Environmental Science

http://www.aimspress.com/

Research article

Spatial assessment of sewage discharge with urbanization in 2004–2014, Beijing, China

Huixuan Li[1],*, Cuizhen Wang[1], Yuqin Jiang[1], Andrew Hug[2] and Yingru Li[3]

[1] Department of Geography, University of South Carolina, Columbia, SC 29208, USA
[2] GIS Specialist, Georgia-Alabama Land Trust, Piedmont, AL36272, USA
[3] Department of Sociology, University of Central Florida, Orlando, FL. 32816, USA

* **Correspondence:** Email: huixuan@email.sc.edu; Tel: +1-334-750-5019.

Abstract: Beijing, China's cultural political, and economic center, is facing critical water pollution-related challenges warranting global attention. This study used remote sensing and geographic information systems to analyze the impact of urbanization on wastewater discharge in Beijing. Two influencing factors—urban index and environment index—were created from remote sensing image classifications to better reflect the impacts from urbanization and green-cover changes on wastewater discharge. The impacts of urban land uses on the volume of wastewater discharge were examined in localized areas (i.e., the so-called unit watersheds delineated from topography and stream segments). Geostatistical results showed that urbanization was primarily responsible for spatial variations of wastewater discharge. While vegetation helped ameliorate the pollution, increased urban areas on the outskirts of the city resulted in accelerated wastewater discharge. Analytical findings of this study could provide spatially explicit information for policy-makers to initiate and adjust protocols and strategies for protecting water resources and controlling wastewater emission, thus improving quality of living environments in Beijing.

Keywords: wastewater discharge; urban index; environment index; spatial analysis; Beijing

1. Introduction

Water nourishes life. Useable fresh water comprises less than 1% of the Earth's total water resources, and water quality has been increasingly degraded by anthropogenic activities. China is facing critical water resource shortages and severe water pollution issues, creating serious public

health concerns [1-4]. This issue has attracted increasing attention since the 1990s, [5] along with China's rapid economic growth and industrialization process [6]. Furthermore, lead levels in Chinese rivers are approximately 44 times higher than water quality standards in the 2007 National Academy of Sciences (NAS) report [7]. Beijing, the capital and a megacity of China, faces especially severe water pollution hazards. In one of Beijing's drinking reservoirs, lead levels over the past three years were 20 times higher than World Health Organization's (WHO) standards [8]. Contamination in water bodies can cause long-term human health risks, creating significant consequences in economic development, thereby damaging its environmental sustainability [3,4].

Conventional studies of water pollution in China have focused on in-situ sampling of a particular area [2,9-13] and have mainly contributed to identifying water pollution sources and controlling contaminators [14-17]. Both point source [18] and non-point source water pollutions [5] have been examined. In these studies, water samples collected from a local scale water system were often used to analyze parameters including temperature, pH, turbidity, etc. by employing physical-chemical experiments. These experiments are often labor intensive and time consuming. Miyun Reservoir is the most important drinking water source for the highly urbanized capital. Previous studies, such as Wang et al. (2002), studied non-point source water pollution using a hydrological model and Material Flow Analysis (MFA) to demonstrate that agricultural activities, including soil erosion and pesticide and fertilizer runoff, were primary sources of pollution [20]. Zhao et al. (2010) investigated heavy metal pollution due to urban runoff and road-deposit sediment [24].

With the popularity and development of Remote Sensing (RS) and Geographic Information System (GIS) technologies, an increasing number of studies have focused on water quality and pollution from a spatial perspective. Common examples include mapping water distribution, modeling sewer networks, detecting leakage, identifying and quantifying point and non-point water pollution sources, managing drinking water quality, as well as monitoring water pollution levels [13,14,25,26]. The Soil and Water Assessment Tool (SWAT), a very powerful watershed-based model, has been commonly applied to simulate environmental impacts of human and natural processes including land management, climate change, and agricultural activities [27]. Statistical methods including linear regression [28,29] spatial regression [6] and statistical hypothesis testing [30,31] have been widely applied to measure contamination levels and analyze relationships between water contamination levels and relevant factors [13]. Inverse Distance Weighting (IDW) is another common GIS-based spatial estimation method employed for analyzing water pollution [32,33].

For human-induced water pollution, industrial wastewater discharge, agricultural irrigation (with extensive use of pesticides and fertilizers), and urban sewage are ranked the top three pollution-contributing sources due to their high toxicities and untraceable characteristics [4,6,18,34]. In urban environments, municipal sewage significantly influences the quality of water bodies along with the urbanization process--especially considering China has the largest population density and is experiencing rapid social development. Vegetation coverage, including parks, forests, grasslands, etc., could have a positive influence on surface and ground water quality [35,36]. Researchers have identified over a dozen influencing factors to evaluate human and natural impacts to water quality [3,4]. Wang et al. (2008) investigated water crises associated with human activity in rural areas of China [18]. Huang et al. (2008) employed remote sensing images and technologies merely to analyze Beijing land cover/land use changes [37]. However, using both RS and GIS technologies to analyze influencing human and environmental factors on water quality in watershed based urban ecosystems, still remains limited and merits further study.

This study aims to investigate the spatiotemporal variation of wastewater discharge with respect to land use change, especially in the watershed-based urban area of Beijing.

2. Materials and Methods

2.1. Study Area

The study area covers the administrative district of China's capital city, Beijing. Northwest regions are dominated by mountains, forests, bare land, agricultural land, and water. South and southeast areas are highly urbanized and generate high levels of wastewater discharge. They are composed of 16 urban districts, in addition to suburban and rural townships totaling of 16,411 km^2; 1268 km^2 of which is urbanized. Total population reached 21.5 million in the southern and southeastern regions in 2014 with 18 million living in urban districts [38]. According to the Beijing Municipal Water Affairs Bureau and the Beijing Municipal Bureau of Statistics (2011), there are 425 rivers, 41 lakes, and 88 reservoirs in Beijing (Figure 1). In Li's report (2013), nearly half of the 88 river systems in the study area were listed in the worst water quality category [39]. While 60% to 70% of urban areas are outfitted with wastewater treatment plants, their limited capabilities leave a significant amount of water untreated, resulting in direct discharge into river systems.
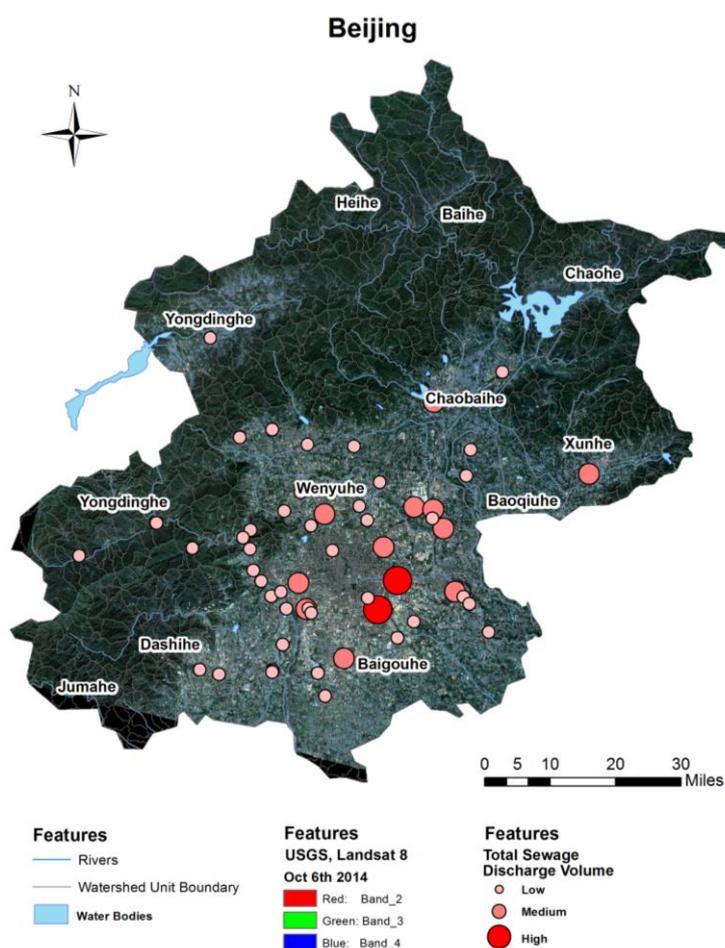

Figure 1. Study area and major water bodies, watershed boundaries and locations of wastewater discharge data collection sites.

2.2. Data

Five types of data were collected and analyzed, including water pollution, socioeconomic, hydrologic, and RS and GIS (Table 1). Hydrologic and RS data, including Landsat 5 TM, Landsat 8 OLI, and 30 meter resolution digital elevation model (DEM) from Shuttle Radar Topography Mission (STRM) were obtained from U.S. Geological Survey (USGS) at a Clearinghouse [40]. DEM data were used to generate watersheds around Beijing. Ground truth data points were randomly collected through imagery in Google Earth around the similar date of the TM and OLI images (Google Earth was chosen for its abundance of high resolution images around Beijing, China). A total of 210 random points over different classes were generated for validation of image classification. Beijing's socioeconomic information was obtained from China City Statistical Yearbooks, in which urbanization was represented by population density [41,42]. The wastewater discharge data was provided by Beijing City Drainage Refco Group Ltd., available from 2004 onward. This study thus examines urbanization of the study area and wastewater-induced water pollution from 2004–2014. Figure 1 shows the point data for a total of 59 sewage data sites at multiple locations. Points are displayed in graduated colors and sizes to indicate the pollution influencing power on surrounding areas. The higher the discharged wastewater volume, the larger the influencing power.

Table 1. Data sets utilized in this study.

Data	Data Type	Data Source	Description
RS image	Raster	Landsat 5 TM	Sep 4, 2004
		Landsat 8 OLI	Oct 6, 2014
		(Path 123 Row 032)	
		Google Earth image	Dec 5, 2005; April 20, 2015
GIS layer	Vector/Shape file	China Data Center	Administrative division and major rivers in Beijing
Water Pollution records	Point Shape file	Beijing city drainage Refco Group Ltd	Total treated sewage wastewater in 2004 and 2014
Hydrologic data	Raster	SRTM 1 Arc-Second Global	DEM data

2.3. Methods

This research was built upon the following analytical structure (Figure 2). Two Landsat images were first classified into the following seven land cover and use types: water, forest, agricultural land, bare soil, low-density urban area, medium-density urban area, and high-density urban area. Accuracy assessment was performed using ground truth points collected in Google Earth. Two influencing factors, urbanization index and environmental index, were developed with a class-specific weighting system and remote sensing satellite image classification. Those indices were statistically analyzed with wastewater discharged volume records to examine relationships between water quality and human activities on urbanization and natural land cover types.
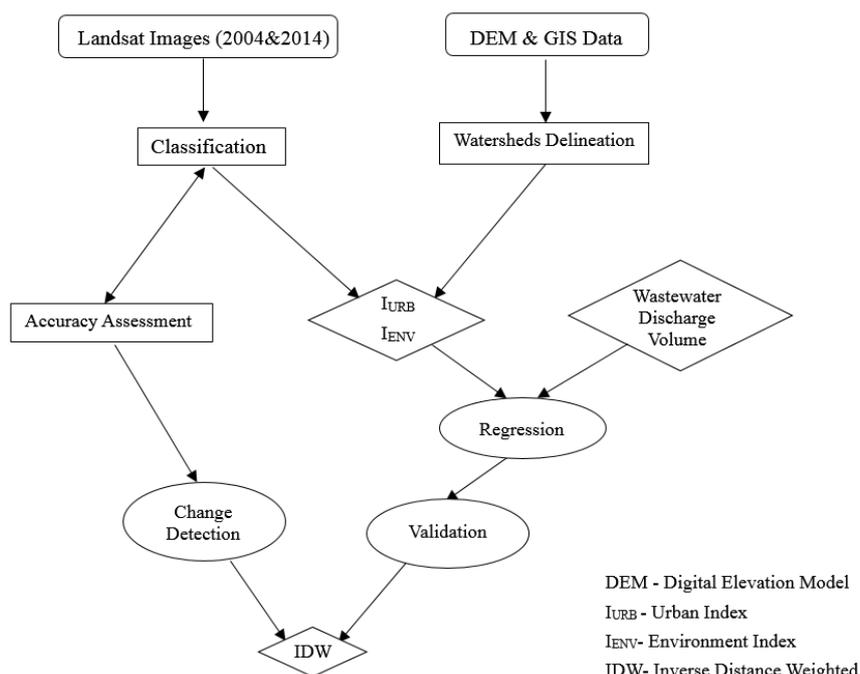
Figure 2. Flow charts of data analysis structure.

2.3.1. Satellite image classification and change detection

Two satellite images of Beijing, dated September 4, 2004 (Landsat 5 TM) and October 6, 2014 (Landsat 8 OLI) were clipped to the study area. With an unsupervised/supervised hybrid classification approach [43,44], 200 unsupervised clusters were generated and assigned one of the following seven classes: water, forest, agriculture, bare land, low-density urban area, medium-density urban area, and high-density urban area. Both row crops and perennial grasses were included as part of the agriculture class their differences were insignificant for this study. 'Bare land' influenced by human activity and naturally 'open land' were classified into low-density urban area to avoid overlap with the natural barren land. For accuracy assessment, the error matrix was extracted to evaluate classification results by using the following parameters: overall accuracy, commission/omission errors, and Kappa coefficient. The accuracy, or agreement, represented the likelihood that image pixels were assigned correct ground truth classes (i.e., correctly classified). Commission errors indicate overestimation, i.e., pixels that belong to a different land type but are assigned the present/current class. Omission error reveals underestimation, meaning the classified pixels belong to a specific land type but were assigned an incorrect class by the classification. Kappa coefficient (ranges between 0 and 1, or 0 to 100%) is a discrete multivariate indicator for classification accuracy. Higher Kappa indicates a higher accuracy of classification results.

Validated classification results were used to display the land use changes over this study period, and a change detection was performed to extract the changes between 2004 and 2014. Changes in urban area were the primary interest of this study. When quantifying the changes, corresponding bands from 2004 and 2014 images were compared in ArcMap using the Image Analysis tool. The changes were cell based, and every cell contained a digital number (DN) ranging from 0 to 255. The changes

were then calculated and quantified by DN values before applying a graduated colored scheme to visually display change detection results.

2.3.2. Watershed delineation

The Hydrology model in Arc Toolbox provides functions to construct stream polylines from DEMs. Said hydrology model calculates water flow direction and accumulation based on elevation change. With respect to watershed delineation in ArcMap, all cells exhibiting flow accumulation values above an assigned threshold value were considered a channel—a path which contains moving water (or connects two bodies of water). After assigning the threshold value, all upstream area of 180,000 m^2 or more, was defined as a channel. The area within one cell was 900 m^2, and the area within 200 cells was 180,000 m^2. Therefore, it was determined that any cell with an upstream area of 200 cells was a channel. Next the Raster Calculator was employed to create a layer of those cells meeting this criteria. Considering the study area size, more than 50 different cell sizes were tested to optimize the resulting watershed. Testing these values indicated a threshold of 25,000 cells, or 22.5 km^2 produced the best results for the catch water area of our study.

The Beijing DEM data were further divided into unit watersheds according to stream distributions. Each junction point of two or more streams was considered an outlet for one unit of watershed, which contained no more than one river segment. Therefore, each watershed unit was assumed to be a piece of relatively homogeneous land with similar topographical features and water quality. The Network Analyst module in ArcMap automatically generated the junction points in the river network, allowing for delineation of 790 unit watersheds within the whole study area (Figure 3). The striped areas in the northwest corner resulted from lack of stream data. These areas were excluded from data analysis.
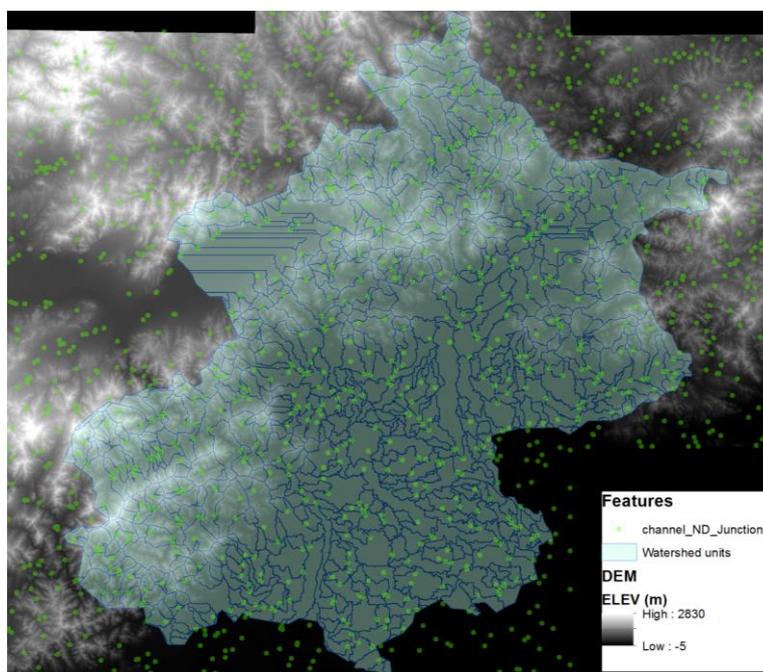

Figure 3. The DEM, stream outlet points, and self-delineated unit watersheds of the study area.

After obtaining classification results, pixel numbers of the seven land use types within each watershed unit were counted. These counted cells were then used to calculate the urban and environmental index.

2.3.3. Urban index, environmental index and spatial analysis

Sewage discharge was expected to be positively related to urban areas and negatively to natural lands, with respect to human-induced wastewater. For this reason, two indices were created from image-classified land covers to evaluate the spatial-temporal influence of anthropogenic activities on water quality--specifically, total sewage value. Within each watershed unit, both urban index (I_{URB}) and environmental index (I_{ENV}) were calculated from the classification results as follows:

$$I = (\sum i \times j)/(t \times \Sigma j) \qquad (1)$$

where i represents the area of a specific class, j is the score associated with class i, and t is the area of the watershed unit. The score of each class represents its influence on either an urban area (I_{URB}) or natural environment (I_{ENV}). The scores of each class in each index are summarized in Table 2.

Table 2. Scores of different land classes in two indices.

Class(i)	Urban Score(j_{URB})	Environment Score(j_{ENV})
Water	0	2
Forest	0	3
Bare Land	0	0
Low Density Urban	1	0
Medium Density Urban	2	0
High Density Urban	3	0
Agriculture Land	0.5	1

The I_{URB}, score value in Table 2 represents the weight of each land cover contributing to urban environments. Urban lands contribute more to urban environments. For example, high-density urban area was marked with 3. The greater the influence, the larger the weighted score was. Agriculture also carries an urban score considering human forces involved. Human impacts regarding other classes such as water, forest, and bare land are limited, and therefore, their scores were defined as 0.

For I_{ENV}, the score of j_{ENV} value in Table 2 quantifies the contribution of each land cover to natural environments, especially green cover. Forest, water bodies, and agricultural land were assigned weighted scores of 3,2,1 respectively, while bare land, low urban density area, medium urban density area, and high urban density area were assigned 0. The agriculture land was given a score of 1 considering crops had higher vegetation coverage compared to other land types, especially urban and bare land. Urban and environment indices in every watershed unit were calculated employing the same procedures as used in the following example: the sample watershed unit contains a total area of 22 cells (t = 22 × 30 m × 30 m = 19,800 m^2) from the Landsat images, including 10 cells (9000 m^2) of high density urban area, 5 cells of medium density urban area (4500 m^2), and 7 cells of forest (6300 m^2). According to equation 1, the I_{URB} would be (9000 × 3 + 2 × 4500)/19,800 × (2 + 3) = 0.36, while the I_{ENV} would be 0.32.

To examine the potential impact of urban areas and human activities to total sewage discharge in unit watersheds, the simple linear regression approach below was developed:

$$\log Y = aX + b + \varepsilon \tag{2}$$

where the dependent variable, Y, represents the volume of total sewage discharge collected at each data location (ton/day); 59 in total. X is indicated by both environment index (I_{ENV}) and urban index (I_{URB}) for each year of the study. Logarithmic function was applied with a representing the slope of the relationship between sewage discharge and indexes, b a constant number serving as the intercept of the regression, and ε, the error term.

Forty of the fifty-nine records were split into training samples, while the remaining 19 were used as validation samples. Regressions were applied using the training samples, which generated the estimation functions for relationships (slope a) between wastewater discharge volume and urban and environment indices. The 19 samples were used for statistical validation through Root Mean Square Error (RMSE) as seen below.

$$RMSE_y = \sqrt{\frac{\sum_n^1 (y_{predicted} - y_{observed})^2}{N}} \tag{3}$$

Error for each point refers to the difference between actual wastewater discharge volume and the predicted volume from the regression result. Final validation results yielded the averaged calculation from 5 experiments, the results of which are summarized in Table 5.

Finally, the inverse-distance-weighted (IDW) method was applied to create a continuous spatial layer to predict areas influenced by sewage discharge based on observed wastewater data. The IDW results were examined to demonstrate the overlap between predicted wastewater discharge and urban area distributions.

3. Results and Discussion

3.1. Classification and urban area changes

Beijing is highly urbanized in the central and southeast areas [40]. The north and southwest regions of Beijing are mountainous and are mostly covered with forests and agricultural lands (Figure 4A). The hybrid classification procedure resulted in an overall classification accuracy of 78% and 80% for 2004 and 2014, respectively (Tables 3 & 4). Some error developed due to ambiguity between mountain shadows and water and bare lands. Additionally, the delineation between low-density, medium-density, and high-density urban land was somewhat subjective resulting in possible additional bias. For urban lands specifically, the three urban classes presented 67–95% producer's accuracy and 65–100% user's accuracy in two years. Overall, the high-density urban category was better classified than low-and-medium-densities. The high commission and omission errors in low-and-medium-density urban classes may result from the similar spectral signatures between these two classes due to their commonly mixed pixels.

The change detection analysis focused on changes in urban areas, providing better insight of increases in the three urban classes in Beijing from 2004 to 2014 (Figure 4B). The more intense urban area changes were expressed through deeper tones of blue. Urbanization has been expanding from Beijing's urban core in the southwest of the study area. The development of Beijing Capital International Airport can also be identified in the Figure 4B in the central east area with the darkest blue tone.
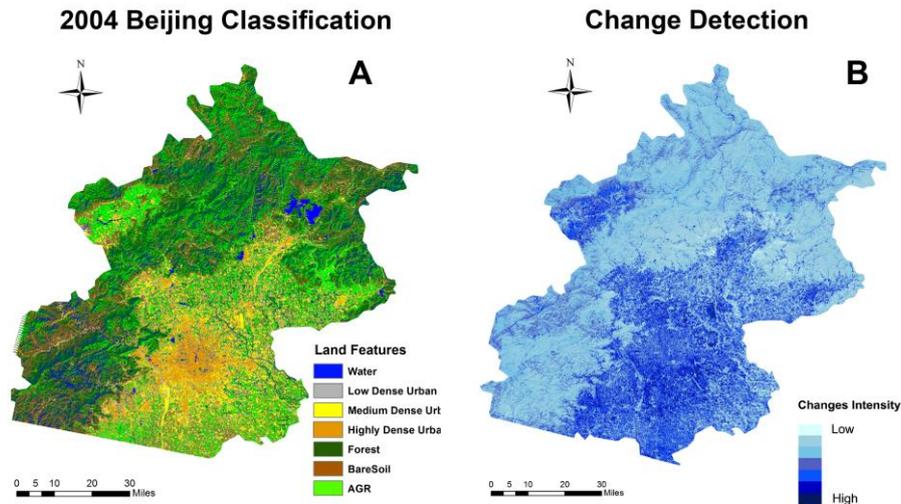
Figure 4. The 2004 class map and the 2004–2014 urban area changes intensity in Beijing.

Table 3. Accuracy assessment of image classification in 2004.

Class	Producer Accuracy	Omission Error	User's Accuracy	Commission Error	Conditional Kappa
Water	80%	20%	80%	20%	0.78
Forest	74%	26%	87%	13%	0.81
Bare soil	86%	16%	73%	27%	0.68
Agricultural land	76%	24%	81%	19%	0.76
Low urban	80%	20%	67%	33%	0.65
Medium urban	71%	29%	65%	35%	0.61
High urban	88%	12%	79%	21%	0.77
Overall Accuracy	78.16%				
Overall Kappa	0.73				

Table 4. Accuracy assessment of image classification in 2014.

Class	Producer Accuracy	Omission Error	User's Accuracy	Commission Error	Conditional Kappa
Water	94%	6%	88%	12%	0.87
Forest	82%	18%	79%	21%	0.68
Bare soil	100%	0%	88%	12%	0.86
Agricultural land	60%	40%	78%	22%	0.73
Low urban	67%	33%	80%	20%	0.79
Medium urban	67%	33%	70%	30%	0.67
High urban	95%	5%	78%	22%	0.75
Overall Accuracy	80.09%				
Overall Kappa	0.75				

3.2. Urban and environment indices

A total of 790 watershed units were delineated in the study area. Urban index and environment index for each unit were calculated in both years. The example maps of the two indices are demonstrated through Figure 5. As expected, the indices corresponded with and reflected the urban land use (I_{URB}) and natural land cover (I_{ENV}) patterns of Beijing. The high urban index values were observed in the central and southeast area (Figure 5A), while high environment index values were observed in the north and southwest (Figure 5B). Urban index values ranged from 0.0008 to 46.15 (between 2004 and 2014). Environment index values ranged from 0.08 to 50. Changes of urban areas were observed through increasing index values, from the highest level to the lowest, specifically 0.0008 to 0.02, and 41.03 to 46.15, respectively (Figure 5C). Natural land cover changes could be observed from the environment index, in which the lowest level dropped from 0.25 to 0.08 between 2004 and 2014 (Figure 5D).
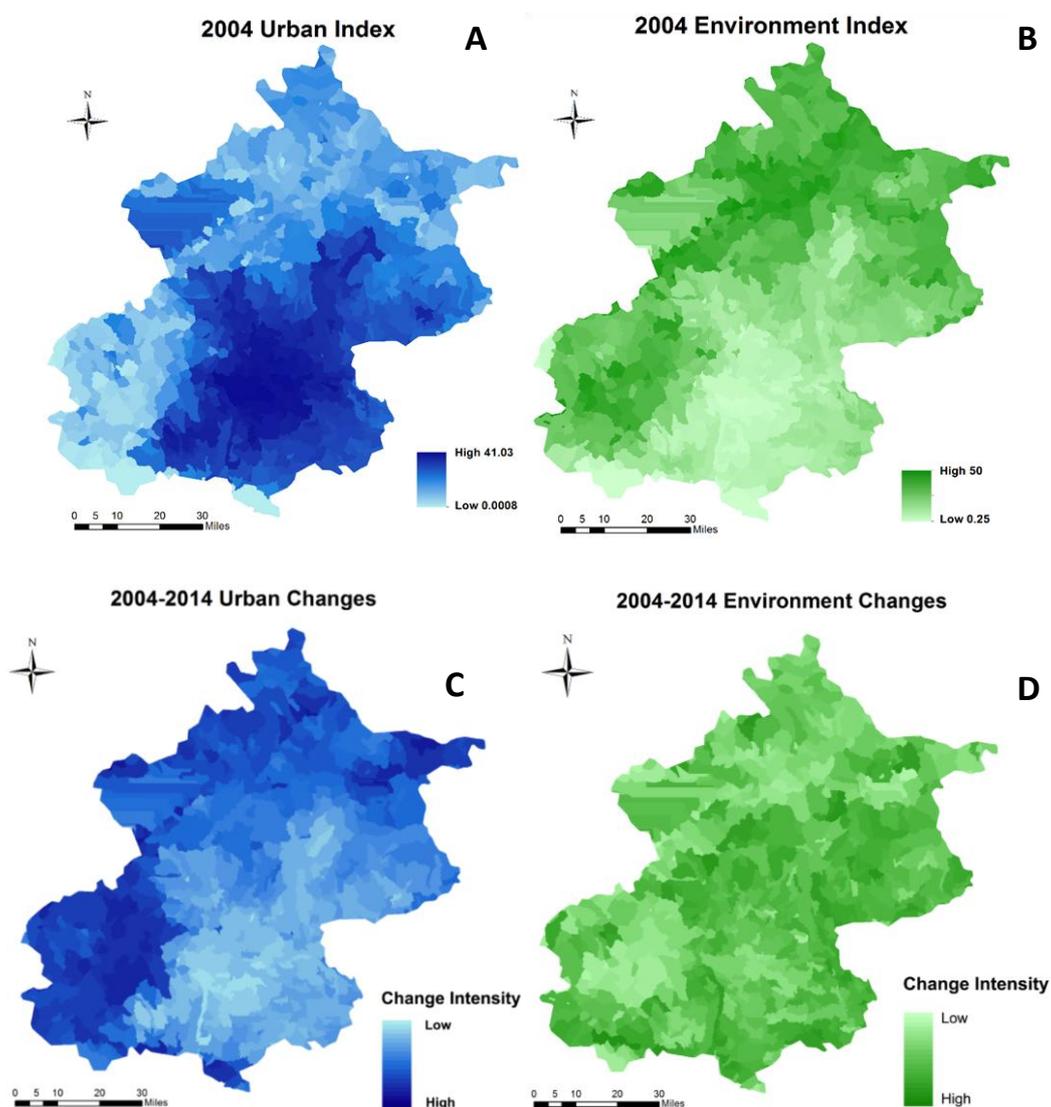

Figure 5. The two indices and their changes between 2004 and 2014. A: the 2004 Urban Index (I_{URB}); B: the 2004 Environment Index (I_{ENV}); C: the 2004-2014 Urban Index Changes. D: the 2004-2014 Environment Index Changes.

3.3. Wastewater discharge patterns and urban area change impacts

Logarithmic transformation was applied on data to scale large values to a comparable magnitude while preserving relative ranking. Therefore the influence of the significant disparity of wastewater discharge volume can be eliminated and better regression results achieved. Logarithmic transformation was applied to minimize bias caused by the magnitudes of wastewater discharge volume at different locations. For instance, some locations' wastewater discharge was around 2000 tons per day, while other locations were 200,000 tons or more per day due to different land types and urbanization intensities. Table 5 shows averaged regression and validation results from 5 experiments. Figure 6 shows regression plots of one experiment to indicate relationships between wastewater discharge volume and urban and environment indices. All regressions were significant at $p < 0.0001$ level (Table 5). 53% of the variance ($R^2 = 0.53$) of wastewater discharge was explained through both urban and environment conditions in 2004. The 2014 models showed 29% and 41% of the variance, respectively through urban and environment indices. Regression errors (RMSE) ranged from 0.75 to 1.13 in two years. Urban index and environment index were comprehensive indicators of land cover types that influence human and natural environments. Positive relationships were found between wastewater discharge volume and urban index in both 2004 and 2014. Environment index, however, was negatively associated with wastewater discharge, as demonstrated in Figure 6. Because environment index was calculated from natural land types, such relationships indicated less wastewater discharged in less urbanized areas. Previous studies have proved that vegetation coverage in cities positively contributed to better water quality [36]. Upstream regions of the study area exhibited better water quality than downstream regions, supporting results found in Li et al.'s research [3].
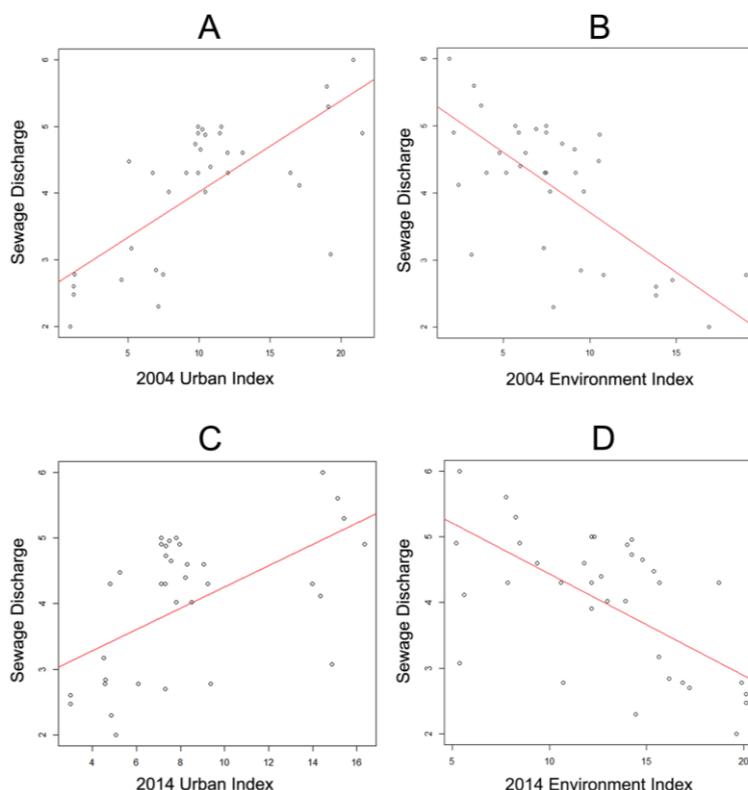

Figure 6. Regression results of the correlation between I_{URB}, I_{ENV} and sewage discharge.

Table 5. Regression Results.

logY	a	b	p-values	Multiple R-Squared	Adjusted R-Squared	RMSE
$I_{URB2004}$	0.14	2.65	<0.0001	0.53	0.51	1.12
$I_{ENV2004}$	−0.18	5.50	<0.0001	0.53	0.52	1.13
$I_{URB2014}$	0.16	2.64	<0.0001	0.29	0.27	0.75
$I_{ENV2014}$	−0.15	5.98	<0.0001	0.41	0.39	0.81

After confirming the regressions were statistically valid, spatial distributions of wastewater discharge were predicted using the IDW interpolation technique (Figure 7). IDW analysis was continued to further visualize display of spatial distribution between wastewater discharge and urbanization over the leveled change detection results. Across the study area, one large hot spot (in light gray) was identified together with several small, scattered hot spots. These high-discharge areas were in agreement with the urban areas in the classification maps in Figure 4A & Figure 5A. The largest hot spot, for example, is located within the urban core of Beijing. Associated with regression results and the IDW map, the expansion of urban areas from 2004 to 2014 influenced the sewage discharge. Thus water pollution crises may increase due to the increasing wastewater discharge of this megacity. According to our wastewater discharged data, the sewage discharge increased 40 % (from 2.55 million tons per day to 4.25 million per day), while the overall treatment rate increased only from 53.9% to 86.1%. There was still 0.34 million tons of untreated wastewater discharged per day. These results matched the records in Wu et al. in 2016, which reported an amount of half a million tons of wastewater discharged every day in Beijing [45].
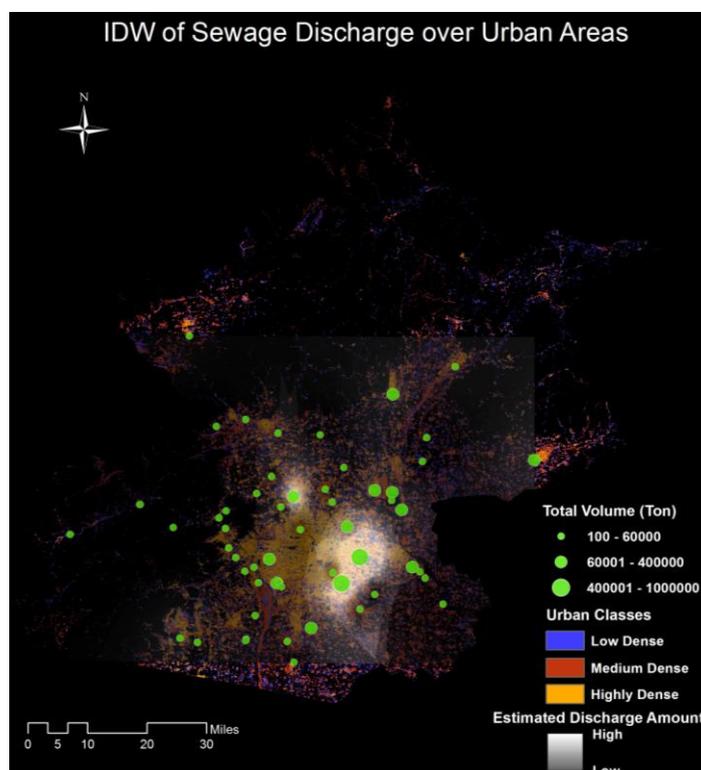

Figure 7. The IDW-interpolated wastewater discharge map overlaid on urban areas.

This study contributes to water quality literature in urban environments by investigating the relationships between wastewater discharge and different land uses associated with human activities from a spatial perspective. Results confirmed that highly urbanized areas contributed to more severe sewage discharge, while increasing vegetation cover could help leverage the pollution problem. The research findings will provide valuable references for governmental agencies in protecting water resources as well as improving the health of citizens. This study may be biased from the imperfect classification accuracy from satellite images, which could affect the indices and therefore the regression results. For example, shadows from clouds and skyscrapers may be misclassified as water, mixed forest or agriculture land, and the delineation between urban classes, especially the low density and medium density urban areas. To minimize said bias, clearest images (cloud free) of the year were selected to eliminate shadows as much as possible. A hybrid classification method was used for this purpose as well. However, additional sources could cause the classification error, such as geometric error, incomplete atmospheric correction etc. Furthermore, none of the classification methods are perfectly accurate. Most importantly, classification is very limited by RS image resolution. In future, other classification methods and higher resolution image data will be considered to improve the accuracies. When more detailed wastewater data sets become available, waterborne disease data could be further examined to investigate the health outcome of urban wastewater discharge. Analyzing water pollution problems is very complicated and challenging, especially considering the interwoven factors from both natural and human influences. Only a few major human factors and physical conditions were focused on. Except the analyzed land types, factors including influences from soil pollution characters, local government control, law force intervention, etc. were not included but may improve the results. This model is simplified to make the first step analysis combining the RS and GIS methods. There are numerous scenarios worth analyzing in the future.

4   Conclusions

This study extracted urban land use changes of Beijing, China from satellite images in 2004 and 2014 and explored its impacts on wastewater discharge of the city. Classification, accuracy assessment, and change detection were performed using remote sensing images and Google Earth. Watershed units were delineated from DEM by identifying junction points between stream segments. Urban Index (I_{URB}) and Environment Index (I_{ENV}) were created according to the weighed influence from different land cover types. At such level, larger volumes of wastewater discharged in a more urbanized area, and also increased with urbanization, while negative association exited between I_{ENV} and wastewater. With the IDW technique, hot spots of high wastewater discharge locations were mapped across the study area, which deserves strong attention from local authorities in protecting the city's water quality and public health. This study indicated that, while point collection of pollution data is often insufficient for sophisticated examination, its integration with satellite imagery improves the spatial representations of pollution sereneness across a large city. Considering time and cost consumption, this is a great stepping stone for a national scale water analysis utilizing satellite images in future.

Acknowledgments

We would like to thank our colleagues, the anonymous reviewers, and the editor for their comments on an earlier version of the manuscript that greatly improved its quality.

Author Contributions

Huixuan Li designed the analytical framework of this study, wrote the majority of the manuscript, and created the maps and tables. Cuizhen Wang provided methodological advice and made major revisions to the manuscript. Yuqin Jiang conducted GIS and statistical analysis and revised the methodology section. Andrew Hug collected and processed data, provided both GIS and RS analysis, and revised and proofread the manuscript. Yingru Li provided advice in initial research design.

Conflicts of interest

The authors declare no conflict of interest.

References

1. World Resource Institute, 2007. Available from: http://www.wri.org/publication/wri-annualreport-2006–2007 (accessed on 21 April 2015).
2. Bao LJ, Maruya KA, Snyder SA, et al. (2012) China's water pollution by persistent organic pollutants. *Environ Pollut* 163: 100-108.
3. Li H, Li Y, Lee MK, et al. (2015) Spatiotemporal analysis of heavy metal water pollution in transitional China. *Sustainability* 7: 9067-9087.
4. Li Y, Li H, Liu Z, et al. (2016) Spatial Assessment of Cancer Incidences and the Risks of Industrial Wastewater Emission in China. *Sustainability* 8: 480.
5. Zhang R, HamerlinckJD, Gloss SP, et al. (1996) Determination of Nonpoint-Source Pollution Using GIS and Numerical Models. *J Environ Qual* 25: 411-418.
6. Ebenstein A (2012) The consequences of industrialization: Envidence from water pollution and digestive cancers in China. *Rev Econ Stat* 94: 186-201.
7. National Academy of Sciences, 2007. Available from: https://www.koshland-science-museum.org/water/html/en/Treatment/Agricultural-and-Industrial-Pollution-in-China.html.
8. Chen S (2015) Beijing drinking water reservoir had lead levels'20 times WHO standard' for at least three years. Available from: http://www.scmp.com/news/china/article/1839337/beijing-drinking-water-reservoir-found-contain-levels-lead-20-times-who.
9. Meng W, Qin Y, Zhang B, et al. (2008) Heavy metal pollution in Tainjin Bohai Bay, China. *J Environ Sci* 7: 814-819.
10. Huang F, Wang X, Lou L, et al. (2010) Spatial Variation and source apportionment of water pollution in Qiantang River (China) using statistical techniques. *Water res* 44: 1562-1572.
11. Jiang J, Wang P, Lung W, et al. (2012) A GIS-based generic real-time risk assessment framework and decision tools for chemical spills in the river basin. *J Hazard Mater* 227: 280-291.
12. Sheng J, Wang X, Gong P, et al. (2012) Heavy metals of the Tibetan top soils. *Environ Sci Pollut Res* 19: 3362-3370.

13. Wang Y, Wang P, Bai Y, et al. (2013) Assessment of surface water quality via multivariate statistical techniques: A case study of the Songhua River Harbin region, China. *J Hydro-environ Res* 7: 30-40.

14. Coskun HG, Gulergun O, Yilmaz L (2006) Monitoring of protected bands of Terkos drinking water reservoir of metropolitan Istanbul near the Black Sea coast using satellite data. *Int J Appl Earth Observ Geoinfor* 8: 49-60.

15. Asadi SS, Vuppala P, Anji Reddy M (2007) Remote Sensing and GIS Techniques for Evalution of Groundwater Quality in Municipal Corporation of Hyderabad (Zone-V), India. *Int J Environ Res Public Health* 4: 45-52.

16. Satapathy DR, Salve PR, Katpatal YB (2009) Spatial distribution of metals in ground/surface waters in the Chandrapur district (Central India) and their plausible sources. *Environ Geol* 56: 1323-1352.

17. Dede OT, Telci IT, Aral MM (2013) The use of water quality index models for the evaluation of surface water quality: a case study for Kirmir Basin, Ankara, Turkey. *Water Qual, Expos Health* 5: 41-56.

18. Wang M, Webber M, Finlayson B, et al. (2008) Rural industries and water pollution in China. *J Environ Manage* 86: 648-659.

19. Ou Y, Wang X (2008) Identification of critical source areas for non-point source pollution in Miyun reservoir watershed near Beijing, China. *Water Sci Technol* 58: 2235-2241.

20. Wang X, Wang Y, Li T, et al. (2002) Characteristics of non-point source pollution in the watershed of Miyun Reservoir, Beijing, China. *Chinese J Geochem* 21: 89-96.

21. Wang X, Ou Y, Dou P, et al. (2009) Relationship between the variation of water quality in rivers and the characteristics of watershed at Miyun, Beijing, China. *Chinese J Geochem* 28: 112-118.

22. Zhu X, Ji H, Chen Y, et al. (2013) Assessment and sources of heavy metals in surface sediments of Miyun Reservoir, Beijing. *Environ monitor assess* 185: 6049-6062.

23. Li Z, Li X, Xu Z (2010) Impacts of water conservancy and soil conservation measures on annual runoff in the Chaohe River Basin during 1961–2005. *J GeographSci* 20: 947-960.

24. Zhao H, Li X, Wang X, et al. (2010) Grain size distribution of road-deposited sediment and its contribution to heavy metal pollution in urban runoff in Beijing, China. *J hazard mater* 183: 203-210.

25. Foster JA, McDonald AT (2000) Assessing pollution risks to water supply intakes using geographical information systems (GIS). *Environ Model Software* 15: 225-234.

26. Yang T, Liu J (2012) Health Risk Assessment and Spatial Distribution Characteristic on Heavy Metals Pollution of Haihe River Basin. *J Environ Analy Toxicol* 2: 152.

27. Soil and Water Assessment Tool, SWAT. Available from: http://swat.tamu.edu/

28. Zhang J, Mauzerall D, Zhu T, et al. (2010) Environmental health in China: progress towards clean air and safe water. *Lanset* 375: 1110-1119.

29. Liu Q, Xie W, Xia J (2013) Using Semivariogram and Moran's I Techniques to Evaluate Spatial Distribution of Soil Micronutrients. *Commun Soil Sci Plan* 44: 1182-1192.

30. Lee CC, Chiu YB, Sun CH (2010) The environmental Kuznets curve hypothesis for water pollution: Do regions matter? *Energy Policy* 38: 12-23.

31. Liu J, Zhang X, Tran H, et al. (2011) Heavy metal contamination and risk assessment in water, paddy soil, and rice around an electroplating plant. *Environ Sci Pollut Res* 18: 1623-1632.

32. Zhou F, Guo H, Liu Y, et al. (2007) Identification and spatial patterns of coastal water pollution sources based on GIS and chemometric approach. *J Environ Sci* 19: 805-810.

33. Murphy RR, Curriero FC, Ball WP (2009) Comparison of spatial interpolation methods for water quality evaluation in the Chesapeake Bay. *J Envirol Engineer* 136: 160-171.

34. Schwarzenbach RP, Egli T, Hofstetter TB, et al. (2010) Global water pollution and human health. *Annu Rev Environ Resour* 35: 109-136.

35. Alexander RB, Boyer EW, Smith RA, et al. (2007) The role of headwater streams in downstream water quality. *JAWRA J Am Water Resour Assoc* 43: 41-59.

36. Liu Z, Li Y, Li Z (2009) Surface water quality and land use in Wisconsin, USA- a GIS approach. *J Integr Environ Sci* 6: 69-89.

37. Huang W, Liu H, Luan Q, et al. (2008) Detection and prediction of land use change in Beijing based on remote sensing and GIS. *The Intern Archi of the Photogram, Rem Sens and Spat Inform Sci* 37: 75-82.

38. China. ORG. CN. China's Political system, 2013. Available from: http://www.china.org.cn/english/Political/28842.htm

39. Li S (2013) All Five of Beijing's Major Water Systems Seriously Polluted. Available from: http://www.theepochtimes.com/n3/251030-all-five-of-beijings-major-water-systems-seriously-polluted/.

40. United States Geological Survey, 2008. Available online: http://earthexplorer.usgs.gov/

41. Yearbook CS, China Statistical Yearbook; Beijing, China, 2015.

42. The Central People's Government of the People's Republic of China, *The People's Republic Of China Yearbook,* 2005. Available from: http://www.gov.cn/test/2005-07/27/content_17403.htm.

43. Wang C, Bentivegna DJ, Smeda RJ, et al. (2010) Comparing classification approaches for mapping cut-leaved teasel in highway environments. *Photogramm Eng Rem S* 76: 567-575.

44. Rozenstein O, Karnieli A (2011) Comparison of methods for land-use classification incorporating remote sensing and GIS inputs. *Appl Geogr* 31: 533-544.

45. Sina News, 2016. Available from: http://news.sina.com.cn/green/roll/2016-08-24/doc-ifxvcsrn9082407.shtml