

Article

Urban Big Data Analytics: A Novel Approach for Tracking Urbanization Trends in Sri Lanka

Nimesh Akalanka ¹, Nayomi Kankanamge ¹, Jagath Munasinghe ¹ and Tan Yigitcanlar ^{2,*}

¹ Department of Town and Country Planning, University of Moratuwa, Katubedda, Moratuwa 10400, Sri Lanka; 1823501@uom.lk (N.A.); nayomie@uom.lk (N.K.); jagathnm@uom.lk (J.M.)

² City 4.0 Lab, School of Architecture and Built Environment, Queensland University of Technology, 2 George Street, Brisbane, QLD 4000, Australia

* Correspondence: tan.yigitcanlar@qut.edu.au; Tel.: +61-7-3138-2418

Abstract: The dynamic nature of urbanization calls for more frequently updated and more reliable datasets than conventional methods, in order to comprehend it for planning purposes. The current widely used methods to study urbanization heavily depend on shifts in residential populations and building densities, the data of which are static and do not necessarily capture the dynamic nature of urbanization. This is a particularly the case with low- and middle-income nations, where, according to the United Nations, urbanization is mostly being experienced in this century. This study aims to develop a more effective approach to comprehending urbanization patterns through big data fusion, using multiple data sources that provide more reliable information on urban activities. The study uses five open data sources: national polar-orbiting partnership/visible infrared imaging radiometer suite night-time light images; point of interest data; mobile network coverage data; road network coverage data; normalized difference vegetation index data; and the Python programming language. The findings challenge the currently dominant census data and statistics-based understanding of Sri Lanka's urbanization patterns that are either underestimated or overestimated. The proposed approach offers a more reliable and accurate alternative for authorities and planners in determining urbanization patterns and urban footprints.

Keywords: urbanization patterns; urbanization dynamics; urban analytics; urban informatics; big data; urban big data; data fusion; Sri Lanka



Citation: Akalanka, N.; Kankanamge, N.; Munasinghe, J.; Yigitcanlar, T. Urban Big Data Analytics: A Novel Approach for Tracking Urbanization Trends in Sri Lanka. *Land* **2024**, *13*, 888. <https://doi.org/10.3390/land13060888>

Academic Editor: Adrianos Retalis

Received: 19 April 2024

Revised: 12 June 2024

Accepted: 13 June 2024

Published: 19 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Urbanization is an ongoing process which is conventionally understood as referring to the influx of populations towards urban centers from surrounding rural areas. In the present context, the magnitude of urbanization around the globe is generally discussed with the total number of megacities, where populations are more than 10 million, increasing from 23 million to 31 million over the last decade and projected to be 41 by 2030 [1]. In addition to this, a large number of new urban centers are also expected to emerge to accommodate this influx of populations. However, this ‘urban nuclei’ centered, socio-demographic, and statistics-based view is less likely to provide a comprehensive picture of the state and trends of urbanization processes.

Population statistics mostly come from surveys conducted once every 2–10 years based on countries’ census year policy [2,3]. The conventional methods used by most countries consider residential population distribution as the only or the main criterion for the identification of urban areas and the patterns of urbanization [4,5]. Yet, these methods inherit a few limitations with them. First, they are time-consuming and incur reasonable costs, requiring lots of human resources involved in them, and therefore, limiting the ability to update frequently [6]. Second, the effects of urbanization and urban activities are essentially not limited to residential populations, as commuter populations too have sizeable impacts on urban functions. Third, the delineation of urban areas is challenging

due to the absence of a universally accepted definition and methodologies. Above all these, there is a parallel process of ‘hidden urbanization’ that is not necessarily reflected in the form of conventional statistics. The increased mobility, use of advanced technology, and changing aspirations of people inevitably change their lifestyles, irrespective of the place of their residence [7]. Thus, urbanization needs to be understood more as a ‘way of life’ [8] than a matter of the movement of populations.

This complex pattern of urbanization will have a profound impact on the spatial characteristics of cities in future [9–11]. Under such circumstances, managing urban areas and providing their inhabitants with a pleasant urban life is becoming increasingly challenging for local governments. For national and provincial governments, the fair allocation of necessary resources is equally challenging, for which a proper understanding of the state of urban areas is crucial, particularly when it comes to sustainable urban development [12,13].

As discussed, it is difficult to identify urban areas and urbanization patterns by purely relying on survey and census data, which only provide information for a specific point in time [14–16]. Still, the identification of urban areas with some level of precision is important for any country or state with multiple functions such as governance, planning and implementation, and resource allocation, and more importantly, for formulating national- and regional-level spatial policies to guide future urban development scenarios [17–19]. Furthermore, the identification of the magnitude and nature of the complex and hidden urbanization in low- and middle-income countries is crucial for formulating appropriate development strategies to assure sustainable urban futures for these countries [20,21]. The use of alternative data sources to identify urban areas and their growth patterns is worth considering in this big data era, where data related to living things, devices, machines, and almost all objects can be traced and recorded with convenience and on a daily basis [22]. Nonetheless, utilizing open-source big data sources to study urban phenomena is not yet popular [23].

Against this backdrop, this paper is focused on a more effective and novel approach to determining the patterns of urbanization with the fusion of big data with other multiple datasets obtained from reliable sources of information on urbanization-related activities. Such an approach is particularly critical for low and middle-income countries, as towns in these nations mostly reflect complex socio-economic and land use dynamics. For instance, informal urban spaces, small boutiques, and third places at the junctions of neighborhoods mostly provide demanded services for local neighborhoods, unlike high-income countries’ shopping malls and urban squares [24]. However, local community demands can be well met in such townships in low- and middle-income countries [17,23,25,26]. This study uses Sri Lanka as a testbed as a low and middle-income country context example to demonstrate the approach proposed. For the study, multiple openly available data sources, which reflect urban lifestyles and urban facilities in urban areas, instead of population statistics as a single dataset are used. The developed approach is straightforward, user-friendly, and cost-effective and will help to provide a better understanding about urbanization patterns and processes for authorities and planners.

2. Literature Background

2.1. Definitions of Urban, Urbanization and Urban Dynamics

The terms “Urban”, “Urbanization”, and “Urban dynamics” mostly go hand in hand. “Urban” refers to the characteristics and attributes associated with city environments, including a high population density and extensive infrastructure [1,27]. “Urbanization”, on the other hand, describes the process by which rural areas transform into urban areas, marked by population migration and city expansion. “Urban dynamics” describes the development and change patterns and processes that occur in urban settings. This encompasses the interplay among social, economic, and environmental elements that impact the development, configuration, and operations of urban areas. Urban dynamics is a broad term that includes a variety of phenomena, including changes in the population, the economy, land use, infrastructure, and social trends. Comprehending urban dynamics is

imperative for proficient urban planning and administration, as it facilitates the anticipation and resolution of issues pertaining to urbanization, sustainability, and the standard of living in urban areas [28].

‘Urban’ may not be a universally defined term and, therefore, the delineation of ‘urban areas’ also lacks a commonly agreed criterion [1,3,27]. In the absence of a widely agreed definition or methodology to identify urban areas, each country has adopted its own criterion to identify urban areas and urban shares of populations. Such criteria can be categorized into four main groups, based on the core units they consider for this purpose: (a) population based; (b) administrative boundary based; (c) land use based; and (d) multicriteria based.

Out of them, demographic characteristics and land uses are the proxies most adopted to delineate urban areas [29]. However, urbanization is a more complex phenomenon than what has been captured with head counts and land uses [30], and must account for a combination of the economic, demographic, social, and technological processes that lead to an increasing share of populations embracing urban characteristics [8,31]. Louis Worth (1938) suggested that “urbanism is a way of life”, denoting the process by which an increasing number of individuals embrace urban lifestyles and linking urbanization to the transition from rural or less populated regions to urban environments, impacting social attitudes, lifestyles, and the overall organization of society [8].

Sri Lanka has adopted a local government institutions-based approach to delineating urban areas, in which municipal councils (MCs) and urban councils (UCs) are classified as urban areas, while the rest of the Pradeshiya Sabhas are considered as ‘rural’ areas [31]. India considers Municipal Corporation Areas as urban areas. Still, to be considered as urban, Indian Municipal Corporation Areas must have a population over 5000, a population density over 400 people per km², and 75% of the male workforce engaged in non-agricultural activities [5].

Bangladesh follows a more comprehensive definition compared to Sri Lanka and India. It defines areas as urban if the respective area is densely populated, developed around a central place with the necessary infrastructure, and most of the population is engaged in the non-agricultural sector [29,32]. Areas that are densely developed and have commercial, residential, and other non-residential land uses are considered as urban in the US [33].

Two basic drawbacks were observed in the aforementioned approaches. The first is that most of these definitions rely on statistical figures such as the population density and type of economic activities, and the second is the use of administrative boundaries, which may not necessarily trace the complex behavioral characteristics and dynamic nature of urban areas and their functions. Hence, research on urbanization has to be based on more sophisticated and complex methods which are competent enough to capture the lifestyles and spatial practices that propel urban dynamics, rather than static spatial elements and periodic statistics.

2.2. Contemporary Tools and Techniques for Understanding Urbanization Patterns

The literature revealed some attempts to adopt different approaches to identify the dynamics of urban areas and the patterns of their evolution. Remote sensing and geographic information system (GIS) technologies have been widely used in this regard, maybe due to their conveniently acquired and reliable data [34–36]. Remote sensing data with a medium to high resolution i.e., Landsat, SPOT, AVHRR, Quick Bird, IKONOS, WorldView, and Sentinel, are commonly used in urban monitoring and detection satellites [37,38]. There, urban areas are identified at the pixel level.

Nevertheless, the areas covered by such medium- and high-resolution imageries are very small when considering the regional and the national scales [39]. In addition to this, state organizations seem to be reluctant to accept satellite-image-based methods to identify urban areas, maybe because of their cost and the high-end technological competencies required for their use [40]. Therefore, it is a challenging task to frequently use satellite imagery to study urban areas and urbanization patterns at the regional and national scales.

Yet, the virtues of satellite imagery in understanding urbanization patterns cannot be underestimated.

Based on satellite bands, different urban indices have been used by contemporary research studies to identify urban areas. Among them, the Normalized Difference Built-Up Index (NDBI) and Normalized Differentiate Vegetation Index (NDVI) are the most used. Still, these indices simply consider the spatial distribution of building footprints, which may lead to many misidentifications such as the classification of abandoned built-up areas as urban. The Human Settlement Composite Index (HSCI) normalizes Nighttime Light (NTL) satellite images with the MODIS Normalized Difference Vegetation Index (NDVI) [41,42]. The Vegetation-adjusted NTL urban index (VANUI) and enhanced vegetation-adjusted NTL index (EVANTLI) are based on the conceptualization that vegetation and urban built-up areas are inversely related [42,43]. The use of NTL images can be identified as an emerging approach to identifying functional urban areas, as it differentiates functioning buildings from non-functioning buildings based on Digital Numbers (DN).

Apart from remote sensing and GIS-based approaches, the use of urban big data is an emerging but understudied resource that can be used in urban research. Urban big data refers to the large and complex datasets generated by urban environments, and, as a result, these data can be used for numerous applications [44]. Urban big data can be used to gain insights into a wide range of subject areas such as urban area extraction [45], urban traffic prediction [46], urban air pollution [47], urban disaster management [48,49], and so on. However, they have not yet been widely used to study the urbanization processes of countries.

2.3. Urban Big Data and Urbanization

Urban big data are massive amounts of dynamic and static data generated from the subjects and objects, including various urban facilities, organizations, and individuals, which have been collected and collated by city governments, public institutions, enterprises, and voluntary individuals using new-generation information technology [44,50,51]. Urban big data have been used for a wide range of purposes such as to study the urban spatial structure and function division [16,52,53], landscape analysis and design [54,55], and the connection intensity between cities [22,56,57].

Refs. [58–60] used urban big data to extract urban areas and then to understand the temporal variations in the urban spatial structure. Still, most studies limit their investigations to one or two data sources. As urbanization is a dynamic and complex process, such complexities may not be captured with a single or limited number of data types. An analysis of urban big data of multiple types from a variety of sources is the only way to properly interpret urbanization based on the lifestyles of people, as elucidated by [8] in his seminal work on ‘urbanism as a way of life’.

3. Materials and Methods

3.1. Case Study

Sri Lanka, as a middle-income country, forms an ideal testbed for the proposed novel approach to determining urbanization patterns with urban big data fusion for the following reasons:

First, according to the widely accepted classification by international organizations (e.g., the UN and World Bank), Sri Lanka is a middle-income country with a notably high urbanization rate [61]. However, a puzzling trend in its urbanization pattern has been observed during the past decades. For instance, according to official statistics, Sri Lanka’s urban population dropped from 21.5% in 1981 to 14.6% by 2001, and marginally increased up to 18.2% by 2012. This sudden drop in urbanization level, which was quite inconsistent with the ground reality [62], was a result of a change of the definition of ‘urban areas’ used by the Department of Census and Statistics (DCS).

Second, the ‘urban’ areas considered for national census before 1987 included municipal councils (MCs) and urban councils (UCs) in addition to town councils (TCs), which

were abolished in 1987 and merged with village councils to form a set of new units called 'Pradeshiya Sabha', and these are considered as 'rural' by the DCS [20]. The distribution of urban areas as per the DCS definition (MCs and UCs) is given in Figure 1.

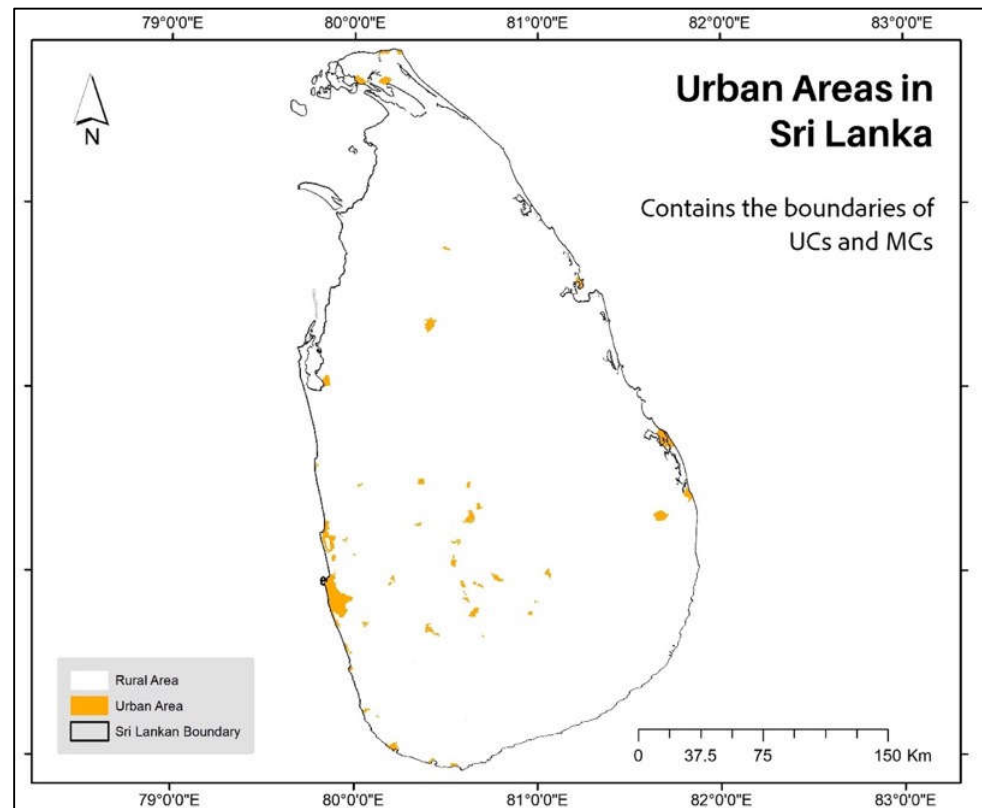


Figure 1. Officially identified urban local government areas in Sri Lanka.

Third, this local-government-type-based urban area classification, however, is misleading and has caused several repercussions in the areas' policy decisions, prioritization of urban investments, planning of future development activities, and so on. In this context, more revealing and vividly descriptive methods are essential to better comprehend the complexities in capturing the urbanization patterns and urban footprint in Sri Lanka.

3.2. Methodology

Having realized the complexity demanded by this task, this study adopted the big data fusion methodology to identify the urbanization patterns in the testbed of a low- or middle-income country context i.e., Sri Lanka. The Python programming language is used to handle large-sized nighttime light (NTL) images and for the segmentation of urban areas. A raw dataset of about 65 GB for the years of 2013, 2017, and 2021 was collected, analyzed, and fused through the Geometric Mean fusion technique. The Adaptive Threshold Segmentation algorithm was used as the urban area extraction method. These methodological steps are elaborated as follows. As given in Figure 2, the methodological framework consists of six main phases. They are: (a) data acquisition; (b) data pre-processing; (c) pre-data fusion; (d) data fusion; (e) urban area extraction; and (f) accuracy assessment. The detailed methodological framework is given in Figure 2.

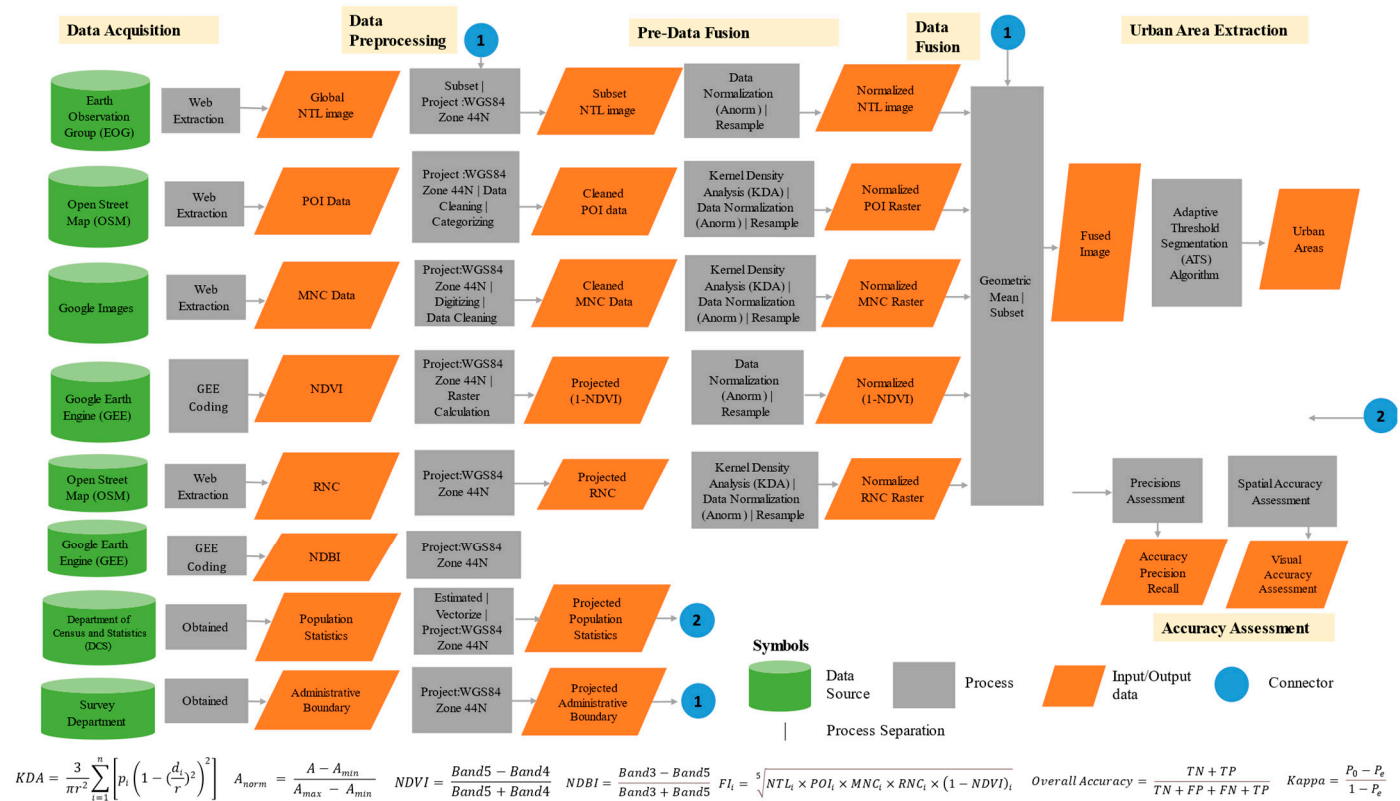


Figure 2. Methodological framework.

3.2.1. Data Acquisition

Table 1 shows the datasets used to examine Sri Lanka's urbanisation. Accordingly, NTL images, point of interest (POI) data, mobile network coverage (MNC) data, NDVI data, and road network coverage (RNC) data were utilized for the data fusion. In line with Louis Wirth's "Urbanism as a way of life", this study intended to capture information that reflected the different aspects of urban lifestyles.

Table 1. Datasets used to examine Sri Lanka's urbanization.

Dataset	Description	Time	Resolution	Data Sources
NTL satellite images	NPP/VIIRS Visible Light data annual average composite	2013 2017 2021	15 arc seconds (~500 m)	National Aeronautics and Space Administration (NASA)
POI data	Point data with latitude, longitude, and type of location		-	Open Street Map (OSM)
MNC data	Mobile 4G network coverage spatial information		-	Local Mobile Network Service Provider
NDVI	NDVI annual average composite prepared using Landsat 8 bands		NIR 30 m Red 30 m	United States Geological Survey (USGS)
RNC	All types of roads, including major, minor, and highway roads		-	Open Street Map (OSM)
NDBI	NDVI is an annual composite of the Landsat 8		Green 30 m NIR 30 m	United States Geological Survey (USGS)

Table 1. Cont.

Dataset	Description	Time	Resolution	Data Sources
Population Statistics	GN wise population data for Sri Lanka	2001 * 2012 * 2013 2017 2021	-	DCS
Administrative Boundary	Administrative boundary layer containing the MC, UC, and PS boundaries	2010	-	Survey Department

In this table census years are denoted using “*” mark.

The included datasets on point of interest (POI), Normalized Difference Vegetation Index (NDVI), road network coverage (RNC), mobile network coverage (MNC), and Nighttime Light (NTL) data are instrumental for understanding the heterogeneity of urban spaces, reflecting the diversity of services, amenities, and environmental features that characterize urban life. Meanwhile, MNC data, by detailing the distribution and intensity of mobile network usage, offer insights into the anonymity and impersonality prevalent in urban settings, highlighting the non-physical interactions that define modern urban life. Additionally, NTL data provide a unique perspective on urban areas’ vibrancy and activity levels, indirectly showcasing the diversity and intensity of urban existence. These datasets, carefully chosen for this study, enable a multifaceted analysis of urbanization, convincingly extrapolating critical aspects of urban life and providing a comprehensive understanding of its complexity.

These datasets are particularly effective for capturing the unique urbanization patterns in Sri Lanka. NTL images from NPP/VIIRS reveal active urban areas and infrastructure growth, crucial for identifying urban expansion in densely populated regions [27]. POI data from OpenStreetMap highlight the distribution of essential services and amenities, reflecting their functional diversity [63]. MNC data from local providers are used to identify the urbanization pattern with a higher accuracy [64]. RNC data map the extensive road infrastructure, critical for analyzing connectivity and accessibility, which are indicators of the growth of urban areas, especially in low- and middle-income countries [65]. Finally, NDVI data from Landsat 8 provide insights into land use changes and the environmental impacts of urbanization [66]. These datasets are important as they are cost effective and readily accessible for urbanization mapping. Further, the aforementioned data sources make them ideal for a comprehensive, accurate, and economically feasible study of urbanization in any country, ensuring that the findings are relevant and actionable for local urban planning and development.

Nighttime Light Satellite Images

Compared to the satellite images of Landsat, SPOT, AVHRR, and Quick Bird, NTL images help to identify functioning buildings [67]. NTL images can be obtained as Visible Infrared Imaging Radiometer Suite (VIIRS) NTL images from the National Aeronautics and Space Administration (NASA) and Defense Meteorological Satellite Program Operational Linescan System (DMSP/OLS). NTL images can also be acquired from the National Oceanic and Atmospheric Administration (NOAA). The NTL image dataset used in this study was derived from the National Polar-orbiting Partnership (NPP)/VIIRS satellite. To understand urbanization trends, an NPP/VIIRS annual nighttime light (VNL) 2.1 average composite dataset was downloaded for the years of 2013, 2017, and 2021. The composite NTL images had a spatial resolution of 15 arc seconds (~500 m). The spatial resolution of NPP/VIIRS is two times higher than that of DMPS/OLS. The NPP/VIIRS NTL images also had a higher radiometric resolution of 14-bit compared to the 6-bit resolution of DMPS/OLS. As a result of having a higher radiometric resolution, NPP/VIIRS managed to reduce the saturation

issues that exist with DMPS/OLS NTL images with a 6-bit resolution [68]. Figure 3 shows the NTL images of Sri Lanka for 2013, 2017, and 2021.

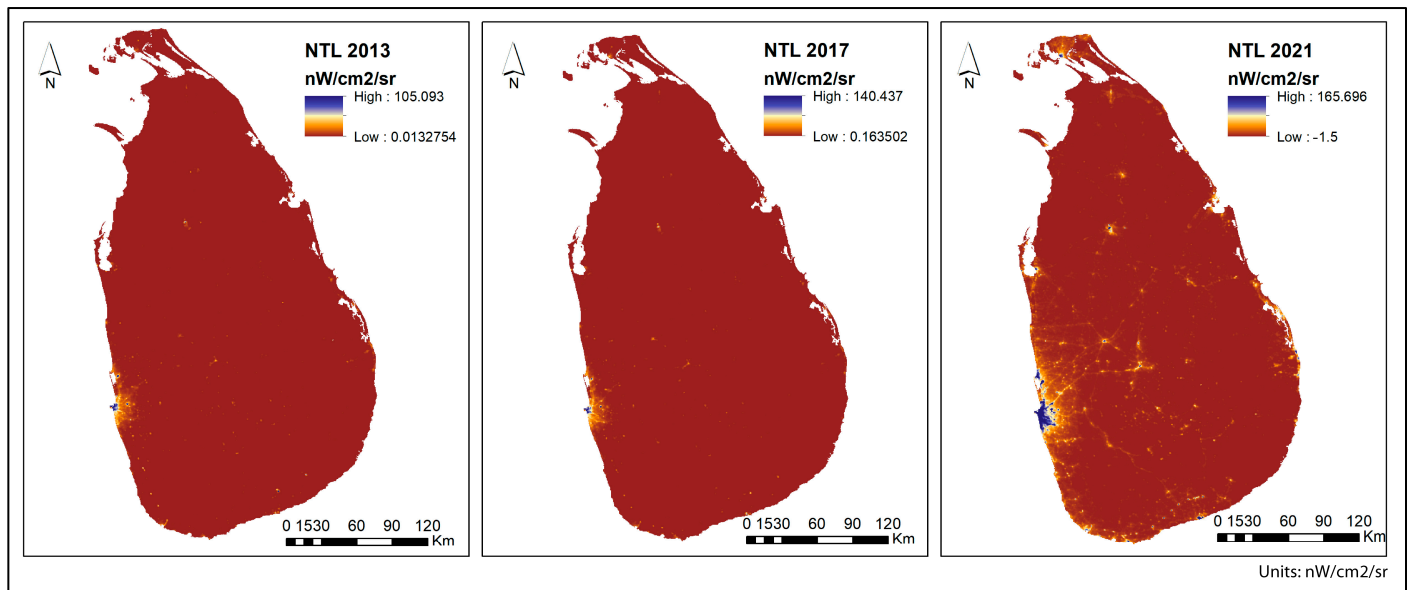


Figure 3. Images of Sri Lanka for 2013, 2017, and 2021.

Point of Interest (POI) Data

POI data refer to the locations of all the places within an urban area, which were obtained from OSM on 4 February 2023. The coordinates, category, and name of the place are included in the POI dataset. The POI data included data about public places, education, health, leisure, catering, accommodation, shopping, financial, tourism, and other locations within Sri Lanka. A total of 37,822 POI data (2013 = 4178; 2017 = 13,624; and 2021 = 20,020) were initially collected. Figure 4 shows the spatial distribution of the POI data for 2013, 2017, and 2021.

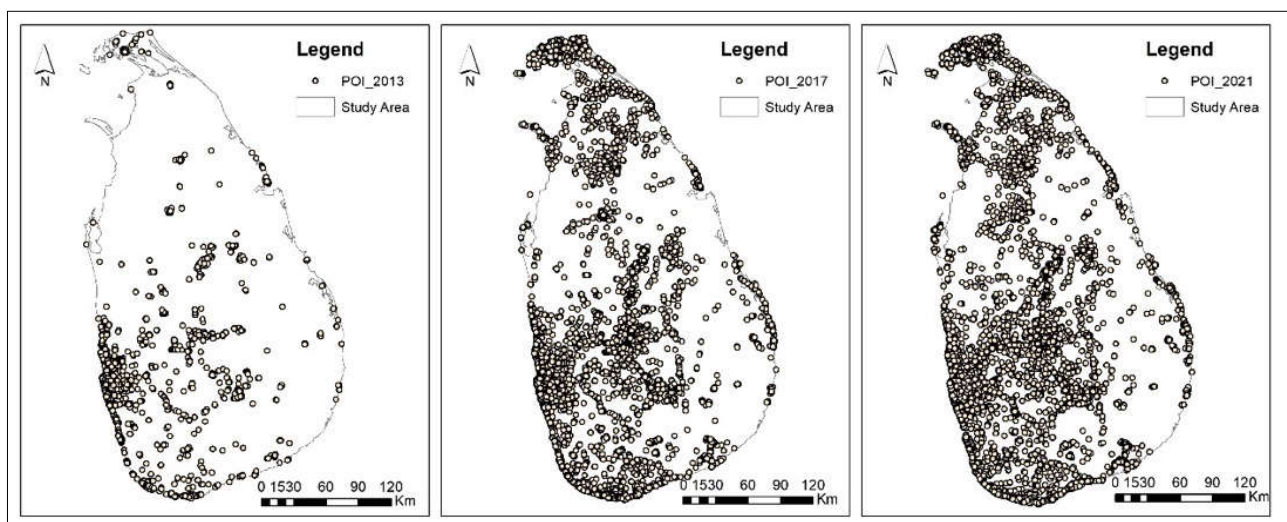


Figure 4. Spatial distribution of the POI data for 2013, 2017, and 2021.

Mobile Network Coverage Data

Mobile network coverage maps were developed considering the openly available data published by different mobile service providers in Sri Lanka. Figure 5 shows the mobile network coverage maps for Sri Lanka for 2013, 2017, and 2021.

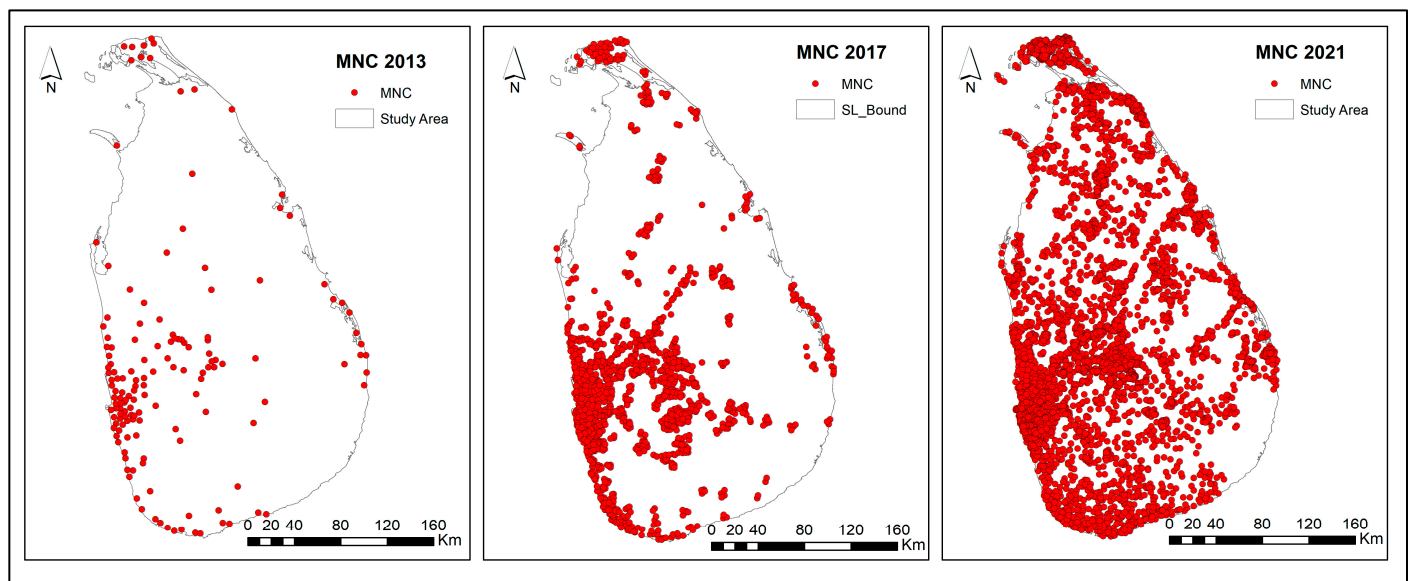


Figure 5. Mobile network coverage map for Sri Lanka for 2013, 2017, and 2021.

Normalized Difference Vegetation Index Data

The NDVI was calculated using Landsat 8 Collection 2 Tier 1 images. The NDVI dataset was composite considering all the available images within the year and obtaining the average value for the considered. All water bodies were masked out in calculating the NDVI. The equation used to calculate NDVI is given in Equation (1).

$$\text{NDVI} = (\text{Band5} - \text{Band4}) / (\text{Band5} + \text{Band4}) \quad (1)$$

The following maps in Figure 6 show the NDVI images for Sri Lanka for the years 2013, 2017, and 2021.

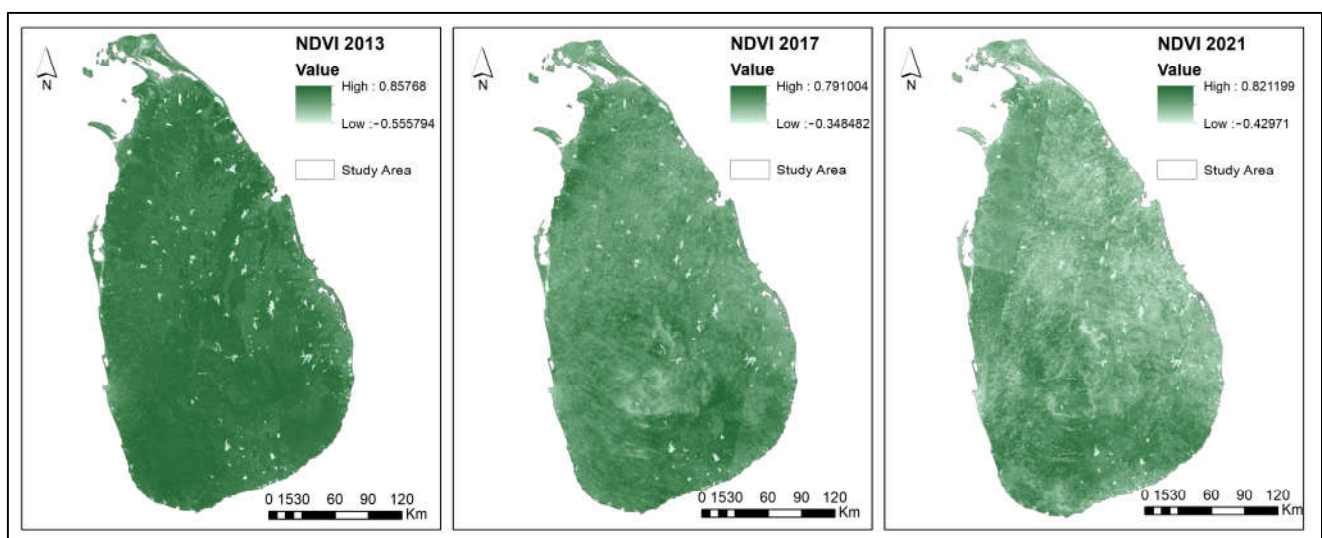


Figure 6. NDVI maps for 2013, 2017, and 2021.

Road Network Coverage Data

RNC data for the whole of Sri Lanka were extracted from the OSM for the years of 2013, 2017, and 2021 and are given in Figure 7.

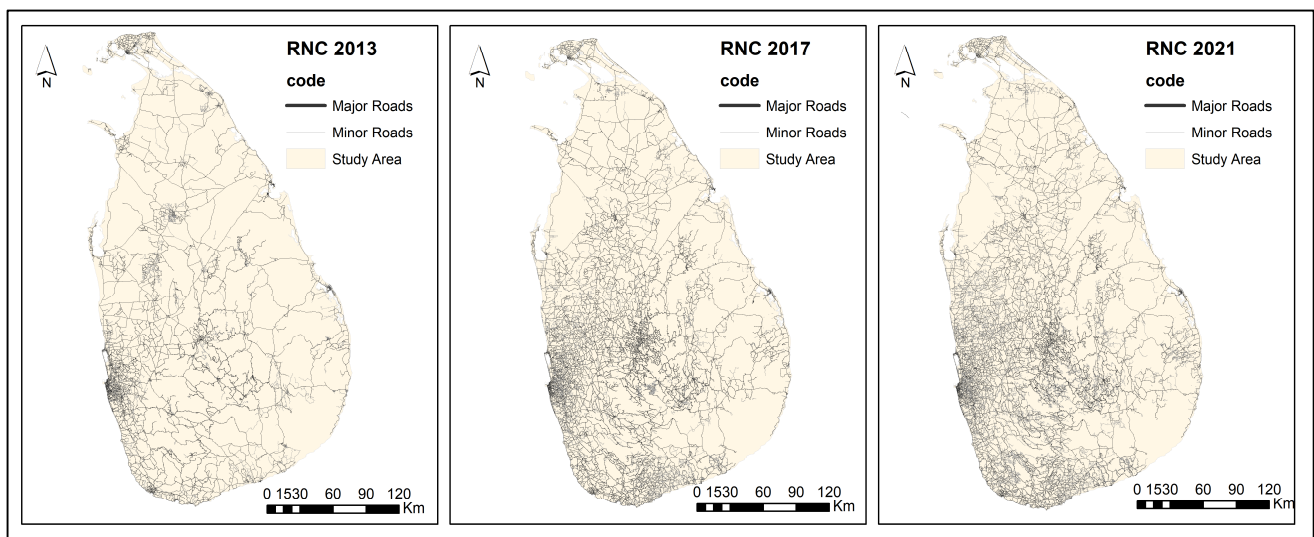


Figure 7. Road network maps of Sri Lanka for 2013, 2017, and 2021.

Normalized Difference Vegetation Index Data

The NDBI given in Figure 8 was calculated using Landsat 8 Collection 2 Tier 1 images. The NDBI dataset was composite considering all the available images within the year and obtaining the average value for the considered year. The following Equation (2) was used to extract the NDBI images for the whole of Sri Lanka covering the study period.

$$\text{NDBI} = (\text{Band3} - \text{Band5}) / (\text{Band3} + \text{Band5}) \quad (2)$$

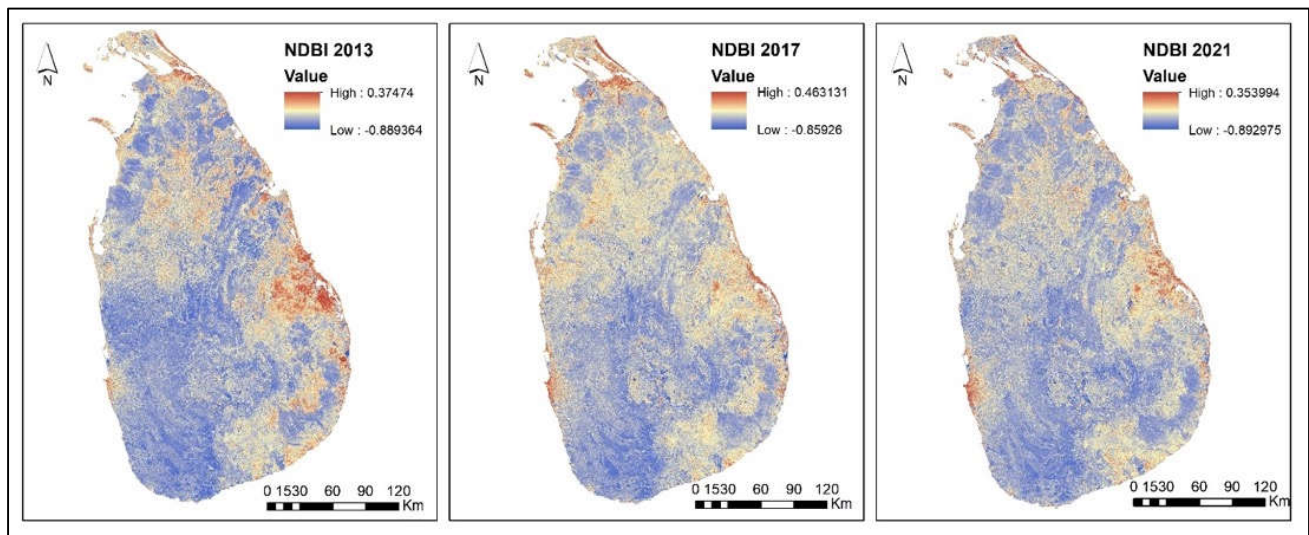


Figure 8. NDBI maps for 2013, 2017, and 2021.

Population Density Distribution

Population density distribution maps given in Figure 9 were developed using the DCS data for 2013, 2017, and 2021.

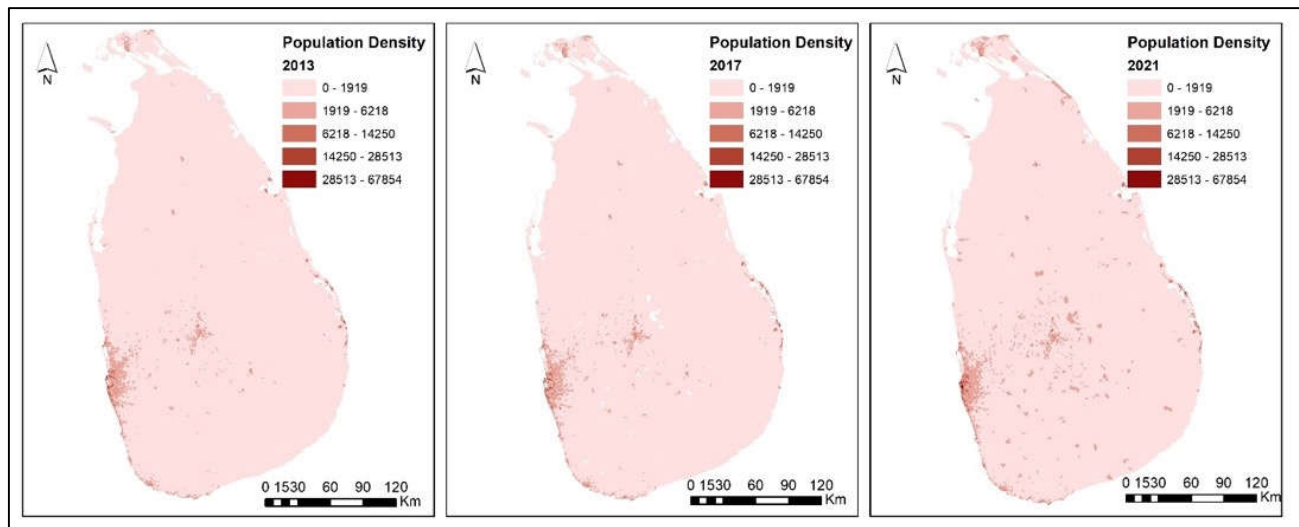


Figure 9. Population density distribution for 2013, 2017, and 2021.

Administrative Boundaries

Adhering to the DCS definition of urban areas in Sri Lanka, MC and UC boundaries (Figure 1) were considered to examine the changes between the study outcomes and the existing urban areas.

3.2.2. Data Preprocessing

Data preprocessing was conducted to enhance the quality of the raw dataset to make it suitable for use in the fusion exercise. Firstly, the downloaded NTL images were subset to the Sri Lankan area using the administrative boundary layer. Secondly, the raw POI dataset was cleaned and categorized into the categories of public places, education, health, leisure, catering, accommodation, shopping, and financial. Under the cleaning process, duplicates and wrongly geocoded POI data were removed. A comprehensive summary of the cleaned dataset is given in Table 2.

Table 2. A comprehensive summary of the cleaned POI dataset.

Primary Classification	Types	Total Points					
		2013		2017		2021	
		Input = 4178		Input = 13,624		Input = 20,020	
		Count	%	Count	%	Count	%
Public Places	Post Office, Police Station, Town Hall, Court House, Prison, etc.	401	10.1%	1078	8.8%	1392	8.3%
Education	University, School, etc.	721	18.2%	2146	17.5%	2510	14.9%
Health	Pharmacy, Hospitals, Clinic, Dentist, Veterinary	285	7.2%	667	5.4%	814	4.8%
Leisure	Theatre, Night Club, Cinema, Park, Playground, Sport Centers, etc.	135	3.4%	293	2.4%	293	1.7%
Catering	Restaurant, Fast Food, Pub, Bar, Food Court, etc.	412	10.4%	1618	13.2%	2996	17.8%
Accommodation	Hotel, Motel, Guesthouse	593	15.0%	2615	21.4%	3374	20%
Shopping	Supermarket, Bakery, Shopping Mall, Department Store, Beverage, Jewelry, etc.	658	16.7%	2096	17.1%	3494	20.7%
Financial	Bank, ATM	746	18.9%	1727	14.1%	1998	11.8%
Output—Sum of cleaned data		3951	100%	12240	100%	16871	100%

Thirdly, MNC data, which were received as GeoTIFF files, were converted into points. Fourthly, the NDVI datasets were converted into (1-NDVI) using the raster calculator in ArcMap, as, usually, the vegetation density is lower in urban areas. Hence, (1-NDVI) had a positive relationship with urban areas [42]. Fifthly, population data obtained for the census years were estimated for the years of 2013, 2017, and 2021. Finally, all the raster data and shapefiles were projected to a WGS 1984 Zone 44N coordinate system.

3.2.3. Pre-Data Fusion

This process of identifying urbanization in Sri Lanka required a pre-data fusion stage for two reasons: (a) to give equal weight to the five datasets used in this study, and (b) to prepare various types of datasets for the data fusion. In this study, min–max normalization and resampling were used to give all datasets equal weights, and a Kernel Density Analysis (KDA) was used for Kernel Density Estimations to convert all datasets into the same raster format prior to the stage of data fusion.

In the pre-data fusion stage, KDA was applied for the POI, RNC, and MNC data, and min–max normalization and resampling processes were conducted for the NTL images and POI, MNC, NDVI, and RNC data to obtain accurate and more reliable data fusion results.

Areas with a higher agglomeration of POI data, MNC data, and RNC were identified through the KDA. For instance, the POI KDA reflected higher values in urban areas than rural areas and lower ones in transition zones between urban and rural areas [69]. All the raster layers were normalized using the min–max normalization equation before the data fusion exercise. According to research conducted by [70], it has been established that performing data resampling before fusion can lead to improved outcomes. Specifically, the utilization of nearest neighbor resampling has proven to be advantageous due to its ability to retain the original values of the raster data, thereby minimizing the introduction of errors. In line with this understanding and to mitigate the potential for errors, this study employed the nearest neighbor resampling technique. In this study, a 100 m spatial resolution was selected for all five fusion raster layers: the NTL raster, POI raster, MNC raster, (1-NDVI) raster, and RNC Raster.

3.2.4. Data Fusion

Data fusion is the integration of information from multiple sources through the application of advanced techniques, with the objective of deriving more precise and valuable insights than those obtainable from any singular data source [71]. The advantage of the data fusion method is that fused data contain more details than single-sourced data. As a result, data fusion methods have been shown to improve study accuracy and reliability [67,72]. Fusing different types of data, especially with NTL images, (a) reduces the blooming effect [73]; (b) reduces the over saturation effect [23]; and (c) increases the accuracy [37,66]. There are different data fusion methods used by different studies such as Wavelet Transform [27,67] Geometric Mean [12], Multi-Level Data fusion [37], and so on.

The use of the geometric mean over other methods discussed, such as Wavelet Transform and Multi-Level Data Fusion, lies in its simplicity and greater accuracy. Unlike Wavelet Transform, which requires complex multi-scale image fusion processes [37], or Multi-Level Data Fusion, which involves intricate steps of sample selection, pixel resolution unification, and feature weighting [45], the geometric mean provides a straightforward yet effective way to fuse different data types. This method effectively eliminates the impact of extreme values and retains the original information of the datasets, ensuring accurate representations of urbanization patterns [12]. Consequently, the geometric mean method is not only easier to implement, but also ensures a high accuracy in capturing the dynamic urbanization patterns crucial for planning and development and urban sprawl identification, not only as a real-time monitoring tool, but also as a predictive tool [74].

This research used the ‘geometric mean’ data fusion method based on the literature to combine the NTL, POI, MNC, RNC, and (1-NDVI) raster data into one detailed raster to identify urban areas. Equation (3) presents the calculation for the geometric mean.

$$GM_n = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n} \quad (3)$$

where GM is the geometric mean of the n number of variables, x_1 , x_2 , and x_n represent the values of the variables, and n is the number of variables.

The geometric mean is a frequently used method in data fusion that can efficiently minimize the influence of image extremum while preserving the original raster’s details [18].

In this study, the datasets were assigned weights based on three criteria: the reliability of the data source, the direct usability of the data, and potential data errors, with the values of 1 (high) and 0 (low). The NTL dataset, sourced from NASA, received a weight of 1.5 due to its high reliability, direct usability from annual average maps, and low errors. The POI data from OpenStreetMap were assigned a weight of 0.5, reflecting their low reliability and potential errors, despite being directly usable. The NDVI data from the USGS were weighted at 1.0, given their credible source and low error rates, though they required merging layers, which could introduce errors. The RNC data, also from OpenStreetMap, received a weight of 0.5 due to similar issues of low reliability and potential data errors. The MNC dataset from local network service providers was assigned a weight of 0.25, as it required manual digitization, introducing errors despite being from a credible source. Table 3 shows the reliability and weights of the each of five datasets.

Table 3. Reliability and weights of the each of five datasets.

Dataset	Type	Reliability (0.25)	Direct Usability (1.0)	Data Errors (0.25)	Weight
NTL Satellite Images	Raster	1	1	1	1.5
POI Dataset	Vector	0	1	0	1
NDVI	Raster	1	0	1	1
RNC	Vector	0	1	0	1
MNC Data	Vector	1	0	0	0.25

Equation (4) for the fused image (FI) using geometric mean data fusion is presented below.

$$FI_i = \sqrt[5]{(1.5 \times NTL_i) \times POI_i \times (0.25 \times MNC_i) \times RNC_i \times (1 - NDVI)_i} \quad (4)$$

where FI_i is the composite index of i , NTL_i is the i th nighttime light DN value, POI_i is the i th POI density raster value, MNC_i is the i th MNC raster value, and RNC_i and $(1 - NDVI)_i$ are the kernel density values of the RNC and NDVI of point i .

Under data fusion, all five datasets (NTL, POI, RNC, MNC, and NDVI) were fused into one raster and then subset for Sri Lanka for the years of 2013, 2017, and 2021 [57].

3.2.5. Urban Area Extraction

The Adaptive Threshold Segmentation algorithm was used from the Open-Source Computer Vision Library (OpenCV) python library to extract the urban areas from the Fused Image (FI). In the ATS algorithm, there are two ways to calculate the optimal threshold values [75]. They are: (a) `Cv2.ADAPTIVE_THRESH_MEAN_C`—calculates the threshold by calculating the mean value of the given block size; and (b) `Cv2.ADAPTIVE_THRESH_GAUSSIAN_C`—calculates the threshold value by taking the weighted sum of the pixels, with weights assigned using the Gaussian window technique.

In this algorithm, the block size can be set based on the area that needs to be scanned for the threshold. If the block size is higher, a large area will be considered when calcu-

lating the threshold value. As this study was conducted for the whole of Sri Lanka, the block size was given as 6001 and the Gaussian method (Option b) was used to obtain the threshold value. The 6001 value was selected based on a visual inspection of the urban areas, and, accordingly, NTL, POI, MNC, RNC, and (1-NDVI) were separated by the image segmentation algorithm.

3.2.6. Accuracy Assessment

Two methods were used to verify the accuracy of this data-fusion-based method of monitoring urbanization in Sri Lanka. These were Precision Assessment and Spatial Accuracy Assessment. The formula for the confusion matrix for Accuracy, Precision, and Recall assessments is given below:

$$\text{Overall Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FP} + \text{FN} + \text{TP}} \quad (5)$$

$$\text{Kappa} = \frac{P_0 - P_e}{1 - P_e} \quad (6)$$

where TP is the number of times a predicted yes was an actual yes, TN is the number of times where a predicted no was actually a no, FP is the number of times where a predicted yes was an actual no, and FN is the number of time where a predicted no was an actual yes. In the Kappa Equation (6), P_0 denotes observed agreement and P_e denotes expected agreement.

The spatial accuracy assessments were conducted by comparing the identified Urban Patches (UPs) with satellite imagery.

3.2.7. Definition of Urban Growth and Rate of Urban Growth

To evaluate the spatial distribution and rate of urban expansion, this study adopted two indicators: urban growth (UG) and rate of urban growth (RUG). Urban growth can be used to represent changes in the urban area within the study period. The rate of urban growth was calculated for the per unit time that the urban growth happened. So, they were two key indices for evaluating the spatial changes in urban expansion in Sri Lanka. The UG and RUG are defined as follows,

$$UG (Y_A - Y_B) = \text{Urban Area of } Y_B - \text{Urban Area of } Y_A \quad (7)$$

$$RUG (Y_A - Y_B) = \frac{UG (Y_A - Y_B)}{(Y_B - Y_A)} \quad (8)$$

In the above equations, Y_A denotes year 1 and Y_B denotes year 2.

4. Results

4.1. Understanding Urbanization Patterns in Sri Lanka through Data Fusion

As shown in Figure 10, the urban areas identified through the big data fusion approach were divided into 61 Urban Patches (UPs). After individually screening the characteristics of the UPs, Wathupitiwala UP (UP 48) was excluded as it referred to the Wathupitiwala Export Processing Zone ('PZ') boundaries. Finally, 60 UPs were identified.

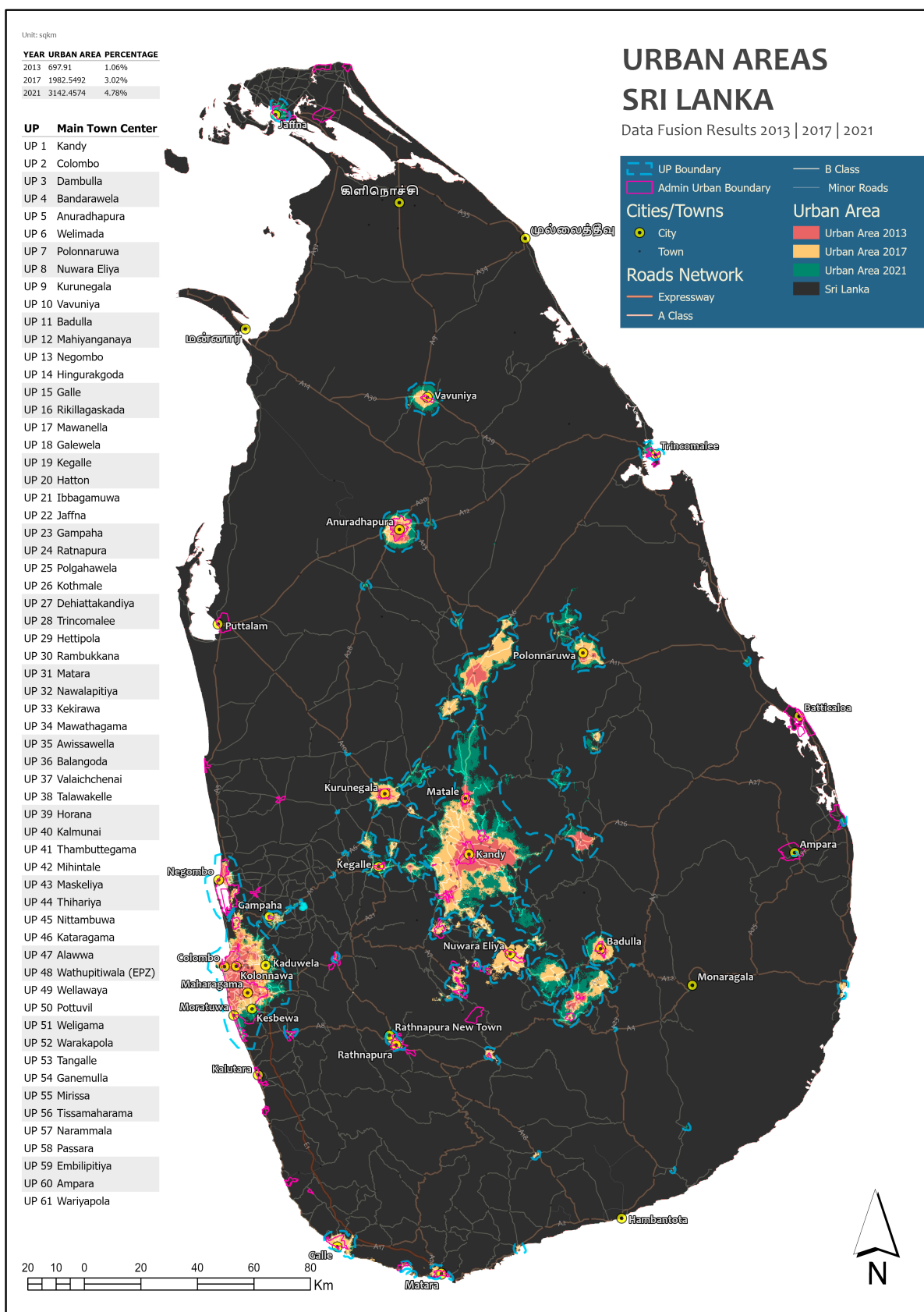


Figure 10. Identified 60 UPs in Sri Lanka.

In this study, the term ‘rate of urban growth (RUG)’ refers to the annual average urban growth (UG) of a particular UP. Accordingly, all the above-identified UPs were categorized as large towns, medium-sized towns, and small towns. The distributions of towns under the above three categories are given in Figures 10 and 11. Figure 11 shows detailed descriptions of the selected UPs in Sri Lanka. The UPs were organized based on the area extent of the identified UPs in 2021. Accordingly, all the aforementioned UPs were categorized based on their RUG and urban area extent. This was to understand the urbanization patterns of different UPs.

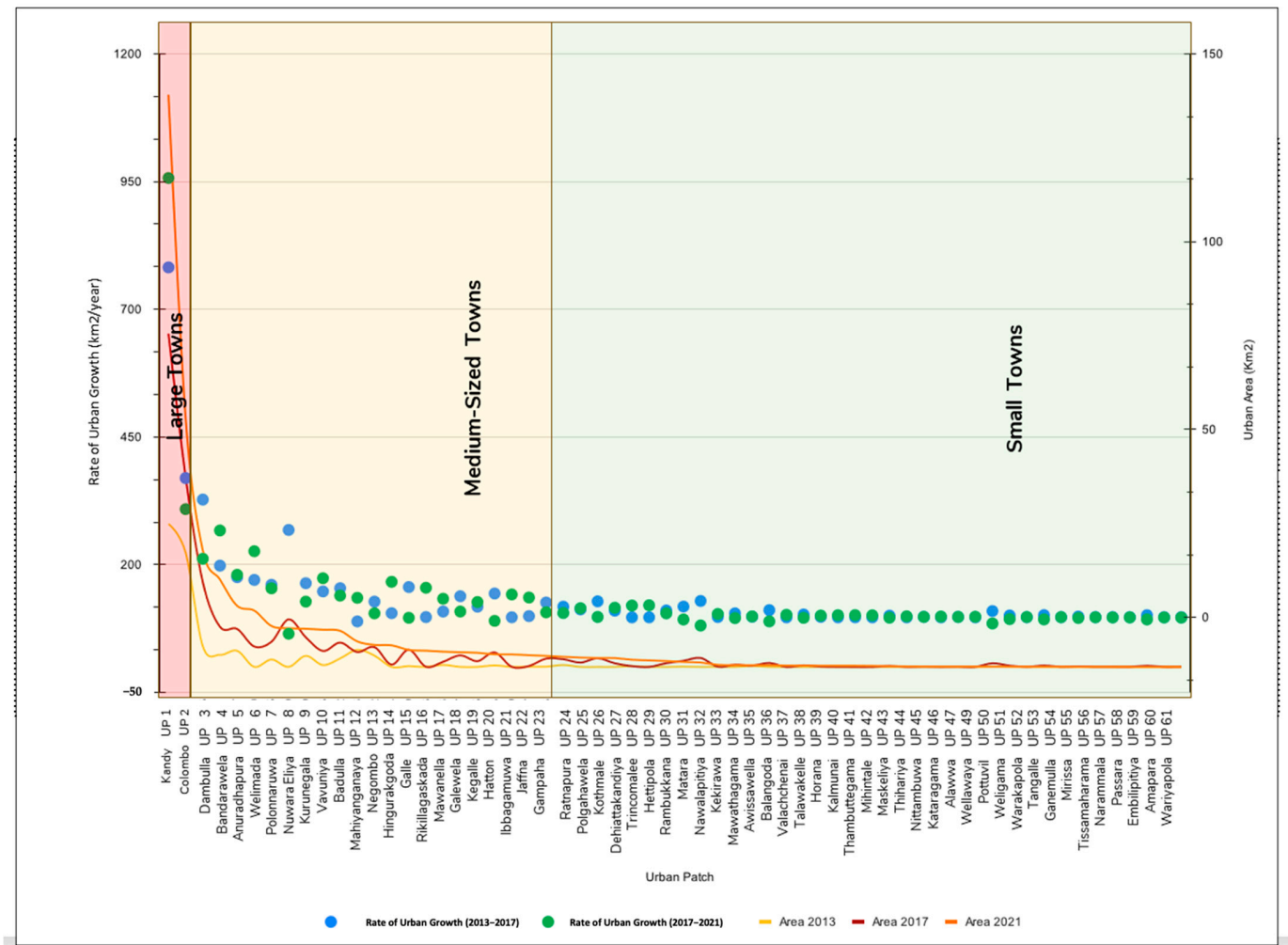


Figure 11. Chart of rate of urban growth (RUG) and the urban area extent.

Figure 11 indicates UPs based on the status of urbanization patterns with reference to the current definition of urbanization of the country: ‘under-bounded (identified UP goes beyond the urban area identified by the CSD)’, ‘over-bounded’ (identified UP does not go beyond the urban area identified by the CSD), and ‘not-within (identified UP is not identified as an urban area by the CSD)’.

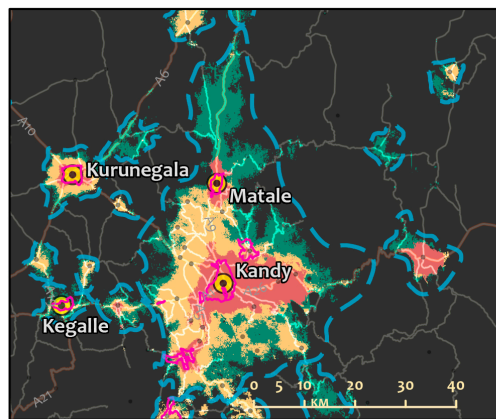
Kandy town (UP1) and Colombo city (UP2) had a higher RUG and expansion that separated these two main areas from the rest. Also, a clear pattern was observed in terms of the RUG rate and expansion identified in Gampaha town (UP 23) and Rathnapura town (UP 24). Balanced UPs had relatively small variations in terms of both RUG and urban area extent. Table 4 shows the estimated accuracy and kappa values of the methodology adopted.

Table 4. Estimated Accuracy and Kappa values.

Year	Class	Rural	Urban	Total	Accuracy	Kappa
2013	Rural	842	145	987	0.8530	0.11
	Urban	2	11	13		
	Total	844	156	1000		
2017	Rural	891	78	969	0.9221	0.42
	Urban	0	32	32		
	Total	891	110	1001		
2021	Rural	833	116	949	0.8810	0.40
	Urban	3	48	51		
	Total	836	164	1000		

The developed approach using data fusion to understand the urbanization patterns in Sri Lanka demonstrated a high efficiency and accuracy across the evaluated years according to the estimate Accuracy and Kappa values. The method consistently produced strong accuracy metrics and moderate to substantial Kappa values, indicating a reliable classification performance for urban areas. Figure 11, given below, categorizes the identified UPs based on the rate of urban growth (RUG) and urban area extent.

Figure 12 visualizes all the identified UPs based on the category they belong to—large towns (L), medium towns (M), and small towns (S). Further, it displays 15 indicators related to each UP. They are: (1) name of the main town center; (2) UP code with reference to the codes given in Figure 10; (3) urban area extent in 2013; (4) 2017; and (5) 2021—this was to understand the temporal changes in each UP in terms of the urban area extent; (6) urban growth (UG) in 2013–2017; (7) UG in 2017–2021; (8) rate of urban growth (RUG) in 2013–2017; (9) RUG in 2017–2021; (10) MC/UC name—this was to understand the related administrative boundaries of each UP. MC/UC areas are known to be urban areas as per the CSD definitions; (11) MC/UC area—this indicator showed the MC/UC area extent relevance to the identified UP; (12) status—this indicator categorized the identified UPs into three statuses. They were: (a) ‘under-bounded’, where the identified UP went beyond the urban area identified by the CSD, (b) ‘over-bounded’, where the identified UP did not go beyond the urban area identified by the CSD, and (c) ‘not-within’, where the identified UP was not identified as an urban area by the CSD; (13) cities and towns in 2013—this indicator showed the names of towns and cities existing within the UP identified in 2013; (14) cities and towns added to the UP by 2017—this showed the names of newly added cities and towns to the same UP identified in 2013, due to the UP expansion; and (15) cities and towns added to the UP by 2021—this showed the names of newly added cities and towns to the same UP identified in 2021, due to the UP expansion. However, none of the UPs identified in 2013 were observed to be shrinking. The letters “L”, “M”, and “S” in Figure 13 refer to the “large”, “medium”, and “small” town categories, as identified in Figure 12. Except for the said categories, distribution of Expressways, A, B, C and Minor roads are visualized in the map. Detailed explanations of subclasses of A,B,C and Minor road categories can be found through http://www.rda.gov.lk/source/rda_roads.htm (accessed on 18 April 2024).



1. Main Town Center	Kandy
2. UP Code	UP 1
3. Urban Area 2013:	279.82 km ²
4. Urban Area 2017:	652.47 km ²
5. Urban Area 2021:	1120.32 km ²
6. UG (2013–2017):	372.64 km ²
7. UG (2017–2021):	467.86 km ²
8. RUG (2013–2017):	93.16 km ² /Year
9. RUG (2017–2021):	116.96 km ² /Year
11. MC/UC Area:	47.7 km ²
12. Status:	Under-bounded

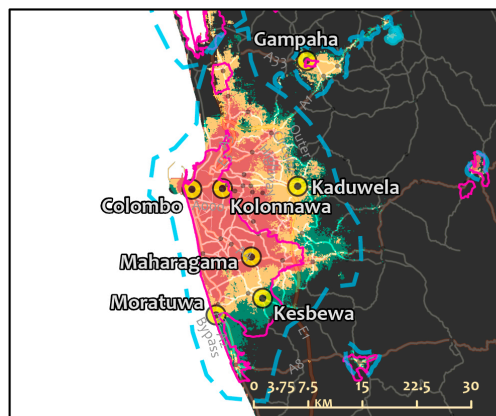
10. MC/UC Name:
Kandy MC, Gampola UC, Wategama UC

13. Cities and Towns in 2013
Kandy, Matale, Katugasthota, Kadugannawa, Digana, Gampola, Theldeniya, Geliyo

14. Cities and Towns added by 2017
Hangurakketa, Pussellawa, Pilimathalawa, Galaha, Delthota, Geliyo, Barigama, Medawala, Hedeniya, Naula

15. Cities and Towns added by 2021
Hunnasgiriya, Rattota, Galagedara, Medamahenuwara, Ududumbara

L



1. Main Town Center	Colombo
2. UP Code	UP 2
3. Urban Area 2013:	223.79 km ²
4. Urban Area 2017:	372.04 km ²
5. Urban Area 2021:	487.25 km ²
6. UG (2013–2017):	148.24 km ²
7. UG (2017–2021):	115.21 km ²
8. RUG (2013–2017):	37.06 km ² /Year
9. RUG (2017–2021):	28.80 km ² /Year
11. MC/UC Area:	40.21 km ²
12. Status:	Under-bounded

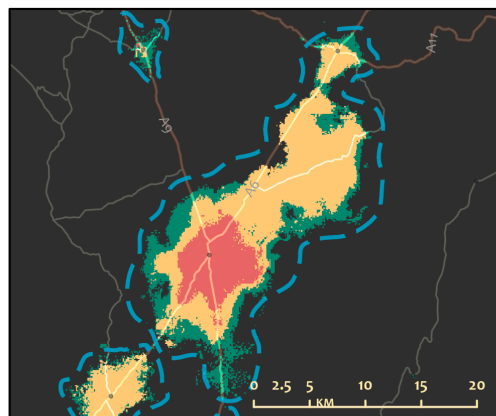
10. MC/UC Name:
Colombo MC

13. Cities and Towns in 2013
Colombo, Panadura, Kolonnawa, Homagama, Nugedoda, Maharagama, Baththaramulla, Pannipitiya, Borelesgamuwa, Dehiwala, Mt. Lavinia, Kottawa Peliyagoda, Wattala, Kiribathgoda, Kadawatha, Ja-Ela, Kandana, Ragama

14. Cities and Towns added by 2017
Moratuwa, Welivita, Ekala, Kaduwela, Malabe, Kothalawala, Biyagama, Piyandala, Hokandara, Rukmale

15. Cities and Towns added by 2021
Kesbewa, Madapatha, Keselwatta, Athurugiriya, Meegoda

L



1. Main Town Center	Dambulla
2. UP Code	UP 3
3. Urban Area 2013:	40.57 km ²
4. Urban Area 2017:	165.97 km ²
5. Urban Area 2021:	228.31 km ²
6. UG (2013–2017):	125.40 km ²
7. UG (2017–2021):	62.34 km ²
8. RUG (2013–2017):	31.35 km ² /Year
9. RUG (2017–2021):	15.59 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

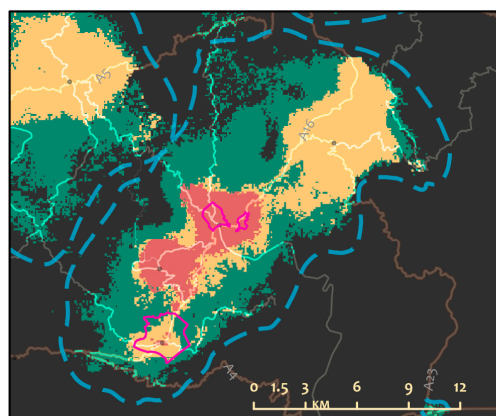
10. MC/UC Name:

13. Cities and Towns in 2013
Dambulla

14. Cities and Towns added by 2017
Inamaluwa, Sigiriya, Habarana

15. Cities and Towns added by 2021

M



1. Main Town Center	Bandarawela
2. UP Code	UP 4
3. Urban Area 2013:	23.52 km ²
4. Urban Area 2017:	78.57 km ²
5. Urban Area 2021:	171.05 km ²
6. UG (2013–2017):	55.05 km ²
7. UG (2017–2021):	92.48 km ²
8. RUG (2013–2017):	13.76 km ² /Year
9. RUG (2017–2021):	23.12 km ² /Year
11. MC/UC Area:	3.44 km ²
12. Status:	Under-bounded

10. MC/UC Name:
Bandarawela UC

13. Cities and Towns in 2013
Bandarawela, Diyatawala, Haputale

14. Cities and Towns added by 2017
Ella

15. Cities and Towns added by 2021
Ballaketuwa

M

Figure 12. Cont.

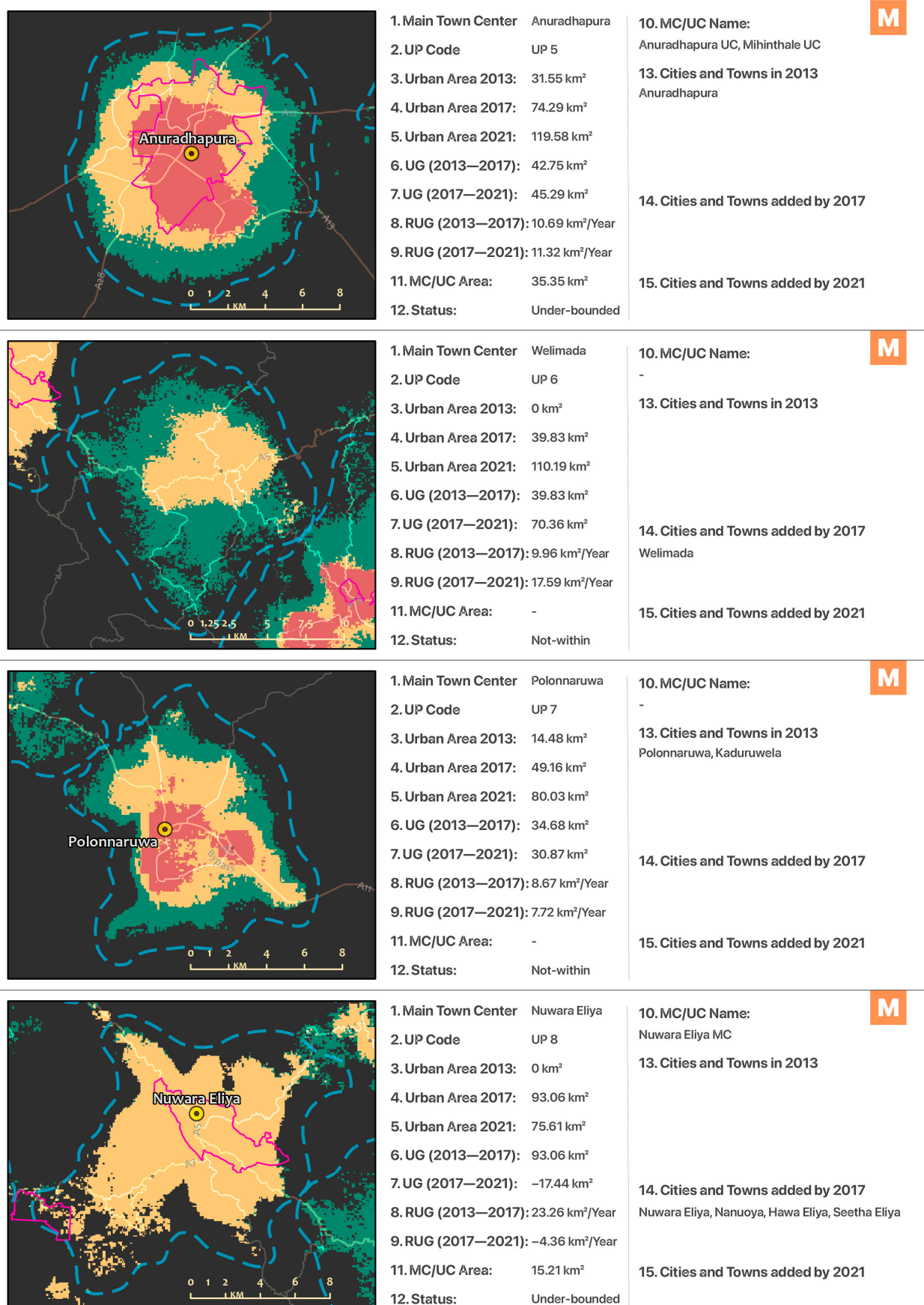
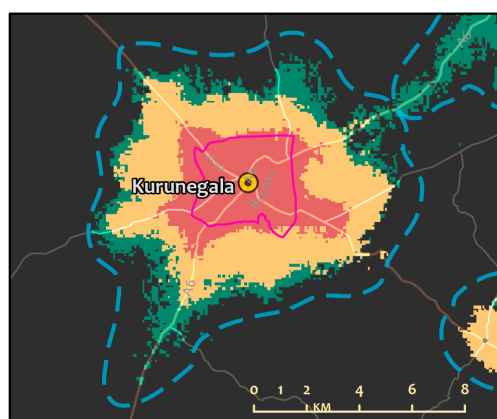


Figure 12. Cont.



1. Main Town Center	Kurunegala
2. UP Code	UP 9
3. Urban Area 2013:	21.6 km ²
4. Urban Area 2017:	57.92 km ²
5. Urban Area 2021:	74.67 km ²
6. UG (2013—2017):	36.32 km ²
7. UG (2017—2021):	16.75 km ²
8. RUG (2013—2017):	9.08 km ² /Year
9. RUG (2017—2021):	4.19 km ² /Year
11. MC/UC Area:	11.87 km ²
12. Status:	Under-bounded

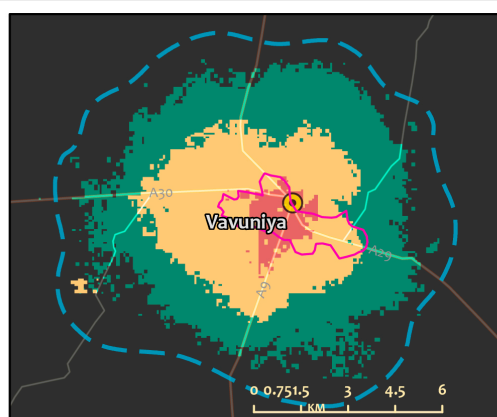
10. MC/UC Name:
Kurunegala MC

13. Cities and Towns in 2013
Kurunegala, Mallawapitiya, Yanthampalawa, Malkaduwwa

14. Cities and Towns added by 2017
Wettaragoda, Heraliyawala, Badagamuwa

15. Cities and Towns added by 2021
Dambokka, Rathgalla

M



1. Main Town Center	Vavuniya
2. UP Code	UP 10
3. Urban Area 2013:	3.55 km ²
4. Urban Area 2017:	31.14 km ²
5. Urban Area 2021:	72.77 km ²
6. UG (2013—2017):	27.58 km ²
7. UG (2017—2021):	41.64 km ²
8. RUG (2013—2017):	6.90 km ² /Year
9. RUG (2017—2021):	10.41 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

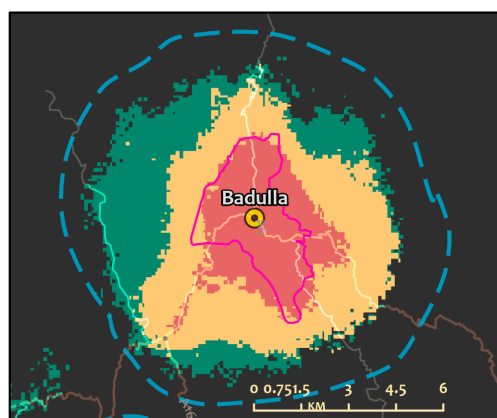
10. MC/UC Name:
-

13. Cities and Towns in 2013
Vavuniya

14. Cities and Towns added by 2017

15. Cities and Towns added by 2021

M



1. Main Town Center	Badulla
2. UP Code	UP 11
3. Urban Area 2013:	16.55 km ²
4. Urban Area 2017:	47.54 km ²
5. Urban Area 2021:	70.65 km ²
6. UG (2013—2017):	30.99 km ²
7. UG (2017—2021):	23.11 km ²
8. RUG (2013—2017):	7.75 km ² /Year
9. RUG (2017—2021):	5.78 km ² /Year
11. MC/UC Area:	10.61 km ²
12. Status:	Under-bounded

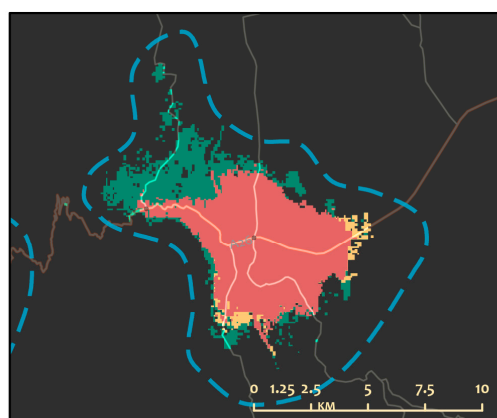
10. MC/UC Name:
Badulla MC

13. Cities and Towns in 2013
Badulla

14. Cities and Towns added by 2017
Hali Ela

15. Cities and Towns added by 2021

M



1. Main Town Center	Mahiyanganaya
2. UP Code	UP 12
3. Urban Area 2013:	33.43 km ²
4. Urban Area 2017:	29.05 km ²
5. Urban Area 2021:	49.82 km ²
6. UG (2013—2017):	-4.37 km ²
7. UG (2017—2021):	20.76 km ²
8. RUG (2013—2017):	-1.09 km ² /Year
9. RUG (2017—2021):	5.19 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

10. MC/UC Name:
-

13. Cities and Towns in 2013
Mahiyanganaya

14. Cities and Towns added by 2017

15. Cities and Towns added by 2021

M

Figure 12. Cont.

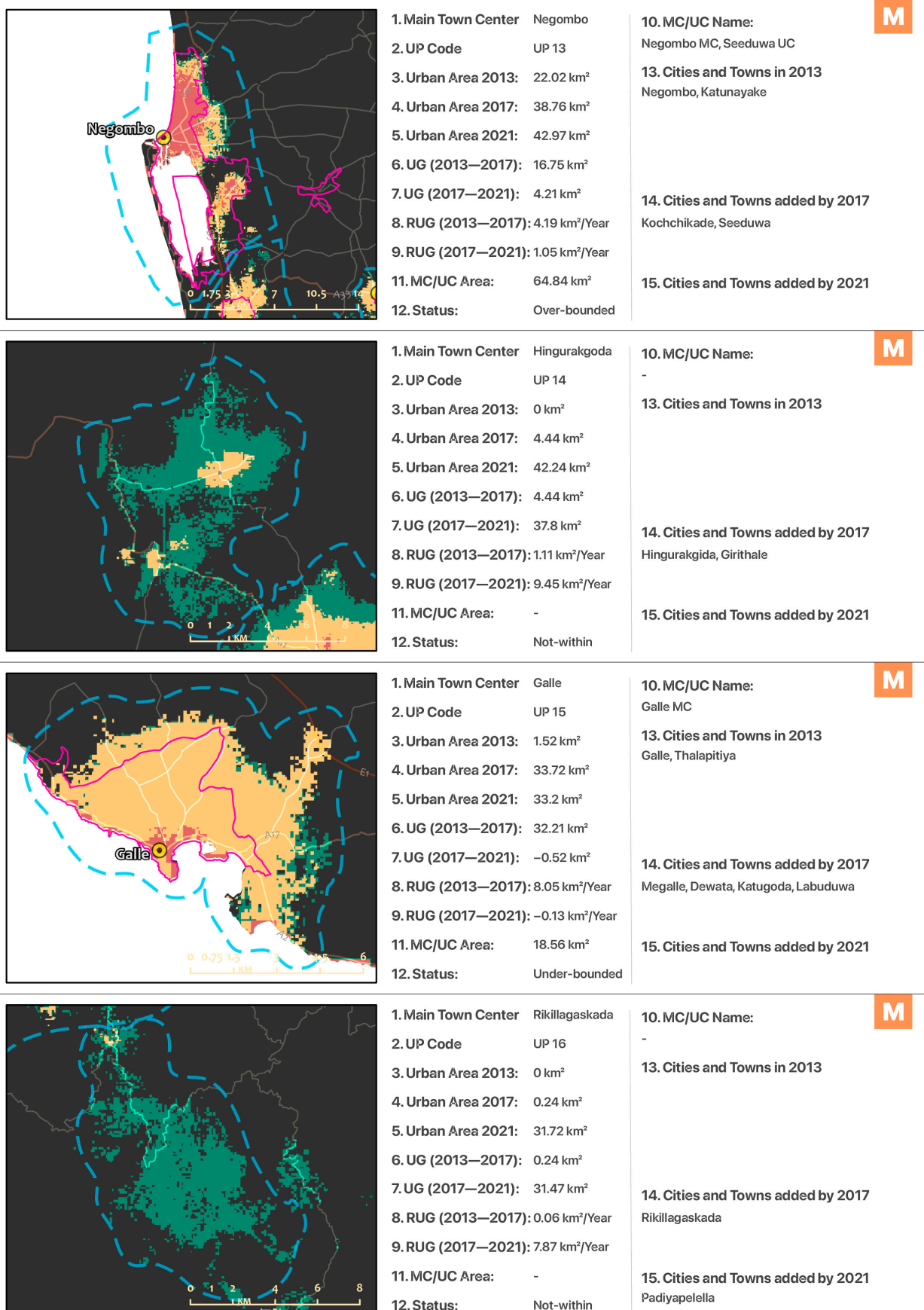
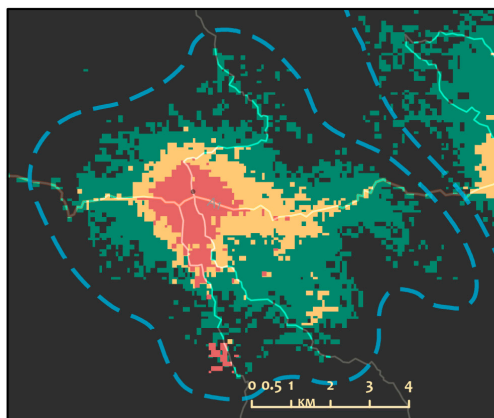


Figure 12. Cont.



1. Main Town Center	Mawanella
2. UP Code	UP 17
3. Urban Area 2013:	3.88 km ²
4. Urban Area 2017:	10.05 km ²
5. Urban Area 2021:	29.83 km ²
6. UG (2013–2017):	6.17 km ²
7. UG (2017–2021):	19.77 km ²
8. RUG (2013–2017):	1.54 km ² /Year
9. RUG (2017–2021):	4.94 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

10. MC/UC Name:

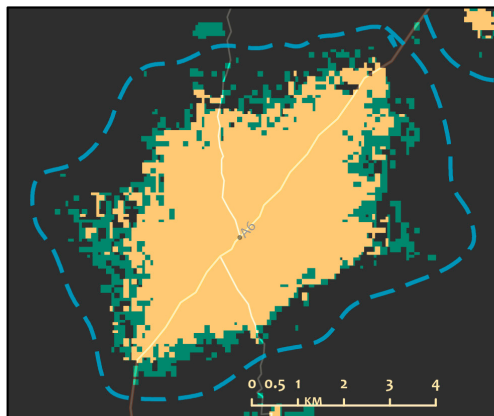
-

13. Cities and Towns in 2013
Mawanella, Hingula

14. Cities and Towns added by 2017

15. Cities and Towns added by 2021
Uthuwankanda

M



1. Main Town Center	Galewela
2. UP Code	UP 18
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	22.48 km ²
5. Urban Area 2021:	28.65 km ²
6. UG (2013–2017):	22.48 km ²
7. UG (2017–2021):	6.17 km ²
8. RUG (2013–2017):	5.62 km ² /Year
9. RUG (2017–2021):	1.54 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

10. MC/UC Name:

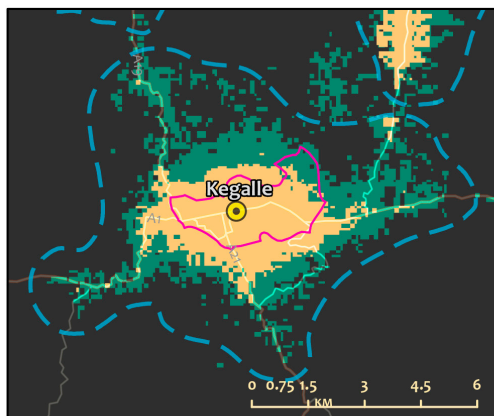
-

13. Cities and Towns in 2013

14. Cities and Towns added by 2017
Galewela

15. Cities and Towns added by 2021

M



1. Main Town Center	Kegalle
2. UP Code	UP 19
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	11.32 km ²
5. Urban Area 2021:	27.63 km ²
6. UG (2013–2017):	11.32 km ²
7. UG (2017–2021):	16.32 km ²
8. RUG (2013–2017):	2.83 km ² /Year
9. RUG (2017–2021):	4.08 km ² /Year
11. MC/UC Area:	5.72 km ²
12. Status:	Under-bounded

10. MC/UC Name:

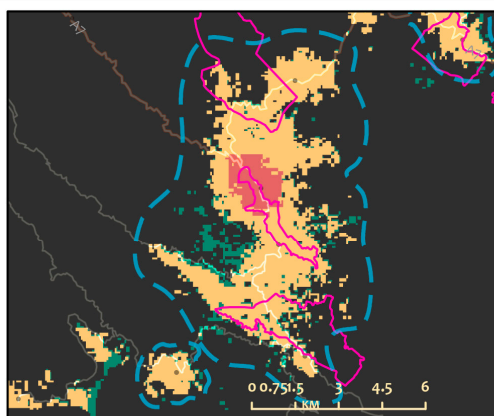
Kegalle UC

13. Cities and Towns in 2013

14. Cities and Towns added by 2017
Kegalle

15. Cities and Towns added by 2021

M



1. Main Town Center	Hatton
2. UP Code	UP 20
3. Urban Area 2013:	2.9 km ²
4. Urban Area 2017:	28.28 km ²
5. Urban Area 2021:	24.5 km ²
6. UG (2013–2017):	25.38 km ²
7. UG (2017–2021):	-3.78 km ²
8. RUG (2013–2017):	6.34 km ² /Year
9. RUG (2017–2021):	-0.94 km ² /Year
11. MC/UC Area:	27.2 km ²
12. Status:	Over-bounded

10. MC/UC Name:

Hatton-Dikoya UC

13. Cities and Towns in 2013
Hatton14. Cities and Towns added by 2017
Dick Oya, Kotagala, Norwood

15. Cities and Towns added by 2021

M

Figure 12. Cont.

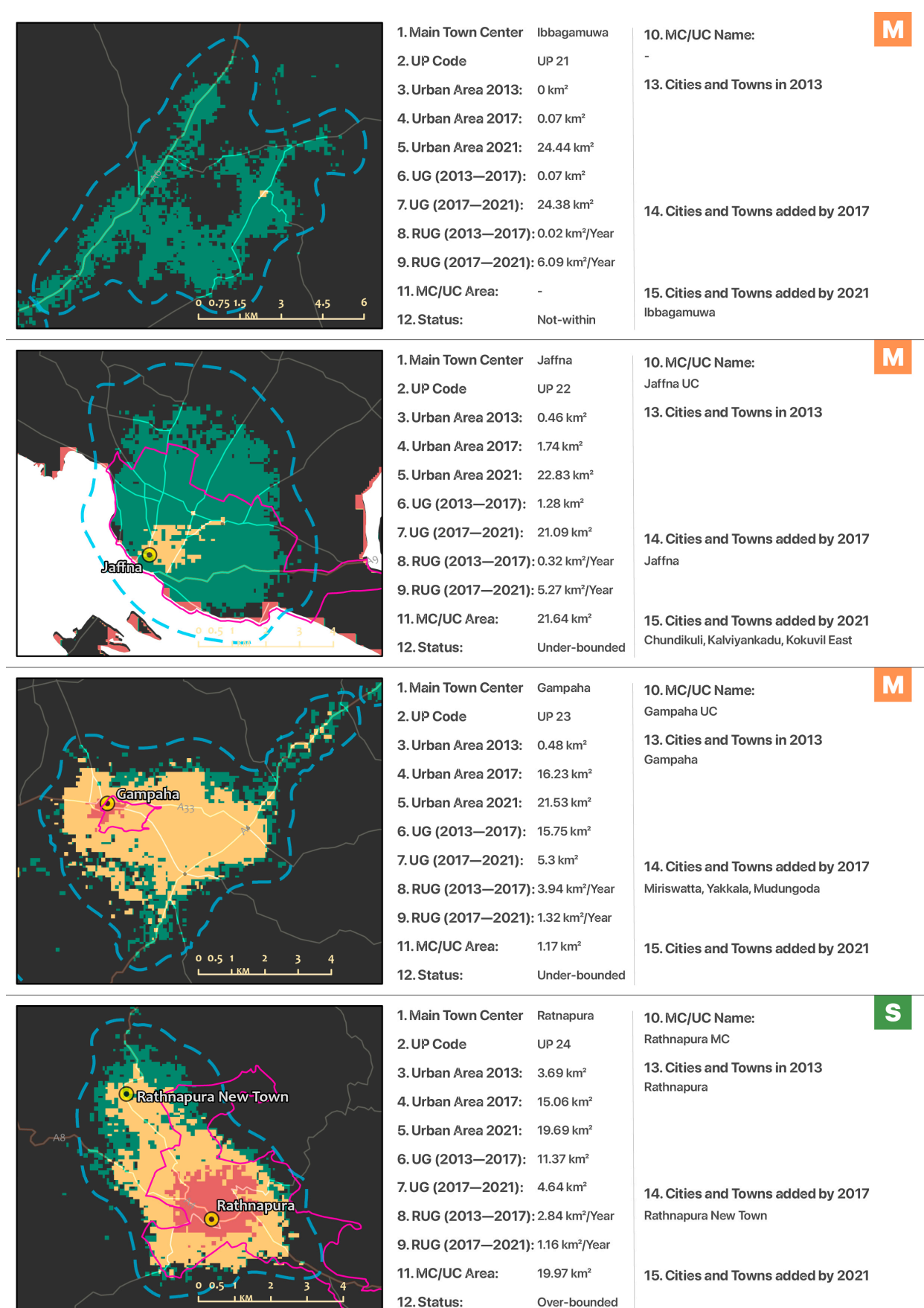


Figure 12. Cont.

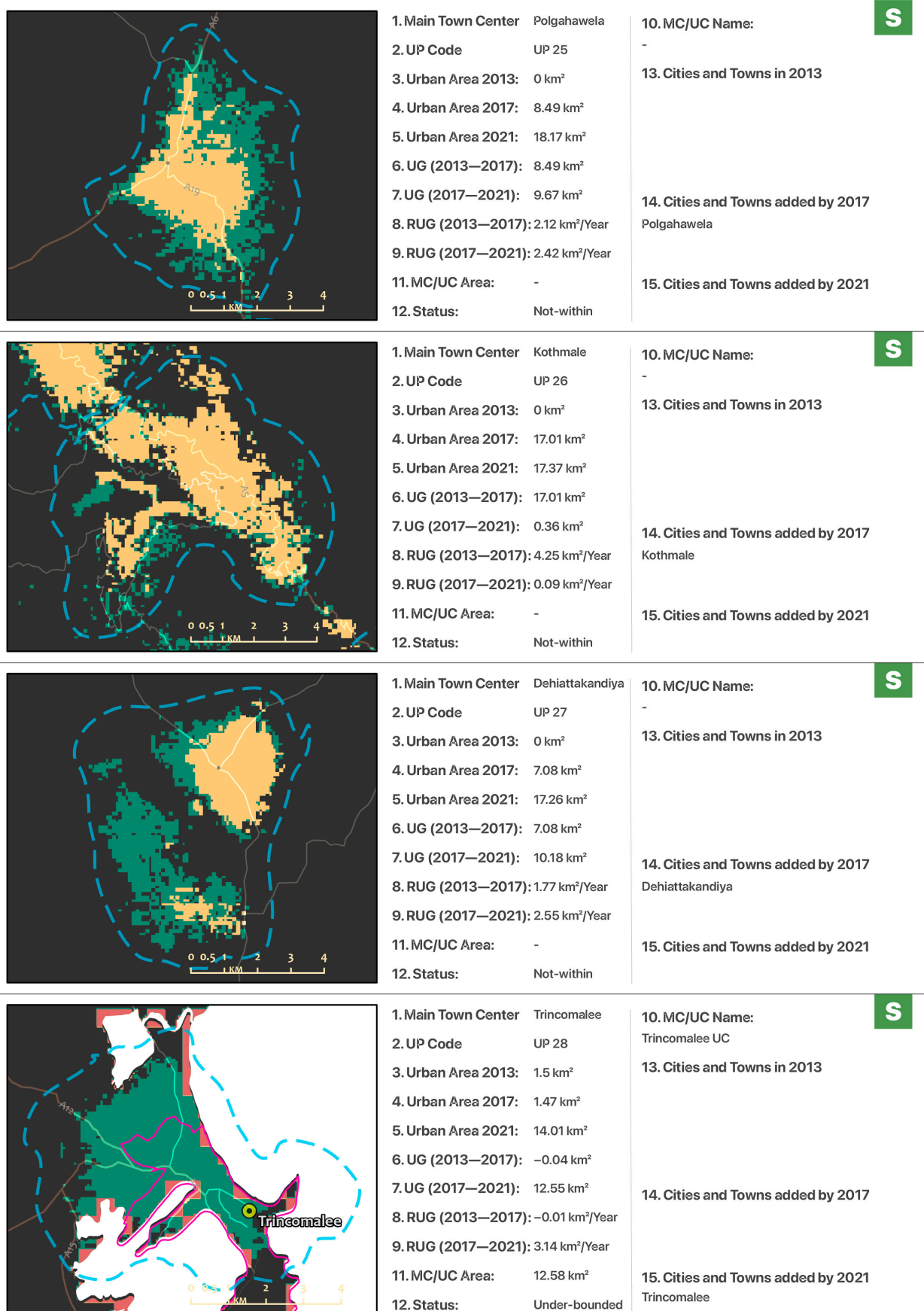


Figure 12. Cont.

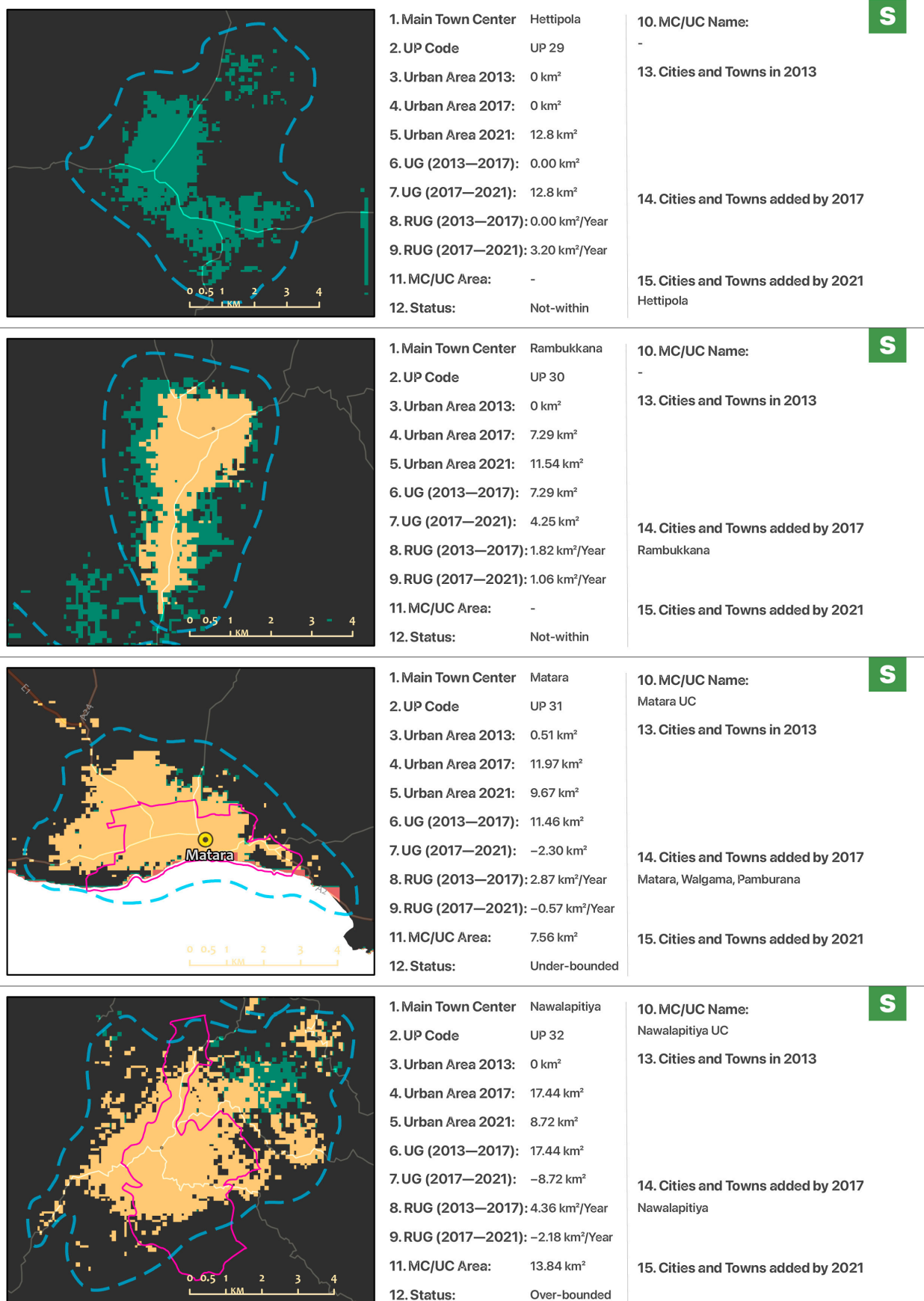
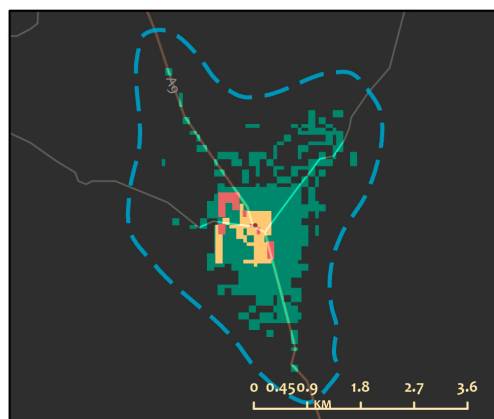


Figure 12. Cont.



1. Main Town Center	Kekirawa
2. UP Code	UP 33
3. Urban Area 2013:	0.18 km ²
4. Urban Area 2017:	0.48 km ²
5. Urban Area 2021:	4.04 km ²
6. UG (2013–2017):	0.30 km ²
7. UG (2017–2021):	3.55 km ²
8. RUG (2013–2017):	0.08 km ² /Year
9. RUG (2017–2021):	0.89 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

10. MC/UC Name:

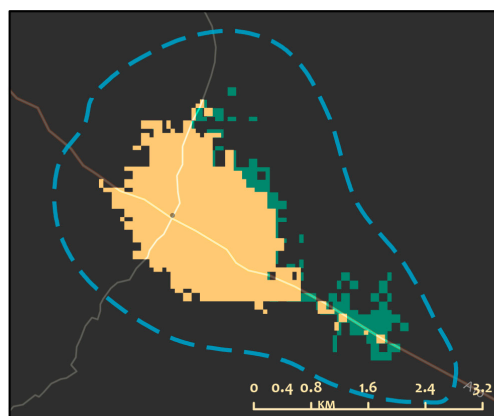
S

-

13. Cities and Towns in 2013
Kekirawa

14. Cities and Towns added by 2017

15. Cities and Towns added by 2021



1. Main Town Center	Mawathagama
2. UP Code	UP 34
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	4.22 km ²
5. Urban Area 2021:	3.43 km ²
6. UG (2013–2017):	4.22 km ²
7. UG (2017–2021):	-0.79 km ²
8. RUG (2013–2017):	1.06 km ² /Year
9. RUG (2017–2021):	-0.20 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

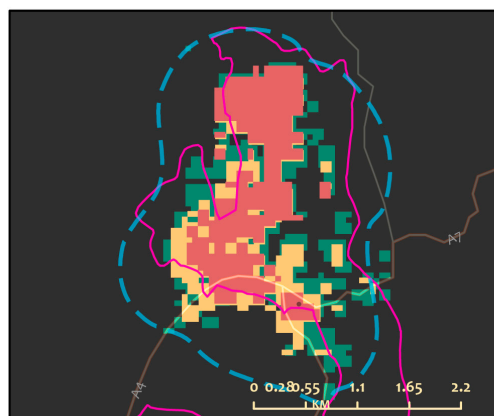
10. MC/UC Name:

S

-

13. Cities and Towns in 2013
Mawathagama14. Cities and Towns added by 2017
Mawathagama

15. Cities and Towns added by 2021



1. Main Town Center	Awissawella
2. UP Code	UP 35
3. Urban Area 2013:	1.76 km ²
4. Urban Area 2017:	2.46 km ²
5. Urban Area 2021:	3.41 km ²
6. UG (2013–2017):	0.70 km ²
7. UG (2017–2021):	0.95 km ²
8. RUG (2013–2017):	0.18 km ² /Year
9. RUG (2017–2021):	0.24 km ² /Year
11. MC/UC Area:	8.89 km ²
12. Status:	Over-bounded

10. MC/UC Name:

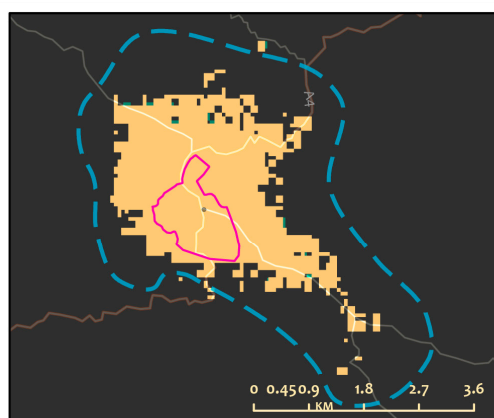
S

Seethawakapura UC

13. Cities and Towns in 2013
Awissawella, Seethagama

14. Cities and Towns added by 2017

15. Cities and Towns added by 2021



1. Main Town Center	Balangoda
2. UP Code	UP 36
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	7.66 km ²
5. Urban Area 2021:	3.33 km ²
6. UG (2013–2017):	7.66 km ²
7. UG (2017–2021):	-4.33 km ²
8. RUG (2013–2017):	1.91 km ² /Year
9. RUG (2017–2021):	-1.08 km ² /Year
11. MC/UC Area:	1.32 km ²
12. Status:	Under-bounded

10. MC/UC Name:

S

Balangoda UC

13. Cities and Towns in 2013
Balangoda14. Cities and Towns added by 2017
Balangoda

15. Cities and Towns added by 2021

Figure 12. Cont.

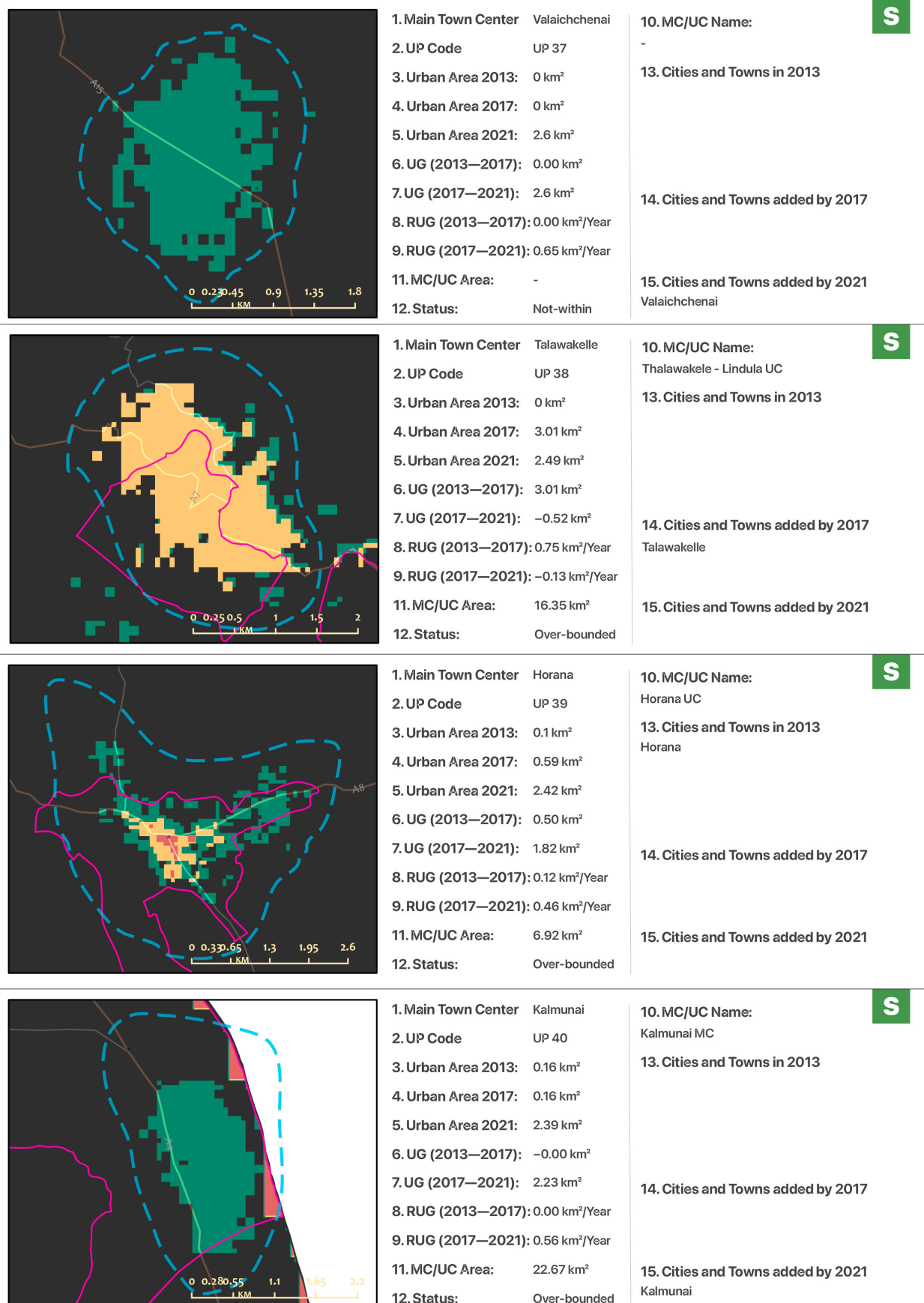
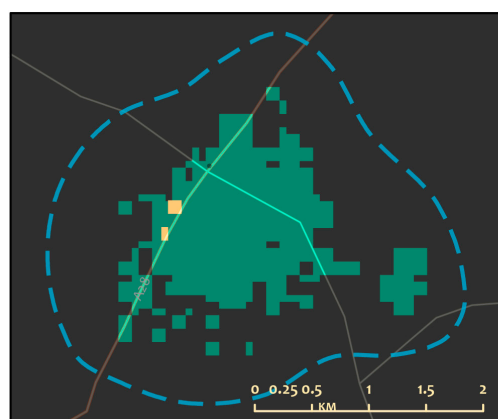


Figure 12. Cont.



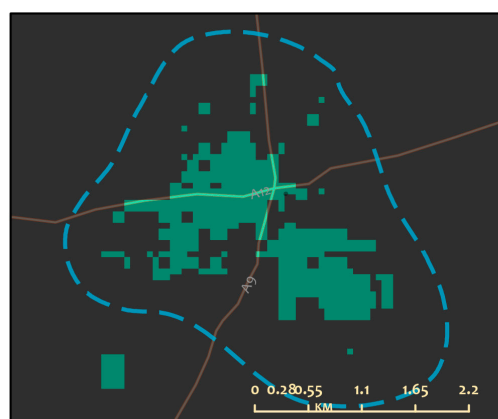
1. Main Town Center	Thambuttegama
2. UP Code	UP 41
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	0.02 km ²
5. Urban Area 2021:	2.33 km ²
6. UG (2013–2017):	0.02 km ²
7. UG (2017–2021):	2.31 km ²
8. RUG (2013–2017):	0.01 km ² /Year
9. RUG (2017–2021):	0.58 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

10. MC/UC Name:

S

13. Cities and Towns in 2013

14. Cities and Towns added by 2017

15. Cities and Towns added by 2021
Thambuttegama

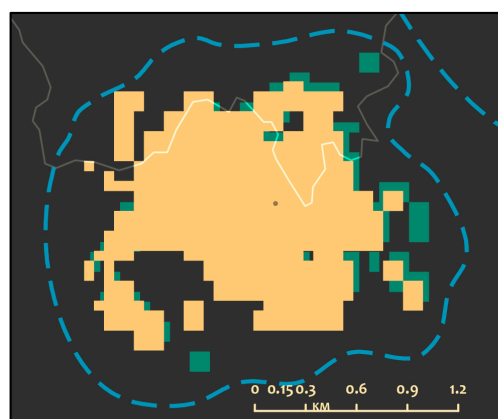
1. Main Town Center	Mihintale
2. UP Code	UP 42
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	0 km ²
5. Urban Area 2021:	2.08 km ²
6. UG (2013–2017):	0.00 km ²
7. UG (2017–2021):	2.08 km ²
8. RUG (2013–2017):	0.00 km ² /Year
9. RUG (2017–2021):	0.52 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

10. MC/UC Name:

S

13. Cities and Towns in 2013

14. Cities and Towns added by 2017

15. Cities and Towns added by 2021
Mihintale

1. Main Town Center	Maskeliya
2. UP Code	UP 43
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	1.78 km ²
5. Urban Area 2021:	1.43 km ²
6. UG (2013–2017):	1.78 km ²
7. UG (2017–2021):	-0.35 km ²
8. RUG (2013–2017):	0.44 km ² /Year
9. RUG (2017–2021):	-0.09 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

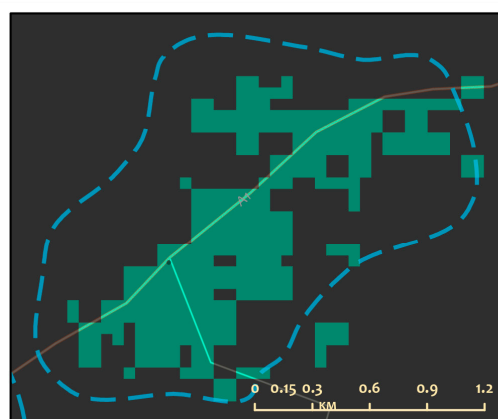
10. MC/UC Name:

S

13. Cities and Towns in 2013

14. Cities and Towns added by 2017
Maskeliya

15. Cities and Towns added by 2021



1. Main Town Center	Thihariya
2. UP Code	UP 44
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	0 km ²
5. Urban Area 2021:	1.05 km ²
6. UG (2013–2017):	0.00 km ²
7. UG (2017–2021):	1.05 km ²
8. RUG (2013–2017):	0.00 km ² /Year
9. RUG (2017–2021):	0.26 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

10. MC/UC Name:

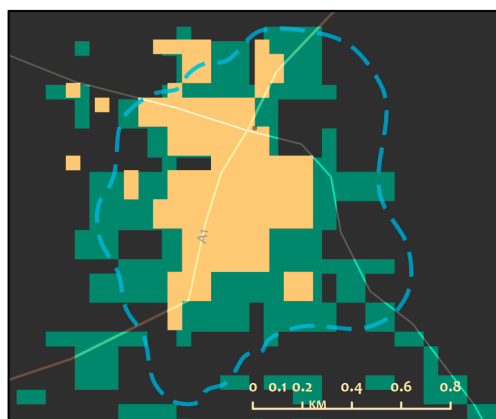
S

13. Cities and Towns in 2013

14. Cities and Towns added by 2017

15. Cities and Towns added by 2021
Thihariya

Figure 12. Cont.



1. Main Town Center	Nittambuwa
2. UP Code	UP 45
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	0.41 km ²
5. Urban Area 2021:	0.88 km ²
6. UG (2013–2017):	0.41 km ²
7. UG (2017–2021):	0.47 km ²
8. RUG (2013–2017):	0.10 km ² /Year
9. RUG (2017–2021):	0.12 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

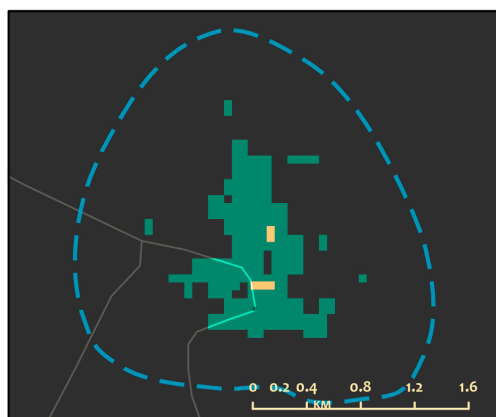
10. MC/UC Name:

S

13. Cities and Towns in 2013

14. Cities and Towns added by 2017
Nittambuwa

15. Cities and Towns added by 2021



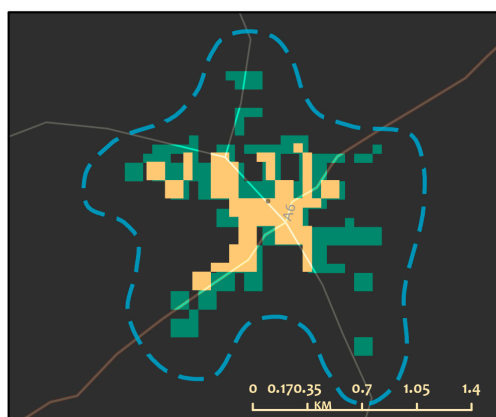
1. Main Town Center	Kataragama
2. UP Code	UP 46
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	0.02 km ²
5. Urban Area 2021:	0.79 km ²
6. UG (2013–2017):	0.02 km ²
7. UG (2017–2021):	0.78 km ²
8. RUG (2013–2017):	0.00 km ² /Year
9. RUG (2017–2021):	0.19 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

10. MC/UC Name:

S

13. Cities and Towns in 2013

14. Cities and Towns added by 2017

15. Cities and Towns added by 2021
Kataragama

1. Main Town Center	Alawwa
2. UP Code	UP 47
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	0.34 km ²
5. Urban Area 2021:	0.78 km ²
6. UG (2013–2017):	0.34 km ²
7. UG (2017–2021):	0.44 km ²
8. RUG (2013–2017):	0.08 km ² /Year
9. RUG (2017–2021):	0.11 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

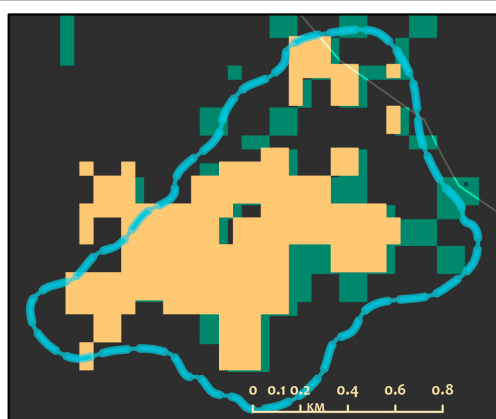
10. MC/UC Name:

S

13. Cities and Towns in 2013

14. Cities and Towns added by 2017
Alawwa

15. Cities and Towns added by 2021



1. Main Town Center	Wathupitiwala
2. UP Code	UP 48
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	0.67 km ²
5. Urban Area 2021:	0.7 km ²
6. UG (2013–2017):	0.67 km ²
7. UG (2017–2021):	0.02 km ²
8. RUG (2013–2017):	0.17 km ² /Year
9. RUG (2017–2021):	0.01 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

10. MC/UC Name:

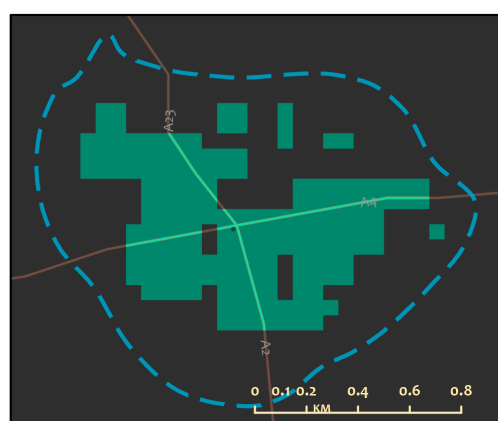
EPZ

13. Cities and Towns in 2013

14. Cities and Towns added by 2017
Wathupitiwala

15. Cities and Towns added by 2021

Figure 12. Cont.



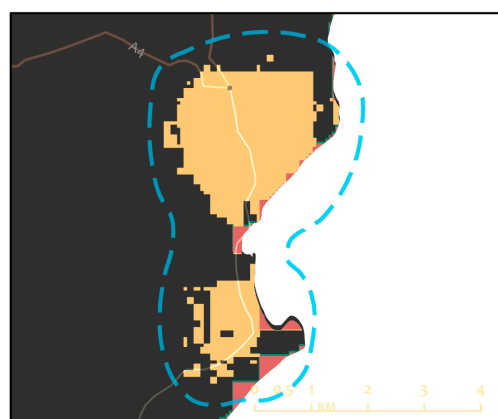
1. Main Town Center	Wellawaya
2. UP Code	UP 49
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	0 km ²
5. Urban Area 2021:	0.6 km ²
6. UG (2013–2017):	0.00 km ²
7. UG (2017–2021):	0.6 km ²
8. RUG (2013–2017):	0.00 km ² /Year
9. RUG (2017–2021):	0.15 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

10. MC/UC Name:

S

13. Cities and Towns in 2013

14. Cities and Towns added by 2017

15. Cities and Towns added by 2021
Wellawaya

1. Main Town Center	Pottuvil
2. UP Code	UP 50
3. Urban Area 2013:	0.45 km ²
4. Urban Area 2017:	7.11 km ²
5. Urban Area 2021:	0.58 km ²
6. UG (2013–2017):	6.66 km ²
7. UG (2017–2021):	-6.53 km ²
8. RUG (2013–2017):	1.67 km ² /Year
9. RUG (2017–2021):	-1.63 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

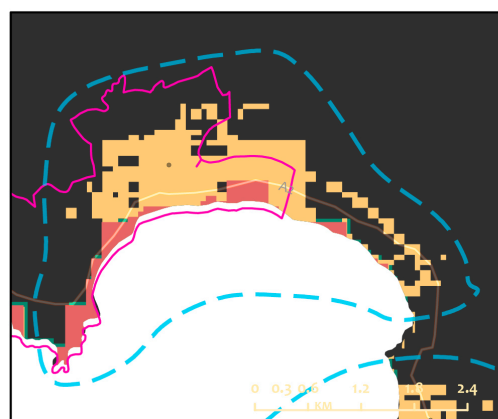
10. MC/UC Name:

S

13. Cities and Towns in 2013

14. Cities and Towns added by 2017
Pottuvil

15. Cities and Towns added by 2021



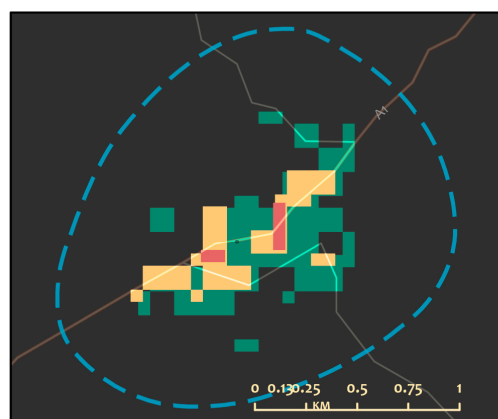
1. Main Town Center	Weligama
2. UP Code	UP 51
3. Urban Area 2013:	0.46 km ²
4. Urban Area 2017:	2.46 km ²
5. Urban Area 2021:	0.55 km ²
6. UG (2013–2017):	2.00 km ²
7. UG (2017–2021):	-1.91 km ²
8. RUG (2013–2017):	0.50 km ² /Year
9. RUG (2017–2021):	-0.48 km ² /Year
11. MC/UC Area:	5.4 km ²
12. Status:	Over-bounded

10. MC/UC Name:

S

Weligama UC

13. Cities and Towns in 2013

14. Cities and Towns added by 2017
Weligama15. Cities and Towns added by 2021
Weligama

1. Main Town Center	Warakapola
2. UP Code	UP 52
3. Urban Area 2013:	0.02 km ²
4. Urban Area 2017:	0.17 km ²
5. Urban Area 2021:	0.45 km ²
6. UG (2013–2017):	0.15 km ²
7. UG (2017–2021):	0.28 km ²
8. RUG (2013–2017):	0.04 km ² /Year
9. RUG (2017–2021):	0.07 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

10. MC/UC Name:

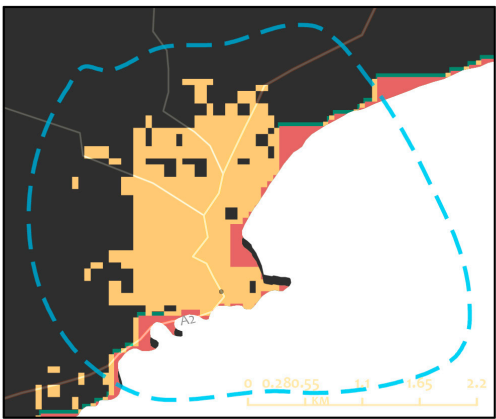
S

13. Cities and Towns in 2013

14. Cities and Towns added by 2017
Warakapola

15. Cities and Towns added by 2021

Figure 12. Cont.



1. Main Town Center	Tangalle
2. UP Code	UP 53
3. Urban Area 2013:	0.27 km ²
4. Urban Area 2017:	2.63 km ²
5. Urban Area 2021:	0.4 km ²
6. UG (2013—2017):	2.36 km ²
7. UG (2017—2021):	−2.23 km ²
8. RUG (2013—2017):	0.59 km ² /Year
9. RUG (2017—2021):	−0.56 km ² /Year
11. MC/UC Area:	−
12. Status:	Not-within

10. MC/UC Name:

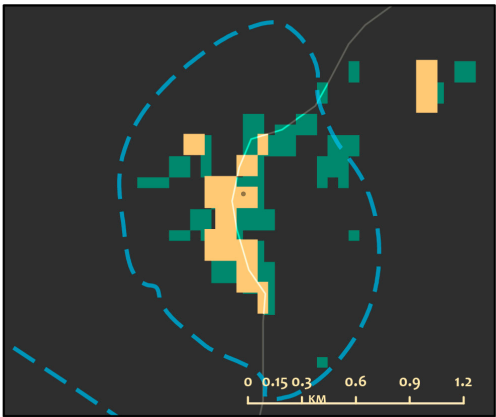
−

13. Cities and Towns in 2013

14. Cities and Towns added by 2017
Tangalle

15. Cities and Towns added by 2021

S



1. Main Town Center	Ganemulla
2. UP Code	UP 54
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	0.17 km ²
5. Urban Area 2021:	0.38 km ²
6. UG (2013—2017):	0.17 km ²
7. UG (2017—2021):	0.2 km ²
8. RUG (2013—2017):	0.04 km ² /Year
9. RUG (2017—2021):	0.05 km ² /Year
11. MC/UC Area:	−
12. Status:	Not-within

10. MC/UC Name:

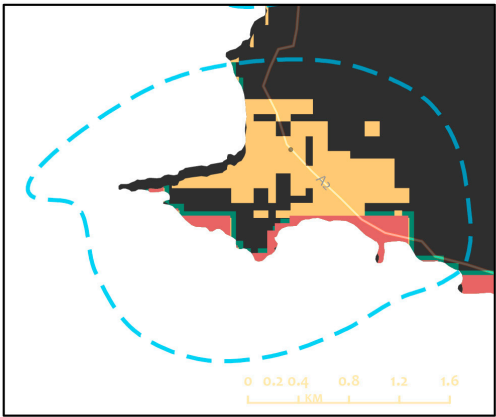
−

13. Cities and Towns in 2013

14. Cities and Towns added by 2017
Ganemulla

15. Cities and Towns added by 2021

S



1. Main Town Center	Mirissa
2. UP Code	UP 55
3. Urban Area 2013:	0.17 km ²
4. Urban Area 2017:	0.94 km ²
5. Urban Area 2021:	0.27 km ²
6. UG (2013—2017):	0.77 km ²
7. UG (2017—2021):	−0.67 km ²
8. RUG (2013—2017):	0.19 km ² /Year
9. RUG (2017—2021):	−0.17 km ² /Year
11. MC/UC Area:	−
12. Status:	Not-within

10. MC/UC Name:

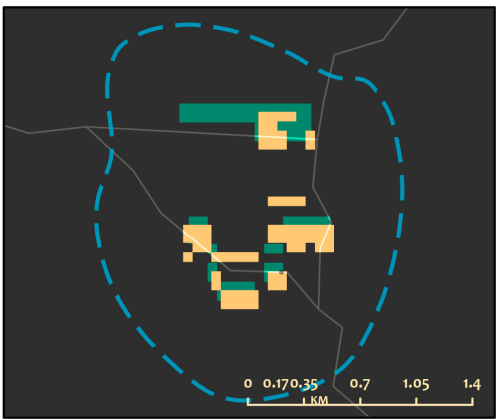
−

13. Cities and Towns in 2013

14. Cities and Towns added by 2017
Mirissa

15. Cities and Towns added by 2021

S



1. Main Town Center	Tissamaharama
2. UP Code	UP 56
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	0.22 km ²
5. Urban Area 2021:	0.26 km ²
6. UG (2013—2017):	0.22 km ²
7. UG (2017—2021):	0.04 km ²
8. RUG (2013—2017):	0.05 km ² /Year
9. RUG (2017—2021):	0.01 km ² /Year
11. MC/UC Area:	−
12. Status:	Not-within

10. MC/UC Name:

−

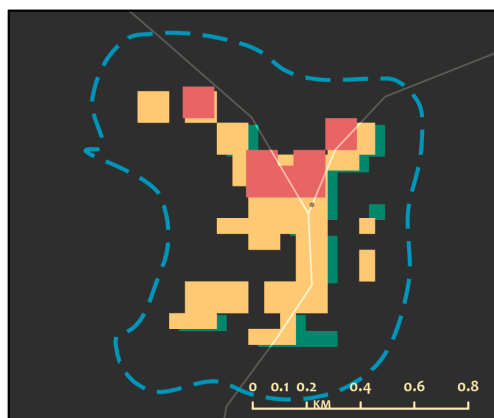
13. Cities and Towns in 2013

14. Cities and Towns added by 2017
Tissamaharama

15. Cities and Towns added by 2021

S

Figure 12. Cont.



1. Main Town Center	Narammala
2. UP Code	UP 57
3. Urban Area 2013:	0.07 km ²
4. Urban Area 2017:	0.29 km ²
5. Urban Area 2021:	0.17 km ²
6. UG (2013–2017):	0.21 km ²
7. UG (2017–2021):	−0.11 km ²
8. RUG (2013–2017):	0.05 km ² /Year
9. RUG (2017–2021):	−0.03 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

10. MC/UC Name:

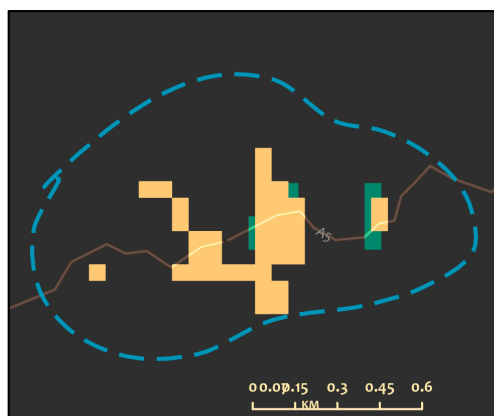
S

-

13. Cities and Towns in 2013

14. Cities and Towns added by 2017
Narammala

15. Cities and Towns added by 2021



1. Main Town Center	Passara
2. UP Code	UP 58
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	0.13 km ²
5. Urban Area 2021:	0.05 km ²
6. UG (2013–2017):	0.13 km ²
7. UG (2017–2021):	−0.08 km ²
8. RUG (2013–2017):	0.03 km ² /Year
9. RUG (2017–2021):	−0.02 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

10. MC/UC Name:

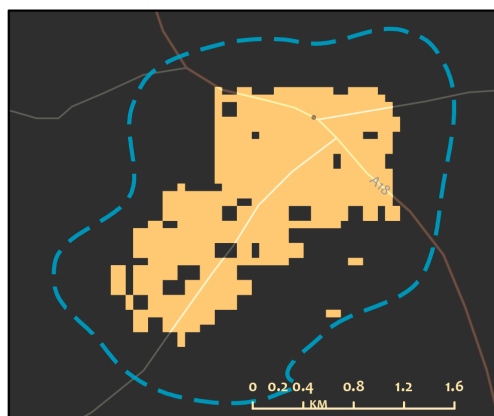
S

-

13. Cities and Towns in 2013

14. Cities and Towns added by 2017
Passara

15. Cities and Towns added by 2021



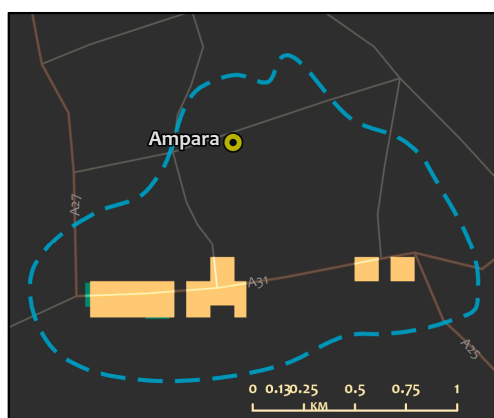
1. Main Town Center	Embilipitiya
2. UP Code	UP 59
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	2.3 km ²
5. Urban Area 2021:	0.05 km ²
6. UG (2013–2017):	2.30 km ²
7. UG (2017–2021):	−2.25 km ²
8. RUG (2013–2017):	0.57 km ² /Year
9. RUG (2017–2021):	−0.56 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

10. MC/UC Name:

S

-

13. Cities and Towns in 2013

14. Cities and Towns added by 2017
Embilipitiya15. Cities and Towns added by 2021
Embilipitiya

1. Main Town Center	Ampara
2. UP Code	UP 60
3. Urban Area 2013:	0 km ²
4. Urban Area 2017:	0.16 km ²
5. Urban Area 2021:	0.02 km ²
6. UG (2013–2017):	0.16 km ²
7. UG (2017–2021):	−0.13 km ²
8. RUG (2013–2017):	0.04 km ² /Year
9. RUG (2017–2021):	−0.03 km ² /Year
11. MC/UC Area:	-
12. Status:	Not-within

10. MC/UC Name:

S

-

13. Cities and Towns in 2013

14. Cities and Towns added by 2017

15. Cities and Towns added by 2021
Ampara

Figure 12. Cont.

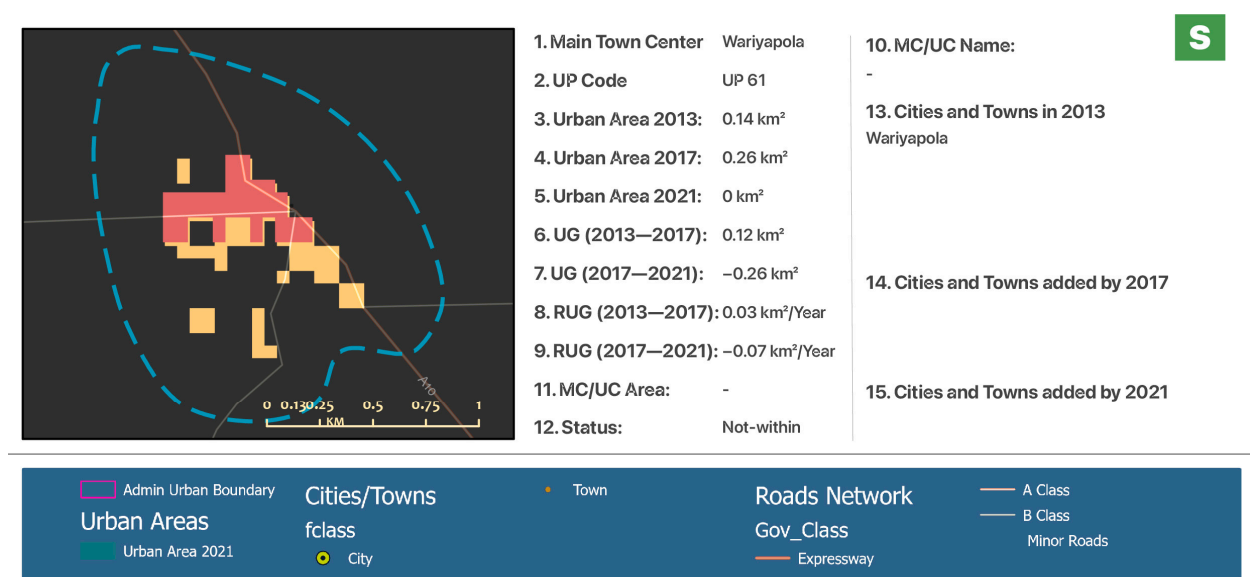


Figure 12. Maps of identified 60 UPs in Sri Lanka with UGs and RUGs.

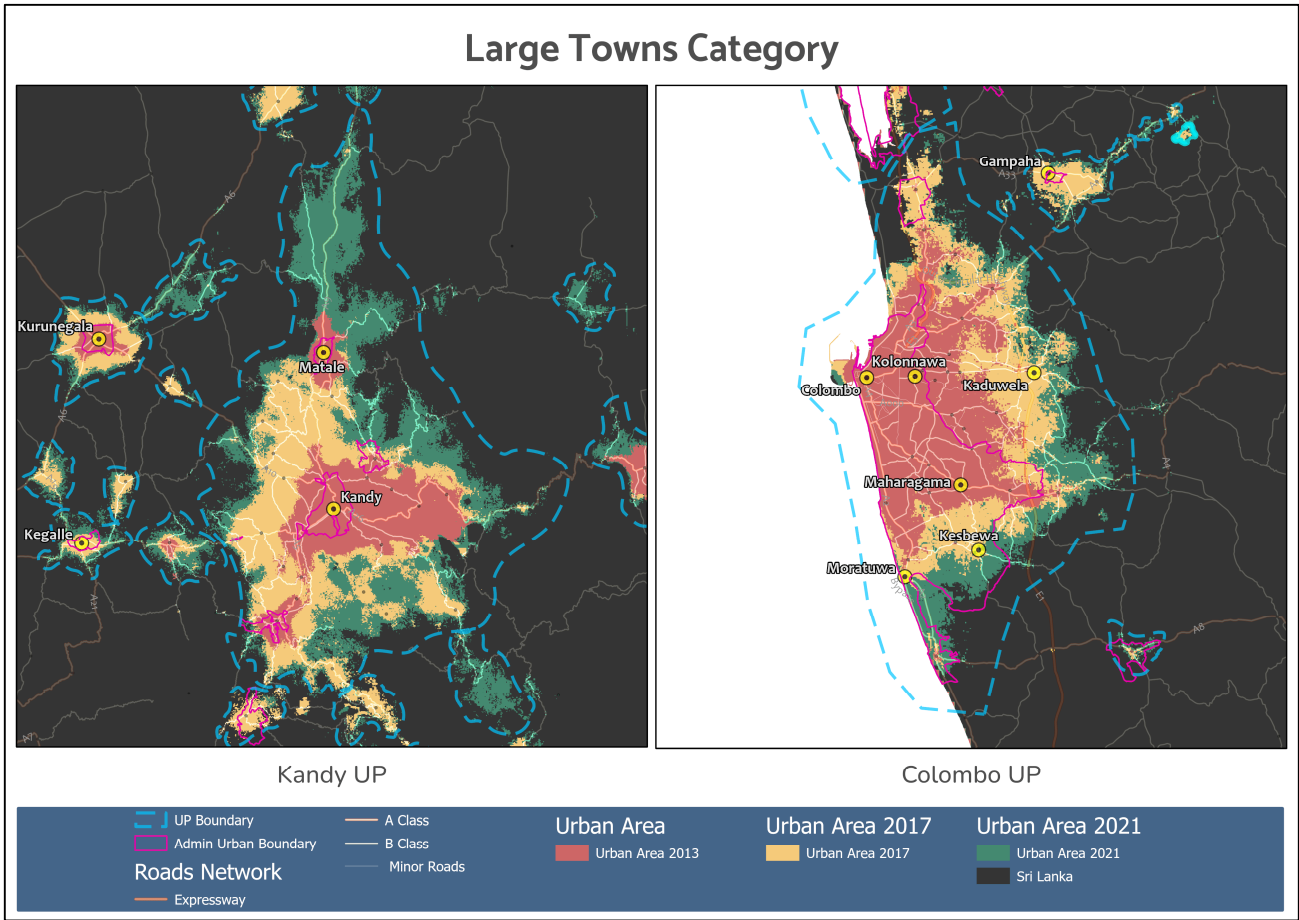


Figure 13. UPs identified as large towns in Sri Lanka.

4.1.1. Large Towns (Urban Area (km²) > 250)

Only two UPs were identified in Sri Lanka to be in the category of large towns. They were Kandy (UP 1) and Colombo (UP 2). These can be considered as renowned urban metro regions in Sri Lanka. Even though both UPs had a larger urban area extent and

higher RUG, the analysis indicated that the RUG slowed down in the two considered time periods—from 2013 to 2017 (Period 1) and from 2017 to 2021 (Period 2). Further, both UPs have well expanded beyond their current administrative boundaries—the MC and UC areas integrating the surrounding urban areas into forming urban conurbations. By 2021, the Colombo UP and Kandy UP accounted for 15.51% and 35.65% of the country's total urban area, respectively. Therefore, around 57% of the country's urban area was located inside these two major UPs.

However, the Colombo UP showed a contiguous urban development along the main arteries, where the Kandy UP was surrounded by many other UPs and no contiguous urban growth was observed. This is where the data fusion approach enters to rectify the misinterpretations made so far in identifying urban areas, either based on administrative boundaries or by built-up footprints. Instead, the data fusion approach fused data sources reflecting both visible and invisible urban infrastructure, i.e., schools, hospitals, cinemas, ATMs or banking facilities, mobile coverage, and so on.

4.1.2. Medium-Sized Towns ($250 > \text{Urban Area (km}^2) > 20$)

Figure 14 shows the medium-sized towns identified. Dambulla (UP 3), Bandarawela (UP 4), Anuradhapura (UP 5), Welimada (UP 6), Polonnaruwa (UP 7), Nuwara Eliya (UP 8), Kurunegala (UP 9), Vavuniya (UP 10), Badulla (UP 11), Mahiyanganaya (UP 12), Negombo (UP 13), Hingurakgoda (UP 14), Galle (UP 15), Rikillagaskada (UP 16), Mawanella (UP 17), Galewela (UP 18), Kegalle (UP 19), Hatton (UP 20), Ibbagamuwa (UP 21), Jaffna (UP 22), and Gampaha (UP 23) were considered as medium-sized towns based on their RUG and urban area. This category has an urban area extent ranging from 20 km² to 250 km².

These medium-sized towns has the characteristic of being formed around a central urban core. Unlike the Colombo and Kandy large UPs and except for the Negombo, Kurunegala, Badulla, Nuwara Eliya, and Gampaha medium-sized UPs, the balanced towns experienced higher urbanization growth during period 1 (2013–2017) and period 2 (2017–2021). Most importantly, 4 of the 21 medium-sized UPs—Dambulla, Welimada, Polonnaruwa, Nuwara Eliya, Vavuniya, Mahiyanganaya, Rikillagaskada, Higurakgoda, Rikillagaskada, Mawanella, Galewela, and Ibbagamuwa—were not even located within the official urban boundaries identified by the DCS. They were meant to be non-urban areas based on the DCS definition. This shows that the DCS definitions have not considered the political and socio-economic changes that have happened over the considered time period. For instance, after the end of the civil war, Jaffna and Vavuniya experienced rapid urban growth, which could be the reason behind identifying them as urban areas because of the fusion exercise.

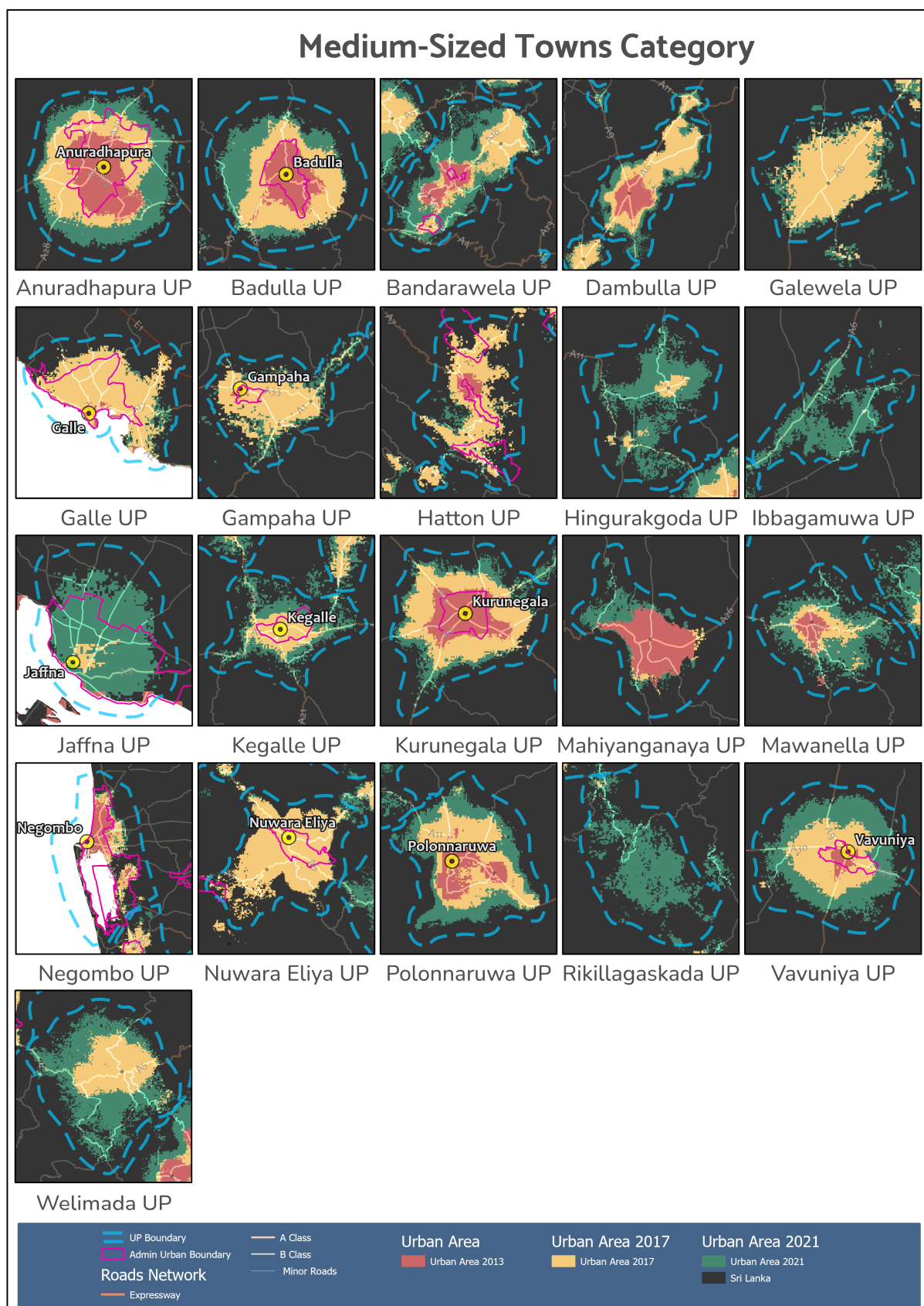


Figure 14. UPs identified as medium-sized towns in Sri Lanka.

4.1.3. Small Towns

UPs with a small RUG variation and urban area extent were identified as small towns. These UPs had an urban area of less than 20 km². Out of the 60 UPs, 37 UPs were categorized as small towns. Figure 15 shows all the identified small towns in Sri Lanka based on the data fusion approach. The majority of the small towns identified in this study do not belong to the current definition of urban areas presented by the DCS.

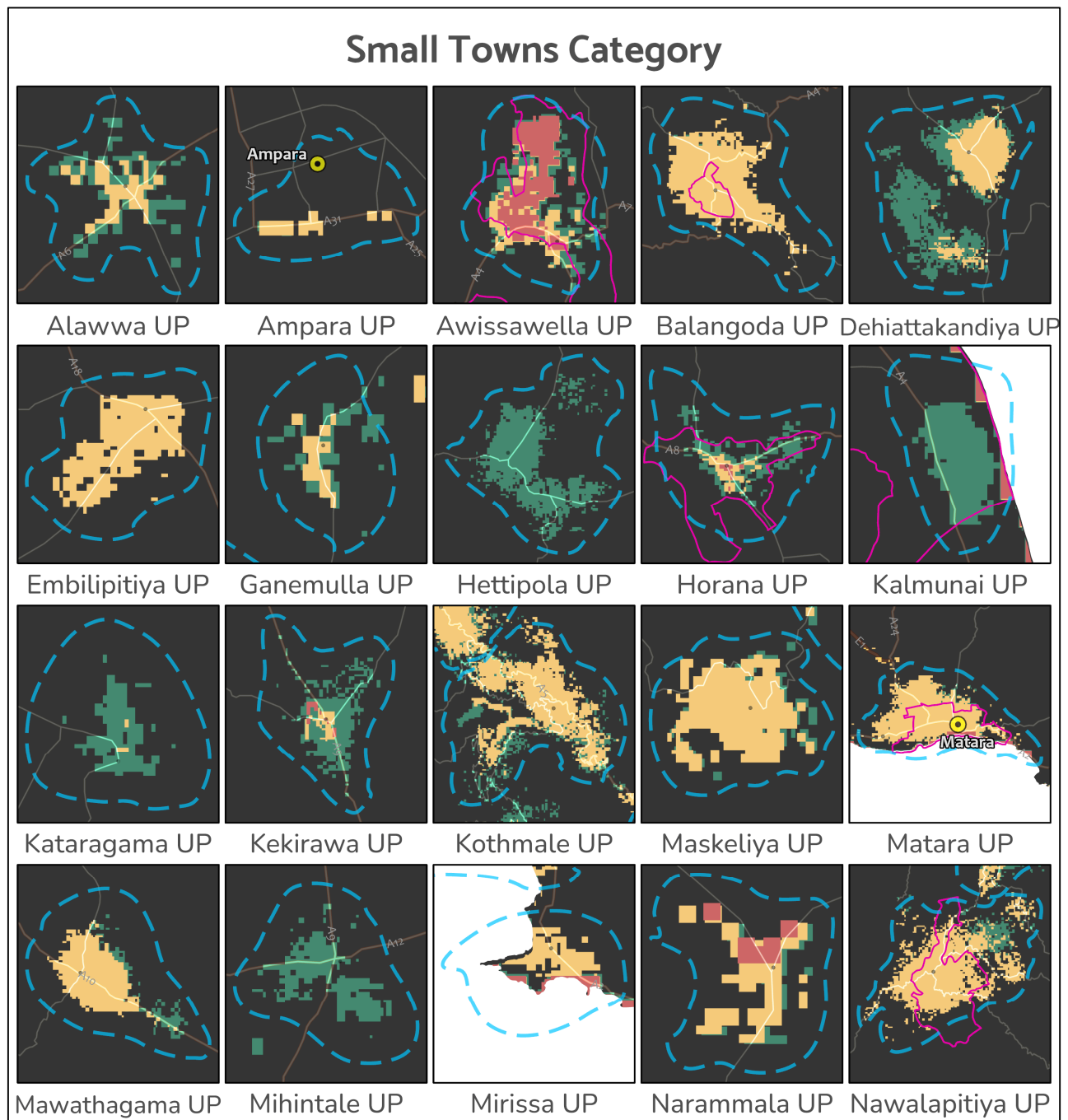


Figure 15. Cont.

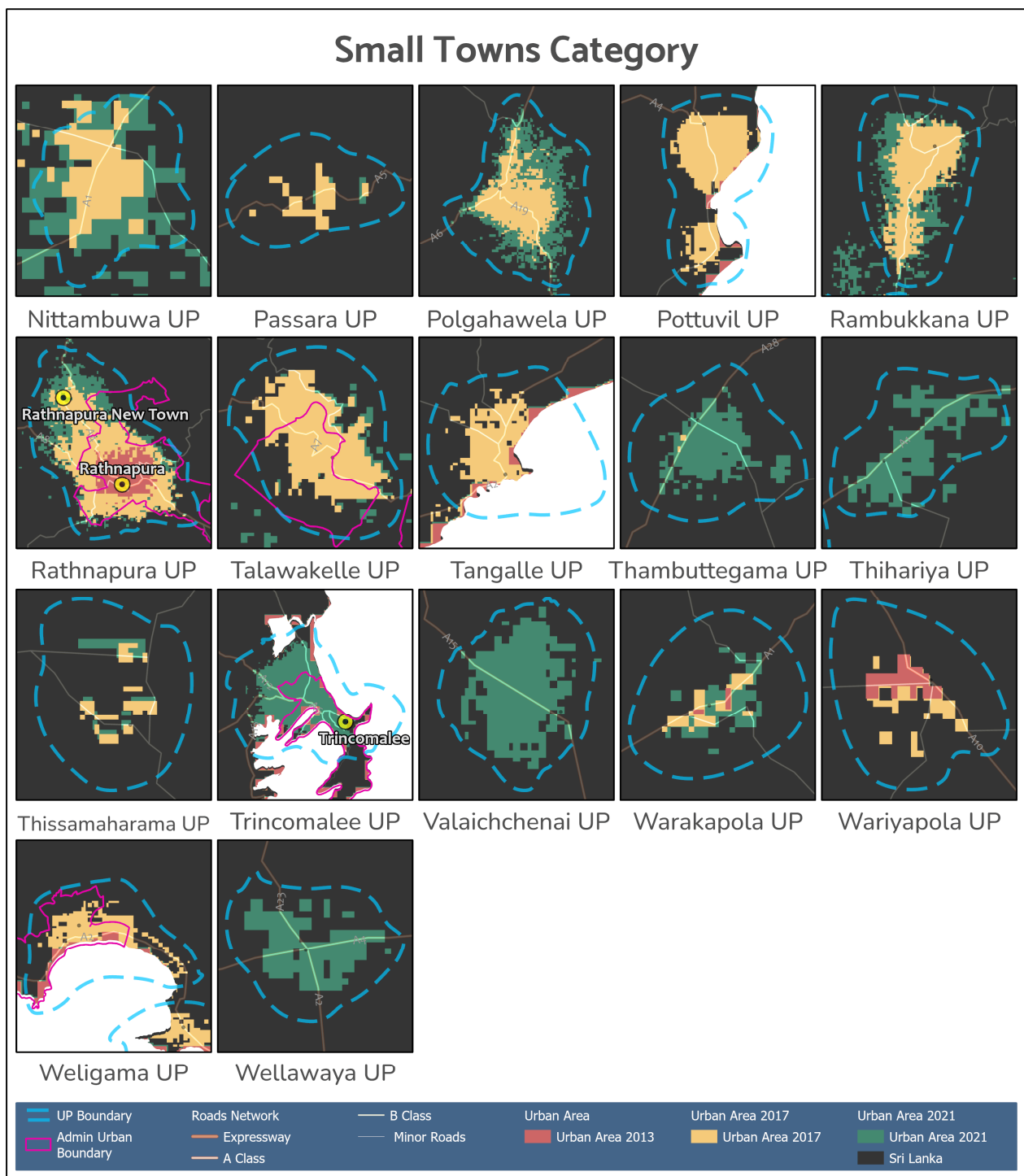


Figure 15. UPs identified as small towns in Sri Lanka.

4.2. Assessment of the Accuracy of the Findings

To assess the accuracy of the identified urbanization pattern, a reliable reference dataset is needed, which will enable comparing and contrasting the differences between the urban areas identified through this data fusion approach with the already established areas. However, in the absence of such a dataset, this study adopted a limited validation, which was employing satellite imagery to visually assess the accuracy of the urban areas identified through data fusion.. Figure 16 shows the Colombo and Kandy urban areas in 2021.

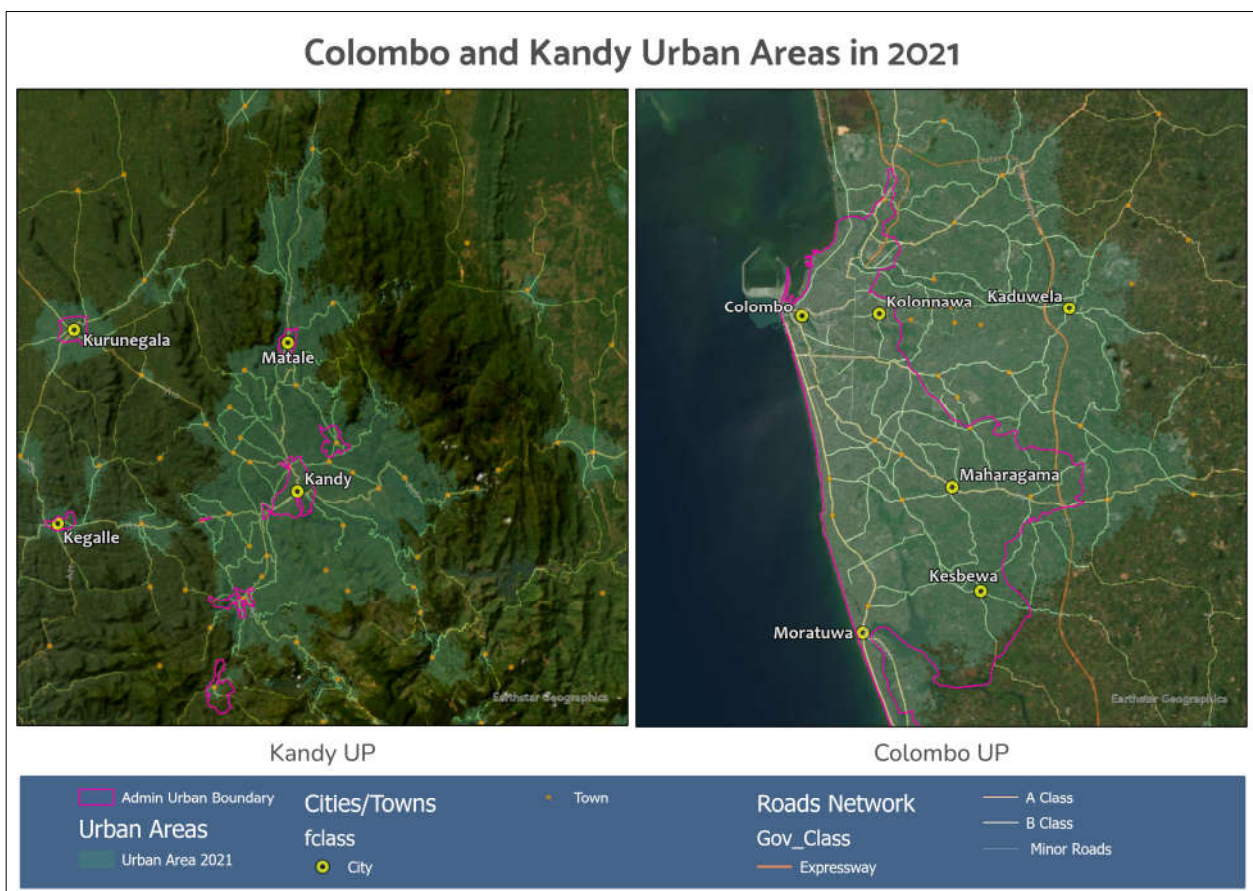


Figure 16. Colombo and Kandy urban areas in 2021.

When examining Colombo (UP 2), a similar pattern was revealed between the urban areas identified by the data fusion model and those observable in the satellite images. This alignment hinted that the data fusion approach was able to accurately capture not only the urban sprawl of Colombo, but also the urban conurbation of Colombo, showcasing its effectiveness as a tool for urban area determination. The analysis of Kandy (UP 1) was challenging due to its diverse topography. Although the data fusion approach occasionally extended into less inhabitable hilly areas, the identified urban area closely resembled the actual urban footprint captured in the satellite overlays and NTL images. This finding was significant, as it demonstrated the data fusion approach's capacity to adapt and provide meaningful insights into urban dynamics, even in geographically complex areas.

The statistical analysis further highlights the effectiveness of the data fusion approach and its potential as a suitable approach for capturing the dynamic urbanization patterns in low- and middle-income countries. For instance, compared to the urban area extent identified by the DCS, Colombo UP (UP 2) increased by 232.12, 604.76, and 1072.62 square kilometers for the years of 2013, 2017, and 2021, respectively. Similarly, the Kandy UP (UP 1) was 183.58, 331.82, and 447.03 square kilometers larger than the urban area extent identified by the DCS in the same time periods. These figures highlight the suitability of the data fusion approach in delineating urban areas to obtain a broader understanding of the dynamic urban growth of a country. Table 5 shows the underestimation/overestimation of urban areas identified through data fusion.

Table 5. Underestimation/overestimation of urban areas identified through data fusion.

UP	Main Town Center	MC/UC		Area (Approx.)			Difference in Urban Area (Admin Urban Area—UP Area)		
		Name	Area	2013	2017	2021	2013	2017	2021
UP 1	Kandy	Kandy MC, Gampola UC, Wattegama UC	47.70	279.82	652.46	1120.32	−232.12	−604.76	−1072.62
UP 2	Colombo	Colombo MC	40.21	223.79	372.03	487.24	−183.58	−331.82	−447.03

Even though the data fusion approach showed accuracy of an acceptable level, there are further areas that need to be improved. Yet, this data fusion approach has the relative advantages of quick adaptation, usability in situations where updated census data are scarce, especially low- and middle-income countries, and revealing the urban lifestyles of citizens that extend beyond formally demarcated administrative boundaries. Based on the context of use and data availability, this method is useful for understanding the dynamic nature of urbanization.

5. Discussion

Urbanization is a complex process that occurs at different scales, from the local to the regional level [23,75]. Understanding the urbanization process often helps authorities to appropriate urban development policies and infrastructure development projects. The use of one data source does not allow for capturing the full extent of the urbanization process and might not provide adequate details to understand the process at the local level. Urban areas are mostly delineated based on building footprints and density, but not on urban facilities such as mobile network coverage and accessibility to facilities such as schools, cinemas, and hospitals, etc. To overcome said drawbacks, this study suggested the data fusion approach using openly available data. The study identified 2 UPs as large towns, 21 UPs as medium-sized towns, and 37 UPs as small towns. Accordingly, the majority of the urban characters were small towns.

When it comes to the growth of all UPs, the Colombo and Kandy UPs had the highest urban area extents from 2013 to 2021. These seemed to have grown over the last few decades into urban metro regions forming conurbations assimilating adjacent townships and localities. Colombo port was the catalyst for it to rapidly expand as the commercial capital of the country. Unlike the other medium-sized towns of Anuradhapura, Polonnaruwa, and Kurunegala, which are historical cities, Kandy continued to grow, even after its attraction in the socio-political context ceased, because of the continuation of the prominence that it received under British ruling of the island.

Except for Welimada, Nuwara Eliya, Higurakgoda, Rikillagaskada, Galewela, Kegalle, and Ibbagamuwa, all the other medium-sized towns—Dambulla, Bandarawela, Anuradhapura, Polonnaruwa, Kurunegala, Vavuniya, Badulla, Mahiyanganaya, Negombo, Galle, Mawanella, Hatton, Jaffna, and Gampaha Ups—had an urban footprint since 2013. Among them, Dambulla (Overall RUG = 23.467), Bandarawela (Overall RUG = 18.44), and Anuradhapura (Overall (RUG) = 11.00) recorded the highest RUGs among the medium-sized towns. Anuradhapura was an ancient capital city of the island that lost its prominence due to the shift of the monarchy to other locations. But from the early 1950s, it has continued to grow as a famous pilgrim destination with the government's implementation of the Anuradhapura Sacred Area Plan. Dambulla is predominantly a commercial center with Sri Lanka's largest wholesale produce market. It is also famous for the historical Royal Cave Temple, which is a popular tourist destination.

When compared with the study findings, it seems that the current delineation of urban areas has not integrated an adequate understanding of the real ground dynamics. Thus, the urban extents of these areas may need to be revisited to guide the direction of development and to provide the needed infrastructure and basic services, assuring that

they will encompass environments conducive for human habitation, business promotion, and proper urban functions.

Small towns can be identified as being in the infant stage of an urbanity that might grow into a metropolitan region of the country in the future. However, the life cycle of a small town will change based on the internal and external political and socio-economic factors that might emerge in and around the town. For instance, the Kandy UP, which was identified as a large-scale town, emerged as a riverine small town. Hence, the urban character of 'small towns' is not exclusive to Sri Lanka, it is evident in numerous other countries such as Bhutan (Paro), Malta (Mdina), Luxembourg (Vianden), and Andorra (Ordino).

The positive RUGs in Trincomalee, Dambulla, Sigiriya, Kurunegala, Mawanella, and Kegalle can be attributed the potential for the development of the proposed Colombo–Trincomalee economic corridor in Sri Lanka (NPP, 2019). This corridor is aimed at enhancing economic activities and connectivity, which will lead to improving infrastructure and transportation networks, attracting businesses and industries along the route.

This study indicates that, in Sri Lanka, the urbanization process, pattern, and rate are governed by both natural and behavioral factors and socio-political and economic forces. This phenomenon may be common to many other nation states of the same category and with similar conditions supportive of urbanization. Accordingly, understanding urbanization cannot be simply confined to instantaneous and periodic statistical information and static boundaries. Hence, new approaches that capture the ground realities and dynamics of urbanization are required to serve planning and development purposes. At the same time, the trend patterns of urbanization demand robust planning and development policies, guiding frameworks, and strategies to circumvent adverse impacts and ensure sustainable urban growth in the future.

The findings in this context of a case study of the nation of Sri Lanka revealed that the proposed approach is invaluable in determining urbanization patterns adequately and efficiently, despite some limitations mentioned in the next section. The approach also holds strong value to be adopted in other low- and middle-income country contexts. This methodology is especially beneficial in low- and middle-income countries where conventional census data may be limited or outdated. The approach's high efficiency and thorough data analysis make it an effective tool for urban planning and policy development. The significance of this issue extends beyond Sri Lanka, as several low- and middle-income countries encounter comparable difficulties in monitoring and controlling urbanization. Implementing this approach can greatly improve the comprehension of growth patterns, facilitating better decision making and promoting sustainable solutions for urban development.

However, this study also showed a few technical limitations. Some obvious urban areas such as Kalutara, Beruwala/Aluthgama, Minuwangoda, Batticaloa, Kaththankudy, Chilaw, Puttalam, Kilinochchi, Medawachchiya, Chavakachcheri, and Point Pedro seem to have gone missing, while a few less obvious smaller urban areas popped out. For instance, the UPs reflect a high NTL and POI density in the Kalutara area. The limited UP coverage likely resulted from NTL's 500 m resolution, indicating a resolution constraint in the data fusion dataset.

Further, Figures 17 and 18 elaborates the Colombo and Kandy urban areas in 2021 and Comparison of fusion results with the real ground realities—cases of Batticaloa and Kaththankudy.

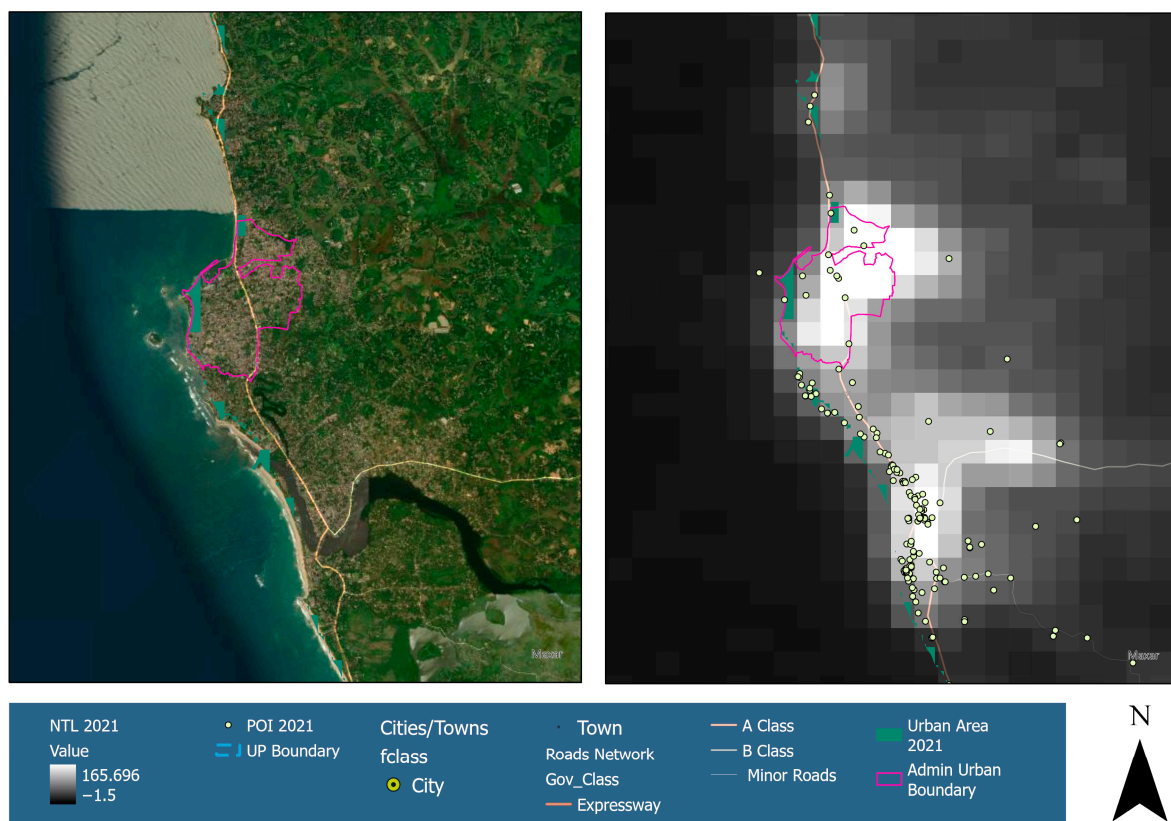


Figure 17. Colombo and Kandy urban areas in 2021.

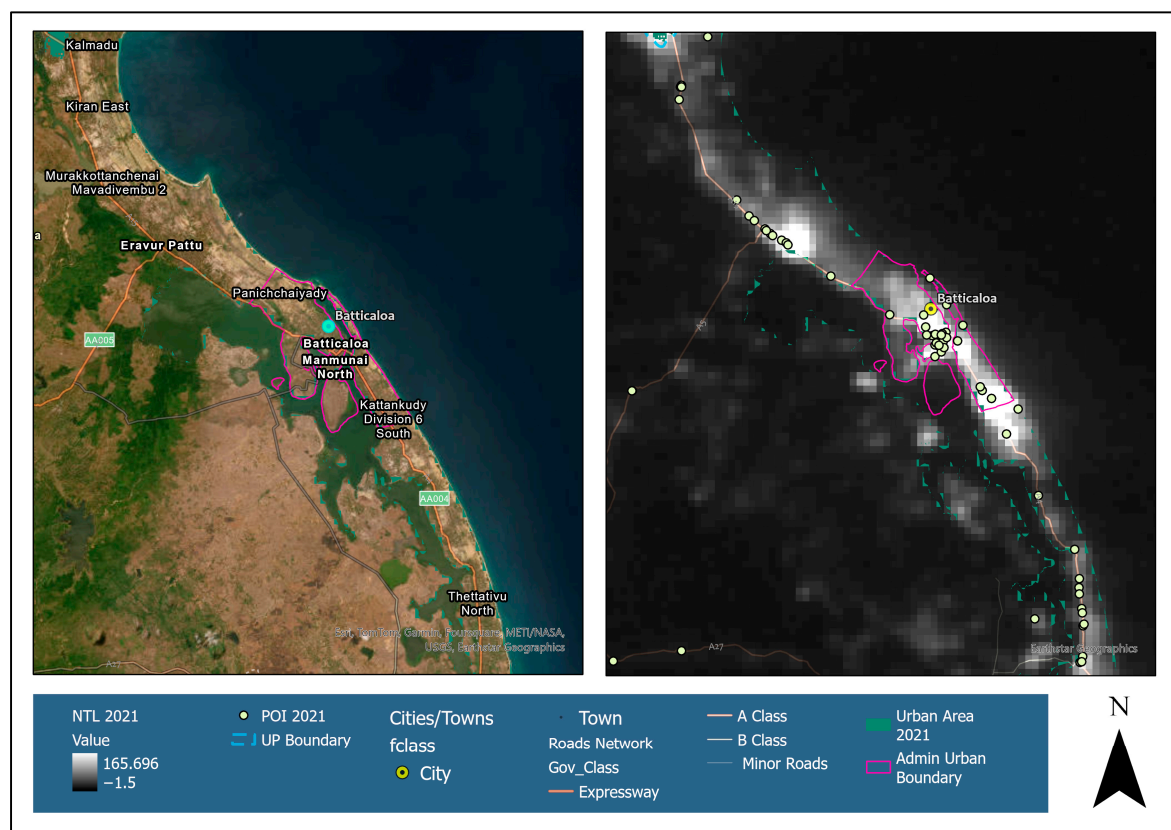


Figure 18. Comparison of fusion results with the real ground realities—cases of Batticaloa and Kaththankudy.

In the cases of both Batticaloa and Kaththankudy, small urban patches were identified by the data fusion approach in areas which had a higher concentration of NTL and POI. However, the results were not prominent, as the study was carried out at the national scale. Therefore, their neglected potentiality was most likely due to the scale factor that the study worked with. Fusing different data sources such as building heights and giving them different weights might lead to bridging the gap between the real ground scenarios and the outputs of the data fusion methodology to identify urban areas.

6. Conclusions

The study reported in this paper aimed to develop an effective approach to determining urbanization patterns through the novel big data fusion approach using multiple data sources that provide reliable information on urbanization activities, particularly in the context of low- and middle-income countries, where urbanization challenges, e.g., development control and disasters risks, are higher.

The study findings, considering the testbed case of Sri Lanka, challenged the current definitions used to delineate urban areas and the methods used to understand urbanization patterns. For instance, from the demarcated urban areas delineated by the DCS, 21.6% were considered as under-bounded, 15% were over-bounded, and 63.3% were not within the urban areas defined by the DCS. As urbanization or urban areas help to understand the development status of a country, inaccurate and inappropriate understandings about the urbanization level would misguide all decisions related to the development of a country. Accordingly, this study emphasizes the need to move towards a data-driven approach to delineate urban areas and understand the dynamic nature of urbanization processes. In particular, the extant definitions used to identify urban areas have failed to understand the role of small towns in shaping the future urbanization pattern of a country.

The study used different data sources, which were employed to trace different functional dimensions of cities, i.e., mobile 4G network coverage data, crowdsourced data, and POI data. Cities are evolving and expanding while adding new layers to these cities. The data sources used to understand the urban extents and present urbanization process may need to be changed in future, and such sources must also be accessible without constraints. Yet, the approach adopted in this study based on big data fusion using openly available data is a suitable way to understand the urbanization pattern of a country. While the findings primarily concern Sri Lanka, the method is also transferable to many other low- and middle-income country contexts. However, the adoption of the novel approach proposed in this paper might need careful tailoring to other national circumstances, e.g., considering planning regulations, governance system differences, data availability, and so on.

This study also has a few limitations. First, the study segmented images using the accurate LOT algorithm from the OpenCV library. However, it was indicated that image segmentation using a Fully Convolutional Neural Network (FCNN) would have produced significantly more accurate findings for identifying urban areas. Second, the POI dataset was used with equal weights for each category, but allocating higher weights for locations that are more closely associated with urban area will be able to help with the accuracy of identifying urban areas. Third, only spatial verification and precision verification could be conducted in this study. However, ground verification would have significantly contributed to the verification results. Fourth, the method struggled to clearly distinguish the urban pattern in the coastal belt, possibly due to the national scale of the study. For instance, coastal towns like Panadura and Kalutara were not effectively captured using this methodology. Finally, there was an uncertainty associated with the boundaries, as these may change with changing datasets and the accuracy levels of the fusion methodology. Our prospective research will concentrate on addressing these limitations and fine-tuning the approach for applicability to other low-, middle-, and high-income country contexts.

Author Contributions: Conceptualization, N.K.; methodology, formal analysis, investigation, data curation, writing—original draft preparation, N.A. and N.K.; writing—review and editing, J.M. and T.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. UN-DESA. *World Urbanization Prospects: The 2014 Revision*; UN-DESA: New York, NY, USA, 2015.
2. Baffour, B.; King, T.; Valente, P. The modern census: Evolution, examples and evaluation. *Int. Stat. Rev.* **2013**, *81*, 407–425. [CrossRef]
3. Weeks, J.R. Defining Urban Areas. In *Remote Sensing of Urban and Suburban Areas*; Springer: Dordrecht, The Netherlands, 2010; pp. 33–45. [CrossRef]
4. Insee. Urban Unit. Available online: <https://www.insee.fr/en/metadonnees/definition/c1501> (accessed on 9 January 2020).
5. Coelho, K.; Sood, A. Urban studies in India across the millennial turn: Histories and futures. *Urban Stud.* **2022**, *59*, 2613–2637. [CrossRef]
6. Thakuriah, P.; Tilahun, N.Y.; Zellner, M. Big data and urban informatics: Innovations and challenges to urban planning and knowledge discovery. In *Seeing Cities Through Big Data*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 11–45.
7. Butler, L.; Yigitcanlar, T.; Paz, A. Smart urban mobility innovations: A comprehensive review and evaluation. *IEEE Access.* **2020**, *29*, 196034–196049. [CrossRef]
8. Wirth, L. Urbanism as a way of life. *Am. J. Sociol.* **1938**, *44*, 1–24. [CrossRef]
9. Abeynayake, T.; Meetiayagoda, L.; Kankanamge, N.; Mahanama, P.K.S. Imageability and legibility: Cognitive analysis and visibility assessment in Galle heritage city. *Int. J. Archit. Urban* **2022**, *46*, 126–136. [CrossRef]
10. Goonetilleke, A.; Yigitcanlar, T.; Ayoko, G.; Egodawatta, P. *Sustainable Urban Water Environment: Climate, Pollution and Adaptation*; Edward Elgar: Cheltenham, UK, 2014.
11. Ioppolo, G.; Cucurachi, S.; Salomone, R.; Shi, L.; Yigitcanlar, T. Integrating strategic environmental assessment and material flow accounting: A novel approach for moving towards sustainable urban futures. *Int. J. Life Cycle Assess.* **2019**, *24*, 1269–1284. [CrossRef]
12. He, X.; Zhang, Z.; Yang, Z. Extraction of urban built-up area based on the fusion of night-time light data and point of interest data. *R. Soc. Open Sci.* **2021**, *8*, 210838. [CrossRef]
13. Yigitcanlar, T. *Sustainable Urban and Regional Infrastructure Development: Technologies, Applications and Management*; IGI Global: Hershey, PA, USA, 2010.
14. Balk, D.; Leyk, S.; Jones, B.; Montgomery, M.R.; Clark, A. Understanding urbanization: A study of census and satellite-derived urban classes in the United States, 1990–2010. *PLoS ONE* **2018**, *13*, e0208487. [CrossRef]
15. Firman, T. Demographic patterns of Indonesia’s Urbanization, 2000–2010: Continuity and change at the macro level. *Demogr. Transform. Socio Econ. Dev.* **2015**, *5*, 255–269. [CrossRef]
16. Zhang, B.; Zhang, J.; Miao, C. Urbanization level in Chinese counties: Imbalance pattern and driving force. *Remote Sens.* **2022**, *14*, 2268. [CrossRef]
17. Abesinghe, S.; Kankanamge, N.; Yigitcanlar, T.; Pancholi, S. Image of a City through Big Data Analytics: Colombo from the Lens of Geo-Coded Social Media Data. *Future Internet* **2023**, *15*, 32. [CrossRef]
18. Jun, Z.; Xiao-Die, Y.; Han, L. The extraction of urban built-up areas by integrating night-time light and POI data: A case study of Kunming, China. *IEEE Access* **2021**, *9*, 22417–22429. [CrossRef]
19. Li, K.; Chen, Y. A genetic algorithm-based urban cluster automatic threshold method by combining VIIRS DNB, NDVI, and NDBI to monitor urbanization. *Remote Sens.* **2018**, *10*, 277. [CrossRef]
20. Ellis, P.; Roberts, M. *Leveraging Urbanization in South Asia: Managing Spatial Transformation for Prosperity and Livability*; World Bank Group: Chicago, IL, USA, 2016. [CrossRef]
21. Yigitcanlar, T.; Fabian, L.; Coiactetto, E. Challenges to urban transport sustainability and smart transport in a tourist city: The Gold Coast, Australia. *Open Transp. J.* **2008**, *2*, 29–46. [CrossRef]
22. He, X.; Cao, Y.; Zhou, C. Evaluation of polycentric spatial structure in the urban agglomeration of the Pearl River Delta (PRD) based on multi-source big data fusion. *Remote Sens.* **2021**, *13*, 3639. [CrossRef]
23. Priyashani, N.; Kankanamge, N.; Yigitcanlar, T. Multisource Open Geospatial Big Data Fusion: Application of the Method to Demarcate Urban Agglomeration Footprints. *Land* **2023**, *12*, 407. [CrossRef]
24. Yigitcanlar, T.; Guaralda, M.; Taboada, M.; Pancholi, S. Place making for knowledge generation and innovation: Planning and branding Brisbane’s knowledge community precincts. *J. Urban Technol.* **2016**, *23*, 115–146. [CrossRef]
25. Mortoja, M.; Yigitcanlar, T.; Mayere, S. How does peri-urbanization trigger climate change vulnerabilities? An investigation of the Dhaka megacity in Bangladesh. *Remote Sens.* **2020**, *12*, 3938. [CrossRef]
26. Mortoja, M.; Yigitcanlar, T.; Mayere, S. What is the most suitable methodological approach to demarcate peri-urban areas? A systematic review of the literature. *Land Use Policy* **2020**, *95*, 104601. [CrossRef]
27. Chen, Y.; Zhang, J. Extraction of urban built-up areas based on data fusion: A case study of Zhengzhou, China. *Int. J. Geo-Inf.* **2022**, *11*, 521. [CrossRef]
28. Sanchez, T. Planning on the Verge of AI, or AI on the Verge of Planning. *Urban Sci.* **2023**, *7*, 70. [CrossRef]

29. McGranahan, G.; Satterthwaite, D. Urbanization Concepts and Trends. IIED. 2014. Available online: <http://pubs.iied.org/10709IIED> (accessed on 18 April 2024).
30. Trask; Sherif, B. *Migration, Urbanization, and the Family Dimension*; United Nations Department of Economic and Social Affairs (UN-DESA): New York, NY, USA, 2022.
31. Weeraratne, B. Can We Produce Better Estimates of Urbanization in Sri Lanka? Available online: <https://www.ips.lk/talkingeconomics/2016/04/05/can-we-produce-better-estimates-of-urbanization-in-sri-lanka/> (accessed on 5 April 2016).
32. Rahman, M.; Mohiuddin, H.; Kafy, A.; Sheel, P.; Di, L. Classification of cities in Bangladesh based on remote sensing derived spatial characteristics. *J. Urban Manag.* **2019**, *8*, 206–224. [\[CrossRef\]](#)
33. United States Census Bureau. Urban and Rural. Available online: <https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html> (accessed on 9 January 2023).
34. Pandey, B.; Joshi, P.; Seto, K.C. Monitoring urbanization dynamics in India using DMSP/OLS night time lights and SPOT-VGT data. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 49–61. [\[CrossRef\]](#)
35. Yu, B.; Tang, M.; Wu, Q.; Yang, C.; Deng, S.; Shi, K.; Chen, Z. Urban built-up area extraction from log-transformed NPP-VIIRS nighttime light composite data. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1279–1283. [\[CrossRef\]](#)
36. Zhou, Y.; He, X.; Zhu, Y. Identification and Evaluation of the Polycentric Urban Structure: An Empirical Analysis Based on Multi-Source Big Data Fusion. *Remote Sens.* **2022**, *14*, 2705. [\[CrossRef\]](#)
37. Ma, X.; Li, C.; Tong, X.; Liu, S. A new fusion approach for extracting urban built-up areas from multisource remotely sensed data. *Remote Sens.* **2019**, *11*, 2516. [\[CrossRef\]](#)
38. Xu, T.; Coco, G.; Gao, J. Extraction of urban built-up areas from nighttime lights using artificial neural network. *Geocarto Int.* **2018**, *35*, 1049–1066. [\[CrossRef\]](#)
39. Hu, T.; Yang, J.; Li, X.; Gong, P. Mapping urban land use by using Landsat images and open social data. *Remote Sens.* **2016**, *8*, 151. [\[CrossRef\]](#)
40. Kuc, G.; Chormański, J. Sentinel-2 imagery for mapping and monitoring imperviousness in urban areas. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2019**, *42*, 43–47. [\[CrossRef\]](#)
41. Ma, T.; Xu, T.; Huang, L.; Zhou, A. A human settlement composite index (HSCI) derived from nighttime luminosity associated with imperviousness and vegetation indexes. *Remote Sens.* **2018**, *10*, 455. [\[CrossRef\]](#)
42. Zhang, Q.; Schaaf, C.; Seto, K.C. The Vegetation Adjusted NTL Urban Index: A new approach to reduce saturation and increase variation in nighttime luminosity. *Remote Sens. Environ.* **2013**, *129*, 32–41. [\[CrossRef\]](#)
43. Zhou, Y.; Smith, S.J.; Elvidge, C.D.; Zhao, K.; Thomson, A.; Imhoff, M. A cluster-based method to map urban area from DMSP/OLS nightlights. *Remote Sens. Environ.* **2014**, *147*, 173–185. [\[CrossRef\]](#)
44. Pan, Y.; Tian, Y.; Liu, X.; Gu, D.; Hua, G. Urban Big Data and the Development of City Intelligence. *Engineering* **2016**, *2*, 171–178. [\[CrossRef\]](#)
45. Zhang, J.; Zhang, X.; Tan, X.; Yuan, X. Extraction of urban built-up area based on deep learning and multi-sources data fusion—The application of an emerging technology in urban planning. *Land* **2022**, *11*, 812. [\[CrossRef\]](#)
46. Meng, C.; Yi, X.; Su, L.; Gao, J.; Zheng, Y. City-wide traffic volume inference with loop detector data and taxi trajectories. In Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Redondo Beach, CA, USA, 7–10 November 2017; pp. 1–10.
47. Gendron-Carrier, N.; Gonzalez-Navarro, M.; Polloni, S.; Turner, M.A. Subways and urban air pollution. *Am. Econ. J. Appl. Econ.* **2022**, *14*, 96–164. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Kankanamge, N.; Yigitcanlar, T.; Goonetilleke, A. How engaging are disaster management related social media channels? The case of Australian state emergency organisations. *Int. J. Disaster Risk Reduct.* **2020**, *48*, 101571. [\[CrossRef\]](#)
49. Kankanamge, N.; Yigitcanlar, T.; Goonetilleke, A.; Kamruzzaman, M. Determining disaster severity through social media analysis: Testing the methodology with South East Queensland Flood tweets. *Int. J. Disaster Risk Reduct.* **2020**, *42*, 101360. [\[CrossRef\]](#)
50. Kharrazi, A.; Qin, H.; Zhang, Y. Urban big data and sustainable development goals: Challenges and opportunities. *Sustainability* **2019**, *8*, 1293. [\[CrossRef\]](#)
51. Tu, W.; Zhu, T.; Xia, J.; Zhou, Y.; Lai, Y.; Jiang, J.; Li, Q. Portraying the spatial dynamics of urban vibrancy using multisource urban big data. *Comput. Environ. Urban Syst.* **2019**, *80*, 101428. [\[CrossRef\]](#)
52. Han, J.; Liu, J. Urban spatial interaction analysis using inter-city transport big data: A case study of the Yangtze river delta urban agglomeration of China. *Sustainability* **2018**, *10*, 4459. [\[CrossRef\]](#)
53. Wang, S.J.; Moriarty, P. A Human-Centered Perspective. In *Big Data for Urban Sustainability*; Springer: Cham, Switzerland, 2019.
54. Lu, Y.; Xu, S.; Liu, S.; Wu, J. An approach to urban landscape character assessment: Linking urban big data and machine learning. *Sustain. Cities Soc.* **2022**, *83*, 103983. [\[CrossRef\]](#)
55. Wang, M. Investigation of remote sensing image and big data analytic for urban garden landscape design and environmental planning. *Arab. J. Geosci.* **2021**, *14*, 473. [\[CrossRef\]](#)
56. Hu, S.; Gao, S.; Luo, W.; Wu, L.; Li, T.; Xu, Y.; Zhang, Z. Revealing intra-urban hierarchical spatial structure through representation learning by combining road network abstraction model and taxi trajectory data. *Ann. GIS* **2023**, *29*, 499–516. [\[CrossRef\]](#)
57. Zhuo, L.; Zheng, J.; Zhang, X.; Li, J.; Liu, L. An improved method of night-time light saturation reduction based on EVI. *Int. J. Remote Sens.* **2015**, *16*, 4114–4130. [\[CrossRef\]](#)

58. Cai, J.; Huang, B.; Song, Y. Using multi-source geospatial big data to identify the structure of polycentric cities. *Remote Sens. Environ.* **2017**, *202*, 210–221. [\[CrossRef\]](#)
59. Huang, X.; Yang, J.; Li, J.; Wen, D. Urban functional zone mapping by integrating high spatial resolution nighttime light and daytime multi-view imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 403–415. [\[CrossRef\]](#)
60. Song, J.; Tong, X.; Wang, L.; Zhao, C.; Prishchepov, A.V. Monitoring finer-scale population density in urban functional zones: A remote sensing data fusion approach. *Landsc. Urban Plan.* **2019**, *190*, 103580. [\[CrossRef\]](#)
61. Knox, P. Urbanization. In *Urbanization: An Introduction to Urban Geography*; Knox, P.L., McCarthy, L., Eds.; Prentice Hall: London, UK, 2009.
62. De Silva, M.; Dharshani, G.; Munasinghe, J. Defining ‘urban’ among urbanizing rural: The case of Sri Lankan urbanization. In Proceedings of the 9th International Conference of Faculty of Architecture Research Unit (FARU), Moratuwa, Sri Lanka, 9–10 September 2016; pp. 288–303. Available online: <http://dl.lib.mrt.ac.lk/handle/123/13036> (accessed on 12 March 2024).
63. Furno, A.; Fiore, M.; Stanica, R.; Ziemlicki, C.; Smoreda, Z. A Tale of Ten Cities: Characterizing Signatures of Mobile Traffic in Urban Areas. *IEEE Trans. Mob. Comput.* **2017**, *16*, 2682–2696. [\[CrossRef\]](#)
64. Zhao, G.; Zheng, X.; Yuan, Z.; Zhang, L. Spatial and Temporal Characteristics of Road Networks and Urban Expansion. *Land* **2017**, *6*, 30. [\[CrossRef\]](#)
65. Zheng, Y.; Zhou, Q.; He, Y.; Wang, C.; Wang, X.; Wang, H. An optimized approach for extracting urban land based on log-transformed DMSP-OLS nighttime light, NDVI, and NDWI. *Remote Sens.* **2021**, *13*, 766. [\[CrossRef\]](#)
66. Chen, Y.; Deng, A. Using POI data and Baidu migration big data to modify nighttime light data to identify urban and rural area. *IEEE Access* **2022**, *10*, 93513–93524. [\[CrossRef\]](#)
67. Shi, K.; Yu, B.; Huang, Y.; Hu, Y.; Yin, B.; Chen, Z.; Wu, J. Evaluating the ability of NPP-VIIRS nighttime light data to estimate the gross domestic product and the electric power consumption of China at multiple scales: A comparison with DMSP-OLS data. *Remote Sens.* **2014**, *6*, 1705–1724. [\[CrossRef\]](#)
68. Li, F.; Liu, X.; Liao, S.; Jia, P. The modified normalized urban area composite index: A satellite-derived high-resolution index for extracting urban areas. *Remote Sens.* **2021**, *13*, 2350. [\[CrossRef\]](#)
69. Ping, B.; Meng, Y.; Su, F. An enhanced linear spatio-temporal fusion method for blending Landsat and MODIS data to synthesize Landsat-like imagery. *Remote Sens.* **2018**, *10*, 881. [\[CrossRef\]](#)
70. Ounoughi, C.; Yahia, S. Data fusion for ITS: A systematic literature review. *Inf. Fusion* **2023**, *89*, 267–291. [\[CrossRef\]](#)
71. Zhang, J.; Yuan, X.; Tan, X.; Zhang, X. Delineation of the urban-rural boundary through data fusion: Applications to improve urban and rural environments and promote intensive and healthy urban development. *Int. J. Environ. Res. Public Health* **2021**, *18*, 7180. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Li, X.; Song, Y.; Liu, H.; Hou, X. Extraction of urban built-up areas using nighttime light (NTL) and multi-source data: A case study in Dalian city, China. *Land* **2023**, *12*, 495. [\[CrossRef\]](#)
73. Khan, S.; Nazir, S.; García-Magariño, I.; Hussain, A. Deep learning-based urban big data fusion in smart cities: Towards traffic monitoring and flow-preserving fusion. *Comput. Electr. Eng.* **2021**, *89*, 106906. [\[CrossRef\]](#)
74. Sahai, S. What Is Adaptive Thresholding in OpenCV. 2022. Available online: <https://www.projectpro.io/recipes/what-is-adaptive-thresholding-opencv> (accessed on 11 January 2024).
75. Weeraratne, B. *Re-Defining Urban Areas in Sri Lanka*; Institute of Policy Studies of Sri Lanka: Colombo, Sri Lanka, 2016.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.