Long-term Evolution Patterns of Qingdao City based on Historical Maps and Remote Sensing Images

Xuechi Li [a], Shan Zhao [b] and Hongli Zhang [c]

School of Geoscience and Technology, Zhengzhou University, Zhengzhou Henan, 450001, China

[a] 15695253627@163.com, [b] zhaoshan4geo@zzu.edu.cn, [c] 1760456734@qq.com

Abstract. Current studies on urban evolution heavily rely on either remote sensing imagery or historical maps, seldom integrating both sources. Additionally, there is a scarcity of research on urban long-term evolution exceeding half a century. This study employs an innovative approach that combines historical maps and remote sensing imagery. Pre-remote sensing era, data from modern historical maps were extracted to reveal urban evolution patterns. In the 1960s-1980s, using early panchromatic remote sensing imagery due to limited urban maps in China, the study explored urban evolution. Post-1980s, higher-resolution Landsat series imagery was utilized. The integration of historical maps and remote sensing imagery forms a coherent, long-term urban evolution sequence, facilitating the discovery of comprehensive urban change patterns. The study involves extracting construction areas for urban expansion analysis, employing a combination of visual interpretation and random forest extraction methods for enhanced efficiency and accuracy.

Keywords: Urban Evolution; Historical Maps; Long-term Evolution; Construction Areas; Random Forest Extraction Methods.

1. Introduction

This study explores the morphological evolution of Qingdao's old city from the early Republican period to the present. It employs historical map restoration and remote sensing images to showcase the city's historical patterns across different time periods. The research takes a broad perspective, investigating the "object representation" and "internal value" of the old city's modernization. It delves into the phenomena, mechanisms, and motivations driving spatial transformations in Qingdao's urban space, reassessing the impact of early and recent modernization on urban form and social space. Led by Wang Fang [1], researchers reconstructed the area and capacity of Lake Bositeng in 1935 using military topographic maps and other data from the Xinjiang Military Region. The study highlights the high measurement accuracy of military topographic maps from this period, especially those with a scale of 1:50,000 and above, making them suitable for reconstructing historical geographic features. In 2018 [2], Xu Peng and colleagues addressed the information gap on urban development in Xi'an, a newly established national central city. They utilized remote sensing methods and Landsat TM/ETM/OLI satellite images from 1993, 2003, 2013, and 2017 to extract information on land use and cover changes. Various methods, including expansion index, quadrant analysis, concentric circle analysis, and vegetation coverage, were employed to analyze the spatiotemporal characteristics of recent urban expansion in Xi'an and assess changes in urban expansion and land use [3]. Research on urban evolution by scholars has predominantly focused on spatial morphological pattern evolution, expansion patterns, driving mechanisms, ecological effects, and other aspects of the urban built-up area.

2. Method

The flowchart in Figure 1 illustrates the four main components of this article: Landsat satellite data processing, Keyhole satellite image processing, night light data processing, and ancient map processing.

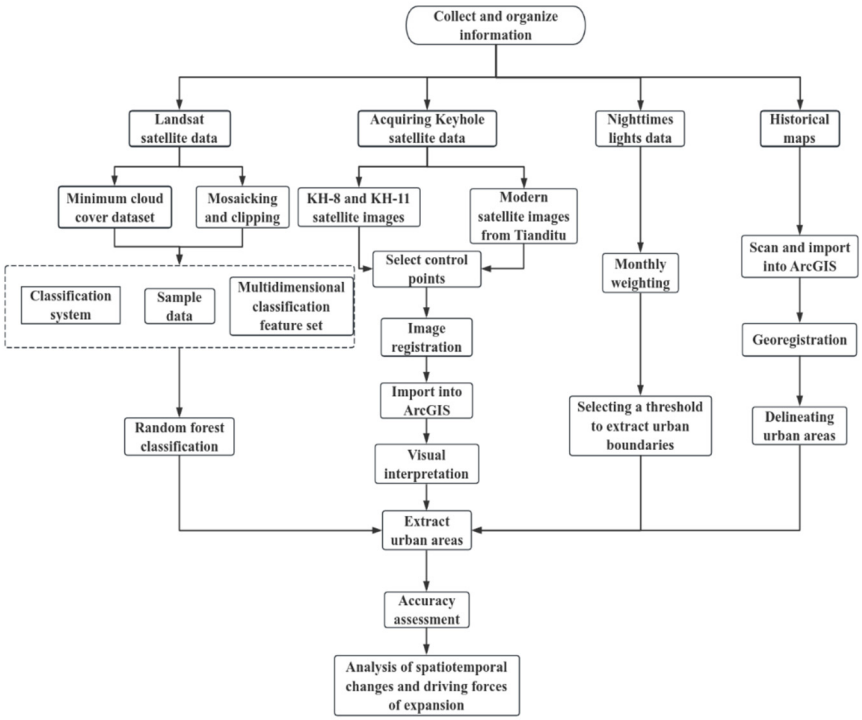


Fig 1. Flow chart of this article

2.1 Random Forest Method

The random forest method is an enhancement of the decision tree algorithm, combining both classification and regression decision trees. The integrated classification algorithm employs bootstrap resampling technology to randomly resample the original training sample set with replacement. Multiple samples are drawn to create a new training sample set, and then multiple decision trees are constructed based on these new training sets. These decision trees collectively form a random forest, and the ultimate classification result is determined by the voting scores from all decision tree classification results. The decision formula is:

$$C = \arg\max_{c}\left(\sum_{k=1}^{n} I\left(h\left(x,\theta_{k}\right)=C\right)\right) \tag{1}$$

in the formula: C is the category to which the test sample belongs; n is the number of decision trees; $I(A)$ is an indicator function, that is, when A is true, the value is 1, otherwise it is 0;$h(x,\theta_k)$is a decision tree classifier; x is a training sample;$\theta_k$is an independent and identically distributed random vector.

This method exhibits high accuracy and computational efficiency, enabling the testing of various feature attributes during the classification process. Concerning the significance of classification results, the random forest effectively mitigates the impact of individual decision trees by averaging their outcomes, thus minimizing the risk of overfitting. Simultaneously, it remains robust and stable, unaffected by the results of individual decision trees.

2.2 Spectral Feature Classification Method

This article utilizes the GEE cloud platform to extract classification features from images within the study area, encompassing multispectral bands and three spectral indices. To fully leverage the spectral information within Landsat-5 TM and Landsat-8 OLI images and enhance classification accuracy, the B1-B7 band of the Landsat-5 TM image and the B1-B11 band of the Landsat-8 OLI image were chosen as classification spectral bands. Additionally, the normalized vegetation index

(NDVI), normalized water index (NDWI), and normalized building index (NDBI) were computed to emphasize information related to vegetation, water, and buildings.

The Normalized Difference Vegetation Index (NDVI) is defined as the normalized ratio index of near-infrared band reflectance and visible light red band reflectance. It is widely used in the extraction and monitoring of vegetation coverage areas. The formula is as follows:

$$NDVI = \frac{NIR - R}{NIR + R} \tag{2}$$

In the formula, *NIR* refers to the reflectivity of the near-infrared band, and *R* refers to the reflectivity of the red light band.

The Normalized Water Index (NDWI) is defined as the normalized ratio index of the visible light green band reflectance and the near-infrared band reflectance. It is widely used in the extraction and monitoring of water bodies. The formula is as follows:

$$NDWI = \frac{G - NIR}{G + NIR} \tag{3}$$

In the formula, *G* refers to the reflectivity of the green light band, and *NIR* refers to the reflectivity of the near-infrared band.

The Normalized Building Index (NDBI) is defined as the normalized ratio index of near-infrared band reflectance and mid-infrared band reflectance. It is widely used in the extraction and monitoring of urban building land. The formula is as follows:

$$NDBI = \frac{MIR - NIR}{MIR + NIR} \tag{4}$$

In the formula, *MIR* refers to the reflectivity in the mid-infrared band, and *NIR* refers to the reflectivity in the near-infrared band.

2.3 City Outline Extraction Method based on Night Light Data

Utilizing the processed NPP/VIIRS annual average nighttime light data, the threshold segmentation method is employed for urban area extraction. Initially, a DN (Digital Number) value is set, and urban areas are extracted using Landsat imagery as the reference. The DN value is iteratively adjusted by 0.01 through visual interpretation, with the optimal threshold being the DN value closest to the urban area. Mask processing is applied to filter out non-urban areas, resulting in the extraction of urban areas. Given the annual average nature of the nighttime light data, the extracted area represents the annual urban extent. Subsequently, urban area elements within this range are delineated. This sequential approach of first extracting urban areas and then urban area elements helps mitigate interference from irrelevant features like saline-alkali land, bare land, and woodland in the classification process.

2.4 Visual Interpretation Method

The Geographic Information System (GIS) method is applied in this study. To minimize unnecessary deformation during the projection process and streamline area calculations, the topographic map is employed, and the projection and registration are conducted following the WGS-1984-Qingdao projection method. The Georeferencing module in ArcGIS 10.6 is utilized, specifically the Add Control Point tool, for point-to-point registration. Landmark points from remote sensing images are incorporated for meticulous fine-tuning. Ultimately, the interpretation and area calculations are performed using Polygon tools and Geometry tools, categorizing features into different types, with units presented in square kilometers.

2.5 Classification Accuracy Evaluation Method

The accuracy of land use classification results profoundly impacts the precision of land use change analysis. Consequently, this paper employs the confusion matrix method to scrutinize the outcomes of land use classification, introducing three evaluation metrics: Producer Accuracy (PA), Overall

Accuracy (OA), and the Kappa coefficient for a comprehensive accuracy assessment. Producer Accuracy denotes the likelihood that actual features of a specific category are accurately classified. For k-category producer accuracy, the formula is:

$$P_{PA} = \frac{P_{kk}}{\sum_{j=1}^{n} P_{kj}} \tag{5}$$

The overall accuracy is defined as the percentage of pixels that are correctly classified out of the total number of pixels and is expressed as:

$$P_{OA} = \frac{1}{N} \sum_{i=1}^{n} P_{ii} \tag{6}$$

The Kappa coefficient combines two accuracy indicators, producer accuracy and user accuracy, and analyzes the consistency between classification results and prior category information. It is a quantitative evaluation method. The Kappa coefficient overcomes the problem that the confusion matrix depends on samples and sampling methods through discrete multivariate analysis methods, and can be expressed as:

$$K = \frac{N \sum_{k=1}^{n} P_{kk} - \sum_{k=1}^{n} \left(\sum_{i=1}^{n} P_{ki} \sum_{j=1}^{n} P_{kj} \right)}{N^2 - \sum_{k=1}^{n} \left(\sum_{i=1}^{n} P_{ki} \sum_{j=1}^{n} P_{kj} \right)} \tag{7}$$

In the above formula, n is the total number of columns in the confusion matrix, that is, the total number of categories; P_{ii} represents the number of samples in the i-th row and i-th column of the confusion matrix, that is, the number of correctly classified samples; $\sum_{i=1}^{n} P_{ki}$ and $\sum_{j=1}^{n} P_{kj}$ represent the total number of samples in column i and j, respectively; N represents the total number of samples. Generally speaking, a Kappa coefficient greater than 0.8 indicates that the classification results are highly consistent with the actual features, that is, the classification accuracy is high. A Kappa coefficient between 0.4 and 0.8 is considered to have moderate agreement, and a Kappa coefficient below 0.4 indicates that the classification result has a large error and cannot be used.

3. Experiment and Analysis

The outcomes of extracting city boundaries based on nighttime light data are illustrated in Figure 2.

In the case of Qingdao City, the random forest method was employed to extract urban construction land from Landsat declouded monthly composite images in 1988, 1998, 2008, and 2022. The errorMatrix() function was utilized to construct a confusion matrix, and accuracy was evaluated and averaged. The specific evaluation results are presented in Table 1. Due to its high accuracy in extracting construction land in urban areas and superior performance compared to other methods, the random forest algorithm was chosen for extracting construction land in urban areas within Qingdao municipal districts for the years 1988, 1998, 2008, and 2022. Visual interpretation was employed to extract urban construction land from Keyhole satellite data in 1968 and 1978, and from 1922 to 1956 based on old maps. The extraction results are depicted in Figure 3.

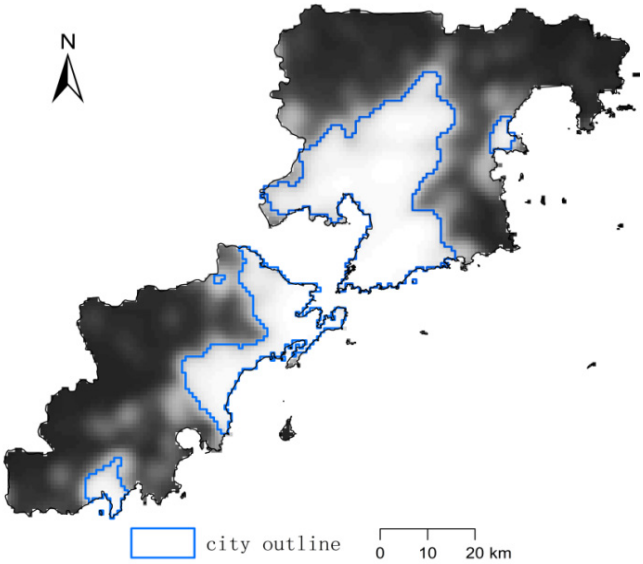


Fig 2. City outline extracted from nighttime light data

Table 1. Kappa coefficient of random forest classification per year.

Years	1988	1998	2008	2022
Kappa coefficient	0.84	0.83	0.78	0.81

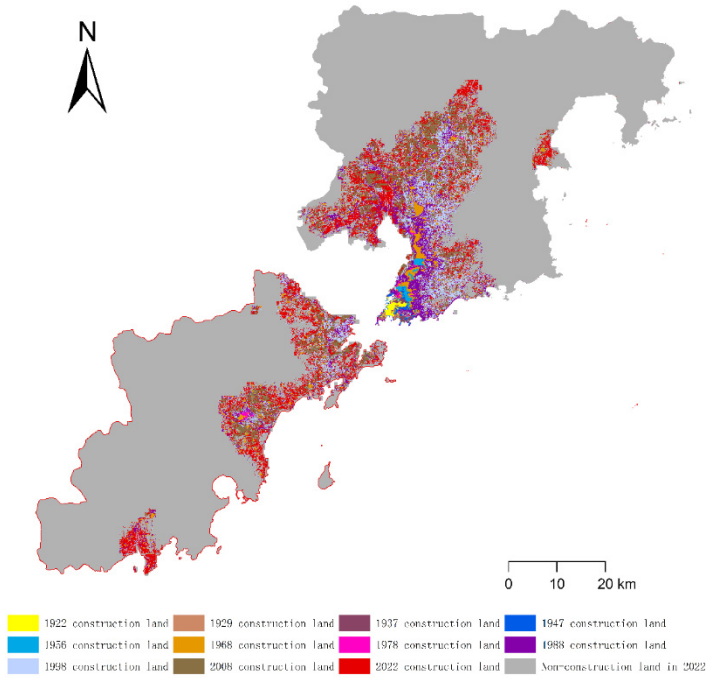


Fig 3. Construction land in Qingdao from 1922 to 2022

4. Conclusion

Statistics and analysis reveal that the urban construction land area in Qingdao was 6.60 km² in 1922 and is projected to reach 1033.31 km² by 2022. Over the past century, the total construction land area of all cities (including counties) has experienced continuous growth, expanding by 1026.71 km², marking a remarkable 155.56-fold increase over the course of 100 years. Notably, from 2008 to 2022, Qingdao witnessed the most rapid expansion in urban construction land, averaging a rate of 27.42

km² per year. Economic, demographic, and transportation factors are the primary driving forces behind the expansion of urban construction land, exerting a decisive influence on this expansion.

References

[1] Wang Fang, Pan Wei. Application experiments of three-dimensional technology in historical geomorphology research - Changes in Bosteng Lake, Xinjiang since 1935. Journal of Earth Environment, 2017, 3:253-262.

[2] Wan Zhiwei, Shao Haiyan. Reconstruction of all elements of land use in counties during the Republic of China period based on 1:50000 topographic maps: A case study of Qingjiang County, Jiangxi Province. Chinese Historical Geography. 2020, 35 (4): 32-42.

[3] Xu Peng, Du Ping. Remote sensing monitoring analysis of urban expansion and land use change in Xi'an city. Journal of Beijing United University, 2018,7.