Using Geospatial Data to Track Changes in Urbanization

HIGHLIGHTS

- Analysis of Earth observation (EO) big data from satellites and sensors can help stakeholders track and understand urban development over time.
- Working with providers of geospatial data, the World Bank has carried out analytics to measure the qualitative and quantitative aspects of urban transformation, such as the distribution and density of urban sprawl, changes in land use and the growth rate of built-up areas. This allows analysts to begin to understand the drivers of land consumption.
- Combining EO data with other data on population and growth reveals dramatic insights on the overall economic viability, inclusivity, resilience, sustainability and quality of life in urban areas. This enables stakeholders and policymakers to develop and implement informed policies in response, to create thriving cities of the future.

PUBLIC POLICY USES FOR EO BIG DATA

- Track the growth of urban areas and understand economic drivers.
- Evaluate the current state of amenities and identify opportunities and priorities for developing dynamic, equitable, sustainable and resilient cities.
- Underpin smart policymaking to promote optimal spatial and transportation links between jobs, affordable property, health and education services, and recreational areas.

Practice Areas:
Resilient Cities, Inclusive Cities

Countries Involved:
Various in South Asia, East Asia and Africa

Data Types:
Remote sensing, satellite imagery
The World Bank has been using big data to track and study changes in urbanization in low- and middle-income countries, to help ensure that it is sustainable, equitable and supports economic growth. As the world’s population becomes more urbanized, reducing unemployment and promoting sustainability and resilience in urban economies is vital. However, much urbanization is not well planned or managed in developing countries, resulting in cities which have grown rapidly but lack critical infrastructure and are unable to take advantage of economies of scale. For citizens, this means poor transport and services, weak links between job opportunities and the workforce, increased inequality, low resilience to shocks and a lack of sustainability.

Successful urbanization depends on the coordination of three distinct but interdependent processes: public investment in infrastructure, private investment in productive capital and household investment in housing. However, the speed of development and the lack of information available to each set of actors often prevents this coordination. Big data can play an important role filling this gap and facilitating improved coordination. It is especially valuable where the speed and pace of urbanization outstrip the authorities’ ability to understand how cities are growing and changing. Without this baseline information, policymakers cannot meet the challenges and opportunities of urbanization through properly informed decisions.

**Visualizing urban landscapes**

In South and East Asia, the World Bank has explored the patterns, consequences and policy implications of cities’ spatial development by drawing on the increasing availability of spatial data and developments in analytics. Satellite or Earth observation (EO) data can deliver quality results to measure urban growth over a wide range of spatial and temporal scales, particularly when combined with data from other sources. The resulting digital urban maps provide an up-to-date, accurate and cost-effective resource to help national, regional and city governments understand the nature of urban development and make informed decisions. EO datasets allow harmonized and standardized measurements, enabling spatially and temporally consistent comparisons, as well as global assessment. Such data is particularly important for monitoring and understanding the evolution of cities – for example, allowing officials to see when built-up areas spill across formal administrative boundaries. This indicates the need to work with adjoining administrative areas on issues such as connective infrastructure (roads, water mains) or collecting garbage.

Drawing on analysis of EO data, the World Bank has built a database that describes
the speed, magnitude and spatial form of urbanization. Using these data, Bank teams have examined the drivers and impacts of the nature of urbanization and how the urban landscape has evolved to its current state. This provides a baseline from which to understand the effects of policy change and identify priorities for new initiatives. In particular, the Bank focused on exploring the institutional frameworks for urban management (such as mechanisms to coordinate service delivery across administrative jurisdictions), investment (in transport, services and other network infrastructure) and regulation (such as land use, zoning and pricing of services).

In 2008, in close collaboration with the European Space Agency (ESA), the World Bank launched the “Earth Observation for Development” initiative. This provides data on urban and other trends in areas where data are traditionally scarce and often unreliable. Such information can be used to establish project baselines against which progress can be gauged, mitigation measures determined and high-priority issues identified. The initiative focuses on a number of areas, including urban development and related fields such as disaster risk management, the environment, water and energy. To facilitate greater collaboration towards these objectives between policymakers and other development stakeholders, the Bank also developed the Platform for Urban Management and Analysis (PUMA)\textsuperscript{4}. This tool allows users with no prior GIS experience to access, analyze and share urban spatial data in an interactive and customizable way.

These initiatives have led to more than 30 World Bank technical assistance projects delivered to national or city partners or due for completion in the 2008-18 period. The results have been used to maintain and strengthen policy dialogues, as well as to guide the design of new projects. They have led to highly specialized big data mapping products and monitoring systems that leverage EO data for South Asian cities – for example, mapping urban extent and analysis of the internal spatial structures of different cities, or monitoring land subsidence in a city with a high reliance on tube wells for water, lowering the water table. The World Bank-ESA partnership has been expanded until 2018, with Urban Development among its top three priority themes. It aims to provide a systematic source of development information, so that stakeholders can draw on state-of-the-art EO capabilities to develop best practices and sustainability plans.

**Mapping the megacities**

Under the World Bank’s South Asia Megacities Improvement Program, EO big data was used to analyze 20 years of urban expansion in the metropolitan areas of Delhi, Mumbai and Dhaka. These data enabled measurement of the qualitative and quantitative aspects of urban transformation, such as the distribution and density of urban sprawl, changes in urban land use and the growth rate of built-up areas. Using this information, analysts can see how informal settlements grow outside the cities’ administrative boundaries, and can begin to understand the drivers of land consumption.

\textsuperscript{4} Available at http://puma.worldbank.org/
The analysis revealed important insights into land cover and use in the three cities (see Figure 1), revealing the percentage of land taken by residential build-up, industrial build-up, agriculture, natural or semi-natural vegetation and forest. This helps city planners and development stakeholders understand existing requirements and plan for future needs.

In Delhi, for example, the maps show the urban sprawl is accelerated by industrial development. This mainly took place between 2003-10, although a significant increase in construction sites indicates that it will continue in the future – and must therefore be planned for.

Digitized spatial data allows analysis at different administrative levels: metropolitan, city, district or sub-district, as well as other non-administrative units. Such datasets allow flexible aggregation, for instance, showing the proportion of sprawl by district, its distribution and density, class evolution within urban areas and the drivers of urban change. (If housing comes before roads, for example, this indicates informal and incremental city building. If roads come first, development is formal.) Combined with environmental or socio-economic data, the data can provide information concerning the ratio of population growth to urban growth, and can measure indicators such as compactness (as a function of city density), the ratio of green space to citizens, and the proximity and accessibility of green areas.

Figure 1: Sample visualizations from the South Asia geospatial analysis.
In 2014, the World Bank used satellite imagery and demographic data to measure expansion and population change between 2000-10 in East Asian urban areas of 100,000 people or more. Analysts used change-detection methods that draw on satellite data from the Landsat remote sensing project operated by the US Geological Survey and the National Aeronautics and Space Administration (NASA). These maps rely on a geophysical definition of built-up areas as landscape units with more than 50 percent coverage of non-vegetative, human-constructed elements. These areas were combined with the AsiaPop map, from the world mapping project\(^5\). The refined land cover datasets were then combined with population density information derived from census data, and used to disaggregate population counts to a grid of 100 meters squared. This approach has allowed the Bank to understand the entire region so as to establish systematically where urbanization is occurring, to what degree and how quickly. This highlights the responses required from stakeholders, such as meeting needs for services (water, sanitation or transport) and regulation. Analysts also quantified the relationship between urbanization, income growth and inequality.

The approach provided critical information on the growth of urbanization in East Asia. For example, by 2010, the region had 869 urban areas with more than 100,000 people (600 of these in China). Populations in these cities had grown dramatically, lifting the region’s new urban population to approximately 200 million people. However, despite this growth, only 36 percent of the region’s total population lives in urban areas. This suggests that the trend in urban growth is likely to continue for many decades. Lower-middle-income countries such as Indonesia, the Philippines and Vietnam had the fastest urban population growth, whereas upper-middle-income countries such as China had the fastest spatial growth (hence most urban areas outside China became denser, while the density of many Chinese urban areas declined). Most of this growth occurred in small and medium-sized cities rather than the megacities. Unsurprisingly, urban density was high, and increased through the period.

Findings from EO big data can help coordination between public investment in infrastructure, private investment in productive capital and household investment in housing. They can allow policymakers to promote optimal spatial and transportation links between jobs, affordable housing and business units, health and education services and recreational areas. These insights can also be used to help support recent rural-to-urban migrants, bringing them to the attention of city authorities and ensuring that rapid urbanization is inclusive. As EO-based big data techniques spread across Africa and other developing regions, and are refined and adapted, they will provide valuable tools and insights to policymakers, and even greater benefits for the citizens of the future.

**Creating livable cities**

Further EO big data approaches are also helping drive sustainable urban development. Research into using high-resolution satellite data for poverty mapping in Sri Lanka draws on emerging techniques that can profile fast-
changing urban areas in near-real-time. These techniques can identify built-up area, building and car density, and types of roofing and road. Using open-source image-processing algorithms, they can even calculate whether buildings are more rectangular or have more chaotic angles (indicating higher poverty) and construct poverty indicators such as the percentage of paved roads in an area. This helps stakeholders target their interventions precisely where they are most needed.

In Kosovo, after a period of chaotic urban expansion, geo-spatial data collected by drones has been used to secure property rights quickly and cheaply – vital to household security and economic growth. The drones record imagery which is processed into high-resolution orthophotographs (aerial photographs corrected to have the same lack of distortion as a map) in a fraction of the time of conventional aerial surveys. The orthophotos are used to gather property boundary information from local residents, which can be used for formal property registration. In the fast-growing city of Ferizaj, this approach was used to support a government program for unregistered land owners to legalize their property rights. The data from the drone flights were processed in 24 hours using two local high-end desktop computers, resulting in orthophoto maps from which land owners could easily identify their property. The initiative can be scaled up globally, especially to secure land rights in low-income countries.

All these projects demonstrate that analysis of EO big data can be a significant tool for managing urban development in low- and middle-income countries. It can measure, baseline and track the growth of urban areas, and highlight the drivers of economic growth. This allows policymakers and stakeholders to better understand the factors leading to inefficiencies and inequality in urban areas, and to develop informed policies in response. They can also build resilience into urban environments, so that residents, institutions, businesses and systems can adapt to chronic stresses or acute shocks. And they can create livable cities that fulfill their residents’ needs.

An Earth observation project shows a successful approach to tracking changes in urbanization using geospatial data

HIGHLIGHTS

• New cloud-based computational platforms such as Google Earth Engine (GEE) enable urbanization to be monitored using multi-spectral imagery from multiple satellites over extended time periods.

• Current and historic satellite data can be ground-truthed by manually labeling areas as “built-up” or “not built-up”. A machine-learning algorithm allows for the conversion of these data into highly-accurate classification of
While urbanization in rapidly growing nations is helping lift hundreds of millions of people out of poverty, it is also creating immense societal challenges. It is expanding greenhouse gas emissions, destabilizing fragile ecosystems, and creating new demands on education, health, and transportation infrastructure. Despite the importance of understanding the drivers of urban growth, it is still not possible to quantify the magnitude and pace of urbanization at a global scale. Standard empirical approaches use data from household surveys, which are costly to carry out, produced infrequently and subject to often-severe measurement problems. Reliable and up-to-date data on urbanization – particularly from developing countries – remain scarce.

The coming revolution in geospatial data holds the potential to transform the way in which we study cities. As satellite imagery at ever-improving spatial and temporal resolutions becomes available, new approaches to machine learning are being developed that convert these images into meaningful information about the nature and pace of change in urban landscapes. Research on land use is rapidly shifting towards remote-sensing methods designed to capture urban features as they are observed in terrestrial Earth observation data.

Despite the promise of satellite imagery for urban analysis, current information on land use is still subject to many drawbacks. Existing satellite-based classifications of urban areas cover limited geographic extents and time periods, and frequently disagree in terms of the size and shape of particular cities. Progress is further inhibited by lack of large datasets that give the “ground truth” regarding urbanization, which are essential for validating the micro-detailed maps of urban areas that remote-sensing methods produce. These deficiencies mean that urbanization still cannot be tracked with a high degree of precision across space and time.

Global-scale analysis of satellite imagery
As powerful new cloud-based computational platforms become available to the research
community, it is becoming feasible to monitor urbanization using multi-spectral imagery from multiple satellites over an extended time-period. One such platform is Google Earth Engine (GEE). GEE leverages cloud-computational services for planetary-scale analysis and consists of petabytes of geospatial and tabular data. This includes a full archive of scenes from the US remote sensing project Landsat (as well as other satellites), together with a JavaScript, Python-based application programming interface (API), and algorithms for supervised and unsupervised image classification.

Recent World Bank research demonstrates the applicability of GEE for studying urban areas at scale. It provides, for the first time, reliable and comprehensive open-source, ground-truth data for supervised image classification that delineates urban areas – in this case, in India. As a large, geographically diverse nation undergoing a rapid urban transition, India represents an ideal context in which to illustrate the applicability of new approaches for mapping urban expanse. The team leveraged the computational power of GEE and its full Landsat archive to introduce a practical and adaptable procedure for analysis of urban areas at a global scale.

The dataset consists of 21,030 polygons that were manually labeled as “built-up” urban areas or “not built-up” areas. To generate a machine-learning algorithm that allows for the conversion of these ground-truth data into a classification for the country as a whole, the team assessed alternative supervised classifiers (Random Forest, Classification and Regression Tree (CART) and Support Vector Machine) and examined the effects of various inputs and class combinations on the performance of the classifiers. They proposed a methodology – “spatial k-fold cross-validation” – to evaluate the extent to which the classifiers...
could be generalized in spatial terms, and their performance in a large and geographically heterogeneous context.

**Accurate mapping across space and time**

The team found that their ground-truth dataset can be used in GEE to produce high-quality maps of built-up areas in India, across space and time. Their classification captures the fabric of built-up urban areas, as well as the fine boundaries between cities and their peripheries.

In validating the classification results, the team showed that when used with standard classifiers available in GEE, they achieve a high overall accuracy rate of around 87 percent in identifying built-up areas for grid cells with a dimension of 30 meters. Of the three types of classifiers examined, Random Forest achieved the best performance (a balanced accuracy rate of 80 percent). However, the performance of the classifiers strongly improves as the size of the training dataset increases (especially with CART). With this methodology, which is designed to evaluate the spatial generalizability of classifiers, the team showed that the classifiers also perform well when the training examples are sampled from areas with heterogeneous land cover (such as a mix of dense vegetation and bare ground).

As inputs for the classifiers, the team used Landsat 7 and Landsat 8, launched in 1999 and 2013 respectively. Although Landsat 7 is of a lower spectral resolution than Landsat 8, it allows for a longer time horizon over which to study urbanization. The project demonstrated that the addition of two pixel-level indices to Landsat 7 as inputs to the classifiers – the Normalized Difference Vegetation Index and the Normalized Difference Built-up Index – improves classifier performance so it approaches that when Landsat 8 is used for the input. This shows that the classifiers can be used to detect urban areas in historic data. When the extent of urban areas in 2000 was mapped in this way, the team found a high overall classification accuracy of 86 percent.

**Towards precise, real-time urban measurement**

Urbanization is a fundamental driver of economic growth in the 21st century. Although it implies massive productivity gains for an economy, it is also accompanied by congestion, pollution and heavy demands on public-sector resources. Understanding the ecological, environmental, social and economic impacts of these changes is essential for preserving a sustainable society.

As parallel computational platforms become accessible to researchers, it is now possible to expand urban research across space and time. In this study the team developed a large-scale geo-referenced dataset that was used to facilitate the detection of urban areas at a national level, and to provide a useful and reliable tool for temporal analysis of urban zones and their rural peripheries. The study highlights the potential of GEE for urban research and illustrates the applicability of the dataset for the detection of urban areas in a country with a large population and a diverse land cover. The methodology and the evaluation procedure are suitable for studies that analyze
large-scale regions, and can easily be applied to other countries and contexts.

However, the team noted several limitations for their approach and made recommendations for future studies. First, the dataset was labeled according to 2014-15 imagery using a visual-interpretation method, which, by its nature, may be subject to idiosyncratic variation across individuals performing the manual classification. It is necessary to assure that across the dataset each example is labelled by multiple people and to account for the agreement between them. Second, the analysis is limited to India. Creating manually labeled ground-truth data is expensive and time consuming. However, crowd-sourcing platforms may allow researchers to scale – at low cost – the labeling method and to construct larger and more comprehensive ground-truth datasets. Third, the sampling method used in this study was designed to detect the boundaries between built-up areas and their periphery. The majority of the labeled examples were sampled from highly populated areas and from their adjacent, low-population environs. This approach may create a risk of false-positive detections when classifying distant or remote areas. It is therefore suggested that future projects taking this approach include in the training set examples from remote areas that are less populated.

New studies that take into account these limitations and successfully exploit emerging approaches such as crowd-sourced information will enable the measurement of urbanization in close to real time. This has the potential to transform how public policy is designed, helping planners to achieve successful urban growth and address many of the most persistent challenges of economic development.

*Big Pixel, Case Study of India: University of California San Diego. Big Pixel Research Team: Ran Goldblatt with Wei You, Gordon Hanson and Amit K. Khandelwal*


*Google Earth Engine: [https://earthengine.google.com/](https://earthengine.google.com/)*