Journal of Urban Management xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Urban Management

journal homepage: www.elsevier.com/locate/jum

Classification of cities in Bangladesh based on remote sensing derived spatial characteristics

Md. Shahinoor Rahman[a,b], Hossain Mohiuddin[c,d], Abdulla-Al Kafy[d], Pintu Kumar Sheel[e], Liping Di[a,*]

[a] Center for Spatial Information Science and Systems, George Mason University, USA
[b] BUET-Japan Institute of Disaster Prevention and Urban Safety, Bangladesh University of Engineering and Technology, Bangladesh
[c] School of Urban and Regional Planning, The University of Iowa, USA
[d] Department of Urban and Regional Planning, Rajshahi University of Engineering and Technology, Bangladesh
[e] Urban and Rural Planning Discipline, Khulna University, Bangladesh

ARTICLE INFO

Keywords:
City classification
Urbanization
Bangladesh
Urban growth
Urban extent
Remote sensing
Clustering

ABSTRACT

In Bangladesh, cities are conventionally classified based on population size and revenue collection. This conventional city classification system neglects the spatial characteristics inherit in cities. Providing a more comprehensive city classification system is essential for the country's future budget allocation and infrastructure development. Five spatial features: city size (area), urban form (AWMPFD), the ratio of built-up and non-built-up areas, urban growth rate, and total night lights intensity for 331 cities of Bangladesh are derived from remote sensing data. This study classifies these cities into six classes using a hierarchical clustering algorithm based on five selected spatial characteristics. The six categories are named for their levels of spatial development, with Cluster 1 being the highest level and Cluster 6 being the lowest level. The share of employment in the primary sector (agriculture) gradually rises from Cluster 1 to Cluster 6. In contrast, the employment share of the service sector follows a reverse trend from Cluster 2 to Cluster 6. Both per capita income and expenditure is higher for the large cities of Cluster 2 than for the metropolitans of Cluster 1. Comparisons across the six classes with non-spatial attributes validate the classification system. Findings also reveal that remote sensing derived spatial information can explain non-spatial characteristics of cities. Therefore, remote sensing derived spatial attributes of cities can be used for city classification where census data are scarce.

1. Introduction

Cities are the engines of growth. The journey towards growth accelerates urbanization. The earth already crossed a momentous threshold: more than half of the world population is now living in urban areas, and projected to be 60% (nearly 5 billion) by 2030 (Burrows, 2012). Though a predominantly rural country, Bangladesh, is transforming through urbanization at a remarkable pace compared to other South Asian countries (Muzzini & Aparicio, 2013; Rahman, Islam, & Ahmed, 2012). Bangladesh is unique for its combination of large population, high density, and low levels of development. In the past two decades, the urban population of Bangladesh grew from 28.61 million to nearly 42 million (BBS, 2014). By 2020, approximately 36.5% of the country is expected to be urbanized (BBS, 2014). The country's rapid transformation, however, has resulted in uncontrolled growth, inadequate housing, and

* Corresponding author.
E-mail addresses: mrahma25@masonlive.gmu.edu (Md. S. Rahman), hossain-mohiuddin@uiowa.edu (H. Mohiuddin), ldi@gmu.edu (L. Di).

https://doi.org/10.1016/j.jum.2018.12.001
Received 26 June 2018; Received in revised form 30 November 2018; Accepted 8 December 2018

an acute shortage of infrastructure, water, and sewerage facilities (Khan, 2008; Roberts & Kanaley, 2006).

In 1991, the Bangladesh Census broadened the definition of an urban area to include any developed areas around an identified central place with amenities such as paved roads, electricity, gas, water supply, sewerage, and sanitation; that is densely populated; where most of the population is employed in nonagricultural sectors; and where a sense of community is well developed (Chatterji & Yang, 2016). For the past two decades, the number of urban centers have been continuously increasing, reaching 570 in 2011 (BBS, 2014). The population size in each of these urban centers have been rapidly increasing as well (BBS, 2014). Contemporary studies on urbanization in Bangladesh have primarily focused on population dynamics, housing, sanitation, and health issues given the country's pace of urbanization (Khan & Islam, 2010; Planning Commission, 2005; Sen & Ali, 2009). However, the spatial and temporal patterns of urbanization are not well studied in a national context.

Traditional development planning in Bangladesh takes a sectoral form rather than a spatial approach (Planning Commission, 2005, 2011). As such, a report from the Planning Commission revealed considerable spatial disparities in most sectors of the country (Planning Commission, 2011). The cities lagging behind in urban development have fewer infrastructure facilities, are limited in electricity and gas provisions, and receive smaller development allocations. Employment in these cities is more dependent on agriculture than industry (Planning Commission, 2008). There has been a recent trend indicating a reduction in income inequality between the capital city of Dhaka and the other cities in eastern Bangladesh; nevertheless, socio-economic indicators remain disparate across cities, especially for the cities in the western region (Planning Commission, 2008). The division of the country based on these regional inequalities may induce social conflict which may then hinder sustainable development (World Bank, 2007). Balancing the distribution of public expenditure can play a significant role in reducing the inequalities among cities (Faguet & Shami, 2008). For instance, the driving force behind the recent progress in health and education is the rapid rise in government spending on infrastructure development. However, most of the administrative decisions, capital investments, and technological innovations are concentrated in a few cities, which contributes to the imbalance in spatial growth and development (Mahmoud, Wadood, & Ahmed, 2008). Thus, vast spatial inequalities in the rates of urbanization, economic growth, and development persist among cities in Bangladesh (Rahman, 2004; Rahman, 2012).

City classification is important for understanding functionality, resource allocation, employment, and development (Harris, 1943; Heikkila & Xu, 2014; Sobolevsky et al., 2014). The concept and practice of city classification based on spatial development level are new in Bangladesh. Historically, the Bangladesh Bureau of Statistics (BBS) classified urban centers into six categories based on population size: urban areas, big urban agglomerations, medium towns, and small centers were termed "city corporations", "municipalities" or *Paurashavas*, and "growth centers", respectively (LGED, n.d.). There are now a total of 12 city corporations and 327 municipalities in Bangladesh (LGED, n.d.). The Local Government Engineering Department (LGED) implements another classification system, where municipalities are placed into either A, B or C categories based on the minimum annual revenue collected over the past three years. (LGED, n.d.). These two methods of urban classification rely only on a single, non-spatial metric rather than a set of comprehensive spatial indicators. Thus, the current classification systems is unable to capture the variation in spatial development levels across cities. Additionally, the LGED allocates resources based on the class (status) of *Paurashavas* (LGED, n.d.). There is a tendency among *Paurashavas* to report augmented revenue potentials to uplift their status from lower categories to higher categories. Moreover, many municipal authorities try to change the status of their *Paurashava* by using their political influence for getting additional fund allocation from the central government. The proposed city classification system may address various inefficiencies of the conventional urban classifications in Bangladesh. The proposed classification system utilizes spatial characteristics of cities. The new classification may also be helpful towards urban decentralization and reducing spatial inequalities by alleviating urban pressures on the larger cities and promoting development of small and medium-sized cities in Bangladesh (Rahman & Islam, 2013).

Since proper city classification is useful for resource allocation, decentralization, and balanced development, this study aims to classify 331 cities in Bangladesh with cluster analysis according to spatial characteristics of cities. These five variables are directly related to city's expansion, economic activities, population growth, and overall development. For instance, the city size (area) serves as a baseline indicator for a city's spatial extent. The built-up area reveals the amount of infrastructure development that has been occurring within a jurisdiction. The ratio of built-up area to the total area captures the intensity of infrastructure development that has been taking place in relation to the total area of city. For example, a city may have a large area but may have only small built-up area compared to the city's size. In contrast, a city may have a small size, but most of the land is built-up. Urban growth captures the temporal growth potentialities of a city since some cities grow faster compared to others. City night light intensities provide a useful proxy variable in measuring the economic activity of a space (Doll, 2010; Mellander, Lobo, Stolarick, & Matheson, 2015). Spaceborne night light intensity data become popular in research and applications. Many studies such as urban extent, a proxy for development as well as socio-economic indicators, population studies, and health studies rely on night light intensity, especially where field level data are scarce (Elvidge, Baugh, Anderson, Sutton, & Ghosh, 2012; Lo, 2002; Noor, Alegana, Gething, Tatem, & Snow, 2008). Noor et al. (2008) found a higher association of night light with poverty indices over the 37 African countries. They concluded that the night light data is a robust and inexpensive alternative to explore the regional inequity. Ma, Zhou, Pei, Haynie, and Fan (2012) has investigated the association between night light data and four variables namely, population, gross domestic product (GDP), built-up area, and electric power consumption, and concluded that night light data could accurately use as an explanatory variable for these selected urban dynamics variables at the local level. Human population count and population density can also be estimated through night light intensity (Prosperie & Eyton, 2000; Sutton, Roberts, Elvidge, & Meij, 1997). The night light data can also be used for measuring the human wellbeing and development (Elvidge et al., 2012; Ghosh, Anderson, Elvidge, & Sutton, 2013). Dugoua, Kennedy, and Urpelainen (2018) examined the rural electrification status at the village level in India using night light intensity. Likewise, Bhandari and Roychowdhury (2011) found that GDP at the district level in India can be significantly explained by the corresponding night light data of that area. Zhou, Hubacek, and Roberts (2015a, 2015b) developed a 1 km spatial resolution urban area map based on the night light data and found that use of the night light data provides a reliable and robust estimate of the global urban land area. Henderson, Yeh, Gong, Elvidge, and Baugh (2003) mapped the development level of three cities: San Francisco, Beijing, and Lhasa. Hence, Night light intensity derived from satellite remote sensing is increasingly used to capture the spatial extent and magnitude of development for regions where data on economic activities are scarce (Huang, Yang, Gao, Yang, & Zhao, 2014). This proxy variable can overcome the absence and inconsistencies of government census data on economic development (Mellander et al., 2015). Thus, night light intensity can be the useful indicator to capture nonspatial characteristics of cities in Bangladesh. Another spatial characteristic of cities, urban form, influences daily life and is an important factor for both quality of life and environmental factor (Schwarz, 2010). Urban form also describes compactness and sprawl characteristics of cities (Huang, Lu, & Sellers, 2007; Rahman, 2012). There are many landscape functionality-based indicators for urban form. Area weighted mean patch fractal dimension (AWMPFD) is one of the important urban form indicators in spatial domain, which has widely been used in urban landscape studies(Vanderhaegen & Canters, 2017). AWMPFD is a class level landscape metric, which indicates the raggedness of the urban boundary. AWMPFD is a complexity metric, which approaches 1 for simple urban forms and 2 for complex urban forms (Herold, Scepan, & Clarke, 2002; Huang et al., 2007). Huang et al. (2007) compares urban form of 77 metropolitans across the world and they found higher AWMPFD for cities in developed countries compared to developing countries. Herold et al. (2002) examines structure and changes in urban land uses in Santa Barbara, California through landscape functionality. Their result shows higher AWMPFD for high density residential area compare to low density residential area. Hence, AWMPFD can be utilized to capture the characteristics of cities (Goerlich Gisbert, Cantarino Martí, & Gielen, 2017; Schwarz, 2010).

The proposed city classification system based on spatial characteristics of cities may explain city's non-spatial characteristics, such as, development income and expenditure, revenue income and expenditure of a city, total population, population growth, literacy rate, and sector wise employment. Although remote sensing data are widely used in many urban studies such as urban growth monitoring, development pattern, urban environment, landscape functionalities, this study may be the first approach to use remote sensing derived spatial features for city classification at national scale. This new classification system may be applicable for implementation by other developing countries, where census data on urban spatial attributes are scarce and unreliable.

2. Materials and methods

The analytical framework of this study is divided into three major sections. The first section reviews data and their properties. The second part describes data preparation process and remote sensing tools for extracting information on the selected spatial variables. The final part describes the clustering process for classifying the cities in Bangladesh. The proposed classification system is then validated through the comparison among clusters using non-spatial attributes of cities: development income and expenditure of a city, revenue income and expenditure of a city, population, population growth, literacy rate, and sector wise employment share.

2.1. Data

The majority of the data used in this study is primarily based on remote sensing data as well as secondary data from the different government and non-government organizations of Bangladesh. City boundaries are collected from the LGED, city authorities, and consulting firms, since, there is not a single source which can provide boundaries for all the cities. Organizations in Bangladesh use different geographic projection systems. Therefore, all city boundaries are projected to World Geographic System 1984, a standard geographic projection system. Non-spatial data—population, literacy rate, sector-wise employment share – is collected from the Bangladesh Bureau of Statistics (BBS) for the years 2001 and 2011. The rationale for selecting these two years is the nationwide census was conducted in these two years. Budget, revenue income, and expenditure, as well as development income and expenditure of each municipality, are collected from LGED, National Board of Revenue, and city authorities for the financial year 2015–2016.

Satellite remote sensing imageries have proved useful sources for accurate multi-temporal landcover information (Ahmed, Kamruzzaman, Zhu, Rahman, & Choi, 2013; Jeff and Star n.d.). The Built-up and non-built-up data of the selected cities are extracted from Landsat image analysis. Landsat 4–5 thematic mapper (TM) images are downloaded from the United States Geological Survey (USGS) earth explorer for the years 2001 and 2011, to match the two-census years of Bangladesh. Images were chosen from the same season (late autumn) of both years to avoid seasonal variation. Images from the late autumn are suitable for built-up area extraction, because, this season is generally cloud free and trees are not in their leaf-off condition, and crop fields are cleared out in a tropical area like Bangladesh (Rahman, Ahmed, & Di, 2017). The annual night light data for the year 2011 are downloaded from the National Center for Environmental Information portal of the National Oceanic and Atmospheric Administration (NOAA). The annual night light data product is the composite cloud-free data processed at 1 km spatial resolution from visible and infrared imagery of the Operational Linescan System of the Defense Meteorological Satellite Program (DMSP-OLS) (Baugh, Elvidge, Ghosh, & Ziskin, 2010; NOAA, n.d.). The night light data product includes intensity from any artificial light sources in urban and rural areas but excludes noise, for instance, moonlit, fire, and lightning. Fig. 1 shows the average night light intensity of settlements in Bangladesh: urban areas appear white because of a higher night light intensity. The digital number (DN) values represent the average light intensity on a 1 km by 1 km grid. The DN value ranges from 0 to 63 on a relative scale, where zero indicates a completely dark area and any other value represents relative light intensity (Ma et al., 2017).

The relative intensity of light can represent variation in economic activities coupled with variation of population density within and across cities and regions. The level of development of a city can be captured by comparisons of total night light intensity with other cities. This variable may help to explain both vertical and horizontal development patterns of Bangladesh cities. The night light intensity shows a concentration of economic activities in different parts of Bangladesh. Fig. 1 shows a concentration of bright areas in

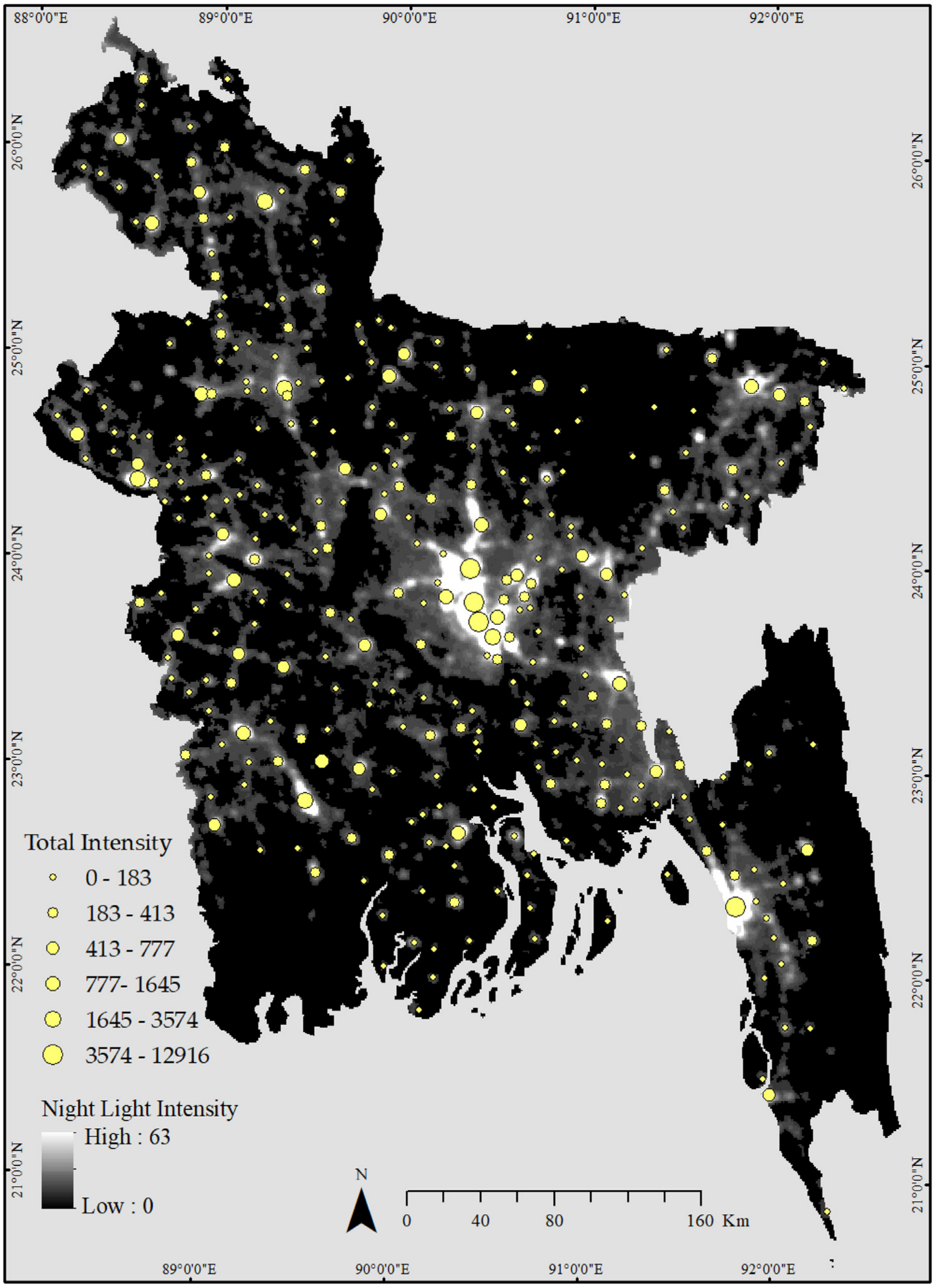

Fig. 1. Composite night light intensities of cities in Bangladesh in 2001.

the center and in the lower south-eastern part of the country, pointing to large urban agglomerations in the areas: the capital Dhaka, the industrial hub Gazipur, and the port city Chittagong. Visual analysis of night light intensity confirms that this method can be used for detecting areas of high levels of economic activities and infrastructure concentration in Bangladesh. As one of the most densely populated countries in the world, Bangladesh, has seen most of its land for horizontal infrastructure development limited in the recent years. Many parts of the country have moved towards vertical development by building high rise buildings for residential and commercial purposes. However, it is difficult to have data on vertical development for each city in Bangladesh because most city authorities do not produce data on vertical development. As vertical development becomes a common trend in cities of Bangladesh, it is necessary to devise some tools to extract information on vertical development at the city level. Night light intensity can represent some features of vertical development since light intensity is higher in dense areas compared to sparse areas. Therefore, night light intensity can assist the city classification system to describe the economic activity and dense vertical development level across the cities of Bangladesh.

2.2. Preparation of feature vectors

Six feature variables –total area, built-up area, buildup ratio, urban growth, AWMPFD and total night light intensity– of cities are used as feature vectors in the clustering algorithm. The total area is directly calculated from the city boundaries. Built-up areas are extracted through the supervised land cover classification of Landsat imageries. Images are classified into four general land covers: built-up area, water, vegetation and bare soil. The three types of land covers except the built-up area are then grouped together to calculate the total non-built-up area of each city. The ratio between built-up and non-built-up area is taken as one variable of the five-dimensional vector space. The difference between the built-up area of 2001 and 2011 was used an indicator for urban growth. Total night light intensity was calculated as the sum of night light intensity within a city. AWMPFD is calculated for the built-up area of each city in 2011. The fractal dimension as a measure of the fragmentation for every patch of built-up area was applied as a derived metric called area weighted mean patch fractal dimension (AWMPFD) which can be expressed by the following equation (McGarigal & Marks, 1995):

$$AWMPFD = \sum_{j=1}^{n} \left[\left(\frac{2ln(0.25p_{ij})}{lna_{ij}} \right) \left(\frac{a_{ij}}{\sum_{j=1}^{n} a_{ij}} \right) \right]$$

Where, a_{ij} and p_{ij} are the area and perimeter of patch j in class i, and n is the total number of patches

Supervised image classification is performed on the Landsat image stack for each city separately to produce the highest possible result. Landsat Thematic Mapper (TM) visible green, visible red, near infrared (NIR), and shortwave infrared (SWIR) bands are used in the landcover classification process. The visible blue band is excluded because of the higher atmospheric scattering in this band. Fig. 2 shows examples of landcover classes and night light intensities of three cities. Although, urban growth rates are not same, the amount of total built-up area increased from 2001 (panel a) to 2011 (panel b) for all the three examples. There is a significant difference in total night light intensity among cities (Fig. 2: panel c) due to the variation in the size of the area as well as in the magnitude of intensity. Therefore, both landcover information and night light intensities are calculated for all cities for the purpose of clustering.

Since three variables –built-up area, built-up ratio, and growth entirely rely on the quality of the landcover classification, it is essential to perform an accuracy assessment of the image classification process. More than 3950 samples are sampled using equalized stratified random sampling method. Ground truth of the landcover type at a sampling point was collected from visual interpretation of high-resolution Google Earth images. The overall accuracy of the 2001 and 2011 classification are 92.67% and 91.68% respectively (Tables 1 and 2). Since this study is only concerned about the built-up area, the classification error among the non-built-up class can be ignored. Therefore, classification accuracy is increased to 96.98% and 95.93% for 2001 and 2011 respectively, after merging water, vegetation, and bare soil into the non-built up class. Another measure of the accuracy of classification is Kappa statistics, which evaluates the classification values compared to valued assigned by chance (Rossiter, 2004; Sim & Wright, 2005). Kappa values are in the range of 0 to 1, -with 0 indicating no agreement between the classified image and the reference image and 1 indicating both the classified and the reference image are identical. The higher Kappa value indicates higher classification accuracy of a classified image. The Kappa Statistics of the classifications are 0.92 and 0.89 for the year 2001 and 2011, respectively when only built-up and non-builtup area are considered (Tables 1 and 2). Therefore, both the overall classification accuracy and kappa value indicate significant classification accuracy of the classified image with ground truth. This higher level of accuracy validates the utility of the extracted information from image classification for the clustering process.

2.3. Clustering process

Hierarchical cluster analysis is performed to identify groups of similar cities based on selected variables. Data mining algorithms for cluster analysis use distance function to determine dissimilarity between observation. Multicollinearity among variables and the scales of variable have a significant impact on the clustering process. Collinearity is the phenomenon among variables when two or more variables measure the same attribute of an object (Dormann et al., 2013; Sambandam, 2003). Therefore, multicollinearity is usually needing to be assessed and taken care of to reduce the redundancy impact of variables in the clustering process. Multicollinearity inflates the variance among variables and hence potentially leads to the wrong identification of relevant predictors in a statistical model. Thus, the degree of multicollinearity among variables can be measured by variance inflation factors (VIF) and tolerance. A small VIF value indicates low correlation among variables, where VIF value 1 indicates no correlation among variables and a VIF value over 10 is a clear signal of multicollinearity among variables. Although Hair, Black, Babin, Anderson, and Tatham (2006), suggests the maximum acceptable level of VIF is 10, many researchers mentioned a VIF greater than 5 indicates high correlation among variables (Kline, 2015; Kock & Lynn, 2012; O'brien, 2007). Therefore, the degree of multicollinearity among initially six selected variables is measured through VIF for this study. The VIF values of city area, built-up area, built-up and non-built-up ratio, urban growth, AWMPFD, and night light intensity are 4.93, 33.54, 5.09, 5.23, 7.68, and 16.03, respectively. The VIF values for six variables are clearly indicating the high correlation between urban built-up area and night light intensity. Thus, variable with highest VIF 33.54, urban built-up area, is dropped from clustering analysis. The VIF values are again calculated for the remaining five variables to avoid multicollinearity in the clustering process. City area, the ratio between built-up and non-built-up area, urban growth, AWMPFD, and night light intensity corresponds to VIF values 4.47, 4.60, 3.78, 4.65, and 5.07, respectively.

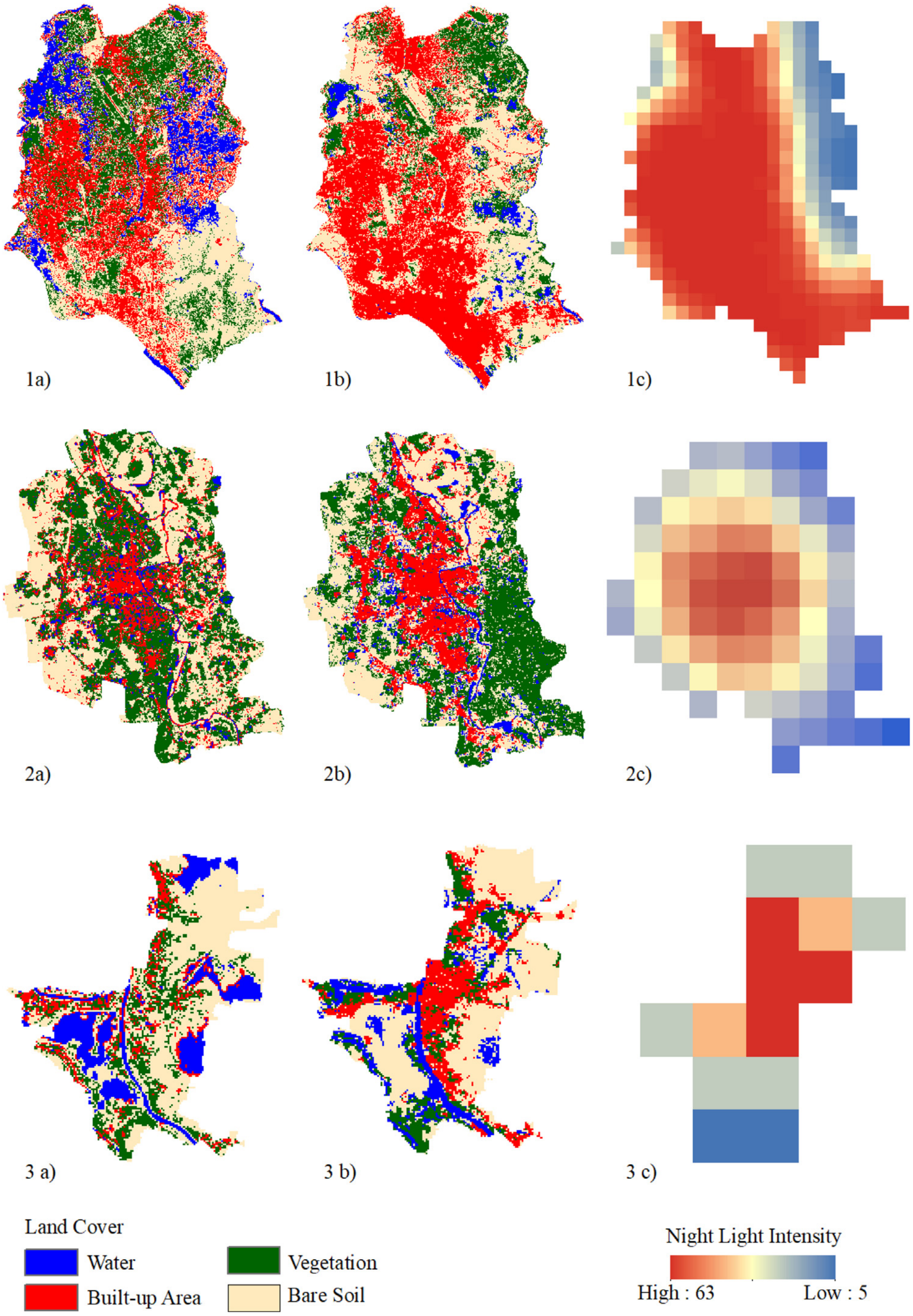

Fig. 2. Three examples of land cover classifications and night light intensity from large, medium, and big cities. Panel a, panel b, and panel c represent landcover 2001, landcover 2011 and night light intensity 2011 respectively for 1) Dhaka, 2) Bogra, and 3) Singra.

Although, night light intensity has VIF value little over 5, this study did not drop night light intensity considering the importance of this variable and the VIF value is very close to 5. Another important aspect in clustering process is the scales of variables. All variables need to be in the same measurement scale since the clustering algorithm depends on the distance function. Normalization of the variables is required before performing the hierarchical clustering process to ensure that the variables have the same weight (Rahman, Yang, & Di, 2018). The five vectors are normalized to the scale between 0 to 1 by max-min normalization.

$$\text{Normalized variable} = (X_{ij} - X_{i\_min})/(X_{i\_max} - X_{i\_min})$$

Table 1
Accuracy assessment for the image classification of 2001.

Ground Truth							
Classified	Class	Water	Urban	Vegetation	Bare Soil	Total	User Accuracy
	Water	917	3	53	2	975	94.05
	Urban	3	964	1	32	1000	96.40
	Vegetation	44	4	920	35	1003	91.72
	Bare Soil	4	77	33	880	994	88.53
	Total	968	1048	1007	949	3972	
	Producer Accuracy	94.73	91.98	91.36	92.73	Overall Accuracy 92.67	Kappa Statistics 0.90

Table 2
Accuracy assessment for the image classification of 2011.

Ground Truth							
Classified	Class	Water	Urban	Vegetation	Bare Soil	Total	User Accuracy
	Water	929	4	15	41	989	93.93
	Urban	11	848	32	98	989	85.74
	Vegetation	5	6	921	57	989	93.12
	Bare Soil	9	10	41	929	989	93.93
	Total	954	868	1009	1125	3956	
	Producer Accuracy	97.38	97.70	91.28	82.58	Overall Accuracy 91.68	Kappa Statistics 0.90

Where, X_{ij} is the value of i^{th} variable of city j; X_{i_min} and X_{i_max} are the minimum and the maximum values of i^{th} variable.

A hierarchical cluster analysis procedure applies an agglomerative hierarchical clustering using distance measures and numerous methods. The agglomerative clustering process starts by considering each of the 331 cities as individual clusters and then, through iteration, joins the two closest clusters until there is one cluster that represents all the previous observations (Gross, 1975; Kaufman & Rousseeuw, 2009). The clustering outcome is highly depending on distance measures and methods, and there are no clear guidelines for the selection of measures, and methods. Therefore, this study uses four different methods –average link, centroid, median and ward—separately for different sets of clustering outcomes. Maximum distance and minimum distance methods are avoided because of the impact of extreme values on these two methods. Total five distance measures including Euclidean distance, Manhattan distance, Squired Euclidean distance, Mahalanobis, and Chebyshev are utilized in this study. Since centroid, median, and ward methods only allow Euclidian distance measure, total eight combination of methods and distance measures are applied for eight sets of clustering outcomes. The discriminant analysis then be performed on clustering outcomes. The degree of success in the clustering process can be determined through the discriminant analysis. (Frenkel, 2004). The discriminant scores are 0.95, 0.94, 0.95, 0.93, 0.95, 0.91, 0.89, 0.81 for average link with Euclidian, Manhattan, squired Euclidian, Mahalanobis, and Chebyshev, centroid-Euclidean, median-Euclidian, and ward-Euclidian combination, respectively. Three combinations –average link method with Euclidian distance, squired Euclidean distance, and Chebyshev distance –among eight combinations of method and distance measures have highest scores (0.95) in discriminant analysis. These three combinations have exactly same clustering output. The cophenetic correlation is another measure which determines the strength of a clustering structure with a dendrogram. The value of the cophenetic correlation metric reflects how well the dendrogram represents the real clustering structure of the dataset. The cophenetic values are 0.92, 0.90, 0.87, 0.86, 0.90, 0.91, 0.79, 0.75 for average link with Euclidian, Manhattan, squired Euclidian, Mahalanobis, and Chebyshev, centroid-Euclidean, median-Euclidian, and ward-Euclidian combination, respectively. Considering the discriminant score and cophenetic metric the combination average link method and Euclidian distance measure is the clear choice for this study.

2.4. Statistical significance test

It is important to examine the robustness of the similarities and the dissimilarities of the clusters on different metrics through statistical significance tests. The statistical significance of clusters is tested using two non-parametric tests, the Kruskal Wallies test and the pairwise Mann-Whitney U test, both the selected non-spatial variables at 95 percent confidence level. The Kruskal Wallies test reports on the overall significant difference among clusters, while the Man-Whitney U test determines the similarities and the dissimilarities between two clusters (Burt, Barber, & Rigby, 2009). The Kruskal Wallies test is suitable since this study requires investigating the significance among six clusters of cities with different sample sizes. In addition, the Kruskal Wallies test is applicable for both continuous and ordinal variables. The number of observations in each of the cluster is more than five. Thus, the chi-square table was used to determine the test result. As the Kruskal Wallies test is conducted on six clusters, the degree of freedom is five. The critical value at 95 percent confidence level for five degrees of freedom (5 DoF) is 11.07. The limitation of the Kruskal Wallies test is that if any of the samples set is different than others, the result is statistically significant. Therefore, it is not possible to determine the similarities or the dissimilarities between two clusters with the Kruskal Wallies test. On the other hand, the pairwise Mann-Whitney U test determines whether two independent samples were selected from populations having the same distribution. Therefore, to find the similarities and dissimilarities between clusters the pairwise Mann Whitney U test is applied for all the selected variables.

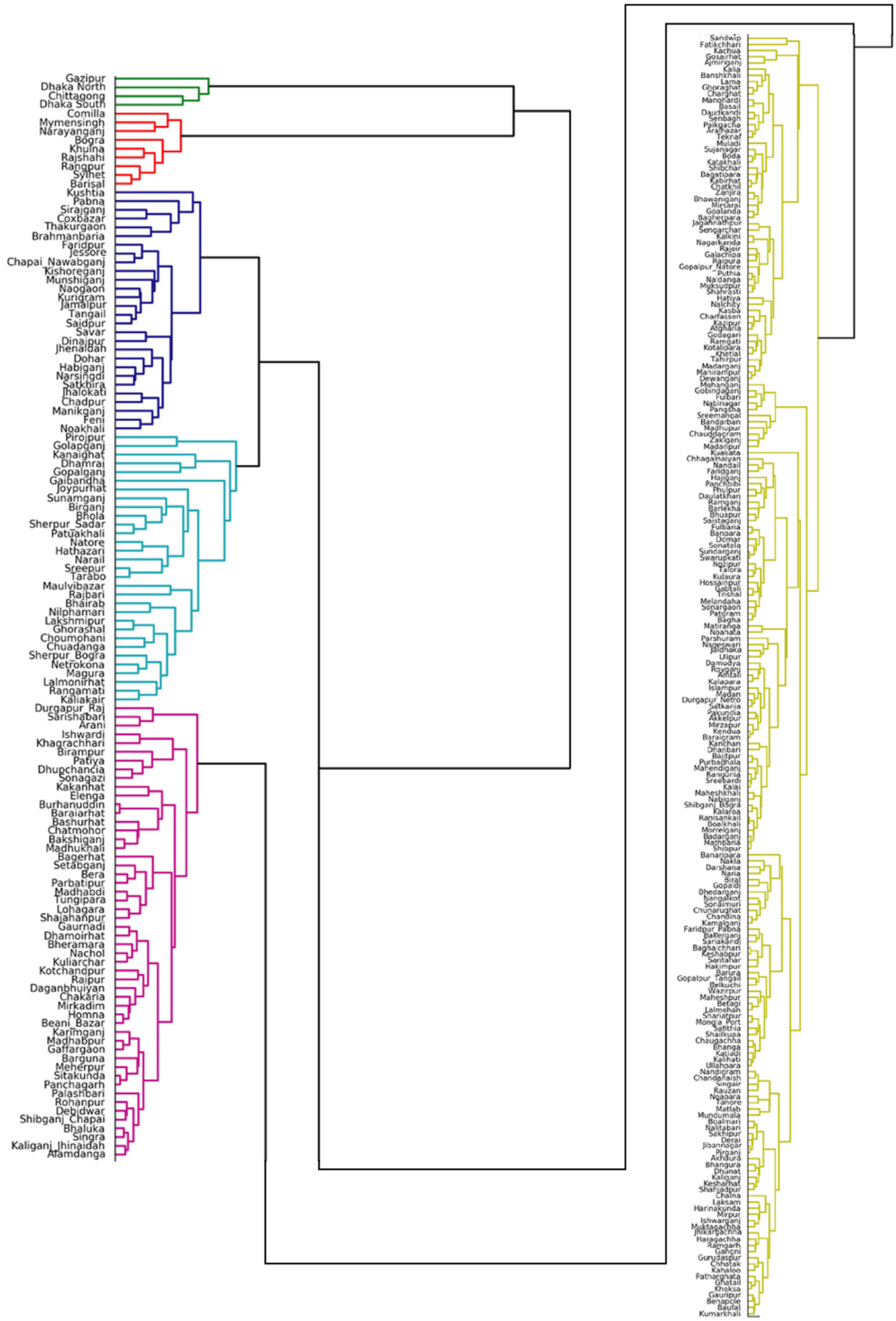

Fig. 3. Dendrogram showing the clustering process and the six clusters for this study. Colors represent different clusters of cities.

Table 3
Composition of different types of cities in the six extracted clusters.

	City Corporation	Class A	Class B	Class C	Total
Cluster 1	4	0	0	0	4
Cluster 2	7	2	0	0	9
Cluster 3	1	27	0	0	28
Cluster 4	0	27	2	2	31
Cluster 5	0	31	13	7	51
Cluster 6	0	67	84	57	208
Total	**12**	**154**	**99**	**66**	**331**

3. Results discussion

3.1. Cluster results and dimensions

The dendrogram, Fig. 3, illustrates the overall clustering of 331 cities in Bangladesh. The last stage of the clustering process is to decide the optimal number of clusters to retain according to the purpose of this study. Although it is difficult to determine the optimal number of clusters, there are several methods of selecting the appropriate number of clusters. The cophenetic correlation is used for this study. If the Cophenetic correlation value exceeds 0.8, then visual inspection of the dendrogram is necessary to determine the number of the clusters (Rohlf & Fisher, 1968; Singh, 2008). The Cophenetic correlation value for this study was 0.92, so visual inspection is conducted to determine the appropriate number of clusters by dividing the dendrogram at the location where average dissimilarity between the clusters jumps considerably between iterations. A six-cluster structure is found after visual inspection of the dendrogram. The average dissimilarity value significantly rises when the clustering process goes from a six-cluster structure to three-cluster structure. Therefore, a six-cluster solution is better suited for this study. These clusters can represent separate classes of cities each having homogenous characteristics. Overall, the cluster analysis process divides 331 cities into meaningful six homogenous classes which are represented by six different clusters. In Cluster 1 to Cluster 6, there are 4, 9, 28, 31, 51, and 208 cities respectively.

Table 3 shows the current status of cities in each cluster. Cluster 1 consists mainly of four large metropolitan cities and Cluster 2 is composed of nine big cities. Most of the moderate sized cities are grouped in Cluster 3 and Cluster 4. Cluster 5 has relatively smaller cities and the cities in Cluster 6 can be denoted as very small cities. Apart from seven city corporations, Cluster 2 also contains two class A municipalities. Moreover, Cluster 3 includes one city corporation and twenty-seven class-A municipalities. The result of Cluster 1 to Cluster 3 clearly indicates that all twelve city corporations are not homogenous in their spatial attributes. One city corporation (Cox-Bazar) is grouped in Cluster 3 with class-A municipalities. From the existing classification system, it is expected that there is a hierarchy (large to small) in spatial attributes from city corporation, Class-A, Class-B and Class-C municipalities. However, the current study finds there is neither such hierarchy nor homogeneity among cities in existing classification system. Therefore, this study again justifies spatial attributes based classification system. A significant number of class-B and Class-C municipalities have similar spatial characteristics with Class A municipalities. In the Cluster 5, a significant number of class B and class C municipalities exist. A large number of class A municipalities are also grouped with small urban areas which indicates considerable disagreement with the existing revenue-based classification system. Consequently, this classification system highlights the fact that the existing classification system does not capture the economic and infrastructure condition of cities and supports the necessity of this newer method of classification. Fig. 4 illustrates the spatial location of cities in different clusters. Cities within a cluster are mostly spatially scattered in different parts of the country, except for these in Cluster 1. Three out of the four cities in Cluster 1 are concentrated around the capital of Bangladesh. In contrast, the cities belonging to the other clusters are scattered throughout the country. None of the other clusters show the same concentration in a particular geographic area as Cluster 1.

The development level of the clusters may have a relationship with spatial attributes of the clusters. Fig. 5 shows a series of three-dimensional scatter diagrams of the six classes created using the selected five spatial variables. Plot 5a illustrates the position of each class of cities concerning urban growth, built up area and total city area. The urban growth is very high for cities in Cluster 1, which indicates a massive level of infrastructure development and conversion of non-built up area into built-up area in the recent decade. Cluster 1 consists of four city corporations (Table 3), Dhaka North City Corporation, Dhaka South City Corporation, Gazipur City Corporation, and Chittagong City Corporation. Cluster 1 can be described by its high built up and non-built up area ratio. Most of Bangladesh's industries and administrative activities are concentrated in the cities of Cluster 1. A very high built up ratio coupled with a very high level of urbanization is found in Cluster 1 compared to the other clusters (Fig. 5c). Plot 5b shows the relationship among night light intensity, urban growth, and the total city area. This diagram confirms a clear separation between Cluster 1 and compared to the other clusters. The value of night light intensity is very high in this area due to the high levels of economic development, administrative activities and vertical development (Fig. 5b). In summary, Cluster 1 shows a significant level of development in all the selected variables. Cluster 1 can be defined as the most developed class of cities in Bangladesh.

Cluster 2 can be characterized by big cities with high built and non-built up ratio and reasonable night light intensity which indicating the development activities are concentrated in these cities. Cities in Cluster 2 also experienced high urban growth in the recent decade. Cluster 2 can be ranked as the second based on the development characteristics. Cluster 3 is characterized by their moderate built up area with a high built up and non-built up ratio (Fig. 5c). Moderate urban growth can also be seen among cities in Cluster 3. The night light intensity is low for the cities in Cluster 3 (Fig. 5b), which indicates a lower level of economic activity in

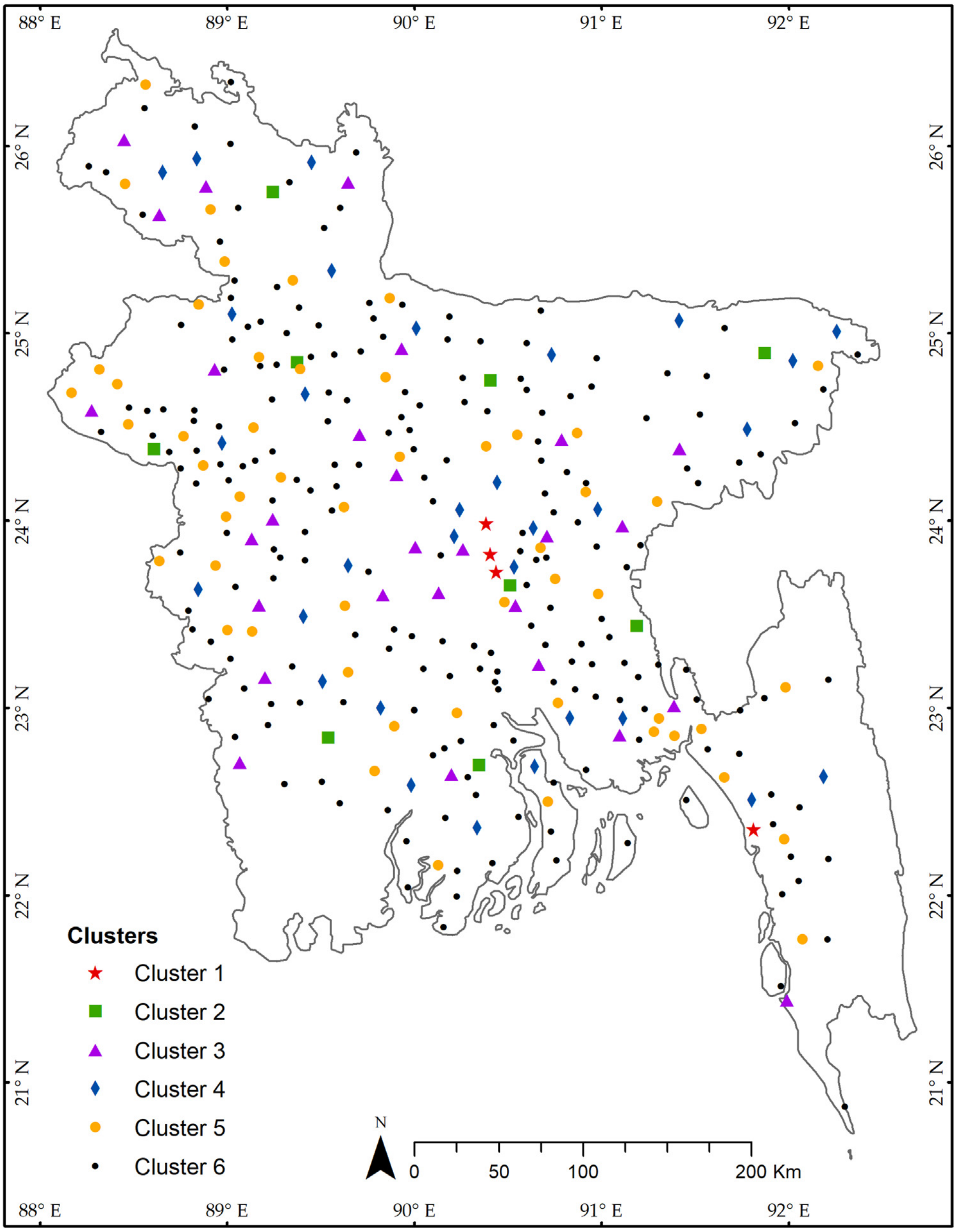

Fig. 4. Distribution of the six classes of cities in Bangladesh.

cities of this class. Thus, Cluster 3 can be denoted as a group of moderate developed cities, considering their spatial attributes and development conditions. Cluster 4 is also composed of moderately sized cities and can be characterized by a lower level of built-up area and built up ratio (Fig. 5c), but also experienced higher urban growth in the last decade (Fig. 5a). Furthermore, the night light intensity indicates a low level of economic activity (Fig. 5b). Cluster 5 consists of small sized cities with a lower level of total city area coupled with a higher built up and non-built up ratio (Fig. 5c). The night light intensity is very low for Cluster 5 (Fig. 5b), indicating minimal levels of economic activity and physical infrastructure in these cities. Cluster 6 includes very small sized cities with very small city areas (Fig. 5d), tiny built-up areas with low built up and non-built up ratios. Most of the cities in Cluster 6 have negligible night light intensity (Fig. 5b) and experienced a very low rates of urban growth in recent past. Thus, cities of this cluster can be termed as the most underdeveloped cities in Bangladesh. This cluster is mainly composed of Category-B and Category-C Municipalities.

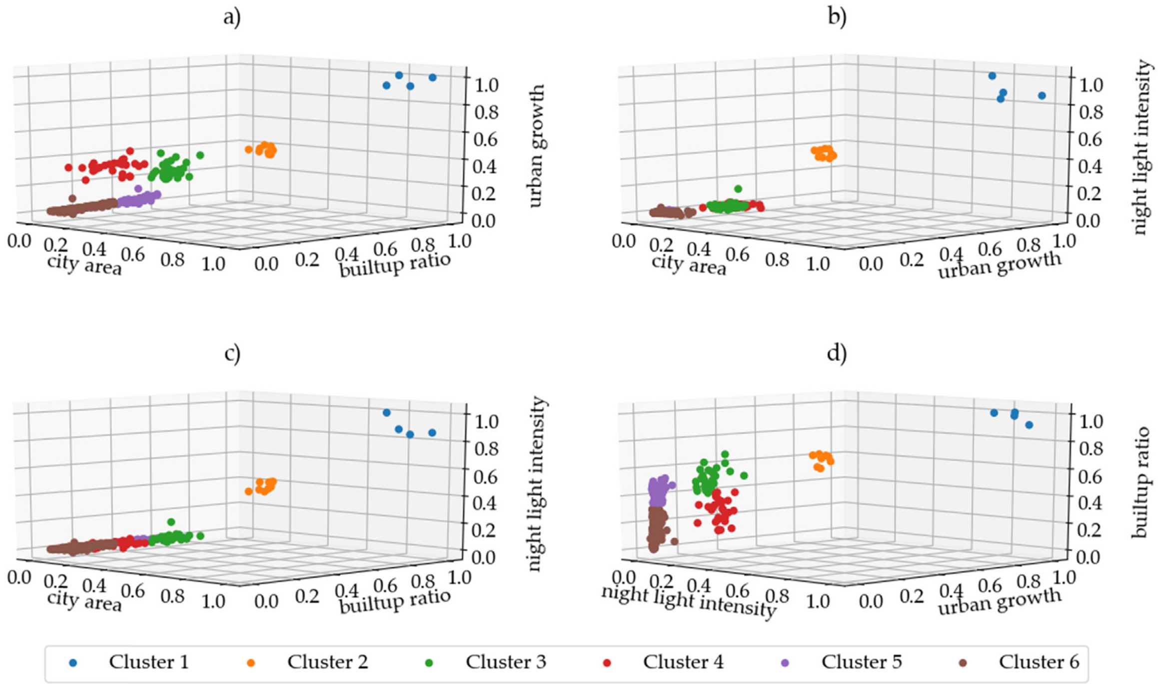

Fig. 5. 3D scatter plots showing the position of the cities of each cluster based on the selected five variables.

3.2. *Analysis of the non- spatial characteristics of the clusters*

The relation of six clusters with non-spatial attributes indicates the relevancy of the proposed classification of cities because of the homogenous non-spatial characteristics of cities within a cluster. This study selects non-spatial variables covering demography, employment, as well as income and expenditure to discuss the classification outcomes through non-spatial attributes. The relationship between clusters and non-spatial attributes explains the usefulness of the proposed city classification system.

3.2.1. *Demography*

The first selected non-spatial variable is the number of people in each city. Fig. 6 illustrates the population distribution of the clusters. The interquartile ranges of Cluster 1 are very high compared to the other clusters. Although Cluster 1 consists of only four cities, the larger interquartile ranges indicate large variation in population size among cities. In contrast, narrow interquartile ranges in all other clusters indicate small variations in population size among the cities. The median population size of Cluster 2 drops eight

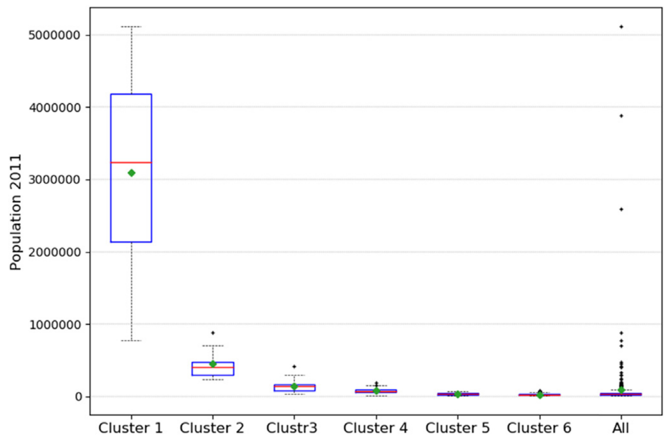

Fig. 6. Box plot of population distribution in each cluster. The upper and lower bounds of the box indicate the 75th and 25th percentiles of the distribution, respectively. The mean (green diamond mark), median (red bar at the middle of the box), and minimum-maximum bound (dash line) of each cluster are also shown in the figure. Black asterisk marks indicate outliers of the distribution (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

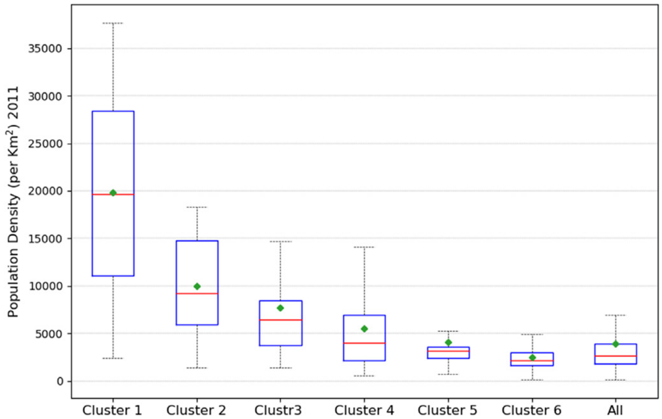

Fig. 7. The distribution of population density in each cluster. The upper and lower bounds of the box indicate the 75th and 25th percentiles of the distribution, respectively. The mean (green diamond mark), median (red bar at the middle of the box), and minimum-maximum bound (dash line) of each cluster are also shown in the figure (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

times from Cluster 1 and is followed by a gradual decline in the subsequent clusters. The median population size of clusters indicates very high population concentration in a few cities. Therefore, policymakers need to pay attention to population decentralization to achieve balanced growth. This classification system identifies clusters of small cities which require attention to accelerate the growth in future. The box plots in Fig. 6 show most of the cities in Cluster 4, Cluster 5, and Cluster 6, have much smaller populations compared to big urban areas. Therefore, it is necessary for policymakers to shift their focus towards those small and medium-sized cities, so that these cities can absorb migrants from rural areas.

Big cities may have large populations, therefore, it is difficult to compare cities based on total population. Population density per square kilometer can capture the magnitude of population concentration. Fig. 7 illustrates the distribution of population density in each cluster. Cluster 1 has mean density of nearly 20,000 person per square kilometer, whereas Cluster 2 has a population density less than half of Cluster 1. Small cities in Cluster 5 and Cluster 6 have very low population densities compared to big urban areas (Cluster 1 and Cluster 2). Both mean and median population density gradually decreases from Cluster 1 through Cluster 6. The result in Fig. 7 shows that city classification based on spatial attributes can capture the population dynamics of many cities.

The population growth trend in the most recent decade reveals another interesting characteristic of cities in Bangladesh and again validates the classification system. Cities in Cluster 1 experienced population growth over 3%, while cities in Cluster 2 and Cluster 3 experienced population growth just above 2%. Population growth is higher in large metropolitans (Cluster 1) compared to cities in other clusters (Fig. 8). The longest interquartile ranges indicate large variation in population growth among cities in Cluster 2. A higher mean population growth than median population growth in Cluster 2 and Cluster 3 indicate the right-skewed distribution of the population growth. Therefore, few cities have exceptional high population growth rate in these two clusters. Higher population growth in major cities may indicates in-migration of people for better opportunities. Therefore, cities in Cluster 2 are the potential options for people to migrate to search of jobs after the Cluster 1, where new infrastructure developments and job opportunities were generated in the recent past. Cluster 4 and Cluster 5 have almost similar population growth of about 2% per year (Fig. 8). Cluster 6

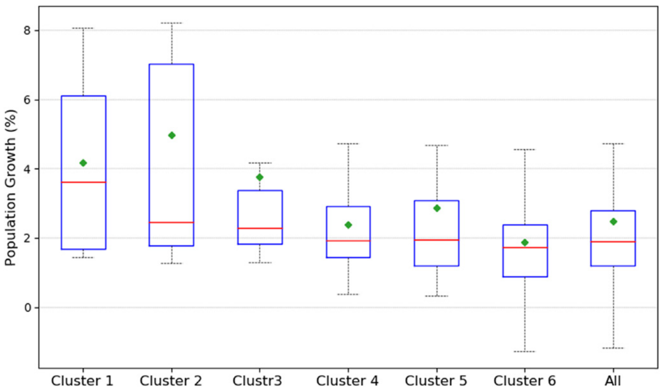

Fig. 8. Population growth rate distribution of clusters. The upper and lower bounds of the box indicate the 75th and 25th percentiles of the distribution, respectively. The mean (green diamond mark), median (red bar at the middle of the box), and minimum-maximum bound (dash line) of each cluster are also shown in the figure (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

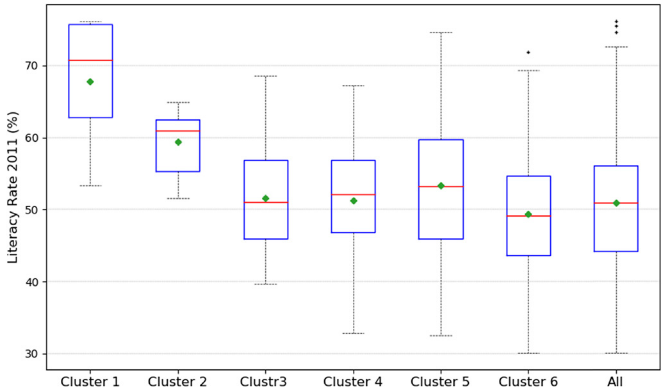

Fig. 9. Distribution of literacy rate in each cluster. The upper and lower bounds of the box indicate the 75th and 25th percentiles of the distribution, respectively. The mean (green diamond mark), median (red bar at the middle of the box), and minimum-maximum bound (dash line) of each cluster are also shown in the figure. Black asterisk marks indicate outliers of the distribution (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

has the lowest median and mean population growth.

Similar to population size and growth rate, Cities of Cluster 1 exhibit an exceptional literacy rate, more than 70% compared to cities in the other clusters. Cluster 2 also has a relatively higher literacy rate, just above 60%, than all the moderate and small sized cities (Fig. 9). The median literacy rate is similar for other clusters of small and medium-sized cities of Bangladesh. The similar median literacy rate of the clusters can be explained by the widespread activities of the government to reduce illiteracy throughout the country(Al Amin, 2010). Both educated migrants and better education facilities are the reason for high literacy rate in major cities (M. Ahmed, Chabbott, Joshi, Pandi, & Prather, 1993). This balanced condition in the literacy rate across the small and medium-size cities is the outcome of the widespread efforts of the government across the cities of Bangladesh. Similar interquartile ranges in most of the clusters also indicate low variety in literacy rate among cities in each cluster.

3.2.2. *Employment*

Sector-wise employment share is one of the important city characteristic. The definition of an urban area in Bangladesh explicitly defines the dominance of nonprimary sector employment. Fig. 10 illustrates the comparison of sector-wise employment among clusters. Cluster 1 consists of large metropolitans and, has a negligible share of agricultural sector employment. The share of the agriculture sector increases gradually from Cluster 1 to Cluster 6. The mean share of the agriculture sector in Cluster 1 is less than 1% whereas one-third of the working class in small cities (Cluster 6) are engaged in agricultural activities. An opposite scenario exists in the service sector, where the employment share in the service sector gradually decreases from Cluster 2 to Cluster 6. The mean share of the service sector in total employment for Cluster 1, Cluster 2 and Cluster 3 is near 75%, whereas it is only about 60% in Cluster 5

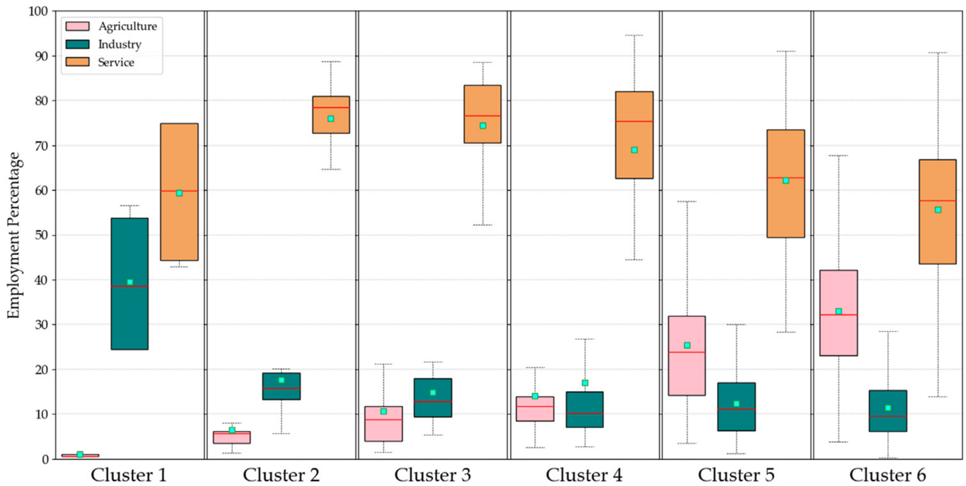

Fig. 10. Sector-wise employment share in each cluster. The upper and lower bounds of the box indicate the 75th and 25th percentiles of the distribution, respectively. The mean (cyan square mark), median (red bar at the middle of the box), and minimum-maximum bound (dash line) of each cluster are also shown in the figure (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

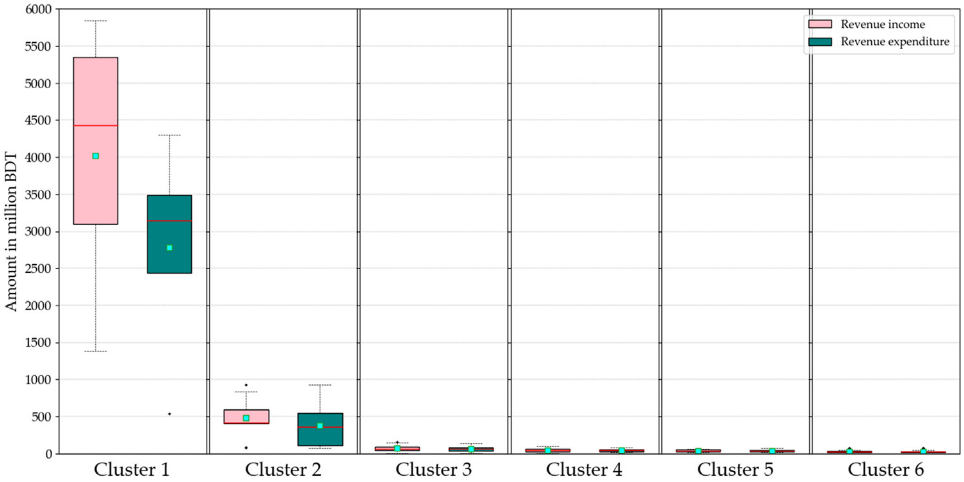

Fig. 11. Revenue income and expenditure distribution of clusters. The upper and lower bounds of the box indicate the 75th and 25th percentiles of the distribution, respectively. The mean (cyan square mark), median (red bar at the middle of the box), and minimum-maximum bound (dash line) of each cluster are also shown in the figure (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

and Cluster 6. Relatively longer interquartile ranges in both the agricultural and the service sectors in Cluster 5 and Cluster 6 indicate variation among cities in these two clusters. Although Cluster 1 consists of only four cities, the longer interquartile ranges points to large variations that exist in service and industrial employment share among the cities. Cluster 1 has 60% employment in service sector because of its large share of employment in the industrial sector (Fig. 10). Cluster 3 and Cluster 4 consisting moderate size cities have a close percentage of job share in the agricultural and industrial sectors. Most of the industrial establishments are concentrated in four major cities in Bangladesh. Thus, there is an enormous disparity in industrial sector employment between Cluster 1 and other clusters. The mean percentage share of industrial sector employment in Cluster 1 is 40%, whereas this sector contributes only between 10–15% of the jobs in other clusters. Cluster 1 is the only case where the interquartile ranges of service sector and industrial sector are overlapped. All other clusters show a large difference between employment share of the two nonprimary sectors. Employment share in each cluster clearly justifies that the proposed city classification system can capture the employment characteristics of cities.

3.2.3. Income and expenditure

Revenue income of a city is the income from its own sources. Development income is mainly government and nongovernment grants from external sources. Development expenditure corresponds to the expenditure which is allocated for economic and infrastructure development. In contrast, revenue expenditure is money spent on running local government departments and services. Fig. 11 shows that the mean and median revenue incomes are significantly lower from Cluster 3 to Cluster 6, compared to Cluster 1 and Cluster 2. Also, there is a large difference in median income between Cluster 1 and Cluster 2. Likewise, Fig. 12 shows similar distribution in development income expenditure across clusters. Cluster 1 has considerably higher development income and expenditure than the other clusters. The presence of outliers in Cluster 2 indicates one city in this cluster attracts exceptional government and nongovernment support for development projects. The revenue income is very low in moderate sized and small cities compared to cities in Cluster 1 and Cluster 2. The median revenue goes through a sharp decline after the second cluster and flattens for the rest of the clusters. The very narrow interquartile ranges in revenue income indicate very close revenue generation potentialities of cities in Cluster 3 to Cluster 6. Although the spatial and demographic attributes among the clusters are different, similar revenue potentials of Cluster 3 to Cluster 6 indicates revenue generation does not rely on these attributes. A similar pattern in both revenue income and development budget also indicates if a city has low revenue potentials, meaning it receives few resource for the development. Two major mechanism for local governments in generating economic growth are through revenue expenditure and development expenditure. The median values of both of these variables are much higher in the developed cities and lower in the least developed areas. The analysis of government total expenditure not only validates the new classification approach for cities in Bangladesh but also directs towards imbalanced spatial development across the country.

It is obvious that cities with large populations require a large budget. Therefore, discussion of total income and expenditure may not capture the disparity among cities properly. Per capita income and expenditure may better explain public income and expenditure of cities. Therefore, the combined spending from both revenue and development sector is converted into per capita expenditure by dividing total revenue by the population in 2011. Fig. 13 illustrates per capita income and expenditure distributions across the clusters. Cluster 2 has the highest median per capita income and expenditure, whereas Cluster 1 has highest total income and expenditure. Higher per capita income of major cities indicates that the largest share of the expenditure is allocated for the development of bigger cities which is contributing to the rapid growth and development of these areas compared to other cities in Bangladesh. Long interquartile ranges of Cluster 2 also indicate variation in per capita expenditure among big cities. Cluster 3

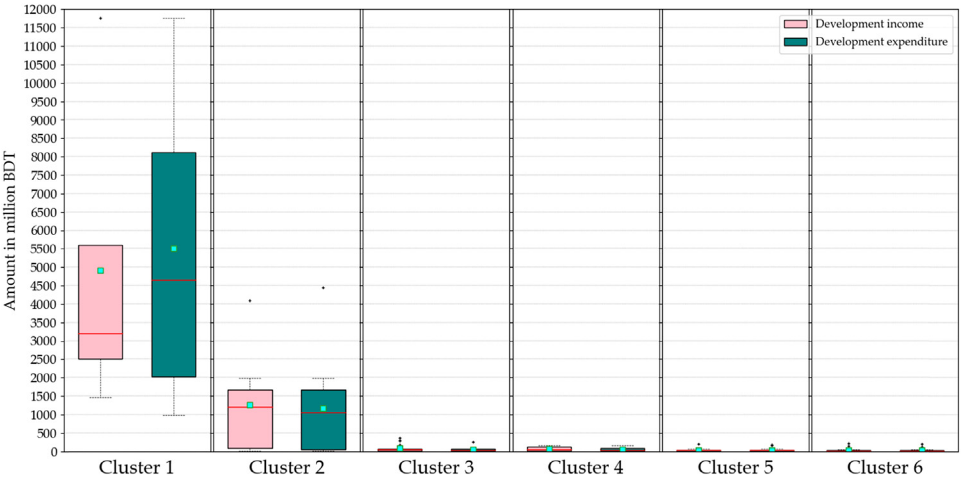

Fig. 12. Development income and expenditure distribution of clusters. The upper and lower bounds of the box indicate the 75th and 25th percentiles of the distribution respectively. The mean (cyan square mark), median (red bar at the middle of the box), and minimum-maximum bound (dash line) of each cluster are also shown in the figure. Black asterisk mark indicates outliers of the distribution (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

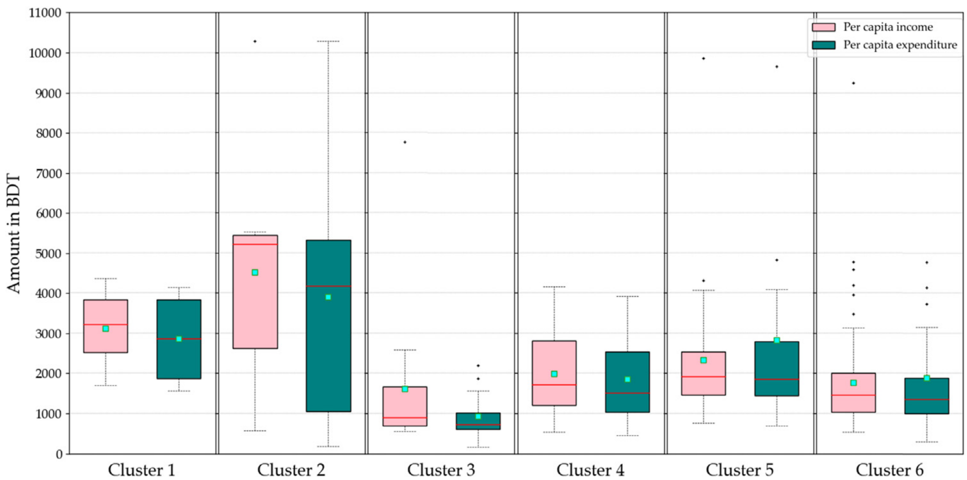

Fig. 13. Cluster-wise median per capita total income and expenditure. The upper and lower bounds of the box indicate the 75th and 25th percentiles of the distribution respectively. The mean (cyan square mark), median (red bar at the middle of the box), and minimum-maximum bound (dash line) of each cluster are also shown in the figure. Black asterisk mark indicates outliers of the distribution (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

consisting of medium-sized cities has the lowest median per capita income and expenditure. The total income and the total expenditure of cities are similar mostly because the expenditure depends on income. A significant difference between income and expenditure can be seen only in Cluster 2 which indicates more spending capabilities of these cities for development. Also, the lower per capita income indicates the limitations of small and medium-sized cities in generating their own resources for local development purposes. This is a critical policy concern affecting the balanced spatial development of cities in Bangladesh. The major cities are receiving the bulk share of development expenditures which may be utilized for infrastructure development consequently built-up areas are increased in these cities. The small and medium-sized cities are lagging due to the lack of revenue potential and development expenditure allocation.

3.3. Discussion on cluster significance

Table 4 summarizes the median values of the selected variables across the six clusters and their respective H values of the Kruskal Wallies test. The H values for all the selected non-spatial variables are higher than the critical value, and the p-value indicates results are significant at the 95 percent confidence level. Therefore, the null hypothesis that results are from an identical population can be rejected. The median population of Cluster 1 is roughly eight times larger than Cluster 2, the nearest group of cities to Cluster 1.

Table 4
The Kruskal Wallies Statistical Significance test values for seven selected non-spatial variables across the six clusters.

Variables		Unit	C 1	C 2	C 3	C 4	C 5	C 6	H	p-value
Demography	Population	Number (thousand)	3235	401	144	67.83	31.91	26.19	143.73	0.00000
	Population Density	Number per Km^2	19655	9180	6449	3967	3138	2150	77.20	0.0000
	Population growth	%	3.61	2.45	2.28	1.92	1.95	1.73	17.81	0.0031
	Literacy rate	%	70.75	60.9	51	52.1	53.2	49.2	27.69	0.00004
Employment	Agriculture	%	0.63	5.66	8.75	11.73	23.87	32.16	107.05	0.0000
	Industry	%	38.63	15.72	12.9	10.18	11.1	9.56	20.02	0.0012
	Service	%	59.84	78.47	76.68	75.43	62.82	57.56	57.59	0.000
Income and Expenditure	Revenue income	BDT (Crore)	442.6	41.2	5.94	4.53	2.97	2.34	61.40	0.00000
	Revenue expenditure	BDT (Crore)	314.4	36.1	5.8	4.3	2.92	2.12	59.81	0.00000
	Development income	BDT (Crore)	320.2	120.5	4.6	4.4	2.3	1.8	42.54	0.00000
	Development expenditure	BDT (Crore)	464.3	104.6	2.9	3.7	2.3	1.9	32.78	0.00000
	Per capita Income	BDT	3215	5215	892	1714	1924	1462	16.27	0.006
	Per capita expenditure	BDT	2861	4180	724	1508	1862	1358	23.73	0.0002

Cluster 5 and Cluster 6 have median populations one-hundredth of Cluster 1. The median population growth is higher in major cities (Cluster 1 and Cluster 2) than the other clusters. Cluster 1 has more than one percent additional growth rate than the nearest cluster, Cluster 2. The multiplier effect of a larger population size and higher growth rate in a few cities will increase population in these areas. The median population gradually decreases from Cluster 1 to Cluster 6, which indicates the strong relationship between a city's spatial attributes and population density. There is also a large difference observed among big, medium, and small cities in literacy rate. The difference in literacy rate points to the possible two scenarios: the big cities provide better education facilities or the migration of educated people towards big agglomerations. Other important aspects of cities are income and expenditure. Table 4 indicates, that the median revenue income of cities in Cluster 1 is roughly two-hundred times higher than these in Cluster 6, while population size is one-hundred time larger. However, revenue income is almost proportional to population considering the relation of Cluster 1 and Cluster 2, where both clusters consist of city corporations. This result implies that revenue income is not proportional to the population of the city when comparing small cities. The difference between revenue income and expenditure is prominent only in Cluster 1, which means only big cities have a surplus in their revenue income. Development budget which is mainly grants from the central government and international aid, is proportionally higher in Cluster 2 considering the population size and revenue income. Putting government investment in big cities (Cluster 2) outside of large metropolitans (Cluster 1) may point to the government's initiative of decentralization. Cluster 3, consisting mostly medium size cities, has the lowest median per capita income and expenditure which indicating negligence towards these medium cities.

Employment brings an interesting relationship between city classification and sector wise employment. The share of the agricultural sector in employment gradually increases from less than 1% in Cluster 1 to more than 31% in Cluster 6. The agriculture sector has one-third share of the job market in Cluster 6, whereas the share of the agricultural sector is negligible in Cluster 1. An opposite scenario exists for the service sector except in Cluster 1. A gradual decrease in the percentage of service sector job can be observed from about 78% in Cluster 2 to 58% in Cluster 6. Cluster 1 is out of this trend because of its significant share in the industrial sector. While Cluster 1 has nearly 39% employment related to industry, only 10% jobs are created by the industrial sector in other clusters. Therefore, industrial establishment only concentrate in and around few large metropolitans of Bangladesh.

Table 5 shows the results of the pairwise Mann-Whitney U test to describe the similarities and dissimilarities between classes of

Table 5
The results of the pairwise Mann–Whitney *U* tests between cluster pairs for selected non-spatial variables.

	Variables	C1	C2	C2	C3	C3	C4	C4	C5	C5	C6
		U	*p*	U	*p*	U	*p*	U	*P*	U	*p*
Demography	Population	1	**0.005**	7	**0.000**	187	**0.000**	202	**0.000**	4119	**0.018**
	Population Density	11	0.157	89	*0.098*	291	**0.015**	623	*0.070*	3366	**0.000**
	Population Growth	16	0.438	114	0.342	270	0.115	414	0.395	1913	*0.053*
	Literacy Rate	8	*0.071*	45	**0.003**	430	0.478	732	0.244	3964	**0.002**
Employment	Agriculture	1	**0.005**	87	*0.086*	259	**0.023**	303	**0.004**	2774	**0.000**
	Industry	2	**0.008**	94	0.136	327	0.197	552	0.367	3744	0.241
	Service	6	**0.037**	123	0.464	325	0.190	406	**0.017**	3171	**0.015**
Income and Expenditure	Revenue Income	0	**0.003**	11	**0.000**	144	**0.044**	198	0.111	550	**0.005**
	Revenue Expenditure	3	**0.012**	12	**0.000**	158	*0.091*	202	0.130	507	**0.002**
	Development Income	5	**0.026**	27	**0.002**	195	0.351	170	**0.032**	666	*0.056*
	Development Expenditure	6	**0.038**	37	**0.006**	171	0.160	195	*0.099*	642	**0.036**
	Per capita Income	10	0.175	33	**0.043**	85	**0.036**	185	0.286	552	**0.024**
	Per capita expenditure	14	0.399	31	**0.032**	59	**0.003**	151	0.072	476	**0.004**

*Values are statistically significant at $\alpha = 0.05$ are in bold and values are significant at $\alpha = 0.1$ are in italic.

cities. Only five pairwise comparisons are shown among 15 possible comparisons since if a class is different from its neighboring class, it must be different from other classes. There is a significant difference in population size between cluster pair. Since population growth is close among clusters, none of the pairs shows a significant difference. There are significant differences in literacy rate between Cluster 2 and Cluster 3 as well as for Cluster 5 and Cluster 6. Since industries are only concentrated near large cities, the pair of Cluster 1 and Cluster 2 is significant in this regard. All other clusters have almost similar industrial concentration, therefore, there is no significant difference among them. Agriculture activities are significantly different among most of the cluster pairs. The share of service sector in employment shows a significant difference between large metropolitan and big cities (the pair of Cluster 1 and Cluster 2). However, there is no significant difference in service sector employment share between big cities (Cluster 2) and moderate cities (Cluster 3). Interestingly, small cities show significant dissimilarity in service sector employment share. Revenue income and expenditure are significantly different in most of the cases except the pair of moderate sized cities (Cluster 3 and Cluster 4) and pair of moderate and small sized cities (Cluster 4 and Cluster 5). Development income and expenditure show similar results as of revenue income and expenditure. There is also a significant difference in per capita income and expenditure among cluster pairs except for the pair of moderate and small sized cities.

4. Conclusion

Cities in Bangladesh have different sizes and rates of urban growth. Based on conventional city classification system, development budget allocation and infrastructure development often rely on the city's existing status. Therefore, proper city classification based on spatial characteristics of cities is essential for optimum resource allocation and balanced development. This study classifies 331 cities in Bangladesh based on five parameters: city size, urban growth, the built-up area ratio, urban form (AWMPFD), and night light intensity. Since information on the built-up area and urban spatial growth are not available for Bangladesh, this study relies on landcover information extracted from satellite imageries. Remote sensing based urban growth and extent information is useful not only in Bangladesh but also anywhere in the world where field level data are scarce. The current study grouped 331 cities into six clusters with significant differences in development conditions among them. Four of 12 city corporations (also known as major cities) are grouped as Cluster 1. These four large urban agglomerations have different characteristics than other city corporations. Two class-A municipalities are grouped together with seven other city corporations, which indicates that these two cities have similar potentialities as city corporations. A number of existing class-A municipalities are grouped with small and very small cities, indicating that the statuses of these cities are overestimated. These six clusters are then reviewed with regards of their demographic, employment as well as income and expenditure aspects. The significant difference among clusters on the selected non-spatial attributes justifies the appropriateness of the proposed clustering procedure used in this study. The characteristics of the clusters can be described by the selected non- spatial variables. Population, population growth rate, literacy rate, employment, income, and expenditure all correspond to the spatial characteristics of the clusters.

City authorities in developing countries often have very limited data on the growth and development of their cities. This study demonstrated the usefulness of remote sensing data for developing countries, where data on cities' spatial characteristics and development are scarce. Therefore, the use of remote sensing data can be a cost-effective solution in developing countries to understand the spatial development condition and to preparing appropriate policy decisions for balanced spatial development. This city classification approach based on development level would be very helpful for local level policy decisions. The small municipalities of Bangladesh are heavily dependent on the central government for resource allocation. However, it is a policy question for the central government during each budgetary period to determine where to place resources for efficient allocation and alleviation of regional disparities in the country. This classification can guide government policymakers on budgetary resource allocation. The LGED allocates resources based on municipality class, which is completely based on the revenue potential of the municipality and ignores spatial attributes of cities. Therefore, present resource allocation may not be efficient nor equitable, as the system does not capture the spatial development patterns of the area. Consequently, the classification methodology of this study would support long-term efficiency as well as equity in the resource allocation and development of urban areas in Bangladesh. For the efficient decentralization of Bangladesh to reduce the pressure on larger cities, the development of these small and medium-size cities is necessary. This classification can guide policymakers to focus on the development of the cities in different clusters. The cluster results show the levels of development as well as of resource allocation are deficient in small and medium-sized urban centers. Thus, it necessary to reconsider the strategy of allocating resources for these small and medium-sized urban areas to assist in their development of these areas. The methodology applied in this study can hopefully super new and innovative methods for resource allocation to small and medium-sized urban areas.

This proposed city classification technique is advantageous over the existing classification system since it considers important spatial aspects of cities instead concentrating on a single aspect (population size or revenue). Therefore, policymakers can use proposed city classification system for better budget allocation and development decision making for cities in Bangladesh. Since there is a colossal disparity between big cities (mostly city corporations) and municipalities, future classification studies only on municipalities (Paurashava) would be helpful. Indicators of city form (compact vs. sprawl) such as Gini coefficient or, Moran's I and landscape functionalities can also be considered for city classification in future research.

Acknowledgments

This study was partially supported by grants from the NASA Applied Science Program (Grant #NNX14AP91G and #NNX15AK27A, PI: Liping Di). We would like to express our sincere appreciation to the United States Geological Department (USGS) for Landsat archives, the Bangladesh Bureau of Statistics (BBS) for population and household data, and the Local Government Engineering Department (LGED) for revenue and budget data for municipalities. This research used data from different sources. Therefore, many names need to be mentioned who helped at different stages of the study. Shahriar Amin and Bivash Das of LGED. Abu Musa Md. Abdullah, Abu Saleh Md. Shahidullah, and Sonia Rahman of Sheltech Consultants (Pvt.) Limited. Md Nasim Haider and Nessar Uddin Ahmed of Development Design Consultants Ltd. S.K. Towhidur Rahaman from Save the Children International. Bibi Khadiza of Engineering and Planning Consultants (EPC). Sanjib Saha of AQUA Consultants and Associate. A. N. M. Rashed Hasan of Bangladesh Rural Electrification Board. Mohammad kamruzzman and Md. Golam Mortoja of Institute of Water Modeling (IWM). Al Mobin Rahman of Sreebardi Paurashava, Al Mehedi Hasan of Bogra Paurashava, Rumana Naznin of Manikgonj Paurashava. Nandita Basu of Barisal City Corporation. Md. Nazrul Islam of Rangpur City Corporation. Kutub Uddin Chisty of Coxbazar Development Authority. Sanjida Hoque of Gazipur City Corporation. Finally, the authors also would like to give special thanks to Zhi Chen of the School of Urban and Regional Planning, The University of Iowa for initial proofreading the manuscript.

References

Ahmed, B., Kamruzzaman, M., Zhu, X., Rahman, M. S., & Choi, K. (2013). Simulating land cover changes and their impacts on land surface temperature in Dhaka, Bangladesh. *Remote Sensing, 5*(11), 5969–5998.

Ahmed, M., Chabbott, C., Joshi, A., Pandi, R., & Prather, C.J. (1993). Primary Education for All: Learning from the BRAC Experience, A Case Study.

Al Amin, M.M. (2010). Factors behind internal migration and migrant's livelihood aspects: Dhaka City, Bangladesh.

Baugh, K., Elvidge, C. D., Ghosh, T., & Ziskin, D. (2010). Development of a 2009 stable lights product using DMSP-OLS data. *Proceedings of the Asia-Pacific Advanced Network, 30*, 114–130.

BBS (2014). *Bangladesh population and Housing Census 2011, National Report, Volume-3: Urban area report* (1st edition). Dhaka: Bangladesh Bureau of Statistics.

Bhandari, L., & Roychowdhury, K. (2011). Night lights and economic activity in India: A study using DMSP-OLS night time images. *Proceedings of the Asia-Pacific Advanced Network, 32*, 218–236.

Burrows, M. (2012). *Global trends 2030: Alternative worlds.* US National Intelligence Council.

Burt, J. E., Barber, G. M., & Rigby, D. L. (2009). *Elementary statistics for geographers* (Third Edition). Guilford Press.

Chatterji, M., & Yang, K. (2016). *Regional science in developing countries.* Springer.

Doll, C. N. (2010). Population detection profiles of DMSP-OLS night-time imagery by regions of the world. *Proceedings of the Asia-Pacific Advanced Network, 30*, 190–206.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., & Leitão, P. J. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography, 36*(1), 27–46.

Dugoua, E., Kennedy, R., & Urpelainen, J. (2018). Satellite data for the social sciences: Measuring rural electrification with night-time lights. *International Journal of Remote Sensing, 39*(9), 2690–2701.

Elvidge, C. D., Baugh, K. E., Anderson, S. J., Sutton, P. C., & Ghosh, T. (2012). The Night Light Development Index (NLDI): A spatially explicit measure of human development from satellite data. *Social Geography, 7*(1), 23–35.

Faguet, J.-P., & Shami, M. (2008). Fiscal policy and spatial inequality in Latin America and beyond (LSE Research Online Documents on Economics No. 27162). *London School of Economics and Political Science*. (LSE Library. Retrieved from) 〈https://ideas.repec.org/p/ehl/lserod/27162.html〉.

Frenkel, A. (2004). Land-use patterns in the classification of cities: The Israeli case. *Environment and Planning B: Planning and Design, 31*(5), 711–730.

Ghosh, T., Anderson, S. J., Elvidge, C. D., & Sutton, P. C. (2013). Using nighttime satellite imagery as a proxy measure of human well-being. *Sustainability, 5*(12), 4988–5019.

Goerlich Gisbert, F. J., Cantarino Martí, I., & Gielen, E. (2017). Clustering cities through urban metrics analysis. *Journal of Urban Design, 22*(5), 689–708.

Gross, A. L. (1975). Analyzing multivariate data. *Psyccritiques, 20*(9), 743–744.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis*, *6*. Upper Saddle River, NJ: Pearson Prentice Hall.

Harris, C. D. (1943). A functional classification of cities in the United States. *Geographical Review, 33*(1), 86–99.

Heikkila, E., & Xu, Y. (2014). Seven prototypical Chinese cities. *Urban Studies, 51*(4), 827–847.

Henderson, M., Yeh, E. T., Gong, P., Elvidge, C., & Baugh, K. (2003). Validation of urban boundaries derived from global night-time satellite imagery. *International Journal of Remote Sensing, 24*(3), 595–609.

Herold, M., Scepan, J., & Clarke, K. C. (2002). The use of remote sensing and landscape metrics to describe structures and changes in urban land uses. *Environment and Planning A, 34*(8), 1443–1458.

Huang, J., Lu, X. X., & Sellers, J. M. (2007). A global comparative analysis of urban form: Applying spatial metrics and remote sensing. *Landscape and Urban Planning, 82*(4), 184–197.

Huang, Q., Yang, X., Gao, B., Yang, Y., & Zhao, Y. (2014). Application of DMSP/OLS nighttime light images: A meta-analysis and a systematic literature review. *Remote Sensing, 6*(8), 6844–6866.

Jeff C. and Star Y. (n.d.). Data Science for Nighttime Lights. Retrieved May 12, 2018, from 〈https://commercedataservice.github.io/tutorial_viirs_part1/〉.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis*, *344*. John Wiley & Sons.

Khan, A. M., & Islam, I. (2010). Education and development in Bangladesh: A study from spatial perspective. *Journal of Bangladesh Institute of Planners, 3*.

Khan, H. (2008). *Challenges for sustainable development: Rapid urbanization, poverty and capabilities in Bangladesh (MPRA Paper No. 9290)*. Germany: University Library of Munich. (Retrieved from) 〈https://ideas.repec.org/p/pra/mprapa/9290.html〉.

Kline, R. B. (2015). *Principles and practice of structural equation modeling.* Guilford publications.

Kock, N., & Lynn, G. (2012). Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations.

LGED. (n.d.). Urban Management Unit. Retrieved May 12, 2018, from 〈http://www.lged.gov.bd/UnitAbout.aspx?UnitID=10〉.

Lo, C. P. (2002). Urban indicators of china from radiance-calibrated digital dmsp-ols nighttime images. *Annals of the Association of American Geographers, 92*(2), 225–240.

Ma, T., Zhou, C., Pei, T., Haynie, S., & Fan, J. (2012). Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: A comparative case study from China's cities. *Remote Sensing of Environment, 124*, 99–107.

Ma, X., Tong, X., Liu, S., Luo, X., Xie, H., & Li, C. (2017). Optimized sample selection in SVM classification by combining with DMSP-OLS, Landsat NDVI and GlobeLand30 products for extracting urban built-up areas. *Remote Sensing, 9*(3), 236.

Mahmoud, C.S., Wadood, S.N., & Ahmed, K.S. (2008), March 13). Addressing Regional Inequality Issues in Bangladesh Public Expenditure [MPRA Paper]. Retrieved May 12, 2018, from 〈https://mpra.ub.uni-muenchen.de/14568/〉.

McGarigal, K., & Marks, B.J. (1995). FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. Gen. Tech. Rep. PNW-GTR-351. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 122 P, 351.

Mellander, C., Lobo, J., Stolarick, K., & Matheson, Z. (2015). Night-time light data: A good proxy measure for economic activity? *PLoS One, 10*(10), https://doi.org/10.1371/journal.pone.0139779.

Muzzini, E., & Aparicio, G. (2013). *Bangladesh: The path to middle-income status from an urban perspective.* World Bank Publications.

NOAA (n.d.). Earth Observation Group - Defense Meteorological Satellite Progam, Boulder | ngdc.noaa.gov. (Retrieved 13 May 2018), from 〈https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html〉.

Noor, A. M., Alegana, V. A., Gething, P. W., Tatem, A. J., & Snow, R. W. (2008). Using remotely sensed night-time light as a proxy for poverty in Africa. *Population Health Metrics, 6*(1), 5.

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality Quantity, 41*(5), 673–690.

Planning Commission (2005). *Unlocking the potential: National strategy for accelerated poverty reduction*. General Economics Division, Planning Commission, Government of the People'States Republic of Bangladesh.

Planning Commission (2008). *A strategy for poverty reduction in the lagging regions of Bangladesh*. General Economics Division, Planning Commission, Govt. of the People's Republic of Bangladesh.

Planning Commission (2011), G. of B. P. (Ed.). Sixth Five Year Plan of Bangladesh, FY 2011-FY 2015, Part 1, 2, 3. Planning Commission, Government of Bangladesh.

Prosperie, L., & Eyton, R. (2000). The relationship between brightness values from a nighttime satellite image and Texas county population. *Southwestern Geographer, 4*, 16–29.

Rahman, M. M. (2004). Regionalization of urbanization and spatial development: Planning regions in Bangladesh. *Journal of Geo-Environment, 4*, 31–46.

Rahman, M. S. (2012). Quantifying urban form-compactness versus sprawl: An analysis of Chittagong and Sylhet metropolitan cities. *Jahangirnagar University Planning Review, 10*, 39–50.

Rahman, M. S., Ahmed, B., & Di, L. (2017). Landslide initiation and runout susceptibility modeling in the context of hill cutting and rapid urbanization: A combined approach of weights of evidence and spatial multi-criteria. *Journal of Mountain Science, 14*(10), 1919–1937.

Rahman, M. S., Islam, B., & Ahmed, B. (2012). An overview on rapid urbanization and induced disaster risk factors in Bangladesh. *World Town Planning Day,* 56–64.

Rahman, M. S., & Islam, M. A. (2013). Decentralization in urban land-use planning in Bangladesh: Rationality of the scene behind a screen. In Golam Rahman (Ed.). *Planned Decentralization: Aspired Development, Bangladesh Institute of Planners* (pp. 5–14). .

Rahman, M. S., Yang, R., & Di, L. (2018). Clustering Indian Ocean tropical cyclone tracks by the standard deviational ellipse. *Climate, 6*(2), 39. https://doi.org/10.3390/cli6020039.

Roberts, B., & Kanaley, T. (2006). *Urbanization and Sustainability in Asia: Case Studies of Good Practice*. Asian Development Bank.

Rohlf, F. J., & Fisher, D. R. (1968). Tests for hierarchical structure in random data sets. *Systematic Biology, 17*(4), 407–412.

Rossiter, D. G. (2004). *Technical note: Statistical methods for accuracy assessment of classified thematic maps.* the Netherlands: Enschede.

Sambandam, R. (2003). Cluster analysis gets complicated. *Marketing Research, 15*(1), 16–21.

Schwarz, N. (2010). Urban form revisited—Selecting indicators for characterising European cities. *Landscape and Urban Planning, 96*(1), 29–47.

Sen, B., & Ali, Z. (2009). Spatial inequality in social progress in Bangladesh. *The Bangladesh Development Studies,* 53–78.

Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy, 85*(3), 257–268.

Singh, W. (2008). Robustness of three hierarchical agglomerative clustering techniques for ecological data. Access Mode: 〈Http://www.Unuftp.is/static/files/rannsoknarritegrdir/WarshaSingh_MastersThesis.Pdf〉.

Sobolevsky, S., Sitko, I., Grauwin, S., Combes, R.T.D., Hawelka, B., Arias, J.M., & Ratti, C. (2014). Mining urban performance: Scale-independent classification of cities based on individual economic transactions. arXiv Preprint arXiv:1405.4301.

Sutton, P., Roberts, D., Elvidge, C., & Meij, H. (1997). A comparison of nighttime satellite imagery and population density for the continental United States. *Photogrammetric Engineering and Remote Sensing, 63*(11), 1303–1313.

Vanderhaegen, S., & Canters, F. (2017). Mapping urban form and function at city block level using spatial metrics. *Landscape and Urban Planning, 167*, 399–409.

World Bank (2007). *Bangladesh – Strategy for sustained growth*, *No. 41245*, The World Bank1–202. (Retrieved from) 〈http://documents.worldbank.org/curated/en/806041468208781718/Bangladesh-Strategy-for-sustained-growths〉.

Zhou, N., Hubacek, K., & Roberts, M. (2015a). Analysis of spatial patterns of urban growth across South Asia using DMSP-OLS nighttime lights data. *Applied Geography, 63*, 292–303.

Zhou, Y., Smith, S. J., Zhao, K., Imhoff, M., Thomson, A., Bond-Lamberty, B., ... Elvidge, C. D. (2015b). A global map of urban extent from nightlights. *Environmental Research Letters, 10*(5), 54011.