The ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 4. Issue 2 (2024)
https://doi.org/ 10.62019/abbdm.v4i02.12

ASIAN BULLETIN OF BIG DATA MANAGEMENT
http://abbdm.com/
ISSN (Print): 2959-0795
ISSN (online): 2959-0809

Optimizing Urban Delimitations in Pakistan Through Exploratory Data Analytical Approach

Sahar Zia, Aqsa Shabbir*, Ali Hussain Kazim, Aysha Hanif, Huma Tauseef, Naveed Iqbal

Chronicle

Article history

Received: April, 20,2024
Received in the revised format: April 28, 2024
Accepted: April 31, 2024
Available online: May 6, 2024

Sahar Zia & Aysha Hanif are currently affiliated with the Department of Geography, Lahore College for Women University, Lahore, Pakistan.
Email: sahar.zia@lcwu.edu.pk
Email: drayshahanif@gmail.com

Aqsa Shabbir is currently affiliated with the Department of Electrical Engineering, Lahore College for Women University, Lahore, Pakistan.
Email: aqsa.shabbir@lcwu.edu.pk

Ali Hussain Kazim are currently affiliated with the Department of Automotive Engineering, University of Engineering and Technology, Lahore, Pakistan.
Email: ali.h.kazim@uet.edu.pk

Huma Tauseef is currently affiliated with the Department of Computer Science, Lahore College for Women University, Lahore, Pakistan.
Email: huma.tauseef@lcwu.edu.pk

Naveed Iqbal is currently affiliated with the Department of Mass Communication, Lahore College for Women University, Lahore, Pakistan.
Email: naveed.iqbal@lcwu.edu.pk

Corresponding Author*

Abstract

In Pakistan, the lack of a standardized definition for urban areas hampers effective urban planning and policy development. Thus, this research study employs data science by integrating the Geographic Information System (GIS) employed to find the weightage of the urban threshold that can redefine the urban areas in Pakistan. With diverse socio-economic and demographic factors considered, exploratory data analysis (EDA) incorporates population density, infrastructure development, economic indicators, and land use patterns to demark urban boundaries. As there is no consensus on what an “urban area” in Pakistan should be, this study takes a data-focused approach to determine the criteria that best capture the complexities of urbanization. It has been done through spatial analytics and statistical modeling in a data science framework. The integration of a data science framework enables to set-up threshold for development of urban landscape. The proposed redefinition of urban areas has implications for urban planning, resource allocation, and policy development, fostering a more accurate representation of the urban fabric in Pakistan. This study contributes to filling the existing gap in defining urban areas in Pakistan. It identifies new urban percentages that more accurately reflect the socio-economic and demographic realities of Pakistan's urban landscape. The findings shows that there are 36 percent of urban grid cells retrieved from the EDA approach. The outcomes provide a foundation for informed decision-making and policy interventions, aligning urban definitions with Pakistan's urbanization process's complex and dynamic nature. This research will also ensure a comprehensive understanding of urban landscapes in the face of diverse challenges particularly in developing countries.

Keywords: Urban Areas, Pakistan, GIS technology, redefining criteria.

INTRODUCTION

For any country's planning and development, the defined urban areas play a key role. These areas serve as economic engines, hubs of innovation, and centers of social and cultural activities (Yigitcanlar, 2009). The concentration of advanced infrastructure, abundant resources, and a diverse, high-tech workforce produce efficiency, enhanced connectivity, and multiple opportunities in any country (Komninos, 2011). This, in turn, makes a substantial and indispensable contribution to the overarching progress of the entire nation. Policy-makers and researchers rely on the delineation between urban and rural populations to analyze and comprehend regional disparities, identify specific needs, and tailor development interventions accordingly. In essence, the official urban definitions significantly shape policy decisions, with government strategies, including employment quotas, being influenced by these classifications (Ma and Su, 2024; Zhong, 2015). Moreover, research outcomes and socio-economic indicators heavily rely on accurate urban planning through GIS integration (Gonzalez *et al.*, 2011; Pham *et al.*, 2011), emphasizing the importance of in-depth understanding for effective urban planning and development. Ironically, the absence of a universally agreed-upon compilation of "urban challenges" is not exclusive to developing countries; it is also observed in developed nations. The definition of these challenges is contingent upon the perspectives and values held by various social groups and actors within the urban environment, regardless of the level of development (Ullah, 2022; Garau, 2009). Therefore, urban area definitions vary across countries, and the criteria used for classification may include population density, infrastructure, administrative boundaries, or multi-criteria. Population density serves as a cornerstone for many countries' urban planning. According to the U.S. Census Bureau, 2021.

The United States, for example, relies on population density and land use characteristics (Chen *et al.*, 2021). High population density is indicative of concentrated human activities, providing a foundational metric for recognizing urbanization. However, previously it is argued that a singular focus on population density may oversimplify urban complexity (Angel, 2012; Montgomery, 2008; Davis, 2006). On the other hand, the Office for National Statistics in the United Kingdom (ONS) incorporates administrative boundaries and land use planning in its urban area definitions, as highlighted in the publication, "Population Density: Urban and Rural Areas" (Pateman, 2011; Harrison and Hoyler, 2015). This approach acknowledges the role of physical structures and spatial organization in identifying urbanized regions, reflecting the importance of infrastructure as a defining criterion.Furthermore, China's urban area definition is closely tied to administrative boundaries(Ding et al., 2020; Yu et al., 2018). Similarly, the Statistical Handbook of Japan, published by the Ministry of Internal Affairs and Communications in 2021, serves as a primary source acknowledging Japan's urban planning based on administrative criteria (Statistics Bureau, 2021).

The handbook provides a comprehensive overview of the administrative considerations that influence the spatial organization of human settlements in the country. Some studies that are administrative-based, may also face some criticism for neglecting functional urban regions and lead to limitations in capturing the true extent of urban influence (Ferrão and Fernández, 2013). Furthermore, several scholars have pointed out inaccurate results for relying on administrative criteria only. Thus, this view emphasized the importance of considering the functional aspects of urban areas for further delineation of urban areas (Logan, 2011; Batty, 2013). Therefore, the European Union adopts a multi-criteria approach, considering factors such as population density, infrastructure, and administrative boundaries collectively (Hanlon *et al.*, 2009; Breckenkamp *et al.*, 2017; Zasada *et al.*, 2013; Moreno-Monroy *et al.*, 2021). This holistic approach aims to capture the complexity of urbanization and provide a comprehensive understanding of urban areas.

However, in the context of Pakistan, the criteria frequently utilized to distinguish between 'rural' and 'urban' regions in the previous census are predominantly based on administrative or municipal status (Ullah, 2022). Before 1981, any region with a population of 5,000 or above, as well as all municipal and town committees, were considered urban. However, since 1981, urban areas have been identified based on administrative criteria, specifically encompassing individuals residing in metropolitan and municipal corporations, municipal committees, and cantonments. The remaining population is classified as rural. Furthermore, according to Zaidi (2017), the officially designated boundaries of cities are labeled as 'rural,' despite being equally urbanized compared to the remainder of the city (Nasir and Khan, 2018).

In the last three decades, the urban population in Pakistan has tripled, and projections suggest that the country will predominantly be urban by 2025 within the South Asian region (Hussnain *et al.*, 2020). However, the previous census records in Pakistan have consistently classified urban and rural areas using criteria that lack clear discernibility (Ullah, 2022; Nasir and Khan, 2018). Clarity regarding discernibility can only be achieved through mapping, which was absent in previous studies. Therefore, the current study aims to address this gap by incorporating detailed mapping techniques to enhance the clarity of urban and rural categorizations, providing a more accurate representation of the evolving urban and rural landscape. This approach will contribute valuable insights into the ongoing urbanization trends and facilitate a comprehensive understanding of urban spatial planning.

MATERIAL AND METHODS

Study area

Pakistan will be the fourth most populous nation in the world by 2030 (Jan and Iqbal, 2018). The census 2017 data reveals that the urban population was around 37%, marking an increase from 33% in the 1998 census and indicating a consistent trend of rural-to-urban migration. However, the classification of rural and urban areas in Pakistan, along with the distinction between them, heavily relies on inconsistent and incomplete definitions, as highlighted earlier in a report published by National Institute of Urban Infrastructure Planning (Haider *et al.*, 2010). Therefore, this study adopts an exploratory data analytical approach to optimize urban delimitations in Pakistan by integrating Geographic Information System (GIS) Mapping techniques for spatial analysis and employs data science methodologies for quantitative assessment.

Criteria Identification

The foundational step in current research methodology involves a process of delineating and defining the key factors that will serve as the basis for identifying urban areas in Pakistan. In the context of Pakistan, the current study adopts a departure from the traditional administrative or municipal-based criteria that have historically distinguished between 'rural' and 'urban' regions, as highlighted by Ullah (2022). As, it is estimated that this urban proportion of the total population will significantly rise to over 50% by 2025 based on the 'administrative definition'(Abdul and Yu, 2020). In addition, continuous change in administrative boundaries remain yet another problem.

Addressing these challenges, the study aligns with the recommendations proposed by experts, particularly focusing on the need to adopt a multi-criteria definition of 'urban'. Thus, study presents the official definition of urban that should be based on a comprehensive analysis of socio-economic, demographic, and physical characteristics, rather than relying solely on administrative status. There are several factors to accurately determine the urban nature of an area but for the current study definition ambient population size and density, economic-Demographic aspects (Correia Filho *et al.*, 2022; Stark, 1978)as well as infrastructure data (Bobylev, 2008; Heshmati and Rashidghalam, 2020)have been selected.

Data collection

Data collection for the study includes multi-dimensional data incorporating spatial, demographic, economic, and infrastructure perspectives to redefine and optimize urban delimitations in the country. Each dataset is mentioned in detail below:

Spatial Data

Spatial data is an integral part of this research study. This type of data encompasses data obtained through open access data sources. These open data sources aid in capturing the geographical characteristics of urban areas with precision, allowing for detailed spatial analysis including population and land use patterns. This approach has been chosen for the past five years in comparable research studies recently conducted across the world (Kumar *et al.*, 2017; Aslam *et al.*, 2023). Therefore, in current time, GHSL-DUG technology is considered as crucial for delineating the arrangement of urban areas. This method is also aligned with recommendations from previous studies conducted in different parts of the world. Therefore, this approach is distinct from conventional approaches. However, it has not been extensively utilized until now and recommended to incorporate in Pakistan (Nasir and Khan, 2018).

Economic-Demographic Data

This section compiles the information related to the population from a census report. Understanding population dynamics is crucial for establishing criteria in defining urban areas, as it reflects the human dimension of urbanization and its impact on spatial boundaries. Thus, it is another important category to execute the analysis. For current study, tehsil level data have been incorporated to see the urban delineation through geospatial technology. Data was acquired from the Pakistan Bureau of Statistics, that is a federal agency under the Government of Pakistan.

Infrastructure Data

The required infrastructure data have been obtained from the spatial raster dataset. These datasets depicts built-up surfaces and its distribution in square meters. The data include the total built-up surface, Residential and non-residential surfaces. This data is available for spatial-temporal Data that is interpolated from 1975 to 2030 with 5 years intervals on different resolutions. For the current study following tiles of Global Human Settlement Layer- GHS-BUILT-S have been obtained for the year 2023 on Mollweide projection as mentioned in table 1 and shown in Figure 1.

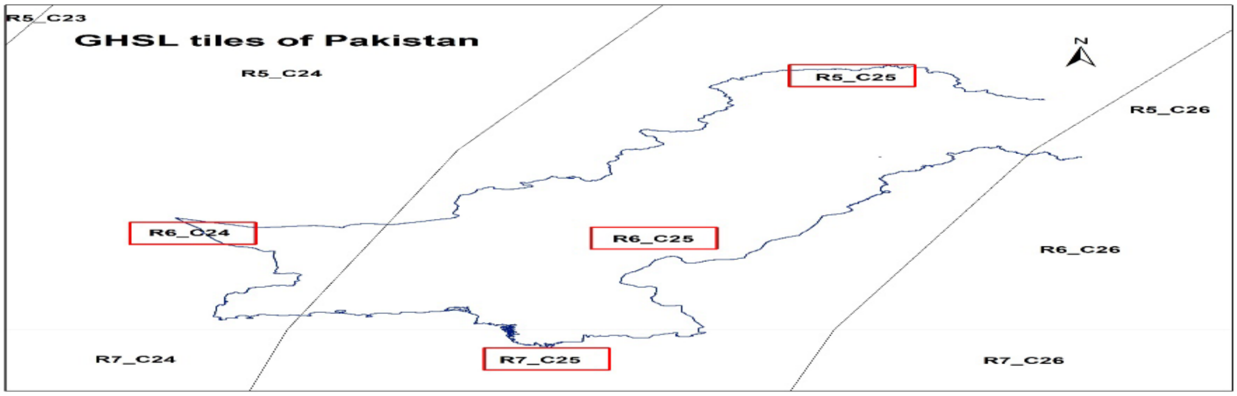

Figure 1.
Global Human Settlement Layer- GHS-BUILT-S tiles coverage of Pakistan

Table 1.
Specification of GHS-BUILT-S tiles of Pakistan

Tile no.	Resolution	Coverage Area
R6- C24	100 meters	South of Balochistan
R7- C25	100 meters	Rann of Kuch
R6- C25	100 meters	Central Pakistan
R5-C25	100 meters	Northern Areas of Pakistan

Exploratory data analysis (EDA) GIS integration

In the final step, an exploratory data analysis (EDA) approach is used for calculating the grid-based degree of urbanization. Descriptive statistics are calculated to characterize urbanization's degree within each grid or spatial unit at the district level in Pakistan. Spatial analysis further computes employing clustering analysis to identify patterns and spatial dependencies.

Following four GIS tools have been incorporated as mentioned below:

The Population To Grid (GHS-POP2G)

It is a versatile tool used for generating geospatial population grids in GeoTIFF format from vector census data, which can be in the form of polygons or points. It facilitates the conversion of census data into gridded format. It enables easier spatial analysis and visualization.

The Population Warping tool (GHS-POPWARP)

It is specifically designed to maintain the total volume of population data during the warping process. This functionality is an important step for preserving the integrity of population grids when undergoing reprojection. This step also ensures that the population distribution remains accurate despite changes in spatial reference.

The Degree of Urbanisation Grid (GHS-DUG)

It is adaptable GIS tool aimed at creating a geospatial settlement classification based on population grid data, according to the Degree of Urbanisation (DEGURBA). This tool provides insights into urban classes by categorizing areas based on their level of urban development.

The Degree of Urbanisation – Territorial units classifier (GHS-DU-TUC)

The GIS tool serves as a tool for classifying local units from a settlement classification grid, aligning them with the Degree of Urbanisation (DEGURBA). This classifier assists in identifying and categorizing different territorial units based on their urbanization status. Later, the cross-tabulation table is optionally computed statistics about the classified polygons grouped by the cross-tabulation land use patterns.

RESULTS

This section presents geospatial settlement classification grid. This classification system comprises two hierarchical levels. At the first level, there are three designated primary classes, including urban center grid cells, urban cluster grid cells, and rural grid cells. Level 2 classification refines the distinctions with eight classes including urban center grid cells, dense urban cluster grid cells, semi-dense urban cluster grid cells, suburban or peri-urban grid cells, rural cluster grid cells, low-density rural grid cells, very low-density rural grid cells, and water grid cells.

GHS Degree of Urbanization Level 1

The classification level 1 features three primary categories at the initial level, each offering a unique perspective on population distribution and spatial characteristics. The first category, comprising rural areas, accounts for 12.66% of the total population, with 25,095,132 individuals spread over 2,314,566 square kilometers. The second class, representing "Towns and semi-dense areas," boasts a substantial population share of 50.67%, accommodating 100,452,408 people within an area of 112,488 square kilometers. Lastly, the third class, designated as "Cities," encompasses 72,682,080 individuals across 14,346 square kilometers, contributing to a population share of 36.67% as shown in table 1.

Table 1.
Results of geospatial settlement classification grid level 1

L1_class	Areas	Population_count	Area_km2	Population_share
1	Rural areas	25095132	2314566	12.65963
2	Towns and semi-dense areas	100452408	112488	50.67477
3	Cities	72682080	14346	36.6656

GHS Degree of Urbanization Level 2

The second level of classification further refines the analysis, encompassing eight distinct classes with varying population densities and spatial characteristics. From "Water grid cells," representing uninhabited areas spanning 168,018 square kilometers, to "Urban center grid cells," housing the largest population of 72,682,080 individuals over 14,346 square kilometers (36.67% of the total), the classification offers a detailed breakdown. Results revealed that "Suburban or peri-urban grid cells" emerges as the highest populated and significant category with 76,950,784 individuals and covering 98,482 square kilometers, representing approximately 38.82% of the total population. This is followed by "Urban center grid cells" where 36.67% of population is residing in Pakistan.

Table 2.

Results of geospatial settlement classification grid level 2

L2_class	Description	Population	Area_km2	%
10	Water grid cells	0	168018	
11	Very low-density rural grid cells	1650686.75	2031222	0.832714
12	Low-density rural grid cells	16287793	104411	8.216628
13	Rural cluster grid cells	7156652	10915	3.610284
21	Suburban or peri-urban grid cells	76950784	98482	38.81901
22	Semi-dense urban cluster grid cells	5729521.5	7947	2.890346
23	Dense urban cluster grid cells	17772106	6059	8.965413
30	Urban center grid cells	72682080	14346	36.6656

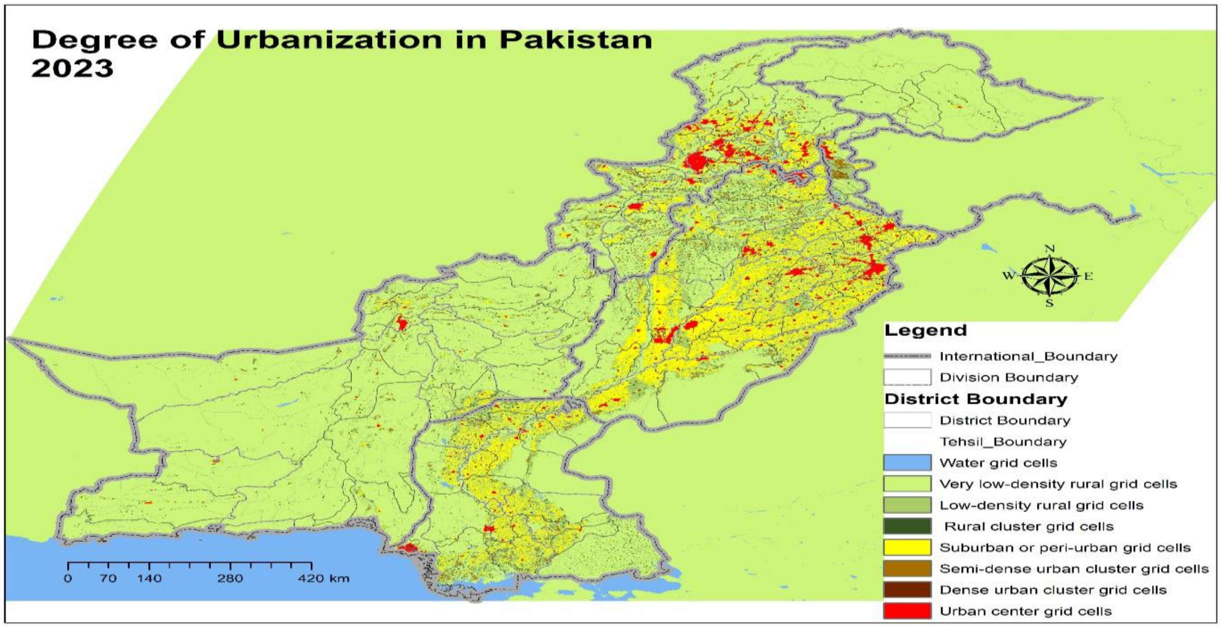

Figure 2.

Degree of urbanization- DEGURBA in Pakistan (2023)

Figure 2 is showing that the degree of urbanization classified on basis of local administrative units (LAUs) in eight distinctive classes. These classes are based on cell based geographical contiguity and population size, measured by minimum population thresholds applied in each LAU belongs exclusively to one of these eight classes. For example, all red color grid cells are showing an urban centre that consists of contiguous (using four-point contiguity) grid cells with a density of at least 1,500 inhabitants per km2. These urban centres are having at least 50,000 population with at least 50 % built-up.

CONCLUSION AND DISCUSSION

This research study presents the **Urban Delimitations in Pakistan**, adopting a two-level approach, delineates urban and rural areas. As proposed in previous studies ((Maffenini *et al.*, 2020). Current study showed that the first level categorizes regions into rural areas, towns and semi-dense areas, and cities, contributing 12.66%, 50.67%, and 36.67% to the overall population, respectively. Moving to the second level, a refinement into eight categories reveals substantial variations in population densities and spatial characteristics, aligning with findings in urbanization studies (McGranahan *et al.*, 2023).

Noteworthy is the prominence of "Suburban or peri-urban grid cells" as the most populated class at 38.82%, as corroborated by recent research on urban dynamics (Jan and Iqbal, 2018). The study's data-driven approach to redefine urban areas integrates GIS, a methodology supported by previous works on urban delineation (Aslam, *et al.*, 2023). By considering factors such as population density, infrastructure, and economic indicators, the research aligns with recommendations for a comprehensive understanding of urban landscapes (Haider, *et al.*, 2010). This contributes to urban planning and policy development, offering a more nuanced representation of Pakistan's evolving urban dynamics, as suggested by scholars in the field (Nasir and Khan, 2018). The proposed redefinition, addressing the dynamic nature of urbanization, provides valuable insights for decision-making and interventions in the context of contemporary challenges, echoing sentiments from urban studies (Garau, 2009). By aligning urban development efforts with the objectives of SDG 11 through an exploratory data analytical approach, Pakistan can make significant strides towards creating inclusive, safe, resilient, and sustainable cities by the year 2030.

DECLARATIONS

Acknowledgment: We appreciate the generous support from all the supervisors and their different affiliations.

Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.

Availability of data and material: In the approach, the data sources for the variables are stated.

Authors' contributions: Each author participated equally in the creation of this work.

Conflicts of Interests: The authors declare no conflict of interest.

Consent to Participate: Yes

Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCE

Abdul, L., & Yu, T.-f. J. U. S. (2020). Resilient urbanization: a systematic review on urban discourse in Pakistan. 4(4), 76.

Angel, S. (2012). *Planet of cities*: Lincoln Institute of Land Policy Cambridge, MA.

Aslam, R. W., Shu, H., & Yaseen, A. (2023). Monitoring the population change and urban growth of four major Pakistan cities through spatial analysis of open source data. *Annals of GIS,* 29(3), 355-367. doi: 10.1080/19475683.2023.2166989

Batty, M. (2013). *The new science of cities*: MIT press.

Bobylev, N. (2008). *Urbanization and environmental security: infrastructure development, environmental indicators, and sustainability*. Paper presented at the Environmental Change and Human Security: Recognizing and Acting on Hazard Impacts.

Breckenkamp, J., Patterson, L., Scharlach, M., Hellmeier, W., & Verma, A. J. T. E. J. o. P. H. (2017). Definitions of urban areas feasible for examining urban health in the European Union. 27(suppl_2), 19-24.

Chen, B., Tu, Y., Song, Y., Theobald, D. M., Zhang, T., Ren, Z., . . . Sensing, R. (2021). Mapping essential urban land use categories with open big data: Results for five metropolitan areas in the United States of America. 178, 203-218.

Correia Filho, W. L. F., de Oliveira-Júnior, J. F., dos Santos, C. T. B., Batista, B. A., de Barros Santiago, D., da Silva Junior, C. A., . . . Environment. (2022). The influence of urban expansion in the socio-economic, demographic, and environmental indicators in the City of Arapiraca-Alagoas, Brazil. 25, 100662.

Davis, M. J. O. H. I. (2006). Planet of slums. 8(5).

Ding, C., Zhao, B. J. C. P., Northern, E. o., & Asia, W. (2020). Urbanization in China: History, development and challenges. 29(2/3), 305-346.

Ferrão, P., & Fernández, J. E. (2013). *Sustainable urban metabolism*: MIT press.

Garau, P. J. R. o. a. R. d. h. g. g. o. m. (2009). Revisiting urban planning in developed countries.

Gonzalez, A., Donnelly, A., Jones, M., Klostermann, J., Groot, A., Breil, M. J. J. o. E. A. P., & Management. (2011). Community of practice approach to developing urban sustainability indicators. 13(04), 591-617.

Haider, M., Badami, M. G. J. E., & ASIA, u. (2010). Urbanization and local governance challenges in Pakistan. 1(1), 81-96.

Hanlon, B., Short, J. R., & Vicino, T. (2009). *Cities and suburbs: New metropolitan realities in the US*: Routledge.

Harrison, J., & Hoyler, M. (2015). *Megaregions: Globalization s New Urban Form?* : Edward Elgar Publishing.

Heshmati, A., & Rashidghalam, M. J. J. o. I. S. (2020). Measurement and analysis of urban infrastructure and its effects on urbanization in China. 26(1), 04019030.

Hussnain, M. Q. u., Waheed, A., Wakil, K., Pettit, C. J., Hussain, E., Naeem, M. A., & Anjum, G. A. (2020). Shaping up the Future Spatial Plans for Urban Areas in Pakistan. 12(10), 4216.

Jan, B., & Iqbal, M. J. S. J. o. A. (2018). Urbanization trend and urban population projections of Pakistan using weighted approach. 24(1).

Komninos, N. J. I. B. I. (2011). Intelligent cities: Variable geometries of spatial intelligence. 3(3), 172-188.

Kumar, P., Sajjad, H., Alare, R. S., Elvidge, C. D., Ahmed, R., & Mandal, V. P. (2017). Analysis of Urban Population Dynamics Based on Residential Buildings Volume in Six Provinces of Pakistan Using Operational Linescan System Sensors. *IEEE Sensors Journal,* 17(6), 1656-1662. doi: 10.1109/JSEN.2017.2652720

Logan, J. (2011). *The new Chinese city: globalization and market reform*: John Wiley & Sons.

Ma, X., & Su, W. J. T. M. (2024). Local government intervention in tourism-driven rural gentrification: Types and interpretative framework. 100, 104828.

Maffenini, L., Schiavina, M., Melchiorri, M., Pesaresi, M., & Kemper, T. (2020). GHS-DUG User Guide: Luxembourg: Publications Office of the European Union. https://doi. org/10 ….

McGranahan, G., Balk, D., Colenbrander, S., Engin, H., MacManus, K. J. E., & Urbanization. (2023). Is rapid urbanization of low-elevation deltas undermining adaptation to climate change? A global review. 35(2), 527-559.

Montgomery, M. R. J. s. (2008). The urban transformation of the developing world. 319(5864), 761-764.

Moreno-Monroy, A. I., Schiavina, M., & Veneri, P. J. J. o. U. E. (2021). Metropolitan areas in the world. Delineation and population trends. 125, 103242.

Nasir, J., & Khan, M. J. U. (2018). Redefining Urban; A Case for Multi-Criteria Definition of Urban Areas in Pakistan.

Pateman, T. J. R. t. (2011). Rural and urban areas: comparing lives using rural/urban classifications. 43, 11-86.

Pham, H. M., Yamaguchi, Y., Bui, T. Q. J. L., & Planning, U. (2011). A case study on the relation between city planning and urban growth using remote sensing and spatial metrics. 100(3), 223-230.

Stark, O. (1978). *Economic-demographic interactions in agricultural development: The case of rural-to-urban migration* (Vol. 6): Food & Agriculture Org.

Statistics Bureau, J. (2021). Statistical Handbook of Japan.

Ullah, I. J. E. e. E. R. b. d. g. e. (2022). Re-identifying the Rural/Urban: A case Study of Pakistan.

Yigitcanlar, T. J. J. o. K. M. (2009). Planning for knowledge-based urban development: global perspectives. 13(5), 228-242.

Yu, H., Deng, Y., & Xu, S. J. C. G. S. (2018). Evolutionary pattern and effect of administrative division adjustment during urbanization of china: empirical analysis on multiple scales. 28, 758-772.

Zasada, I., Loibl, W., Berges, R., Steinnocher, K., Köstl, M., Piorr, A., . . . Europe, m. f. l. u. c. i. (2013). Rural–urban regions: A spatial approach to define urban–rural relationships in Europe. 45-68.

Zhong, Y. (2015). *Local government and politics in China: Challenges from below*: Routledge.