



Article

# Study on Urban Heat Island Intensity Level Identification Based on an Improved Restricted Boltzmann Machine

Yang Zhang <sup>1</sup> , Ping Jiang <sup>2,\*</sup>, Hongyan Zhang <sup>3</sup> and Peng Cheng <sup>4</sup>

<sup>1</sup> College of Urban Economics and Public Administration, Capital University of Economics and Business, Beijing 100070, China; geozhangyang@yeah.net

<sup>2</sup> College of Resources and Environmental Science, Wuhan University, Wuhan 430079, China

<sup>3</sup> State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan 430079, China; zhanghongyan@whu.edu.cn

<sup>4</sup> College of Public Administration, Guangxi University, Nanning 530004, China; rsandgis@yeah.net

\* Correspondence: longkangkang@whu.edu.cn; Tel.: +86-186-2793-1818

Received: 5 December 2017; Accepted: 16 January 2018; Published: 23 January 2018

**Abstract:** Thermal infrared remote sensing has become one of the main technology methods used for urban heat island research. When applying urban land surface temperature inversion of the thermal infrared band, problems with intensity level division arise because the method is subjective. However, this method is one of the few that performs heat island intensity level identification. This paper will build an intensity level identifier for an urban heat island, by using weak supervision and thought-based learning in an improved, restricted Boltzmann machine (RBM) model. The identifier automatically initializes the annotation and optimizes the model parameters sequentially until the target identifier is completed. The algorithm needs very little information about the weak labeling of the target training sample and generates an urban heat island intensity spatial distribution map. This study can provide reliable decision-making support for urban ecological planning and effective protection of urban ecological security. The experimental results showed the following: (1) The heat island effect in Wuhan is existent and intense. Heat island areas are widely distributed. The largest heat island area is in Wuhan, followed by the sub-green island. The total area encompassed by heat island and strong island levels accounts for 54.16% of the land in Wuhan. (2) Partially based on improved RBM identification, this method meets the research demands of determining the spatial distribution characteristics of the internal heat island effect; its identification accuracy is superior to that of comparable methods.

**Keywords:** improved restricted Boltzmann machine; urban heat island; intensity level identification; green island; Wuhan; China

## 1. Introduction

Currently, the world is experiencing rapid and high-intensity urbanization. Along with the creation of the enormous social and economic benefits that urbanization brings, urbanization also profoundly changes city, regional and even global ecological environments [1–4]. The phenomenon known as urban heat island is one of the worst manifestations of urban ecological problems [5–9]. Howard L introduced the heat island effect in the London climate book in 1833, and the study of the urban heat island effect has interested scholars around the world over the last 200 years [10–15]. In recent years, remote sensing information technology has developed rapidly. With the advantages of time synchronization, wide monitoring and continuous detection, thermal infrared remote sensing has become one of the main technical methods for urban ecological and environmental monitoring [16–18].

The study of urban heat islands focuses on the distribution of spatial changes in the urban land surface temperature. Urban heat island intensity is a relative-temperature data concept, in opposition to the absolute temperature standard which is strictly divided by the absolute value of the temperature. In the current study on heat island intensity, there are many kinds of calculation methods to determine heat island intensity, which can be mainly summarized as four kinds. The first method defines the difference between the average temperature of the urban suburbs and the average temperature of the city center as the heat island intensity [19–22]. This method cannot resolve the spatial distribution of urban high-temperature zones. The second method defines the difference between the typical temperature of the urban suburbs and the typical temperature of the highest zones of the city center as the heat island intensity. The typical temperature values of the suburbs needed by this method are difficult to obtain. Regarding these two methods for the calculation of the heat island intensity, they are subject to data and observation conditions, the observation results are different, comparability is poor, and the calculations have much uncertainty. The third method uses the difference between the urban temperature and a reference green space as the heat island intensity [23,24], but this method also has problems, such as the method's computational complexity and the complicated background support needed for the large-scale urban heat island intensity level identification. The fourth method is the comparative light temperature method, which is divided by the intensity level to give the urban heat island effect. This method began with the work of Sun [25]. They derived the comparative light temperature method based on metaphorical criteria, which were proposed in the remote sensing information model by Ainai [26]. However, this method did not carry out any land surface temperature inversions, and the applicable conditions rely on the surface light temperature. In the division of the heat island intensity level, the majority of studies adopted the method of equal space based on a certain principle [27–29]. The disadvantage of this approach is the subjectivity when determining the best split point. The different division points divide the urban heat island structure, resulting in different spatial distributions and variation patterns, which are bound to cause some uncertainty in the research and analysis.

In light of this, this paper proposes a method that identifies the urban heat island intensity level based on an improved RBM (Restricted Boltzmann machine) model. The urban heat island intensity level identifier is obtained by applying a weak supervised learning technique. The identifier automatically initializes the annotation and optimizes the model parameters sequentially until the target identifier is completed, with very little information about the weak labeling of the target training sample, and an urban heat island intensity spatial distribution map is generated.

## 2. Research Area and Data

### 2.1. Research Area

Wuhan is the central city of the central region and the core city of the middle reaches of the Yangtze River city group. Wuhan is located in the central region of China, with a geographic location in east longitude of  $113^{\circ}41'$ – $115^{\circ}05'$  and north latitude of  $29^{\circ}58'$ – $31^{\circ}22'$ . The four boundaries of Wuhan form the shape of a butterfly, from west to east, in the spatial distribution model. The Yangtze River flows through Wuhan and intersects here with its largest tributary, the Hanjiang River. Wuhan is divided into the three parts—Hanyang, Hankou and Wuchang—which are commonly known as the “Wuhan three towns”, by the two rivers. Most of the terrain in Wuhan is dominated by plains; in the middle of Wuhan, residual hills are scattered from east to west, and there are many lakes and pond weirs in Wuhan. Wuhan is known as the “Hundred Lake City”, with more than 166 existing large and small lakes, of which 43 lakes are in the central city. The areas of water in Wuhan, up to 2217.6 km<sup>2</sup>, account for more than one-quarter of the total area of Wuhan, of which the lake area is 803.17 km<sup>2</sup>; the ratio of lake water is ranked first among the major cities of China. Wuhan is a privileged center in terms of China's economic geography; water and land transportation conditions are very developed.

“The centre of nine provinces” reputation has existed since the Qing Dynasty. The elevation of the study area is shown in Figure 1.

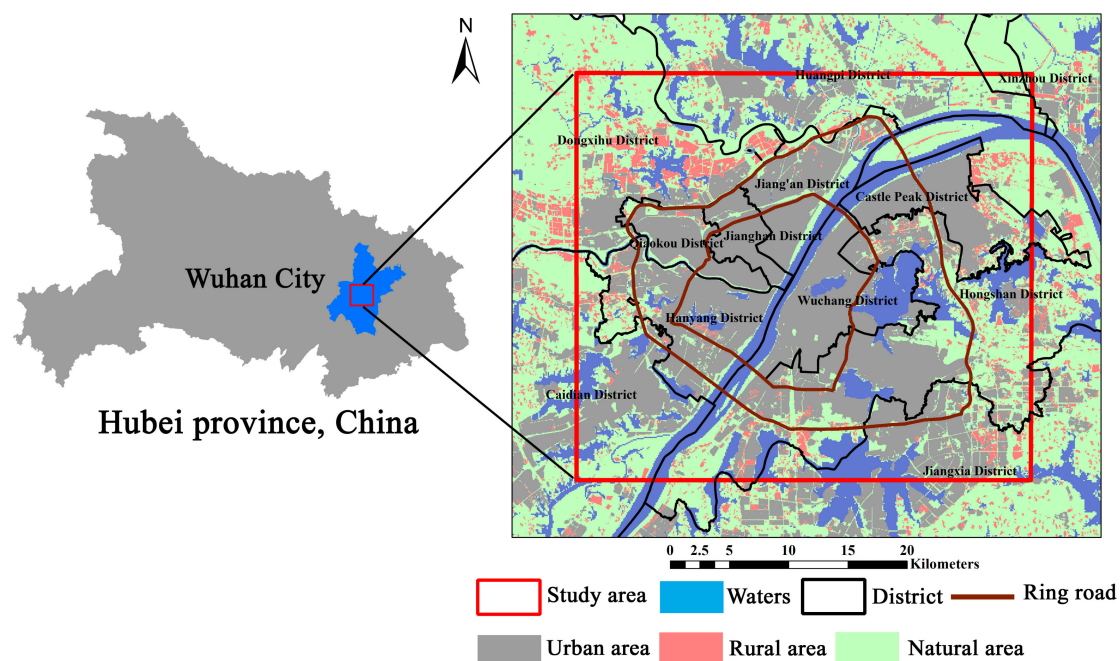


Figure 1. Map of study area.

2.2. Data

Because Landsat/TM (Thematic Mapper), ETM+ (Enhanced Thematic Mapper) and TIRS (Thermal infrared scanner) image data have the advantages of high spatial resolution and easily obtain rich archive time series and data, they have been widely used in land surface temperature inversion of urban heat islands. Therefore, Landsat 8 satellite TIRS image data were selected for this study’s data; the information related to the image acquisition is shown in Table 1. The collected remote sensing images of the study area have no cloudiness, which eliminates the error caused by the cloud cover factor in land surface temperature inversion.

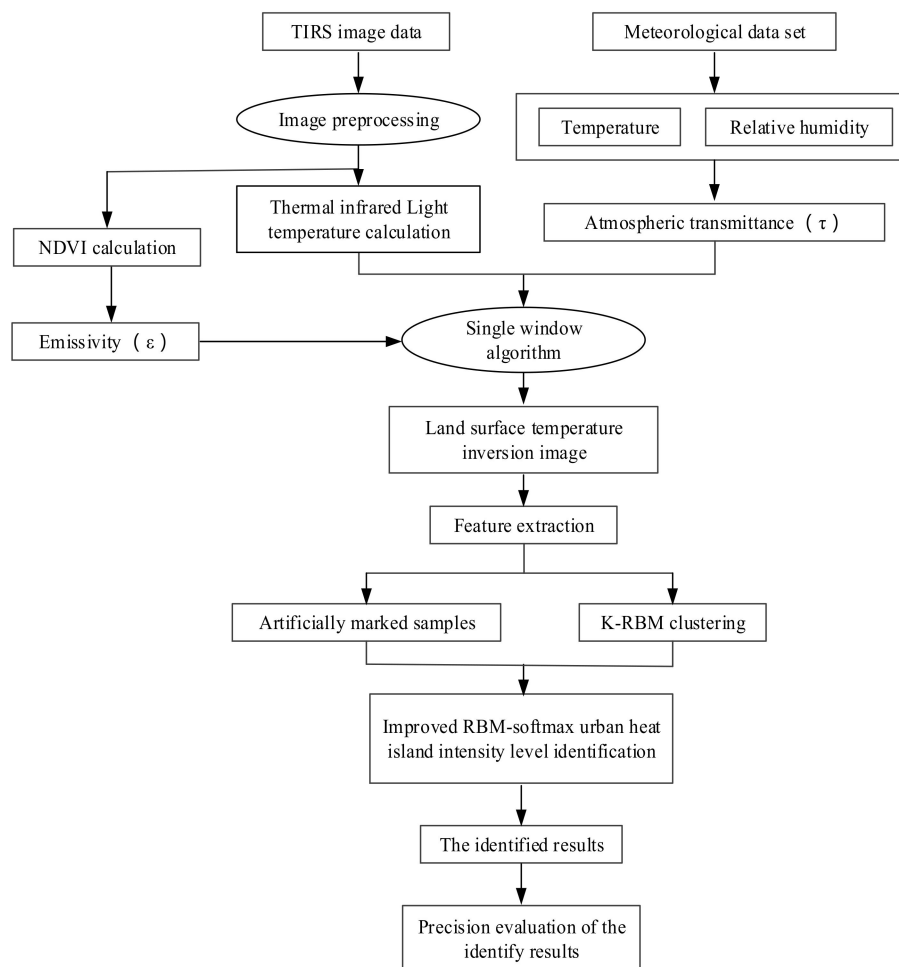
Table 1. Acquisition information of remote sensing image data.

Acquisition Time	Line	Number	Image Type	Thermal Infrared Band Spatial Resolution
23 July 2016	123	39	Band 1 Coastal	30 m
			Band 2 Blue	30 m
			Band 3 Green	30 m
			Band 4 Red	30 m
			Band 5 NIR (Near infrared)	30 m
			Band 6 SWIR 1 (Short-wave infrared)	30 m
			Band 7 SWIR 2	30 m
			Band 8 Pan	15 m
			Band 9 Cirrus	30 m
			Band 10 TIRS 1	100 m
			Band 11 TIRS 2	100 m

### 3. Research Methods

#### 3.1. Technical Route

The core steps of urban heat island intensity level identification include land surface temperature inversion, K-RBM clustering (K means clustering of RBM), improved RBM, and reconstruction of urban heat island intensity levels based on the improved RBM. First, the image is pre-processed with the 2016 TIRS data images (23 July 2016, Goddard Space Flight Center, Greenbelt, MD, USA), with the Landsat 8 satellite (National Aeronautics and Space Administration, Washington, DC, USA) as the data source. The projection mode is the Universal Mercator Projection, WGS (World Geodetic System) 1984/UTM (Universal Transverse Mercator) Zone 49N, and the resolution resampling is 30 m by cubic convolution method. Then, in accordance with Rozenstein [30], the single window algorithm can be used for urban land surface temperature inversion. After land surface temperature inversion, K-RBM clustering is performed. The samples with a high confidence level are marked in the clustering result. Based on the principle of minimum error, the most valuable sample set is composed of samples with an artificial mark, which were entered into the improved RBM-softmax urban heat island intensity level identification method. During training, the model parameters are determined by the steepest descent method until convergence, thus finally obtaining the identification results, and the accuracy of the classification is calculated. The core flow of the urban heat island intensity level identification process is shown in Figure 2.



**Figure 2.** Technical route. RBM: restricted Boltzmann machine; NDVI: normalized vegetation index.

### 3.2. Land Surface Temperature Inversion

The thermal infrared band of the TIRS remote sensing image of Wuhan is introduced into ENVI software (5.3, Harris Corporation, Melbourne, FL, USA). This is the basic data source for urban heat island analysis. While using ArcGIS software (10.2, Environmental Systems Research Institute, Redlands, CA, USA) to convert the required vector data to raster data, a basic image database is established, in a unified format, using the ENVI remote sensing software.

This study aims to quantitatively evaluate the spatial and temporal patterns and changes of the urban heat island. Therefore, surface temperature is the core indicator for this urban heat island research. In addition, Landsat, ETM+ and TIRS remote sensing images are used as data sources, and the inversion of the land surface temperature is performed using a single window algorithm.

#### (1) Calculate the radiation light temperature

Using the Landsat8 data users handbook provided by NASA (National Aeronautics and Space Administration) landsat science official website to get the model for calculating the radiation brightness temperature [31]:

$$L_{\lambda} = Gains \times DN + Biases \quad (1)$$

$$T = K_2 / \ln(K_1 / L_{\lambda} + 1) \quad (2)$$

where  $L_{\lambda}$  is the radiation intensity received by the TM remote sensor. Gains is the gain factor [unit:  $(W \cdot m^{-2} \cdot sr^{-1} \cdot \mu m^{-1}) / DN$ ]; DN is the digital number. Biases is the offset coefficient (unit:  $W \cdot m^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$ ).  $T$  is the radiation light temperature (unit: K);  $K_1$  and  $K_2$  are preset constants.  $L_{\lambda}$  is the luminance value of the unit spectral range (unit:  $mW \cdot cm^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$ ). The values of  $K_1$  and  $K_2$  of band 10 of Landsat8 TIRS are 774.89 and 1321.08.

#### (2) Calculate the specific emissivity

The specific emissivity of object  $\varepsilon$  is a manifestation of the ability of the object to emit electromagnetic radiation, which is a key parameter in surface temperature inversion. The calculation of emissivity is based on the empirical emissivity ration formula proposed by Van de Griend et al. [32].

$$\varepsilon = 1.009 + 0.047 \ln(NDVI) \quad (3)$$

where  $NDVI$  is the normalized vegetation index [33],  $NDVI > 0$  or  $\varepsilon = 0$ .

#### (3) Land surface temperature inversion

Real surface temperature inversion, using the single window algorithm model, was proposed by Rozenstein [30]. The inversion model is:

$$T_s = A_0 + A_1 T_{10} - A_2 T_{11} \quad (4)$$

where  $LST$  is the land surface temperature (unit: K),  $T_{10}$  and  $T_{11}$  are the radiation light temperatures (unit: K),  $\varepsilon$  is the emissivity of the object, and  $\tau$  is the atmospheric transmittance. The specific calculation process is as follows:

$$\begin{aligned} A_0 &= a_{10} E_1 - a_{11} E_2 \\ A_1 &= 1 + A + b_{10} E_1 \\ A_2 &= A + b_{11} E_2 \\ E_1 &= D_{11} (1 - C_{10} - D_{10}) / E_0 \\ E_2 &= D_{10} (1 - C_{11} - D_{11}) / E_0 \\ A &= D_{10} / E_0 \\ E_0 &= D_{11} C_{10} - D_{10} C_{11} \end{aligned} \quad (5)$$

where  $C_{10}$ ,  $C_{11}$ ,  $D_{10}$  and  $D_{11}$  are calculated as follows:

$$\begin{aligned} C_i &= \varepsilon_i \tau_i \\ D_i &= [1 - \tau_i][1 + (1 - \varepsilon_i) \tau_i] \end{aligned} \tag{6}$$

where the values of  $a_i$ ,  $b_i$  and  $\tau_i$  are shown in Tables 2 and 3. We can get these parameters in [30].

**Table 2.** Inversion regression coefficients of TIRS in different temperature ranges ( $r_{10}^2$  and  $r_{11}^2$  are the determinants of the fit).

Temperature Range/ $^{\circ}$ C	$a_{10}$	$b_{10}$	$r_{10}^2$	$a_{11}$	$b_{11}$	$r_{11}^2$
0–30	−59.139	0.421	0.9991	−63.392	0.457	0.9991
0–40	−60.919	0.428	0.9985	−65.224	0.463	0.9985
10–40	−62.806	0.434	0.9992	−67.173	0.47	0.9992
10–50	−64.608	0.44	0.9986	−69.022	0.476	0.9986

**Table 3.** Relationships between atmospheric transmittance ( $\tau$ ) and water vapor ( $\omega$ ) content when the water vapor content is from 0.5 to 3 g/cm<sup>2</sup>.

Atmospheric Mode	Atmospheric Transmittance Estimation Equation	$r^2$
The United States in 1976 standard atmosphere	$\tau_{10} = -0.1146\omega + 1.0286$	0.9982
	$\tau_{11} = -0.1568\omega + 1.0083$	0.9947
Mid-latitude summer	$\tau_{10} = -0.1134\omega + 1.0335$	0.9986
	$\tau_{11} = -0.1546\omega + 1.0078$	0.9996

### 3.3. K-RBM Clustering

According to the technical route, K-RBM clustering is required before the urban heat island intensity level is identified. The clustering algorithm is as follows:

- (1) Initialize K-RBMs, and initialize the rate of error change (EC = 15) of the sum of the reconstruction errors of the eigenvectors of all the pixels in the sample. Initialize all elements in the class chart  $Y$  to 0,  $Y = \{y_{m,n} \mid 1 \leq m \leq M, 1 \leq n \leq N\}$ ;  $y_{m,n}$  indicates the class of any pixel ( $m, n$ ) in the sample. The number of the subscript  $k$  in the  $3 \times 3$  neighborhood of ( $m, n$ ) in the initialized graph  $Y$  is  $N3_{m,n}(k) = 0$ .

- (2) Input the feature vectors of all the pixels ( $A$ ) into the K-RBMs.

The feature vectors of all the pixels ( $m, n$ ) are input into the K-RBMs, and the reconstruction error ( $E_{m,n}^k$ ) of the K-RBMs for  $F_{m,n}$  is calculated. Based on the principle of minimum error, the extracted feature vector ( $F_{m,n}$ ) is classified into K groups ( $S_k$ ).

$$F_{m,n} \in S_k = \left\{ F_{m,n} \mid k = \operatorname{argmin} E_{m,n}^k \right\} \tag{7}$$

- (3) Correct the error of the preliminary classification results. In this step, the mean value ( $P_k$ ) of the eigenvector ( $F_{m,n}$ ) in  $S_k$  is calculated; then, the distance ( $D_{m,n}^k$ ) between  $F_{m,n}$  and the cluster center is calculated.

$$D_{m,n}^k = \|F_{m,n} - P_k\|_2 + \beta N3_{m,n}(k) \tag{8}$$

$$\beta = \begin{cases} 0 & \text{if } Q > T1 \\ 0.4 & \text{else} \end{cases}$$

where  $T1$  represents the threshold.

Based on the principle of error minimization, the feature vector ( $F_{m,n}$ ) is divided into  $K$  groups ( $G_k$ ). The sample class chart  $Y$  is updated according to the  $k$  class of the corresponding pixel of  $F_{m,n}$  in  $G_k$ .

$$F_{m,n} \in G_k = \left\{ F_{m,n} \mid k = \arg \min E_{m,n}^k \right\} \tag{9}$$

- (4) Train the  $K$ -RBMs by the feature vector ( $F_{m,n}$ ) in  $G_k$ . If the algorithm converges, the training is terminated, otherwise perform step 2.

### 3.4. Improved Restricted Boltzmann Machine

The restricted Boltzmann machine is a bipartite graph with two layers of structure. The hidden unit of the RBM models the distribution, based on the visible unit. There is no connection within the layer, and the layers are fully connected. Hinton [34] developed the basic RBM model as a tectonic depth belief network, which brought about widespread concern. Because the RBM units are random, the restricted Boltzmann machine can also be regarded as a random neural network. This model requires strong computing power and a fast learning algorithm, which is widely used in machine learning.

The limited Boltzmann model is given as the sample  $X_{I \times L} = \{x^{(1)}, x^{(2)}, \dots, x^{(L)}\}$  with  $x^{(l)} = \{x_1^{(l)}, x_2^{(l)}, \dots, x_I^{(l)}\} (1 \leq l \leq L)$ , which represents training sample  $l$ . The weight matrix of the visible and hidden layers of the RBM is  $W_{J \times I}$ , and when the sample input is  $X_{I \times L}$ , the state of the hidden layer is  $H_{J \times L} = W_{J \times I} \times X_{I \times L}$  and can be expressed as follows:

$$H_{J \times L} = W_{J \times I} \times X_{I \times L} = \begin{bmatrix} h_1^{(1)} & h_1^{(2)} & \dots & h_1^{(L)} \\ h_2^{(1)} & h_2^{(2)} & \dots & h_2^{(L)} \\ \dots & \dots & h_j^{(l)} & \dots \\ h_j^{(1)} & h_j^{(2)} & \dots & h_j^{(L)} \end{bmatrix} \tag{10}$$

In the formula, each column represents a characteristic corresponding to an input sample, and each row corresponds to the state of one of the hidden units in the hidden layer for all input samples.  $h_j^{(l)}$  represents the state of hidden unit  $j$  when the input is sample  $i$ .

The traditional standard RBM objective function is as follows:

$$\max_{w,b,c} \sum_{l=1}^L \log P(X^{(l)}) \tag{11}$$

$P$  is a small constant, and it can be used to control the sparseness, that is, to improve the RBM via sparse filtering.

$$P = s.t. \frac{1}{L} \sum_{l=1}^L h_j^{(l)}, \quad \forall j = 1, \dots, J \tag{12}$$

where  $h_j^{(l)}$  is the normalization of each row of the matrix of Formula (12) by  $l_2$ .

The re-normalized matrix is now the input. The new matrix now has new row  $j$  and row  $l$  elements. In Formula (12), the corresponding constraint on each hidden node is realized. The hidden node is now available to all samples. However,  $h_j^{(l)}$  is the result of normalization of the column, so the mean of the square of elements in each column is 1. In the context of different samples, the mean of the activated hidden nodes is basically the same. Because all the hidden nodes in the whole sample have achieved selectivity, only a small part of the hidden nodes is activated. This suggests that sparseness has been achieved.

The process of calculating  $h_j^{(l)}$  is as follows:

- (1)  $l_2$  is normalized for each row of matrix ( $H_{J \times L}$ ), using  $h_j^{(l)}$  to denote the value of row  $j$  and column 1 in the matrix after row normalization. The formula is expressed as follows:

$$\overline{h_j^{(l)}} = \frac{h_j^{(l)}}{\sum_{l=1}^L h_j^{(l)}} \quad (13)$$

- (2)  $l_2$  is normalized for the columns of the normalized matrix, using  $h_j^{(l)}$  to denote the value of row  $j$  and column 1 in the matrix after row normalization. The formula is expressed as follows:

$$\hat{h}_j^{(l)} = \frac{\overline{h_j^{(l)}}}{\sum_{j=1}^J \overline{h_j^{(l)}}} \quad (14)$$

Using this improved RBM in combination with a softmax classifier. The newly developed algorithm is called urban heat island intensity level identification based on RBM-softmax. This classifier will use the hidden layer in an RBM as the input layer of the softmax classifier for the classification.

### 3.5. Construction of the Urban Heat Island Intensity Level Identification Algorithm Based on the Improved RBM

The construction progress of the urban heat island intensity level identification algorithm, based on the improved RBM, is as follows:

- (1) A small amount of sample S1 is extracted from the image that has been subjected to the land surface temperature inversion and it is manually marked. The number of artificial markers in a small sample of S1 of each type of clustering center is calculated. At the same time, the unmarked samples (S2) are clustered based on the K-means algorithm.
- (2) The Euclidean distance between the two types of cluster centers is compared with that of S1 and S2. Based on the minimum error criterion, the set of assignments with higher confidence, achieved via the K-RBM algorithm, is selected and combined with S1 samples to form the most valuable samples.
- (3) For pre-training to improve the RBM, the unsupervised learning of the RBM is improved by using the full sample (S). The next step is carried out when the error between the modified training samples of the RBM and the reconstructed data is quite small. At the same time, the relevant parameters obtained from the pre-training improvement RBM are used in the next step to improve the RBM; otherwise, continue to perform pre-training operations.
- (4) The most valuable sample obtained in step 2 is entered into the improved RBM-softmax urban heat island intensity level identification model, training the model parameters by the steepest descent method until convergence.
- (5) The data are input into the trained RBM-softmax urban heat intensity level identifier; then, the identification result is obtained, and the classification accuracy is calculated.

The construction process of the RBM model is shown in Figure 3.



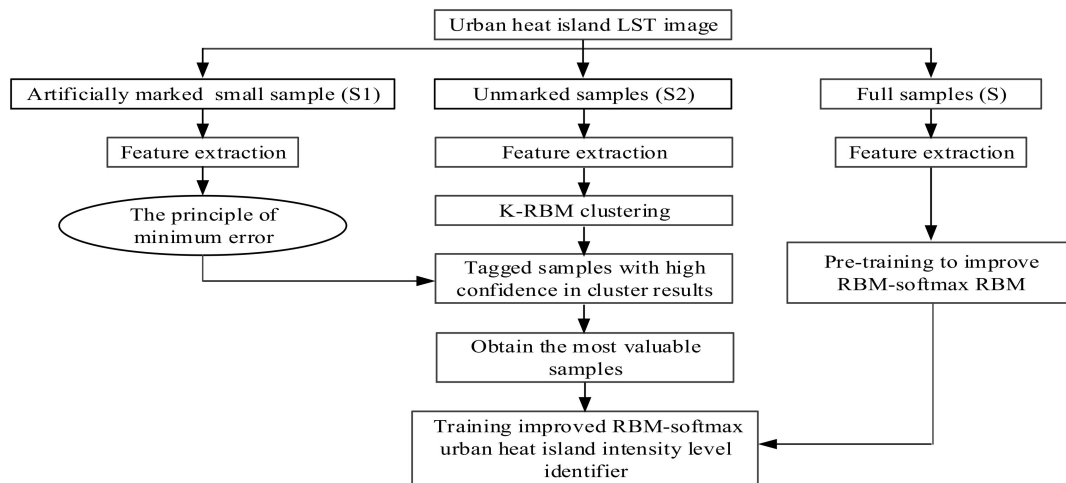


Figure 3. Construction process of the RBM model.

### 4. Results Analysis

#### 4.1. Spatial Distribution of Surface Urban Heat Island

The spatial distribution of the surface urban heat island (SUHI) is drawn, in accordance with the to the land surface temperature inversion method. The Wuhan urban heat island phenomenon is very significant, as shown in Figure 4. High temperatures have an obvious aggregation phenomenon in urban space, but water is a low value area for temperature. The maximum surface temperature is 310 K. This temperature will make people feel very hot.

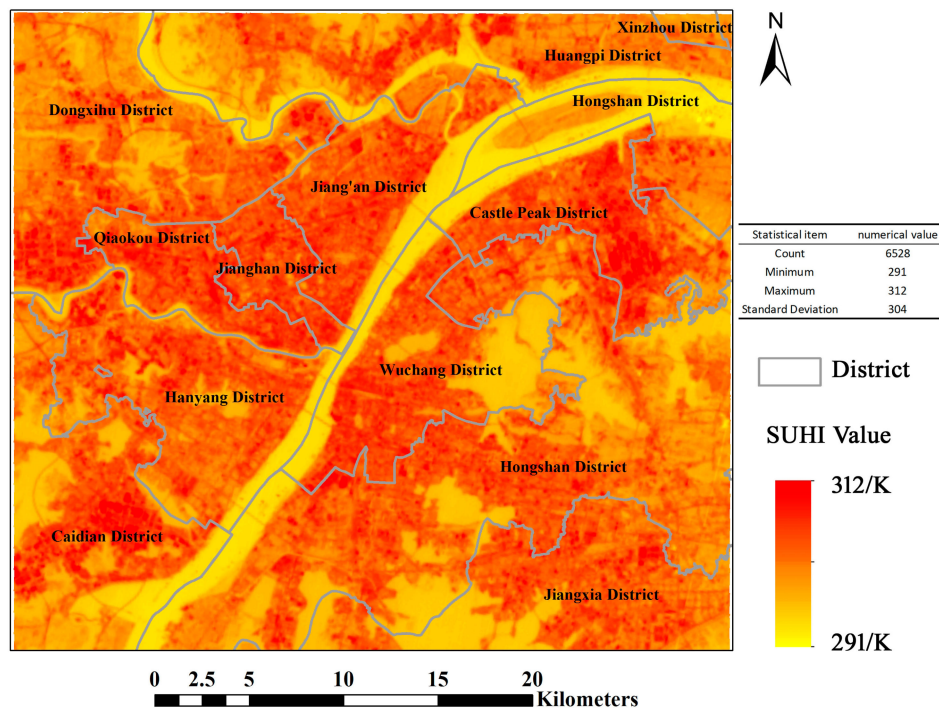


Figure 4. Spatial distribution of the surface urban heat island (SUHI) in Wuhan.

In order to clearly describe the variation in SUHI temperature, we plotted Figure 5.

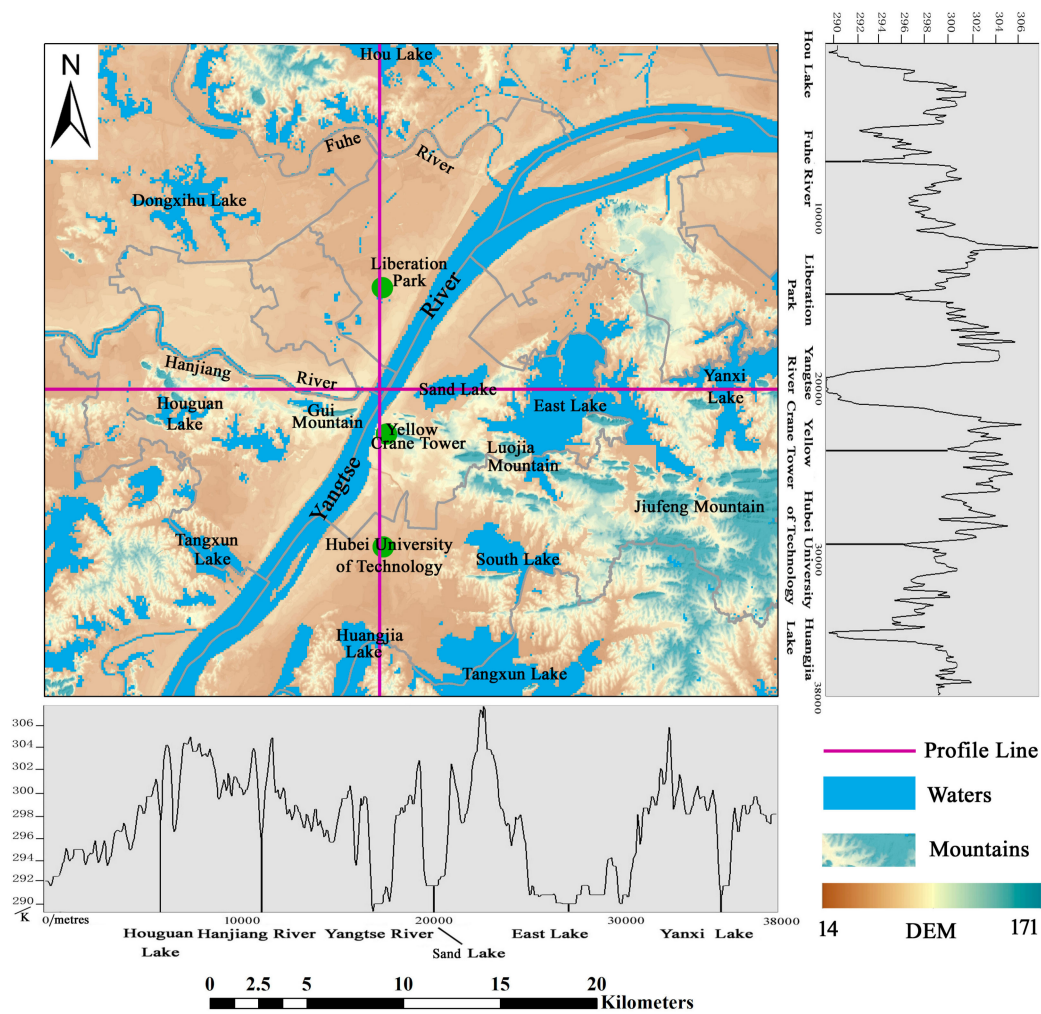


Figure 5. Land surface temperature changes in the profile line.

The land surface temperature profile for the east–west line in Wuhan, starts from Houguan Lake, through the Hanjiang River, Yangtze River, Sand Lake, East Lake, until Yanxi Lake and its surrounding farmland. In the land surface temperature profile from the east–west in Wuhan, a high, a middle, and a low are shown on both sides of the characteristics. This indicates the existence of urban heat island phenomenon. In addition, the east–west profile has the Yangtze River as its center. The temperature range on both sides of the east–west profile is asymmetric. The main reason for this is that the west area of urban buildings is widely distributed, while the east region of the river area is widely distributed.

From the east–west profile line, the overall urban center temperature is higher than the urban suburbs’ temperatures. The details show a “peak” or “valley” intertwined pattern. The main reason for this is the type of city under the surface. “Valley” mostly appears in farmland, water, and park green space and near the vegetation-covered university.

The north–south profile line, starts from Hou Lake, runs through the Fuhe River, Liberation Park, the Yangtze River, Yellow Crane Tower, Hubei University of Technology, and finally, Huangjia Lake to the surrounding farmland. From the land surface temperature of the north–south profile in Wuhan, a high, a middle, and a low on both sides of the characteristics can also be observed. It can be seen in Figure 5 that the extreme value of the profile line appears in the Jiang’an economic and technological development zone, on the north bank of the river, but the temperature range in the south is higher than that in the north. From the range of changes along the profile line of view, there are still high and low peaks and valleys staggered along the distribution trend. However, the intensity of the north–south profile is not as dramatic as that of the east–west profile.

From the range of the north–south profile line, the waters can form a trough of temperature. Some areas with small building densities and well-covered areas also form large and small cryogenic zones. This includes Hubei University of Technology, Wuhan University campus and other colleges and universities.

4.2. Identify the Results

Based on the improved RBM algorithm, an urban heat island intensity level identification map was generated. According to the regional distribution characteristics of the urban heat island and the thermal environment principle, the four categories were named green island, sub-green island, heat island and strong heat island.

The green island, sub-green island, heat island and strong heat island areas in Wuhan are 21,341 hm<sup>2</sup>, 37,753 ha, 50,525 hm<sup>2</sup> and 19,403 ha. The largest area is that of the heat island level, accounting for 39.19% of the total area of Wuhan, followed by the sub-green island, accounting for 29.28% of the total area of Wuhan. Green island accounts for 16.55% of the total area of Wuhan; the smallest area is that of the strong heat island, accounting for 14.97% of the total area of Wuhan. The total area of the heat island and strong island levels accounted for 54.16% in Wuhan (Table 4); the phenomenon of the heat island effect in Wuhan is obvious, and the heat island areas are widely distributed.

Table 4. Urban heat island level statistical results in Wuhan.

Heat Island Level	Area	Area Ratio
Green island	21,341	16.55%
Sub-green island	37,753	29.28%
Heat island	50,525	39.19%
Strong heat island	19,304	14.97%

Figure 6 shows that the urban heat island effect in Wuhan is obvious. The main urban areas form an island-like high-temperature area with a clear boundary, which is in sharp contrast to the urban fringe.

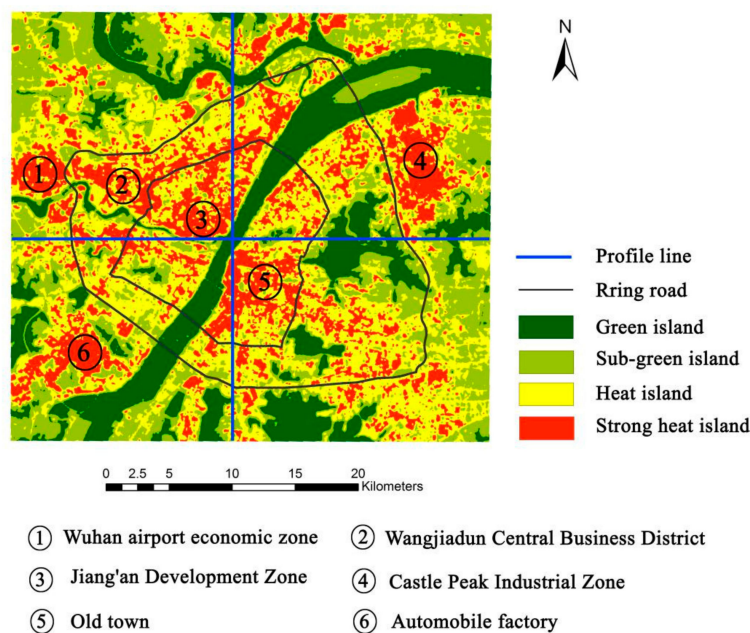


Figure 6. Urban heat island intensity level identification results in Wuhan.

The green island is mainly distributed in the Yangtze River, Han River, East Lake, Sand Lake, Tangxun Lake and other water areas of Wuhan. The sub-green island is mainly distributed in Wuhan Jiufeng Mountain, Luoja Mountain, Guishan and other mountains, Yellow Crane Tower, Liberation Park, Evergreen Park and the surrounding areas of farmland and other areas. The hot island is located in the central city of Wuhan and has a large distribution area. The boundary between the heat island and green island, the sub-green island, has a clear line, forming a sharp contrast. The strong heat island is mainly distributed along the Castle Peak Industrial Zone, Wuhan Airport Economic Zone, Jiang'an Development Zone, Wangjiadun Central Business District, the automobile factory area and urban residential intensive areas of Wuhan.

The cause of the heat island effect, in addition to the underlying surface of the media, is the city's unique heat source conditions, such as plant-based industrial areas, densely populated commercial and living areas, steel mills and power plants. These will only increase and deepen the heat intensity of certain areas.

The surface temperature of industrial-intensive areas was significantly higher than that of other areas in Wuhan. The Castle Peak Industrial Zone is the main distribution area of the strong heat island because it is in the city's unique heat source zone, where the Wuhan Iron and Steel Plant and Wuhan cooking plants are located. Strong heat islands are mainly distributed in the Jiang'an Development Zone and Wuhan Airport Economic Zone. The Hankou Wangjiadun Central Business District and the old town also form strong heat islands, mainly because these places are in the city where resident activity is intensive.

Regarding the hottest areas that can be seen, most of the "heat" comes from the industrial building areas; secondary heat sources are mainly distributed in dense commercial areas and residential areas. Heavy industry production companies, steel mills and other enterprises in the production industry produce a lot of heat. On the other hand, with the rapid development of the city, the emerging urban areas block the cooling space of these factories and groups, so that the heat island effect in the factory areas is greatly strengthened. At the same time, the vegetation in the factory areas is relatively small, and the solar radiation is strong, which further increases the agglomeration degree of the heat island effect.

From the heat island spatial distribution map, the distribution of water and green space is clearly seen. This distribution of water and green space has a high impact on the urban surface light temperature, which is obvious. Around the several large lakes and urban parks of the main city of Wuhan, there is a green island center in Wuhan. In addition, the large area of water and green areas around some of the urban land also showed a lower temperature, due to an edge effect. The cooling effects of water and vegetation on the urban heat island are of great practical significance for segmenting and controlling the spatial distribution of the heat island within the city.

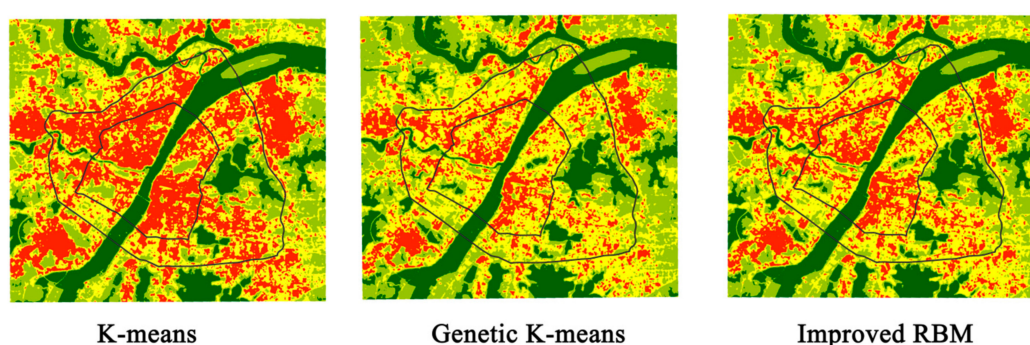
#### *4.3. Comparison of Identification Results*

The improved RBM identification algorithm was compared with two other models: K-means clustering and genetic K-means clustering [35]. All three models were tested 100 times, and the optimal results were taken. The comparative analysis of three-model is shown in Table 5 and Figure 7, respectively. Under the same experimental conditions, the improved RBM identification algorithm has the highest classification accuracy, with a total accuracy of 93.31% and a Kappa coefficient of 0.8861, which are higher than those for the K-means clustering algorithm and genetic K-means clustering. Among the three algorithms, the improved RBM identification algorithm had the shortest run time at 0.72 s. The genetic K-means clustering algorithm was the second best; however, the K-means clustering had the worst performance. K-means clustering algorithms easily fall into the phenomenon of local optimization because the initial clustering center is more sensitive, so the generated image clustering effect is poor. The differences between the internal temperatures of the city are not obvious, and it cannot achieve the research goals. However, the genetic K-means clustering algorithm does not fall into the local optimum. The algorithm has a good classification effect, and the computational efficiency

is higher than that of the K-means clustering algorithm. Compared with the K-means and the genetic K-means, the improved RBM identification algorithm is more obvious in terms of the division of the temperature levels within the city, and the boundary of the categories are obvious. The clustering result of the algorithm can meet the research goal of determining the spatial distribution of the heat island effect within the city. This research method is more practical.

**Table 5.** Identification accuracy comparison of different methods.

Identification Model	K-Means Clustering	Genetic K-Means Clustering	Improved RBM Identifier
Total Accuracy	73.39%	91.22%	93.31%
Kappa	0.6725	0.8735	0.8861
Testing time	2.87	0.91	0.72



**Figure 7.** Comparison of identification results obtained by different methods.

## 5. Discussion

This study reveals that the close relationship between urban heat island intensity and urban land types. In addition, the urban heat island intensity is also strongly correlated with urban morphology. For example, the geometry of different blocks will affect people's outdoor thermal comfort [36,37]. The research method proposed in this study can also be applied to the study of urban microclimate, which is also a new direction worthy of further study.

The RBM is a variant of the BM (Boltzmann machine, BM), but its limited model must be binary chart. The DBM (Deep Boltzmann machine, DBM) overcomes the problem. The DBM is made up of many RBM stacks. The DBM layers the pre-training RBM with an unsupervised greed pre-training method and the calculation can be made from the initial value of the probability model of supervised learning training. In this way, the learning performance of the whole model can be greatly improved. This will be a continuous exploration direction for this research issue and it can continue to improve the accuracy and computational efficiency of heat island intensity level identification.

## 6. Conclusions

The heat island intensity in an urban space should be a relative temperature concept. It is not an absolute temperature standard that is strictly divided by the absolute value of temperature. The identification model based on the improved RBM heat island strength level was established to solve the problem of subjectivity when dividing the optimal segmentation point of the current heat island. The division points of the heat island intensity under different standards change the urban heat island structure and produce different spatial differentiation rules, which inevitably cause some uncertainty in the research and analysis. Based on the improved RBM model, the urban heat island intensity level identifier is constructed using the idea of weak supervised learning. The identifier automatically initializes the annotation and optimizes the model parameters step-by-step until the target identifier is found. Using very little information about the weak labeling of the target training

samples, the urban heat island intensity spatial distribution map is generated. This method also has certain advantages, in comparison with other classification models. Experimental results showed the following:

- (1) The largest heat island area is that of the heat island level in Wuhan. Regarding the hottest areas that can be seen, most of the “heat” comes from the industrial building area; secondary heat sources are mainly distributed in dense, commercial areas and residential areas. Heavy industry production companies, steel mills and other enterprises in production projects produce a lot of heat. From the heat island spatial distribution map, the distribution of water and green space can also be clearly seen; the impact of the distribution of water and green space on urban surface light temperature is very obvious. Around the several large lakes and urban parks of the main city of Wuhan, a green island center of Wuhan is formed.
- (2) For the RBM identification algorithm, the total accuracy was 93.31%, the kappa coefficient was 0.8861, and the test time was 0.72 s. The algorithm had a good classification effect, and the computational efficiency was higher than that of the K-means clustering algorithm. Compared with K-means and the genetic K-means, the improved RBM identification algorithm is more obvious in terms of the division of the temperature level within the city, and the boundary of the category is obvious. The clustering effect of the algorithm can determine the spatial distribution of the heat island effect within the city, and this method is more practical than others.
- (3) Based on the improved RBM model, this paper analyzed the urban heat island and compared it to those obtained with other methods. The focus of the study the spectral characteristics of surface temperature after inversion, which includes texture features for further feature extraction. This method may be enhanced by adding other methods to complete the feature extraction, such as an automatic encoder method. In the process of RBM learning, it is possible to obtain a positive training sample image and a negative training sample image, through the use of a sliding window, to obtain a better distribution rule of the data.

**Acknowledgments:** The authors would like to thank the anonymous editor and reviewers, who gave valuable comments and helped to improve the quality of the paper. This work is supported by the National key Research and Development Program of China: Geographic big data mining and spatiotemporal pattern discovery (ID: 2017YFB0503600), the National Social Science Fund (ID: 14CGL054) and Open Research Fund Program of Key Laboratory of Digital Mapping and Land Information Application Engineering NASG (ID: GCWD201808).

**Author Contributions:** Y.Z. and P.J. conceived and designed the experiments; P.J. and P.C. performed the experiments and analyzed the data; H.Z. contributed analysis tools; Y.Z. wrote the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Nomenclature

$L_{\lambda}$	Radiation intensity received by the TM remote sensor
DN	Digital number
$\varepsilon$	Specific emissivity
$T$	Radiation light temperature
$A_0, A_1, A_2$	Coefficient of Formula (5)
$\tau$	Atmospheric transmittance
$\omega$	Water vapor
$a_i, b_i$	Regression coefficients of TIRS
$E_0, E_1, E_2, C_i, D_i$	Procedure parameter
$k, K$	Classification sequence number, number of classifications, respectively
$m, n$	Line sequence number of pixels, column sequence number of pixels, respectively
$M, N$	Maximum number of lines and columns in pixels
$i, I$	Sequence number of $X^{(l)}$ , total number of $X^{(l)}$ , respectively
$l, L$	Training sample sequence number, total number of training samples, respectively
$j, J$	Sequence number of hidden units, total number of hidden units, respectively

## References

1. Stoll, M.J.; Brazel, A.J. Surface-air temperature relationships in the urban environment of Phoenix, Arizona. *Phys. Geogr.* **1992**, *13*, 160–179.
2. Gosling, S.N.; McGregor, G.R.; Paldy, A. Climate change and heat related mortality in six cities part 1: Model construction and validation. *Int. J. Biometeorol.* **2007**, *51*, 525–540. [[CrossRef](#)] [[PubMed](#)]
3. Imhoff, M.L.; Zhang, P.; Wolfe, R.E.; Bounoua, L. Remote sensing of the urban heat island effect across biomes in the continental USA. *Remote Sens. Environ.* **2010**, *114*, 504–513. [[CrossRef](#)]
4. Zhang, Y.; Jiang, P.; Chen, Y.Y. Study on heat island effect in Wuhan city based on land sat TM remote sensing. *Ecol. Environ. Sci.* **2012**, *21*, 884–889. (In Chinese)
5. 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* **2008**, *319*, 756–760. [[CrossRef](#)] [[PubMed](#)]
6. Gosling, S.N.; Lowe, J.A.; McGregor, G.R. Associations between elevated atmospheric temperature and human mortality: A critical review of the literature. *Clim. Chang.* **2009**, *92*, 299–3411. [[CrossRef](#)]
7. Buyantuyev, A.; Wu, J. Urban heat islands and landscape heterogeneity: Linking spatiotemporal variations in surface temperatures to land-cover and socioeconomic patterns. *Landsc. Ecol.* **2010**, *25*, 17–33. [[CrossRef](#)]
8. Tan, J.G.; Zheng, Y.F.; Tang, X. The urban heat island and its impact on heat waves and human health in Shanghai. *Int. J. Biometeorol.* **2010**, *54*, 75–84. [[CrossRef](#)] [[PubMed](#)]
9. Clinton, N.; Gong, P. MODIS detected surface urban heat islands and sinks: Global locations and controls. *Remote Sens. Environ.* **2013**, *134*, 294–304. [[CrossRef](#)]
10. Voogt, J.A.; Oke, T.R. Thermal remote sensing of urban climates. *Remote Sens. Environ.* **2003**, *86*, 370–384. [[CrossRef](#)]
11. Fung, W.Y.; Lam, K.S.; Nichol, J.; Wong, M.S. Derivation of nighttime urban air temperatures using a satellite thermal image. *J. Appl. Meteorol. Climatol.* **2009**, *48*, 863–872. [[CrossRef](#)]
12. Rizwan, A.M.; Dennis, L.Y.; Chunho, L.I.U. A review on the generation, determination and mitigation of urban heat island. *J. Environ. Sci.* **2008**, *20*, 120–128. [[CrossRef](#)]
13. Peng, B.F.; Shi, Y.S.; Wang, H.F. The impacting mechanism and laws of function of urban heat islands effect: A case study of Shanghai. *Acta Geogr. Sin.* **2013**, *68*, 1461–1471.
14. Ren, Z.; He, X.; Zheng, H.; Zhang, D.; Yu, X.; Shen, G.; Guo, R. Estimation of the relationship between urban park characteristics and park cool island intensity by remote sensing data and field measurement. *Forests* **2013**, *4*, 868–886. [[CrossRef](#)]
15. Haashemi, S.; Weng, Q.H.; Darvishi, A.; Alavipanah, S.K. Seasonal variations of the surface urban heat island in a semi-arid city. *Remote Sens.* **2016**, *8*, 352. [[CrossRef](#)]
16. Zhou, H.M.; Zhou, C.H.; Ge, W.Q. The surveying on thermal distribution in urban based on GIS and remote sensing. *Acta Geogr. Sin.-Chin. Ed.* **2001**, *56*, 197–204. (In Chinese)
17. Weng, Q.H.; Rajasekar, U.; Hu, X.F. Modeling urban heat islands and their relationship with impervious surface and vegetation abundance by using ASTER images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4080–4089. [[CrossRef](#)]
18. Xie, Q.J. Analysis on characteristics and influencing factors of urban heat island effect in Wuhan. *Resour. Environ. Yangtze Basin* **2016**, *25*, 463–469. (In Chinese)
19. Li, L.G.; Liu, X.M.; Zhao, X.L.; Wang, H.B.; Wang, X.Y.; Zhang, Y.S. Characteristics of heat island effect in inner and outer suburbs of Shenyang and the relationships with urbanization. *J. Appl. Ecol.* **2010**, *21*, 1609–1613. (In Chinese)
20. Li, L.G.; Wang, H.B.; Jia, Q.Y.; Guo, H.L.; Wang, X.Y.; Zhang, Y.S.; Ai, J.F. Urban heat island intensity and its grading in Liaoning province of Northeast China. *J. Appl. Ecol.* **2012**, *23*, 1345–1350. (In Chinese)
21. Zhang, Y.J.; Alan, T.M.; Turner, B.L. Optimizing green space locations to reduce daytime and nighttime urban heat island effects in Phoenix, Arizona. *Landsc. Urban Plan.* **2017**, *165*, 162–171. [[CrossRef](#)]
22. Dos Santos, A.R.; de Oliveira, F.S.; da Silva, A.G.; Gleriani, J.M.; Gonçalves, W.; Moreira, G.L.; Silva, F.G.; Branco, E.R.F.; Moura, M.M.; da Silva, R.G.; et al. Spatial and temporal distribution of urban heat islands. *Sci. Total Environ.* **2017**, *605*, 946–956. [[CrossRef](#)] [[PubMed](#)]
23. Yokobori, T.; Ohta, S. Effect of land cover on air temperatures involved in the development of an intra-urban heat island. *Clim. Res.* **2009**, *39*, 61–73. [[CrossRef](#)]

24. Yan, H.; Fan, S.; Guo, C.; Wu, F.; Zhang, N.; Dong, L. Assessing the effects of landscape design parameters on intra-urban air temperature variability: The case of Beijing, China. *Build. Environ.* **2014**, *76*, 44–53. [[CrossRef](#)]
25. Sun, S.M.; Lu, C.Y. Study on monitoring intensity of urban heat island and taking it as an indicator for urban ecosystem by remote sensing. *J. Xiamen Univ. Nat. Sci.* **2002**, *41*, 70–73. (In Chinese)
26. Ainai, M.A. Remote sensing information model and geographic mathematics. *Acta Sci. Nat. Univ. Pekin.* **2001**, *37*, 558–562.
27. Xu, H.Q.; Chen, B.Q. An image processing technique for the study of urban heat island changes using different seasonal remote sensing data. *Remote Sens. Technol. Appl.* **2003**, *18*, 129–133. (In Chinese)
28. Zhang, Z.M.; He, G.J.; Xiao, R.B.; Wang, W.A. A study of the urban heat island changes of Beijing city based on remote sensing. *Remote Sens. Inf.* **2005**, *6*, 46–48. (In Chinese)
29. Yang, W.R.; Li, F.; He, Y. Characteristic change and analysis of urban heat island in Beijing, China, Summer, 2003–2011. *Acta Ecol. Sin.* **2014**, *34*, 4390–4399.
30. Rozenstein, O.; Qin, Z.H.; Derimian, Y.; Karnieli, A. Derivation of land surface temperature for Landsat-8 TIRS using a split window algorithm. *Sensors* **2014**, *14*, 5768–5780. [[CrossRef](#)] [[PubMed](#)]
31. Zanter, K. Landsat8 (L8) Data Users Handbook. Landsat Science Official Website. Available online: <https://landsat.usgs.gov/landsat-8-l8-data-users-handbook> (accessed on 20 January 2018).
32. Van de Griend, A.A.; Owe, M. On the relationship between thermal emissivity and the normalized difference vegetation index for nature surfaces. *Int. J. Remote Sens.* **1993**, *14*, 1119–1131. [[CrossRef](#)]
33. Black, A.; Stephen, H. Relating temperature trends to the normalized difference vegetation index in Las Vegas. *GISci. Remote Sens.* **2014**, *51*, 468–482. [[CrossRef](#)]
34. Hinton, G.E.; Osindero, S.; Teh, Y.W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [[CrossRef](#)] [[PubMed](#)]
35. Zhang, H.Y.; Zhai, H.; Zhang, L.P.; Li, P.X. Spectral-Spatial Sparse Subspace Clustering for Hyperspectral Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3672–3684. [[CrossRef](#)]
36. Tania, S.; Koen, S.; Andreas, M. Analysis of microclimatic diversity and outdoor thermal comfort perceptions in the tropical megacity Dhaka, Bangladesh. *Build Environ.* **2015**, *94*, 734–750.
37. Marcel, I.; Nyuk, H.W.; Steve, K.J. Urban microclimate analysis with consideration of local ambient temperature, external heat gain, urban ventilation, and outdoor thermal comfort in the tropics. *Sustain. Cities Soc.* **2015**, *19*, 122–135.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).