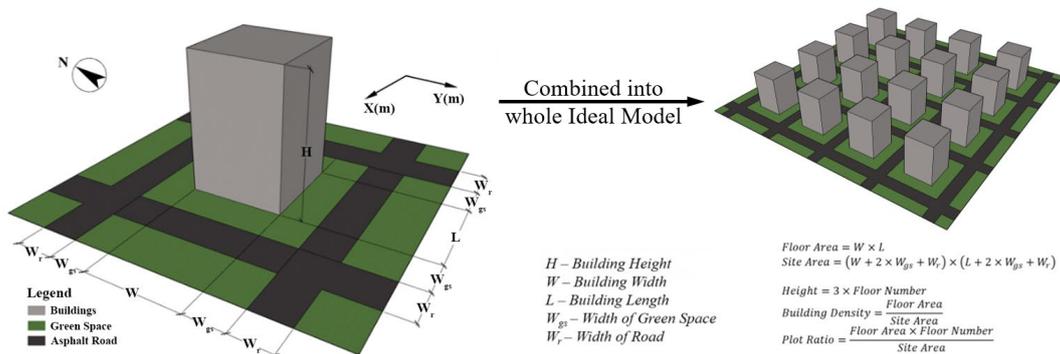
Figure 3-5. The idealized model of the urban structure and calculation method of several key parameters.

In Table 3-3, LCZ-1 represents a high-density high-rise model, LCZ-2 represents a high-density mid-rise model, LCZ-3 represents a high-density low-rise model, Mid represents a mid-density mid-rise model, LCZ-4 represents a low-density high-rise model, and LCZ-5 represents the low-density mid-rise model.

Table 3-3. Model characteristic of various LCZs. (Note: *same density, **same height, ***same plot ratio.)

LCZ ID	Floor area	Width of Green Space	Width of Road	Site area	Floor Number	Height	Building Density	Plot Ratio
LCZ-1	100	2	4	324	11	**33**	**0.31***	**3.40**
LCZ-2	100	2	4	324	7	**21****	**0.31***	**2.16**
LCZ-3	100	2	4	324	4	**12**	**0.31***	**1.22*** **
Mid	100	5	4	576	7	**21****	**0.17**	**1.22*** **
LCZ-4	100	8	4	900	11	**33**	**0.11**	**1.22*** **
LCZ-5	100	8	4	900	7	**21****	**0.11**	**0.78**

To better analyze and compare the microclimate changes in the simulation space under the same meteorological conditions, June 20th, was selected as the simulation day since June 21st is the summer solstice with the longest sunshine time and the largest solar zenith. The simulation time ran from 00:00 Local Time on June 20th to 00:00 Local Time on June 22nd, 48 hours in total (all times listed hereafter are in Local Time). The same meteorological data that were used for the initial data were input into the configuration files. Table 3-4 shows the input configuration used in the simulation for Tokyo. By combining the date, meteorological data, and location, the solar radiation details could be determined, including the solar elevation angle and building shadow. According to the historical meteorological data of the Japan Meteorological Agency [72], the average air temperature in June 2019 was approximately 21.8 °C, so the initial temperature of the atmosphere was set to 295 K. Due to software limitations, only the average wind speed and primary wind directions were selected for the initial wind velocity [73]; this has been recognized as a major limitation to this approach [48]. We set the initial wind speed to 5m/s and the wind direction southerly [72].

Table 3-4. Input configuration used in simulation with ENVI-met

Start Date (DD.MM)	20.06
Start Time (HH:MM:SS)	00:00:00
Total Simulation Time (h)	48
Wind Speed in 10m height (m/s)	5
Wind Direction (0=from North…180=from South)	180
Initial Temperature of Atmospheric (K)	295
Special Humidity at model top (g/kg)	7
Relative Humidity in 2m (%)	75

3.2.4. Heat stress assessment

The outputs of the simulation were air temperature, mean radiant temperature, solar radiation intensity, relative humidity, and wind speed simulation data during the study period at 1.5 m above the ground. In the first 24 h, the thermal interactions and physical processes that take place are entirely used to the environmental components are more balanced in the second 24 h. As a result, the simulation results are more accurate [74].

In outdoor thermal environment research, a comprehensive indicator is necessary for a complex and variable environmental situation, to evaluate the outdoor thermal environment. The WBGT is proposed as a simple, purely physical heat stress indicator [75]. Traditionally, WBGT has been obtained mainly from three indicators: wet-bulb temperature (T_w), black-globe temperature (T_g) and dry-bulb temperature (T_a), as shown in Equation (3-7) [76,77]:

$$T_{WBGT} = 0.7T_w + 0.2T_g + 0.1T_a \ ,$$

(3-7)

However, WBGT has the following deficiency in numerical simulation forecasts: T_w, T_g and T_a are not directly obtained by the simulation, but, the air temperature, relative humidity, wind speed, and solar radiation can be simulated directly.

Ono and Tonouchi [51] established a heat balance equation for the three parameters in the WBGT calculation formula based on the climate characteristics of Japan. By solving the heat balance equation, the WBGT calculation formula is obtained using air temperature, relative humidity, solar radiation intensity, and wind speed, which are the ENVI-met simulation results. After verification, the average estimation errors average is ~0.11°C, and the WBGT can be estimated using Equation (3-8):

$$T_{WBGT} = 0.735T_a + 0.0374RH + 0.00292 \times T_a \times RH + 7.619 \times 10^{-3}SR - 4.557 \times 10^{-3}SR^2$$
$$- 0.0572v - 0.4064$$

(3-8)

where T_a represents air temperature, *SR* represents solar radiation intensity, *RH* represents relative humidity, and *v* represents wind speed.

3.3. Results

3.3.1. LCZ classification results

The distribution of LCZs is shown in Figure 3.6a, which includes two parts: the DID (Figure 3.6b) and SA (Figure 3.6c). With the help of the pie chart and histogram in Figure 3.7, we can obtain the specific urban structure cluster in Kanto MMA. Figure 3.7a shows the percentage of each LCZ in the whole region (WR), DID, and SA. The innermost pie chart represents WR, the middle is the DID, and the outermost ring is the SA. The part of the chart with a black frame indicates the building type and the other island cover type. Figure 3.7b shows the proportion of each LCZ in the DID and SA.

(a)

(b)   (c)
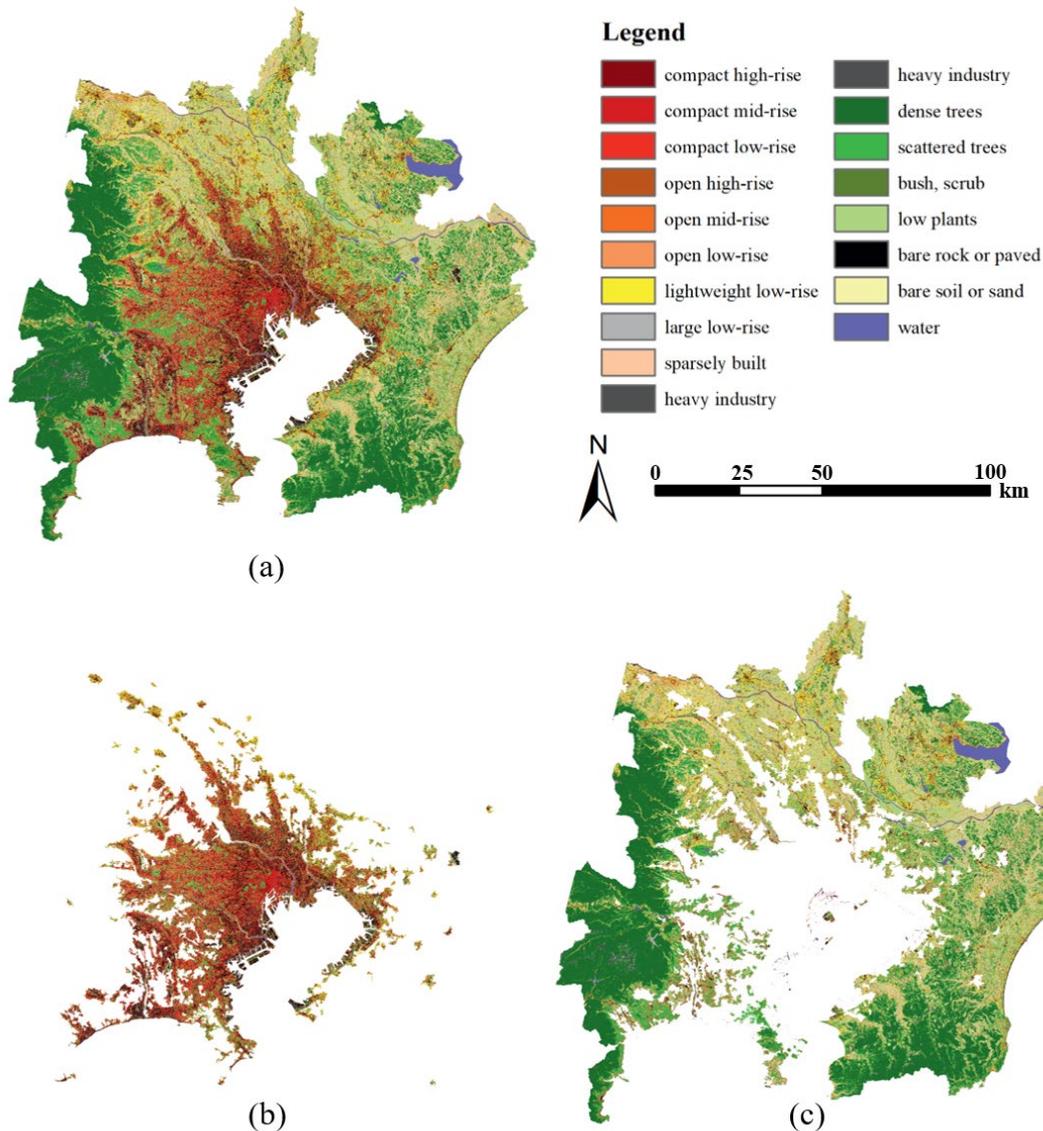
Figure 3.6. Distribution of Local Climate Zones: (a) Distribution of LCZs of the whole region (WR); (b) Distribution of LCZs of the densely inhabited district (DID); (c) Distribution of LCZs of the surrounding area (SA).

LCZ ID	WR	DID	SA
LCZ 1	0.40%	1.09%	0.16%
LCZ 2	2.25%	6.49%	0.82%
LCZ 3	5.79%	21.63%	0.47%
LCZ 4	2.62%	2.92%	2.53%
LCZ 5	2.53%	3.94%	2.06%
LCZ 6	4.55%	12.90%	1.75%
LCZ 7	8.75%	8.23%	8.92%
LCZ 8	0.64%	0.83%	0.57%
LCZ 9	9.85%	1.04%	12.81%
LCZ 10	3.69%	7.73%	2.33%
LCZ A	18.97%	0.02%	25.33%
LCZ B	14.72%	9.67%	16.42%
LCZ C	4.83%	2.43%	5.64%
LCZ D	11.44%	1.23%	14.87%
LCZ E	5.47%	17.95%	1.28%
LCZ F	2.22%	1.48%	2.47%
LCZ G	1.28%	0.45%	1.55%
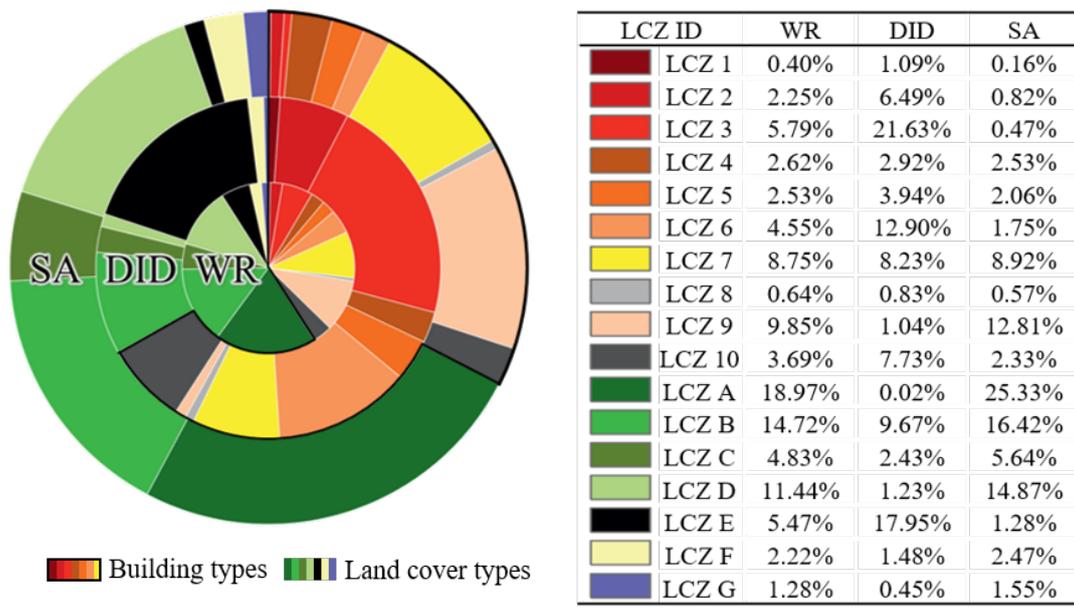
(a)
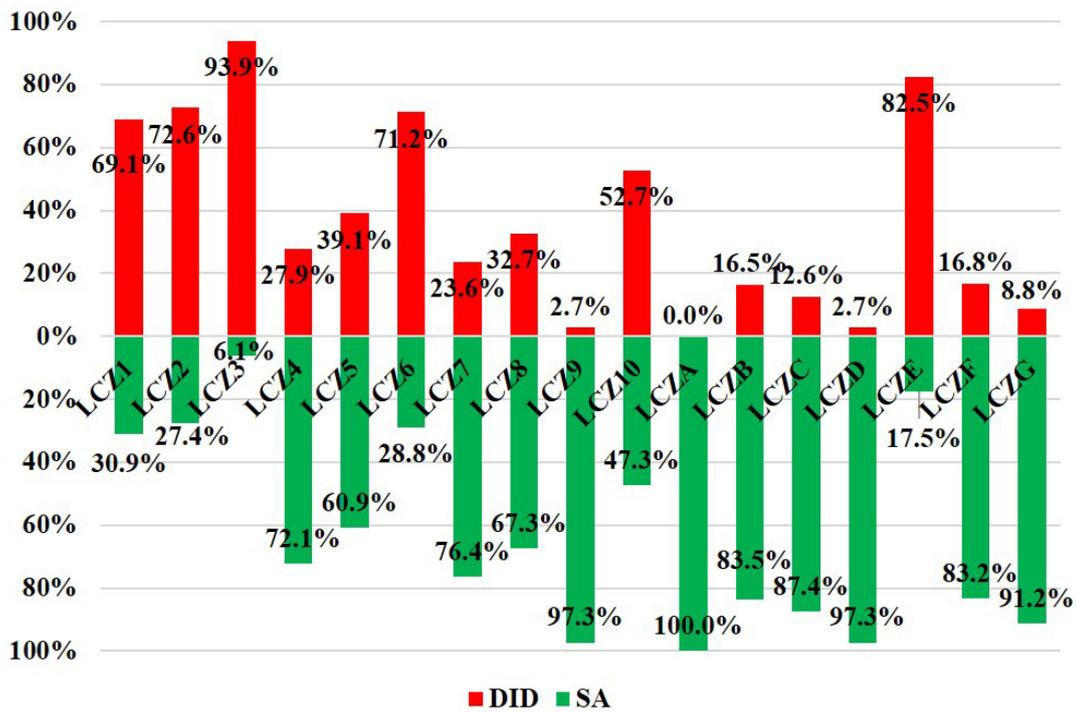
(b)

Figure 3.7. Graph showing the distribution of each LCZ in the WR, DID, and SA: (a)Pie chart showing the percentage of each LCZ in the WR, DID, and SA; (b) Histogram representing the proportion of each LCZ in WR, DID, and SA.

DIDs are distributed along the coast of Tokyo Bay and are mainly composed of building types (LCZ-1 to LCZ-10). There is a small component of LCZ-E (bare rock or paved) in the DID, which is largely in the surrounding area of the urban center and near the southern coast. From Central Tokyo to the suburban areas, the pixel color changes from dark red to orange, indicating that the building density and building height gradually drop. In the SA, the west side and southeast corners are mostly LCZ-A (dense trees) and LCZ-B (scattered trees), and the north side is covered with LCZ-D (low plants). There is a large area of LCZ-G (water) in the northeast corner.

Figure 3.7a shows that building types represent about 70% of the land cover, much more than that of other land cover types. Among them, LCZ-3 (compact low-rise) and LCZ-6 (open low-rise) are the top two building types, representing 21.63% and 12.90% of the buildings, respectively. LCZ-8 (large low-rise) has the smallest representation at only ~0.83%. Among the LCZ of the land cover types, LCZ-E was greatest. In contrast to the DID, the ratio of land cover types in the SA is about 68%, which is higher than that of the building types. The first is LCZ-A, the second is LCZ-B, and the third is LCZ-D. LCZ-9 (sparsely built) accounts for 12.81% of building types, which represents the majority. According to Figure 3.7b, LCZ-1, LCZ-2, LCZ-3, LCZ-6, and LCZ-E had larger proportions in the DID than they did in the SA. Conversely, the percentages of LCZ-4, LCZ-5, LCZ-7, LCZ-8, LCZ-9, LCZ-A, LCZ-B, LCZ-C, LCZ-D, LCZ-F, and LCZ-G in the SA were higher than those in the DID. The proportions of LCZ-5 and LCZ-10 in the DID and SA were similar.

The confusion matrix can be obtained mathematically. In Table 3-5, the bold data indicate that the selected test point has the same LCZ in both the Google Earth images and identification results, meaning that these test points were correctly identified. Among all the LCZs, the identification accuracy of LCZ-8 (large low-rise) is the lowest. A large number of LCZ-8 building types are identified as LCZ-10 (heavy industry) because there are quite a few factory buildings similar to LCZ-8 in the factories representing LCZ-10. After further calculation, the overall accuracy was ~80.2%, and the Kappa index reached 0.7798. This meets our precision requirements.

Table 3-5. Confusion matrix of identification results

LCZ	1	2	3	4	5	6	7	8	9	10	A	B	C	D	E	F	G	Sum	UA
1	**4**	1		1														6	66.7
2	2	**13**	1		2													18	72.2
3		2	**55**			4	1	1										63	87.3
4	2	1		**19**		2												24	79.2
5		3		1	**13**		1		1									19	68.4
6	1		2		1	**37**	4	1	1									47	78.7
7		2	1			6	**78**	3										90	86.7
8						1	1	**3**		3					1			9	33.3
9			1		3	5		3	**87**	2						2		103	84.5
10						1		11	1	**21**					1			35	60.0
A									3		**167**	13	3	1				187	89.3
B						1	3		2		11	**127**	6					150	84.7
C						1		3	1		4	7	**27**					43	62.8
D								2	5			3	10	**93**	1			114	81.6
E								2		3				5	**37**	12		59	62.7
F								1		3				1	3	**10**		18	55.6
G														3		1	**11**	15	73.3
Sum	9	22	60	21	19	58	88	30	101	32	182	150	46	103	43	25	11	1000	
PA	44.4	59.1	91.7	90.5	68.4	63.8	88.6	10.0	86.1	65.6	91.8	84.7	58.7	90.3	86.0	40.0	100		

Overall Accuracy/%: 80.20   **Kappa index: 0.7798**

In order to study the relationship between the air temperature of different LCZs in the DID and SA, we retrieved the land surface temperature on November 15, 2019, whose data were the same as those in Table 3-1. The land surface temperature in June is shown in Figure 3.8, and the statistical data of each LCZ in the DID and SA are shown in Figure 3.9. Thus, it can be seen that, for the same building type, the air temperature varies considerably with location. The value of each building type in the DID is generally higher than that in the SA. The difference in LCZ-10 was the largest, and the differences among LCZ-2, LCZ-3 and LCZ-9 were relatively small. Combining the temperature difference data from LCZ-1 to LCZ-6, we found that, as the building height decreases, the value rises significantly, but it is minimally affected by building density changes.

(a)   (b)   (c)
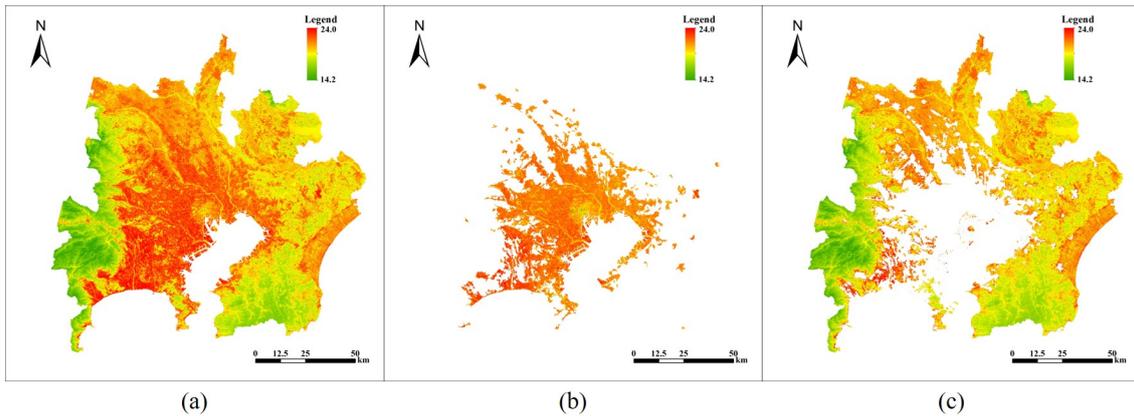
Figure 3.8. Land surface temperature of study area on 15th November 2019: (a) Air temperature of the WR; (b) Air temperature of the DID; (c) Air temperature of the SA.
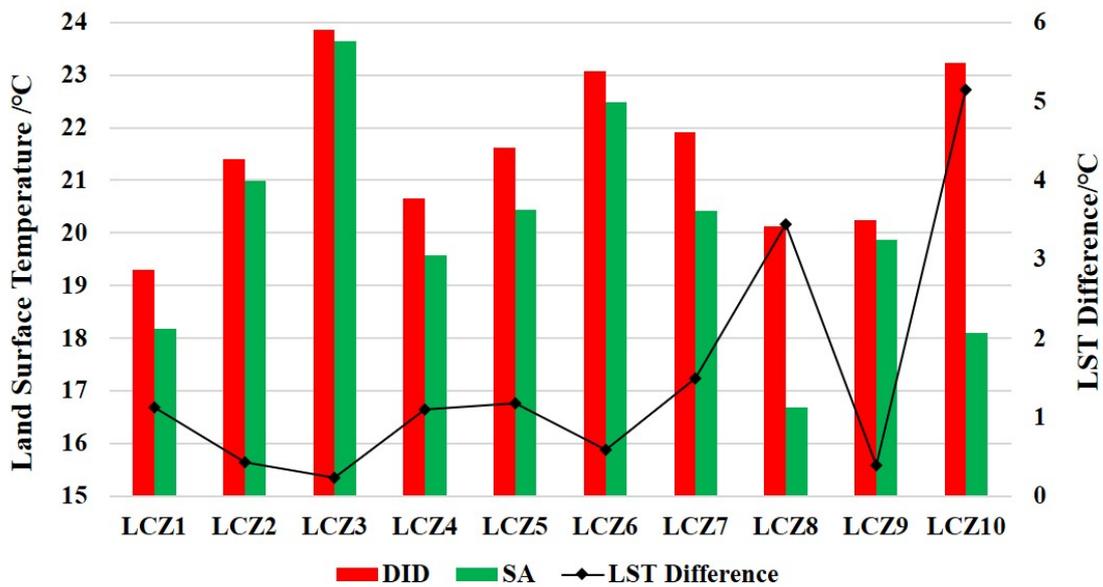
Figure 3.9. Land surface temperature of each LCZ in DID and SA.

3.3.2.   Simulation results

The WBGT clearly reflects the local climate of each LCZ. Figure 3.10 shows the WBGT changes in the open space between buildings of each idealized model for one day. This figure is divided into three parts, where the WBGT changes under the same density, height, and plot ratio. Daytime is active from 05:00 to 19:00, while nighttime occurs between 19:00 and 05:00. The WBGT is very different between the daytime and nighttime, exhibiting a Kuznets curve in the daytime and decreasing linearly at night. The WBGT reaches its highest value around 15:00 and approaches its lowest value around 05:00. The WBGT difference among the various models during the daytime is significantly greater than that at night.

LCZ ID	Density	Height	Plot Ratio
LCZ1	0.31[*]	33	3.40
LCZ2	0.31[*]	21	2.16
LCZ3	0.31[*]	12	1.23

LCZ ID	Density	Height	Plot Ratio
LCZ2	0.31	21[**]	2.16
Mid	0.17	21[**]	1.22
LCZ5	0.11	21[**]	0.78

LCZ ID	Density	Height	Plot Ratio
LCZ3	0.31	12	1.22[**]
Mid	0.17	21	1.22[**]
LCZ4	0.11	33	1.22[**]
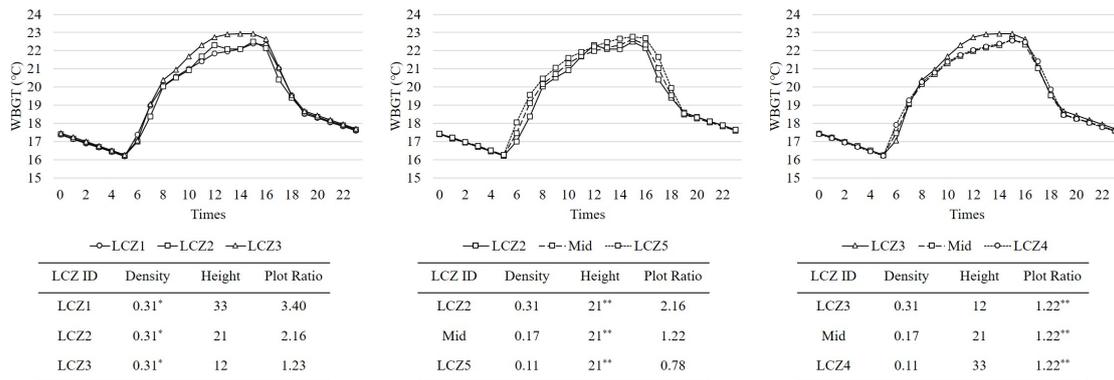
Figure 3.10. WBGT changes of each idealized model during one day: (a)WBGT changes with different height and plot ratios when the density is the same; (b) WBGT changes with different density and plot ratios when the height is the same; (c) WBGT changes with different density and height when the plot ratio is the same;

From three different constraints, this research compared the WBGT changes. The left figure shows the WBGT when the density is the same, as the height and plot ratio increasing. Overall, the WBGT of LCZ-3 was the highest, and that of LCZ-1 was the lowest. This phenomenon is particularly pronounced during the daytime, especially at 12:00; however, at other times, the WBGT values of LCZ-1 and LCZ-2 are mostly the same. The situation at the same height as shown in the middle figure of Figure 3.10. The WBGT decreases as the building height increases, and the WBGT change curve is mostly consistent when the density is the same. At 12:00, the values of Mid and LCZ-2 are almost the same and are lower than those of LCZ-5. When the plot ratio is the same, the WBGT difference is smaller than in the other two cases. Similarly, the WBGT of Mid was only slightly smaller than that of LCZ-4 (basically the same) and much larger than that of LCZ-3. The WBGT maximum difference occurred at 15:00 under the same plot ratio, height, and density, this occurs at 12:00 under the same density.

Four typical time points were selected—06:00, 12:00, 18:00, and 24:00—to analyze the WBGT spatial distribution. The noon WBGT results are shown in Figure 3.11, the midnight results are shown in Figure 3.13, and those at 06:00 and 18:00 are shown in Figures 3.15. The WBGT spatial distributions are also shown in Figures 3.11 (12:00), 3.13 (24:00), and 3.16 (06:00 and 18:00). The color changes from blue to yellow as the WBGT rises, as shown in the legend. The three columns represent the distribution under the same density, same height, and same plot ratio. The value along the x-axis increases in the east direction, and that along with the y-axis increases in the north direction. Figures 3.12, 3.14, and 3.16 represent the maximum, minimum, and average WBGT at different times.

As Figure 3.11 shows, the peaks and valleys of the WBGT are all surrounding areas of the buildings, with the peaks to the south of the buildings, the valleys mainly on the northern sides, and some peaks in the northeastern direction. The largest WBGT difference is between two adjacent buildings in the

north–south direction because of the peaks to the north of buildings and the valleys to the south of buildings. In comparison, the change in the WBGT is relatively gentle between two buildings in the east–west direction. Under the same density, the mid-rise models differ from the high-rise and low-rise models. There is a particularly high peak in the mid-rise models, and the peaks of the high-rise and low-rise models are more balanced. When comparing the peaks, the valleys of an idealized model are the same. When the height is the same, as the density decreases, the WBGT is slower, and the peaks are more average. The higher the height is and the lower the density is, the lower the WBGT peak is under the same plot ratio.
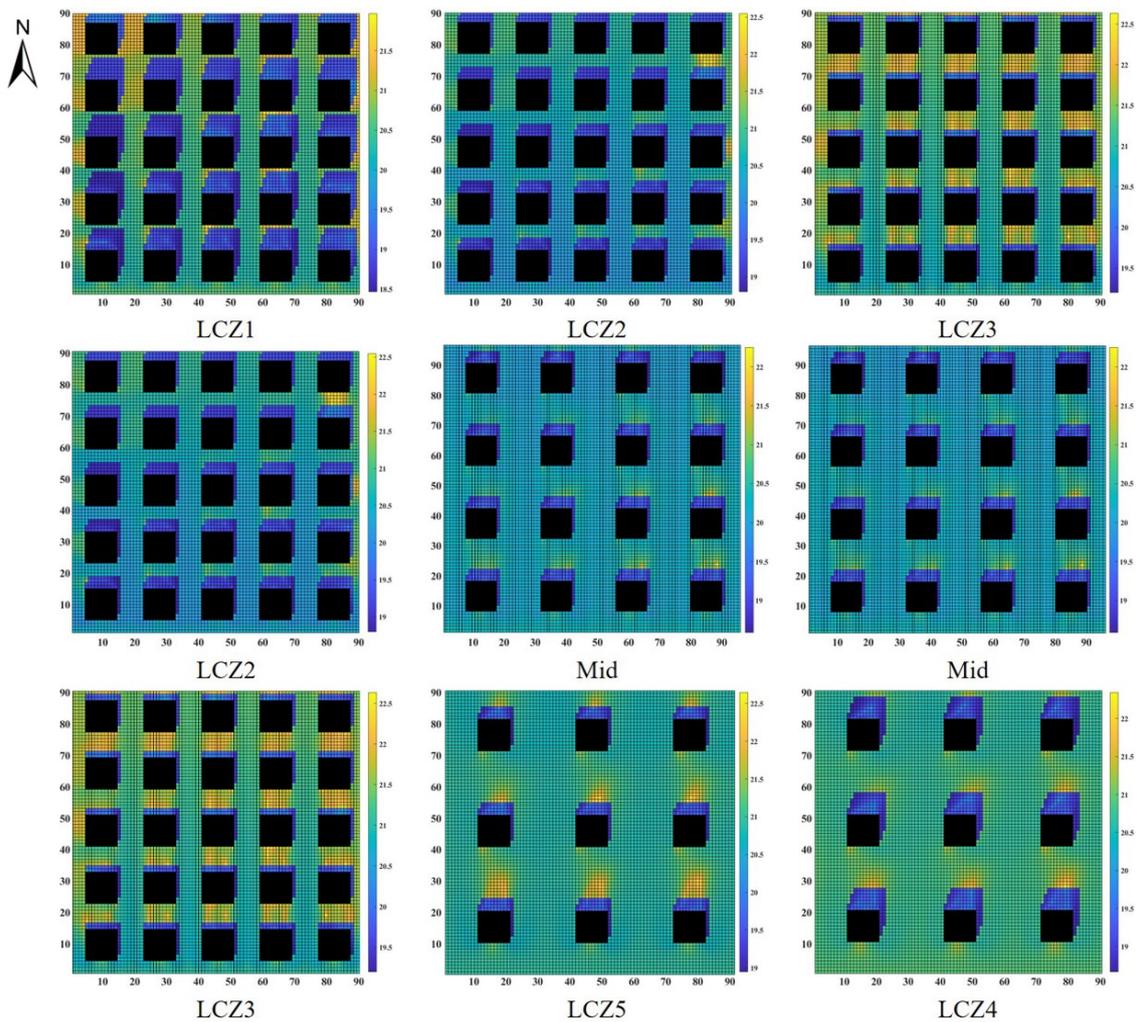
Figure 3.11. WBGT spatial distribution at 12:00 of each idealized model (black blocks indicate buildings).

According to Figure 3.12, under the three model conditions, the trends of the maximum, minimum, and average are similar. When the building density is the same, as the density increases, the WBGT

gradually increases. In the case of the same building height, the changing trend of the average value is similar to that of the previous case, but the maximum and minimum values are the minima of the Mid model. When the floor area ratio does not change, the Mid model has a significant decrease compared with the LCZ-3 model, but the Mid and LCZ-4 models are the same.
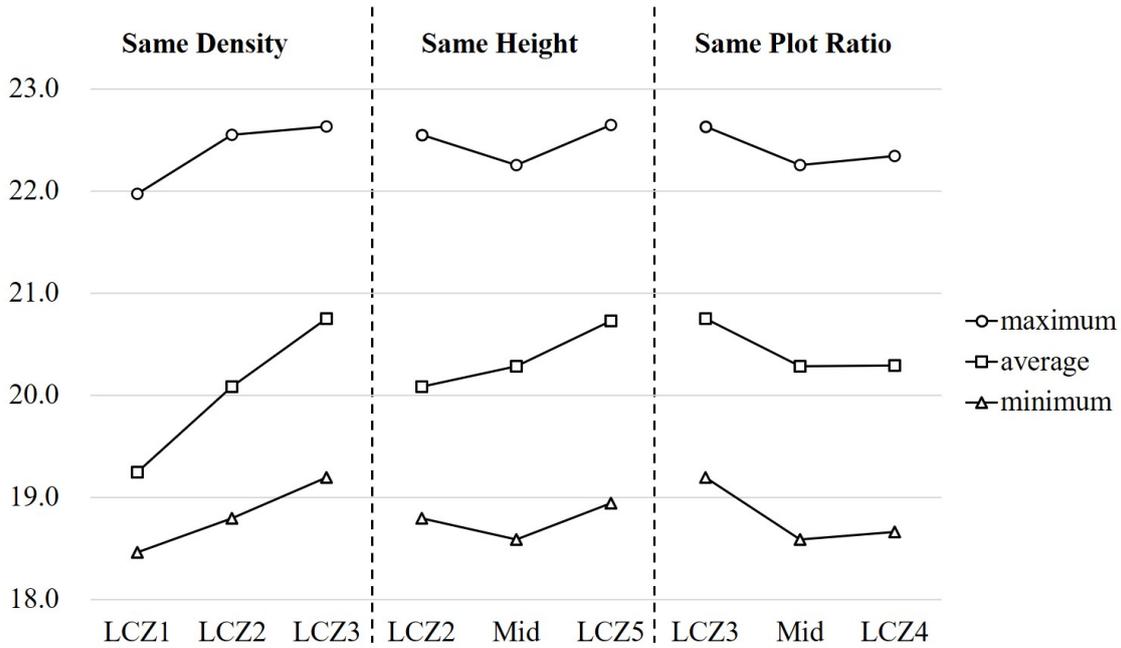
Figure 3.12. The maximum, minimum and average of WBGT at 12:00 of each idealized model

Figure 3.13 shows the spatial distribution at 24:00. At night, the WBGT has little data, which means that the WBGT is generally low and there are almost no valleys. In contrast, the peaks are obvious, and they are all north of the buildings. When the building density is the same, there is a prominent peak in LCZ-2 that is similar to that during daytime. The peaks in LCZ-1 were higher than those in LCZ-3; at the same height, the peaks of LCZ-2 are higher than those of Mid and LCZ-5. The single peak area of LCZ-5 is higher than the Mid peak. In the last case, the WBGT distribution around each building was similar between Mid and LCZ-4.
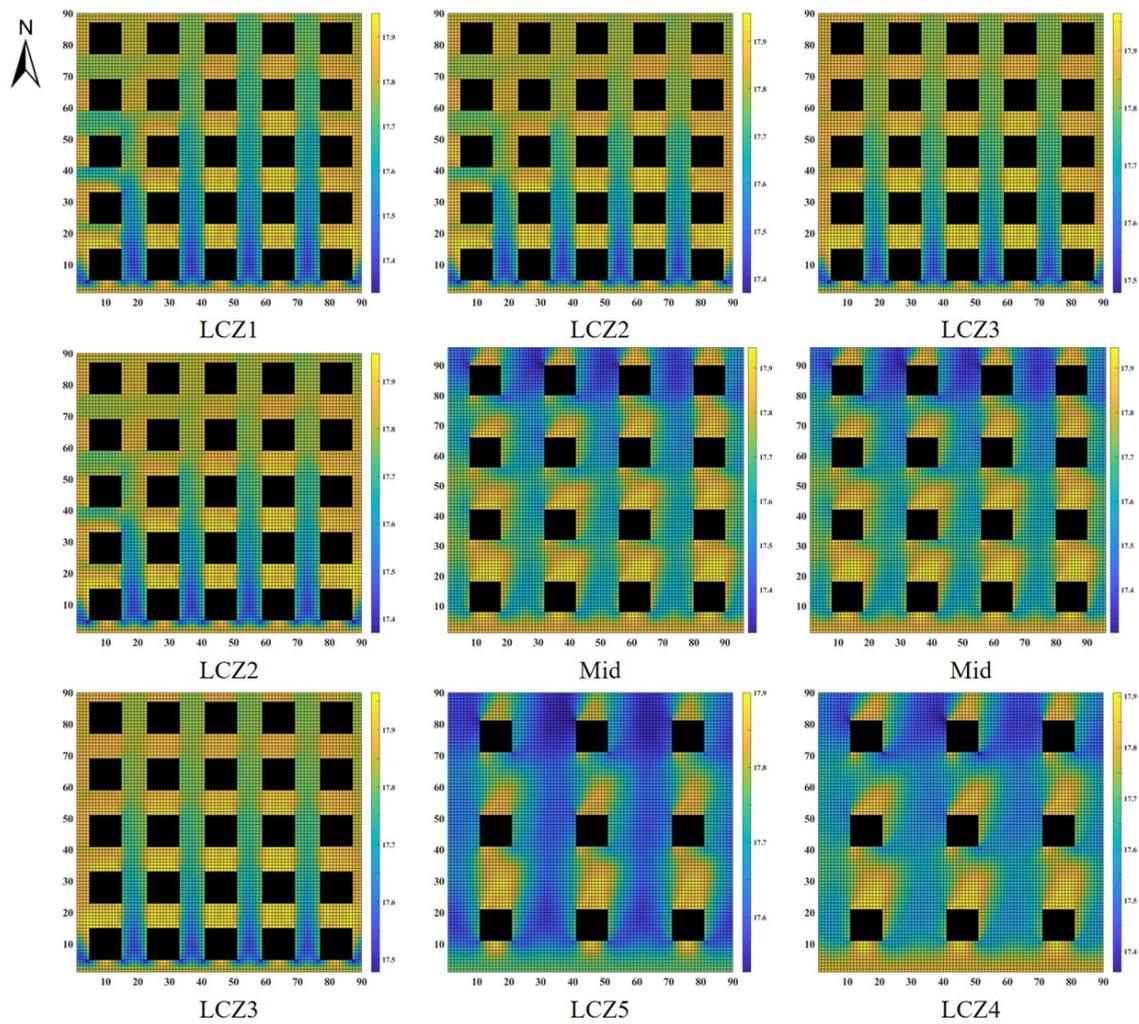
Figure 3.13. WBGT spatial distribution at 24:00 of each idealized model (black blocks indicate buildings).

Unlike the daytime, the average WBGT at night is generally closer to the maximum, as shown in Figure 3.14. The minimum of each model is different, and the maximum is almost similar. In particular, when the building height and plot ratio are the same, the minimum value of the Mid model is significantly smaller than that of the other two models. In both cases, WBGT is inversely proportional to the building density. When the building density is the same, WBGT decreases as the building height increases.
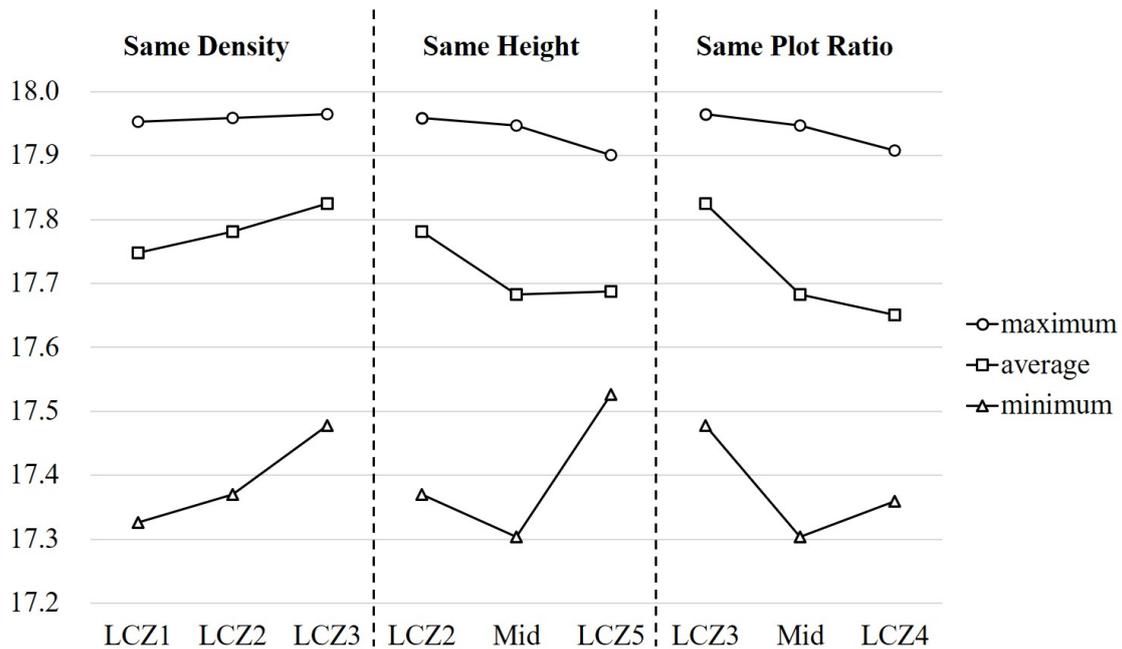
Figure 3.14. The maximum, minimum and average of WBGT at 24:00 of each idealized model

Comparing the WBGT distribution during the day and at night, there are more data available during the daytime than at night-time. The peak position was reversed during the day and night. It is mostly concentrated on the south side of the buildings in the daytime but on the north side at night. The peak height is also larger at night. The valley's position during the day is on the north side, but the valleys at night are difficult to discern. According to the two tables above, it is obvious that the WBGT maximum, minimum, and average at night are much lower than they are during the day. Moreover, the difference at night is also much less. Under the three conditions, the changes in the WBGT were the same during the day and night.

Compared to the results at 12:00 and 24:00, the WBGT spatial distribution at 06:00 and 18:00 (shown in Figure 3.15) are significantly different. In contrast to the above two cases at specific times, the distribution in the early morning and at nightfall depends on the building shadow due to sunrise and sunset. The difference between the WBGT in the shadows and in the sunlight is not obvious. Comparisons of the maximum, minimum, and average WBGT between 06:00 and 18:00 (shown in Figure 3.16), demonstrate that the WBGT in the afternoon was significantly higher than that in the morning. The change curve at the two time points is basically the same when the building density, height, and plot ratio are the same, especially for the average WBGT. There will be some places with lower WBGTs in Mid and LCZ-4.

(a)

(b)
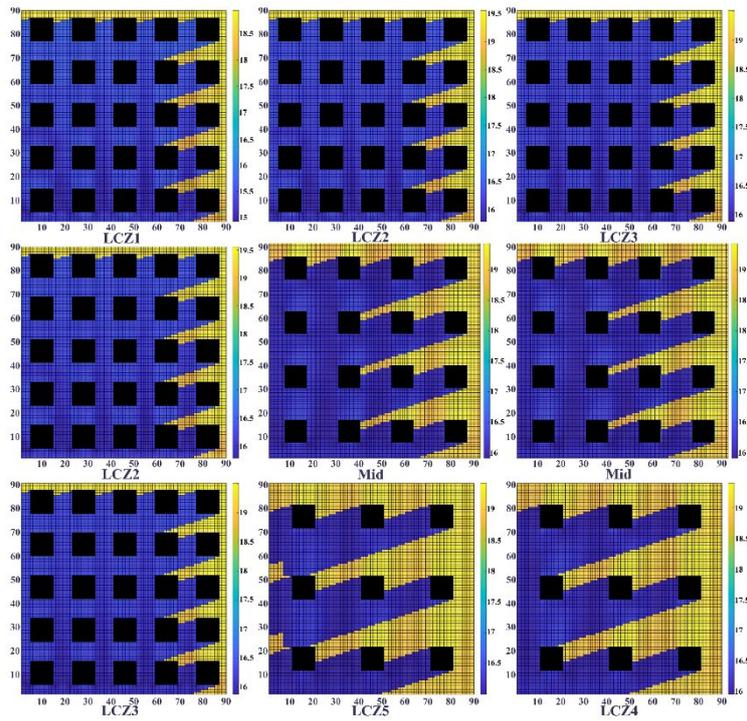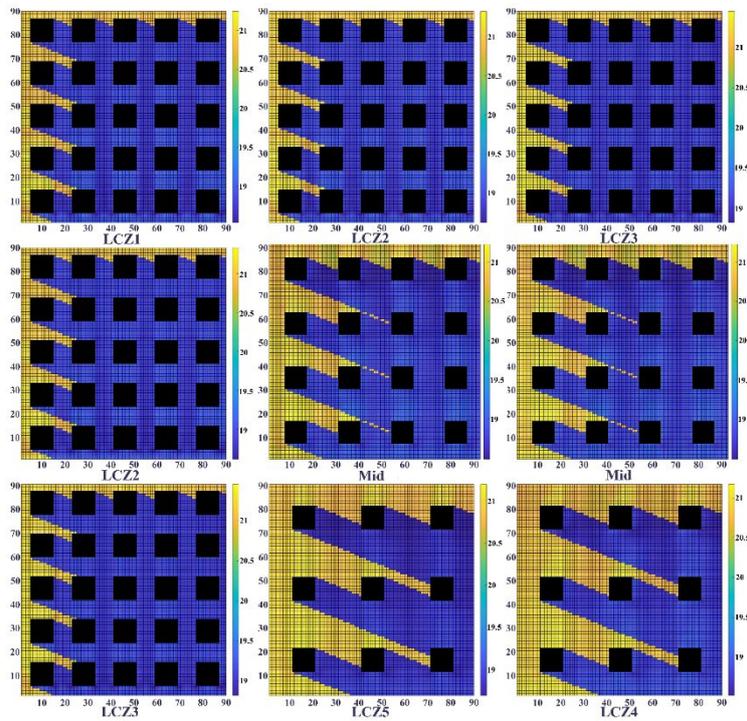
Figure 3.15. WBGT spatial distribution at 06:00 (a) and 18:00 (b) of each idealized model (black blocks indicate buildings)
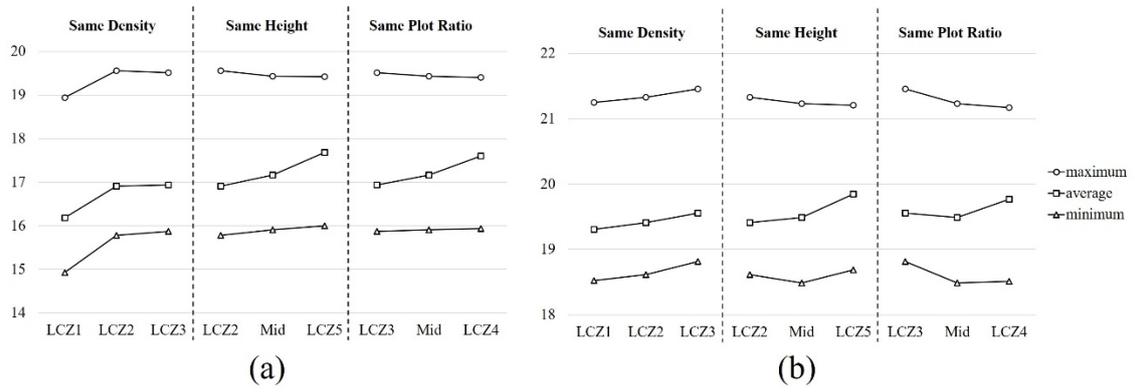
Figure 3.16. The maximum, minimum and average of WBGT at 06:00 (a) and 18:00 (b) of each idealized model

3.4. Discussion

A Each type of LCZ has corresponding microclimate characteristics through the ENVI-met simulation. WBGT was introduced to evaluate urban outdoor heat stress; the calculation results demonstrate that there is a significant relationship between the WBGT and LCZs throughout the day. This means that the proposed LCZ classification system could be used to identify urban structures, and urban structures can affect urban heat stress.

3.4.1. Application of remote sensing in urban structure classification

This classification system divided the urban structures into more detailed categories based on local characteristics [78]. The MLC method used in this research is a reliable and fully developed method, according to previous research [79]. This research combined the classification system with a method to identify urban structures, and a confusion matrix was selected to validate the accuracy of the results. In the case study, which looked at urban structures in the Kanto MMA, Japan, the land use and land cover in the study region were identified. The identification results show that the building types in the DID of the Kanto MMA are significantly larger than the land cover types, and, in the SA, the results are opposite, with the proportion of land cover types being higher than that of building types.

In the DID, compact low-rise buildings constitute the most prevalent type and are located around the urban area. Large low-rise buildings are the least prevalent. In the urban surrounding area not far from the urban center, there are also several compact high-rise cluster areas. This fits the characteristics of a "satellite city" in the five urban expansion models proposed by Camagni et al. [80] Seven auxiliary city centers are distributed around the urban center as satellite cities [81]. In the SA, vegetation occupies the most area, including dense trees, scattered trees, and low-lying plants. Agricultural land is the most important component of low-lying plants and is widely distributed in the northern part of the Kanto MMA. Sparsely built-in agricultural land represents a large proportion of building types. Combined with the air temperature of the research region, the same building type has different urban microclimate in the DID and SA. The air temperature on the southwest side of the study area, consisting of dense trees and scattered trees, is significantly lower than elsewhere. The SA, in which vegetation accounts for a large proportion of land cover types is obviously cooler than the DID, which means that the vegetation has a significant impact on air temperature.

This research only identifies urban structures in a qualitative manner, however, not a quantitative one. For example, in the process of selecting the training area, this research was based only on our subjective perspective, especially with respect to the LCZ types. In our future work, each LCZ will have its own characteristics, such as a specific sky view factor, aspect ratio, building surface fraction, impervious surface fraction, pervious surface fraction, building height, and terrain roughness class. According to these indexes, the training area will be selected more scientifically, which will produce results closer to reality.

3.4.2.   The influence of urban structure on urban thermal stress

There is a connection between urban heat stress and urban structures [82]. In this research, urban microclimate simulations were performed based on idealized models under specific meteorological conditions and dates, without considering human activities. By calculating and analyzing WBGT, this study evaluates the urban heat stress. Evaluations of heat stress need to consider multiple factors, including air temperature, humidity, wind speed and solar radiation [83]. During the day, building shadows can effectively reduce the WBGT, resulting in WBGT valleys on the northern side of buildings. Considering the effects of wind velocity, according to a study on the wind environment of cubic buildings [84,85], the wind speed in the north-south direction is much higher than that in the east–west direction. This causes the WBGT on the south side of buildings to be significantly higher than on the east and west sides. Similarly, as building height increases, the shadow area increases. When the height and density rise, the wind speed increases. At night, without solar radiation, wind becomes a major factor affecting the WBGT. The wind environment, such as wind speed and direction, is similar to that during the daytime, so this situation leads to a WBGT peak on the north side of buildings.

In the microclimate simulation, owing to the limitations of the simulation model, it is difficult to simulate the impact of human social activities, such as vehicles, air conditioning, and population movement, on air temperature. It is also arduous to simulate the microclimate of land cover LCZs. At the same time, because only a few idealized models were established for the simulation, we did not evaluate the urban microclimate in all cases, such as for different building layouts and combinations. In the future, we will build more varieties of idealized models to simulate and analyze the possible variations of future urban microclimates.

With the rapid development of remote sensing, more advanced remote sensing images provide more powerful data support for the LCZ classification system. This research combined the new classification system and ready-made developed methods to identify the urban microclimate structure of the Kanto MMA and validate the identification accuracy. Then, according to the land surface temperature retrieved with the same remote sensing data, we found that the temperatures of the same building types in the DID and SA are also quite different. ENVI-met was used in this research to simulate microclimates of some idealized LCZ models to verify that various urban structures have different impacts on urban heat stress.

3.4.3.   Limitations of the current research

We found that the LCZ classification system has some shortcomings in the research process. As a land use and land cover classification system, the "climate" component of the "local climate zone" is misleading. Through this study, we found that various urban structures will affect the urban microclimate and urban heat stress when all external conditions are the same; however, the urban

environment is also disturbed by many other factors, such as the location, landscape, and distance from the ocean [86,87]. This study confirms that location also has an impact on the local climate. In addition, the LCZ classification system cannot accurately divide the combined urban structure; for example, compact high-rise buildings can contain compact mid-rise or low-rise buildings.

Because idealized models were used to simulate the microclimate, it is difficult to verify the accuracy of the simulation through measured data. Previous studies have shown that ENVI-met is a robust and reliable simulation method, and the results were verified by measured data [88–92]. We plan to verify the accuracy of ENVI-met in further studies of urban structure and urban microclimate.

3.5. Summary

This research analyzed the feasibility of a new urban structure classification system, the LCZ classification system, and combined it with the ENVI-met simulation model to confirm the implications of urban structure on heat stress. Taking the Kanto MMA in Japan, as a case study, we used the LCZ classification system to identify urban structures based on Google Earth imagery and Landsat 8 OLI with the help of the MLC method. Through accuracy validation, we confirmed that the MLC method can identify each LCZ with high accuracy. Combined with the temperature distribution in the research region, we found that the same LCZ in the DID and SA differed substantially. In addition, the ENVI-met simulation model can be used to simulate urban microclimates. This study also used WBGT to evaluate and verify the impact of various urban structures on urban heat stress.

Based on the reflectance differences in each land use and land cover type for various bands with different wavelengths, urban structures were identified with the MLC method, according to the LCZ classification system, and the identification accuracy met the precision requirements for the process. The same urban structure types in the DID and SA have different microclimates. By combining the ENVI-met simulation tool with the WBGT, this study evaluated the urban microclimate. Under idealized conditions (including idealized simulation models and specific weather conditions), without considering the impact of human activities, it is clear that urban structures have a substantial impact on urban heat stress. When high-rise buildings have a moderate building density, it is beneficial to improve the comfort and safety of the microclimate for human beings. Under the same plot ratio, the temperature density and height or high-rise/mid-density can effectively reduce the WBGT. In summary, urban structures have a significant influence on urban heat stress.

Reference

1.   Grimm, N.B.; Faeth, S.H.; Golubiewski, N.E.; Redman, C.L.; Wu, J.; Bai, X.; Briggs, J.M. Global change and the ecology of cities. *Science (80-.).* 2008, *319*, 756–760.

2.   Xu, C.; Liu, M.; Hong, C.; Chi, T.; An, S.; Yang, X. Temporal variation of characteristic scales in urban landscapes: an insight into the evolving internal structures of China's two largest cities. *Landsc. Ecol.* **2012**, *27*, 1063–1074.

3.   Yang, X.; Yao, L.; Jin, T.; Peng, L.L.H.; Jiang, Z.; Hu, Z.; Ye, Y. Assessing the thermal behavior of different local climate zones in the Nanjing metropolis, China. *Build. Environ.* **2018**, *137*, 171–184.

4.   Mills, G. Cities as agents of global change. *Int. J. Climatol. A J. R. Meteorol. Soc.* **2007**, *27*, 1849–1857.

5.   Li, J.; Li, C.; Zhu, F.; Song, C.; Wu, J. Spatiotemporal pattern of urbanization in Shanghai, China between 1989 and 2005. *Landsc. Ecol.* **2013**, *28*, 1545–1565.

6.   JieGu Origins and review of urban time-space structure studies. *WORLD Reg. Stud.* **2016**, *25*, 69–79.

7.   Inostroza, L.; Baur, R.; Csaplovics, E. Urban sprawl and fragmentation in Latin America: A dynamic quantification and characterization of spatial patterns. *J. Environ. Manage.* **2013**, *115*, 87–97.

8.   Burgalassi, D.; Luzzati, T. Urban spatial structure and environmental emissions: A survey of the literature and some empirical evidence for Italian NUTS 3 regions. *Cities* **2015**, *49*, 134–148.

9.   Xie, Z.; Ye, X.; Zheng, Z.; Li, D.; Sun, L.; Li, R.; Benya, S. Modeling polycentric urbanization using multisource big geospatial data. *Remote Sens.* **2019**, *11*, 310.

10.   Liu, Z.; Liu, S. Polycentric development and the role of urban polycentric planning in China's mega cities: An examination of Beijing's metropolitan area. *Sustainability* **2018**, *10*, 1588.

11.   Banzhaf, E.; Hofer, R. Monitoring urban structure types as spatial indicators with CIR aerial photographs for a more effective urban environmental management. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2008**, *1*, 129–138.

12.   Stewart, I.D. A systematic review and scientific critique of methodology in modern urban heat island literature. *Int. J. Climatol.* **2011**, *31*, 200–217.

13.   Wilson, J.S.; Clay, M.; Martin, E.; Stuckey, D.; Vedder-Risch, K. Evaluating environmental influences of zoning in urban ecosystems with remote sensing. *Remote Sens. Environ.* **2003**, *86*, 303–321.

14.   Puissant, A.; Weber, C. The utility of very high spatial resolution images to identify urban objects. *Geocarto Int.* **2002**, *17*, 33–44.

15.   Herold, M.; Liu, X.; Clarke, K.C. Spatial metrics and image texture for mapping urban land use. *Photogramm. Eng. Remote Sens.* **2003**, *69*, 991–1001.

16. Montanges, A.P.; Moser, G.; Taubenböck, H.; Wurm, M.; Tuia, D. Classification of urban structural types with multisource data and structured models. In Proceedings of the 2015 Joint Urban Remote Sensing Event, JURSE 2015; Institute of Electrical and Electronics Engineers Inc., 2015.

17. Kalnay, E.; Cai, M. Impact of urbanization and land-use change on climate. *Nature* **2003**, *423*, 528–531, doi:10.1038/nature01675.

18. Pielke Sr, R.A.; Marland, G.; Betts, R.A.; Chase, T.N.; Eastman, J.L.; Niles, J.O.; Niyogi, D.D.S.; Running, S.W. The influence of land-use change and landscape dynamics on the climate system: relevance to climate-change policy beyond the radiative effect of greenhouse gases. *Philos. Trans. R. Soc. London. Ser. A Math. Phys. Eng. Sci.* **2002**, *360*, 1705–1719.

19. Hendel, M.; Azos-Diaz, K.; Tremeac, B. Behavioral adaptation to heat-related health risks in cities. *Energy Build.* **2017**, *152*, 823–829.

20. Roaf, C.; Crichton, D. Nicol (2006) ADAPTING BUILDINGS AND CITIES FOR CLIMATE CHANGE: A 21st century survival guide.

21. Mazhar, N.; Brown, R.D.; Kenny, N.; Lenzholzer, S. Thermal comfort of outdoor spaces in Lahore, Pakistan: Lessons for bioclimatic urban design in the context of global climate change. *Landsc. Urban Plan.* **2015**, *138*, 110–117.

22. Hajat, S.; Kovats, R.S.; Lachowycz, K. Heat-related and cold-related deaths in England and Wales: who is at risk? *Occup. Environ. Med.* **2007**, *64*, 93–100.

23. Arnfield, A.J.; Herbert, J.M.; Johnson, G.T. Urban canyon heat source and sink strength variations: A simulation-based sensitivity study. In Proceedings of the Sydney: Congress of Biometerology and International Conference on Urban Climate, WMO; 1999.

24. Santamouris, M.; Papanikolaou, N.; Livada, I.; Koronakis, I.; Georgakis, C.; Argiriou, A.; Assimakopoulos, D.N. On the impact of urban climate on the energy consumption of buildings. *Sol. energy* **2001**, *70*, 201–216.

25. Taha, H. Urban climates and heat islands: albedo, evapotranspiration, and anthropogenic heat. *Energy Build.* **1997**, *25*, 99–103.

26. Perini, K.; Magliocco, A. Effects of vegetation, urban density, building height, and atmospheric conditions on local temperatures and thermal comfort. *Urban For. Urban Green.* **2014**, *13*, 495–506.

27. Rizwan, A.M.; Dennis, L.Y.C.; Chunho, L.I.U. A review on the generation, determination and mitigation of Urban Heat Island. *J. Environ. Sci.* **2008**, *20*, 120–128.

28. Blocken, B.; Janssen, W.D.; van Hooff, T. CFD simulation for pedestrian wind comfort and wind safety in urban areas: General decision framework and case study for the Eindhoven University campus. *Environ. Model. Softw.* **2012**, *30*, 15–34.

29. Salata, F.; Golasi, I.; de Lieto Vollaro, R.; de Lieto Vollaro, A. Urban microclimate and outdoor

thermal comfort. A proper procedure to fit ENVI-met simulation outputs to experimental data. *Sustain. cities Soc.* **2016**, *26*, 318–343.

30.   Gál, C. V; Kántor, N. Modeling mean radiant temperature in outdoor spaces, A comparative numerical simulation and validation study. *Urban Clim.* **2020**, *32*, 100571.

31.   Sharmin, T.; Steemers, K.; Matzarakis, A. Microclimatic modelling in assessing the impact of urban geometry on urban thermal environment. *Sustain. cities Soc.* **2017**, *34*, 293–308.

32.   Arnfield, A.J. Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island. *Int. J. Climatol.* **2003**, *23*, 1–26, doi:10.1002/joc.859.

33.   Huang, J.; Cedeno-Laurent, J.G.; Spengler, J.D. CityComfort+: A simulation-based method for predicting mean radiant temperature in dense urban areas. *Build. Environ.* **2014**, *80*, 84–95.

34.   Huttner, S.; Bruse, M. Numerical modeling of the urban climate–a preview on ENVI-met 4.0. In Proceedings of the 7th international conference on urban climate ICUC-7, Yokohama, Japan; 2009; Vol. 29.

35.   Matzarakis, A.; Rutz, F.; Mayer, H. Modelling radiation fluxes in simple and complex environments: basics of the RayMan model. *Int. J. Biometeorol.* **2010**, *54*, 131–139.

36.   Musy, M.; Malys, L.; Morille, B.; Inard, C. The use of SOLENE-microclimat model to assess adaptation strategies at the district scale. *Urban Clim.* **2015**, *14*, 213–223.

37.   Krüger, E.L.; Minella, F.O.; Rasia, F. Impact of urban geometry on outdoor thermal comfort and air quality from field measurements in Curitiba, Brazil. *Build. Environ.* **2011**, *46*, 621–634.

38.   Ng, E.; Chen, L.; Wang, Y.; Yuan, C. A study on the cooling effects of greening in a high-density city: An experience from Hong Kong. *Build. Environ.* **2012**, *47*, 256–271.

39.   Teller, J.; Azar, S. Townscope II—A computer system to support solar access decision-making. *Sol. energy* **2001**, *70*, 187–200.

40.   Lindberg, F.; Holmer, B.; Thorsson, S. SOLWEIG 1.0–Modelling spatial variations of 3D radiant fluxes and mean radiant temperature in complex urban settings. *Int. J. Biometeorol.* **2008**, *52*, 697–713.

41.   Ketterer, C.; Matzarakis, A. Human-biometeorological assessment of heat stress reduction by replanning measures in Stuttgart, Germany. *Landsc. Urban Plan.* **2014**, *122*, 78–88.

42.   Ali-Toudert, F.; Mayer, H. Effects of asymmetry, galleries, overhanging facades and vegetation on thermal comfort in urban street canyons. *Sol. energy* **2007**, *81*, 742–754.

43.   Middel, A.; Häb, K.; Brazel, A.J.; Martin, C.A.; Guhathakurta, S. Impact of urban form and design on mid-afternoon microclimate in Phoenix Local Climate Zones. *Landsc. Urban Plan.* **2014**, *122*, 16–28.

44.   Johansson, E.; Spangenberg, J.; Gouvêa, M.L.; Freitas, E.D. Scale-integrated atmospheric simulations to assess thermal comfort in different urban tissues in the warm humid summer of São Paulo, Brazil. *Urban Clim.* **2013**, *6*, 24–43.

45.   Wang, Y.; Berardi, U.; Akbari, H. Comparing the effects of urban heat island mitigation strategies for Toronto, Canada. *Energy Build.* **2016**, *114*, 2–19.

46.   Perini, K.; Chokhachian, A.; Dong, S.; Auer, T. Modeling and simulating urban outdoor comfort: Coupling ENVI-Met and TRNSYS by grasshopper. *Energy Build.* **2017**, *152*, 373–384.

47.   Ketterer, C.; Matzarakis, A. Comparison of different methods for the assessment of the urban heat island in Stuttgart, Germany. *Int. J. Biometeorol.* **2015**, *59*, 1299–1309.

48.   Acero, J.A.; Herranz-Pascual, K. A comparison of thermal comfort conditions in four urban spaces by means of measurements and modelling techniques. *Build. Environ.* **2015**, *93*, 245–257.

49.   Taleb, H.; Taleb, D. Enhancing the thermal comfort on urban level in a desert area: Case study of Dubai, United Arab Emirates. *Urban For. Urban Green.* **2014**, *13*, 253–260.

50.   Mahgoub, A.O.; Gowid, S.; Ghani, S. Global evaluation of WBGT and SET indices for outdoor environments using thermal imaging and artificial neural networks. *Sustain. cities Soc.* **2020**, 102182.

51.   Ono, M.; Tonouchi, M. Estimation of wet-bulb globe temperature using generally measured meteorological indices. *Jpn J Biometeor* **2014**, *50*, 147–157.

52.   Nations, U. The world's cities in 2016. *New York, United* **2016**.

53.   Matsumoto, J.; Fujibe, F.; Takahashi, H. Urban climate in the Tokyo metropolitan area in Japan. *J. Environ. Sci. (China)* **2017**, *59*, 54–62, doi:10.1016/j.jes.2017.04.012.

54.   Survey, U.S.G. No Title Available online: https://earthexplorer.usgs.gov/.

55.   Lelovics, E.; Unger, J.; Gál, T.; Gál, C. V. Design of an urban monitoring network based on Local Climate Zone mapping and temperature pattern modelling. *Clim. Res.* **2014**, *60*, 51–62, doi:10.3354/cr01220.

56.   Zheng, Y.; Ren, C.; Xu, Y.; Wang, R.; Ho, J.; Lau, K.; Ng, E. GIS-based mapping of Local Climate Zone in the high-density city of Hong Kong. *Urban Clim.* **2018**, *24*, 419–448, doi:10.1016/j.uclim.2017.05.008.

57.   Gamba, P.; Lisini, G.; Liu, P.; Du, P.; Lin, H. Urban climate zone detection and discrimination using object-based analysis of VHR scenes. *Proc. 4th GEOBIA* **2012**, 70–74.

58.   Stewart, I.D.; Oke, T.R. Local climate zones for urban temperature studies. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 1879–1900.

59.   Ali, M.Z.; Qazi, W.; Aslam, N. A comparative study of ALOS-2 PALSAR and landsat-8 imagery for land cover classification using maximum likelihood classifier. *Egypt. J. Remote Sens. Sp. Sci.* **2018**, *21*, S29–S35, doi:10.1016/j.ejrs.2018.03.003.

60.   Strahler, A.H. The use of prior probabilities in maximum likelihood classification of remotely sensed data. *Remote Sens. Environ.* **1980**, *10*, 135–163, doi:10.1016/0034-4257(80)90011-5.

61.   Yilmaz, E.; Al-Rubaie, M.; Chang, J.M. Locally Differentially Private Naive Bayes Classification. *arXiv* **2019**.

62. Lin, Z.; Xu, H. A study of Urban heat island intensity based on "local climate zones": A case study in Fuzhou, China. In Proceedings of the 4th International Workshop on Earth Observation and Remote Sensing Applications, EORSA 2016 - Proceedings; Institute of Electrical and Electronics Engineers Inc., 2016; pp. 250–254.

63. Comber, A.; Fisher, P.; Brunsdon, C.; Khmag, A. Spatial analysis of remote sensing image classification accuracy. *Remote Sens. Environ.* **2012**, *127*, 237–246, doi:10.1016/j.rse.2012.09.005.

64. Gwet, K.L. Computing inter-rater reliability and its variance in the presence of high agreement. *Br. J. Math. Stat. Psychol.* **2008**, *61*, 29–48, doi:10.1348/000711006X126600.

65. Congalton, R.G.; Oderwald, R.G.; Mead, R.A. Assessing Landsat Classification Accuracy Using Discrete Multivariate Analysis Statistical Techniques. *Photogramm. Eng. Remote Sensing* **1983**, *49*, 1671–1678.

66. Canters, F. Evaluating the uncertainty of area estimates derived from fuuy land-cover classification. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 403–414.

67. ENVI-met Available online: https://www.envi-met.com/ (accessed on Apr 29, 2021).

68. Taleb, D.; Abu-Hijleh, B. Urban heat islands: Potential effect of organic and structured urban configurations on temperature variations in Dubai, UAE. *Renew. Energy* **2013**, *50*, 747–762, doi:10.1016/j.renene.2012.07.030.

69. Fahmy, M.; Sharples, S. On the development of an urban passive thermal comfort system in Cairo, Egypt. *Build. Environ.* **2009**, *44*, 1907–1916, doi:10.1016/j.buildenv.2009.01.010.

70. GAO, W. THERMAL EFFECTS OF OPEN SPACE WITH A GREEN AREA ON URBAN ENVIRONMENT : Part I:A theoretical analysis and its application. *J. Archit. Plan. Environ. Eng. (Transactions AIJ)* **1993**, *448*, 15–27, doi:10.3130/aijax.448.0_15.

71. Yin, S.; Lang, W.; Xiao, Y. The synergistic effect of street canyons and neighbourhood layout design on pedestrian-level thermal comfort in hot-humid area of China. *Sustain. Cities Soc.* **2019**, *49*, 101571, doi:10.1016/j.scs.2019.101571.

72. Japan Meteorological Agency Available online: http://www.jma.go.jp/jma/indexe.html (accessed on Jan 31, 2021).

73. Chan, S.Y.; Chau, C.K. On the study of the effects of microclimate and park and surrounding building configuration on thermal comfort in urban parks. *Sustain. Cities Soc.* **2021**, *64*, 102512, doi:10.1016/j.scs.2020.102512.

74. Deng, J.Y.; Wong, N.H.; Zheng, X. The Study of the Effects of Building Arrangement on Microclimate and Energy Demand of CBD in Nanjing, China. In Proceedings of the Procedia Engineering; Elsevier Ltd, 2016; Vol. 169, pp. 44–54.

75. Liang, D.; Qigao, C. Evaluation for Outdoor thermal environment quality. *Environ. Sci. Res. Vol8* **1995**.

76.   Burr, R.E. *Heat illness: a handbook for medical officers*; Army Research Inst of Environmental Medicine Natick MA, 1991;

77.   Zhu, N.; Chong, D. Evaluation and improvement of human heat tolerance in built environments: A review. *Sustain. Cities Soc.* 2019, *51*, 101797.

78.   Wang, R.; Cai, M.; Ren, C.; Bechtel, B.; Xu, Y.; Ng, E. Detecting multi-temporal land cover change and land surface temperature in Pearl River Delta by adopting local climate zone. *Urban Clim.* **2019**, *28*, 100455, doi:10.1016/j.uclim.2019.100455.

79.   Cabral, A.I.R.; Silva, S.; Silva, P.C.; Vanneschi, L.; Vasconcelos, M.J. Burned area estimations derived from Landsat ETM+ and OLI data: Comparing Genetic Programming with Maximum Likelihood and Classification and Regression Trees. *ISPRS J. Photogramm. Remote Sens.* 2018, *142*, 94–105.

80.   Camagni, R.; Gibelli, M.C.; Rigamonti, P. Urban mobility and urban form: The social and environmental costs of different patterns of urban expansion. *Ecol. Econ.* **2002**, *40*, 199–216, doi:10.1016/S0921-8009(01)00254-3.

81.   Liu, K.; Murayama, Y.; Ichinose, T. Using a new approach for revealing the spatiotemporal patterns of functional urban polycentricity: A case study in the Tokyo metropolitan area. *Sustain. Cities Soc.* **2020**, *59*, 102176, doi:10.1016/j.scs.2020.102176.

82.   Jamei, Y.; Rajagopalan, P.; Sun, Q. (Chayn) Spatial structure of surface urban heat island and its relationship with vegetation and built-up areas in Melbourne, Australia. *Sci. Total Environ.* **2019**, *659*, 1335–1351, doi:10.1016/j.scitotenv.2018.12.308.

83.   CHEN, R.; DONG, L. Study on Comfort Level Evaluation of Microclimate of Landscape Architecture in Humid-hot Climate Area [J]. *Build. Sci.* **2013**, *8*.

84.   Watson, D. Climatatic design: Energy efficient building principles and practices. **1983**.

85.   He, B.J.; Ding, L.; Prasad, D. Enhancing urban ventilation performance through the development of precinct ventilation zones: A case study based on the Greater Sydney, Australia. *Sustain. Cities Soc.* **2019**, *47*, 101472, doi:10.1016/j.scs.2019.101472.

86.   Zheng, B.; Myint, S.W.; Fan, C. Spatial configuration of anthropogenic land cover impacts on urban warming. *Landsc. Urban Plan.* **2014**, *130*, 104–111, doi:10.1016/j.landurbplan.2014.07.001.

87.   Estoque, R.C.; Murayama, Y.; Myint, S.W. Effects of landscape composition and pattern on land surface temperature: An urban heat island study in the megacities of Southeast Asia. *Sci. Total Environ.* **2017**, *577*, 349–359, doi:10.1016/j.scitotenv.2016.10.195.

88.   Chen, Y.; Wu, J.; Yu, K.; Wang, D. Evaluating the impact of the building density and height on the block surface temperature. *Build. Environ.* **2020**, *168*, 106493, doi:10.1016/j.buildenv.2019.106493.

89.   Liu, D.; Hu, S.; Liu, J. Contrasting the performance capabilities of urban radiation field between

three microclimate simulation tools. *Build. Environ.* **2020**, *175*, 106789, doi:10.1016/j.buildenv.2020.106789.

90. Forouzandeh, A. Numerical modeling validation for the microclimate thermal condition of semi-closed courtyard spaces between buildings. *Sustain. Cities Soc.* **2018**, *36*, 327–345, doi:10.1016/j.scs.2017.07.025.

91. Ma, X.; Fukuda, H.; Zhou, D.; Wang, M. Study on outdoor thermal comfort of the commercial pedestrian block in hot-summer and cold-winter region of southern China-a case study of The Taizhou Old Block. *Tour. Manag.* **2019**, *75*, 186–205, doi:10.1016/j.tourman.2019.05.005.

92. Tsitoura, M.; Tsoutsos, T.; Daras, T. Evaluation of comfort conditions in urban open spaces. Application in the island of Crete. *Energy Convers. Manag.* **2014**, *86*, 250–258, doi:10.1016/j.enconman.2014.04.059.

Chapter 4. Case Study and Precision Evaluation

4.1. Study Area

Kanto Major Metropolitan Area (shown in Figure 4.1) is the most populous metropolitan area around the world, which located at the Tokyo Bay. It mainly comprises the Tokyo, Kanagawa, Chiba, and Saitama, and some other cities of Ibaraki, Tochigi, Gunma, and Yamanashi. In light of the Statistics Bureau of Japan, the population was about 37.27 million [1]. It has an area of about 17,808 square kilometers and a population density of 2,093 people per square kilometers.
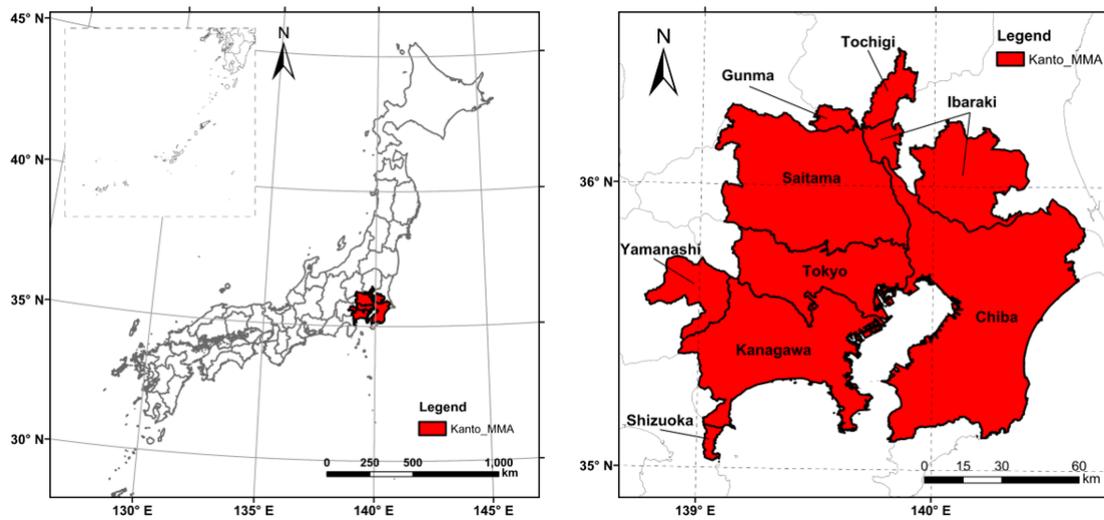
(a)
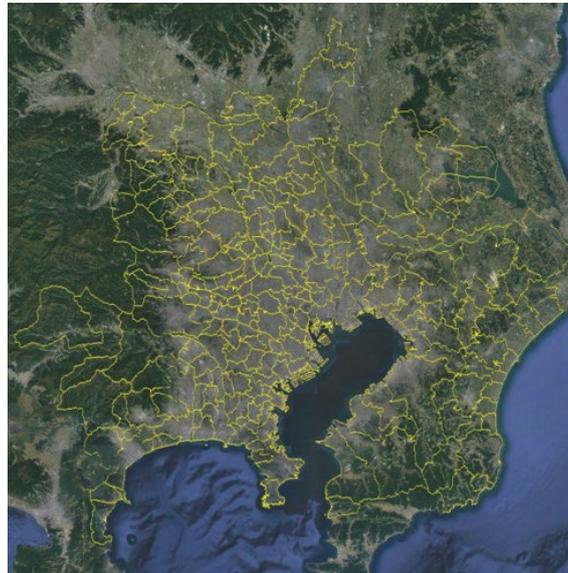
(b)

Figure 4.1. The position (a) and satellite image (b) of the research region.

The center of the Kanto Major Metropolitan Area are the 23 special wards, contains several commercial centers including Shinjuku, Shibuya, Ikebukuro, and Ginza. The 23 special wards are surrounded by numerous suburban cities, which are seamlessly merged into one continuous built-up area. Many rivers flow through the area, and the majors are Ara River and Tama River.

The Kanto Major Metropolitan Area is located in a subtropical humid climate zone [2] with hot and humid summers and cold winters. August is the hottest month of all year, the temperature can reach 25.53 °C, on the contrary, January is the coldest month with only 3.50 °C [3]. The average annual rainfall is nearly 1,530 millimeters, wet in summer and dry in winter [4].

4.2. Data Source and Preprocessing

We used multiple remote sensing data. First, the Landsat 8 OLI/TIRS products in 2010 with a resolution of 30 meters were collected from the NSGS EarthExplorer website [5]. Clouds have a great impact on the accuracy of remote sensing data such as Landsat. Accordingly, we chose a sunny day as research date: October 11, 2010. These Landsat 8 OLI/TIRS data needed to be merged to the size of research region. Table 4-1 shows some Landsat 8 OLI/TIRS data basic information used in this study, including the scene identifier, the worldwide reference system path and row (WRS Path, Row), date, and land cover ratio. Based on remote sensing data, we used the maximum likelihood classification to classify the urban structure combined with satellite images. Previous study results proved that the accuracy of urban structure obtained by this method met the requirements [6]. According to local climate zone (LCZ), the urban structure at 30 m spatial resolution (shown in Figure 4.2) includes two categories: built-up area (10 types) and land cover area (7 types), and built-up area was divided into 10 types: compact high-rise buildings (LCZ 1), compact mid-rise buildings (LCZ 2), compact low-rise buildings (LCZ 3), open high-rise buildings (LCAZ 4), open mid-rise buildings (LCZ 5), open low-rise buildings (LCZ 6), lightweight low-rise buildings (LCZ 7), large low-rise buildings (LCZ 8), sparsely built (LCZ 9), and heavy industry (LCZ 10); similarly, land cover area was divided into 7 types: dense trees (LCZ A), scattered trees (LCZ B), bush and scrub (LCZ C), low plants (LCZ D), bare rock or paved (LCZ E), bare soil or sand (LCZ F), and water (LCZ G) [7].

Table 4-1. Basic information of research region Landsat 8 OLI/TIRS data.

Location	Kanto Major Metropolitan Area
Scene Identifier	LC81070352013260LGN02/
	LC81070362013260LGN02
WRS Path, Row	107, 035/107, 036
Date	2013-09-17
Land Cloud Cover	0.05%/2.88%
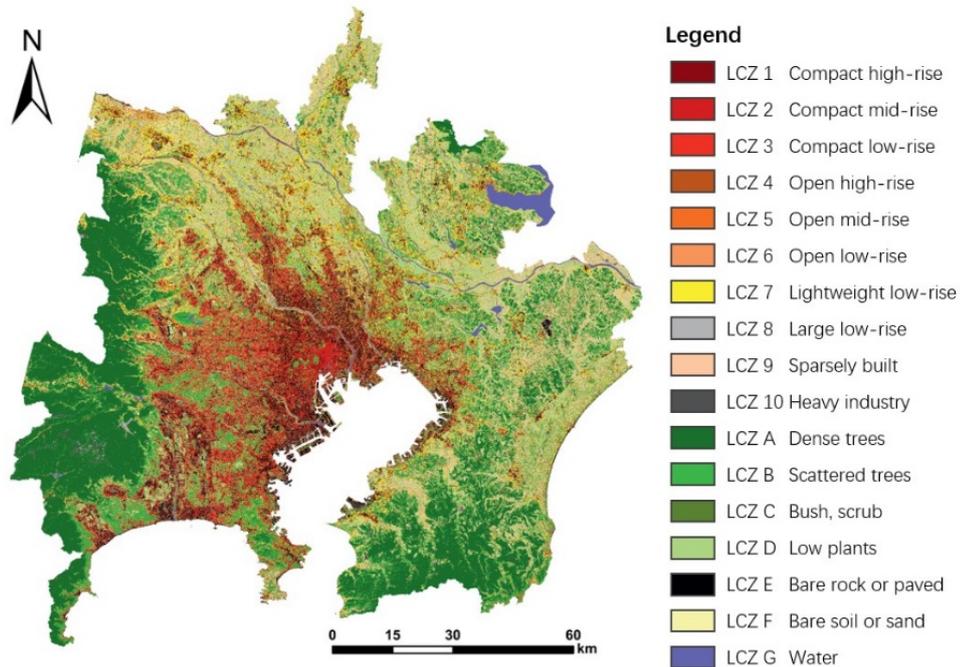
Figure 4.2. The urban structure of the research region obtained from remote sensing data though the maximum likelihood classification method.

Since the urban structure was only qualitative data, that is, it could only indicate which type it was, so how to quantify the urban structure was the key to this study. We calculated the area proportion of each types in various grids in both high and low resolution according to the urban structure with the spatial resolution of 30 m. Then, the urban structure of each grids was determined based on the area proportion of each types.

Second, we downloaded the DEM data with 30 meters spatial resolution (shown in Figure 4.3.) obtained by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model Version 2 (GDEM V2) from the NASA Earthdata website [8]. This model only represented altitude information, without and further definition about the land surface [9].
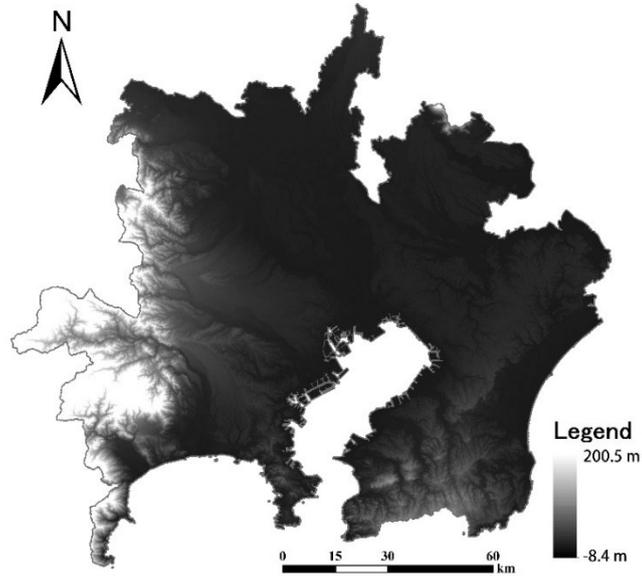
Figure 4.3. The digital elevation model of the research region.

Third, the 2010 annual average air temperature data of Kanto Major Metropolitan Area at 1 km spatial resolution were downloaded from the Japan National Land Numerical Information (NLNI) website [10] which was provided by MLIT. air temperature data was rasterized using a fishnet with a resolution of 1 km (shown in Figure 4.4.).
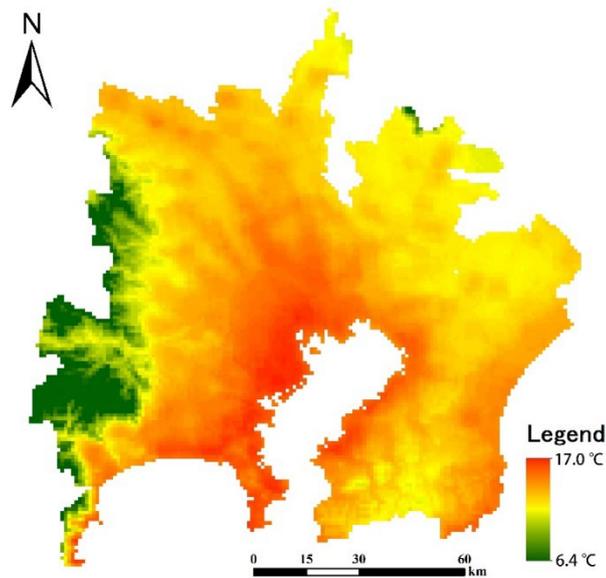
Figure 4.4. The annual average air temperature of the research region at a 1 km spatial resolution.

So as to downscale air temperature from 1 km to 250 m, we resampled the independent and dependent variables with the resolution of both 1 km and 250 m. Then we could statistics the air temperature and urban structure at different resolutions for the next step of downscaling.

4.3. Accuracy Evaluation

4.3.1.   Accuracy Verification by Using K-fold Cross-validation

There were multiple mechanisms to avoid underfitting and overfitting, to build an optimal model [11]. A commonly used method is to separate the sample out two parts, one is training sample for model fitting and the other is test sample for verifying the accuracy. The verification method include leave-one-out CV, and k-fold CV [12]. The k-fold (k=10) CV method was used to calculated precision. The principle of 10-fold cross-validation is shown in Figure 4.5. This method randomly divided the whole samples into ten equally sized sub samples. Then, one of the ten sub samples is reserved to verify the ET model established by the other nine sub samples. The validation process was repeated ten times until each sub sample is used as a test sample. The overall accuracy is the mean of the accuracy of all ten sub samples [13].
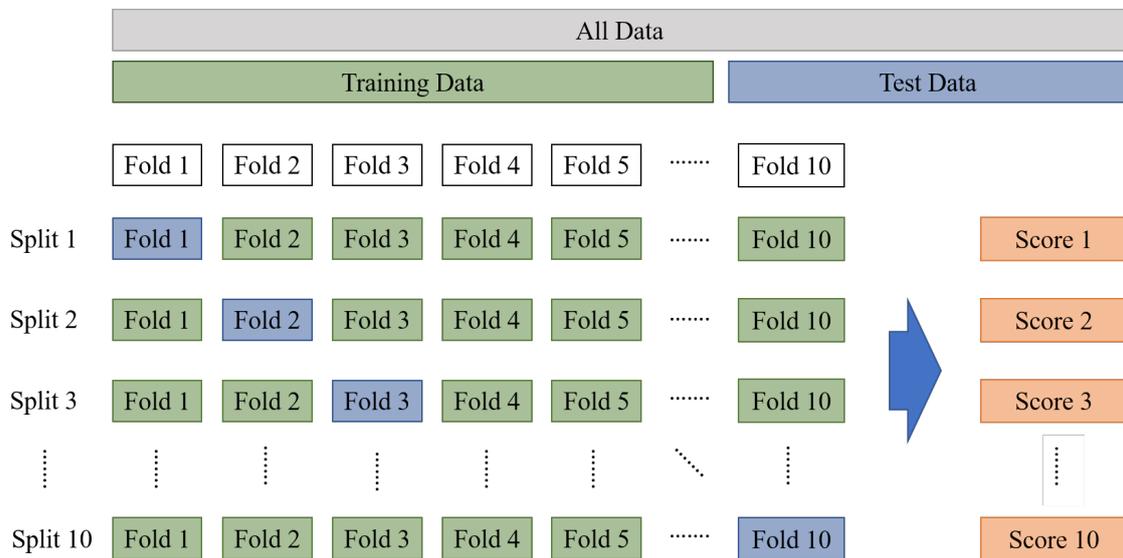
Figure 4.5. The principle of 10-fold cross-validation

4.3.2.   Accuracy Verification by Combining with the Meteorological Station Data

To provide further confirmation of this method accuracy, we introduced the air temperature data from meteorological station. A total of 37 meteorological station that can provide annual average air temperature data in 2010 were located in the research region (shown in Figure 4.6). The meteorological station was evenly distributed in the research region. We count the annual average air temperature of each meteorological station in 2010 and compared with the air temperature after downscaling to verify the accuracy.

No.	Meteorological Station	Latitude	Longitude	Altitude
1	Tokyo	35°41.5'N	139°45.0'E	25.2
2	Edogawarinkai	35°38.3'N	139°51.8'E	5.0
3	Haneda	35°33.2'N	139°46.8'E	6.0
4	Fuchu	35°41.0'N	139°29.0'E	59.0
5	Ome	35°47.3'N	139°18.7'E	155.0
6	Hachioji	35°40.0'N	139°19.0'E	123.0
7	Chiba	35°36.1'N	140°06.2'E	3.5
8	Abiko	35°51.8'N	140°06.6'E	20.0
9	Katori	35°51.5'N	140°30.1'E	37.0
10	Funabashi	35°42.7'N	140°02.6'E	28.0
11	Sakura	35°43.7'N	140°12.7'E	5.0
12	Yokoshibahikari	35°39.4'N	140°28.8'E	6.0
13	Narita	35°45.8'N	140°23.1'E	41.0
14	Mobara	35°26.2'N	140°17.7'E	11.0
15	Ushiku	35°23.8'N	140°08.9'E	30.0
16	Kisarazu	35°21.7'N	139°56.4'E	60.0
17	Sakahata	35°14.1'N	140°05.9'E	120.0
18	Katsuura	35°09.0'N	140°18.7'E	11.9
19	Saitama	35°52.5'N	139°35.2'E	8.0
20	Koshigaya	35°53.0'N	139°45.4'E	3.0
21	Tokorozawa	35°46.4'N	139°24.8'E	119.0
22	Kuki	36°05.2'N	139°38.1'E	12.0
23	Hatoyama	35°59.1'N	139°20.1'E	44.0
24	Kumagaya	36°09.0'N	139°22.8'E	30.0
25	Yorii	36°06.3'N	139°11.0'E	128.0
26	Yokohama	35°26.3'N	139°39.1'E	39.1
27	Miura	35°10.7'N	139°37.8'E	42.0
28	Shijidou	35°19.2'N	139°27.0'E	5.0
29	Ebina	35°26.0'N	139°23.2'E	18.0
30	Odawara	35°16.6'N	139°09.3'E	14.0
31	Ajiro	35°02.7'N	139°05.5'E	66.9
32	Otsuki	35°36.5'N	138°56.3'E	364.0
33	Tatebayashi	36°14.4'N	139°30.8'E	23.0
34	Oyama	36°20.3'N	139°49.8'E	44.0
35	Koga	36°12.1'N	139°43.0'E	20.0
36	Tsukuba	36°03.4'N	140°07.5'E	25.2
37	Tsuchiura	36°06.2'N	140°13.2'E	26.0

Figure 4.6. The location of meteorological station

4.4. Parameters tuning of Extra Trees Model for Case Study

Parameters tuning was an indispensable step to modify the model parameters and to find the optimal model [14]. ET model had the same principle as RF model as mentioned above. The parameter tuning method was similar to RF model. Various parameters combination would lead to different prediction results. Therefore, the single parameters tuning could not optimize the various models [15]. Since n_estimators, max_features, max_depth, and min_samples_leaf were the four most important parameters that could affect the predicted results [16], we tuned the above four parameters and set the other parameters as default values. Figure 4.7 shows the model scores changes when adjusting above four main parameters. Among them, Figure 4.7a shows the change of model scores when n_estimators ranges from 1 to 100 and when max_features is set as None or sqrt; Figure 4.7b shows the partial enlarged detail of score changes when n_estimators ranges from 10 to 50; Figure 4.7c shows the partial enlarged detail of score changes when max_depth ranges from 10 to 20; Figure 4.7d represents the changes of   model score , train MSE, and test MSE when min_samples_leaf ranges from 1 to 30.

In order to prevent underfitting, we added n_estimators to improve the fitting ability of ET model in the tuning process. The value of n_estimators was the ideal value (point A and point B in Figure 4.7b) when the model scores dropped for the first time. At the same time, we set max_features as to None (blue line in Figure 4.7a and 4.9b) and sqrt (orange line in Figure 4.7a and 4.9b). Similarly, when the model scores did not rise significantly for the first time, the value of max_depth was the suitable value (point C in Figure 4.7c). Affected by several parameters tuning, the ET model might also be overfitting[17]. Since the increase of n_estimators and max_depth would cause the model to be complicated and overfitting, it was necessary to rise the value of min_samples_leaf to prevent overfitting and improve the accuracy of model prediction. When train MSE was slightly lower than test MSE, the model was good fitting [18]. Therefore, we introduced Δ to indicate the difference between train and test MSE. When the Δ no longer decreased significantly, it was the optimal set of min_samples_leaf (point D in Figure 4.7d).
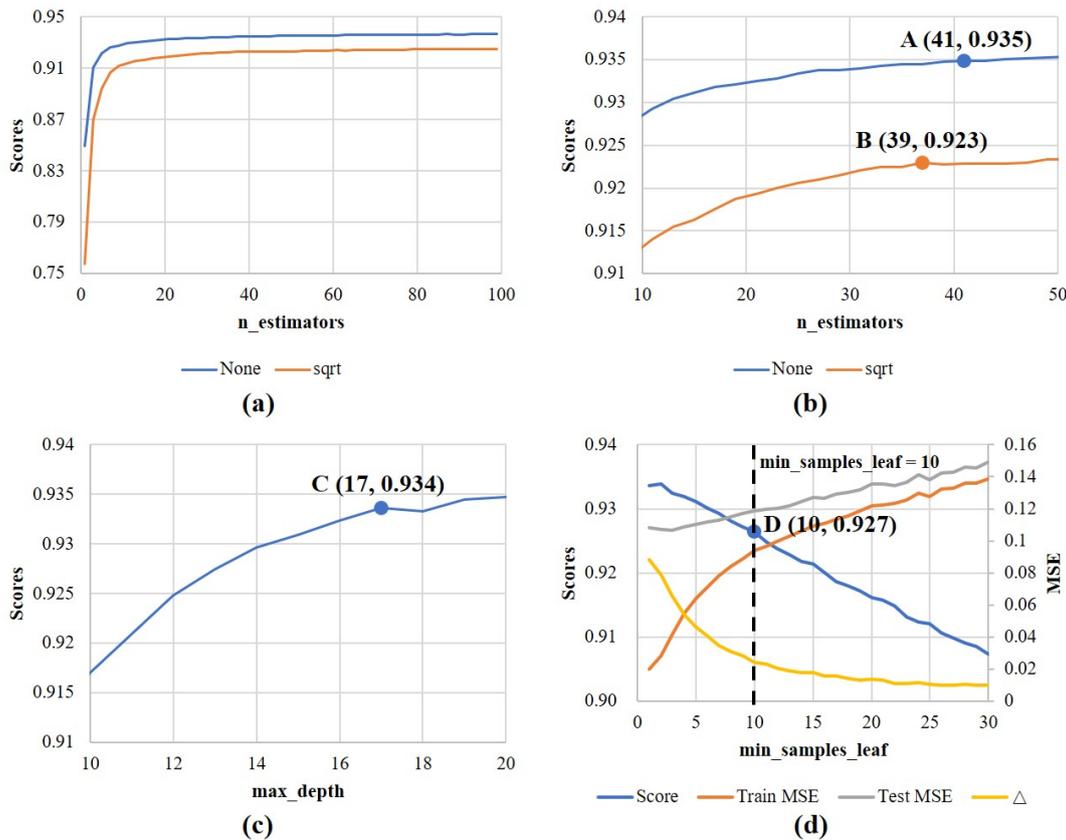
Figure 4.7. The model scores changes (y-axis) when adjusting n_estimators, max_depth, min_samples_leaf (x-axis), and set max_features as None (blue line) or sqrt (orange line): (a) the score changes when n_estimators ranges from 1 to 100 and when max_features is set as None or sqrt; (b) the partial enlarged detail when n_estimators ranges from 10 to 50; (c) the partial enlarged scores when max_depth ranges from 10 to 20; (d) the changes of model score, train MSE, test MSE, and △ (difference between train and test MSE) with min_samples_leaf ranging from 1 to 30.

If we set all the parameters as default values, the model score was 0.912. The model score rose rapidly while n_estimators raised from 1 to 6, then trend to be flat. Due to Figure 4.7b, the orange line reached the first maximum scores (scores = 0.923, point A), when n_estimators was 39. And the first maximum score of blue line was 0.935 (point B), when n_estimators was 41. Thus, point A was the best parameters combination of n_estimators and max_features. In the same way, the first maximum value of max_depth was the optimal solution. Consequently, the max_depth was set as 17, and the score was 0.934. According to Figure 4.7d, when min_samples_leaf exceeded 10, Δ, that is, the difference between train and test MSE did not reduce significantly. When the min_samples_leaf was 10, the score was 0.927 (point D). Currently, the train MSE was 0.106 °C, which was slightly lower than the test MSE with a value of 0.136 °C. Therefore, the ET model was good fitting as min_samples_leaf was 10. After parameters tunning, we obtained the optimal combination (shown in

Table 4-2) corresponding to the model score of 0.927, that was 0.015 higher than the original model score.

Table 4-2. The optimal combination of the four main parameters of the ET model in this study.

Parameter Name in scikit-learn	Ranges	Optimal Value
n_estimators	1,3,5,7…99	41
max_features	None, Auto	None
max_depth	0,1,2,3…30	17
min_samples_leaf	0,1,2,3…30	10

4.5. Precision results of Extra Trees model by using K-fold cross-validation

We used the cross_val_score module in the scikit-learn libraries to combine the previously divided training samples and test samples and conduct 10-fold CV on the optimal model to verify whether the ET model fits well. The results of 10-fold CV are shown in Figure 4.8. Overall, the ET model showed good fitting that the average of scores was about 0.926, ranging from 0.903 to 0.941. The validation results met the requirements. Based on the optimal ET model, we predicted air temperature with low-resolution, and compared between real air temperature and predicted air temperature as shown in Figure 4.9. The predicted air temperature and real air temperature of the scatters falling on the 1:1 line was equal. Most of the scattered points were clustered around the 1:1 line with R^2 of 0.927, that confirmed the ET model have good performance in regression. In summary, this ET model established in section 3.1 can be applied for following downscaling studies.
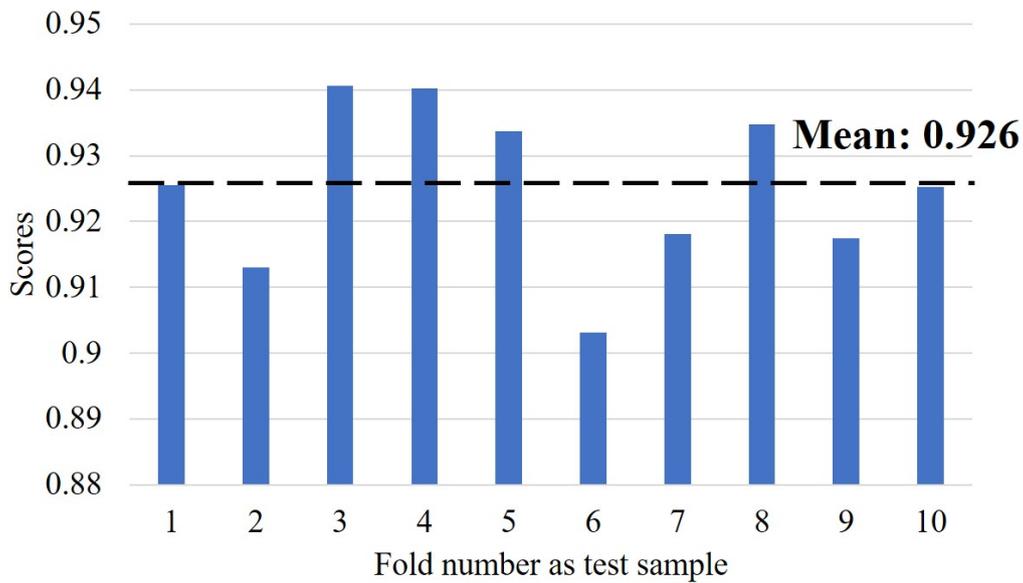
Figure 4.8. 10-fold cross-validation results of the ET model established in section 3.1 (x-axis is the fold number as the test sample; y-axis is the model scores; the black dotted line is the mean of validation results)

4-12
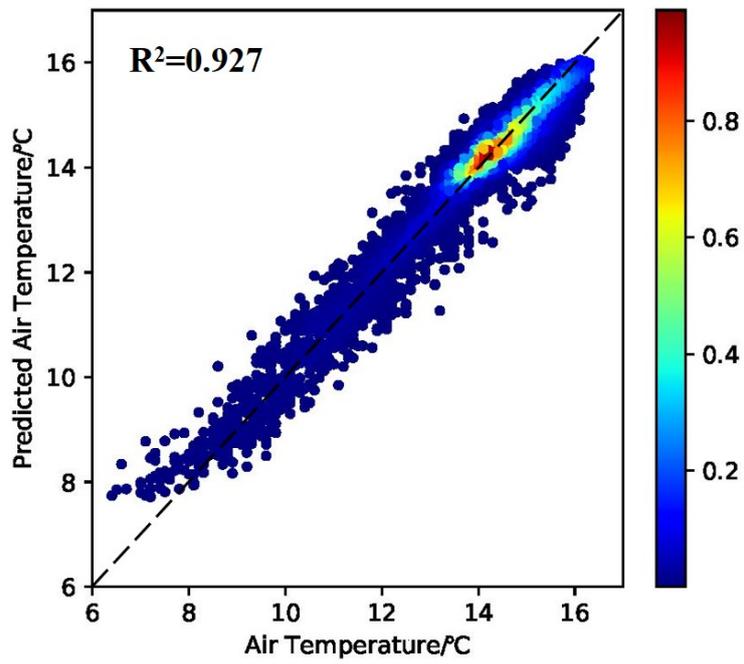
Figure 4.9. Scatter plot of predicted air temperature (y-axis) versus real air temperature (x-axis) (the black dotted line is the 1:1 line; color bar represents the frequency of scatters)

4.6. Air Temperature Downscaling Results

While constructing the downscaling model, the ET model was able to provide the variable importance by observing how much predicted error increased when the train samples of this variable was replaced while those of all other variables remain unchanged [19]. The importance scores were expressed in MSE. The variable importance of the independent variables rose with the increase of MSE [20]. Figure 4.10 shows the importance score of independent variables for the ET model. The x-axis is the MSE, or importance score, and the y-axis is the independent variable. Compared with other variables, DEM scores significantly higher in importance, that is, DEM plays an important part in ET models. The next was LCZ A representing dense trees, which was significantly higher than other urban structure types. However, LCZ 1, LCZ G, LCZ F, and LCZ 8 reflected the lowest importance in this ET model, and their importance scores were less than 0.05.
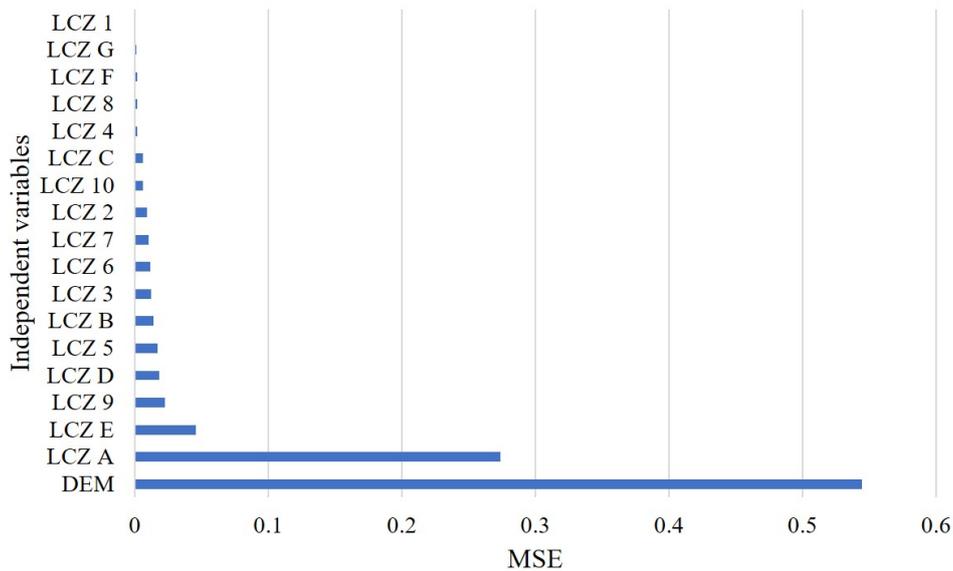
Figure 4.10. Importance score of independent variables (x-axis is MSE, which represents importance scores; y-axis is independent variables)

The purpose of the air temperature downscaling model was to overcome the low-resolution of air temperature data provided by NLNI, so as to furtherly research on the urban thermal environment accurately in Japan. According to the ET model built in Section 4.4, we improved the resolution of air temperature from 1 km to 250 m, namely, we predicted the corresponding air temperature by using independent variables with high resolution. In order to improve the downscaling effect and reduce errors, spline method was used to fit residual, which was introduced in Section 2.6. Figure 4.11 shows the air temperature distribution including the air temperature products from NLNI, low-resolution predicted air temperature (1 km), and high-resolution downscaled air temperature (250m); the latter

two represent the predicted air temperature distribution. The color band of air temperature was represented as green to yellow to red. Green areas are low-temperature area, and red areas are high-temperature area. Comparing low-resolution NLNI/AT and predicted AT, the similarity of high-temperature and low-temperature area distribution was extremely high. The predicted air temperature was more sharpen than air temperature products from NLNI. The results showed that the ET model is fitting for following air temperature downloaded in the Kanto Major Metropolitan Area. Figure 4.11c shows the high-resolution air temperature after residuals fitting. And the air temperature distribution was basically consistent with that of air temperature products from NLNI. Downscaled air temperature ranged from 7.1 ℃ to 16.9 ℃, and the temperature difference was smaller than air temperature products of NLNI which ranged from 6.4 ℃ to 17.0 ℃. From the downscaled results, we found that there was a large piece of red area throughout the Kanto Major Metropolitan Area , which were mainly distributed along the Tokyo Bay. The low-temperature area represented by green mostly covered the west side of the study area. And the downscaled air temperature was generally slightly higher than air temperature products from NLNI. Combined with urban structure, this indicated that the downscaling effect in low-temperature area (mainly mountain areas with high vegetation coverage) is poor. A long and narrow yellow area, that is, the mid-temperature area, lain between the mountain area and the urban area on the west side. Compared with east side, red area on the north side is deeper into the land. The area of the mid-temperature zone in the north and west was basically the same. Based on the high-resolution air temperature, several yellow lines, which were Ara River, Edo River, and Tama River, with mid-temperature passed through the urban center. Meanwhile, there was a small conspicuous orange area, that is, Tokyo Imperial Palace with a lot of vegetation space, in the red area on the northwest bank of Tokyo Bay, and the air temperature was slightly lower than the surrounding area.
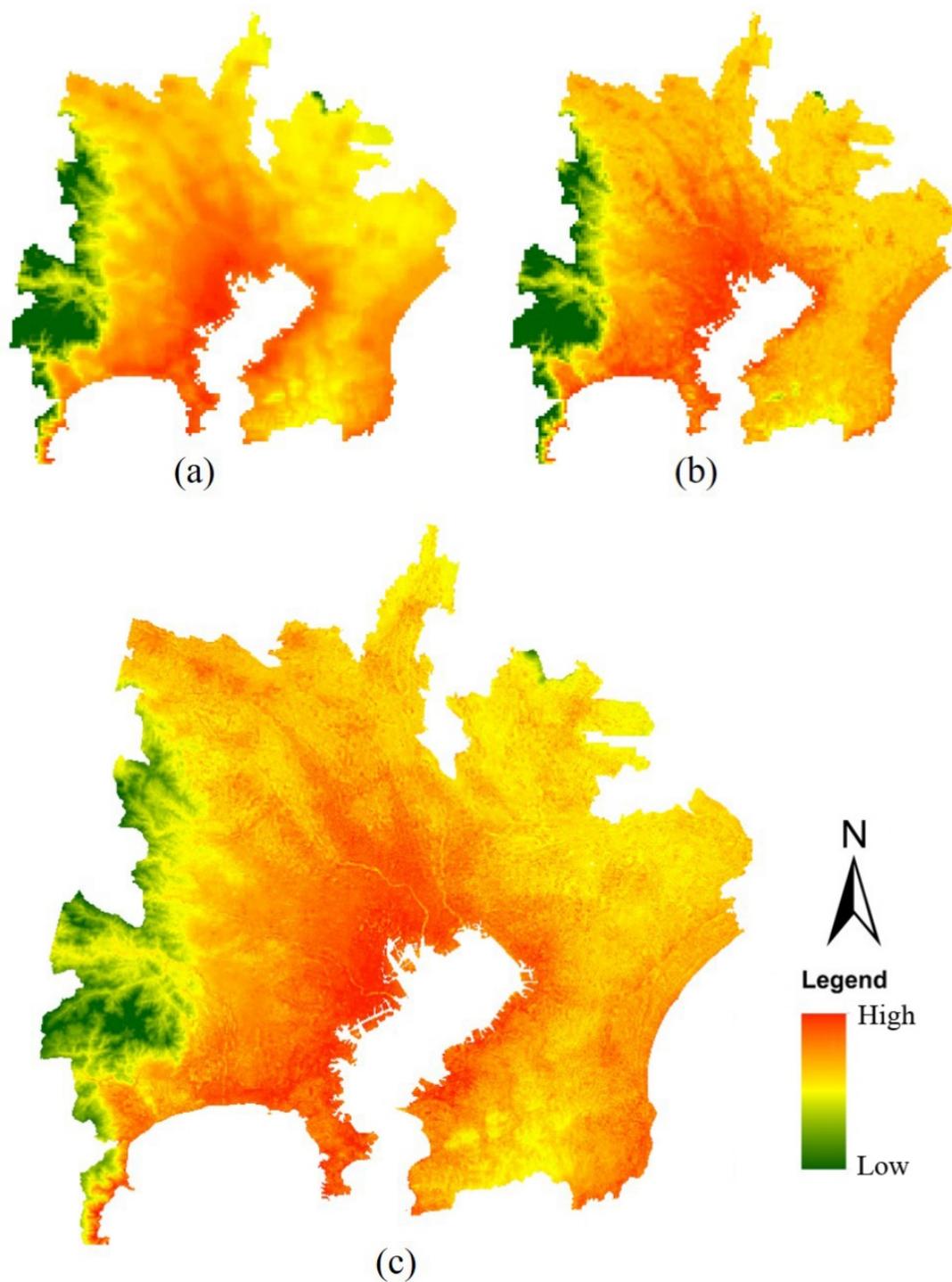
Figure 4.11. Predicted and downscaled results at different resolution from air temperature products from NLNI: (a) low-resolution air temperature products from NLNI (1km); (b) low-resolution predicted AT (1 km); (c) high-resolution downscaled air temperature (250 m).

4.7. Validation Results Comparing Downscaled Result and Meteorological Station Data

According to the measured air temperature from 37 meteorological stations mentioned in Section 2.4.2, we collected the annual average air temperature data of 2010 from Japan Meteorological Agency [3]. Then, we used the location information of the meteorological station to extract the corresponding downscaled air temperature data from the downscaling results in Figure 11c. Finally, we integrated the two set of air temperature data, and the result is shown in Figure 12. The x-axis is the downscaled air temperature, and the y-axis is the measured air temperature. It is obvious that the scatters were distributed near the blue dashed line. The R^2 was about 0.9542. The error did not exceed plus or minus 0.4 ℃. The difference of most data sets was less than or equal 0.1 ℃, and only the two sets differed by 0.4 ℃. A small amount of errors can meet the research requirements, that is, the ET model-based air temperature downscaling method meets the precision requirements and can be applied in air temperature downscaling research.
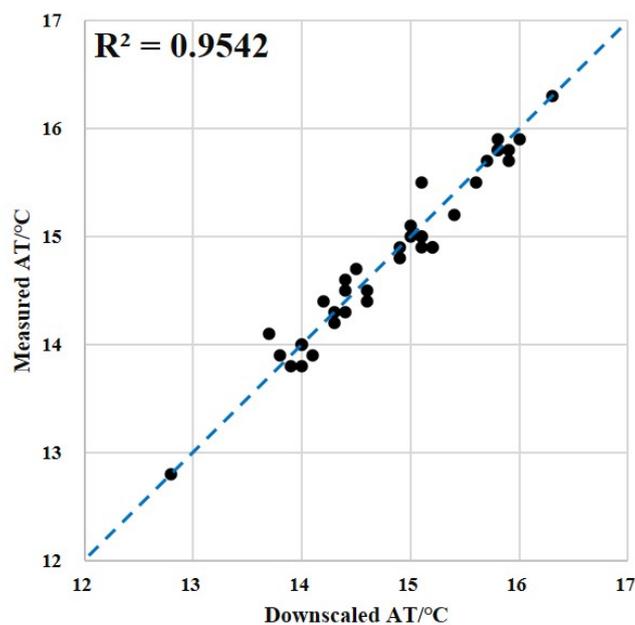
Figure 4.12. Comparison of downscaled (x-axis) and measured (y-axis) air temperature of meteorological station. (the blue dotted line is the 1:1 line)

4.8. Discussion

The objective of this study was to evaluate the effectiveness of extra tree regressor model to downscale air temperature based on urban structure and DEM. The results of the air temperature downscaling experiment in Kanto Major Metropolitan Area demonstrated that usefulness of this method.

4.8.1.   Accuracy analysis of the downscaling results in Kanto Major Metropolitan Area.

According to the downscaling accuracy validation of Section 3.2 and 3.4, the downscaling results could generally reflect the accurate high-resolution air temperature. It was important to point out that the precision validation of the first meteorological station with high vegetation and water coverage, that is, Tokyo meteorological observation point, which located in Tokyo Imperial Palace. Tokyo Imperial Palace, which could be equated to a large-scale urban park, usually exerted a cooling effect in warm areas and played an important part in reducing the air temperature of downtown areas [21,22], that has been confirmed in different region across the globe [23–27]. Similarly, Tokyo Imperial Palace located in the urban center is also surely helpful to reduce air temperature of the urban center, that is, the air temperature here was lower than that in the surrounding areas. Through prediction and residual fitting, the downscaled air temperature of Tokyo meteorological observation point was 16.3 ℃, which was the same as the measured data at this point. The value was lower than the surround areas with 17.0 ℃ (before downscaling) and 16.9 ℃ (after downscaling). However, based on high-resolution air temperature products from NLNI, the air temperature at this point reached 16.8 ℃, because this area had a smaller size than the resolution (1 km), or was between several grids. The downscaling results with a resolution of 250 m provided an accurate description of air temperature at this point. This was the purpose of air temperature downscaling.

However, downscaled air temperature trended to produce deviations in extreme temperature ranges, including high-temperature in downtown and low-temperature areas in mountain region. As mentioned in Section 3.3, the air temperature ranges from 6.4 ℃ to 17.0 ℃ before downscaling to 7.1 ℃ to 16.9 ℃ after downscaling. The range was significantly reduced, especially at low temperatures. Figure 4.11 also shows that in low-temperature area (air temperature was less than 8 ℃), the scatters had a larger deviation from the black dashed line than the other ranges. We thought that this was caused by two reasons. The first is that the number of training samples in these extreme temperature ranges is insufficient to calibrate the ET downscaling model [28]. Secondly, for extremely low-temperature areas, namely mountain areas, the urban structure in this area is relatively simple, mainly LCZ A (dense trees), and even some grids with a resolution of 1 km were all of the same urban structure types. This caused that the urban structure could not fit the ET downscaling model, the air temperature prediction in these areas depended entirely on the independent variable of DEM. This indicates that the accuracy of the air temperature downscaling method based on machine learning model is higher in the area with

high heterogeneity than in the space with homogeneous urban structure. This conclusion had also been confirmed in previous studies [29–31].

4.8.2.   Variables' importance analysis

The variables importance could also reflect the impact of urban structure and altitude on air temperature to a certain extent. According to Figure 4.12, it was obvious that DEM was extremely important in the ET model establishment, that is, altitude had a great influence on temperature. Based on the environmental lapse rate, the air temperature at high altitudes were generally lower than those at low altitude [32,33].

Excluding DEM, in the urban structure types, LCZ A had the greatest impact on temperature on air temperature. This was well understood that vegetation areas, especially dense trees, had a good cooling effect on the urban environment, which corresponds to results of other studies [34–36]. However, we turned up that the importance of LCZ 1 (compact high-rise building), which could also affect the urban thermal environment [37,38], was the lowest. We inferred that the variable importance was also related to the number of variable samples. Therefore, we counted the proportion of each urban structure types. The results showed that the lowest four variables were LCZ 1, LCZ G, LCZ F, and LCZ 8 in the importance ranking. Their corresponding area ratios were 0.4%, 1.1%, 1.9%, and 0.8% respectively. In the urban structure types, the area proportion of the highest importance, namely LCZ A, was 19.3%, which is the largest. Similarly, because DEM existed in the entire study area, it could participate in the model establishing fully, it had the highest importance.

Consequently, it should be advised that the importance score of the independent variables depended not only on the influence of the independent variables on the dependent variable, but also on count and size of variables. For example, removing or replacing independent variables might change the variable importance because different interrelated variables might be act as surrogates [28].

4.8.3.   Limitations of the current research

We found that the ET method has some defects in the research process. It was undeniable that the air temperature downscaling method based on machine learning algorithm and the correlation between urban structure and air temperature had a high accuracy. We used the ET model, set DEM, NDVI, NDBI, and the reflectance of band 1 to band 5 and band 7 in Landsat 8 OLI/TIRS as the independent variables, and performed the similar parameters tuning, then predicted the air temperature in the same study area. It turned out that CV score was about 0.886, which was smaller than the CV results of this study. Obviously, the independent variables choice was more conducive to the model establishing. However, according to the aforementioned, this independent variable combination was more suitable for high heterogeneous areas, and the downscaling accuracy for high homogeneous areas of this combination was not satisfactory, especially in mountain areas in this study. In this study, even if we

added DEM to the independent variable to improve the downscaling effect of high homogeneous regions, the results were still undesirable.

In the measured air temperature verification section, since NLNI provided the annual average air temperature in 2010, we only could use the corresponding meteorological station measured air temperature data to verify the downscaling results. But the location of meteorological station was limited. For example, in this study, there were densely distributed meteorological stations in plain area which could provide measured data. On the contrary, there were no meteorological stations in mountain areas. So, we were hard to verify the accuracy of mountain areas. Meanwhile, remote sensing data could not provide accurate air temperature data for precision verification.

4.9. Summary

In this chapter, we presented an air temperature downscaling method that can be applied for AUHI analysis. Taking the Kanto Major Metropolitan Area, Japan, as a case study, we establish an ET model, a machine learning model, to assist in downscaling research. The basic principle of downscaling was that the regression of air temperature, as dependent variable, and independent variables, including DEM and urban structure remained unchanged at different scale. Based on the resampled low-resolution air temperature that obtained from NLNI, DEM and urban structure data which identified by maximum likelihood classification method from Landsat 8 OLI/TIRS data, we built an optimal ET model by parameters tuning, and the model passed the CV. Then, we inputted DEM and urban structure with a scale of 250 m into the ET model to predict the high-resolution air temperature data. Finally, we fitted the residual to obtain the downscaling air temperature data of study area and compared with measured air temperature from meteorological station to check the downscaling precision.

We concluded that the proposed ET downscaling method for low-resolution air temperature according to DEM, urban structure, was confirmed to be flexible and efficient in downscaling air temperature spatial resolution from 1 km to 250 m in this study area. In accordance with prediction accuracy analysis, 10-fold CV score was sufficient to support the subsequent downscaling. Combining DEM and the urban structure with a resolution of 250 m, we could obtain high-resolution air temperature data through the previous ET model. Finally, the comparison between downscaled air temperature and measured air temperature supported the conclusion that the ET model and the independent variables combination, including DEM and the urban structure, can be used to downscale low-resolution air temperature, especially in high heterogeneity regions.

Although it is not within the scope of this study, the ET model may also be applicable to other regions or other spatial scales and should be further validated. By way of illustration, downscaling the air temperature to a smaller resolution with 100 m, and in other regions in various continents around the world. Meanwhile, how to improve the downscaling precision in high homogeneous areas also needs to be further studied attention in future research.

Reference

1.   Bureau, S. Statistical handbook of Japan 2015. **2015**.

2.   Peel, M.C.; Finlayson, B.L.; McMahon, T.A. Updated world map of the Köppen-Geiger climate classification. *Hydrol. earth Syst. Sci.* **2007**, *11*, 1633–1644.

3.   Japan Meteorological Agency Available online: http://www.jma.go.jp/jma/index.html (accessed on Apr 29, 2021).

4.   Matsumoto, J.; Fujibe, F.; Takahashi, H. Urban climate in the Tokyo metropolitan area in Japan. *J. Environ. Sci. (China)* **2017**, *59*, 54–62, doi:10.1016/j.jes.2017.04.012.

5.   EarthExplorer Available online: https://earthexplorer.usgs.gov/ (accessed on Jan 31, 2021).

6.   Wang, R.; Gao, W.; Zhou, N.; Kammen, D.M.; Peng, W. Urban structure and its implication of heat stress by using remote sensing and simulation tool. *Sustain. Cities Soc.* **2021**, *65*, 102632, doi:10.1016/j.scs.2020.102632.

7.   Stewart, I.D.; Oke, T.R. Local climate zones for urban temperature studies. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 1879–1900.

8.   Earthdata Search Available online: https://search.earthdata.nasa.gov/search (accessed on Jan 31, 2021).

9.   Peckham, R.J.; Gyozo, J. Development and Applications in a Policy Support Environment Series: Lecture Notes in Geoinformation and Cartography. *Heidelberg, Ger.* **2007**.

10.   National Land Numerical Information Available online: https://nlftp.mlit.go.jp/ksj/ (accessed on Jan 31, 2021).

11.   Fassnacht, F.E.; Hartig, F.; Latifi, H.; Berger, C.; Hernández, J.; Corvalán, P.; Koch, B. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. *Remote Sens. Environ.* **2014**, *154*, 102–114, doi:10.1016/j.rse.2014.07.028.

12.   Kuhn, M.; Johnson, K. Over-Fitting and Model Tuning. In *Applied Predictive Modeling*; Springer New York, 2013; pp. 61–92.

13.   Wu, Y.; Duguay, C.R.; Xu, L. Assessment of machine learning classifiers for global lake ice cover mapping from MODIS TOA reflectance data. *Remote Sens. Environ.* **2021**, *253*, 112206, doi:10.1016/j.rse.2020.112206.

14.   Biau, G.; Scornet, E. A random forest guided tour. *Test* **2016**, *25*, 197–227, doi:10.1007/s11749-016-0481-7.

15.   Probst, P.; Boulesteix, A.-L. To Tune or Not to Tune the Number of Trees in Random Forest. *J. Mach. Learn. Res.* **2017**, *18*, 6673–6690.

16.   Srivastava, T. Tuning the parameters of your random forest model. *Anal. Vidhya* **2015**, *9*.

17.   Bishop, C.M. *Pattern recognition and machine learning*; springer, 2006; ISBN 1493938436.

18.   Ballabio, C. Spatial prediction of soil properties in temperate mountain regions using support

vector regression. *Geoderma* **2009**, *151*, 338–350, doi:10.1016/j.geoderma.2009.04.022.

19. Liaw, A.; Wiener, M. *Classification and Regression by RandomForest*; 2001;

20. Zhao, W.; Sánchez, N.; Lu, H.; Li, A. A spatial downscaling approach for the SMAP passive surface soil moisture product using random forest regression. *J. Hydrol.* **2018**, *563*, 1009–1024, doi:10.1016/j.jhydrol.2018.06.081.

21. Guo, S.; Yang, G.; Pei, T.; Ma, T.; Song, C.; Shu, H.; Du, Y.; Zhou, C. Analysis of factors affecting urban park service area in Beijing: Perspectives from multi-source geographic data. *Landsc. Urban Plan.* **2019**, *181*, 103–117, doi:10.1016/j.landurbplan.2018.09.016.

22. Peng, J.; Dan, Y.; Qiao, R.; Liu, Y.; Dong, J.; Wu, J. How to quantify the cooling effect of urban parks? Linking maximum and accumulation perspectives. *Remote Sens. Environ.* **2021**, *252*, 112135, doi:10.1016/j.rse.2020.112135.

23. Chen, X.; Su, Y.; Li, D.; Huang, G.; Chen, W.; Chen, S. Study on the cooling effects of urban parks on surrounding environments using Landsat TM data: A case study in Guangzhou, southern China. *Int. J. Remote Sens.* **2012**, *33*, 5889–5914, doi:10.1080/01431161.2012.676743.

24. Yang, C.; He, X.; Yu, L.; Yang, J.; Yan, F.; Bu, K.; Chang, L.; Zhang, S. The Cooling Effect of Urban Parks and Its Monthly Variations in a Snow Climate City. *Remote Sens.* **2017**, *9*, 1066, doi:10.3390/rs9101066.

25. Algretawee, H.; Rayburg, S.; Neave, M. Estimating the effect of park proximity to the central of Melbourne city on Urban Heat Island (UHI) relative to Land Surface Temperature (LST). *Ecol. Eng.* **2019**, *138*, 374–390, doi:10.1016/j.ecoleng.2019.07.034.

26. Skoulika, F.; Santamouris, M.; Kolokotsa, D.; Boemi, N. On the thermal characteristics and the mitigation potential of a medium size urban park in Athens, Greece. *Landsc. Urban Plan.* **2014**, *123*, 73–86, doi:10.1016/j.landurbplan.2013.11.002.

27. Sugawara, H.; Shimizu, S.; Takahashi, H.; Hagiwara, S.; Narita, K.; Mikami, T.; Hirano, T. Thermal Influence of a Large Green Space on a Hot Urban Environment. *J. Environ. Qual.* **2016**, *45*, 125–133, doi:10.2134/jeq2015.01.0049.

28. Hutengs, C.; Vohland, M. Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sens. Environ.* **2016**, *178*, 127–141, doi:10.1016/j.rse.2016.03.006.

29. Ebrahimy, H.; Azadbakht, M. Downscaling MODIS land surface temperature over a heterogeneous area: An investigation of machine learning techniques, feature selection, and impacts of mixed pixels. *Comput. Geosci.* **2019**, *124*, 93–102, doi:10.1016/j.cageo.2019.01.004.

30. Gao, F.; Kustas, W.; Anderson, M. A Data Mining Approach for Sharpening Thermal Satellite Imagery over Land. *Remote Sens.* **2012**, *4*, 3287–3319, doi:10.3390/rs4113287.

31. Wang, R.; Gao, W.; Peng, W. Downscale MODIS Land Surface Temperature Based on Three Different Models to Analyze Surface Urban Heat Island: A Case Study of Hangzhou. *Remote Sens.* **2020**, *12*, 2134, doi:10.3390/rs12132134.

32.   Jingchao, J.; Junzhi, L.; Chengzhi, Q.; Yamin, M.; A-Xing, Z. Near-surface air temperature lapse rates and seasonal and type differences in China. *Prog. Geogr.* **2016**, *35*, 1538–1548, doi:10.18306/dlkxjz.2016.12.010.

33.   Lookingbill, T.R.; Urban, D.L. Spatial estimation of air temperature differences for landscape-scale studies in montane environments. *Agric. For. Meteorol.* **2003**, *114*, 141–151, doi:10.1016/S0168-1923(02)00196-X.

34.   Shiflett, S.A.; Liang, L.L.; Crum, S.M.; Feyisa, G.L.; Wang, J.; Jenerette, G.D. Variation in the urban vegetation, surface temperature, air temperature nexus. *Sci. Total Environ.* **2017**, *579*, 495–505, doi:10.1016/j.scitotenv.2016.11.069.

35.   Davis, A.Y.; Jung, J.; Pijanowski, B.C.; Minor, E.S. Combined vegetation volume and "greenness" affect urban air temperature. *Appl. Geogr.* **2016**, *71*, 106–114, doi:10.1016/j.apgeog.2016.04.010.

36.   Skelhorn, C.; Lindley, S.; Levermore, G. The impact of vegetation types on air and surface temperatures in a temperate city: A fine scale assessment in Manchester, UK. *Landsc. Urban Plan.* **2014**, *121*, 129–140, doi:10.1016/j.landurbplan.2013.09.012.

37.   Alexander, C. Influence of the proportion, height and proximity of vegetation and buildings on urban land surface temperature. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *95*, 102265, doi:10.1016/j.jag.2020.102265.

38.   Cao, Q.; Luan, Q.; Liu, Y.; Wang, R. The effects of 2D and 3D building morphology on urban environments: A multi-scale analysis in the Beijing metropolitan region. *Build. Environ.* **2021**, *192*, 107635, doi:10.1016/j.buildenv.2021.107635.

Chapter 5. Downscale Air Temperature Prediction of Metropolitan Area by Extra Trees Model

5.1. Study Area Introduction

Japan is an island country, centered on the Japanese archipelago, including four major islands of Hokkaido, Honshu, Shikoku, and Kyushu. It consists of about 6,800 islands, with a total land area of about 377,950 square kilometers. It is the 4th largest island country in the world and the largest island country in East Asia [1]. It is located at the junction of the four plates of the Eurasian plate, the North American plate, the Pacific plate, and the Philippines plate.

Approximately 75% of Japan's land is mountainous and hilly areas. Small-scale mountain basins and plains are scattered throughout the country, becoming a concentration of residential, farming, and economic activities. The mountainous area accounts for the largest land area in the four countries at 79.9%, followed by Kyushu at 64.8%, Honshu at 63.6%, and Hokkaido the lowest at 49%. The plains in Japan are not large, and most of them are small-scale alluvial plains, coastal plains, and alluvial platforms.

Due to the large latitude of Japan, the air temperature difference between north and south is very significant; and the continuous mountain peaks running through the center of the Japanese archipelago make the climate difference between the Pacific coast and the Japan Sea coast. Because Japan is located on the east coast of the mid-latitude mainland, Japan's climate is obviously affected by the confluence of monsoons and ocean currents. Japan is an island country with small annual temperature differences and abundant precipitation, with obvious maritime climate characteristics.

5.1.1. Japan Climate Zones

Japan's climate characteristics are affected by ocean currents and seasonal winds, with four distinct seasons. Summer is the season of high temperature, high humidity and rain, and winter is low temperature and low precipitation. Japan can be divided into six climate zones, namely Hokkaido Zone, Sea of Japan Zone, Central Highland Zone, Pacific Coast Zone, Inland Sea Zone, and Ryukyu Zone, as shown in Figure 5.1. Since the Ryukyu Islands are far away from the Japanese archipelago, this study does not include the Ryukyu Zone.

● Hokkaido Zone

Hokkaido is the northernmost island in Japan, located at the mid-latitude of northern hemisphere and the eastern edge of northeast Asia, and is surrounded by Sea of Japan, the Pacific Ocean, and the Sea of Okhotsk. The climate of Hokkaido is affected by the currents around the island. Summers are cool, winters are harsh, and there is a large air temperature difference between summer and winter. As it is not affected by East Asian rainy season and typhoon, the weather is very dry. The entire area of Hokkaido is heavily snowed. The northern region and the eastern region of Hokkaido are very cold. Except for the coastal areas, the average temperature of the coldest month in almost all areas is -8 ℃ or lower, and usually dropship to -30 ℃.

● Sea of Japan Zone

Sea of Japan Zone is the characteristic of the winter climate. Its scope is the western coastal area of Honshu Island. In winter, it is controlled by high air pressure in Siberia, blowing northwest wind. At the same time, there is often heavy snow because the warm Tsushima Sea current passes through the Sea of Japan in winter and brings a lot of water vapor. There is less precipitation in summer and sometimes abnormally high temperatures occur.

● Central Highland Zone

Central Highland Zone is the climate of a basin surrounded by high-altitude mountainside. It is a typical inland climate with low precipitation. Cold in winter and cool in summer. In Japan, which is largely affected by the warm and moist winds from the Pacific Ocean side in summer and the cold and moist winds that cross the Sea of Japan from mainland China in winter, these seasonal winds are blocked by the surrounding high-altitude mountains. Since it is not affected, the humidity is low and stable throughout the year, and the annual rainfall is low, about 1000 mm. The number of days of precipitation, which is the number of days when precipitation of 1 mm or more is observed, is also small except for the rainy season (including meteorological phenomena similar to the rainy season). Also, in winter, the temperature in the morning and evening is often quite low due to the radiative cooling phenomenon.

● Pacific Coast Zone

Pacific Coast Zone refers to the regional climate characteristics of the Japanese archipelago on the Pacific coast, including the east coast of Honshu, Southern Shikoku, and most of Kyushu. Affected by the southeastern monsoon, there are many strong typhoons in the rainy season in summer. In winter, there is less snow.

● Inland Sea Zone

Inland Sea Zone includes parts of Sanyo, Kita-Shikoku, Kinki, and Kyushu. The weather here is often sunny, with little rainfall, and suffering from drought from time to time. The summer monsoon is blocked by the Shikoku Mountains, and the winter monsoon is blocked by the Chugoku Mountains. For this reason, the weather and humidity are stable throughout the year, and the precipitation months are bimodal in May, June, July (rainy season) and September (autumn rain / typhoon), and 8 between the two peaks. The number of precipitation days throughout the year (the number of days when precipitation of 1 mm or more is observed) is also small except for the rainy season (including meteorological phenomena similar to the rainy season). The situation is different from the area with the climate on the Sea of Japan side. Precipitation is around 1000 to 1600 mm per year.
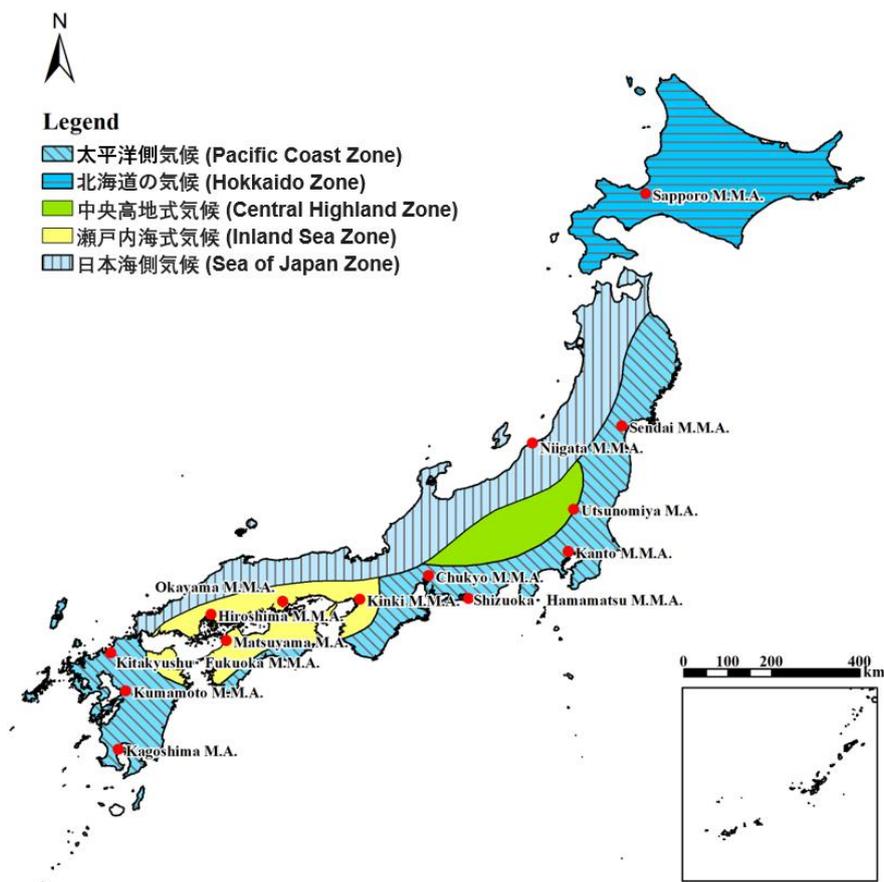
Figure 5.1. Japan Climatic Zones

5.1.2.   Selection of Metropolitan Area as Study Area

Metropolitan Area refers to a region with metropolis as the core and radiating to the surroundings to build a city. The characteristics of the metropolitan area are reflected in the economic ties between nearby cities, the division of labor and cooperation of industries, the mutual influence of transportation and social life, urban planning and infrastructure construction [2]. With the development of social, economic and political systems, the metropolitan areas have become the key regions [3]. Metropolitan areas include satellite cities, towns, and intermediate rural areas. These areas are economically linked to urban core areas and are usually connected by commuting [4]. Most metropolitan areas are based on a core city, but in some cases, metropolitan areas have multiple centers with almost equal importance.

The Statistics Bureau of Japan defines the metropolitan area as one or more central cities and their related outlying cities. There must be at least 1.5% of the permanent population aged 15 or over commuting to school or working in one of the cities to become an outlying municipality. To qualify as a central city, a city must either be a designated city of any population or a non-designated city with a

city proper population of at least 500,000. There are a total of 14 metropolitan areas in Japan, including 11 major metropolitan area as shown in Table 5-1 [5].

Table 5-1 List of Metropolitan Areas in Japan

Climate Zone	Metropolitan Name	Population	Area	Population Density
Pacific Coast Zone	Kanto MMA	37,273,866	14,034	2,656
	Chukyo MMA	9,363,221	7,266	1,289
	Kitakyushu-Fukuoka MMA	5,538,142	5,731	966
	Shizuoka-Hamamatsu MMA	2,842,151	5,494	517
	Sendai MMA	2,256,964	6,506	347
	Kumamoto MMA	1,492,975	4,251	351
	Kagoshima MA	1126639	3458	326
Inland Sea Zone	Kinki MMA	19,302,746	13,033	1,481
	Hiroshima MMA	2,096,745	5,048	415
	Okayama MMA	1,639,414	3,637	446
	Matsuyama MA	706,883	2,272	311
Hokkaido Zone	Sapporo MMA	2636254	4514	584
Sea of Japan Zone	Niigata MMA	1395612	5345	261
Central Highland Zone	Utsunomiya MA	1655673	5455	304

According to Table 5-1, there are the most metropolitan areas located in Pacific Coast Zone, with a total of seven, including Kanto Major Metropolitan Area, Chukyo Major Metropolitan Area, Kitakyushu-Fukuoka Major Metropolitan Area, Shizuoka-Hamamatsu Major Metropolitan Area, Sendai Major Metropolitan Area, Kumamoto Major Metropolitan Area, and Kagoshima Metropolitan Area. Followed by Inland Sea Zone, with a total of four, including Kinki Major Metropolitan Area, Hiroshima Major Metropolitan Area, Okayama Major Metropolitan Area, and Matsuyama Metropolitan Area. Each of the other three climatic zone has only one metropolitan area, which is the

Sapporo Major Metropolitan Area of Hokkaido Zone, Niigata Major Metropolitan Area of Sea of Japan Zone, and Utsunomiya Metropolitan Area of Central Highland Zone. It can be seen that the metropolitan area of Japan is mainly concentrated on the Pacific coast and the Set Inland Sea area.

We selected the largest metropolitan area whose population is most from each climatic zone as the research objects, namely Kanto Major Metropolitan Area, Kinki Major Metropolitan Area, Sapporo Major Metropolitan Area, Niigata Major Metropolitan Area, and Utsunomiya Metropolitan Area.

- Kanto Major Metropolitan Area

The Kanto Major Metropolitan Area is the most populous metropolitan area around the world. And at the same time, Tokyo has the largest metropolitan economy in the world. According to a study conducted, the Tokyo urban area had a total GDP of \$2 trillion in 2012, that topped in the list. Tokyo is also a major international finance center. About the transportation, Tokyo is Japan's largest domestic and international hub for rail, ground, and air transportation. The population density is about 2,656 people/km^2. Tokyo is a subtropical monsoon climate with an annual average air-temperature of 15.6 degrees. It has 4 distinctive seasons with abundant precipitation. There is more precipitation in summer because of the southeastern monsoon. There is less snow in winter.

More detailed information about the Kanto Metropolitan Area has been introduced in Section 4.1, and the study area is also the same as the case study in Chapter 4, as shown in Figure 5.2. The central cities include 23 special wards area (Tokyo), Yokohama, Sagamihara, Kawasaki, Chiba, Saitama.
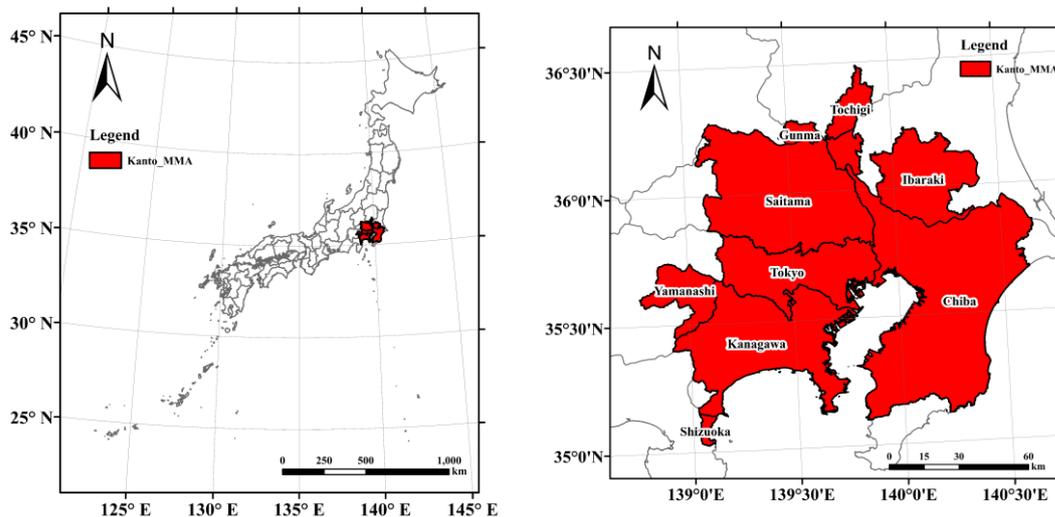
Figure 5.2. The location of Kanto Major Metropolitan Area in Japan

● Kinki Major Metropolitan Area

Kinki (shown in Figure 1) is a metropolitan region in the Kansai region of Japan encompassing the metropolitan areas of the cities of Kyoto in Kyoto Prefecture, Osaka in Osaka Prefecture and Kobe in Hyōgo Prefecture. It is the second-most-populated urban region in Japan (after the Kanto Major Metropolitan Area), containing approximately 15% of Japan's population. The entire region has a population (as of 2010) of 19 million over an area of 13,033 km^2.

The location of Kinki Major Metropolitan Area is shown in Figure 5.3. The central cities are mainly composed of Kyoto, Osaka, Kobe, and Sakai.
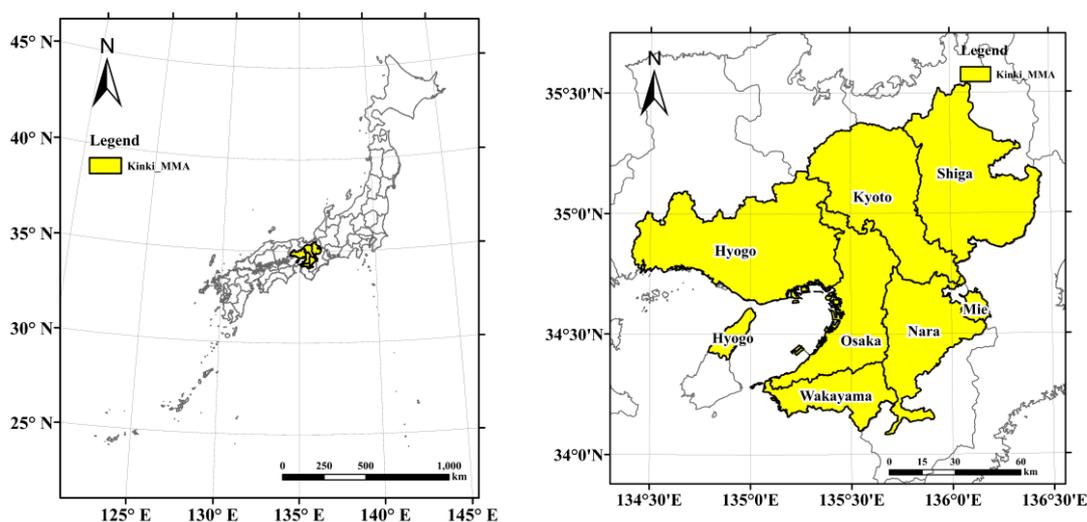
Figure 5.3. The location of Kinki Major Metropolitan Area in Japan

● Sapporo Major Metropolitan Area

Sapporo Metropolitan Area is located in the southwest of the Ishikari Plain, with a municipal area of 4,514 square kilometers. With a population of more than 2.63 million, it is the largest city in Hokkaido and the fifth most populous city in Japan [6]. It contains about half of Hokkaido's population. Although it is located in the western part of Hokkaido, it is the radiating point of Hokkaido's railways, highways, and air routes, so it is also the center of Hokkaido in terms of traffic. In 2014, the GDP of Sapporo Major Metropolitan Area was 6,547.8 billion Japanese yen, and the per capita income of citizens was 2.612 million yen [7]. One of the characteristics of Sapporo's economy is that the proportion of the secondary industry is very low, accounting for only 10.7% of the total economy. The tertiary industry has a very important position, accounting for 88% of the total economy.

Figure 5.4 shows the location of Sapporo Major Metropolitan Area in Japan. Sapporo City is the central city of Sapporo Major Metropolitan Area.
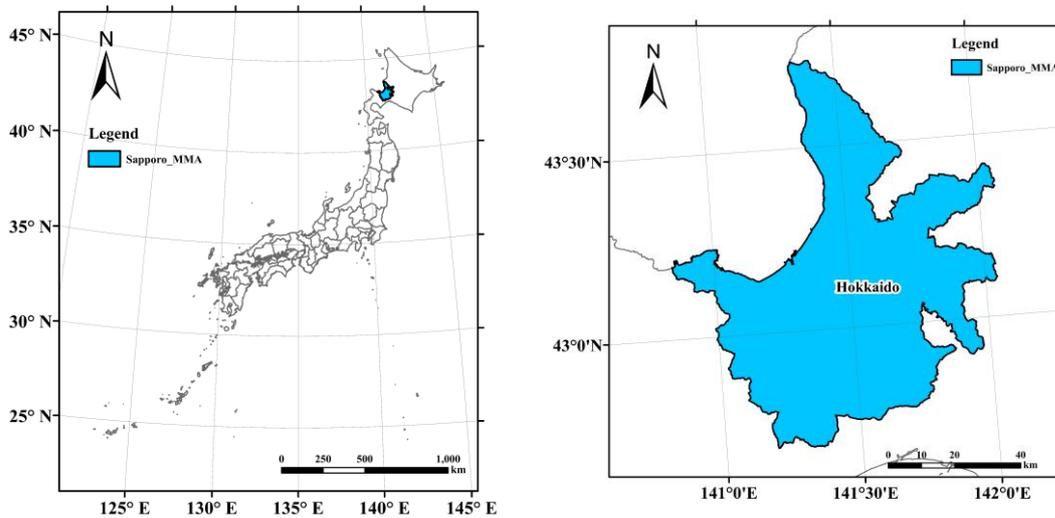
Figure 5.4. The location of Sapporo Major Metropolitan Area in Japan

● Niigata Major Metropolitan Area

Niigata Major Metropolitan Area is located in the northeastern part of Niigata Prefecture, Japan, at the estuary of the Shinano River, which is the longest river in Japan, and is the largest city on the coast in Honshu. It covers an area of 5,345 square kilometers and has a population of nearly 1.4 million.

Figure 5.5 shows the location of Niigata Major Metropolitan Area in Japan, and the central city is Niigata City.
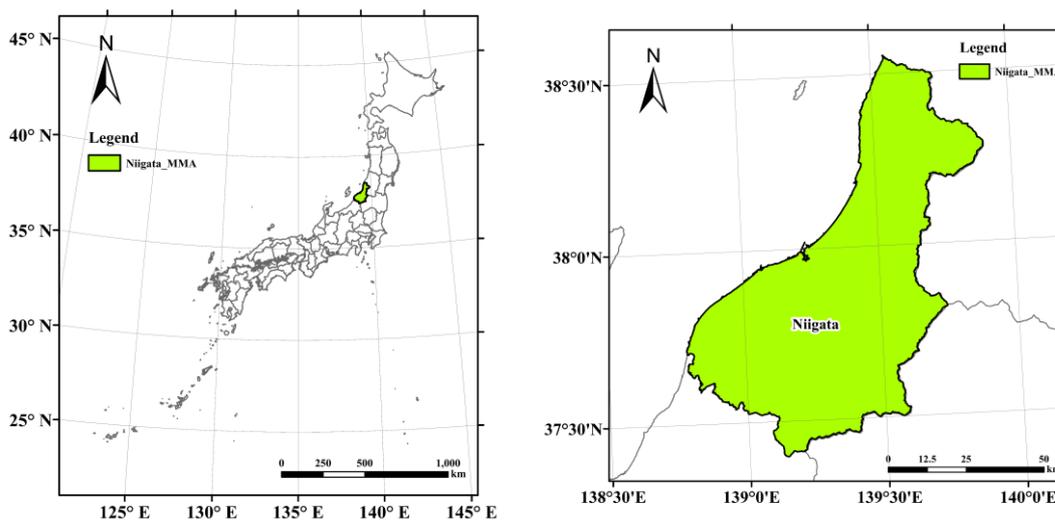
Figure 5.5. The location of Niigata Major Metropolitan Area in Japan

● Utsunomiya Metropolitan Area

Utsunomiya Metropolitan Area is located in the central part of Tochigi Prefecture and is the largest city in the North Kanto with a population of approximately 1.65 million, and a population density of 304 persons per square kilometers [8]. Most commercial and business activities in Utsunomiya take place around the Utsunomiya train station [9].

Figure 5.6 shows the location of Utsunomiya Metropolitan Area whose the central city is Utsunomiya.
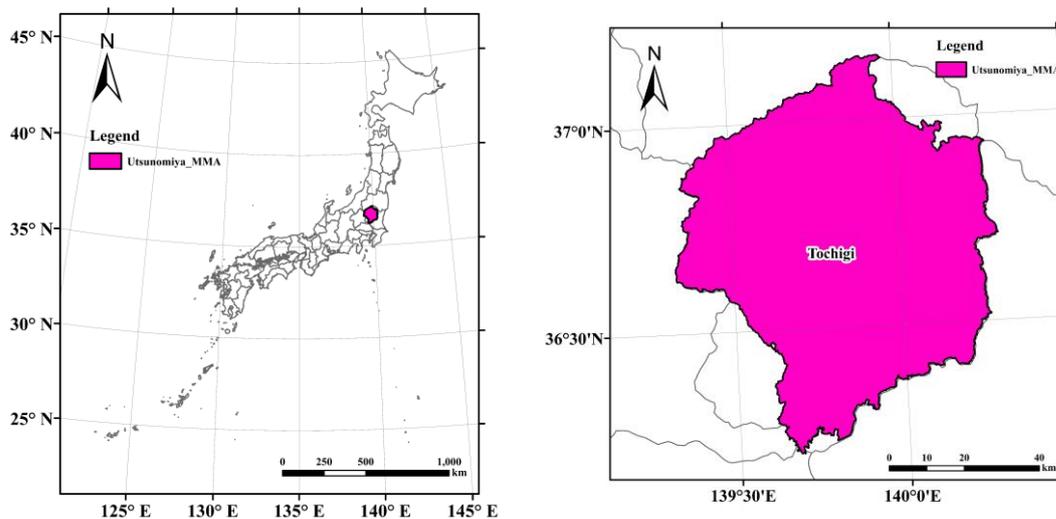
Figure 5.6. The location of Utsunomiya Major Metropolitan Area in Japan

5.1.3.   Urban Climate of Study Area

According to the meteorological data from Japan Meteorological Agency [10], we separated counted the monthly average of air temperature and precipitation of each metropolitan areas in 2010, as shown in Figure 5.7. The histogram represents the changes of air temperature of each metropolitan areas during 2010, and the line chart represents the changes of precipitation. The air temperature changes trends of study area are similar. The highest temperature is August, and the lowest temperature is January. Therefore, we selected monthly average air temperature of January and August, and annual average air temperature as the research object. Different from air temperature, the precipitation curve of each metropolitan area can be roughly divided into two categories: the same periods of high temperature and rainy season (Kanto Major Metropolitan Area, Kinki Major Metropolitan Area, and Utsunomiya Metropolitan Area) and different periods of high temperature and rainy season (Sapporo Major Metropolitan Area and Niigata Major Metropolitan Area).
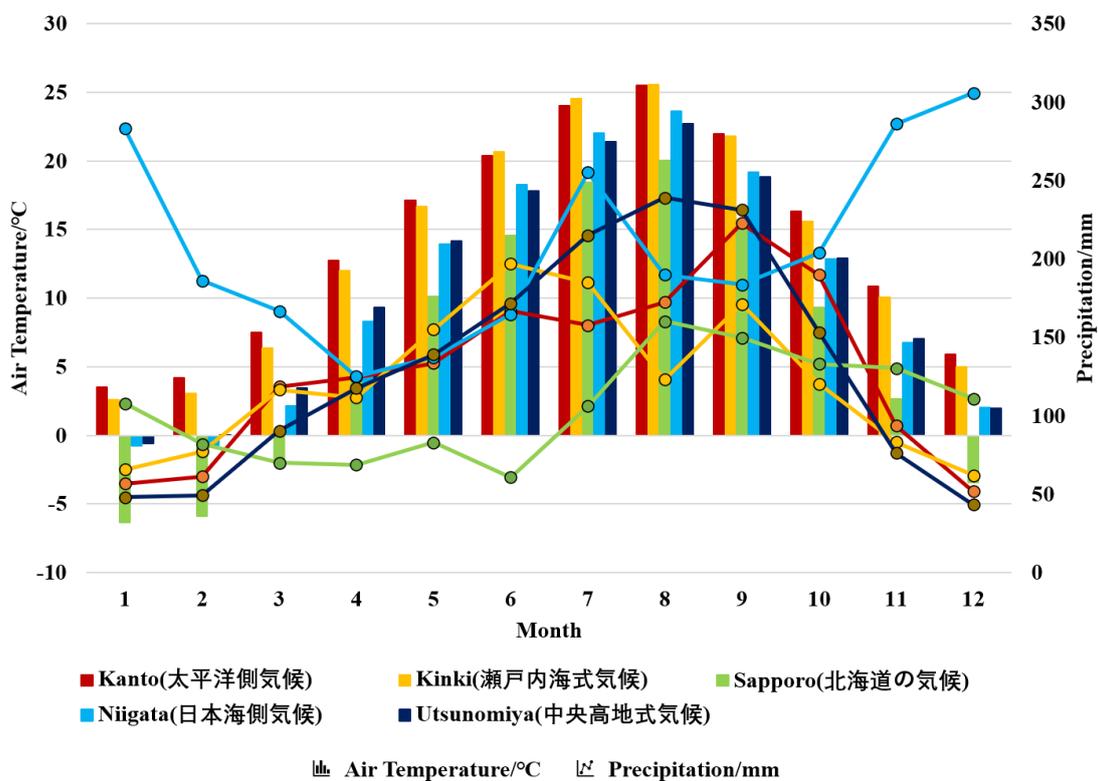
Figure 5.7. The climate characteristics of each metropolitan areas

5.2. Data Source and Research Data

5.2.1.   Data Source

According to the ET model establishment process provided in Figure 2.22, in order to build the regression model between surface parameters (including DEM and urban structure) and air temperature, we need low-resolution independent variables after resampling from high-resolution independent variables and low-resolution air temperature data. The data source for air temperature downscaling is shown in Figure 5.8.
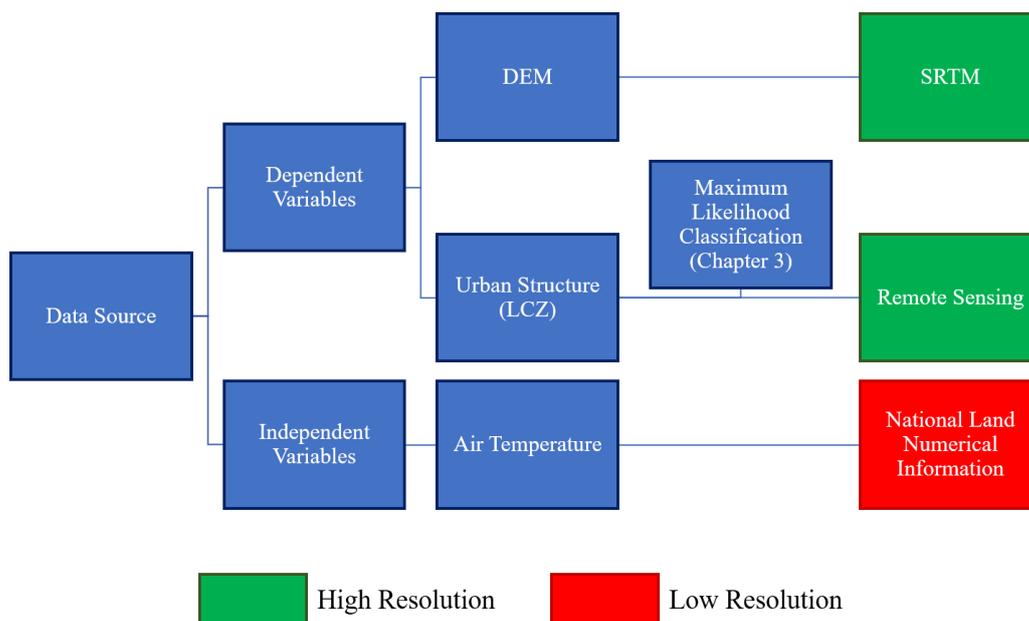
Figure 5.8. The data source of ET model establishing

As shown in Figure 5.8, the green box represents the high-resolution data with a spatial resolution of 30 m, and the orange box represents the low-resolution data with a spatial of 1 km. Data source are mainly divided into two categories, namely dependent variables, and independent variables. The independent variables are air temperature data in 2010, which can be obtained from National Land Numerical Information. And dependent variables consist of DEM and urban structure. DEM data is provided from Shuttle Radar Topography Mission (SRTM). The urban structure was classified based on remote sensing data with the help of maximum likelihood classification method introduced in Chapter 3.

In order to obtain high-resolution urban structure, and minimize the impact of time on urban structure, we need remote sensing data around 2010 to classify the urban structure. After entering the 21st century, Japan's urbanization process has slowed down, we assumed that the urban structure has not undergone major changes over time in recent years. For improving the identification accuracy of urban structure, we used Landsat 8 Operational Land Imager (OLI) and the Thermal Infrared Sensor

(TIRS) data with 11 bands. In summary, the remote sensing data provided by Landsat 8 OLI/TIRS satellite is the best choice. Landsat 8 OLI/TIRS is a low earth orbit satellite launched in February of 2013 to collect images of the earth surface [11]. The satellite collects images of the Earth with a 16-day repeat cycle. And the approximate scene size is 170 kilometers north-south by 183 kilometers east-west. Landsat 8 OLI/TIRS image data files consist of seven spectral bands [12]. Table 5-2 shows the Spectral Bands of Landsat 8 OLI/TIRS.

Table 5-2 The Spectral Bands of Landsat 8 OLI/TIRS

Sensor	Band No.	Band	Wavelength /μm	Resolution/m
OLI	1	Ultra-Blue (coastal/aerosol)	0.43-0.45	30
	2	Blue	0.45-0.51	30
	3	Green	0.53-0.59	30
	4	Red	0.64-0.67	30
	5	Near Infrared (NIR)	0.85-0.88	30
	6	Shortwave Infrared (SWIR) 1	1.57-1.65	30
	7	Shortwave Infrared (SWIR) 2	2.11-2.29	30
	8	Panchromatic	0.50-0.68	15
	9	Cirrus	1.36-1.38	30
TIRS	10	Thermal Infrared (TIRS) 1	10.60-11.19	100
	11	Thermal Infrared (TIRS) 2	11.50-12.51	100

5.2.2.   Research Data

● Digital Elevation Model Data

A digital elevation model (DEM) is a 3D computer graphics representation of altitude data to represent terrain. The Dem data used in this study is obtained from Shuttle Radar Topography Mission (SRTM). The spatial resolution of SRTM is about only 1 arc-second for global coverage (~30 meters) [13]. The DEM data of each metropolitan area is shown in Figure 5.9.
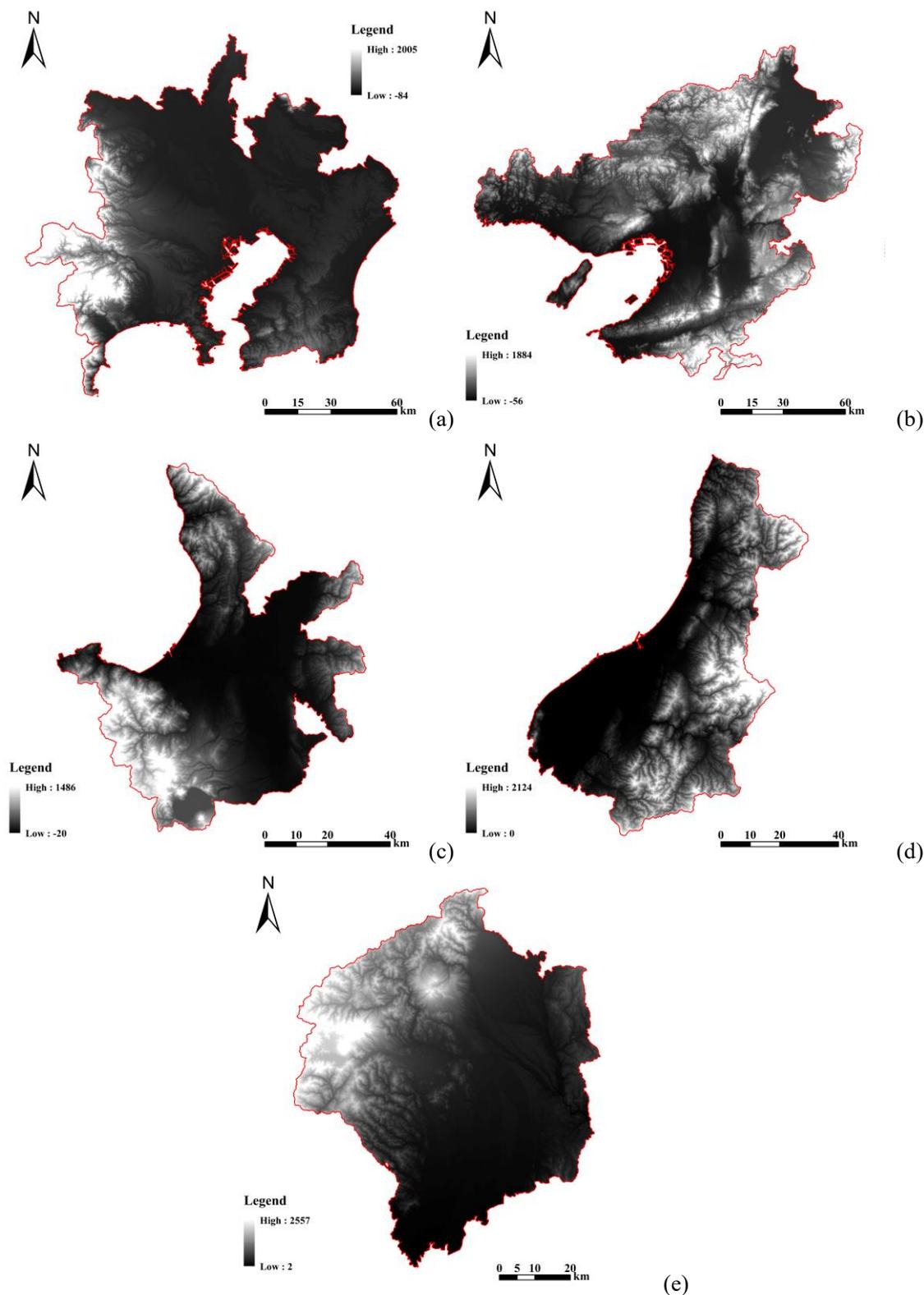
Figure 5.9. The DEM data of each metropolitan area

(a: Kanto Major Metropolitan Area; b: Kinki Major Metropolitan Area; c: Sapporo Major Metropolitan Area; d: Niigata Major Metropolitan Area; e: Utsunomiya Metropolitan Area)

● Landsat 8 OLI/TIRS Data

The remote sensing data are greatly influenced by clouds and weather. And we choose sunny days without clouds above the research region. Table 5-3 shows the basic information of Landsat 8 OLI/TIRS data. Meanwhile, in order to reduce the impact of snow on the top of the mountain on the accuracy of urban structure classification, we chose the date without snow as the original data. Then, we also need to merge and clip the original image before it can be used in the next research.

Table 5-3 Basic information of Landsat 8 OLI/TIRS data of each metropolitan area

Location	Kanto Major Metropolitan Area
Landsat Scene Identifier	LC81070352013260LGN02, LC81070362013260LGN02
WRS Path/ WRS Row	107/035,107/036
Date	2013-09-17
Land Cloud Cover	0.05%, 2.88%
Sun Elevation	51.88°, 52.95°
Sun Azimuth	148.06°, 146.60°

Location	Kinki Major Metropolitan Area
Landsat Scene Identifier	LC81100352014076LGN01, LC81100362014076LGN01
WRS Path/ WRS Row	110/035,110/036
Date	2014-03-17
Land Cloud Cover	7.46%, 2.20%
Sun Elevation	46.85°, 47.87°
Sun Azimuth	145.26°, 143.96°

Location	Sapporo Major Metropolitan Area
Landsat Scene Identifier	LC81080302015193LGN02
WRS Path/ WRS Row	108/030
Date	2015-07-12
Land Cloud Cover	4.47%
Sun Elevation	62.76°
Sun Azimuth	134.38°

Location	Niigata Major Metropolitan Area
Landsat Scene Identifier	LC81080332013155LGN01,
	LC81080342013155LGN01
WRS Path/ WRS Row	108/033,108/034
Date	2013-06-04
Land Cloud Cover	3.22%, 2.05%
Sun Elevation	66.69°, 67.36°
Sun Azimuth	129.34°, 126.09°

Location	Utsunomiya Metropolitan Area
Landsat Scene Identifier	LC81070342013260LGN02,
	LC81070352013260LGN02
WRS Path/ WRS Row	107/034,107/035
Date	2013-09-17
Land Cloud Cover	2.71%, 0.05%
Sun Elevation	50.79°, 51.88°
Sun Azimuth	149.45°, 148.06°

Location	Kanto Major Metropolitan Area
Landsat Scene Identifier	LC81070352013260LGN02,
	LC81070362013260LGN02
WRS Path/ WRS Row	107/035,107/036
Date	2013-09-17
Land Cloud Cover	0.05%, 2.88%
Sun Elevation	51.88°, 52.95°
Sun Azimuth	148.06°, 146.60°

Location	Kinki Major Metropolitan Area
Landsat Scene Identifier	LC81100352014076LGN01,
	LC81100362014076LGN01
WRS Path/ WRS Row	110/035,110/036
Date	2014-03-17
Land Cloud Cover	7.46%, 2.20%
Sun Elevation	46.85°, 47.87°
Sun Azimuth	145.26°, 143.96°

Location	Sapporo Major Metropolitan Area
Landsat Scene Identifier	LC81080302015193LGN02
WRS Path/ WRS Row	108/030
Date	2015-07-12
Land Cloud Cover	4.47%
Sun Elevation	62.76°
Sun Azimuth	134.38°

Location	Niigata Major Metropolitan Area
Landsat Scene Identifier	LC81080332013155LGN01,
	LC81080342013155LGN01
WRS Path/ WRS Row	108/033,108/034
Date	2013-06-04
Land Cloud Cover	3.22%, 2.05%
Sun Elevation	66.69°, 67.36°
Sun Azimuth	129.34°, 126.09°

Location	Utsunomiya Metropolitan Area
Landsat Scene Identifier	LC81070342013260LGN02,
	LC81070352013260LGN02
WRS Path/ WRS Row	107/034,107/035
Date	2013-09-17
Land Cloud Cover	2.71%, 0.05%
Sun Elevation	50.79°, 51.88°
Sun Azimuth	149.45°, 148.06°

● Air Temperature Data

Air temperature data of 2010 with a spatial resolution of 1 km was downloaded from Japan National Land Numerical Information (NLNI) website [14] which was provided by Ministry of Land, Infrastructure, Transport and Tourism (MLIT). The downloaded air temperature data is the shapefile of a fishnet with a side length of 1km. In order to process the air temperature data efficiently, we need to covert the polygon file to low-resolution raster data with a spatial resolution of 1 km. The processed air temperature data is shown below, including than annual average air temperature and the monthly average air temperature of January and August in 2010.
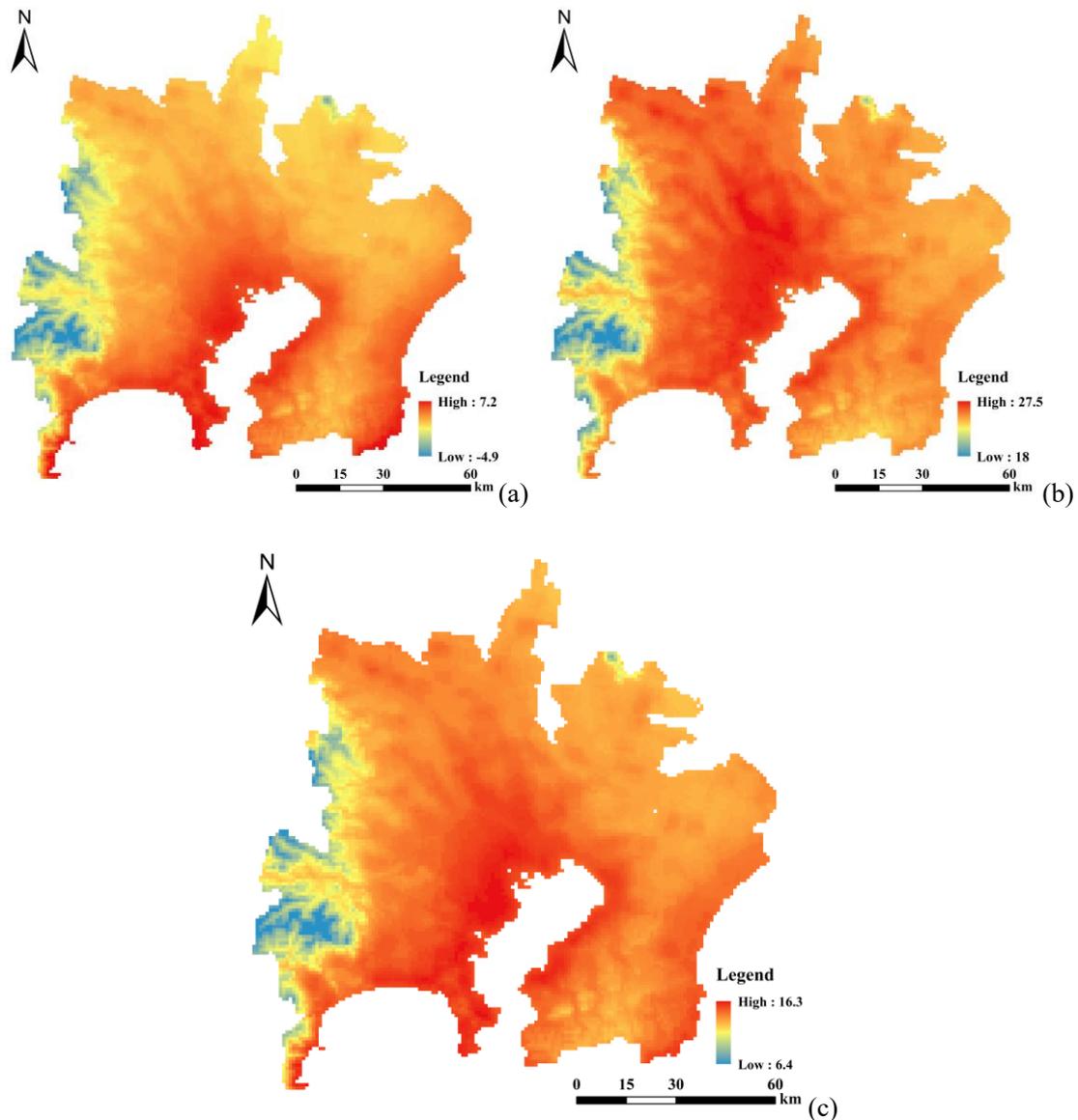
Figure 5.10. Air temperature of Kanto Major Metropolitan Area. (a) Monthly average air temperature in January; (b) Monthly average air temperature in August; (c) Annual average air temperature

Figure 5.10 shows the air temperature of Kanto Metropolitan Area. The red area represents high-temperature area, yellow area represents mid-temperature area, and blue area represents low-temperature area. The annual average temperature ranges from 6.4 °C to 16.3 °C. the high-temperature areas are mainly concentrated in the Tokyo Bay. And the air temperature ranges of monthly average air temperature in January and August are from -4.9 °C to 7.2 °C and from 18.0 °C to 27.5 °C. The high-temperature area in January is significantly smaller than that in August. The area of the low-temperature zone is roughly the same.
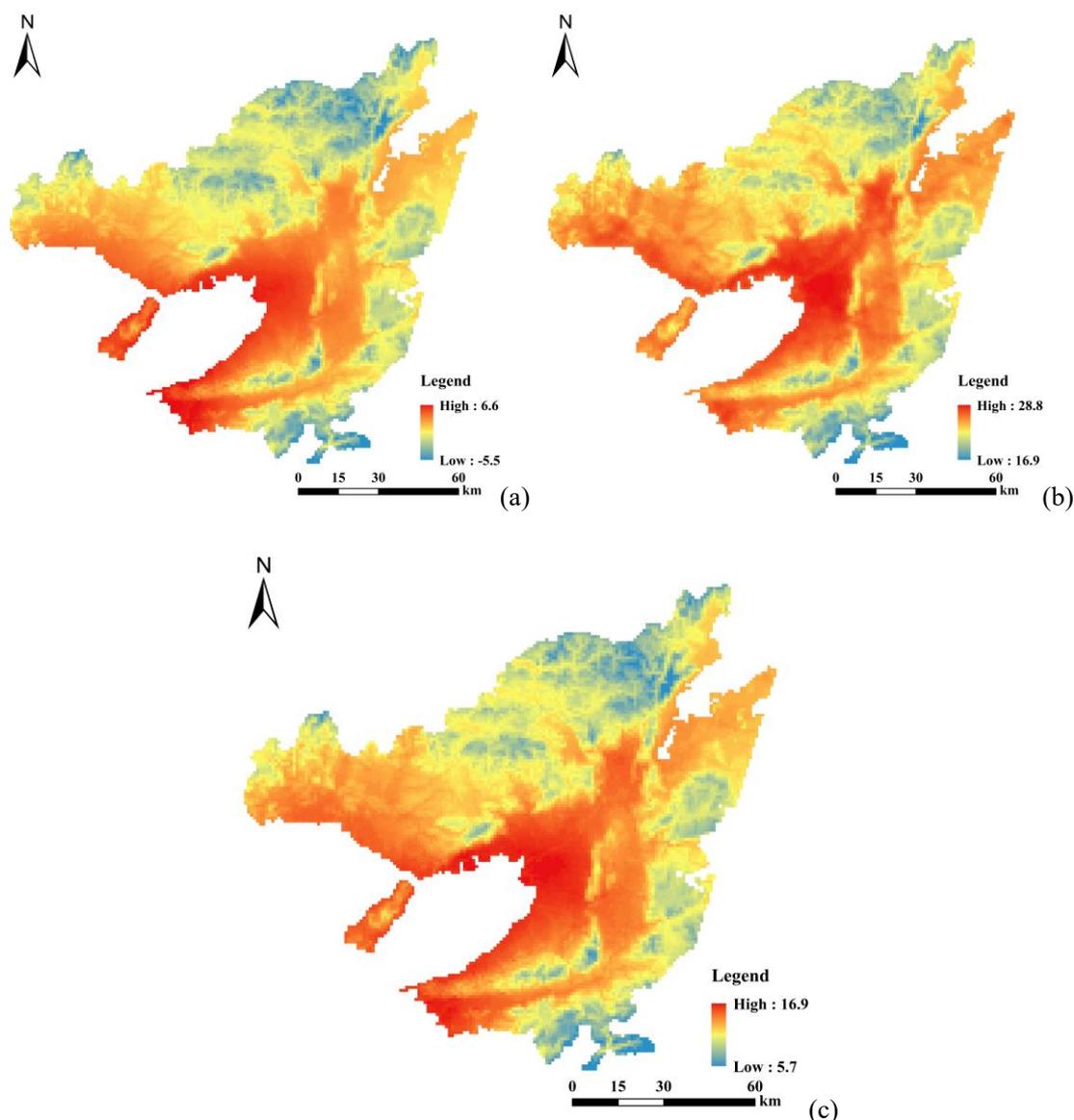
Figure 5.11. Air temperature of Kinki Major Metropolitan Area. (a) Monthly average air temperature in January; (b) Monthly average air temperature in August; (c) Annual average air temperature

Figure 5.11 shows the air temperature of Kinki Major Metropolitan Area. And the Figure 5.11a represents the annual average air temperature, Figure 5.11b and 5.11c represent the monthly average air temperature in January and August. The air temperature ranges of annual averages and monthly average in January and August are from 5.7 °C to 16.9 °C, from -5.5 °C to 6.6 °C and from 16.9 °C to 28.8 °C. Unlike the Kanto Major Metropolitan Area, the distribution of high-temperature, mid-temperature and low-temperature areas is almost similar in January and August. But the color in the high-temperature areas is obviously darker, that is, atmosphere urban heat island in August is serious.
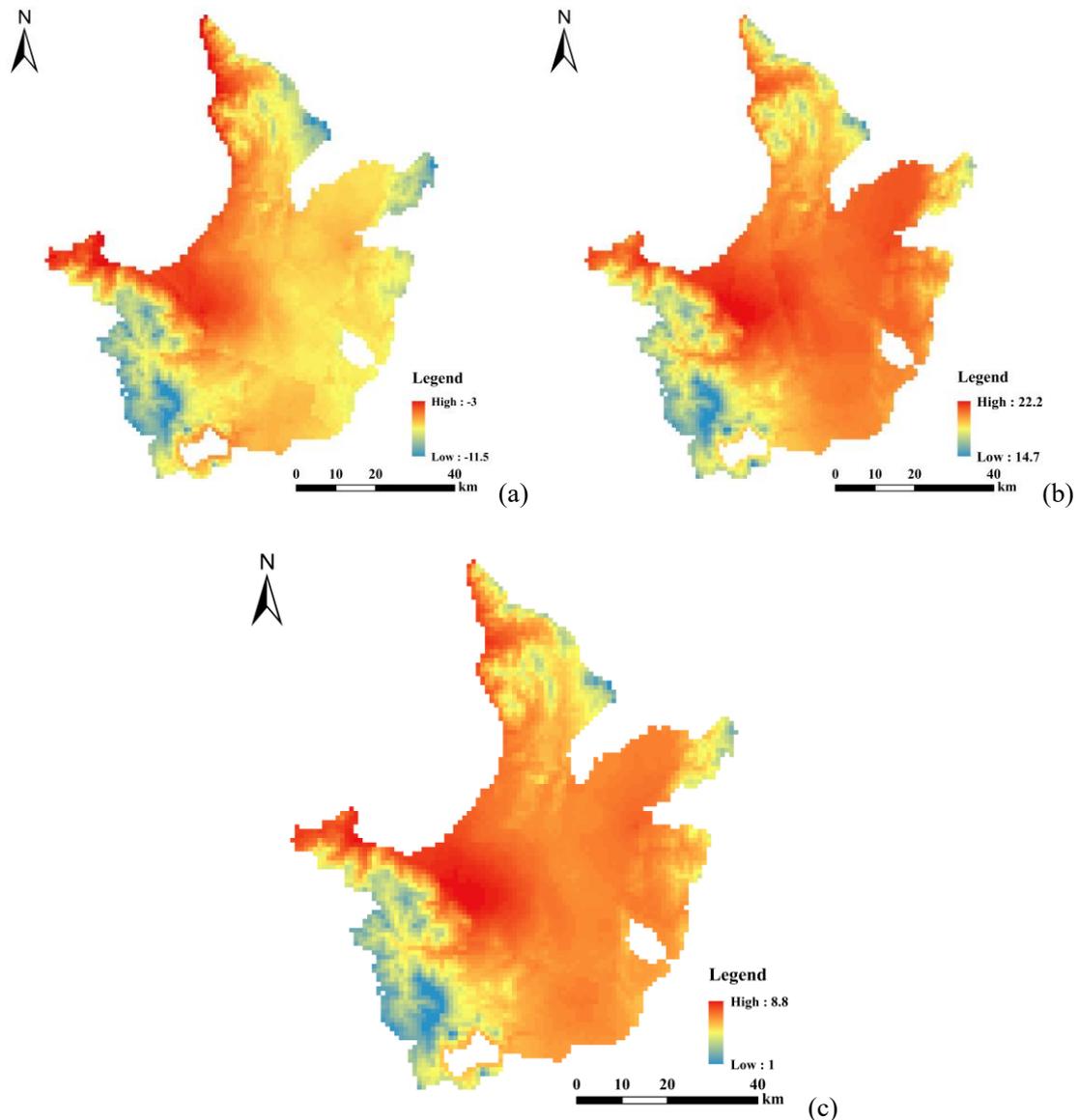
Figure 5.12. Air temperature of Sapporo Major Metropolitan Area. (a) Monthly average air temperature in January; (b) Monthly average air temperature in August; (c) Annual average air temperature

Figure 5.12 shows the air temperature distribution of Sapporo Major Metropolitan Area. The air temperature is significantly lower than that of other metropolitan areas. The annual air temperature is only from 1.0 °C to 8.8 °C, and the range in January is even lower than -11.5 °C to -3.0 °C. Compared to the others, the air temperature in August is similar with a range from 14.7 °C to 22.2 °C. And there are obviously more high-temperature areas in August than in January. The air temperature in the urban center in January is higher than that in the surround suburb plains.
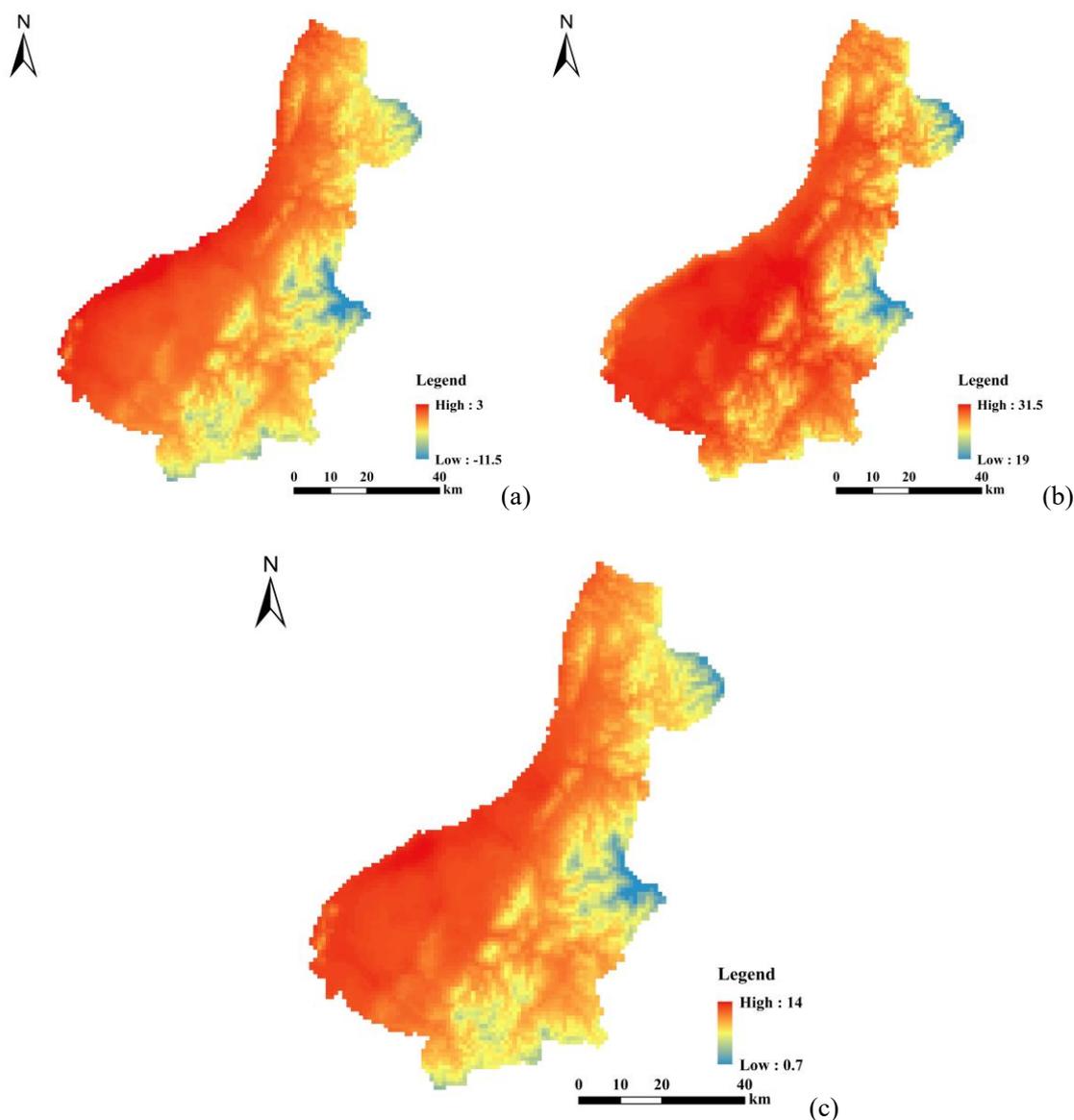
Figure 5.13. Air temperature of Niigata Major Metropolitan Area. (a) Monthly average air temperature in January; (b) Monthly average air temperature in August; (c) Annual average air temperature

Figure 5.13 represents the air temperature of Niigata Major Metropolitan Area. The ranges of annual average is from 0.7 °C to 14 °C, and the ranges in January and August are from -11.5 °C to 3.0 °C and from 19.0 °C to 31.5 °C.
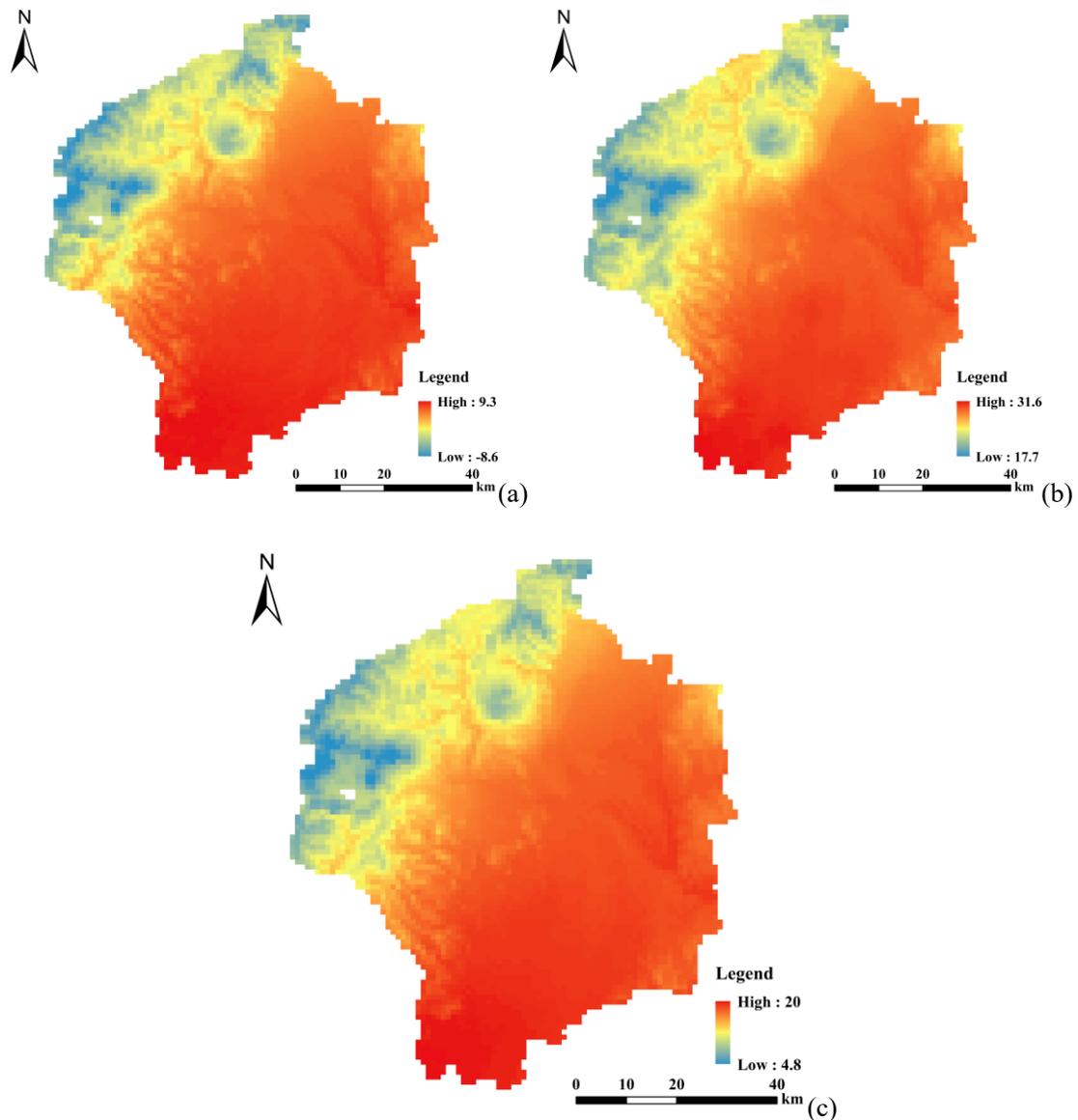
Figure 5.14. Air temperature of Utsunomiya Metropolitan Area. (a) Monthly average air temperature in January; (b) Monthly average air temperature in August; (c) Annual average air temperature

Figure 5.14 shows the air temperature of Utsunomiya Metropolitan Area. The maximum, minimum, average air temperature and difference of each metropolitan area are shown in Table 5-4. The air temperature difference in January is higher than that in August. And the difference of Utsunomiya Metropolitan Area is the largest and of Sapporo Major Metropolitan Area is the smallest.

Table 5-4 The maximum, minimum, average air temperature, and air temperature difference of each metropolitan area

Location	Period	Min	Max	Average	Difference
Kanto	Annual	6.4	16.3	14.2	9.9
	January	-4.9	7.2	3.5	12.1
	August	18.0	27.5	25.6	9.5
Kinki	Annual	5.7	16.9	13.7	11.2
	January	-5.5	6.6	2.6	12.1
	August	16.9	28.8	25.6	11.9
Sapporo	Annual	1.0	8.8	6.5	7.8
	January	-11.5	-3.0	-6.3	8.5
	August	14.7	22.2	20.0	7.5
Niigata	Annual	0.7	14.0	10.7	13.3
	January	-11.5	3.0	-0.7	14.5
	August	19.0	31.5	29.0	12.5
Utsunomiya	Annual	4.8	20.0	16.2	15.2
	January	-8.6	9.3	4.9	17.9
	August	17.7	31.6	27.8	13.9

5.3. Urban Structure Identification of Study Area

We used the method introduced in Chapter 3 to identify the urban structure of the five metropolitan areas based on remote sensing data. The verification of the error matrix shows that the maximum likelihood classification method can be applied in urban structure identification with high accuracy, which can be used in this section.

In this section, combining with the Landsat 8 OLI/TIRS data mentioned in Section 5.2.2, we identified the urban structure of each metropolitan area. As in Chapter 3, we still used the local climate zone to divide the urban structure. Metropolitan areas will be divided into two types, including the built types and land cover types [15]. Built types include ten parts, namely compact high-rise buildings (LCZ 1), compact mid-rise buildings (LCZ 2), compact low-rise buildings (LCZ 3), open high-rise buildings (LCZ 4), open mid-rise buildings (LCZ 5), open low-rise buildings (LCZ 6), lightweight low-rise buildings (LCZ 7), large low-rise buildings (LCZ 8), sparsely built (LCZ 9), and heavy industry (LCZ 10); land cover types consist of seven parts, namely dense trees (LCZ A), scattered trees (LCZ B), bush and scrub (LCZ C), low plants (LCZ D), bare rock or paved (LCZ E), bare soil or sand (LCZ F), and water (LCZ G). The urban structure identification results are shown in the figure below.
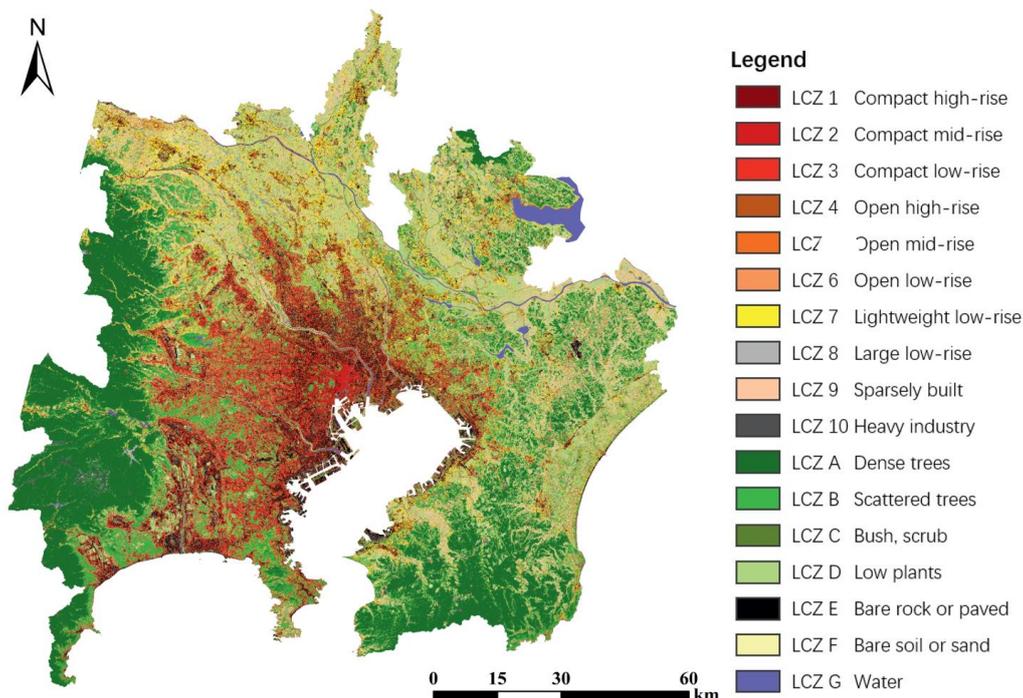
Figure 5.15. The urban structure of Kanto Major Metropolitan Area

Figure 5.15 shows the urban structure of Kanto Major Metropolitan Area. The urban area, that is,

the built type, are almost concentrated on the coast of Tokyo Bay, especially on the west coast of Tokyo Bay. And on the west side of the metropolitan area and the south side of Boso Peninsula is mainly vegetation area. There are large tracts of plains with few houses on the surrounding area of the urban area. Overall, the urban area of the Kanto Metropolitan Area obviously occupies a large part.
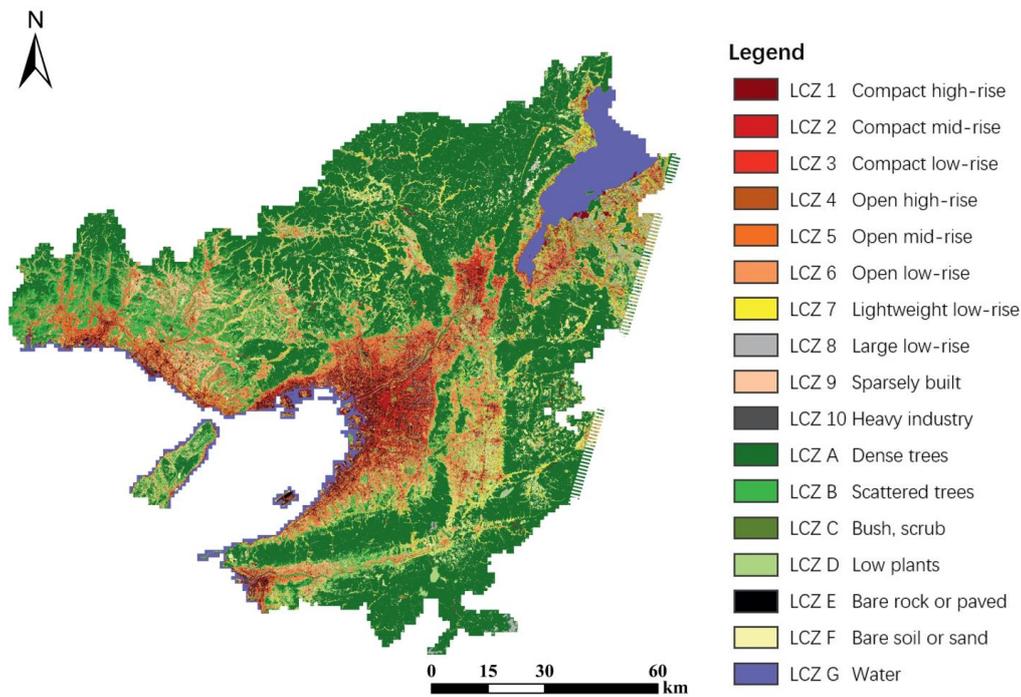
Figure 5.16. The urban structure of Kinki Major Metropolitan Area

Figure 5.16 shows the urban structure of Kinki Major Metropolitan Area. Similar to the Kanto Metropolitan Area, the urban is mainly concentrated along the bay, which is Osaka Bay, but the urban area of Kinki Major Metropolitan Area is more concentrated, mainly on the northeastern coast of Osaka Bay. There is also a relatively concentrated and obvious urban area on the southwest side of Biwa Lake (which is a large area of water surface on the northeast side of the Kinki Major Metropolitan Area), that is Kyoto City. Compared to Kanto Major Metropolitan Area, Kinki Metropolitan Area has significantly more vegetation areas.

Figure 5.17 shows the urban structure of Sapporo Major Metropolitan Area. The main urban area is located between Ishikari River and Mount Moiwa. According to that scale bar, the area of the Sapporo Major Metropolitan Area is much smaller than the Kanto Major Metropolitan and Kinki Metropolitan Area, and is similar to the Niigata Major Metropolitan Area, and slightly larger than the Utsunomiya Metropolitan Area. The Ishikari Plain occupies a large area of the entire metropolitan area. Surrounded by vegetation.
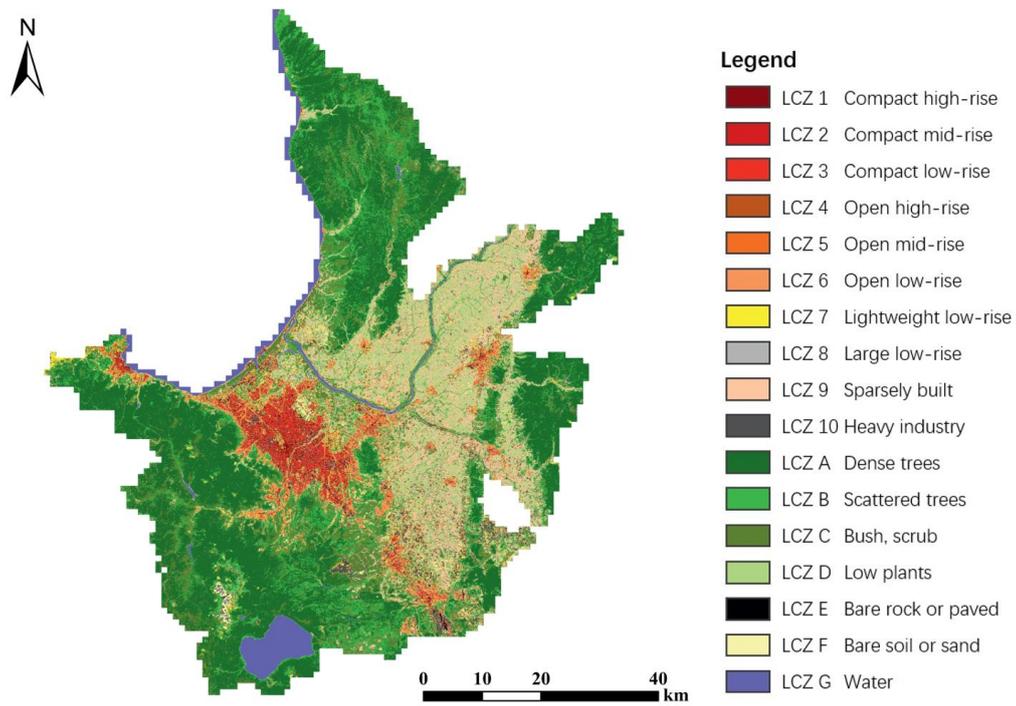
Figure 5.17. The urban structure of Sapporo Major Metropolitan Area
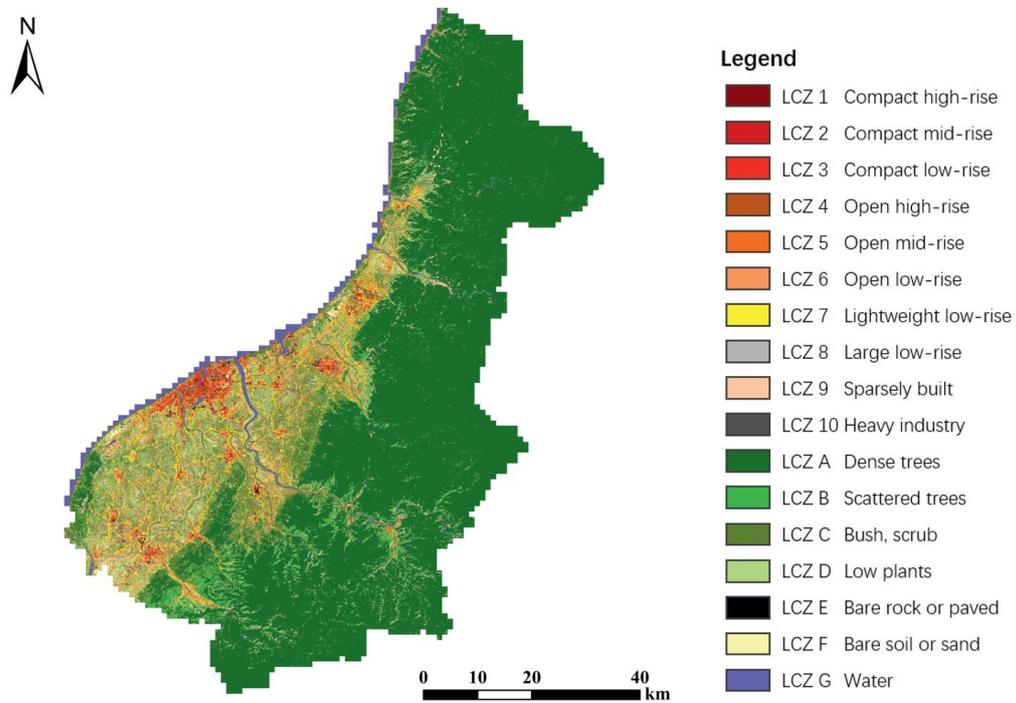
Figure 5.18. The urban structure of Niigata Major Metropolitan Area

Figure 5.18 shows the urban structure of Niigata Major Metropolitan Area. It can be clearly found the vegetation covers more than half of the metropolitan area. And the urban area only occupies a small part.

Figure 5.19 shows the urban structure of Utsunomiya Metropolitan Area. The vegetation covers a large area, and the urban area is also small. According to the results of urban identification, we have calculated the proportion of each urban structure in different metropolitan areas, as shown in Table 5-5. Blue represents the lowest ratio, and the red represents the highest proportion.
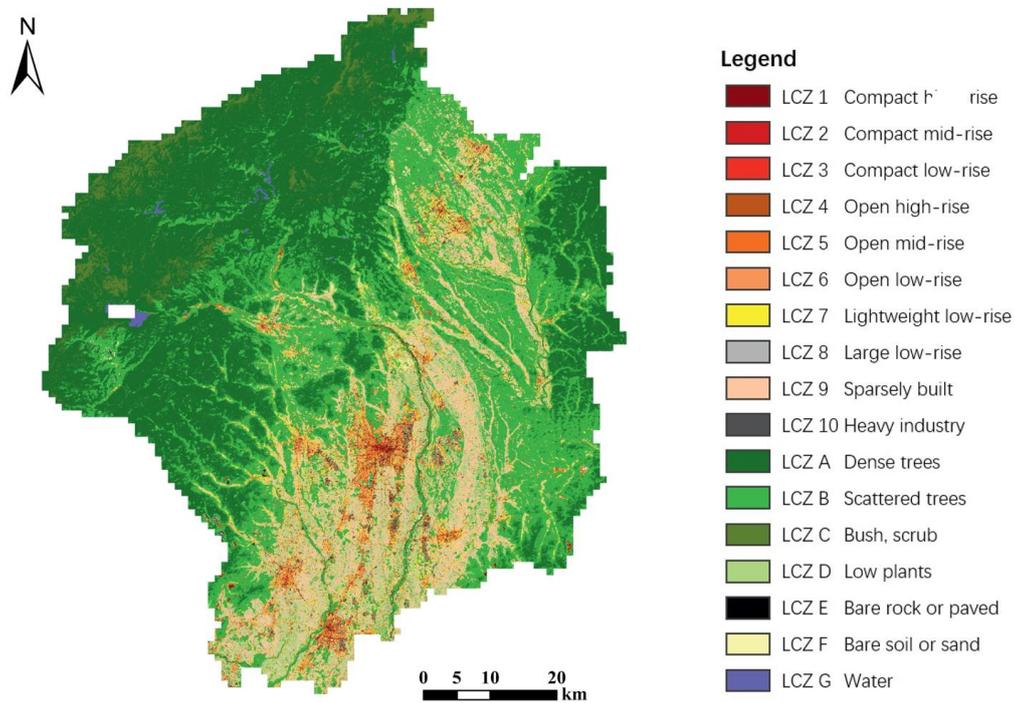
Figure 5.19. The urban structure of Utsunomiya Metropolitan Area

Table 5-5 The proportion of each urban structure types in different metropolitan areas (%)

LCZ ID	1	2	3	4	5	6	7	8	9	10	A	B	C	D	E	F	G
Kanto	0.4	2.3	5.8	2.6	2.5	4.6	8.8	0.6	9.9	3.7	19.0	14.7	4.8	11.4	5.5	2.2	1.3
Kinki	0.4	1.8	3.2	1.8	3.5	5.1	4.0	1.8	3.2	1.0	44.2	8.1	3.3	6.9	1.8	4.3	5.5
Sapporo	0.1	0.3	0.7	0.6	1.0	4.7	3.7	0.3	4.8	1.0	59.9	5.0	10.5	3.7	0.3	1.3	2.1
Niigata	0.2	0.5	1.8	0.4	1.2	2.8	3.0	0.4	10.4	3.6	32.4	14.0	13.6	10.2	0.9	1.6	3.4
Utsunomiya	0.1	0.1	0.4	0.9	1.5	1.6	4.5	0.2	8.8	0.6	36.0	33.0	4.2	5.5	0.8	1.9	0.2

As the Table 5-5 shows that LCZ A is the largest in each metropolitan area. Among them, in Sapporo Major Metropolitan Area, LCZ A accounts for nearly 60%, while the Kanto Major Metropolitan Area, which has the least ratio, accounted for 19%. For all urban structures, the proportion of LCZ 1 is almost the least among all metropolitan areas, accounting for only 0.1% to 0.4%. Comparing LCZ 1 to LCZ 6, no matter which metropolitan area, the proportion decreases with the increase of building height, and decreases with the increase of building density.

Based on the urban structure of each metropolitan area obtained in this section, combined with the DEM data in Section 5.2.2, we used the extra trees model (mentioned in Chapter 2) to downscale the low-resolution air temperature data from 1km to 250 m. The results are shown in the next section.

5.4. Downscale Air Temperature based on Extra Trees Model

5.4.1.　　Accuracy Analysis of Extra Trees Model

According to the method of air temperature downscaling introduced in Chapter 2, the core of downscaling is to establish a regression model between the independent variables (DEM and urban structure) and the dependent variable (air temperature). In order to be able to establish a regression model, we created the fishnet with a side length of 1 km according to the scope of each metropolitan area. Then, we separately counted the value of annual average air temperature of 2010, monthly average air temperature of January and August in 2010, and the average value of DEM in each grid, and calculated the area proportion of each LCZ types in each grid. With the help of statistical data ode independent variables and dependent variable, we used the scikit learning model libraries based on Python 3.8 to establish Extra Trees Model for each metropolitan area and various period. Through parameter tunning we have establish the optimal model for air temperature downscaling. The parameters settings are shown in Table 5-6.

Table 5-6 The parameters settings of each optimal models for each metropolitan area and various period.

Location	Period	Parameter Name			
		n_estimators	max_features	max_depth	min_samples_leaf
Range		1,3,5,7…99	None, Auto	0,1,2,3…30	0,1,2,3…30
Kanto	Annual	30	None	19	9
	January	38	None	20	7
	August	21	None	14	8
Kinki	Annual	41	None	15	10
	January	63	None	17	11
	August	71	None	13	8
Sapporo	Annual	31	None	14	7
	January	21	None	13	6
	August	37	None	13	9
Niigata	Annual	31	None	9	7
	January	21	None	10	8
	August	27	None	10	7
Utsunomiya	Annual	29	None	11	7
	January	17	None	11	8
	August	21	None	10	6

Through the optimal Extra Trees Models, we predicted the low-resolution air temperature data with a spatial resolution of 1 km. This step can detect the accuracy of downscaling model while preparing for the high-resolution residual fitting. The cross-validation scores of each optimal model for each metropolitan area and various period is shown in Figure 5.20. The green bar chart represents the cross-validation scores of annual average air temperature in 2010; the blue bar chart represents the January; and red bar chart represents the August.
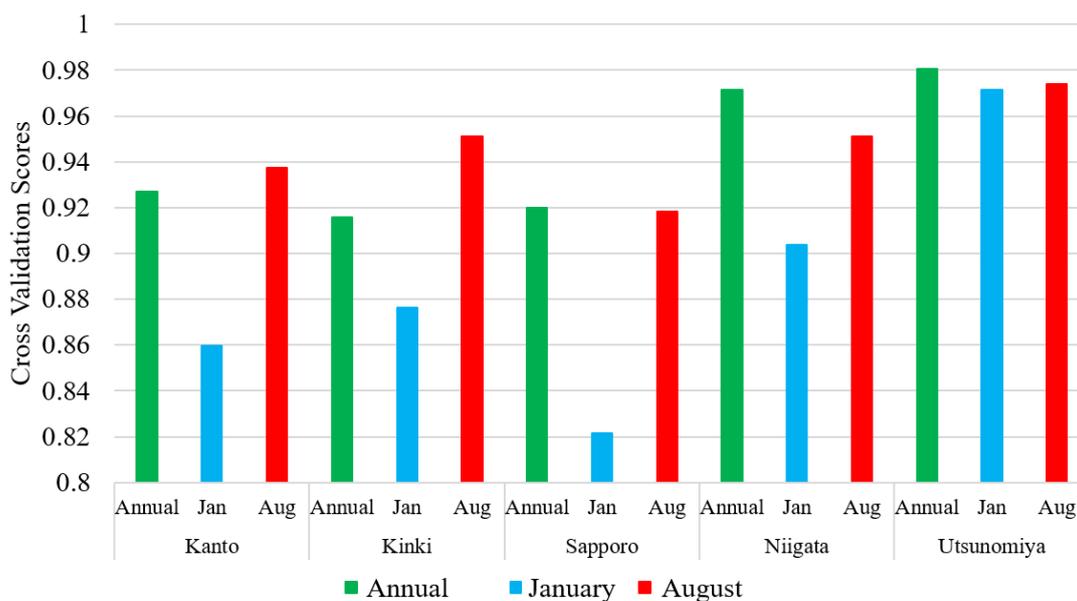
Figure 5.20. The cross-validation scores of each optimal model for each metropolitan area and various period

As shown in Figure 5.20, the cross-validation scores are all higher than 0.82, which meets the accuracy for air temperature downscaling. It is obvious that the prediction accuracy of the regression model in January is significantly lower than that in August. The cross-validation scores of the prediction models in August are all over 0.9. For comparison, only three metropolitan areas with cross-validation scores higher than 0.9 in January area 0.9038 for Niigata Major Metropolitan Area and 0.9713 for Utsunomiya Metropolitan Area. The Sapporo Major Metropolitan Area has the lowest cross-validation score in January, only 0.8126. In general, the prediction effect of the Extra Trees Models for Sapporo Major Metropolitan Area is the worst among the five metropolitan areas, while the prediction accuracy of the Extra Model for Utsunomiya Metropolitan Area is the highest, with all three values exceeding 0.97.

Using the above optimal models, combined with high-resolution DEM and urban structure data with a spatial resolution of 250 m, we predicted high-resolution air temperature data, and then after residual fitting, the downscaled air temperature data with a resolution of 250 m is as follows.

5.4.2.   Downscaled Air Temperature from 1 km to 250 m

Figure 5.20 shows the downscaled air temperature of Kanto Major Metropolitan Area with a spatial resolution of 250 m. Compared to the original air temperature with a spatial resolution of 1 km, the ranges of downscaled air temperature are larger. Among them, the air temperature difference in January changed more, from 12.1 ℃ to 13.3 ℃, while in August, it only increases from 9.5 ℃ to 10 ℃. The air temperature distribution is basically the same as that of the low-resolution data, but the air temperature of the river in urban center is significantly lower than that of surrounding area.
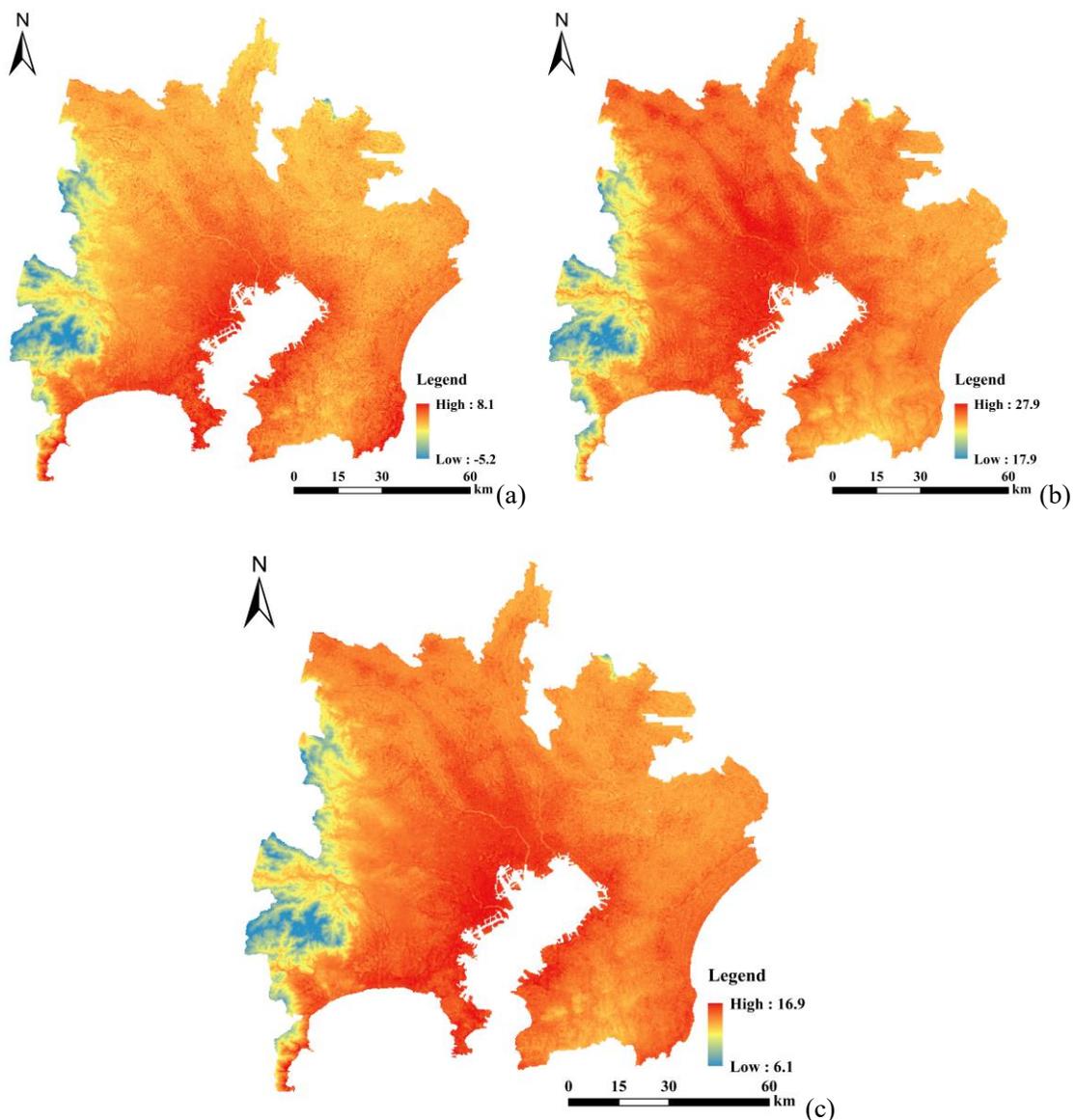
Figure 5.21. Downloaded air temperature data of Kanto Major Metropolitan Area. (a) Monthly average air temperature in January; (b) Monthly average air temperature in August; (c) Annual average air temperature
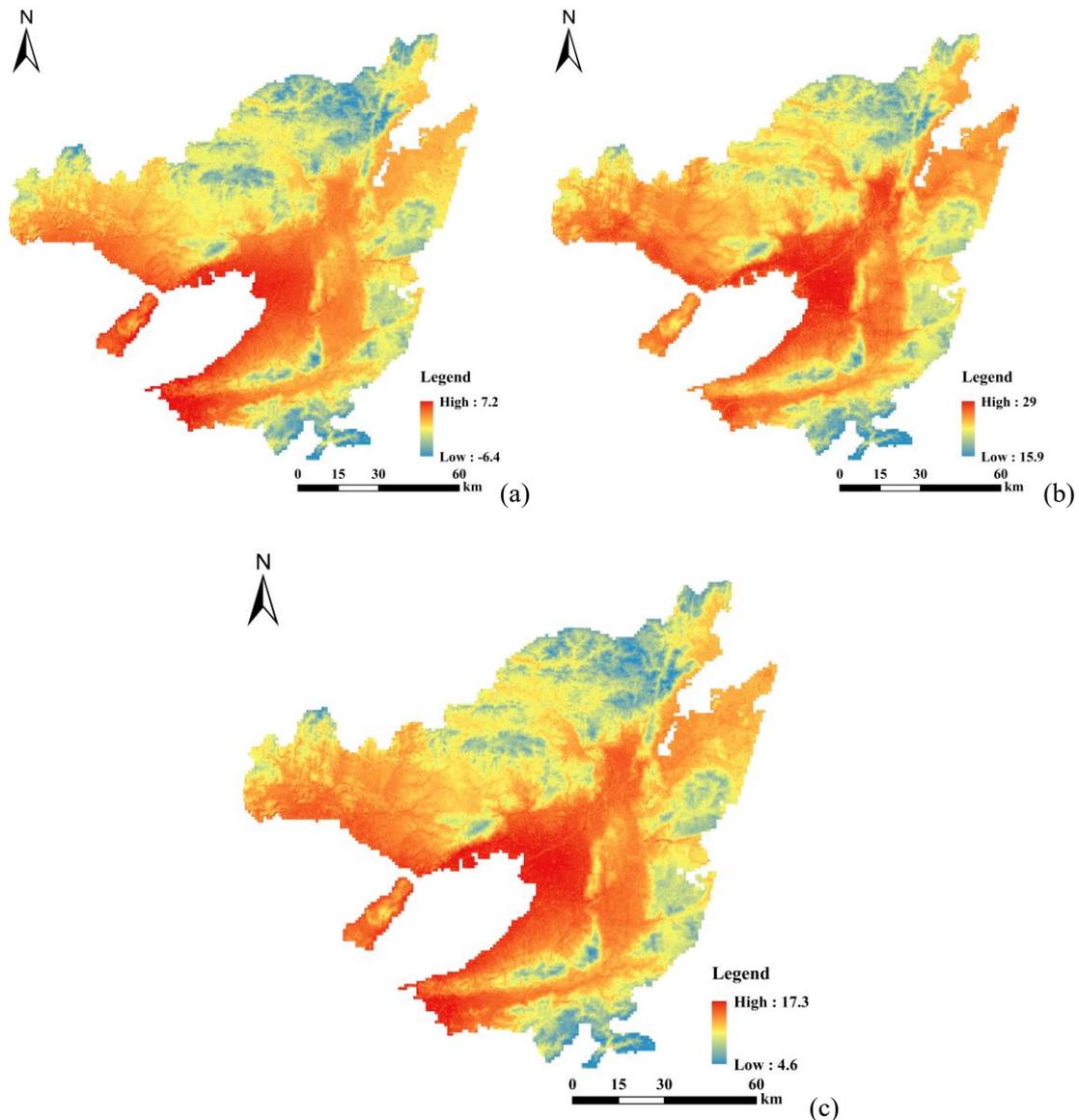
Figure 5.22. Downscaled air temperature of Kinki Major Metropolitan Area. (a) Monthly average air temperature in January; (b) Monthly average air temperature in August; (c) Annual average air temperature

Figure 5.22 shows the downscaled air temperature of Kinki Major Metropolitan Area. Similar to Kanto Major Metropolitan Area, the difference of downloaded air temperature is slightly larger than that of original air temperature. The changes between January and August are similar. The air temperature difference in January increases from 12.1 ℃ to 13.6 ℃, and in August it increases from 11.9 ℃ to 13.1 ℃. And the change of air temperature difference of annual average value is 1.5 ℃, from 11.2 ℃ to 12.7 ℃. From the perspective of air temperature distribution, high-resolution air temperature data can more accurately show the air temperature distribution.
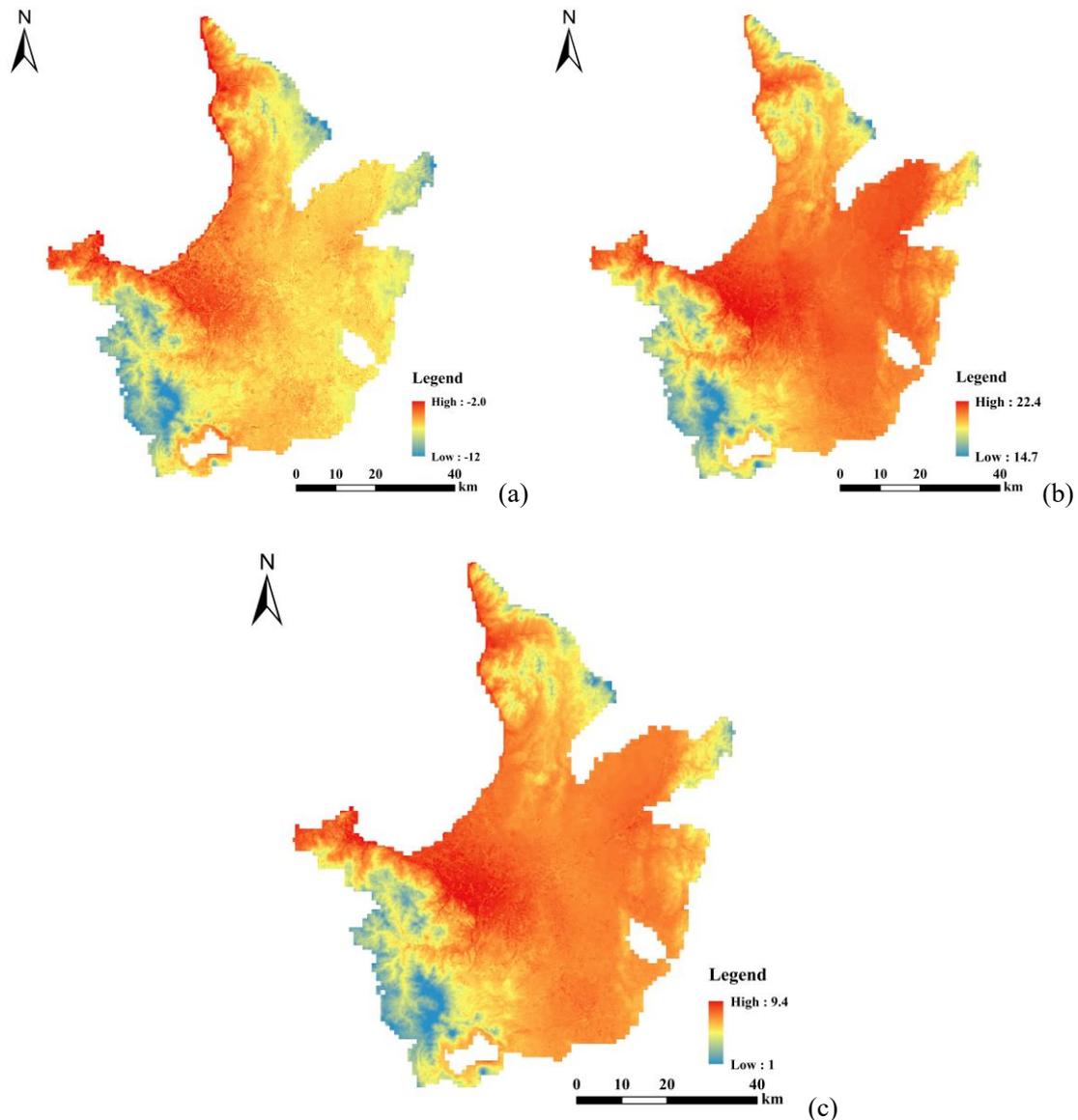
Figure 5.23. Downscaled air temperature of Sapporo Major Metropolitan Area. (a) Monthly average air temperature in January; (b) Monthly average air temperature in August; (c) Annual average air temperature

Figure 5.23 shows the downloaded air temperature of Sapporo Major Metropolitan Area. According to Figure 5.23a and Figure 5.23b, the urban heat island in January is more obvious than that in August. Compared to August, the air temperature difference in January is also larger, at 10 ℃, while the air temperature difference in August is only 7.7 ℃. And no matter in January or August, compared with Kanto Major Metropolitan Area and Kinki Major Metropolitan Area, the cooling effect of the river is not so obvious. There is no clear low temperature zone at the river location.
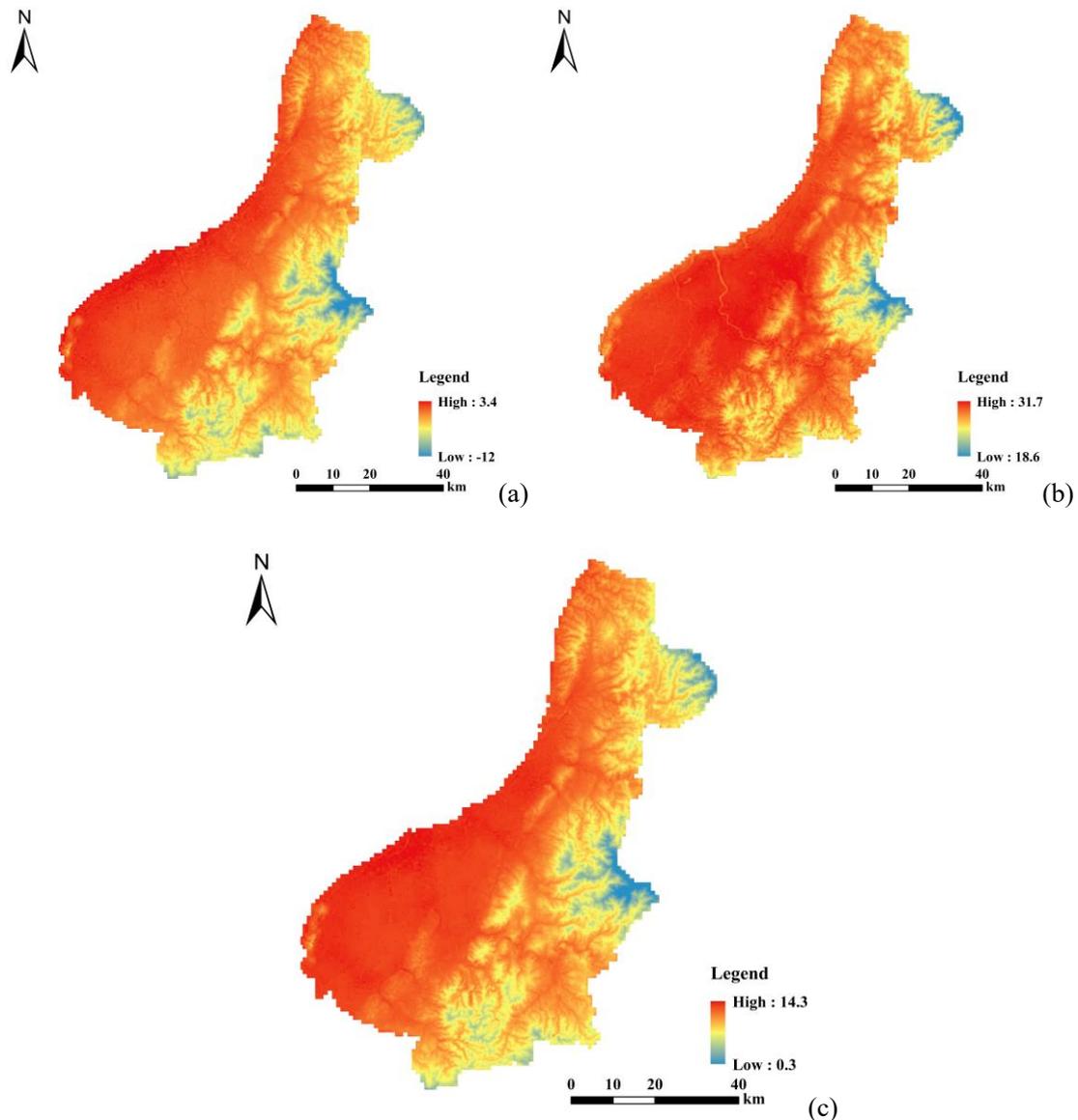
Figure 5.24. Downscaled air temperature of Niigata Major Metropolitan Area. (a) Monthly average air temperature in January; (b) Monthly average air temperature in August; (c) Annual average air temperature

Figure 5.24 shows the downscaled air temperature of Niigata Major Metropolitan Area. Based on the high-resolution air temperature data, it can be clearly seen that the air temperature distribution between January and August is various. The high temperature area in January is distributed along the coast of the Sea of Japan, while the high temperature area in August is distributed in the urban areas in the plain. In addition, the air temperature difference in the Niigata Major Metropolitan Area in January is 15.4 °C, which is significantly higher than that in August. The air temperature difference in August is only 12.5 °C.
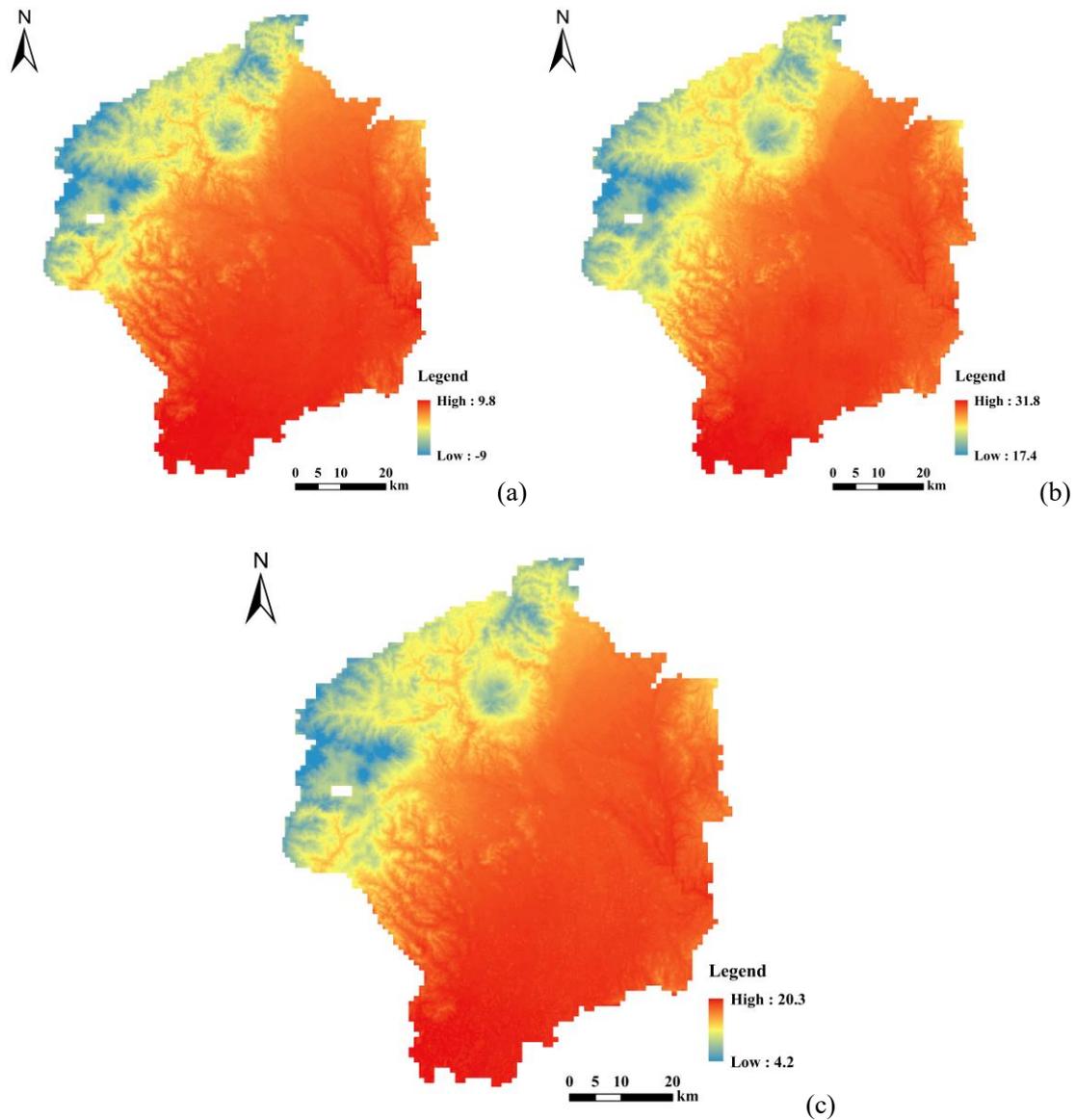
Figure 5.25. Downscaled air temperature of Utsunomiya Metropolitan Area. (a) Monthly average air temperature in January; (b) Monthly average air temperature in August; (c) Annual average air temperature

Figure 5.25 shows the downscaled air temperature of Utsunomiya Metropolitan Area. Since the southern side of Utsunomiya Metropolitan Area is adjacent to Kanto Major Metropolitan Area, the air temperature on the south side is significantly higher than the other areas.

5.5. Summary

In this chapter, we selected the five metropolitan areas which is the largest in each climatic zone in Japan. The downscaling method introduced in the previous chapter was applied to downscaled air temperature, including the monthly average in January and August, and the annual average in 2010, which downloaded from Japan National Land Numerical Information, from 1 km to 250 m. After cross-validation, we concluded that the downscaling accuracy meet the requirements, and the downscaled high-resolution air temperature can be used for the further atmosphere urban heat island analysis.

Reference

1.   Island Countries Of The World - WorldAtlas Available online:
        https://www.worldatlas.com/articles/which-are-the-island-countries-of-the-world.html (accessed
        on May 3, 2021).

2.   Squires, G.D. *Urban sprawl: Causes, consequences, & policy responses*; The Urban Insitute,
        2002; ISBN 0877667098.

3.   Muro, M.; Katz, B.; Rahman, S.; Warren, D. MetroPolicy: Shaping a New Federal Partnership for
        a Metropolitan Nation. **2008**.

4.   World, D.; Areas, U. *Definition of Urban Terms*;

5.   Population Census 2015 Population Census Major Metropolitan Area and Metropolitan Area |
        File | Browse Statistics | Portal Site of Official Statistics of Japan Available online:
        https://www.e-stat.go.jp/en/stat-
        search/files?page=1&layout=datalist&toukei=00200521&tstat=000001080615&cycle=0&tclass1
        =000001110216&tclass2val=0 (accessed on May 4, 2021).

6.   Wang, R.; Murayama, Y. Geo-simulation of land use/cover scenarios and impacts on land surface
        temperature in Sapporo, Japan. *Sustain. Cities Soc.* **2020**, *63*, 102432,
        doi:10.1016/j.scs.2020.102432.

7.   Cabinet Office Home Page Available online: https://www.cao.go.jp/index-e.html (accessed on
        May 4, 2021).

8.   Information |Utsunomiya city formal Web site for foreigners Available online: https://www15.j-
        server.com/LUCUTNMY/ns/tl.cgi/https%3A//www.city.utsunomiya.tochigi.jp/about/1007962.ht
        ml?SLANG=ja&TLANG=en&XMODE=0&XCHARSET=utf-8&XJSID=0 (accessed on May 4,
        2021).

9.   Ishii, S.; Tabushi, S.; Aramaki, T.; Hanaki, K. Impact of future urban form on the potential to
        reduce greenhouse gas emissions from residential, commercial and public buildings in
        Utsunomiya, Japan. *Energy Policy* **2010**, *38*, 4888–4896, doi:10.1016/j.enpol.2009.08.022.

10.   Japan Meteorological Agency Available online: http://www.jma.go.jp/jma/index.html (accessed
        on Apr 29, 2021).

11.   Loveland, T.R.; Dwyer, J.L. Landsat: Building a strong future. *Remote Sens. Environ.* **2012**, *122*,
        22–29, doi:10.1016/j.rse.2011.09.022.

12.   USGS EROS Archive - Landsat Archives - Landsat 8 OLI (Operational Land Imager) and TIRS
        (Thermal Infrared Sensor) Level-1 Data Products Available online:
        https://www.usgs.gov/centers/eros/science/usgs-eros-archive-landsat-archives-landsat-8-oli-
        operational-land-imager-and?qt-science_center_objects=0#qt-science_center_objects (accessed
        on May 5, 2021).

13.   USGS EROS Archive - Digital Elevation - Shuttle Radar Topography Mission (SRTM) 1 Arc-

Second Global Available online: https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1-arc?qt-science_center_objects=0#qt-science_center_objects (accessed on May 4, 2021).

14. National Land Numerical Information Available online: https://nlftp.mlit.go.jp/ksj/ (accessed on Jan 31, 2021).

15. Stewart, I.D.; Oke, T.R. Local climate zones for urban temperature studies. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 1879–1900.

Chapter 6. Atmosphere Urban Heat Island Analysis of Metropolitan Area based on High-resolution Air Temperature

In order to be able to reflect the characteristic of urban heat islands in various metropolitan area and different periods objectively and quantitatively, we introduced urban heat island intensity and urban heat island ratio index to compare, analyze and evaluate the urban heat islands in each maximum area of different periods.

6.1. The Calculation of Urban Heat Island Intensity

Atmosphere urban heat island intensity is generally considered to the air temperature differences between the urban and suburbs, and it is calculated from the air temperature differences between them [1–4]. However, due to the ambiguity of the definition between urban and suburb areas, it is difficult to determine the intensity of the urban heat island.

Therefore, in this section, the densely inhabited districts and the surrounding area are used as the boundary, the densely inhabited districts are defined as the urban area, and other areas are regarded as the suburb area, and the study area is divided between urban and suburb on this basis, using both average air temperature difference as the urban heat island intensity. The equation is as follows:

$$UHII = T_{urban} - T_{suburb}$$

$$(6-1)$$

where *UHII* represents the urban heat island intensity; T_{urban} represents the average air temperature of urban area; and T_{suburb} represents the average air temperature of suburb area.
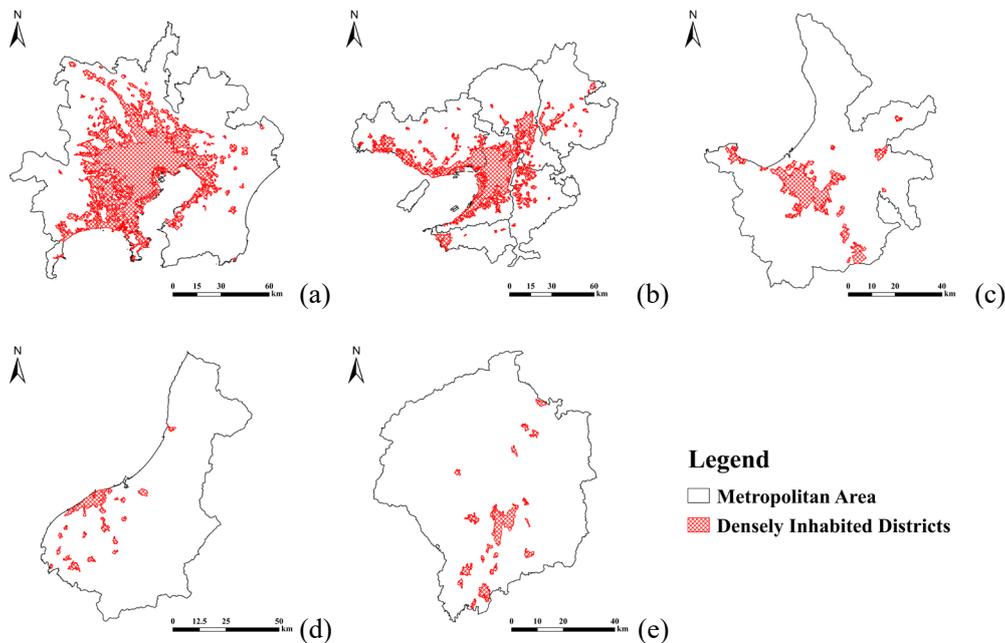
Figure 6.1. The distribution of densely inhabited districts of each metropolitan area:
(a) Kanto ;(b) Kinki ;(c) Sapporo ;(d) Niigata :(e) Utsunomiya

The densely inhabited districts were downloaded from Japan National Land Numerical Information [5] as shown in Figure 6.1. The red grid represents the densely inhabited districts. There is a large area of densely inhabited districts in Kanto Major Metropolitan Area, followed by the Kinki Major Metropolitan Area, and then is the Sapporo Major Metropolitan Area. The smallest proportions are the Niigata Major Metropolitan Area and Utsunomiya Metropolitan Area.

After statistics, we obtained the average air temperature of the urban areas and suburb areas of each metropolitan area in different periods. Based on the statistics results, we calculated the urban heat island intensity, the results are shown in Table 6-1.

Table 6-1 The urban heat island intensity

Area	Period	Urban	Suburb	Urban Heat Island Intensity
Kanto	January	4.7515	3.3028	1.4486
	August	26.5900	25.2519	1.3381
	Annual	15.2560	13.9393	1.3167
Kinki	January	4.6955	2.3603	2.3352
	August	26.6307	25.1982	1.4325
	Annual	15.6676	13.3712	2.2964
Sapporo	January	2.1060	-0.7541	2.8602
	August	30.7647	28.9641	1.8006
	Annual	13.3082	10.6546	2.6536
Niigata	January	-4.9347	-6.4214	1.4867
	August	21.2726	19.9442	1.3284
	Annual	7.8723	6.4005	1.4718
Utsunomiya	January	8.2105	5.0402	3.1703
	August	30.4433	27.8046	2.6387
	Annual	18.9705	16.2409	2.7296

According to Table 6-1, the urban heat island intensity in January is generally higher than that in August. Among them, the urban heat island intensity in Kanto Major Metropolitan Area and Niigata Major Metropolitan Area is relatively weak. The value of them is only from 1.3 ℃ to 1.5 ℃. And their difference in urban heat island intensity between January and August is also the smallest, only 0.11 ℃ and 0.15 ℃. In contrast, the urban heat island intensity of Utsunomiya Metropolitan Area is higher than the other metropolitan areas. The value in January reaches 3.17 ℃. And the value in August is 2.64 ℃, which is significantly higher than the August value of other metropolitan area, the second-ranked, Sapporo Major Metropolitan Area is only 1.80 ℃, and all others are below 1.5 ℃. However,

the difference of Utsunomiya Metropolitan Area between January and August is only 0.5 ℃. The difference between January and August of Kinki Major Metropolitan Area and Sapporo Major Metropolitan is the largest, reaching 0.9 ℃ and 1.0 ℃ respectively. According to the annual average data, Utsunomiya Metropolitan Area has the highest heat island intensity at 2.73 ℃, followed by Sapporo Major Metropolitan Area at 2.65 ℃, then the urban heat island intensity of Kinki Major Metropolitan Area is 2.30 ℃. The smallest are the Niigata Major Metropolitan Area and Kanto Major Metropolitan Area, with 1.47 ℃ and 1.32 ℃ respectively.

6.2. The Calculation of Urban Heat Island Ratio Index

In the last section, we obtained high-resolution air temperature data of different metropolitan areas in various periods. We aimed to study the temporal and spatial changes of atmosphere urban heat island in various metropolitan areas by using high-resolution air temperature data downscaled from air temperature data obtained from Japan Nation Land Numerical Information with a spatial resolution of 1 km. Then, based on the normalization of air temperature data, we introduced urban heat island ratio index to quantify the distribution of urban heat islands, specifically the area ratio of urban heat island, which can also reflect whether the urban heat island zone is concentrated or not [6,7].

However, due to the differences in the periods of the downscaled air temperature images, it is difficult to directly use the air temperature data to compare the urban heat island in different periods. In order to be able to use the air temperature in different periods to compare the urban heat islands in various metropolitan areas, we adopted the method of normalizing the air temperature and calculated the urban heat island ratio index.

Therefore, in order to avoid the error caused by the differences of periods, we normalized the downscaled air temperature, and the range of air temperature is controlled between 0 and 1. The equation is as follow:

$$T_N = \frac{T_i - T_{min}}{T_{max} - T_{min}}$$

(6-2)

where T_N represents the normalized values of air temperature; T_i represents the air temperature of pixel i; T_{max} represents the maximum air temperature; and T_{min} represents the minimum air temperature.

Through the above normalization, we used Jenks natural breaks classification method to classify the downscaled air temperature. Jenks natural breaks classification method is based on the inherent natural grouping in the original data [8,9]. The classification interval is identified, and similar values can be classified appropriately, and the difference between each group can be maximized. In other words, this method tries to reduce the variance in the groups and maximize the variance between groups [10]. This method requires an iterative process. In other words, the calculation must be repeated using different interrupts in the original data to determine which group interrupts have the smallest variance [11]. With the help of Jenks natural breaks classification method, we divided the downscaled air temperature into 7 grades, which are low temperature zone, sub low temperature zone, sub mid temperature zone, mid temperature zone, sub high temperature zone, high temperature zone, and extremely high temperature zone.

Finally, we introduced the urban heat island ratio index to indicate the intensity of the urban heat island:

$$URI = \frac{1}{100m} \sum_{i=1}^{n} w_i p_i$$

(6-2)

where URI represents urban heat island ratio index; m represents the number of normalized downscaled air temperature levels; i represents the level of high temperatures zones and extremely high temperature zones; n represents the number of higher temperature zone levels, including high temperature zone and extremely high temperature zone; w represents the weightd value, takes the level value of level i; p represents the area ratio of the level i.

Essentially, urban heat island ratio index is obtained by calculating the ratio of urban heat island zone to urban total area considering the weighted value of each level and reflects the development of urban heat island. The larger the urban heat island ratio index is, the more serious the urban heat island is.

6.2.1.   Normalize the Downscaled Air Temperature

Before the Jenks natural breaks classification, we first needed to normalize downscaled air temperature, so that the data ranges of each periods and each metropolitan area is unified between 0 and 1.

Based on the above the Jenks natural breaks classification method, the downscaled air temperature has been scaled to the range of 0 to 1. The results are shown in Figure 6.2. In Figure 6.2a, 6.2b, 6.2c, and 6.2d, the upper figure represents the normalized air temperature in January, the mid figure represents that in August; and the bottom figure represents that through the year. And in Figure 6.1e, the upper left figure represents the normalized air temperature in January; the upper right figure represents that in August; and the bottom figure represents that through the year. The color bar is from blue to yellow to red. The blue area is 0; the yellow area is 0.5; and the red area is 1. The normalized air temperature data is similar to the downscaled air temperature. Based on the normalized air temperature data, we used Jenks natural breaks classification to divide the metropolitan area into seven levels.

(a)   (b)

(c)   (d)
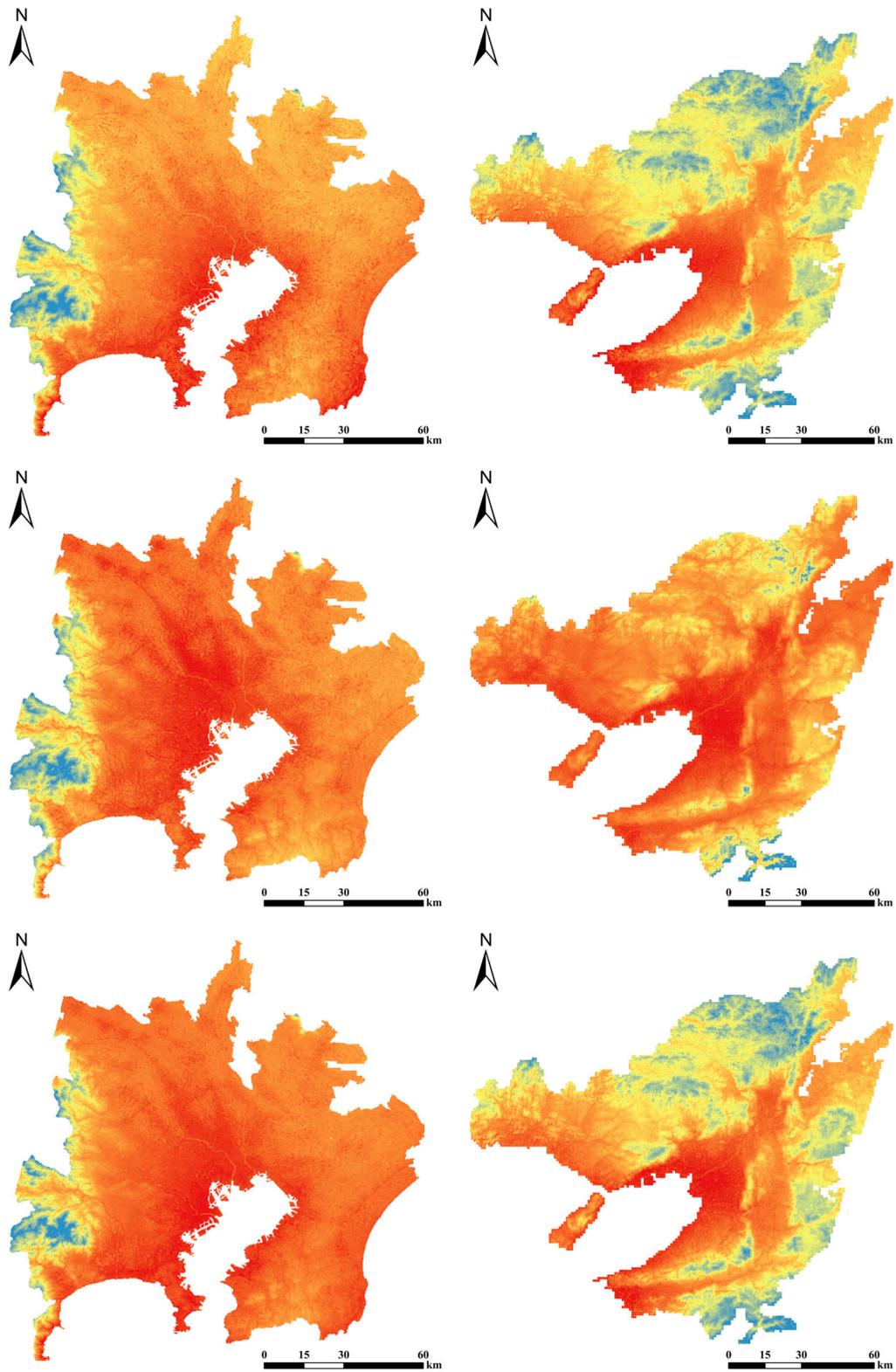
(e)
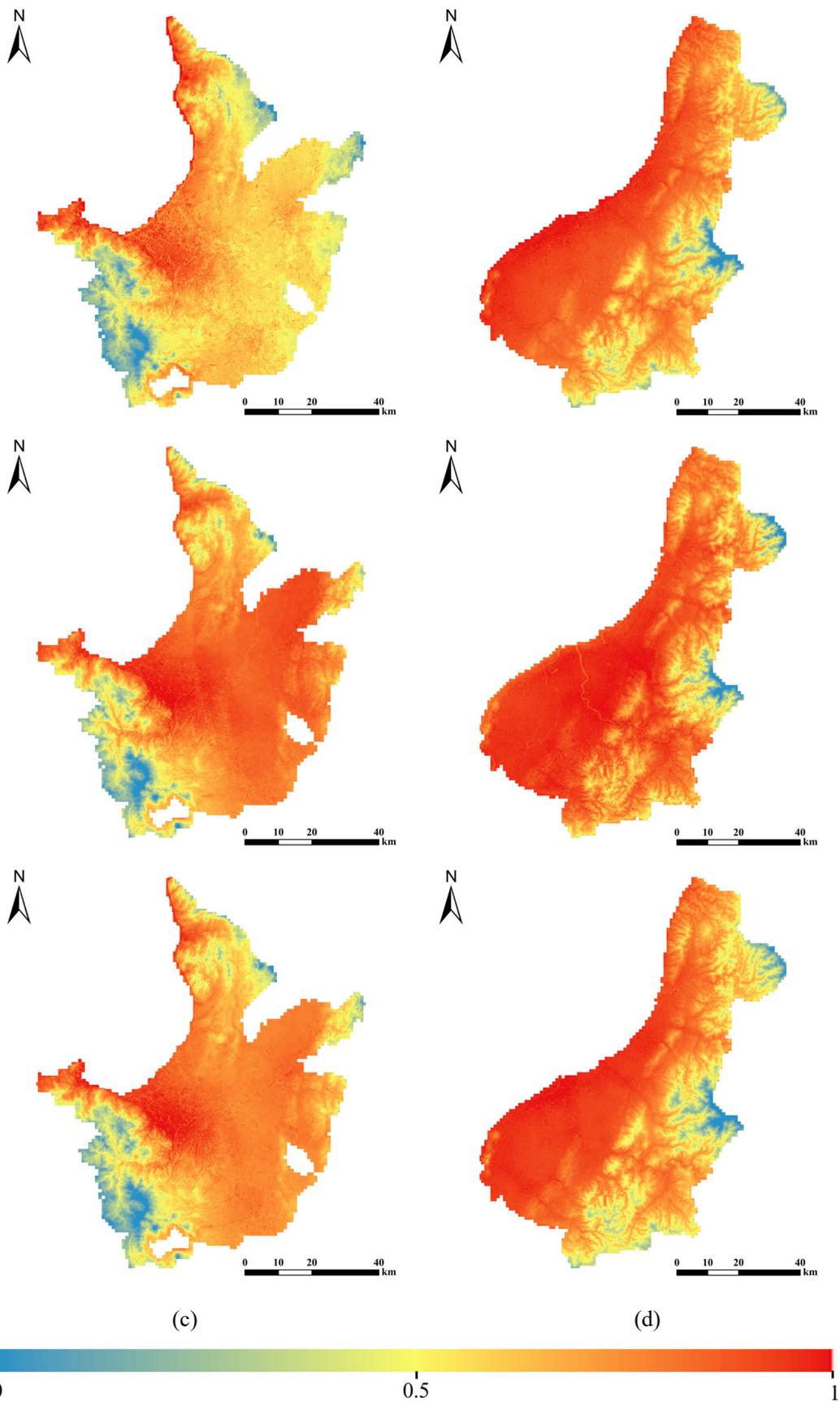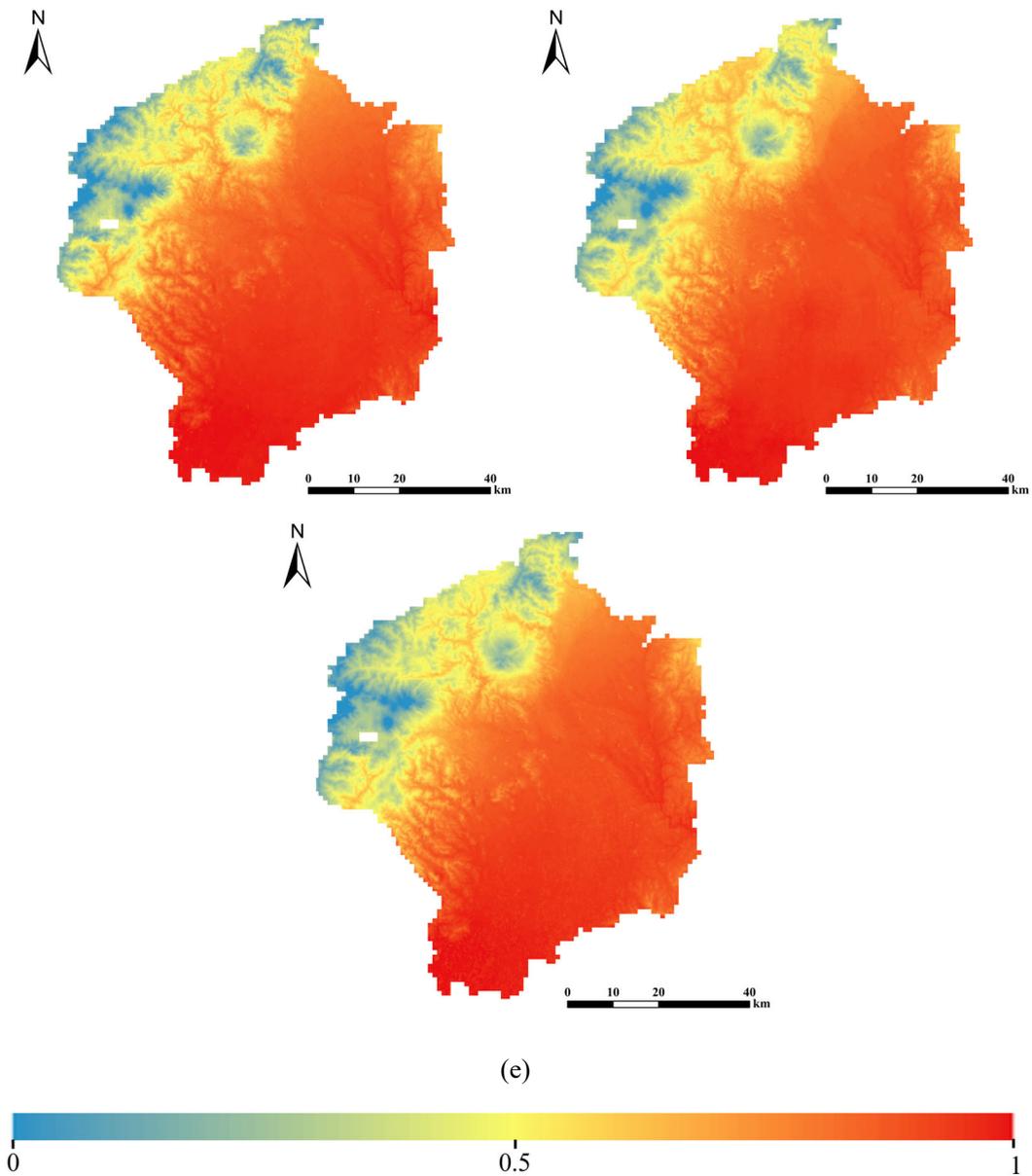
Figure 6.2. The Normalized Air Temperature of each metropolitan area and each period

6.2.2.   Jenks Natural Breaks Classification of Normalized Air Temperature

Based on the above normalized air temperature, the study areas are divided into seven levels by the Jenks natural breaks classification method. Table 6-2 shows the thresholds of each normalized air temperature level. We set the same threshold for different periods of the same metropolitan area.

Table 6-2 The thresholds of each normalized air temperature level

Area	low	sub low	sub mid	mid	sub high	high	extremely high
Kanto	0	0.3053	0.506	0.6378	0.7245	0.7815	0.8681	1
Kinki	0	0.3605	0.5706	0.6931	0.7645	0.8061	0.8775	1
Sapporo	0	0.3745	0.5863	0.7062	0.774	0.8123	0.8801	1
Niigata	0	0.5131	0.7282	0.8184	0.8562	0.872	0.9098	1
Utsunomiya	0	0.5131	0.7282	0.8184	0.8562	0.872	0.9098	1

According to the thresholds of each normalized air temperature level in Table 6-1, the normalized air temperature classification results of each metropolitan areas in different periods are shown in Figure 6.3, 6.4, 6.5, 6.6, and 6.7.

(a)

(b)

(c)
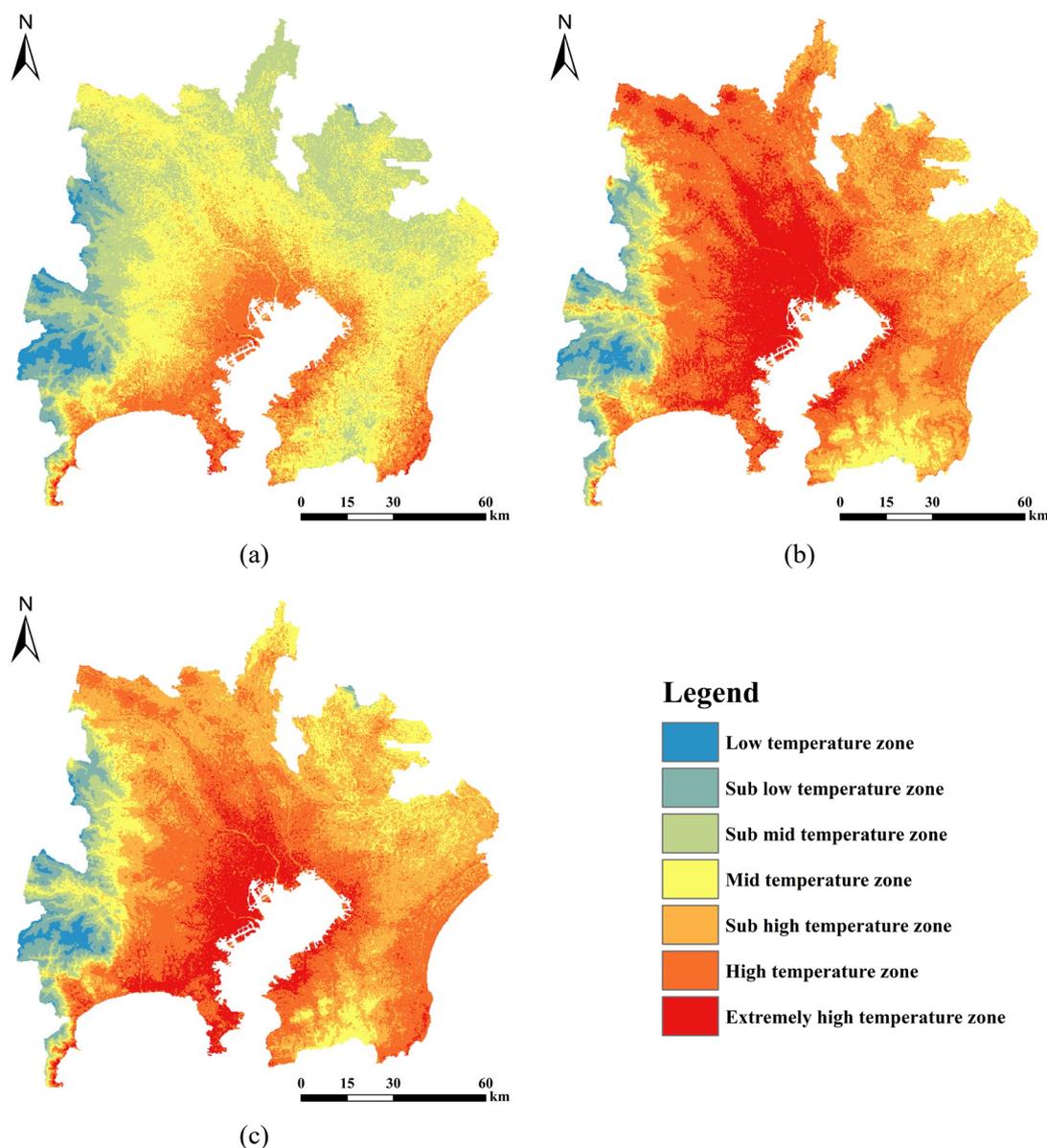
Figure 6.3. The normalized air temperature classification results of Kanto Major Metropolitan Area
(a) January; (b) August; (c) Annual.

Figure 6.3 shows the normalized air temperature classification results of Kanto Major Metropolitan Area. From blue to red respectively represent from the low temperature zone to the extremely high temperature zone. The urban heat island zones are concentrated on the coast and extend to the northwest of metropolitan area. And the urban heat island zone in January is significantly smaller and more concentrated, and it in August basically covers the entire plain.
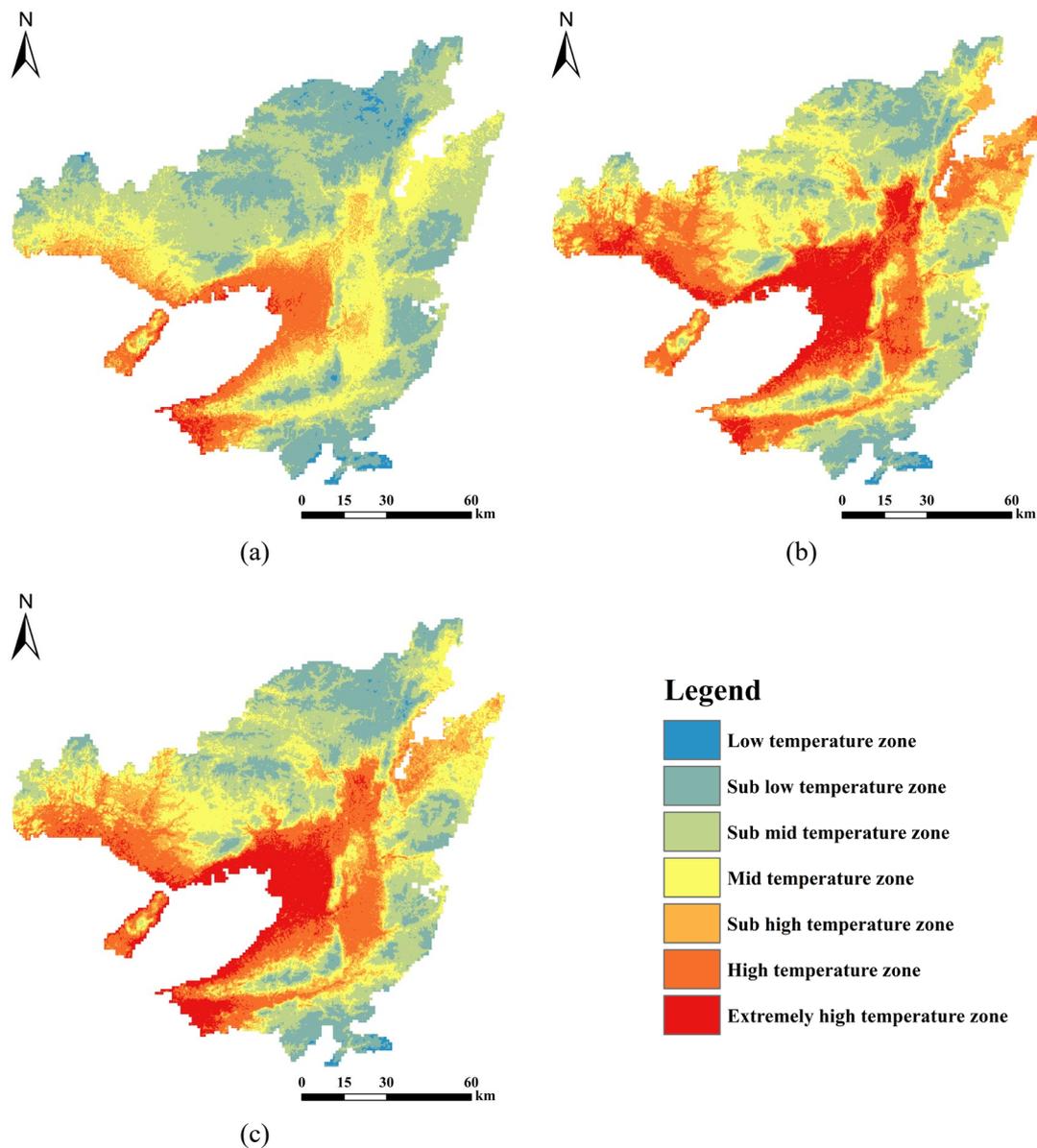
Figure 6.4. The normalized air temperature classification results of Kinki Major Metropolitan Area
(a) January; (b) August; (c) Annual.

Figure 6.4 shows the normalized air temperature classification of Kinki Major Metropolitan Area. Similar to the Kanto Major Metropolitan Area, the urban heat island zone is also concentrated along the coast. And the urban heat island zone in January is also smaller than that in August. And the area of the relatively extremely high temperature area in January is also very small.
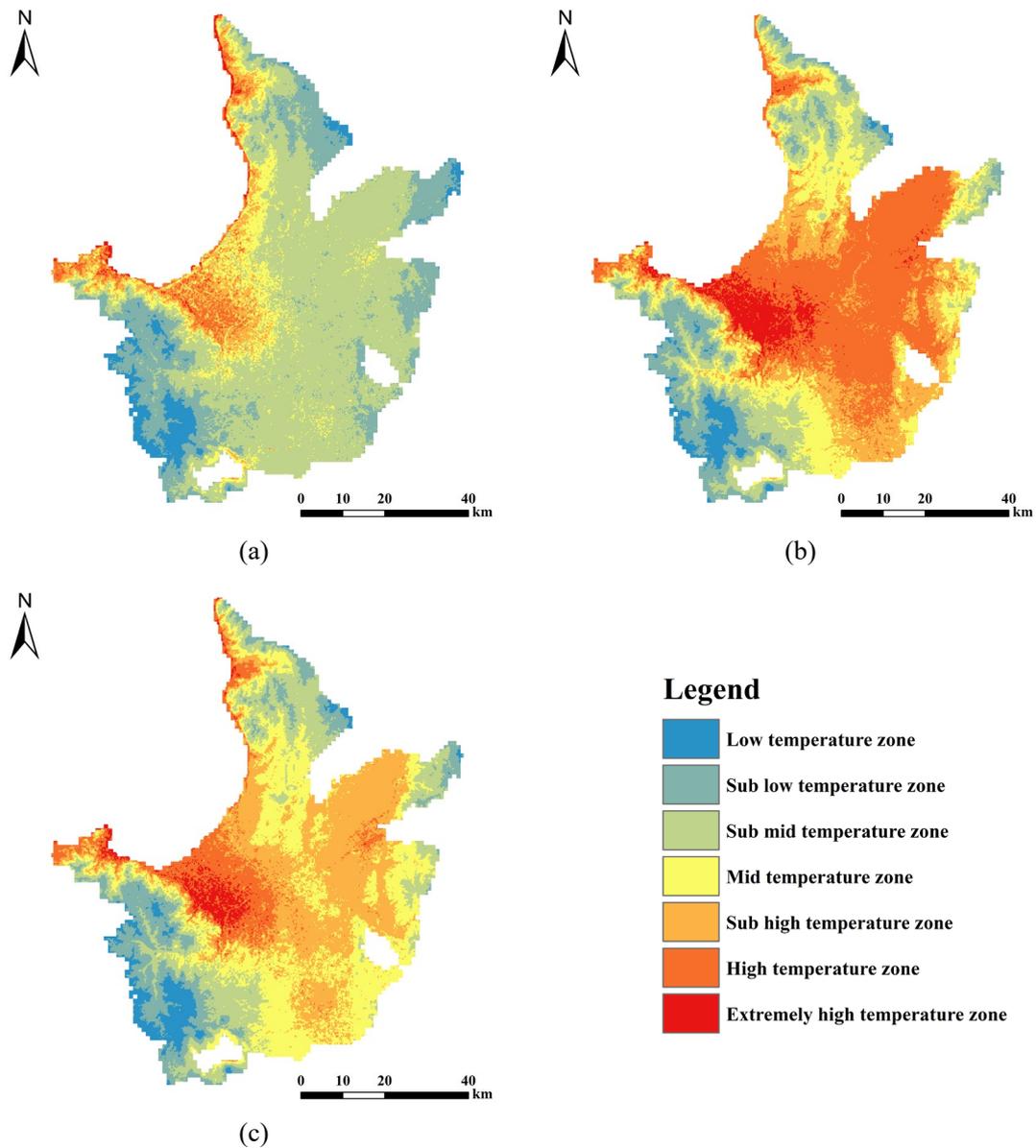
Figure 6.5. The normalized air temperature classification results of Sapporo Major Metropolitan Area (a) January; (b) August; (c) Annual.

Figure 6.5 shows the normalized air temperature classification results of Sapporo Major Metropolitan Area. The results show that the urban heat island zone is basically concentrated in the Sapporo city. In January, the normalized air temperature of the plain area around the Sapporo city is significantly lower than that in the urban center, so it is only a sub mid temperature zone. In contrast, the normalized air temperature of the surrounding area in August is not much lower than that of the urban center, so there is a large area of urban heat island zone. According to the annual average data, it is between January and August, and in the plain area, there is a large sub high temperature and mid temperature zone.
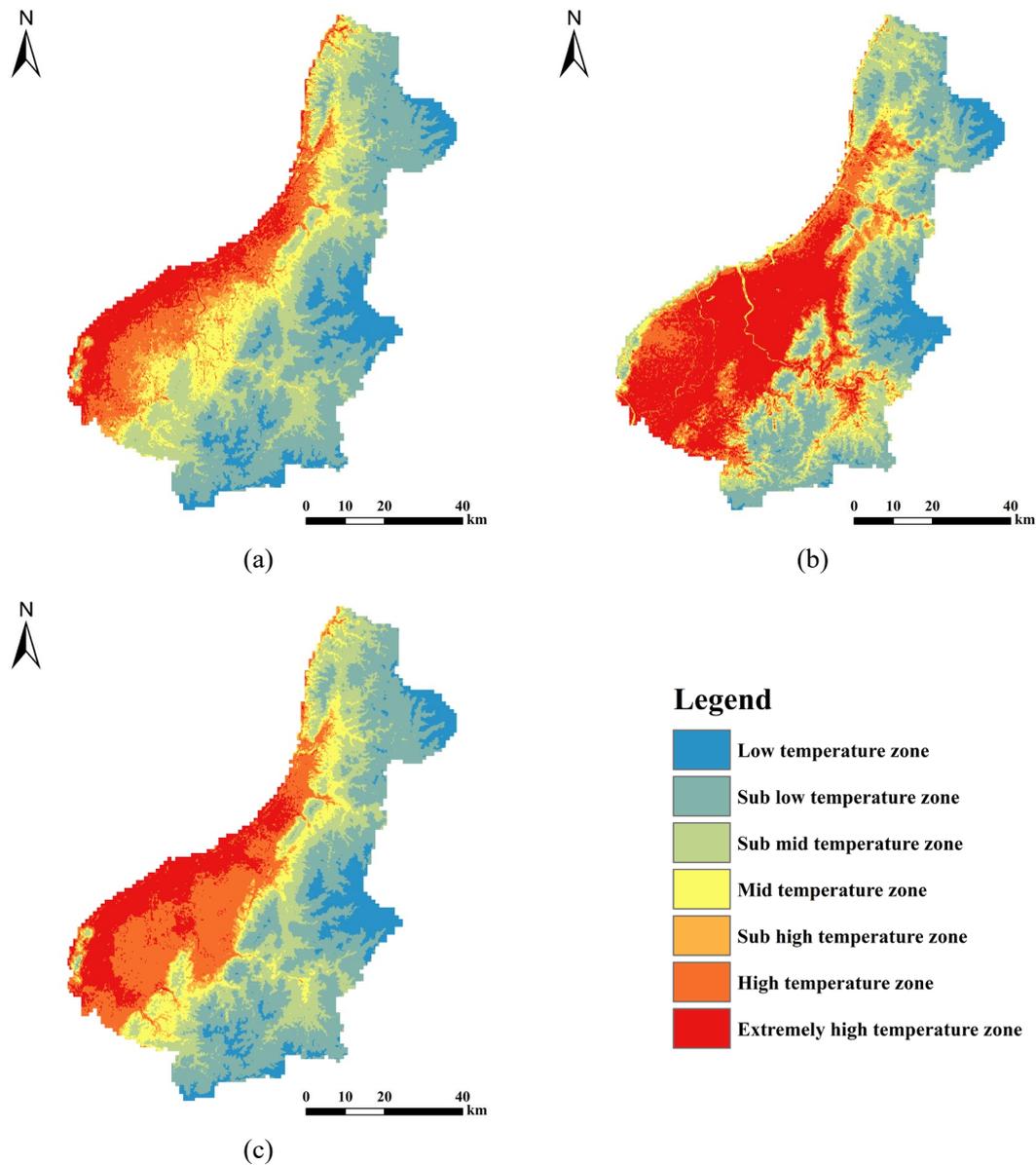
Figure 6.6. The normalized air temperature classification results of Niigata Major Metropolitan Area
(a) January; (b) August; (c) Annual.

Figure 6.6 shows the normalized air temperature classification results of Niigata Major Metropolitan Area. Unlike other metropolitan areas, the classification results of the Niigata Major Metropolitan Area in January and August are quite different. The urban heat island zone in January is obviously more concentrated near the coastline. In contrast, the urban heat island zone in August is concentrated in the plain area. In January, with the deepening of the inland plains, the extremely high temperature zone gradually becomes the high temperature zone, then is sub high temperature zone, and the finally is mid temperature zone. In August, it is not affected by the distance from the sea. The entire plain is basically covered by extremely high temperature zone and high temperature zone. According to the

annual average data, the distribution of urban heat islands is more similar to that in January, which is greatly affected by the distance from the sea.
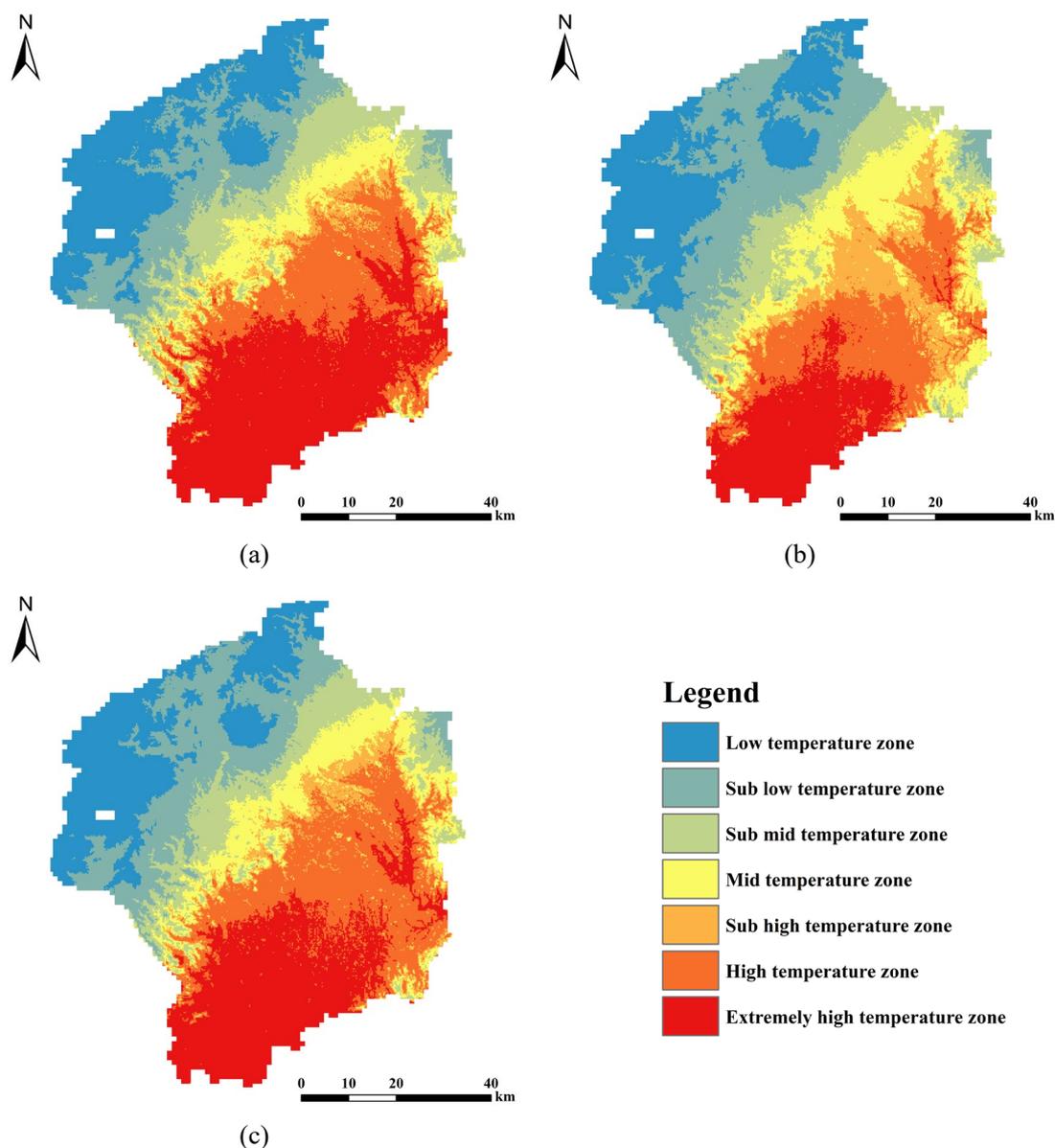
Figure 6.7. The normalized air temperature classification results of Utsunomiya Metropolitan Area

(a) January; (b) August; (c) Annual.

Figure 6.7 shows the normalized air temperature classification results of Utsunomiya Metropolitan Area. Unlike other metropolitan areas, the urban heat island zone in August is smaller than in January. Affected by the Kanto Major Metropolitan Area on the south side, the urban heat island zone is always on the south side of the metropolitan area regardless of the periods.

6.2.3.   Calculate the Urban Heat Island Ratio Index

According to the downscaled air temperature level distribution map obtained in the previous section, we counted the area of each level and calculate the corresponding ratio of each level. The statistical results are shown in Table 6-3.

The results show that the ratio of urban heat island zones (that is, the sum of extremely high temperature zone and high temperature zone) in January is significantly lower than in August. The Kanto Major Metropolitan Area in August accounted for 55.36 % of the urban heat island zone, which is 43.34 % higher than January. For comparison, the urban heat island ratio in Sapporo Major Metropolitan Area in January only accounted for 4.32 %, that is 32.75 % less than that in August. The difference between the urban heat island ratio in Niigata Major Metropolitan Area and Kinki Major Metropolitan Area in January and August is almost the same with about 20 %. Specifically, the ratio of Kinki Major Metropolitan Area is 30.72 % in August and 10.08 % January; and in Niigata Major Metropolitan Area, the ratio in August is 40.51 %, and in January is 20.11 %. The Utsunomiya Metropolitan Area is different from the other metropolitan areas, the ratio of urban heat islands in January is larger than that in August, which is 40.48 % in January and 27.43 % in August.

Based on the area ratio of each level of the normalized air temperature, combined with the urban heat island classification standard, we calculated the urban heat island ratio index with the help of Equation 6.2. We took Kanto Major Metropolitan Area as an example to introduce the whole calculation process.

In this section, the normalized air temperature of Kanto Major Metropolitan Area was divided into 7 levels, including low temperature zone, sub low temperature zone, sub mid temperature zone, mid temperature zone, sub high temperature zone, high temperature zone, and extremely high temperature zone, so m in Equation 6.2 is 7. We set the extremely high temperature zone and the high temperature zone as the urban heat island zone, so n is 2. The level values of the extremely high temperature zone and high temperature zone are 7 and 6, respectively. Combining with the area ratio of each temperature zone in Table 6-3, the urban heat island ratio area of Kanto Major Metropolitan Area can be calculated as follows:

In January

$$URI_{Jan} = \frac{1}{100 \times 7}(7 \times 0.65 + 6 \times 11.37) = 0.10$$

(6-3)

In August

$$URI_{Aug} = \frac{1}{100 \times 7}(7 \times 17.34 + 6 \times 38.03) = 0.50$$

(6-4)

Table 6-3 The area and ratio of each level of various metropolitan area and in different periods.

Area	Period	low area (km²)	low ratio (%)	sub low area (km²)	sub low ratio (%)	sub mid area (km²)	sub mid ratio (%)	mid area (km²)	mid ratio (%)	sub high area (km²)	sub high ratio (%)	high area (km²)	high ratio (%)	extremely high area (km²)	extremely high ratio (%)
Kanto	January	267	2.05	742	5.69	3285	25.18	5171	39.64	2013	15.43	1484	11.37	85	0.65
	August	134	1.03	541	4.14	608	4.66	1189	9.11	3352	25.69	4961	38.03	2262	17.34
	Annual	157	1.20	537	4.12	640	4.90	1780	13.64	4389	33.64	4170	31.97	1372	10.52
Kinki	January	148	1.13	3354	25.77	4468	34.32	2732	20.99	1004	7.72	1167	8.97	144	1.11
	August	53	0.40	1294	9.94	3104	23.85	2764	21.24	1802	13.84	2542	19.52	1459	11.21
	Annual	42	0.32	1794	13.78	3238	24.88	2721	20.90	1705	13.10	2254	17.31	1263	9.70
Sapporo	January	243	4.83	1005	20.01	2676	53.28	521	10.37	361	7.19	186	3.70	31	0.62
	August	150	2.98	521	10.37	748	14.89	860	17.12	882	17.57	1617	32.19	245	4.88
	Annual	206	4.10	561	11.16	832	16.56	1402	27.92	1426	28.39	473	9.41	123	2.44
Niigata	January	527	9.52	1770	31.97	1149	20.75	641	11.57	282	5.09	651	11.76	518	9.35
	August	340	6.14	1176	21.23	997	18.00	526	9.49	256	4.63	716	12.93	1528	27.58
	Annual	502	9.06	1683	30.39	1062	19.17	510	9.21	171	3.08	962	17.38	649	11.72
Utsunomiya	January	1053	18.48	829	14.55	690	12.11	557	9.79	261	4.58	863	15.15	1443	25.33
	August	1074	18.86	1111	19.50	624	10.96	754	13.24	570	10.01	825	14.49	737	12.94
	Annual	984	17.28	962	16.89	601	10.55	588	10.33	323	5.68	1081	18.99	1155	20.29

Annual

$$URI_{Ann} = \frac{1}{100 \times 7}(7 \times 10.52 + 6 \times 31.97) = 0.38$$

(6-5)

Using the same method, we calculated the urban heat island ratio index in different periods of each metropolitan area. The results are shown in Figure 6.8. The blue, red and green histograms represent the urban heat island ratio index for January, August, and the whole year respectively. Urban heat island ratio index can intuitively indicate the ratio of the urban heat island zone in each metropolitan area in different periods.
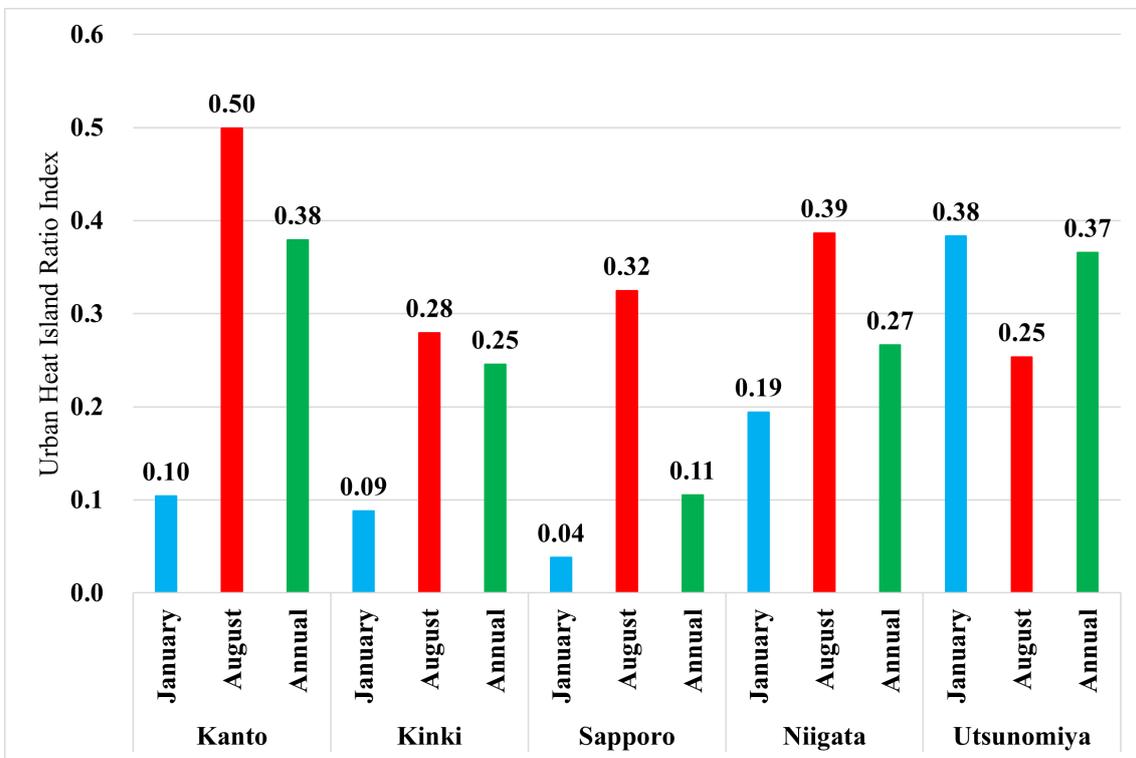
Figure 6.8. The urban heat island ratio index of each metropolitan area in different periods

As the Figure 6.8 shows, the urban heat island ratio indexes of Kanto Major Metropolitan Area, Kinki Major Metropolitan Area, Sapporo Major Metropolitan Area, and Niigata Major Metropolitan Area in January are lower than that in August. Among them, Kanto Major Metropolitan Area has the largest difference between January and August, reaching 0.4. The exception is the Utsunomiya Metropolitan Area, where the urban heat island ratio index in January is larger than that in August, and the difference is the smallest among the five metropolitan areas, which is only 0.13. The lowest urban heat island ratio index is in January in Sapporo Major Metropolitan Area, only 0.04. From the annual

average value, the urban heat island ratio index of Kanto Major Metropolitan Area is the largest, reaching 0.38; followed by the Utsunomiya Metropolitan Area, which has a slightly smaller value than Kanto Major Metropolitan Area at 0.37. The next are Niigata Major Metropolitan Area and Kinki Major Metropolitan Area with 0.27 and 0.25, respectively. The final is Sapporo Major Metropolitan Area at 0.11.

6.3. Atmosphere Urban Heat Island Analysis based on UHII and URI

According to the urban heat island intensity results and urban heat island ratio index calculated in Section 6.1 and 6.2, we collected the values of each metropolitan areas in different, and compared and analyzed the results. The statistical results are shown in Figure 6.9. The line chart represents urban heat island intensity, and the histogram chart represents the urban heat island ratio index.
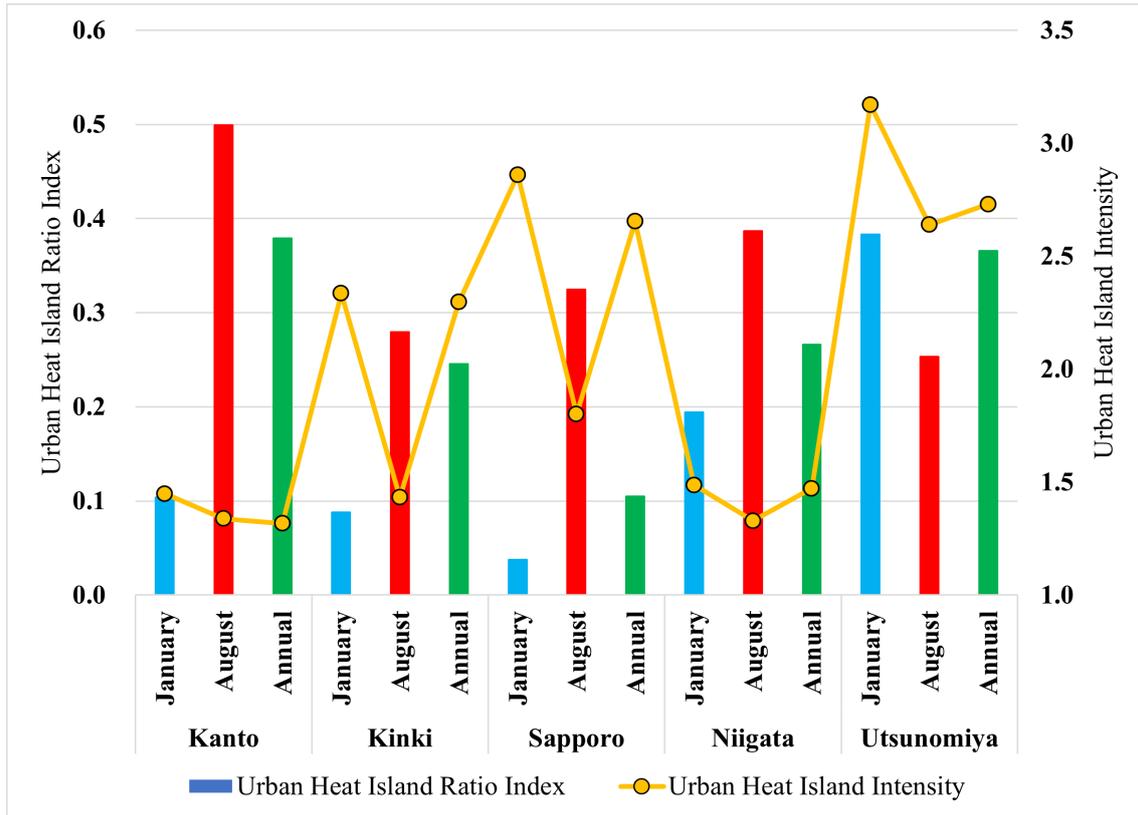
Figure 6.9. The urban heat island intensity and urban heat island ratio index of each metropolitan area in different periods

By comparing the situation of urban heat island intensity and urban heat island ratio index in different periods of the five metropolitan areas, we can find the following summaries.

The first is about the urban heat island intensity. The urban heat island intensity in January is stronger than that in August. The difference between January and August in Sapporo Major Metropolitan Area located in Hokkaido Zone and Kinki Major Metropolitan Area located in Inland Sea Zone is the largest. And Kanto Major Metropolitan Area and Niigata Major Metropolitan has the lowest urban heat island intensity and difference. Both metropolitan areas are located in the climatic zones facing the ocean. Kanto Major Metropolitan Area is located Pacific Coast Zone, and Niigata Major Metropolitan Area is located Sea of Japan Zone. The only metropolitan area located inland

which is Central Highland Zone, Utsunomiya Metropolitan Area has the highest urban heat island intensity and is greatly influenced by Kanto Major Metropolitan Area. The annual average data is similar. Utsunomiya Metropolitan Area has the highest urban heat island intensity, the next are Sapporo Major Metropolitan Area and Kinki Major Metropolitan Area, and the lowest is Niigata Major Metropolitan Area and Kanto Major Metropolitan Area.

Regarding urban heat island ratio index, Utsunomiya Metropolitan Area is different from the other four metropolitan area. The value in January is higher than that in August, while in other metropolitan areas, the urban heat island ratio index in August is higher than that in January. And the value of Kanto Metropolitan Area in August is much larger than that in January.

Combined with the urban heat island intensity and urban heat island ratio index, the five metropolitan areas can be roughly divided into three categories. The Utsunomiya Metropolitan Area is one type, which is characterized by the strong urban heat island intensity, and the urban heat island in January is higher than that in August, that is, the urban heat islands are more concentrated in August, but is not much different from that in January. Then, the Kanto Major Metropolitan Area and Niigata Major Metropolitan Area belong to one type. The characteristic of this type is that the urban heat island is very weak, but the urban heat island zone accounts for Table large ratio in August, which shows that most metropolitan areas are in urban heat island zone in August. However, in January, the urban heat island ratio index is low. This indicates that the urban heat island zone is small in January, and combined with Figure 6.3 and 6.6, we can find that the urban heat island zone in January is greatly affected by the distance from the sea. Finally, the last type includes Sapporo Major Metropolitan Area and Kinki Major Metropolitan Area. In this category, the urban heat island intensity in January and August is large, and the annual average is relatively medium compared to the other two types. However, the regional distribution of urban heat islands is also quite different between January and August. And the urban heat island zone is less affected by the distance from the sea, and more affected by the urban contour.

6.4. Summary

In this chapter, we calculated the urban heat island intensity and urban heat island ratio index. The urban heat island intensity can indicate the strength and weakness of the urban heat island in this area. The larger the value, the more serious this area. On the contrary, it shows that the urban heat island in this area is weak, and even there is no urban heat island. The urban heat island indicates the coverage and concentration of the urban heat island area. The larger the value, the more the urban heat island area is covered, and the wider the coverage of urban heat island is. The smaller the value, the more concentrated the urban heat island and the smaller the area of urban heat island zone.

The urban heat island intensity in the winter in January is generally higher than that in summer in August. However, the regional distribution of urban heat island zone is the opposite. Expect for the Utsunomiya Metropolitan Area, the urban heat island zone of the other four metropolitan area is more concentrated in the winter in January. The urban heat island zones of the Kanto Major Metropolitan Area and Niigata Major Metropolitan Areas area greatly affected by the distribution from the sea, while the urban heat island zones of Sapporo Major Metropolitan Area and Kinki Major Metropolitan Area are mainly concentrated in the center of the metropolitan area. The urban heat island zone of Utsunomiya Metropolitan Area is affected by Kanto Major Metropolitan Area and is concentrated in Utsunomiya city and the southern part of the metropolitan area.

Reference

1.   Martin-Vide, J.; Sarricolea, P.; Moreno-García, M.C. On the definition of urban heat island intensity: The "rural" reference. *Front. Earth Sci.* **2015**, *3*, 24, doi:10.3389/feart.2015.00024.

2.   Kim, Y.-H.; Baik, J.-J. Maximum urban heat island intensity in Seoul. *J. Appl. Meteorol.* **2002**, *41*, 651–659.

3.   Atkinson, B.W. Numerical modelling of urban heat-island intensity. *Boundary-Layer Meteorol.* **2003**, *109*, 285–310, doi:10.1023/A:1025820326672.

4.   Memon, R.A.; Leung, D.Y.C.; Liu, C.H. An investigation of urban heat island intensity (UHII) as an indicator of urban heating. *Atmos. Res.* **2009**, *94*, 491–500, doi:10.1016/j.atmosres.2009.07.006.

5.   National Land Numerical Information Available online: https://nlftp.mlit.go.jp/ksj/ (accessed on Jan 31, 2021).

6.   Han-qiu, X.; Ben-qing, C.; Han-qiu, X.; Ben-qing, C. An Image Processing Technique for the Study of Urban Heat Island Changes Using Different Seasonal Remote Sensing Data. *Remote Sens. Technol. Appl.* **2011**, *18*, 129–133, doi:10.11873/J.ISSN.1004-0323.2003.3.129.

7.   XU, H.; CHEN, B. Remote sensing of the urban heat island and its changes in Xiamen City of SE China. *J. Environ. Sci.* **2004**, *16*, 276–281.

8.   Jenks, G.F.; Caspall, F.C. Error on Choroplethic Maps: Definition, Measurement, Reduction. *Ann. Assoc. Am. Geogr.* **1971**, *61*, 217–244, doi:10.1111/j.1467-8306.1971.tb00779.x.

9.   Jenks, G. The Data Model Concept in Statistical Mapping. *undefined* **1967**.

10.   McMaster, R. In Memoriam: George F. Jenks (1916-1996). *Cartogr. Geogr. Inf. Syst.* **1997**, *24*, 56–59, doi:10.1559/152304097782438764.

11.   FAQ: What is the Jenks optimization method? Available online: https://support.esri.com/en/technical-article/000006743 (accessed on May 15, 2021).

Chapter 7. Comparative Study of Atmosphere Urban Heat Island and Surface Urban Heat Island

7.1. Introduction

With the development of urban around the word, urban heat island is the main subject of urban climate research. It mainly occurs on the surface and in the atmosphere. Based on this, urban heat islands can be divided into two types: atmosphere urban heat island and surface urban heat island. Although these two urban heat island types are formed in different ways, they can be related to the energy exchange between surface and atmosphere [1].

Atmosphere urban heat island refers the difference in air temperature between urban area and suburb area, is the result of direct contact between the heat detector and the air. Atmosphere urban heat island is generally divided into urban canopy layer heat island and urban boundary layer heat island according to the air temperature at different heights [2]. Urban canopy layer heat island is the earliest and most common urban heat island to be studied [3]. It is usually studied based on the air temperature measured at meteorological station and meteorological observation points in urban area and suburb area. The air temperature data used in this study was downscaled from the Japan National Land Numerical Information. According to the product specification [4], the air temperature grid data comes from the data of various meteorological station and meteorological observation points. Therefore, the atmosphere urban heat island studied in this article refers to urban canopy layer heat island.

Surface urban heat island can be represented based on the land surface temperature difference between urban area and suburbs observed directly from the infrared thermography from satellites and aircraft [5]. For a long time, satellite thermal infrared remote sensing data has greatly promoted the surface urban heat island research on the regional or larger scales [2,6–8], and aircraft remote sensing is relatively rare. This is since satellites remote sensing data have periodic observations on a large area and relatively uniform data processing standard. However, it is very expensive to obtain aircraft remote sensing data with periodic and large area observation [9]. In this section, we mainly use satellite thermal infrared remote sensing data, namely Landsat 4-5 Thematic Mapper (TM) data, to retrieve land surface temperature for surface urban heat island research. Based on the limitation of Landsat 4-5 TM on natural factors such as weather and clouds, we only selected Kanto Major Metropolitan Area and Kinki Major Metropolitan Area as the research object to compare the atmosphere urban heat island and surface urban heat island.

In order to compare atmosphere urban heat island and surface urban heat island, we introduced some landscape pattern metrics to compare and analyze the regression relationship between urban structure and urban heat islands.

In this section, we mainly used two landscape patter metrics, that is Percentage of landscape (PLAND) and Aggregation index (AI), to characterize the urban landscape. PLAND represents the percentage of a certain patch type total area to the landscape total area. The larger the PLAND is, the richer coverage of this kind of patch is [10,11]. AI reflects the degree of patch aggregation, and a large value indicated that the landscape is composed of a few large patches that are clustered together [11,12].

We calculated these two metrics and established regression models with land surface temperature and air temperature to compare the relationship between different urban structure on atmosphere urban heat islands and surface urban heat islands.

To sum up, in this chapter, we first used Landsat 4-5 TM remote sensing data to retrieve the land surface temperature, and then analyze the surface urban heat island, and compared it with the atmosphere urban heat island based on the high-resolution downscaled air temperature data obtained in the previous chapters. Meanwhile, we also introduced two landscape pattern metrics to establish the regression relationship between urban structure and urban heat island, so as to compare and analyze the impact of various urban structures on atmosphere urban heat island and surface urban heat island.

7.2. Comparison of AUHI and SUHI: Case Study of Kanto MMA and Kinki MMA

7.2.1.   The Retrieval of Land Surface Temperature from Landsat 4-5 TM

The most important data basis for surface urban heat islands is the land surface temperature. Therefore, the acquisition of land surface temperature is the core of this section.

Land surface temperature is the radiant surface temperature of land derived from solar radiation [13]. Land surface temperature and air temperature are essentially different, and they are complementary, because the temperature difference between the surface and the air above determines the sensible heat flux, and the correlation between them is caused by the surface heat balance [14].

With the continuous development of remote sensing technology, there are many ways to obtain land surface temperature. And the most important and common including two ways. The first is to download land surface temperature products from Moderate Resolution Imaging Spectroradiometer (MODIS). MODIS land surface temperature products are the one of most reliable [15–19]. However, the MODIS land surface temperature products also have some shortcomings, the most notable is the low spatial resolution, which is only about 1 km [20–22]. It is difficult to meet the accuracy requirements of surface urban heat island analysis based on the high-resolution land surface temperature in this section. For comparison, another commonly used method for obtaining land surface temperature is to retrieve from thermal infrared remote sensing. Among them, the most widely used is to retrieve the land surface temperature based on the remote sensing data from the Landsat series satellites. The land surface temperature retrieved based on thermal infrared remote sensing data has relatively high spatial resolution and high retrieval accuracy. But the main problem is that it cannot penetrate the clouds, which limits their practical applications [23]. According to the high-resolution characteristics, we still use the land surface temperature retrieved from thermal infrared remote sensing data as the research object.

What this section needs is the land surface temperature data in 2010. The launch time of Landsat 8 TIRS is 2013 [24]. The scan line corrector in the ETM+ instrument of Landsat 7 failed. The result is that approximately 22% of the data is lost in the Landsat 7 ETM+ when collecting without the scan line corrector [25]. Therefore, we selected Landsat 4-5 TM as the data source to retrieve the land surface temperature.

Landsat 4-5 TM consist seven spectral bands, six bands are in the visible and near infrared bands, and only one band is in the thermal infrared [26]. Band 1 (blue) with a wavelength of 0.45-0.52 μm is used for the coastal water research. Band 2 (green) with a wavelength of 0.52-0.60 μm is used for the identification of crops and vegetation. Band 3 (red, with a wavelength of 0.63-0.69 μm) and Band 4 (near infrared, with a wavelength of 0.77-0.90 μm) are used for the vegetation indexes calculation, for example normalized difference vegetation index (NDVI). Band 5 (short wave infrared, with a wavelength of 1.55-1.75 μm) and Band 7(short wave infrared, with a wavelength of 2.09-2.35 μm) are used for cloud, ice, snow, and geological structure discrimination. Final, Band 6 (thermal infrared,

with a wavelength of 10.40-12.50 μm) is used for the land surface temperature retrieval [27]. In order to retrieve land surface temperature, since Landsat 4-5 TM has only one thermal infrared band, it is not allowed the application of split-window method to retrieve the land surface temperature [28]. Finally, we used single-channel method for retrieving land surface temperature from Landsat 4-5 TM remote sensing data.

Single-channel method consists of 3 separate steps, namely, NDVI Thresholds, Thermal Band Digital Numbers to Brightness Temperature and Retrieve Land Surface Temperature [29].

- NDVI Thresholds

This step estimates land surface emissivity employing Normalized Difference Vegetation Index Thresholds Method (NDVI^{THM}) to distinguish between soil pixels (NDVI < NDVI$_S$), full vegetation pixels (NDVI > NDVI$_V$) and mixed pixels (NDVIS ≤ NDVI ≤ NDVI$_V$). And the threshold values of NDVI$_S$ = 0.2 and NDVI$_V$ = 0.5 make the method applicable for global conditions. The NDVI can be calculated from LANDSAT 4-5 TM by Equation 7-1. And then NDVI are converted to Land Surface Emissivity (LSE) using modified NDVI^{THM}.

$$NDVI = \frac{Band\ 4 - Band\ 3}{Band\ 4 + Band\ 3}$$

(7-1)

In the conversion, the emissivity (ε) of soil and full vegetation pixels should be used. And the emissivity can be defined by user. In general, the default emissivity values are:

soil pixels:

$\varepsilon = \varepsilon_{S\lambda} = 0.96$

full vegetation pixels:

$\varepsilon = \varepsilon_{V\lambda} + C_\lambda = 0.985 + 0.05 = 0.99$

The emissivity of mixed pixels is calculated by Equation 7-2. The proportion of vegetation (P_V) of each pixel is calculated by Equation 7-3 considering the proportion of vegetation. And the cavity effect (C_λ) that due to surface roughness is calculated by Equation 7-4. In this equation, the geometrical factor (F') is the mean value 0.55.

$$E = E_{V\lambda} \times P_V + E_{S\lambda} \times (1 - P_V) + C_\lambda$$

(7-2)

$$P_V = \left(\frac{NDVI - NDVI_S}{NDVI_V - NDVI_S} \right)^2$$

(7-3)

$$C_\lambda = (1 - E_{S\lambda}) \times \varepsilon_{V\lambda} \times F' \times (1 - P_V)$$

(7-3)

Additionally, this step allows the user to define the emissivity for the surface water. Normally, the default emissivity is 0.99. Before then, water bodies mask and the subset area (polygon feature layer) must be defined.

●   Thermal Band Digital Numbers to Brightness Temperature

The next step is using appropriate conversion coefficients to convert pixel values of thermal infrared band (band 6) in Landsat 4-5 TM to at-sensor spectral radiance in $W \cdot m^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$ according to Equation 7-4 and then transforming L_S to at-sensor brightness temperature (T_S) applying inverted Planck's Law and specific calibration constants (K_1 and K_2) as in the Equation 7-5. And this step can only work with LPGS (Level-1 Product Generation System) data [30].

$$L_S = M_L \times DN + A_L$$

(7-4)

where L_S represents top of atmosphere spectral radiance ($W/(m^2 \cdot srad \cdot \mu m)$); M_L represents band-specific multiplicative rescaling factor from the metadata, which can be obtained from Landsat Metadata (MTL) file; A_L represents band-specific additive rescaling factor from the metadata, which can be obtained from Landsat Metadata (MTL) file; DN represents quantized and calibrated standard product pixel values.

$$T_S = \frac{K_2}{\ln \left(\frac{K_1}{L_S} + 1 \right)}$$

(7-5)

where T_S represents top of atmosphere brightness temperature (K); K_1 and K_2 are band-specific thermal

conversion constant from the metadata, which can be obtained from Landsat Metadata (MTL) file.

● Retrieve Land Surface Temperature

This step estimates land surface temperature (LST) with single-channel algorithm according to Equation 7-6a, 6b and 6c.

$$LST = \gamma \left[\frac{1}{\varepsilon} (\psi_1 L_S + \psi_2) + \psi_3 \right] + \delta$$

(7-6a)

$$\gamma = \left[\frac{c_2 L_S}{T_S^2} \left(\frac{\lambda^4 L_S}{c_1} + \frac{1}{\lambda} \right) \right]^{-1}$$

(7-6b)

$$\delta = -\gamma \times L_S + T_S$$

(7-6c)

In these Equations, the necessary data are the calculated brightness temperature (T_s) and the land surface emissivity (ε) datasets as well as some specific atmospheric functions (AF). The AF's (ψ_1, ψ_2 and ψ_3) are used for correction of the atmosphere influence which is very important part of the land surface temperature retrieval algorithm. The AF's are computed from atmospheric parameter with Equation 7-7a, 7b and 7c.

$$\psi_1 = \frac{1}{\tau}$$

(7-7a)

$$\psi_2 = -L^\downarrow - \frac{L^\uparrow}{\tau}$$

(7-7b)

$$\psi_3 = L^\downarrow$$

(7-7c)

where τ represents atmospheric transmissivity, is set to 0.93; $L^{\uparrow}$ represents up-welling atmospheric radiance, is set to 0.5; $L^{\downarrow}$ represents down-welling atmospheric radiance, is set to 0.84.

These three steps are the whole procedure of land surface temperature retrieval from Landsat 4-5 TM remote sensing data.

7.2.2.   Data Source for Land Surface Temperature Retrieval

Because the satellite thermal infrared remote sensing imagery is greatly affected by the weather, in order to obtain the remote sensing data of the cloudless days in January and August, we only selected two metropolitan areas as the research objects, namely the Kanto Major Metropolitan Area and Kinki Major Metropolitan Area. And the revisit period of the Landsat satellite is only 16 days, which is difficult to obtain satisfactory data in a month. So we chose the data range to be plus or minus one month and one year. The basic information of Landsat 4-5 TM data is shown in Table 7-1.

Table 7-1 The basic information of Landsat 4-5 TM data

Location	Kanto Major Metropolitan Area	
Month	January	August
Landsat Scene Identifier	LT51070352009025BJC00, LT51070362009025BJC00	LT51070352010204HAJ00, LT51070362010204HAJ00
WRS Path/ WRS Row	107/035,107/036	107/035,107/036
Date	2009-01-25	2010-07-23
Land Cloud Cover	5%,2%	7%, 11%

Location	Kinki Major Metropolitan Area	
Month	January	August
Landsat Scene Identifier	LT51100352011052BJC00, LT51100362011052BJC00	LT51100352011196BJC00, LT51100362011196BJC00
WRS Path/ WRS Row	110/035,110/036	110/035,110/036
Date	2011-02-21	2011-07-15
Land Cloud Cover	9%,8%	7%, 10%

7.2.3.   The Retrieval Result and Surface Urban Heat Island Analysis

We retrieved the land surface temperature by the method mentioned in Section 7.2.1, and obtain the following results of Kanto Major Metropolitan Area and Kinki Major Metropolitan Area. In order to distinguish the land surface temperature of January and August, we defined January and February as winter, and July and August as summer. The following figures show the land surface temperature of Kanto Major Metropolitan Area and Kinki Major Metropolitan Area in winter and summer.
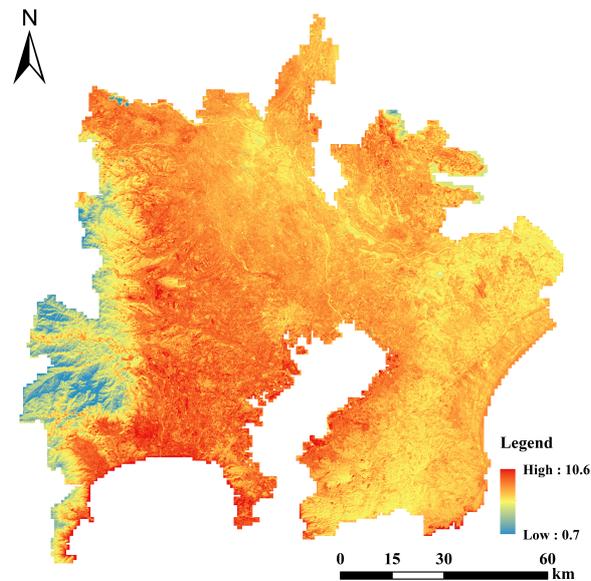
(a)
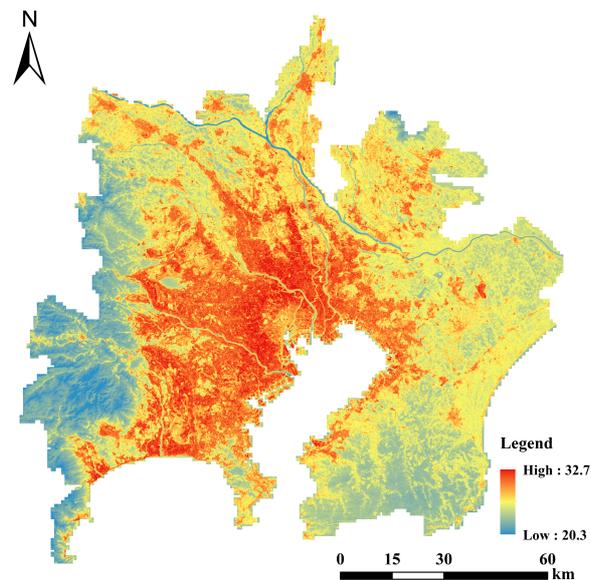
(b)

Figure 7.1 The Land surface temperature of Kanto Major Metropolitan Area
in winter (a) and summer (b)
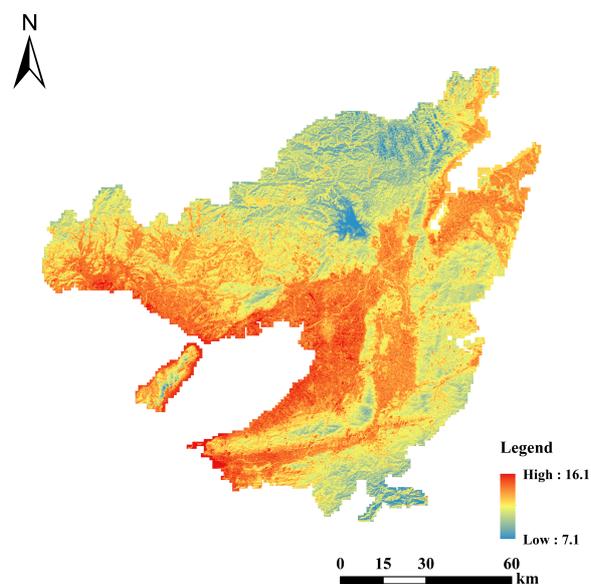
(a)
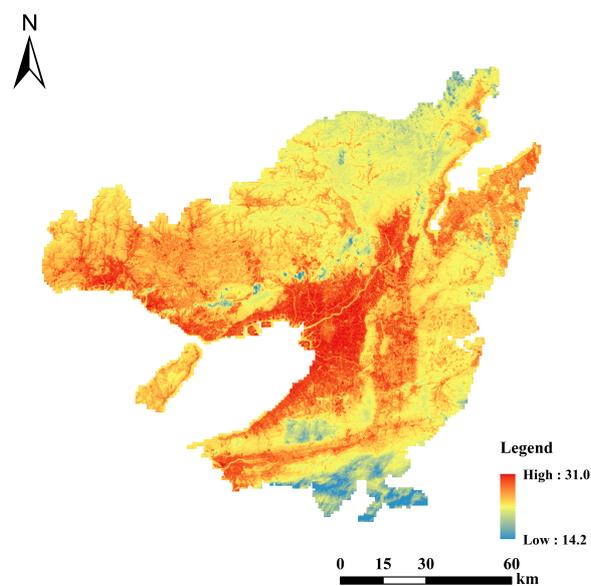
(b)

Figure 7.2 Land Surface Temperature of Kinki Major Metropolitan Area

in winter (a) and summer (b)

During the land surface temperature retrieval, there will be some extremely maximum or minimum values, so we set a confidence interval with 95% to obtain the range of the land surface temperature, as shown in Table 7-2.

The results shows that the land surface temperature distribution and range of Kanto Major Metropolitan Area and Kinki Major Metropolitan Area. The Kanto Major Metropolitan Area has

significantly more red areas in winter than that in summer. For Kinki Major Metropolitan Area, the red area range is similar in winter and in summer, but the color in summer is darker, that is, the land surface temperature of urban area in summer is higher. It can be seen from Table 7-2 that the land surface temperature difference of Kanto Major Metropolitan Area and Kinki Major Metropolitan Area in summer is obviously higher than that in winter. However, because the date values are different of two metropolitan area, they cannot be compared horizontally.

Table 7-2 The ranges of land surface temperature

Area	Period	Maximum	Minimum	Difference
Kanto	Winter	10.6	0.7	9.9
	Summer	32.7	20.3	12.4
Kinki	Winter	16.1	7.1	9
	Summer	31	14.2	16.8

7.2.4.   Comparative Study of Atmosphere Urban Heat Islands and Surface Urban Heat Island

Because the land surface temperature retrieved from Landsat 4-5 TM is a specific value at a certain moment, for example, the satellite thermal infrared remote sensing data of Kanto Major Metropolitan Area in winter was scanned at 10:01 on January 25, 2009, Tokyo Time. For comparison, the air temperature data downscaled from Japan National Land Numerical Information are monthly average data or annual average data. The two sets of temperature data have completely different time dimensions, so land surface temperature data and air temperature data cannot be used directly to compare the difference between atmosphere urban heat island and surface urban heat island in the Kanto Major Metropolitan Area and Kinki Major Metropolitan Area in different seasons. Therefore, we first need to normalize the land surface temperature and air temperature before the further comparison and analysis.

In this section, we used feature scaling method to normalize the land surface temperature and air temperature. Feature scaling method is used to standardize the range of variables or features [31]. Min-max normalization is the simplest method in feature scaling, and used in this section. It rescaled the range of feature to scale the range in [0,1]. The equation is as follows:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

(7-8)

where *x* represents the pixel values; *x'* represents normalized values.

Through standardization, we obtained the normalized air temperature and land surface temperature ranging from 0 to 1 of Kanto Major Metropolitan Area and Kinki Major Metropolitan Area in winter and summer. Next, we divided the entire metropolitan area into urban areas and suburb areas according to the densely inhabit districts mentioned in the previous chapter. Then, we calculated the average normalized air temperature and land surface temperature of urban areas and suburb areas. Final, we compared atmosphere urban heat island intensity and surface urban heat island intensity based on the average normalized temperature. The results are shown in Table 7-3.

Table 7-3 The comparison of atmosphere and surface urban heat island intensity

Area	Period	Atmosphere Urban Heat Island			Surface Urban Heat Island		
		Urban	Suburb	AUHII	Urban	Suburb	SUHII
Kanto	Winter	0.74	0.61	0.14	0.78	0.57	0.21
	Summer	0.87	0.75	0.12	0.84	0.40	0.44
Kinki	Winter	0.91	0.71	0.20	0.81	0.46	0.35
	Summer	0.78	0.63	0.15	0.89	0.45	0.45

The results show that the surface urban heat islands are significantly stronger than the atmosphere urban heat islands. The surface urban heat island intensity can reach a maximum of 0.45, and in contrast, the maximum of atmosphere urban heat island intensity is only 0.20.

Comparing the surface urban heat island and the atmosphere urban heat island, the similarity is that the urban heat island intensity of Kinki Major Metropolitan Area is higher than that of the Kanto Major Metropolitan Area, and the difference of the urban heat island intensity of two metropolitan area is smaller in summer. The surface urban heat island intensity of Kinki Major Metropolitan Area in summer is only 0.01 higher than that of the Kanto Major Metropolitan Area.

The differentiation is also obvious. In winter, the atmosphere urban heat island intensity is higher than in summer, and the difference is not large. The value in the Kinki Major Metropolitan Area is 0.05, and that in the Kanto Major Metropolitan Area is only 0.02. However, the surface urban heat island intensity in winter is obviously lower than that in summer. The difference in Kanto Major Metropolitan Area is obvious. The surface urban heat island intensity in winter is 0.21, while in summer it can reach 0.44, and the difference is 0.23. In Kinki Major Metropolitan Area, the surface urban heat island is higher than that in Kanto Major Metropolitan Area. The value in winter is 0.35, and it is 0.45 in summer, which is 0.10 higher than that in winter.

Combining the land surface temperature distribution of Figures 7.1 and 7.2, we can find that there

is a yellowish area in the urban centers of both Kanto Major Metropolitan Area and Kinki Major Metropolitan Area, that is, the land surface temperature in the urban center is lower than that in the surround area. This shows that there is a small surface urban cool island area in this area. But from a larger scale, the surface urban heat island still exists, that is, the land surface temperature of urban area is higher than that of suburb areas. To sum up, according to the land surface temperature, there is also a small cool island in the surface urban heat island, thus forming a structure similar to that of a volcano. But for atmosphere urban heat island, this phenomenon does not exist.

7.3. The Influence of Urban Structure on Two Kinks of Urban Heat Islands

7.3.1.   The Calculation of Landscape Pattern Metrics

The landscape pattern metrics is a quantitative index that can reflect landscape composition and spatial configuration. According to the previous research [32–34], we used two widely used landscape pattern metrics: percentage of landscape area (PLAND), and the aggregation index (AI). In this section, FRAGSTATS was applied to calculated the landscape pattern metrics [35].

In order to determine the impact of urban structure on atmosphere urban heat island and surface urban heat island, we established the most commonly used multiple linear regression model for urban structure and urban heat island [36–38]. We created the grid data with a side length of 10 kilometers as shown in Figure 7.3 for an example in Kanto Major Metropolitan Area, and remove the grids that accounted for less than 70% of the study area through statistics.
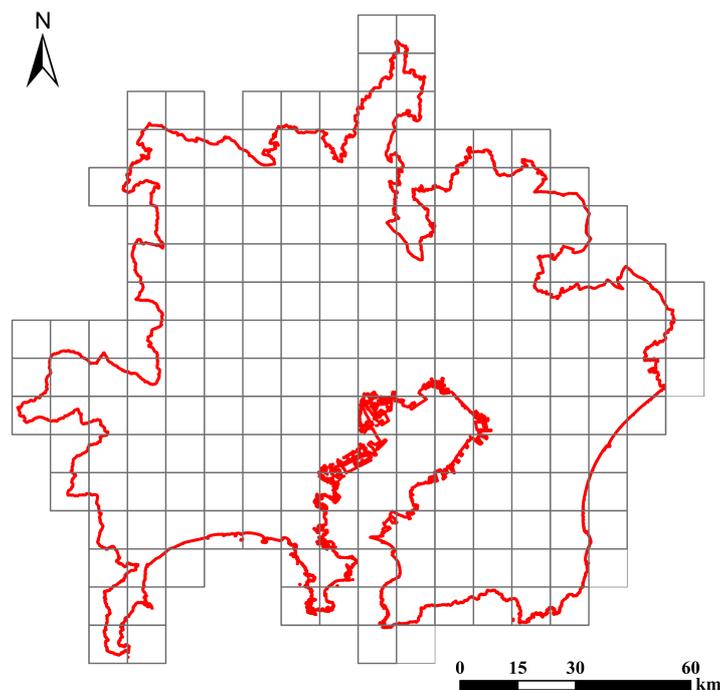
Figure 7.3 The grid data with a side length of 10 kilometers for multiple linear regression model establishment

Then, we calculated the PLAND and AI, with the Equation 7.9 [39] and 7.10 [40], and counted the air temperature and land surface temperature for each grid for the further regression model establishment.

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

(7-9)

where P_i represents the proportion of the landscape occupied by patch type i; a_{ij} represents the area of patch ij; A represents the total landscape area.

$$AI = \left[\sum_{i=1}^{m} \left(\frac{g_{ii}}{max \rightarrow g_{ii}} \right) P_i \right] (100)$$

(7-10)

where g_{ii} represents the number of like adjacencies between pixels of patch type i based on the single count method; $max \rightarrow g_{ii}$ represents the maximum number of the like adjacencies between pixels of patch type i based on the single count method; P_i represents the proportion of landscape comprised of patch type i.

According to the PLAND and AI calculated by FRAGSTATS and the temperature of each pixel, we used SPSS to establish the multiple linear regression for analysis and comparison.

7.3.2.   The Regression Analysis of Urban Structure and Urban Heat Island

The results of the multiple linear regression analysis of the urban structure and air temperature (AT) / land surface temperature (LST) are shown in Table 7-4 and Table 7-5. They indicate the effect of urban structure composition and configuration on air temperature and land surface temperature. The variance inflation factor of each independent variable is less than 5, so there is no multicollinearity in all multiple linear regression model.

The values in tables represent the standardized coefficients. AT represents air temperature and LST represents land surface temperature. The color represents the magnitude of the standardized coefficients, the redder the color, the stronger the positive affect of the independent variables on air temperature or land surface temperature, and the bluer the color, the stronger the negative affect of the independent variables on air temperature or land surface temperature. One asterisk represents the p-value less than 0.05, and two asterisks represent the p-value less than 0.01.

Combining with the last three rows, namely p-value, R^2, and adjusted R^2, the p-value of all multiple linear regression model are less than 0.01, it means that all model is statistically significant. Comparing the R^2 and adjusted R^2, in Kanto Major Metropolitan Area, the value in summer is larger than that in winter, and in Kinki Major Metropolitan Area, the difference between summer and winter is small. Adjusted R^2, are all more than 0.5, it means that all models are good fitting.

Table 7-4 Multiple linear regression analysis of urban structure and air temperature (AT) / land surface temperature (LST) of Kanto Major Metropolitan Area.

	Percentage of Landscape (PLAND)				Aggregation Index (AI)			
	AT_Winter	AT_Summer	LST_Winter	LST_Summer	AT_Winter	AT_Summer	LST_Winter	LST_Summer
LCZ1	0.06	0.026	-0.024	-0.166*	0.201*	0.097	-0.196	-0.225**
LCZ2	0.027	0.11	-0.031	0.143	0.004	0.204*	0.227	0.38**
LCZ3	0.083	0.095	-0.266	0.097	0.333**	0.365**	0.249**	0.268**
LCZ4	-0.152	-0.049*	-0.026	0.183*	0.046	-0.078	-0.042	-0.106
LCZ5	0.487*	0.151	-0.363*	0.005	0.167*	0.081	0.036	0.015
LCZ6	-0.374**	-0.127*	0.468**	0.203**	-0.376**	-0.226**	0.008*	-0.044**
LCZ7	-0.449*	0.153**	0.186*	0.248**	-0.539**	-0.12	-0.087**	-0.015
LCZ8	-0.224*	-0.19**	-0.167**	-0.057*	-0.008	0.068	0.06	0.089
LCZ9	-0.092	-0.103	0.076	-0.24**	0.164	0.158	0.1	-0.056
LCZ10	-0.148	-0.252*	-0.095	-0.142*	0.005	0.037	0.007	-0.016
LCZA	-0.641**	-0.697**	-0.58**	-0.535**	-0.067*	-0.393**	-0.357**	-0.331**
LCZB	0.212*	0.112	0.212*	-0.054	0.171**	0.14	0.38**	0.071
LCZC	0.089	-0.104	-0.001	-0.001	-0.103	-0.209**	-0.053	-0.199*
LCZD	0.095	0.016	-0.188	-0.038	0.099	0.037	0.112	-0.055
LCZE	0.453*	0.366*	0.435*	0.281**	0.124*	0.007	0.15**	0.039
LCZF	-0.11	0.002	0.1	-0.088*	-0.136	-0.174*	-0.094	-0.117
LCZG	-0.082	-0.075*	-0.048	-0.081**	0.03	0.021	-0.016	0.068
p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
R^2	0.712	0.834	0.786	0.972	0.625	0.793	0.525	0.799
Adjusted R^2	0.660	0.805	0.748	0.967	0.533	0.743	0.408	0.749

7-15

Table 7-5 Multiple linear regression analysis of urban structure and air temperature (AT) / land surface temperature (LST) of Kinki Major Metropolitan Area.

	Percentage of Landscape (PLAND)				Aggregation Index (AI)			
	AT_Winter	AT_Summer	LST_Winter	LST_Summer	AT_Winter	AT_Summer	LST_Winter	LST_Summer
LCZ1	0.201*	0.131*	-0.045	-0.063	0.106*	0.053	-0.094	-0.083
LCZ2	-0.105	0.04	-0.126	0.089	-0.019	0.015	-0.017	-0.006
LCZ3	0.254*	0.265**	0.002	0.247*	0.153*	0.289**	-0.007	0.325**
LCZ4	0.179*	0.054	0.108	-0.082	0.13	0.034	0.223**	-0.054
LCZ5	0.037	0.26**	0.036	0.198*	-0.128**	-0.049	-0.13**	-0.033
LCZ6	0.125*	0.184**	0.101	0.225**	0.013	0.04	0.093	0.062
LCZ7	0.189**	0.093**	0.222**	0.103*	-0.036	-0.042	0.033	-0.029
LCZ8	-0.057	-0.083*	-0.02	-0.093	0.017	-0.01	0.043	-0.035
LCZ9	0.017	0.038	0.009	0.088	-0.002	0.108	-0.095	0.106
LCZ10	-0.088	-0.055	0.294**	0.123	0.083	0.092	0.219**	0.171**
LCZA	-0.672**	-0.621**	-0.517**	-0.549**	-0.198**	-0.274**	-0.266**	-0.353**
LCZB	0.133*	0.059	0.029*	0.031	0.114**	0.024**	0.075***	0.031*
LCZC	0.03	0.062	0.102	0.096	-0.225**	-0.114**	-0.126*	-0.053
LCZD	0.088	0.127**	0.026	0.121*	0.006	0.125**	-0.042	0.108*
LCZE	0.351**	0.152*	0.356**	0.138*	-0.02	-0.022	-0.008	-0.061
LCZF	0.007	0.021	0.042	0.03	0.012	0.017	-0.085	0.022
LCZG	0.101	0.081	0.133*	0.068	0.146**	0.131**	0.133**	0.05
p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
R^2	0.843	0.914	0.866	0.867	0.842	0.877	0.789	0.806
Adjusted R^2	0.821	0.901	0.847	0.848	0.818	0.858	0.757	0.777

According to Table 7-4 and 7-5, the most obvious is that no matter which metropolitan area, dense trees have a strong negative effect on both air temperature and land surface temperature, while bare rock and pave have a strong positive effect on these, that is, dense trees will decrease the temperature and bare rock and pave will increase the temperature. However, in Kinki Major Metropolitan Area in summer, the bare rock and pave have less obvious effect on increasing air temperature and land surface temperature. On the other hand, compact high-rise buildings have a positive effect on air temperature and a negative effect on land surface temperature in both Kanto Major Metropolitan Area and Kinki Major Metropolitan Area. Combining with the urban structure, most compact high-rise buildings are concentrated in the urban center, which explains why there is a small cool island in urban center mentioned in Section 7.2.

7.4. Summary

In this chapter, we retrieved the land surface temperature of Kanto Major Metropolitan Area and Kinki Major Metropolitan Area based on the satellite thermal infrared remote sensing data of Landsat 4-5 TM. Because the temporal dimension of land surface temperature is completely different from air temperature, we normalized the land surface temperature and air temperature, and then calculated the atmosphere urban heat island intensity and surface urban heat island intensity. Meanwhile, we introduced two landscape pattern metrics to indicate the spatial composition and configuration of urban structure. The multiple linear regression model is used to indicate the affect of urban structure on atmosphere urban heat island and surface urban heat island.

The results show that there are some similarities and differentiation between atmosphere urban heat island and surface urban heat island. The similarity is that the urban heat island intensity of Kinki Major Metropolitan Area is always higher than that in Kanto Major Metropolitan Area, and the surface urban heat island intensity is always higher than the atmosphere urban heat island intensity. The first and most obvious differentiation is that the surface urban heat island intensity in summer is much higher than that in winter, in contrast, the atmosphere urban heat island intensity in summer is slightly lower than that in winter. The other differentiation is that there is a small cool island in the center of surface urban heat island, which does not exist in the atmosphere urban heat island. The multiple linear regression model in Section 7.3 also confirms this conclusion. The compact high-rise buildings in the urban center have a positive effect on the air temperature and a negative effect on the land surface temperature. Overall, dense trees have a strong negative impact on air temperature and land surface temperature, that is, dense trees can greatly reduce both air temperature and land surface temperature. On the contrary, bare rock and pave greatly increase the air temperature and land surface temperature, and enhance the urban heat island.

Reference

1.   Oke, T.R. The energetic basis of the urban heat island. *Q. J. R. Meteorol. Soc.* **1982**, *108*, 1–24, doi:10.1002/qj.49710845502.

2.   Voogt, J.A.; Oke, T.R. Thermal remote sensing of urban climates. *Remote Sens. Environ.* **2003**, *86*, 370–384, doi:10.1016/S0034-4257(03)00079-8.

3.   Huang, F.; Zhan, W.; Wang, Z.H.; Voogt, J.; Hu, L.; Quan, J.; Liu, C.; Zhang, N.; Lai, J. Satellite identification of atmospheric-surface-subsurface urban heat islands under clear sky. *Remote Sens. Environ.* **2020**, *250*, 112039, doi:10.1016/j.rse.2020.112039.

4.   National Land Numerical Information Available online: https://nlftp.mlit.go.jp/ksj/ (accessed on Jan 31, 2021).

5.   Oke, T.R. The Heat Island of the Urban Boundary Layer: Characteristics, Causes and Effects. In *Wind Climate in Cities*; Springer Netherlands, 1995; pp. 81–107.

6.   Weng, Q.; Fu, P.; Gao, F. Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data. *Remote Sens. Environ.* **2014**, *145*, 55–67.

7.   Weng, Q. Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends. *ISPRS J. Photogramm. Remote Sens.* 2009, *64*, 335–344.

8.   Zhou, D.; Xiao, J.; Bonafoni, S.; Berger, C.; Deilami, K.; Zhou, Y.; Frolking, S.; Yao, R.; Qiao, Z.; Sobrino, J. Satellite Remote Sensing of Surface Urban Heat Islands: Progress, Challenges, and Perspectives. *Remote Sens.* **2018**, *11*, 48, doi:10.3390/rs11010048.

9.   Chen, A.; Sun, R.; Chen, L. Studies on urban heat island from a landscape pattern view: A review. *Shengtai Xuebao/ Acta Ecol. Sin.* 2012, *32*, 4553–4565.

10.   He, P.; Zhang, H. Study on factor analysis and selection of common landscape metrics. *For. Res. Beijing* **2009**, *22*, 470–474.

11.   Mas, J.F.; Gao, Y.; Pacheco, J.A.N. Sensitivity of landscape pattern metrics to classification approaches. *For. Ecol. Manage.* **2010**, *259*, 1215–1224, doi:10.1016/j.foreco.2009.12.016.

12.   Jianbing, W. Landscape Pattern Scale Effect——Taking Shangri-La County as an Example. *Open J. Soil Water Conserv.* **2014**, *02*, 13–28, doi:10.12677/ojswc.2014.22002.

13.   *Urban Heat Island Modeling for Tropical Climates*; Elsevier, 2021;

14.   Dousset, B.; Luvall, J.C.; Hulley, G.C. Surface Temperatures in the Urban Environment. In *Taking the Temperature of the Earth*; Elsevier, 2019; pp. 203–226.

15.   Wan, Z. New refinements and validation of the MODIS Land-Surface Temperature/Emissivity products. *Remote Sens. Environ.* **2008**, *112*, 59–74, doi:10.1016/j.rse.2006.06.026.

16.   Wang, W.; Liang, S.; Meyers, T. Validating MODIS land surface temperature products using long-term nighttime ground measurements. *Remote Sens. Environ.* **2008**, *112*, 623–635, doi:10.1016/j.rse.2007.05.024.

17.   Sabol, D.E.; Gillespie, A.R.; Abbott, E.; Yamada, G. Field validation of the ASTER Temperature-Emissivity Separation algorithm. *Remote Sens. Environ.* **2009**, *113*, 2328–2344, doi:10.1016/j.rse.2009.06.008.

18.   Capelle, V.; Chédin, A.; Péquignot, E.; Schlüssel, P.; Newman, S.M.; Scott, N.A. Infrared continental surface emissivity spectra and skin temperature retrieved from IASI observations over the tropics. *J. Appl. Meteorol. Climatol.* **2012**, *51*, 1164–1179, doi:10.1175/JAMC-D-11-0145.1.

19.   Freitas, S.C.; Trigo, I.F.; Bioucas-Dias, J.M.; Göttsche, F.M. Quantifying the uncertainty of land surface temperature retrievals from SEVIRI/Meteosat. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 523–534, doi:10.1109/TGRS.2009.2027697.

20.   Wang, R.; Gao, W.; Peng, W. Downscale MODIS Land Surface Temperature Based on Three Different Models to Analyze Surface Urban Heat Island: A Case Study of Hangzhou. *Remote Sens.* **2020**, *12*, 2134, doi:10.3390/rs12132134.

21.   Zhu, W.; Lu, A.; Jia, S. Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. *Remote Sens. Environ.* **2013**, *130*, 62–73, doi:10.1016/j.rse.2012.10.034.

22.   Wan, Z. New refinements and validation of the MODIS Land-Surface Temperature/Emissivity products. *Remote Sens. Environ.* **2008**, *112*, 59–74, doi:10.1016/j.rse.2006.06.026.

23.   Duan, S.B.; Li, Z.L.; Leng, P. A framework for the retrieval of all-weather land surface temperature at a high spatial resolution from polar-orbiting thermal infrared and passive microwave data. *Remote Sens. Environ.* **2017**, *195*, 107–117, doi:10.1016/j.rse.2017.04.008.

24.   USGS EROS Archive - Landsat Archives - Landsat 8 OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor) Level-1 Data Products Available online: https://www.usgs.gov/centers/eros/science/usgs-eros-archive-landsat-archives-landsat-8-oli-operational-land-imager-and?qt-science_center_objects=0#qt-science_center_objects (accessed on May 5, 2021).

25.   Landsat 7 Available online: https://www.usgs.gov/core-science-systems/nli/landsat/landsat-7?qt-science_support_page_related_con=0#qt-science_support_page_related_con (accessed on May 18, 2021).

26.   What are the band designations for the Landsat satellites? Available online: https://www.usgs.gov/faqs/what-are-band-designations-landsat-satellites?qt-news_science_products=0#qt-news_science_products (accessed on May 18, 2021).

27.   Sobrino, J.A.; Jiménez-Muñoz, J.C.; Paolini, L. Land surface temperature retrieval from LANDSAT TM 5. *Remote Sens. Environ.* **2004**, *90*, 434–440, doi:10.1016/j.rse.2004.02.003.

28.   Sobrino, J.A.; Li, Z.L.; Stoll, M.P.; Becker, F. Multi-channel and multi-angle algorithms for estimating sea and land surface temperature with atsr data. *Int. J. Remote Sens.* **1996**, *17*, 2089–2114, doi:10.1080/01431169608948760.

29.   Walawender, J.P.; Hajto, M.J.; Iwaniuk, P. A new ArcGIS toolset for automated mapping of land surface temperature with the use of LANDSAT satellite data. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS); 2012; pp. 4371–4374.

30.   Using the USGS Landsat Level-1 Data Product Available online: https://www.usgs.gov/core-science-systems/nli/landsat/using-usgs-landsat-level-1-data-product (accessed on May 18, 2021).

31.   Feature scaling - Wikipedia Available online: https://en.wikipedia.org/wiki/Feature_scaling (accessed on May 20, 2021).

32.   Maimaitiyiming, M.; Ghulam, A.; Tiyip, T.; Pla, F.; Latorre-Carmona, P.; Halik, Ü.; Sawut, M.; Caetano, M. Effects of green space spatial pattern on land surface temperature: Implications for sustainable urban planning and climate change adaptation. *ISPRS J. Photogramm. Remote Sens.* **2014**, *89*, 59–66, doi:10.1016/j.isprsjprs.2013.12.010.

33.   Li, W.; Bai, Y.; Chen, Q.; He, K.; Ji, X.; Han, C. Discrepant impacts of land use and land cover on urban heat islands: A case study of Shanghai, China. *Ecol. Indic.* **2014**, *47*, 171–178, doi:10.1016/j.ecolind.2014.08.015.

34.   Guo, G.; Wu, Z.; Chen, Y. Complex mechanisms linking land surface temperature to greenspace spatial patterns: Evidence from four southeastern Chinese cities. *Sci. Total Environ.* **2019**, *674*, 77–87, doi:10.1016/j.scitotenv.2019.03.402.

35.   FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps Available online: https://www.umass.edu/landeco/research/fragstats/fragstats.html (accessed on May 20, 2021).

36.   Pan, W.; Du, J. Impacts of urban morphological characteristics on nocturnal outdoor lighting environment in cities: An empirical investigation in Shenzhen. *Build. Environ.* **2021**, *192*, 107587, doi:10.1016/j.buildenv.2021.107587.

37.   Akila, M.; Earappa, R.; Qureshi, A. Ambient concentration of airborne microbes and endotoxins in rural households of southern India. *Build. Environ.* **2020**, *179*, 106970, doi:10.1016/j.buildenv.2020.106970.

38.   Liu, Q.; Zhang, S.; Zhang, H.; Bai, Y.; Zhang, J. Monitoring drought using composite drought indices based on remote sensing. *Sci. Total Environ.* **2020**, *711*, 134585, doi:10.1016/j.scitotenv.2019.134585.

39.   O'Neill, R. V.; Krummel, J.R.; Gardner, R.H.; Sugihara, G.; Jackson, B.; DeAngelis, D.L.; Milne, B.T.; Turner, M.G.; Zygmunt, B.; Christensen, S.W.; et al. Indices of landscape pattern. *Landsc. Ecol.* **1988**, *1*, 153–162, doi:10.1007/BF00162741.

40.   He, H.S.; DeZonia, B.E.; Mladenoff, D.J. An aggregation index (AI) to quantify spatial patterns of landscapes. *Landsc. Ecol.* **2000**, *15*, 591–601, doi:10.1023/A:1008102521322.

Chapter 8. Conclusion

8.1. Conclusion

With the development of urbanization, the urban environment is deteriorating. The most notable phenomenon is the formation and enhancement of urban heat island. Urban heat islands are ubiquitous in metropolitan area around the world. Japan, as a developed country, includes 14 metropolitan areas. Most of these metropolitan areas are also affected by urban heat islands

This study introduced a temperature spatial downscaling method based on machine learning algorithm to downscale air temperature from 1 km to 250 m for high-resolution atmosphere urban heat island analysis. The core of this downscaling method is to establish the regression model between urban structure and temperature, and then we used the unchanged characteristics of regression models at different scale to predict high-resolution temperature data with high-resolution resolution urban structure, thereby analyzed atmosphere urban heat island. Finally, we compared the similarity and differentiation between atmosphere urban heat island and surface urban heat island.

The main works and results can be summarized as follows:

In *Chapter One*, PREVIOUS STUDY AND PURPOSE OF THE STUDY, this chapter introduce the research background and significance of urban heat island and the application of temperature spatial downscaling. In addition, we reviewed the previous studies of urban heat island and temperature spatial downscaling method. Finally, we elaborated on the purpose of this study.

In *Chapter Two*, DOWNSCALE AIR TEMPERATURE PREDICTION BASED ON MACHINE LEARNING MODEL, this chapter is the core of this study, mainly about the detailed air temperature spatial downscaling process. Firstly, it is the basic principles of temperature spatial downscaling. This study used statistical downscaling, the principle of which is to use the same regression between local underlying surface features and temperature at different scales. Secondly, we compared three different three different types of spatial downscaling models, including thermal sharpen models, multiple linear regression models, and random forest models. By comparison, we found that the random forest model based on machine learning algorithm has the best downscaling effect. Thirdly, we selected the independent variables, namely the underlying surface characteristic. Based on the previous research, we Finally adopted the urban structure and digital elevation model (DEM) as the independent variables of the model. Fourthly, five models based on machine learning algorithm were chose to test downscaling accuracy by the last independent variables and air temperature. The results showed that the extra trees model has the highest prediction accuracy. Fifthly, we obtained the optimal models for air temperature downscaling prediction through parameters tuning. Finally, we use low-resolution real air temperature and predicted air temperature to fit the residuals to obtain the final air temperature downscaling results.

In *Chapter Three*, URBAN STRUCTURE IDENTIFICATION AND ITS IMPLICATION OF URBAN CLIMATE, we selected Kanto Major Metropolitan Area as a case study to identify the urban structure by using maximum likelihood classification method based on remote sensing data and

satellite images. We divided the study area into 17 types, including 10 built types and 7 land cover types according to local climate zone system. Through accuracy validation based on the error matrix, we confirmed that the maximum likelihood classification method can identify each local climate zone with high accuracy. In addition, we used ENVI-met, a micro-climate simulation tools, to simulate urban micro-climate of some ideal models of different local climate zone types to evaluate and verify the impact of various urban structure on urban climate. The results showed that urban structure have a significant influence on urban climate, especially the air temperature, that is, there is a close relationship between the air temperature and urban structure. It provided the strong theoretical support for the downscaling models in Chapter 2.

In *Chapter Four*, CASE STUDY AND PRECISION EVALUATION, we selected Kanto Major Metropolitan Area as a case study and establish the extra trees model to assist in air temperature spatial downscaling research. Based on the resampled low-resolution air temperature downloaded from Japan National Land Numerical Information, digital elevation model, and urban structure identified by method mentioned in Chapter three, we built an optimal extra trees model by parameters tuning. And then, we inputted digital evaluation model and urban structure with a spatial resolution of 250 m into the extra trees model to predict the high-resolution air temperature data. Finally, we fitted the residual to obtain the downscaling air temperature data of study area and compared with measured air temperature from meteorological station to check the downscaling precision. The results showed that the proposed extra trees model is confirmed to be flexible and efficient in downscaling air temperature spatial resolution from 1 km to 250 m. The results passed the 10-fold cross validation and the verification of the measured air temperature data from the meteorological station.

In *Chapter Five*, DOWNSCALE AIR TEMPERATURE PREDICTION OF METROPOLITAN AREA BY EXTRA TREES MODEL, Japan is roughly divided into 5 climatic zones, and we selected the largest metropolitan area from each climatic zone as the research object, including Kanto Major Metropolitan Area, Kinki Major Metropolitan Area, Sapporo Major Metropolitan Area, Niigata Major Metropolitan Area, and Utsunomiya Metropolitan Area. With the help of the before introduced downscaling model, we selected the monthly average air temperature in January and August, and annual average air temperature in 2010 to downscale for the atmosphere urban heat island analysis in next chapters. Through cross-validation, we found that the downscaling results meet the accuracy requirements. The downscaling accuracy in August is generally higher than that in January. Among all metropolitan areas, Utsunomiya Metropolitan Area has the highest downscaling accuracy, and the downscaling accuracy of Sapporo Major Metropolitan Area in January is the lowest.

In *Chapter Six*, ATMOSPHERE URBAN HEAT ISLAND ANALYSIS OF METROPOLITAN AREA BASED ON HIGH-RESOLUTION AIR TEMPERATURE, we calculated the urban heat island intensity and urban heat island ratio index for the urban heat island evaluation of each metropolitan area according to the high-resolution air temperature downscaled in Chapter Five. The results showed

that the urban heat island intensity in January is generally higher than that in August. However, the regional distribution of urban heat island zone is the opposite. Except for the Utsunomiya Metropolitan Area, the urban heat island zone of the other four metropolitan area is more concentrated in January. The urban heat island zones of the Kanto Major Metropolitan Area and Niigata Major Metropolitan Areas area greatly affected by the distribution from the sea, while the urban heat island zones of Sapporo Major Metropolitan Area and Kinki Major Metropolitan Area are mainly concentrated in the center of the metropolitan area. The urban heat island zone of Utsunomiya Metropolitan Area is affected by Kanto Major Metropolitan Area and is concentrated in Utsunomiya city and the southern part of the metropolitan area.

In *Chapter Seven*, COMPARATIVE STUDY OF ATMOSPHERE URBAN HEAT ISLAND AND SURFACE URBAN HEAT ISLAND, we retrieved land surface temperature of Kanto Major Metropolitan Area and Kinki Major Metropolitan Area for surface urban heat island analysis. For comparing with atmosphere urban heat island with different temporal dimension, urban heat islands were normalized to the range between 0 and 1. Meanwhile, we introduced two landscape pattern metrics and used multiple linear regression model to indicate the affect of urban structure on urban heat island. The results showed urban heat island intensity of Kinki Major Metropolitan Area is always higher than that in Kanto Major Metropolitan Area, and the surface urban heat island intensity is always higher than the atmosphere urban heat island intensity. However, the surface urban heat island intensity in summer is much higher than that in winter, in contrast, the atmosphere urban heat island intensity in summer is slightly lower than that in winter. And there is a small cool island in the center of surface urban heat island, which does not exist in the atmosphere urban heat island. The multiple linear regression also proved this. And dense trees can greatly reduce both air temperature and land surface temperature, on the contrary, bare rock and pave greatly increase the air temperature and land surface temperature, and enhance the urban heat island.

In *Chapter Night*, CONCLUSION, this chapter presented the whole summary of each chapter.