DEMONSTRATION REPORT
SPATIAL GROWTH ANALYTICS

City Form Lab

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

TECHNICAL SUPPORT FACILITY
FOR NATIONAL URBAN DEVELOPMENT PROGRAM IN INDONESIA
DEMONSTRATION OF ANALYTICS

SPATIAL GROWTH ANALYTICS

Authors:

Andres Sevtsuk
PI, City Form Lab
Singapore University of Technology and Design

Reza Amindarbari
Researcher, City Form Lab
Singapore University of Technology and Design

Collaborators:

Taimur Samad
Senior Urban Economist
World Bank-Indonesia

Wilmar Salim
University of Bandung

Chandan Deuskar
Urban Development Analyst
World Bank

Renata Simatupang
Economist
World Bank-Indonesia
SECTIONS

1 INTRODUCTION

2 GEOSPATIAL DATA

3 SPATIAL GROWTH AND CHANGE ANALYTICS

4 PLANNING DECISIONS SUPPORT
INTRODUCTION
1. INTRODUCTION

A key engagement of the World Bank under the National Urban Development Program (P3N) is technical and institutional capacity building for supporting evidence-based spatial and development planning in Indonesia. This activity will take place through new permanent technical assistance facilities – City Planning Labs (CPLs) - that are proposed within Bappeda in four participating cities – Denpasar, Surabaya, Palembang and Balikpapan. Proposed CPL teams, supervised by a director, include civil servants from Bappeda, Dinas Tata Kota and other related city departments, and technical staff with backgrounds in urban planning, geographic information systems (GIS), and spatial analysis hired from outside the local government. In order to build the essential capabilities at CPLs, teams of international World Bank consultants will collaborate closely with CPLs in the first eighteen months of the project.

CPLs’ primary goal is to collect geospatial data, and to conduct analyses that can inform spatial planning decisions in the abovementioned cities. CPLs' analytic activities will be structured around a core module and four supplementary modules. The core module aims to establish the CPL facilities, institutional structures and data platforms, and to train CPL staff to implement basic urban growth and change analysis in each respective city. The optional modules include: (1) city economic competitiveness, (2) slum analytics and management systems, (3) climate and risk resilience planning systems, (4) and monitoring land and real estate markets.

A separate World Bank concept note describes the analytic activities proposed for each module. The purpose of the present report is to demonstrate the implementation and specification of the analyses proposed in the core urban growth analysis module. The issues and opportunities that Indonesian cities in general, and the four pilot cities in particular, face due to rapid urban growth are diverse and locally specific. The list of analytic activities needed to understand and address all of them with appropriate planning tools and policies is too exhaustive to be captured here. This report focuses on a few basic, but useful spatial analysis techniques that are proposed as part of the core CPL module in each city.

Figure 1. Palembang is a city of 1.7M inhabitants in Sumatra. Due to the presence of frequent earthquakes and the lack of investment in commercial real estate, the city center remains horizontally dense but vertically low rise.
The report first discusses essential data requirements for the proposed analytics and the needed geospatial data platform for sharing geographic information among government agencies, stakeholders and the general public. The first section of the report focuses on data preparation, appropriate units of analysis, required attributes for each dataset, and potential sources for the corresponding data in Indonesia. The second section concentrates on the analytic activities. The analytic activities we discuss and demonstrate include spatial growth and change analysis, spatial accessibility analysis, impact analysis, and spatial-statistical models. Technical implementation details and required geo-processing tools are discussed through a series of examples using data from other cities and countries. The third and final part of the report focuses on potential planning applications of the proposed analysis techniques.

Developing an evidence-based planning decision support system that relies on empirical and up-to-date data and utilizes powerful spatial analysis tools could make a significant contribution towards turning the rapid urbanization in Indonesian cities into a vast opportunity for economic growth, equitable resource distribution and access to human development opportunities. The data and analysis techniques described in this report are, however, not only applicable to Indonesian cities - they could also benefit a number of other rapidly urbanizing countries in South East Asia and beyond. The City Planning Labs project in Indonesia will test the implementation of geospatial data systems and analysis techniques in four medium-scale cities, which will offer a unique opportunity to learn and improve from the experience. Another report will be compiled at the end of a 12 or 18 month engagement with the CPLs in Indonesia, describing how the implementation unrolled and what could be done better next time.
2.1. GEOSPATIAL DATA

The first responsibility of CPLs is to collect and disseminate geospatial data required for supporting planning decisions in cities. A large amount of data that CPLs require for their analytical activities already exists in different agencies and departments at the local and national level in Indonesia. We provide a table at the end of this section showing all the datasets that are recommended to be collected at each CPL, where we also indicate whether and where we have witnessed the availability of such data. Not all of these data will be available in each of the four cities. During the first year, the table can be used as a wish-list for geospatial data for the core CPL module.

CPLs would benefit from establishing a sustainable long-term collaboration with agencies who generate or harvest geospatial data in Indonesia, in order to have access to the most up-to-date information and to also share the data they compile with other agencies. Much of the existing data, however, is not digitized or attributed for GIS. Construction and change-of-use permits, land use maps, cadastral records, Masterplans and survey records can be paper-based, making referencing and data sharing difficult. CAD drawings, where they exist, often come without a spatial reference for geographic location. Increasingly, Indonesian cities have started collecting these data digitally and some have access to remote-sensing information, such as medium or high-resolution aerial or satellite imagery, LiDAR scans, or satellite stereo imagery. A number of cities have gone through initial efforts to digitize building footprints, street networks and land-use classification zones, albeit with relatively limited attribute information describing the characteristics of the geometric elements. In the first project phase, CPL staff will need to collect, organize and standardize a broad range of existing datasets available in each city.

As existing data are often partial, CPLs will also need to obtain some data, such as remotely sensed topographic data through specialized service providers. Ground surveys should be prepared for building use and condition maps or additional socio-economic data gathering. CPLs should develop routine processes for implementing annual ground surveys for accuracy.
verification and for completing or updating existing datasets. The World Bank consultants in this project will propose standards and guidelines for such data collection during the first implementation phase.

2.1.1. Units of Analysis

All geospatial data is associated with geometric spatial units (e.g. address points, building or parcel polygons, street centerlines), which are geographically referenced. Each individual spatial unit contains certain attributes about the environmental feature it symbolizes (e.g. building floor area, census block population, parcel value etc.). The spatial units for available data typically depend on both the geometric elements observed in the built environment (streets, buildings, parcels), the ease with which the data can be physically surveyed on ground (census blocks can be typically surveyed in less than a day), as well as the graphic representation constraints (e.g. sloping land is usually symbolized with horizontal elevation lines that are not witnessed in reality). For privacy and security reasons, or technical limitations, agencies that collect data do not always release spatial information at the same spatial resolution or with the same units of analysis that data was originally assembled. Statistic Indonesia (BPS), for instance, collects all census data at a census block level (often comparable to a city block), but only releases data at the village, sub-district and higher levels.

Original spatial units of data, as disseminated by different data collection agencies, do not necessarily match the needs for urban growth analyses proposed in the core module. An important part of CPL’s activities will be dedicated to compiling and deriving new datasets, using geo-processing techniques for aggregating or disaggregating raw data to desired spatial units. For building level analysis, for example, land use data or economic census data should be disaggregated from original tract levels to individual building levels (discussed below).

To classify spatial data in this report we focus on the spatial unit rather than the content of attribute fields. Datasets with different spatial units may have
similar attributes – *parcels, buildings, census blocks* and *villages* can all contain information about the residential population, the built area, or the number of business establishments they accommodate (Figure 3).

The following list of data is seen as most essential for CPLs to collect, in order to carry out the analytics proposed in the core module. For each dataset, desired attributes, required raw data, and potential data sources are explained where possible.

### 2.1.1.1. Buildings and address points

Individual buildings are among the finest spatial unit of analysis commonly used in cities. A substantial part of human activities takes place in buildings; the people that buildings accommodate and the activities they engage in, generate the origins and destinations of most pedestrian and vehicular movement on city streets.

Buildings are typically represented as polygon features (building footprints). Polygon features illustrate the realistic geometry of the actual building footprints and allow valuable post-processing analysis that may not be included in the original attributes (building area or perimeter calculations, 3d height extrusions, volume calculations). However, buildings can also be represented as point features, placed at either the centroid of the actual footprint or at one or more exterior entrance locations of the building. Point representation of buildings can be useful for analyzing accessibilities between buildings on the network of city streets that requires discrete locations (Figure 4). Polygons can easily be converted to points, but not vice versa. Centroid points, may fail however, to capture the accurate relationship between a building and its street(s) or the visual sightlines available between buildings. It is therefore preferable to represent buildings with realistic polygons.

Address points, discussed hereafter, can be placed at entrances to accurately capture the relationship between buildings and streets. The geometry of buildings (footprints or 3D volumes) are typically obtained from satellite imagery or aerial scans (e.g. LiDAR, stereo imagery). A high-resolution geo-referenced satellite image can be used as a base for drawing building...
polygons with vector-based lines in drafting software like AutoCAD, DraftSight or Rhinoceros. An effort should be made to distinguish two autonomously owned or used buildings into two separate polygons whenever possible. This can be challenging to do if buildings on satellite imagery appear to share walls or are otherwise densely spaced, but a careful distinction of individual buildings according to addresses can greatly benefit later analysis. A physical ground check is usually required to cover areas that are poorly visible in the satellite image (e.g. cloudy or obstructed by trees) and to ensure that the drawings match reality.

Generating address point features typically requires a ground survey, using GPS tools for recording the spatial position of each observed address location on the ground (Figure 5). If accurate street network data is available (e.g. from Navteq, Tom Tom) and the total number of addresses on each street known, then address point locations can also be interpolated in GIS (with some spatial error). Tools for performing such addressing are offered in ESRI’s ArcGIS. Managing standardized addressing data is usually a national level activity; the system should be consistent throughout the country. Efforts towards a national address database appear to be under way at the Indonesian Geospatial Information Agency (BIG).

Desired building and address point attributes for proposed CPL analytics include:

*Volume, total floor area and footprint area*

Building volume and floor area indicate the amount of available space for human activities. Although floor area is a more accurate indicator of space for living, working, studying, it is not always easily obtained and requires detailed surveys or registry records. Building volume, however, can be directly computed from the basic geometry information (footprint area and building height). Approximate building floor area can be found by diving the building height by a typical floor height (e.g. 3 meters) and multiplying it by the area of the footprint. Typical floor-to-floor height for residential buildings is 2.8 meters.
for office building 3.5 meters, for institutional buildings 4 meters and for industrial buildings 5 meters.

Address

Street addressing is the practice of assigning unique names to spatial features, typically buildings, plots and business locations, using a consistent hierarchical system. A street addressing system contains several components that are consistent across all individual units. The most typical components of an address are street segment name, street type (road, drive, highway etc.) plot or building number, unit number (if applicable), and an area ID such as ZIP or postal code. The purpose of using such a hierarchical naming system is to allow users to locate an address even when they do not have access to GIS data.

Street addressing is vital for locating facilities and infrastructure (businesses, hospitals, schools etc.), and delivery services (e.g. postal or emergency services) in an urban setting. Developing a standard street addressing system as a common platform among all public and private agencies is also crucial for urban information management. It allows for systematically storing surveyed data at the highest possible resolution (household or business level). It also allows for generating a great deal of spatial data from registry records that contain address attributes and keep them continuously updated with little effort and cost.

Developing a street addressing systems is often a national level effort, but conducted at a local level, where CPLs can play a significant role. There are a number of examples of such efforts in developing countries, including in a series of Sub-Saharan countries in Africa, in collaboration with the World Bank (see Farvacque-Vitkovic et al 2005).

While in the long-run, public and private agencies may update their address information using a standard addressing system, CPLs can also help assemble data from registry records using available street information. This requires bringing presently available addresses into a uniform format. ArcGIS geocoding tools, and Python or VB string functions allow for matching the

Figure 7. RT-RW is currently the smallest addressing unit in Indonesia. Each RT contains several households. Source: John Taylor.
Business establishment and employment data at an individual building level should ideally include the total number of businesses and employees classified by different industry categories (e.g., retail establishments or services), as shown in figure 7. Raw business location data typically comes at address or geographic coordinate (point) level, as explained further below. Aggregating such point data to building footprints, however, provides convenient units of analyses and allows buildings to be used as inputs in multiple types of spatial analyses.

Businesses establishment and employment data are usually available in two different forms. They may be available in aggregate census tract level (or other statistical boundaries), indicating the total number of businesses within each aggregated area. Storing detailed business classification information is not typically feasible in this case. Second, every business can be shown as an individual unit, represented by points with attribute information (see 2.1.1.3. business locations). When business establishments are shown at the building level, the business attributes should be summarized, showing the sum total of all establishments that occupy each building.

BPS collects data on medium and large business establishments with more than 20 employees, which constitutes a small percentage of all businesses in Indonesia. BPS also surveys small samples of all business every year, which cannot be disaggregated lower than city scale. CPLs may need to conduct ground surveys to collect more comprehensive data on business establishments.
Population is the key determinant of demand for a city’s resources. Detailed data on spatial distribution of population – and demographic sub-groups – allows for efficient estimates for a city’s resource needs. Population and demographic data are not commonly disseminated at the building level, but aggregated to census block or tract levels. In Indonesia, BPS conducts household surveys and provides census data at the village (Desa or Kelurahan) level. If individual buildings’ type (e.g. residential, commercial etc.), and floor area or volume are known, then population values from higher-level spatial units can be disaggregated to the building level with reasonable accuracy. The total number of residents in a census block can thereby be allocated between residential structures, weighing the allocations by the size of each building (Figure 8).

Building type and subtype
Building type describes the types of activities that take place in the building (Figure 9). Reliable assessment of the real estate market (asset and use) is not feasible without building type and subtype information. Building types or subtypes do not necessary share the same market. Commercial and residential spaces belong to separate markets and separate demand segments. To determine the supply side of each market, it is essential to keep track of building stock by type.

Building subtypes (e.g. housing) can also have separate markets (Figure 10). The demand for large landed houses is composed of a different socio-economic group of buyers and renters than the demand for small studios or public housing units.

The two main sources for building type data are zoning maps, which usually do not contain subtype information, and ground surveys. Developing and maintaining an accurate building type database can be very labor intensive, but the pay-offs are also high since building level data allows for many useful analyses about city’s real estate market.

Figure 8: Disaggregating population information from census tract level to buildings. The population of the census tract is allocated only among buildings that contain residential uses, weighing the allocation by the building volumes. Naturally some spatial error is generated in the process, but storing population estimates at an individual building level is useful for a number of high-resolution analyses.
Once a reliable building type and subtype database is assembled through ground surveys, it is practical to maintain and update it via building permit, modification permit, demolition permit and change-of-use permit databases. If a new permit is issued, the finished building occupancy permit can automatically signal to the building database managers that a new building has been added to the stock. The building type database can then verify the data and add the new building to the repository. A similar procedure can follow other types of building permits.

2.1.1.2. Parcels

Parcel geometry records land ownership borders. The geometry of parcel borders is often provided by national land agencies (e.g. BPN in Indonesia). In Indonesia, Tax Directorate General also prepares parcel polygon datasets, for its own land value assessment purposes. The two parcel databases currently remain separate, but BPN appears to be working on a joint database. Attributes that parcels should contain include:

Assessed value and transaction history

Parcel is an intuitive unit for land market related analyses, as transactions and value assessments are conducted at the parcel level. In Indonesia, Tax Directorate General keeps track of land transactions, and assesses land values.

Zoning

Parcel is the appropriate unit for containing zoning attributes, such as land use, plot ratio, height limit and setbacks, as building permits are issued for specific parcels. Zoning information presented in masterplans and detailed plans, which are prepared by Bappeda in each city in Indonesia, typically specify zoning regulations for each parcel.

1 Source: personal communication with BPN
Building properties

Most physical building structure or land improvement data can be also aggregated to the parcel level: e.g. total floor area, total building volume, address, total number of businesses.

Frontage

In addition to typical geometrical properties (perimeter and area), it is useful for parcel datasets to contain the length of street frontage: the length of parcel perimeter that is directly connected to a street (Figure 11). Frontage is an important determinant for land value, and essential for developing hedonic pricing models for land.

Parcel type

Parcels can also optionally be classified based on the number of streets that a parcel is directly connected to (Figure 12). A “middle parcel” has access to one street, but a “corner parcel” can have access to 2, 3 or 4 streets. Similar to frontage, parcel type, as defined above, is an important determinant of land value, and useful for developing hedonic pricing models.

2.1.1.3. Business locations

As mentioned above, raw business establishments and employment data is best stored at an individual business establishment level, represented as points (Figure 13). Representing separate business establishments with separate point features is the most robust and useful way of storing the business establishments’ data. Points can always be aggregated or joined to other larger units (e.g. buildings or parcels) if needed.

The attribute information of business locations should typically indicate:

- The legal name of the business
- The name of a parent company (if applicable)
- Detailed industry classification code (e.g. NAICS) at as detailed level as available (e.g. 6 digits), see Figure 7.
- Year established at the present location
- Number of employees
- Longitude and latitude coordinates
- Address
- ZIP code
- Town, Region

CPLs may conduct ground surveys to collect business location data as BPS collects data only on medium and large business establishments with more than 20 employees.

In the longer run, accurate business location data can be collected from DG Tax records that should account for all business locations for income tax and sales tax reasons. A good tax system can produce ample spatial data annually, at almost no extra cost.

2.1.1.4. Transportation network

The movement of goods and people in cities takes place through three layers of networks: vehicular roads, pedestrian paths, and public transit networks. The latter is often used together with the pedestrian network, and forms a multi-modal network. Analyses that help us understand how resources and facilities are accessible to users through the mentioned networks require data. Most cities collect and prepare road centerline datasets (Figure 14) and sometimes public transit network datasets, but often overlook the pedestrian network. Road centerlines are the most commonly used GIS data for accessibility analyses, not only for vehicular movement, but also for pedestrian movement. A large part of pedestrian flow takes place along streets. However, street centerlines do not capture pedestrian routes that are not along roads (e.g. through unoccupied parcels in informal settlements). Purely vehicular routes, such as toll roads, do not have sidewalks. It is, thus, recommended that the CPLs prepare geospatial datasets of street centerlines, public transit lines, as well as sidewalks and other pedestrian paths (Figure 13). A great deal of a
city’s circulation in Indonesia occurs on foot. Beyond accessibility analyses, sidewalk databases will be also useful for sidewalk improvement plans.

Street network centerline data may be available in Indonesian cities via third-party data providers, such as NAVTEQ.

Potential attributes for road network datasets include:
- Width or number of lanes
- Type (e.g. paved or unpaved)
- Street name
- Road classification
- Traffic directionality

Desired attributes for pedestrian network datasets include:
- Width
- Type (e.g. indoor, outdoor, outdoor but sheltered)

Potential attributes for public transport network datasets include:
- List of buses using the segment
- Average time consumed on the segment
- Frequency (of bus or train on the segment)
- Start and end stations of the segment

2.1.1.5. Administrative boundaries

Administrative boundaries are abstract extents that define the spatial authority of governance of communities in a hierarchical order from national level to smallest groupings in neighborhoods e.g. RT or RW in Indonesian cities). Administrative boundaries are common spatial units for storing socio-economic data.
BPS provides census data at the village level, as well as higher aggregation levels such as city, district, or province. Micro data in Indonesian cities is typically collected by the head of village at the RT level (Figure 15). The CPLs should digitize and distribute the following set of administrative boundaries:

- Regency (Kabupaten) or City (Kota)
- Sub-district (Kecamatan)
- Neighborhood/village (Desa or Kelurahan)
- Block (RT/RW)

Each administrative polygon should carry a unique identifier name. Lower level polygons should also indicate the names or IDs of the higher level polygons they are part of. The polygon shapefiles can be likely obtained from the local BPS office up to the neighborhood level. Mapping the RT-RW boundaries could require collaboration with the heads of villages. Such mapping has been previously implemented in Solo.

2.1.1.6. Other spatial data

The datasets discussed above constitutes only the most essential data that can be used in the core module of CPLs. Much of the data is also directly useful for other optional CPL modules. The list of spatial data that cities collect or already have can be very exhaustive. Many of those data can be associated to one or several spatial units mentioned above; e.g. energy consumption can be associated to buildings and parcels, crime rates to any administrative boundaries. However, there are some other data that require their own spatial units: for example, data on water infrastructure or Wi-Fi hotspot. A list of spatial data that CPLs are recommended to collect is provided in the table below.
Not available but desired for proposed analytics

- Available
- Available in some of the participating cities

<table>
<thead>
<tr>
<th>INDONESIA URBAN DATA</th>
<th>BPN</th>
<th>BPS</th>
<th>Bappeda</th>
<th>DG Spatial Planning</th>
<th>DG Tax</th>
<th>Other Potential Sources</th>
<th>BIG</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DATA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1</strong> IMAGERY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-resolution satellite image</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aerial photography</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital Elevation Model (DEM) of urban topography</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall urban extent (built-up area in the metro area)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>World Bank</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2</strong> PLANNING REGULATIONS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zoning plans</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Floor area ratios (gross plot ratios), as specified in regulations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land use as shown in city master plan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Building heights limit, as specified in regulations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STATISTICAL BOUNDARIES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Provincial administrative boundaries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Municipal administrative boundaries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>District administrative boundaries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sub-district administrative boundaries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zip code areas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Census tracts (Village)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>DEMOGRAPHIC CHARACTERISTICS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Residential population (census)</td>
<td></td>
<td></td>
<td>BPS data are aggregated at village level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Residential population by sex</td>
<td></td>
<td></td>
<td>BPS data are aggregated at village level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Residential population by age group</td>
<td></td>
<td></td>
<td>BPS data are aggregated at village level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Residential population by residence type</td>
<td></td>
<td></td>
<td>BPS data are aggregated at village level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Household surveys: household income</td>
<td></td>
<td></td>
<td>BPS data are aggregated at village level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Household surveys: family size</td>
<td></td>
<td></td>
<td>BPS data are aggregated at village level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>URBAN FORM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Observed floor area ratios</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Observed land use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Officially recognized informal settlements</td>
<td>World Bank</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Building floor areas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>City blocks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parcel boundaries, ownership</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Building footprints</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Building heights</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Building ages</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Building uses (e.g. residential, commercial, etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Building types (e.g. walk-up, condo, row-house, kampong, informal)*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Building addresses/ZIP Codes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>INFRASTRUCTURE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-------------------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Transportation infrastructure: roads by category, # lanes, direction, setbacks, centerlines, intersections, traffic lights, toll gates.</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transportation infrastructure: rail *</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transportation infrastructure: bus lines and stations*</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transportation infrastructure: bicycle routes</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transportation infrastructure: pedestrian sidewalks, crossings*</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Service infrastructure: water, sewage, and drainage*</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Service infrastructure: electricity lines, substations</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drinking water supply networks*</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Potable water source locations (e.g. wells)*</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ECONOMIC CHARACTERISTICS</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Municipal/district expenditure by economic categories*</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Province expenditure by economic categories*</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Municipal/district revenue by sources*</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Province revenue by sources*</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>ESTABLISHMENTS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----------------</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Firm distribution (Individual establishment locations)*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job distribution (jobs per area/ type)*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Points of interest (museums, institutions)*</td>
<td>○</td>
<td>○</td>
<td></td>
</tr>
<tr>
<td>Public institutions (hospitals, police stations, libraries etc.)*</td>
<td>○</td>
<td>○</td>
<td></td>
</tr>
<tr>
<td>Public institutions: schools (with levels, and no. students)*</td>
<td>○</td>
<td>○</td>
<td></td>
</tr>
<tr>
<td>Public institutions: hospitals (with specialties, and capacities)*</td>
<td>○</td>
<td>○</td>
<td></td>
</tr>
<tr>
<td>Public institutions: others (detailed)*</td>
<td>○</td>
<td>○</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>NATURAL HAZARD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seismic hazard zones</td>
<td></td>
<td></td>
<td>World Bank</td>
</tr>
<tr>
<td>Flood zones</td>
<td>○</td>
<td>○</td>
<td>World Bank</td>
</tr>
<tr>
<td>Land topography (points / topolines)</td>
<td>○</td>
<td>○</td>
<td>World Bank</td>
</tr>
</tbody>
</table>
## LAND AND REAL ESTATE MARKET

<table>
<thead>
<tr>
<th>Land cadaster: parcels (tax, transactions, etc.)*</th>
<th>●</th>
<th>●</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land prices*</td>
<td>●</td>
<td>Developers &amp; Brokers</td>
</tr>
<tr>
<td>Housing prices*</td>
<td>●</td>
<td>Developers &amp; Brokers</td>
</tr>
<tr>
<td>Commercial real estate prices*</td>
<td>●</td>
<td>Developers &amp; Brokers</td>
</tr>
<tr>
<td>Housing Rents*</td>
<td>●</td>
<td>Developers &amp; Brokers</td>
</tr>
<tr>
<td>Commercial rents*</td>
<td>●</td>
<td>Developers &amp; Brokers</td>
</tr>
<tr>
<td>Land rents*</td>
<td>●</td>
<td>Developers &amp; Brokers</td>
</tr>
<tr>
<td>Housing tenure (vacant, owner-occupied, rental occupied)</td>
<td>Developers &amp; Brokers</td>
<td></td>
</tr>
<tr>
<td>Residential unit sizes/ no. of rooms</td>
<td>●</td>
<td>Developers &amp; Brokers</td>
</tr>
<tr>
<td>Issued building permits within the past x period*</td>
<td>●</td>
<td>Developers &amp; Brokers</td>
</tr>
<tr>
<td>Issued demolition permits within the past x period*</td>
<td>●</td>
<td>Developers &amp; Brokers</td>
</tr>
<tr>
<td>Projects currently in construction (residential/commercial/industrial/office/other)</td>
<td>●</td>
<td>Developers &amp; Brokers</td>
</tr>
<tr>
<td>Sales Prices / Rate of Sales at each new development currently on the market</td>
<td>Developers &amp; Brokers</td>
<td></td>
</tr>
<tr>
<td>List of all permits required for new developments by land use type and typical durations.</td>
<td>●</td>
<td></td>
</tr>
</tbody>
</table>
2.2. DATA PLATFORM

To fulfill their primary goal of assembling, maintaining and distributing large geospatial databases, the City Planning Labs need a digital geospatial data platform that satisfies five fundamental requirements. The platform should:

- Allow data to be efficiently and conveniently stored and managed
- Allow data to be shared across different city departments or with members of the public over Internet browsers
- Enable all data management operations to be performed from a local networked computer
- Enable users to download data layers that they have security clearance to access
- Enable the end-users to interact with the datasets on a web-browser, by querying their attributes, overlaying different data layers, using simple base-maps to situate the information, and overlaying personal information layers on published maps.

The capacity to operate basic spatial functions (e.g. spatial search, measurement or proximity search, overlay function etc.) would be desirable additional functions for the end users, though not a first-order priority.

There is a considerable list of open source and proprietary GIS server technologies available for managing spatial data. ArcGIS Server, ArcGIS Online, and MapInfo Spatial Server are among the most commonly used proprietary options. On the other hand, GeoServer, OpenGeo Suite or GeoNode are some examples of widely used open source web-based geospatial content management platforms. The World Bank’s Platform for Urban Management and Analysis (PUMA), currently under development, is also a potential open source option for the City Planning Labs. Apart from the initial cost difference, fast setup time, and off-the-shelf availability of functions...
are the main advantages of proprietary platforms. On the other hand, open source platforms allow for a higher level of customization.

Some Indonesian cities and agencies have already developed their own platforms (Figure 17) using both proprietary and open source options. For example, for their permitting systems, Surabaya has utilized ArcGIS Online, while Balikpapan has developed its open source platform. At the national level, BIG, has developed an integrated open-source/proprietary (ArcGIS online) data-sharing platform. In their decision for data platform options, CPLs’ should consider platforms the respective cities are currently using.

In addition to geospatial content management software, each CPL requires server space for storing its geospatial content. CPLs can host their geospatial data on local servers or use networked enterprise storage systems, such as cloud storage on Amazon Web Services (AWS). While storing data on a cloud can be more expensive than storage on a local server, outsourcing could also provide other benefits. Cloud storage systems are typically more stable than local hosts, particularly resilient to power shutdowns and human errors, keeping data constantly online. Using cloud storage also shifts the maintenance burden from CPLs to the service provider. Professional data storage systems offer qualified technical assistants with service contracts. Geospatial content management software can be used to set up different levels of security access to different data users.

It is possible that CPLs use a different platforms in short and long run. ArcGIS online in combination with Amazon cloud storage, for instance, could be an option for short term, as it is easy and fast to set up with low initial cost, while CPLs develop their customized open-source permanent platforms.
Figure 19. One of the data platforms of the City of Cambridge - Cambridge City Viewer - through which a large amount of geospatial data including buildings, parcels, paved surfaces, sidewalks, street centerlines, trees, and infrastructure systems is published.

Figure 20. Census block level demographic data of Cambridge, Boston and Somerville, from Census Bureau platform.
SPATIAL GROWTH AND CHANGE ANALYTICS
3.1. SPATIAL GROWTH AND CHANGE

Monitoring trends in spatial data is a very useful analytical technique that can provide valuable information for planners. Effective planning decisions are built upon short and long-term projections of the spatial distribution of resources and demand. These projections are often made by extrapolating current trends in spatial data. Forecasting the spatial distribution of resources (e.g., housing, jobs, agricultural land) and demand (represented, for instance, by population and income) relies on an understanding of current and past trends.

While projected spatial values are crucial inputs for more complex statistical models, a simple comparison of projected values with benchmarks from other cities or existing conditions can provide a ground for evidence-based decision-making. Based on projected values, planners can decide whether intervening actions should be taken to strengthen, slow down or reverse current trends.

Monitoring the shifts in demographic data, for instance, can reveal a significant increase in the number of households with young children in peripheral areas of a city, which can increase the demand for schools, hospitals, or food resources. Evidence of such a growing demand allows planners to decide if new facilities are needed to support this trend or if policies are required slow it down. Understanding the rate of population growth should form the basis for land-use planning.

Although monitoring trends in spatial data does not explain their underlying causes, and thus cannot suggest what interventions will affect them, it can at least frame the issues that planners should focus on.

Figure 21. Indonesian cities are facing rapid growth with an average annual urbanization rate of 4.2% between 1993 and 2007.
3.1.1. Mapping Change in Spatial Values

A convenient way of monitoring, storing, and representing change in spatial properties is to include change values – either absolute change or its rate – in the attributes of each spatial feature (Figure 26). For example, census blocks can contain attributes that represent the change in their building stock, particular demographic group, employment, land use coverage, energy consumption or per capita area of green space in a given period of time.

3.1.2. Spatiotemporal data

If every spatial feature contains start and end time data, we can visualize snapshots of different points of time, using only one dataset (Figure 27). Spatiotemporal databases are useful when spatial units themselves change over time, not merely the value of their attributes. Changes in building stock or business establishments can be best stored by spatiotemporal databases, as the spatial features change over time – some buildings are demolished and some are added to the building stock over time.

3.1.3. The expansion of urban extent

The most fundamental aspect of growth and change in cities is how much and where urban areas are emerging. The expansion in the boundary of cities represents a shift in land use coverage. The main source of land for a city’s growth is usually agricultural or forest land. Given the significant role of agriculture sector in Indonesia’s economy, annual loss of agricultural land can significantly reduce its cities’ food security.

Figure 22. Spatiotemporal datasets. Each feature have start and end time, allowing for monitoring change in spatial features: for example in their size, location or existence.
Overlaying snapshots of the city’s extent over time can capture trends in physical expansion. A future extent can be projected based on the past. However, distinguishing the urban land coverage from non-urban land extent of a metropolitan area is not a trivial task, and depends on the definition of “urban land”. Enhancements in remote-sensing technologies and availability of high-resolution satellite imagery have made accurate geo-referenced maps of urban extent available for many cities. Extent boundaries for most medium to large cities in Asia can be obtained upon request from Annemarie Schneider at the University of Wisconsin. By overlaying different snapshots of urban extent over time can not only capture the growth rate, but also its character. We can distinguish between leapfrog growth, edge growth or infill developments just by overlaying the extent maps of different times (Figure 23).

Projecting the future extent of a city also requires information about buildable area around it. Geographical constraints such as water bodies, steep lands, or protected forests pose a barrier to growth and should be accounted for in growth projections. The availability of space not only affects the rate of possible growth but also its character. Cities that are constrained by geographic features, such as water bodies or steeply sloped land, grow very differently from those with no barriers around them. The former, for instance, leave no room for leapfrog development, set serious limits on sprawl, and tend to develop at higher densities (e.g., Hong Kong). The latter allow for lower density development and fragmented growth (e.g., Los Angeles). Note that

The growth and change mapping we have discussed so far, examined trends in a single variable over time. Analytical techniques that are presented in the following, Accessibility Analysis, Impact analysis, and spatial statistical models, can be used to study change among multiple sets of spatial data.

2 A more accurate projection of boundary requires controlling for determinants such as population, and economic performance of the city. This requires regression models, which are explained later in this report.
Figure 26. Population growth between 1990 and 2000 in Cambridge. Blocks represented in white experienced the lowest change (growth or decline) in the population among all census blocks of Cambridge.
3.2.ACCESSIBILITY ANALYSIS

Understanding how accessible the resources of a city are to people is key to planning new infrastructures. Accessibility analyses investigate how locations of one group of phenomena (origins) are related to another group of phenomena (destinations). These spatial relations can be described in various ways. Accessibility can be assessed along transportation networks, or along idealized continuous space that simplifies constraints to movement (Figure 27). It can also be described geometrically (based on distance) or topologically (e.g. based on number of turns on a network, or number of steps in a topological grid). In this report we focus on examples of accessibility analyses that are conducted over urban transportation networks. Accessibility measures such as Gravity and Reach are computed at a fine resolution for individual buildings or address points over a network of city streets.

Accessibility analysis helps us identify underserved areas, for instance, areas with low pedestrian accessibility to schools (Figure 31). Comparing accessibility values of individual origins across the city, or comparing their values to benchmarks from other cities, allows us to detect areas with problematic accessibility values.

Accessibility analyses also provide inputs to other analytics proposed in this report. They are key determinants of land and real property values, land use patterns, and business clustering. Accessibly values can be used in hedonic pricing models for land and real properties. Potential impacts of new network infrastructure, such as bridges, roads, bus routes, or sidewalks, can be analyzed using before and after accessibility values.

3.2.1.Accessibility Measures

Among accessibility metrics, Reach and Gravity metrics are simple to specify and can be interpreted most intuitively. Reach and Gravity metrics can be computed over real transportation networks, and be implemented for various spatial units, including buildings, address points, parcels or KT zones.
Computing accessibility values over network and at building or address point level can capture a detailed image of proximity to a city’s resources at an individual household level.

**Reach**

The Reach metric counts the number of resources that can be reached from an origin within a given search radius over a network of paths (Figures 28, 31, 32). For individual buildings, for instance, it can tell us how many jobs, schools or wells are available in a 10 minutes walking radius around it. The metric doesn’t capture the variation in distance to different reachable resources; it simply counts all destinations within the given radius. If we compute accessibility to retail spaces, an establishment located 600 meter away from the origin is treated the same as one that is only 50 meter away from the origin. This drawback is addressed by Gravity metric. The advantage of the Reach metric is that it is intuitive to understand and communicate to multiple stakeholders – everyone can understand what it means to have two schools within 10 minutes walking radius as opposed to none.

The reach centrality, $R^r_i$, of a building $i$, in a street network $G$ describes the number of other buildings in $G$ that are reachable from $i$ at a shortest path distance of at most $r$. It is defined as follows:

$$R^r_i = \sum_{j \in G - \{i\}} d[i,j] \leq r \cdot W[j]$$

where $d[i,j]$ is the shortest path distance between nodes $i$ and $j$ in $G$ and $W[j]$ is the weight of destination node $j$. Figure 32 illustrates the implementation of Reach to jobs in Cambridge and Somerville, MA within a 600m walking radius along the available street network.
**Gravity**

Similar to Reach, the Gravity metric counts the number of resources that can be reached from an origin within a search radius over a network, but additionally accounts for their distance from the origin (Figure 30). A building with a few shops located next door will get a higher accessibility value to commercial establishments than a building with a large number of retail establishments that are located far away.

The attraction of destinations does not drop linearly when their distance from the origin increases, but at an exponential rate, and it varies for different modes of transport. The inverse exponent of distance is often used instead of simple inverse distance for weighing destinations, and the distance decay rate is controlled by a coefficient for each mode of transport. Gravity of point $i$, in graph $G$, can be specified as:

\[
Gravity[i]_r = \sum_{j \in G \setminus \{i\}, d[i,j] \leq r} \frac{W[j]}{d[i,j]^b}
\]

where $Gravity[i]_r$ is the gravity index at point $i$ in network $G$ within search radius $r$, $W[j]$ is the weight of destination $j$, $d[i,j]$ is the shortest distance between $i$ and $j$, and $b$ is the exponent for adjusting the effect of distance decay.

These accessibility measures can be specified in the Urban Network Analysis Toolbox in ArcGIS. A more exclusive help document is available with the toolbox to explain the specifications in detail. In order to run the toolbox, ArcGIS 10 and the network Analyst extension are required.

Figure 30. Illustration of the Gravity metric.
Figure 31. Underserved areas; the areas in orange don’t have access to public schools within a 1200 meter network radius. Overlaying the underserved areas and census data shows that approximately 10,000 people don’t have walking access to public schools.
Figure 3.2. Reach to jobs located within 600 meter network radius (approximately 10 minutes walk) in Cambridge, MA.
3.3. SPATIAL-STATISTICAL MODELS

Unlike accessibility and growth analytics, spatial-statistical models can examine relationships between more than two spatial values. Growth and change analytics each capture over-time change in only one property of a spatial unit. Spatial-statistical models, however, can examine the relationship of one spatial value to a number of other variables, and are thereby better suited for projecting changes under more realistic multivariable scenarios. Statistical models can be developed to examine the relationship of land prices to various determinants including accessibility (e.g. to bus stops, retail establishments etc.), frontage, area, or parcel type. Having such explanatory models, one can then predict how the value of each individual parcel is likely to change when, for example, a new bus stop or road is constructed, controlling for covariates.

3.3.1. Regression Analysis

In statistics, regression analysis examines whether and how a dependent variable is related to one or more independent variables. The results of a regression function generate coefficients for the effects that each of the independent variables has on the dependent variable and an indication of whether and how significant these effects are. The models also tell us how much of the total variation in the dependent variable is explained by variations in the given independent variables. Those coefficients that are found to be significant can be used to predict future changes under similar conditions.

Hedonic pricing models, which form one type of regression models, are widely used for projecting land or real estate value. In these models, the selling price of a real property (e.g. a housing unit) is predicted based on a linear function of the characteristics of the unit – age, size, number of rooms, structural quality, accessibility, ownership structure, lot size etc.
An accurate specification of regression models requires the use of specialized software like SAS, Stata or SPSS and necessitates a clear understanding of concepts and assumptions that regression analysis is grounded on. It is recommended that these models be specified by only staff who have had proper training in regression analysis and understand their foundations thoroughly. Simple multiple regressions and bivariate scatter plots can also be specified in MS Excel, using the analysis toolbox.

### 3.3.2. Trend Estimation and Autoregressive Models

Trend estimation is a form of regression analysis, where time is the only linear predictor of the dependable variable. Trend estimation examines a correlation between the outcome values and time at which they took place. Trend estimation is suitable for projecting the long-term trend in variables whose key determinants are not fully known but a pattern in their values can be identified over time. We may not know, for instance, what variables can predict the increase of travelers to the city center, but a trend with a significant yearly time coefficient may be used to observed project the number based on past observations. Even if the data oscillated up and down, a trend regression can help us determine whether a significant long-term increase or decrease is present (Figure 33).

The value of variables sometimes follows a cyclical pattern over time, where variables at one observation period are dependent on values during the previous period. Energy consumption in a census block, for instance, may follow a cyclical pattern, following the winter-summer cycle in the environment. Cyclical patterns in data may be independent of the overall long-term trend. While there may be a winter-summer cycle in the energy consumption at the household level, the long-term trend may be insignificant (the total annual energy consumption not changing), even when the energy consumption at a particular period (spring) may exhibit a cyclical decrease.

Autoregressive models are used to predict over-time changes in variables with cyclical patterns, where independent variables include the value of the
dependent variable in the previous measurement period, as well as a linear time predictor that may or may not be significant for the long-term trend (Figure 34). More than one time lag variable can be used to capture longer cyclical effect and the linear time variable can be squared to capture nonlinear effects.

Both linear trend analysis and autoregressive analysis can be applied to a number of important planning problems in cities. Trend analysis can capture the long-term pattern in key urban growth indicators – annual rural to urban land conversion, increase in residents or jobs, city GDP change, growth in transit ridership, etc. Cyclical trend analysis can capture predicted land and real estate values, seasonal changes in resource consumption or cyclical patterns in construction permit applications.

### 3.3.3. Spatial Regression Analysis

There are various techniques for carrying out regression analysis, but common assumptions underlie most ordinary least squares (OLS) regression techniques. One of the underlying assumptions is that the dependent variable on the left-hand side of the regression equation can interact with independent variables on the right-hand side, but separate observations of the dependent variable are independent of each other. The price of land may depend on several independent factors, such as lot size, location and building height, but it should not depend on the price of land of the neighboring parcel. In reality, this assumption may not hold; land values can depend on neighboring land values around them.

This independent distribution of the dependent variable assumption of OLS regressions is relaxed in spatial lag and error type models. Spatial autocorrelation models allow either the dependent variable to depend on adjacent dependent variables or the error terms of adjacent observations to be correlated. The former case can be modeled by the “spatial lag models,” and the latter by the “spatial error models” (Anselin, 1988).
Spatial regressions can be specified in GeoDa or GeoDa Space software packages that are freely distributed. Many phenomena in spatial analyses exhibit spatial autocorrelation which such models capture. If spatial autocorrelation is present then OLS autoregressive models yield a better explanation to the variations in observed data (Figures 35, 36 & 37).

Figures 35 and 36 provide an example of assessed land value distribution in Cambridge, MA. A spatial lag regression is specified in GeoDa to predict how the per-square-foot value of land depends on four independent variables: plot ratio, parcel size, access to streets and distance from the nearest subway station. In Figure 35 (top) an OLS model is specified with all four variables, but without allowing for autocorrelation between neighboring parcels. Below, a spatial lag model is specified, which adds spatial autocorrelation in the dependent variable (price per square foot) to the estimation ("W_LV_PSF" in the model). The high z-values suggest that land values in Cambridge are indeed strongly correlated with neighbors – beautiful improvements in the neighbors' yard can significantly increase surrounding land values. Figure 36 plots the bivariate effect of proximity to subway stations, showing how land values decrease as the distance to the nearest subway station increases, controlling for other variables.
Figure 36. The relationship between the land value (US$/Sq.Ft) and distance to subway station predicted by the spatial lag model. All other predictors of land value are kept constant.

Figure 37. Containing land value (US$/Sq.Ft.), land area and aggregated building data (such as total floor area), the parcel dataset of Cambridge, MA, provides the basis for the hedonic pricing model for land. Each parcels’ distance to subway stations, is computed using the network analyst of ArcGIS, which requires street network and transit station locations datasets.
3.4. IMPACT ANALYSIS

Effective decisions in policy and spatial planning need to be evaluated before their implementation. Such evaluations require an understanding of the potential impacts of the decisions. Comparing the probable impacts of a series of alternative decisions to the existing conditions allows planners to establish a concrete base for informed decision-making.

Impact analysis includes two broad groups of analytics. The first group includes analytics and statistical models that can predict the impact of a proposed policy (e.g. zoning) or spatial intervention (e.g. sidewalk improvement) on an outcome variable, such as land value, accessibility, crime rate or employment. The second type of analysis keeps track of changes in question variables (e.g. land value, or pedestrian movement) before, during and after an intervention. The latter provides useful empirical evidence upon which future planning decisions can be made.

Accessibility analyses and regressions can be used as inputs to impact analysis, since they can capture changes in a variable when spatial conditions change. To analyze the impact of a new bridge or bus route, for instance, accessibility analyses can be used to measure the change in accessibility values – how much the citizens’ accessibility to jobs changes when a new bridge is built over the river.

Utilizing a hedonic pricing model allows us to assess the impact of the same bridge on land values across the city. The output of accessibility analysis, which computes the changes in accessibility values, can be used in the hedonic pricing model for land, in which accessibility values form key independent variables. If other variables are kept constant, the reflected land value change reveals the pure impact of the new bridge on land values.

Figure 38 illustrates the impact of how a hypothetical highway that cuts through the Geylang neighborhood in Singapore, on the accessibility of buildings to businesses. Comparing the gravity index from buildings to businesses before and after the proposed highway in a local 1km walking
range shows a 56% decline, on average, in accessibility to business establishments. The impact of change in accessibility on other key variables such as land value can be then analyzed by a spatial statistical model (Figure 39).

Figures 39 and 40 below model the potential impact of a new subway station on land values in Cambridge, MA, using the hedonic pricing model for land in Cambridge, based on the present land values (See Figure 35). The example in Figure 39 demonstrates the estimated difference in land values before and after the proposed subway station. Using coefficient estimated in the spatial lag model of Figure 35, the total hypothetical change in land values that could result from adding the new subway stop is around $20 million. The distribution of the new per square foot prices is shown in Figure 40.
Figure 38. Comparison of the accessibility values before (left) and after a highway cut through Geylang, Singapore shows a significant drop in the local gravity to business establishments (center). Such comparisons in accessibility values can be used as input to regression models for predicting the other impacts of a spatial intervention (see Figure 39 and 40). The percentage change in accessibility to businesses as a result of the proposed highway is shown on the right.
Figure 39. Analyzing the impact of a proposed subway station in north Cambridge on the value (US$/Sq.Ft) of lands within 10-minute walk around the proposed station, using the hedonic pricing model for land developed in the previous section (see Figure 35). The figure shows a comparison between the present values (left) and the predicted values (right).
Figure 40. As a result of the new subway, land values increase 8% in dollars per square foot on average, which is approximately $20,000,000 in total for all parcels located within 600 meters from the proposed subway station.
PLANNING DECISIONS SUPPORT
4. PLANNING DECISION SUPPORT

The ultimate goal of CPL data collection and spatial growth analyses is to support cities in their planning decisions with concrete evidence. Informing planning decisions by measurable evidence does not always require complex analytics; planning evidence can sometimes be directly extracted from raw spatial data.

In the previous sections we discussed geospatial data, a number of analytical activities and their potential applications for urban planning. The way these analytical techniques can inform planning decisions can be summarized as follows:

1) By describing qualities of space in measurable terms, analysis of spatial data makes it possible to compare existing condition to certain benchmarks, and to thereby inform planners of present challenges. Spatial data and analytics help cities identify problems and frame questions they should focus on. Benchmarks can be chosen to meet a city’s goals and ideals based on national or international examples, or based on more complex underlying causes of the observed patterns.

   For example, by looking at rental payments as a share of household income in census tracts and comparing that to desired ratios, one can directly assess whether some household income groups are paying too large a portion of their monthly income on housing. In other cases, the assessment may require several analytical steps. Accessibility analysis, for instance, can inform planners whether access to key infrastructures or facilities is underserved in certain area, and if so, informs planners where such areas are located.

2) In addition to identifying existing challenges, spatial data and analytics can be used to identify forthcoming challenges. Trend estimation analyses can describe probable forthcoming issues by comparing the predicted value of a variable to its desired
benchmark value. Keeping track of trends in the number of multi-family building permits that are annually issued, and trends in demographic groups that form typical occupants for such units, planners can predict whether the city is headed toward shortages or oversupply in the market for multi-family housing.

3) Geospatial data and analytics can inform planners of underlying interactions and correlations between spatial variables. Representing spatial qualities with numeric data allows us to utilize statistical regression analysis to explain relationship between such variables. Spatial-statistical models can be used to identify the determinants of observed socio-economic variables. Statistical models can outline spatial conditions that require change in order to improve socio-economic indicators. If a model shows a strong negative correlation between the existence of commercial establishments that face directly to streets and crime rates on these streets, planner may use this evidence for deciding where to allocate commercial space in zoning plans.

A statistical model that analyzes previous sidewalk improvement outcomes may reveal a positive correlation between sidewalk quality and business revenue along sidewalks. One interpretation of this correlation is that sidewalk improvement can be an effective tool for increasing business revenue in areas where proper sidewalks do not exist but other preconditions for commerce are in place. The correlation coefficient of the model can be used to estimate how much business owners could benefit from such public investment, and whether they could be involved in financing sidewalks through taxation.

4) Impact analyses allow planners to assess different future investment or improvement scenarios based on a key outcome variable. A comparison of different alternatives can allow one to identify the most impactful scenario.
The spatial analysis techniques that form the focus of the CPLs’ core module include a) spatio-temporal change mapping, b) accessibility assessment, c) trend analysis, d) spatial regression analysis and e) impact analysis. Rather than elaborating on any one application of these techniques at greater length, we have tried to provide a brief overview of the nature and utility of each technique, pointing towards applications for various urban planning and management tasks. The exact application focus of the techniques in the four participating CPL cities – Surabaya, Denpasar, Palembang and Balikpapan – should be identified together with the local government representatives and CPL staff. The analysis should be chosen to address the most important spatial analysis and planning questions specific to each city.

5. References


