Chapter 2
Review of Literature

Abstract The literature on urban growth and sprawl is bigger than the scope of this chapter. Application of remote sensing and GIS in the study of urban growth/sprawl is also extensive in terms of their variations. A general research review cannot embrace the totality. However, in this chapter, efforts have been made to briefly document the important aspects that mainly focus on urban growth and sprawl, their patterns and processes, and measurement/analysis. It has been found that sprawl as a concept suffers from difficulties in definition. As a result, quantification of this phenomenon is rather complicated and sometimes confusing, especially if we use remote sensing data. There are dozens of metrics that are practiced by the urban planners and administrators in their cities, especially in developed countries. Merits and demerits of these measurements/analytical techniques have also been addressed briefly. It also documents similar studies in India, and proceeds toward the proposed research by documenting the scope of research.

Keywords Urban growth · Sprawl measurement · Remote sensing · GIS · Sprawl technique advantages

2.1 Urban Growth and Sprawl

The spatial configuration and the dynamics of urban growth are important topics of analysis in the contemporary urban studies. Several studies have addressed these issues which have dealt with diverse range of themes (e.g., Acioly and Davidson 1996; Wang et al. 2003; Páez and Scott 2004; Zhu et al. 2006; Hedblom and Soderstrom 2008; Geymen and Baz 2008).

Urban sprawl, as a concept, suffers from difficulties in definition (Johnson 2001; Barnes et al. 2001; Wilson et al. 2003; Roca et al. 2004; Sudhira and Ramachandra 2007; Angel et al. 2007; Bhatta 2010). Galster et al. (2001) critiqued the
conceptual ambiguity of sprawl observing that much of the existing literature is ‘lost in a semantic wilderness.’ Their review of the literature found that sprawl can alternatively or simultaneously refer to: (1) certain patterns of land use, (2) processes of land development, (3) causes of particular land-use behaviors, and (4) consequences of land-use behaviors. They have reviewed many definitions of sprawl from different perspectives. It seems that sprawl is used both as a noun (condition) and as a verb (process), and suffers from lack of clarity even though many would claim to ‘know it when they see it’ (e.g., Ewing 1994).

Although accurate definition of urban sprawl is debated, a general consensus is that urban sprawl is characterized by unplanned and uneven pattern of growth, driven by multitude of processes and leading to inefficient resource utilization. The direct implication of such sprawl is change in land-use and land-cover of the region as sprawl induces the increase in built-up and paved area (Sudhira and Ramachandra 2007). It is worth mentioning that opinions on sprawl held by researchers, policy makers, activists, and the public differ sharply, and the lack of agreement over how to define sprawl certainly complicates the efforts to characterize and restrict this type of land development. Researchers have made many attempts to characterize the sprawl as shown in Table 2.1.

Table 2.1 clearly shows that sprawl can be analyzed from many perspectives and it may vary from case to case. However, most of the researchers have emphasized the density (e.g., built-up density, population density, and housing density). Therefore, it can be considered that the best characterization of sprawl is considering the density criterion.

2.2 Physical Patterns and Forms of Urban Growth and Sprawl

Wilson et al. (2003) identified three categories of urban growth: infill, expansion, and outlying; with outlying urban growth further separated into isolated, linear branch, and clustered branch growth (Fig. 2.1). The relation (or distance) to existing developed areas is important when determining what kind of urban growth has occurred.

Indeed all types of urban growth are not considered as sprawl; one development that can be considered as sprawl by someone may not be considered by others (Roca et al. 2004). Furthermore, urban sprawl has a negative connotation, and not all urban growth is necessarily unhealthy. In fact, some types of urban growth (e.g., infill growth) are generally considered as remedies to sprawl. Therefore, sprawl cannot be characterized by the simple quantification of ‘the amount of land that has changed to urban uses.’ Sprawl phenomenon should be treated separately than the general urban growth. Ewing (1994) reviewed several patterns of urban sprawl, and argued that the pattern of sprawl is ‘like obscenity’, the experts may know sprawl when they see it.
Table 2.1 Varying characterization of sprawl (after Torrens 2008)

<table>
<thead>
<tr>
<th>Growth</th>
<th>Social</th>
<th>Aesthetic</th>
<th>Decentralization</th>
<th>Accessibility</th>
<th>Density</th>
<th>Open space</th>
<th>Dynamics</th>
<th>Costs</th>
<th>Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audirae et al. (1990)</td>
<td>✔️</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>Bae and Richardson (1994)</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>Benfield et al. (1999)</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>Burchell et al. (1998)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Calthorpe et al. (2001)</td>
<td>✔️</td>
<td>✔️</td>
<td>❌</td>
<td>❌</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Clapham (2003)</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>Duany et al. (2001)</td>
<td>✔️</td>
<td>✔️</td>
<td>❌</td>
<td>❌</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>El Nasser and Overburg (2001)</td>
<td>✔️</td>
<td>✔️</td>
<td>❌</td>
<td>❌</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Ewing (1997)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Ewing et al. (2002)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Farley and Frey (1994)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Galster (1991)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Galster et al. (2001)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Gordon and Richardson (1997a)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Gordon and Richardson (1997b)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Hasse and Lathrop (2003a)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Hasse and Lathrop (2003b)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Hasse (2004)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>HUD (1999)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Ledermann (1967)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Lessinger (1962)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Malpezzi (1999)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>OTA (1995)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Peiser (1989)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Pendall (1999)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>RERC (1974)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Sierra Club (1998)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Sudhira et al. (2004)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>
Galster et al. (2001) considered the urban sprawl both as a pattern of urban land use (a spatial configuration of a metropolitan area at a specific time); and as a process, namely as the change in the spatial structure of cities over time. Sprawl as a pattern or a process is to be distinguished from the causes that bring such a pattern about, or from the consequences of such patterns (Galster et al. 2001). If the sprawl is considered as a pattern it is a static phenomenon and as a process the sprawl is a dynamic phenomenon. Some of the researchers have considered sprawl as a static phenomenon, whereas some have analyzed it as a dynamic phenomenon; however, most of the researchers shout for both.

Sprawl, as a pattern, helps us to understand the spatial distribution but as a static phenomenon, in fact areas described as sprawled are typically part of a dynamic urban scene (Harvey and Clark 1965; Ewing 1997). The dynamics of sprawl process can be understood from the theoretical framework of urban growth process. Herold et al. (2005b) present a hypothetical schema of urban growth process using a general conceptual representation as shown in Fig. 2.2. According to them, urban area expansion starts with a historical seed or core that grows and disperses to new individual development centers. This process of diffusion continues along a trajectory of organic growth and outward expansion. The continued spatial evolution transitions to the coalescence of the individual urban blobs. This phase transition initially includes development in the open space in interstices between the central urban core and peripheral centers. This conceptual growth pattern continues and the system progresses toward a saturated state. In Fig. 2.2, this ‘final’ agglomeration can be seen as an initial urban core for further urbanization at a less detailed zoomed-out extent. In most traditional urbanization-studies this ‘scaling up’ has been represented by changing the spatial extent of concentric rings around the central urban core.

The preceding framework suggests that some parts of an urban area may pass through a sprawl stage before eventually thickening so that they can no longer be characterized as sprawl. However, from this point of view, what, when, and where...
it can be characterized as a sprawl becomes complex. Therefore, sprawl as a process without considering the pattern cannot be characterized. Rather, it should be considered as a pattern at multiple dates. “In any event, measuring the respective dimensions of development patterns for an urban area at different times will reveal the process (or progress) of sprawl” (Galster et al. 2001).

One may refer Bhatta (2010) for detailed discussion on urban growth pattern and process. The title also addresses the causes and consequences of urban growth and sprawl, smart and sustainable urban growth, and urban planning policies to restrict the sprawl. The title cites several references that may be useful for the study of urban growth/sprawl. Since the primary goal of current research is not focused on urban planning, rather focused on the analysis of urban growth and sprawl from remote sensing data, the documentation of review on urban growth/sprawl/planning has been kept limited.

2.4 Remote Sensing for Analysis of Urban Growth and Sprawl

Understanding the urban patterns, dynamic processes, and their relationships is a primary objective in the urban research agenda with a wide consensus among scientists, resource managers, and planners, because future development and management of urban areas require detailed information about ongoing processes and patterns.

Remote sensing, although challenged by the spatial and spectral heterogeneity of urban environments (Jensen and Cowen 1999; Herold et al. 2004), seems to be an appropriate source of urban data to support such studies (Donnay et al. 2001). It is irrefutable that the earth observation is a modern science, which studies the earth’s
changing environment, through remote sensing tools such as satellite imagery and aerial photographs (EEA 2002). A report published by NASA highlighted the fact that the advances in satellite-based land surface mapping are contributing to the creation of considerably more detailed urban maps, offering planners a much deeper understanding of the dynamics of urban growth and sprawl, as well as associated matters relating to territorial management (NASA 2001).

In terms of analyzing urban growth, Batty and Howes (2001) believe that remote sensing technology, especially considering the recent improvements, can provide a unique perspective on growth and land-use change processes. Data sets obtained through remote sensing are consistent over great areas and over time, and can provide information at a great variety of geographic scales. The information derived from remote sensing can help to describe and model the urban environment, leading to an improved understanding that benefits applied urban planning and management (Banister et al. 1997; Longley and Mesev 2000; Longley et al. 2001).

Analysis of urban growth from remote sensing data, as a pattern and process, helps us to understand how an urban landscape is changing through time. This understanding includes: (1) the rate of urban growth, (2) the spatial configuration of growth, (3) whether there is any discrepancy in the observed and expected growth, (4) whether there is any spatial or temporal disparity in growth, and (5) whether the growth is sprawling or not.

In the recent years, remote sensing data and geographic information system (GIS) techniques are widely being used for mapping (to understand the urban pattern), monitoring (to understand the urban process), measuring (to analyze), and modeling (to simulate) the urban growth, land-use/land-cover change, and sprawl. The physical expressions and patterns of urban growth and sprawl on landscapes can be detected, mapped, and analyzed by using remote sensing data and GIS techniques. The decision support systems within the GIS can evaluate remote sensing and other geospatial datasets by using multi-agent evaluation (Axtell and Epstein 1994; Parker et al. 2003) which can also predict the possibilities in the subsequent years using the current and historical data. In the last few decades, these techniques have successfully been implemented to detect, analyze, and model the urban growth dynamics. All these issues have been reviewed and documented with a greater detail in Bhatta (2010).

### 2.5 Measurement and Analysis of Urban Growth and Sprawl

The process of mapping urban growth results in the creation of abstracted and highly simplified change maps of the study area (Singh 1989). Recently, urban change detection focus has shifted from detection to quantification of change, measurement of pattern, and analysis of pattern and process of urban growth and sprawl.

Urban growth can be quantified by measuring the built-up change between two dates (Singh 1989; Jensen 2005). However, quantification of sprawl, as a pattern or
process, is a real challenging issue (Bhatta et al. 2010a). Wilson et al. (2003) said that without a universal definition, quantification and modeling of urban sprawl is extremely difficult. They argue that creating an urban growth model instead of an urban sprawl model allows us to quantify the amount of land that has changed to urban uses, and lets the user decide what he or she considers as urban sprawl. Angel et al. (2007) also support this concept. This statement, however, discourages the efforts of quantifying the sprawl and makes the sprawl phenomenon more ambiguous. “Although there have been many studies on the measurement of urban form they have limitations in capturing the characteristics of urban sprawl” (Yeh and Li 2001); the results obtained from such measurement processes are often not easily interpretable. Several limitations of remote sensing data have made the quantification processes more difficult (Du et al. 2002; Prenzel 2004; Paolini et al. 2006; Hardin et al. 2007).

Sprawl can be measured in relative and absolute scales (Bhatta et al. 2010a). Absolute measurements are capable to create a black-and-white distinction between a sprawled city and a compact city. Relative measures, in contrast, quantify several attributes that can be compared among cities, among different zones within a city, or among different time (date) for a city. In case of relative measures, whether the city is sprawled or not is generally decided by the analyst, or even left without characterizing the sprawl. It is important to mention that most of the sprawl measurement techniques, in general, are relative measures, and can be used as indicators of sprawl. Absolute measurement of sprawl is never possible with these measures, unless we define a threshold toward the black-and-white characterization of sprawling and nonsprawling. Defining a threshold is not an easy task. Researchers have made their own assumptions toward defining this threshold, which are even less clear to the scientists (Bhatta et al. 2010a). It is important to realize that relative measures, most often, fail to draw conclusion on sprawl, and a threshold used in one area cannot be used in other areas reliably. These measures may serve the scientific purposes well, but can never become a technology, because to interpret the results one has to be a scientist. Therefore, how these techniques can become a tool for a city administrator is an obvious question.

Many metrics and statistics have been used to measure the sprawl. These metrics are generally known as spatial metrics. Spatial metrics are numeric measurements that quantify spatial patterning of land-cover patches, land-cover classes, or entire landscape mosaics of a geographic area (McGarigal and Marks 1995). These metrics have long been used in landscape ecology [where they are known as landscape metrics (Gustafson 1998; Turner et al. 2001)] to describe the ecologically important relationships such as connectivity and adjacency of habitat reservoirs. Applied to research fields outside of landscape ecology and across different kinds of environments (in particular, urban areas), the approaches and assumptions of landscape metrics may be more generally referred to as spatial metrics (Herold et al. 2005a). Spatial or landscape metrics, in general, can be defined as quantitative indices to describe structures and patterns of a landscape (O’Neill et al. 1988). Herold et al. (2005a) defined it as ‘measurements derived
from the digital analysis of thematic-categorical maps exhibiting spatial heterogeneity at a specific scale and resolution.

Spatial metrics have found important applications in quantifying urban growth, sprawl, and fragmentation (Hardin et al. 2007). Based on the work of O’Neill et al. (1988), sets of different spatial metrics have been developed, modified, and tested (Hargis et al. 1998; McGarigal et al. 2002; Riitters et al. 1995). Many of these quantitative measures have been implemented in the public domain statistical package FRAGSTATS (McGarigal and Marks 1995; McGarigal et al. 2002).

Spatial metrics can be grouped into three broad categories: patch, class, and landscape metrics. Patch metrics are computed for every patch in the landscape, class metrics are computed for every class in the landscape, and landscape metrics are computed for entire patch mosaic. There are numerous types of spatial metrics that are found in the existing literature, for example: area/density/edge metrics (patch area, patch perimeter, class area, number of patches, patch density, total edge, edge density, landscape shape index, largest patch index, patch area distribution); shape metrics (perimeter-area ratio, shape index, fractal dimension index, linearity index, perimeter-area fractal dimension); core area metrics (core area, number of core areas, core area index, number of disjunct core areas, disjunct core area density, core area distribution); isolation/proximity metrics (proximity index, similarity index, proximity index distribution, similarity index distribution); contrast metrics (edge contrast index, contrast-weighted edge density, total edge contrast index, edge contrast index distribution); contagion/interspersion metrics (percentage of like adjacencies, clumpiness index, aggregation index, interspersion and juxtaposition index, mass fractal dimension, landscape division index, splitting index, effective mesh size); connectivity metrics (patch cohesion index, connectance index, traversability index); diversity metrics (patch richness, patch richness density, relative patch richness, Shannon’s diversity index, Simpson’s diversity index, Shannon’s evenness index, Simpson’s evenness index); and many others (McGarigal and Marks 1995; McGarigal et al. 2002).

The question is which spatial metrics are most appropriate for the measurement and analysis of sprawl. Galster et al. (2001) identified eight conceptual dimensions of land-use patterns for measuring the sprawl. Under the name of sprawl metrics, Angel et al. (2007) have demonstrated five metrics for measuring manifestations of sprawl and five attributes for characterizing the sprawl. Under each attribute they have used several metrics to measure the sprawl phenomenon. However, they had not recommended any threshold that could be used for distinguishing a sprawling city from a non-sprawling city. Furthermore, interpretation of results from these metrics is also difficult and confusing since metrics are huge in number and one may contradict with other. Refer Bhatta et al. (2010a) for more detailed discussion.

Sierra Club (1998) ranked major metropolitans in USA by four metrics, including: (1) population moving from inner city to suburbs; (2) comparison of land use and population growth; (3) time cost on traffic; and (4) decrease of open space. USA Today (2001) puts forward the share of population beyond standard metropolitan statistical area as an indicator for measuring sprawl. Smart Growth

Review of Literature
America (Ewing et al. 2002) carried out a research to study the impacts of sprawl on life quality in which four indices had been used to measure urban sprawl: (1) residential density; (2) mixture of residence, employment, and service facilities; (3) vitalization of inner city; and (4) accessibility of road network. All of these metrics were relative measures and failed to black-and-white discrimination of sprawling and nonsprawling.

Some of the researchers have also contributed to measuring sprawl by establishing multi-indices by GIS analysis or descriptive statistical analysis (Nelson 1999; Kline 2000; Torrens 2000; Galster et al. 2000, 2001; Hasse 2004). These indices cover various aspects including population, employment, traffic, resources consumption, architecture aesthetics, living quality, etc. Commonly used indices include: growth rate (of a population or built-up area); density (population density, residential density, employment density); spatial configuration (fragmentation, accessibility, proximity); and others (per-capita consumption of land, land-use efficiency etc.) (e.g., The Brookings Institution 2002; USEPA 2001; Fulton et al. 2001; Masek et al. 2000; Pendall 1999; Sutton 2003; Jiang et al. 2007). However, no one has provided any straight answer to the questions like: what should be the built-up growth rate in a non-sprawling city, or what should be the per-capita consumption of land in a non-sprawling city.

Torrens (2008) argues that sprawl should be measured and analyzed at multiple scales. In his (her) approach to measuring sprawl, (s)he has declared some ground-rules in developing the methodology. Measurements have been made to translate descriptive characteristics to quantitative form. The analysis is focused at micro-, meso-, and macro-scales and can operate over net and gross land. The analysis examines sprawl at city-scale and at intraurban levels—at the level of the metropolitan area as well as locally, down to the level of land parcels. Although interurban comparison and the use of remote sensing data are not focused on in this chapter, the methodology should be sufficient to be generalized to other cities using remote sensing data. The methodology devised a series of 42 measures of sprawl, which have been tracked longitudinally across a 10-year period. Although the author claims that this approach can provide a real insight of urban sprawl, however, the methodology became complex and resulted in confusion owing to the use of many scales and metrics.

Jiang et al. (2007) proposed 13 attributes under the name of ‘geospatial indices’ for measuring the sprawl in Beijing. Finally, they proposed an integrated urban sprawl index that combines the preceding 13 indices. This approach, indeed, minimizes the interpretation effort. However, their approach requires extensive inputs of multi-temporal data such as population, GDP, land-use maps, land-use master planning, floor-area ratio, maps of highways, and maps of city centers. Many developing countries lack such type of temporal data, and therefore, most of these indices are difficult to derive. Furthermore, they did not mention any threshold to characterize a city as sprawling or nonsprawling. However, this type of temporal analysis is useful to compare between cities or different zones within a city or status of a city at different dates. Whether a city is becoming more
sprawling or not, with the change of time, can be well depicted by this type of analysis.

The main problem associated with most of the available sprawl measurement scales is the failure to define the threshold between sprawling and nonsprawling. Although relative comparisons can provide us some insights into sprawl phenomenon and the associated city, but often these measures are not adequate and we need black-and-white characterization of sprawl. The second greatest problem is the number of metrics used for the measurement of sprawl. The preceding discussion shows that many scales and parameters are being used for the measurement of sprawl. The question is what the most stringent tools are or how effective they are. The answer is still awaited. Geoghegan et al. (1997), Alberti and Waddell (2000), Parker et al. (2001), and Herold et al. (2003) propose and compare a wide variety of different metrics for the analysis of urban growth. However, their comparisons do not suggest any standard set of metrics best suited for use in urban sprawl measurement as the significance of specific metric varies with the objective of the study and the characteristics of the urban landscape under investigation.

Important to mention, many metrics are correlated and thereby contain redundant information. Ritters et al. (1995) examined the correlations among 55 different spatial metrics by factor analysis and identified only five independent factors. Thus, many typical spatial metrics do not measure different qualities of spatial pattern. The analyst should select metrics that are relatively independent of one another, with each metric (or grouping of metrics) able to detect meaningful structure of urban landscape that can result in reliable measures of sprawl. It is often necessary to have more than one metric to characterize an urban landscape because one metric cannot say about all. However, the use of many metrics results in many measures which are often difficult to interpret resulting in difficulties for reaching to a black-and-white conclusion. Use of highly correlated metrics does not yield new information, rather makes interpretations more difficult. “Just because something can be computed does not mean that it should be computed” (Turner et al. 2001). Often, different metrics may also result in opposite conclusions; for example, in Herold et al. (2003), ‘number of patches’ within the time span 1929–1976 was increasing (an indication of sprawl); however, if one considers ‘mean nearest neighbour distance between individual urban patches’, it was decreasing (an indication of compactness).

Another challenge is the spatial resolution of remote sensing data. Many metrics, for example patch or spatial heterogeneity analysis, are dependent on spatial resolution. In a low spatial resolution image, individual objects may appear artificially compact or they may get merged together. In an area of low density development where houses are relatively far apart, a spatial resolution of 30 m will produce an estimate of developed land four times that produced using the same underlying data but a spatial resolution of 15 m. Apparently, the most preferred spatial data are those that are sufficiently fine scale to represent individual units, e.g., individual land parcels or houses. Indeed higher spatial resolution provides better interpretability by a human observer; but a very high resolution leads to a high object diversity which may end up in problems when a classification algorithm is applied to the data, or it
may produce a very high number of patches resulting in complications in metric analysis. Due to the increased heterogeneity in high resolution images, analysis of spatial association or spatial heterogeneity will also be influenced at a high degree. Further, for multi-temporal analysis of sprawl or measurement of sprawl as a process may include images from sensors having different resolutions. In such cases, resolution-dependent metrics are no longer usable.

Furthermore, statistical properties and behaviors of some metrics are not well known (Turner et al. 2001). In cases where a single number is reported for a landscape, we may have little understanding of the degree to which landscape pattern must change to be able to detect significant change in the numeric value of the metric. Therefore, the analyst should definitely consider the criteria that will be applied to determine whether the obtained result from a metric is meaningful or not.

The density of built-up and intensity of annual growth can efficiently depict the sprawl features of low density and strong change, but they are still weak in capturing the particular spatial patterns of urban sprawl. These metrics are not usually spatially explicit; for example the growth of built-up area to the growth of households in a city. This type of metrics measures what is present and in their relative amounts, or properties, without reference to where on the landscape they may be located.

Entropy method, another urban sprawl metric, is perhaps the most widely used technique to measure the extent of urban sprawl with the integration of remote sensing and GIS (Yeh and Li 2001; Lata et al. 2001; Li and Yeh 2004; Sudhira et al. 2004; Kumar et al. 2007). Relative entropy can be used to scale the entropy value into a value that ranges from 0 to 1 (Thomas 1981). Yeh and Li (2001) argue that because entropy can be used to measure the distribution of a geographical phenomenon, the measurement of the difference of entropy between time $t_1$ and $t_2$ can be used to indicate the magnitude of change of urban sprawl.

Entropy is more robust, spatial, statistic than the others (Yeh and Li 2001; Bhatta et al. 2010a). Many studies have shown that entropy is better than the spatial dispersal statistics, such as the Gini and Moran coefficients (Tsai 2005), that are often dependent on the size, shape, and number of regions used in calculating the statistics, and the results can change substantially with different levels of areal aggregation (Smith 1975; Thomas 1981). This is a manifestation of the scale problem or modifiable areal unit problem (MAUP) (Openshaw 1984; Openshaw and Alvanardies 1999; Armhein 1995) which may exert unspecified influence on the results of spatial analysis (Openshaw 1991). The effects of the MAUP can be divided into two components: the scale effect and the zoning effect (Armhein 1995). The scale effect is the variation in numerical results that occurs due to the number of zones used in an analysis. The zoning effect is the variation in numerical results arising from different grouping systems of small areas into larger units. Figure 2.3a shows the change in means that occurs when smaller units are aggregated into larger units, and Fig. 2.3b shows the change in mean with the change of zoning system (shape and size).

Thomas (1981) indicates that relative entropy is better than traditional spatial dispersal statistics because its value is invariant with the value of $n$ (number of
Therefore, the use of relative entropy can mitigate the scale effect of MAUP. However, relative entropy is still, to some extent, sensitive to the variations in the shapes and sizes of the regions used for collecting the observed proportions. For example, if there are two scales of analysis for the dispersion of population in a country, such as regions and subregions, different entropy values will be obtained if the data are collected based on regions instead of subregions. The entropy decomposition theorem can be used to identify different components of the entropy that are related to different zone sizes in collecting the data (Batty 1976; Thomas 1981; Yeh and Li 2001). It can alleviate the problem of comparing the results between different zone sizes because the influence of scaling can exactly be measured (Yeh and Li 2001).

Several other sprawl measurement techniques have also been reviewed and documented in Bhatta et al. (2010a) with a greater detail. However, it has been found that the entropy method is robust because of its preceding advantages.

In case of understanding the urban growth, analysis or quantification of sprawl is not adequate. In case of planned cities, the future growth is generally planned and modeled in advance. In many instances, the growth is restricted within a sharp defined boundary so that the actual growth does not exceed the planned growth. Therefore, whether the observed growth meets the planned (expected) growth and to what extent they meet are also necessary to be evaluated. Almeida et al. (2005) have used Chi-square statistics (degree of freedom) to evaluate the observed growth in comparison to expected growth. Their methodology, however, considers

Fig. 2.3 The MAUP effects (Bhatta 2010). a Scale effect. b Zoning effect

zones. Therefore, the use of relative entropy can mitigate the scale effect of MAUP. However, relative entropy is still, to some extent, sensitive to the variations in the shapes and sizes of the regions used for collecting the observed proportions. For example, if there are two scales of analysis for the dispersion of population in a country, such as regions and subregions, different entropy values will be obtained if the data are collected based on regions instead of subregions. The entropy decomposition theorem can be used to identify different components of the entropy that are related to different zone sizes in collecting the data (Batty 1976; Thomas 1981; Yeh and Li 2001). It can alleviate the problem of comparing the results between different zone sizes because the influence of scaling can exactly be measured (Yeh and Li 2001).

Several other sprawl measurement techniques have also been reviewed and documented in Bhatta et al. (2010a) with a greater detail. However, it has been found that the entropy method is robust because of its preceding advantages.

In case of understanding the urban growth, analysis or quantification of sprawl is not adequate. In case of planned cities, the future growth is generally planned and modeled in advance. In many instances, the growth is restricted within a sharp defined boundary so that the actual growth does not exceed the planned growth. Therefore, whether the observed growth meets the planned (expected) growth and to what extent they meet are also necessary to be evaluated. Almeida et al. (2005) have used Chi-square statistics (degree of freedom) to evaluate the observed growth in comparison to expected growth. Their methodology, however, considers
pattern and process in a single metric. Therefore, the urban growth pattern and process cannot be distinguished.

Since the entropy (degree of sprawl) and Chi-square (degree of freedom) are different measures and one may not relate other, it necessitates determining the ‘degree of goodness’ of the urban growth. The degree of goodness actually refers to the degree at which observed growth relates the planned growth and the magnitude of compactness (as opposed to sprawl). However, the review of literature did not find any metric that can quantify the goodness in urban growth in a single quantifiable scale. As stated in Chap. 1, the concept of goodness in urban growth was earlier introduced by the author and his colleagues during this research (Bhatta et al. 2010b). But, the method had several limitations (explained later in Sect. 3.14.5).

2.6 Administrative Versus Natural Boundary for the Analysis

Cities have administrative boundaries associated with them in the sense that city governments have jurisdiction over certain well-defined administrative areas. But the area contained within the jurisdictional boundary of a city has little to do with the metropolitan area of the city. In some cases, this area is very small in comparison with the size of the metropolitan area. The Los Angeles metropolitan area, for example, contains 35 independent municipalities. In other cases, for example in Beijing, the jurisdictional boundary of the municipality contains an area that is much larger than the built-up area of the city. The official area of the municipality is therefore not a very precise measure, neither of the built-up area of the city nor of what we intuitively grasp to be the city (Angel et al. 2005). Furthermore, the extent of a city is a dynamic phenomenon; it changes over time. However, the jurisdictional boundary of the city cannot be changed frequently owing to administrative complexities.

However, the main problem of the natural city extent is the lack of census data. Census data are generally collected and published in respect of census blocks; these blocks are defined in consideration of administrative boundary. Therefore, how the population or other socioeconomic parameters change over time within the natural boundary cannot be analyzed. If the analysis requires consideration of census observations then it must be based on the administrative tracts.

2.7 Determining the Natural Boundary of the City

Delineating the natural boundary or extent of a city is a real difficult task, because urban to rural transact shows a gradient (Fig. 2.4). If one interprets Fig. 2.4, he or she can understand the difficulty in demarcating the line that separates ‘rural’ and ‘urban.’
There are several approaches to determine the natural extent of a city or town (Bhatta 2010). Traditionally, the physical delimitation of urban areas and agglomerations has been characterized by two clearly differentiated approaches. On the one hand, delimitation is based upon physical or morphological criteria, where the continuous built-up area or the density of contiguous ambits comprises the basic mechanisms for the delimitation. On the other hand, studies based upon functional or economic criteria, the emphasis is placed upon the existing relations and flows throughout the urbanized territory, where the relation between place of residence and place of work is fundamental (Roca et al. 2004). Among these two, the former is more practiced. Population density is also another preferred option of defining urban area. In India, an area with a town or city of at least 50,000 people with continuous growth around it, encompassing a number of smaller towns and rural settlements based round the core town, with the possibility of being urbanized within the next couple of decades is defined as a standard urban area (Kundu and Basu 1999). One can understand the difficulty of delineating the urban area considering this definition. Furthermore, remote sensing data are incapable to convey such information. Therefore, in consideration of remote sensing data, physical or morphological criteria are the preferred approach to determine the natural extent of the city.

2.8 Subdivision of Natural City Extent

One of the major limitations of the several metrics (e.g., area index and shape index) is the consideration of the entire city in the analysis. They do not consider the variations of built-up areas in the different parts of the city. Therefore, a further analysis is essential to integrate the internal variations of the built-up area.
understand the variations in built-up pattern, a popular approach is the moving window (Martinuzzi et al. 2007). It measures the spatial frequency\(^1\)—pattern surrounding individual units of observation (i.e., a pixel) within the window. However, this measurement is highly dependent on the size of the window and there is no standard rule to define the window size. Further, it is influenced by the spatial resolution of the image data and can give varying results. For example, using remote sensing imagery with a spatial resolution of 60 m makes it more likely that small clusters of homogeneous classes will be classified as isolated cells; in comparison, measure from 30 m resolution image likely appears more scattered. On the other hand, aggregation (resampling) from 30 to 60 m resolution will also cause the pattern at the patch level to become simplified with fewer edges and a higher area-to-perimeter ratio, revealing a more compact urban pattern. Therefore, the area under analysis should be subdivided into smaller pieces independent of pixel array (window), and the aggregation (of built-up areas) should be done in consideration of these pieces.

There are different approaches to subdivide the study area. In case of analysis considering the administrative boundaries it is rather easier; administrative hierarchical boundaries may become the basis, for example, municipal boundary may be subdivided into boroughs or wards. In case of natural city extents, they are generally subdivided into circular or pie sections or in both (Kumar et al. 2007; Bhatta et al. 2010b). The main problem of this approach is its blindness to the actual shape of the city, for example, if the city is very long and narrow in shape the circular subdivision includes several rural (or other) areas in the analysis as shown in Fig. 2.5. Another better approach may be to construct buffered zones (positive or negative) by using the natural boundary of the city. However, for multitemporal analysis, we shall have multiple natural boundaries for multiple dates. Which boundary should be considered to construct the buffer is unclear. Therefore, we need a different approach for subdivision which can avoid the stated problems.

---

\(^1\) Spatial frequency denotes the rate of changes in land-use/land-cover classes with the change of space.
2.9 Similar Studies in India

Although a chronological history of urban geographic research in India can be found in Thakur and Parai (1993), researches on urban growth, especially by using remote sensing data, have not been well documented. The first appearance of urban growth and sprawl analysis using remote sensing data in a publicly published form can be found in 1989 in the Journal of the Indian Society of Remote Sensing. Two papers, Sokhi et al. (1989) and Uttarwar and Sokhi (1989), were published; both were focused on the city of Delhi. Research of Uttarwar and Sokhi (1989) was aimed to study the urban fringe from aerial remote sensing data. This study was conducted for land-use change detection and inventory of land uses in suburban areas from Survey of India topographical maps of 1979, aerial photographs of 1984, and field check for updating in 1988. These data had been visually interpreted to generate land-use maps via manual mapping. Then these maps had been compared to quantify the urban growth; however, this chapter was not focused on sprawl. The other paper, Sokhi et al. (1989), was focused on mapping and monitoring of sprawl. Satellite imageries of 1975, 1981, 1985, and 1987 had been visually interpreted to prepare land-use/land-cover maps via manual mapping. Then the changes and built-up growth had been quantified from these maps. This chapter was on simple urban growth quantification. The distinction between urban growth and sprawl is not clear in this chapter.

Taragi and Pundir (1997) analyzed the urban growth and sprawl of Lucknow city for the period of 1972–1992. They derived nine major land-use/land-cover classes from four temporal remote sensing images via visual interpretation and manual mapping. They prepared a built-up change map and computed built-up growth rate. Based on these built-up information and land-use/land-cover maps, they tried to identify the sprawl intuitively. However, their ‘sprawl-map’ is actually a built-up change map and this research did not quantify the sprawl. Their characterization of sprawl was mainly limited to rapid growth.

A paper of Sudhira et al. (2003) was focused on the analysis of urban growth pattern in the form of either radial or linear sprawl along the Bangalore-Mysore highway over a period of 1972–1998. They used classified remote sensing image for 1998 and Survey of India topographical maps for the year 1972. Then the Shannon’s entropy was used to determine the sprawl in different directions. Although this chapter successfully used the entropy method, however, consideration of only two temporal data set in a wide temporal gap is a questionable approach.

Sudhira et al. (2004) conducted a research to determine the dynamics of sprawl and to model the future sprawl using remote sensing and other dataset. They considered a study area of 434.2 km² in Mangalore–Udupi region in Karnataka state and the time span was considered 1972–1999. To determine the built-up information, they classified satellite remote sensing data for the year 1999. Built-up information of 1972 had been extracted from Survey of India topographical maps. Then they had analyzed the sprawl via built-up growth versus population growth, Shannon’s entropy, pachiness, built-up density, population density, annual
population growth rate, distance from Mangalore, and distance from Udupi. Metrics analysis had been performed using a $3 \times 3$ kernel window. Finally, they had predicted the built-up area for the years 2020 and 2050 by using regression analysis. However, although the title claimed for the modeling of sprawl, ultimately it resulted in a sum increase of built-up area for future.

Sudhira and Ramachandra (2007) conducted a similar analysis as the aforementioned one. They analyzed the urban growth and sprawl of Bangalore city using two temporal satellite remote sensing images for the time span of 1992–2000. Images had been classified to extract built-up, water, vegetation, and open land. Then they had adopted similar analytical approach as Sudhira et al. (2004), for characterizing the sprawl. This chapter also has similar limitations as in the preceding one (i.e., Sudhira et al. 2004).

Kumar et al. (2007) considered Indore city for a similar analysis. They analyzed the city using three temporal satellite remote sensing data for a period of 1990–2000. The growth patterns of built-up had been studied initially by dividing the area into four zones. The observations had been made with respect to each zone. Then, the study area had been divided into concentric circles of 1 km buffers; the growth patterns had been studied based on urban built-up density with respect to each circular buffer in all four zones. These observations had been integrated with road network to check the influence of infrastructure on haphazard urban growth. They also used entropy and population versus built-up growth for the identification of sprawl. This study had shown a novelty in dividing the zones and integrating the zones with road network.

A paper by Jat et al. (2008) is another example which analyzed the Ajmer city (Rajasthan) for the time span of 1977–2005 using eight temporal satellite remote sensing data. They classified the images using supervised maximum likelihood classifier to extract ten information classes. Then landscape metrics (such as, entropy, patchiness or landscape diversity, built-up density, population vs. built-up index) had been calculated. Landscape metrics had been extracted using a $5 \times 5$ moving window to characterize the sprawl. Finally, they modeled the urban growth for 2011 and 2051. This research was comprehensive in all respects.

A research by Jha et al. (2008) was aimed to study the spatial extent of urbanization in Haridwar, and patterns of periodic changes in urban development (systematic/random) in order to develop future plans. Remote sensing images had been used to map the spatial extent of urbanization for the years 1989, 1998, 2000, and 2002. Entropy was used to study the patterns of urban development (systematic or random) for the time span 1989–2002. The distributed entropy and relative mean entropy values had been evaluated considering two location factors: (i) urban development at peripheries of 1,000 m each from the center of the city, (ii) urban development at peripheries of 1,000 m each from the highway along the upper *Ganga* canal. This research, although not solely focused on sprawl, has shown novelty in using the entropy.

Another recent work, conducted by Farooq and Ahmad (2008), was aimed to analyze the urban sprawl around Aligarh city by using remote sensing data for a time span of 1971–2006. They used Survey of India topographical maps for 1971
and remote sensing images for the other three years (1989, 1999, and 2006). Then built-up areas had been mapped by means of visual inspection. Built-up growth and population data had been used to identify the sprawled area. In this chapter, rapid built-up growth and visual inspection of pattern remained the scales to identify the sprawl. The novelty of this chapter was in discussing the consequences of urban growth for the study area.

2.10 Similar Studies on Kolkata

Although the city of Kolkata is a very old and important urban area in India, analysis of urban growth and sprawl for this city was not found in the existing literature when this research was initiated. Kolkata urban agglomeration is the largest urban agglomeration in eastern India, and second largest in India with a population of 13.2 million as per 2001 census. This area contains more than 50% of the total population of West Bengal. Therefore, while initiating this research, the city of Kolkata has been considered as a study area to understand the pattern and process of urban growth and sprawl.

A research by Richardson et al. (2000) found that cities in developing countries are becoming significantly more compact in spite of decelerating population growth and the beginnings of decentralization. Acioly and Davidson (1996) also reported that there was evidence that a general process of change was leading to more compact cities in developing countries. Whether the city of Kolkata supports their findings or contradicts them would be interesting to know through the current research.

2.11 Scope of the Research

The scope of this research, in view of the observations documented in this chapter, can be listed as:

(1) Analysis of urban growth and sprawl (from remote sensing data) for the city of Kolkata for the time span 1980–2010.
(2) Application of several existing models for the analysis of urban growth and sprawl for the city of Kolkata to evaluate their results and suitability.
(3) Modification of existing models (if require) to analyze the urban growth and sprawl for the city of Kolkata.
(4) Proposal of some simple (in contrast of complex) new approaches that require minimal set of input data to analyze the urban growth and sprawl. The analysis of urban growth is now essential for many layers of the administration, environmentalists, planners, stakeholders, and even individuals. It is not necessary that they will be an expert in remote sensing and GIS. Therefore,
simple approaches are preferred for quick and wide adaptation. Developing countries generally lack consistent and reliable data, especially spatial data (e.g., multi-temporal road network, land-use/land-cover maps, and parcel data). Census data are also highly generalized for the past. Therefore, analysis should not be based on a wide variety of data that are not possible to obtain easily, especially for public consumption.

(5) Identification of simple models that can characterize the sprawl in black-and-white. Relative measures, most often, fail to draw conclusion on sprawl and cannot be used in other areas reliably. These measures may serve the scientific purposes well, but, never can become a technology, because to interpret the results one has to be a scientist.

(6) Introduction of new zoning concept that is not dependent on the administrative block or buffer or circular/pie sectional divisions; rather is based on natural multi-temporal boundary of the city.

References


Bae CHC, Richardson HW (1994) Automobiles, the environment and metropolitan spatial structure. Lincoln Institute of Land Policy, Cambridge


Benfield FK, Raimi M, Chen D (1999) Once there were greenfields: how urban sprawl is undermining America’s environment. Economy and Social Fabric, The Natural Resources Defense Council, Washington, DC
EEA (European Environment Agency) (2002) Towards an urban atlas: assessment of spatial data on 25 European cities and urban areas. EEA, Copenhagen
Galster G (1991) Black suburbanization: has it changed the relative location of the races? Urban Aff Q 26:621–628


Harvey RO, Clark WAV (1965) The nature and economics of urban sprawl. Land Econ 41(1):1–9


Urban Growth Analysis and Remote Sensing
A Case Study of Kolkata, India 1980–2010
Bhatta, B.
2012, XIII, 103 p. 23 illus., 13 illus. in color., Softcover
ISBN: 978-94-007-4697-8