Research Paper

MACHINE LEARNING FOR BUILDING EXTRACTION AND INTEGRATION OF PARTICLE SWARM OPTIMIZATION WITH SLEUTH FOR URBAN GROWTH PATTERN VISUALISATION FOR LIVABLE CITIES

Chandan Mysore Chandrashekar, Research Scholar, RCG School of Infrastructure Design and Management, IIT Kharagpur; India

Aadithyaa J S, Undergraduate student, Department of Architecture and Regional Planning, IIT Kharagpur; India

Prakash P S, Research Scholar, RCG School of Infrastructure Design and Management, IIT Kharagpur; India

Bharath Haridas Aithal, Assistant Professor, RCG School of Infrastructure Design and Management, IIT Kharagpur; India

Abstract

Rapidly increasing population and migration from rural areas to nearby urban agglomerations develop tremendous pressure on system of the existing cities without compromising socio-economic and cultural linkages. Policy interventions, both at global and local scale, have created newer avenues for the researchers to explore real-time solutions for problems world-wide. For instance, the outcome of 2015 United Nations agenda for the achievement of the Sustainable Development Goals (SDGs) by the year 2030 primarily focuses on urbanization issues and probabilistic modelling of future scenarios to obtain a robust alternative for resource utilization and further for maximizing sustainability through land use pattern analysis. This is the clear indication toward the very important role of "ever dormant" urban planning, especially in the case of a rapidly developing country such as India. Remote sensing and geo informatics along with Machine learning can provide extremely relevant information about the pattern change in cities and as input to visualize the future growth pockets. In this context, potential of cellular automata (CA) in urban modelling has been explored by various researchers across the globe. In the recent past, models have been drawing majority of the attention along with geographic CA processes about urban growth and urban sprawl studies. Most recent approaches include optimization of transition rules based on machine learning techniques and evolutionary algorithms that follow nature-inspired mechanism such as Genetic Algorithm, Ant colony optimization, Particle Swarm Optimization (PSO), simulated annealing, Grey Wolf optimizer etc. Irrespective of any modelling technique, model calibration remains one of the challenging and most crucial steps towards obtaining realistic results. This research communication tries to demonstrate a novel idea of integrating PSO with SLEUTH post calibration of the spatial-temporal footprint of urban growth from the year 1990 to 2017 for Kolkata, a historical megacity of Eastern India. Results were evaluated and validated using statistical fit measures reveals PSO-SLEUTH performed substantially better compared to traditional Brute Force calibration method (BFM). Another significant development was in terms of

computation time of optimized values from days (BFM) to hours (PSO). The study identifies Kolkata region to be sensitive to spread and road gravity coefficients during calibration procedure. Results indicate growth along the transport corridors with multiple agents fuelling the growth. Further, with the aid of high spatial resolution data, buildings were extracted to understand the growth parameters incorporating neural networks. Using the results, renewable energy aspects were explored to harness and provide a suitable local solution for energy issues in energy gobbling cities. Pattern of landscape change, development of better process of modeling and extraction of building from machine learning techniques for planning smart cities with self-sustaining energy is presented in this research work.

Keywords

*Cellular Automata, Particle swarm optimization, SLEUTH, Urban modeling, Pattern recognition*

1. Introduction

Asia alone is projected to house 54% of total world urban population (6.3 billion) by the year 2050 against 50% in the year 2010 (UNDESA, 2015) indicates the need for liveable cities that are more likely urban areas. Urban areas, over years have been reorganised and are reorganising due to increased population mostly migration. This ever increasing and irreversible urban growth has given impetus to understand urban morphology in depth and its implications on surrounding nature. Rapid urban growth in Indian scenario can be attributed to rural to urban migration, re-classification of cities, availability of improved health facilities, jobs, improved living, availability of essential amenties etc. in urban areas (Bharath et al., 2018). With the commencement and successful implementation of urban development policies and schemes such as Smart Cities, AMRUT, JNNURM, etc. have promoted migration from rural area to urban core (Shaw & Das, 2017). In 2010, India accounted for 30% of urban areas with 367.5 million people residing within. A clear jump in can be observed in urban population post 2000 (figure 1).

An important decision was taken during September 2015, UNSDS with title "Transforming our world: the 2030 Agenda for Sustainable Development" in thinking of global community in adopting the Agenda for Sustainable Development (SD) through vision 2030. Goal number 11 of 17 Sustainable Development Goals (SDG) elaborates 'Sustainable Cities and Communities' and guidelines to meet these criteria. In addition, article 76 of 2030 agenda pronounces use of geospatial data for tracking and monitoring the progress of various projects under the umbrella of SDG's. In this regard, the creation of big data pertaining to cities, specifically the buildings that are the significant portion will be of paramount importance. The recent development of machine-learning algorithms coupled with geospatial data are among the front-runner technological tools, expected to play a crucial role. Geospatial data sets provide the means to study large area of the earth surface and helps to generate necessary data from complex urban surface features. Geospatial data majorly comprises of satellite data and GIS data. Satellite remote sensing (RS) has been credited for its quickest and lesser cost available method of mapping large areas. The availability of high resolution and multi-spectral satellite imagery provides the best accuracy needed for assessing urban growth pattern (Bharath et al., 2014). Satellite technology and data analysis are adopted by both regional and national governments (Chen et al., 2000; Ji et al., 2001). It helps to map urban areas at finer scales,

both temporally and spatially with consistent images of Earth surface. Remote sensing, its analysis and integration with geographical information systems (GIS) forms a robust technology to monitor temporal urban growth. GIS platform has entered majority of the service sectors by providing facilities to capture, manage, store, retrieve, analyze and display geospatial data on a real-time basis. Further strength to GIS comes from its distinguished built-in database, decision support system and application-specific plugins, makes more reliable tool for urban studies (Tran et al., 2015; Boori et al., 2015).
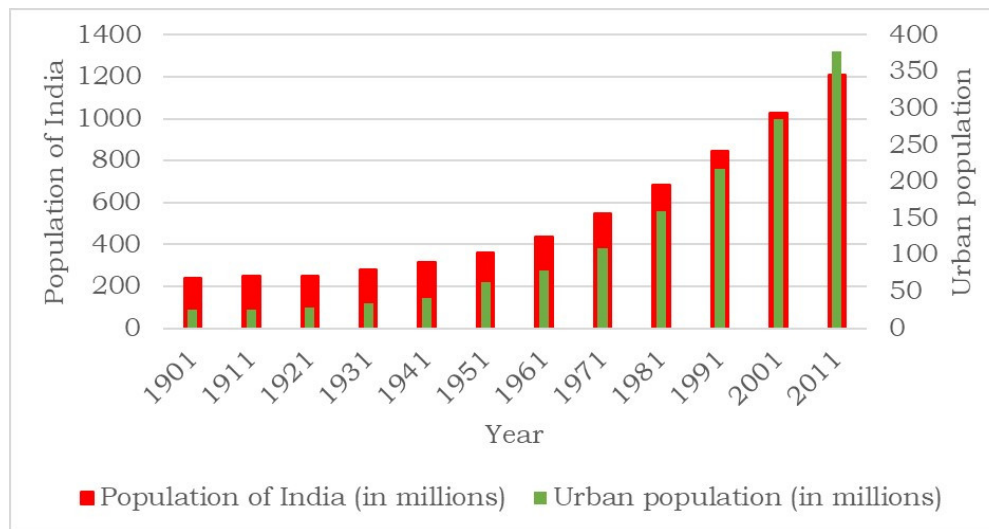
Figure 1 Trends of urbanization in India 1901-2011

Recently, machine learning techniques for building extraction information from remote sensing images has become easier with the advent of integrating various approaches on the same platform. Speed and accuracies of digital image processing have improved significantly with the use of several machine-learning algorithms. Random forest (RF) and Support Vector Machine (SVM) methods of learning algorithms proved efficient in information extraction from satellite imagery (Thanh & Kappas, 2018). Having previous and present years' temporal data, it is necessary to have a vision of development for the near future in terms of urban modeling. It is the process of identifying a theory, could be translated into a mathematical model as well as developing specific computer-aided programs to feed the model with data to calibrate, validate, verify and predict future urban trends (Verburg et al., 2002; Batty, 2009).Theoretical assumptions, method followed, spatial, temporal aspects and geographical extents might vary from each model types, but the final outcome of these models are to understand the complex interrelationships between natural ecosystem and urban environment by observing irreversible heterogeneous patterns of change (Li & Liu, 2006; Crooks, 2010). Wu and Silva, 2010 reviewed urban land dynamics and related models, depicted in figure 2. They highlight the significance of artificial intelligence (AI) and its deep theoretical understanding of urbanization process and relative pressure on land use pattern change. Figure 2 shows broad categories of AI systems used in urban models. In this research paper we discuss briefly about cellular automata (CA), SLEUTH model, optimization techniques such as genetic algorithm (GA) and particle swarm optimization (PSO), its application to urban growth models. The structure of the paper is organized as follows: sub sections 1.1-1.3 provides a brief overview, literature review of Cellular automata and SLEUTH urban growth model, its improvement over last two decades, integration of stochastic optimization

techniques to SLEUTH and building extraction for urban growth application. Section 2 describes the study region selected to implement modeling and extraction of building information along with data sources. In section 3 we describe the novel integrated method adopted to solve time-consuming conventional modeling problem countered with PSO-SLEUTH. Results are discussed with statistics and visualization in section 4. Finally, conclusions are given in section 5.
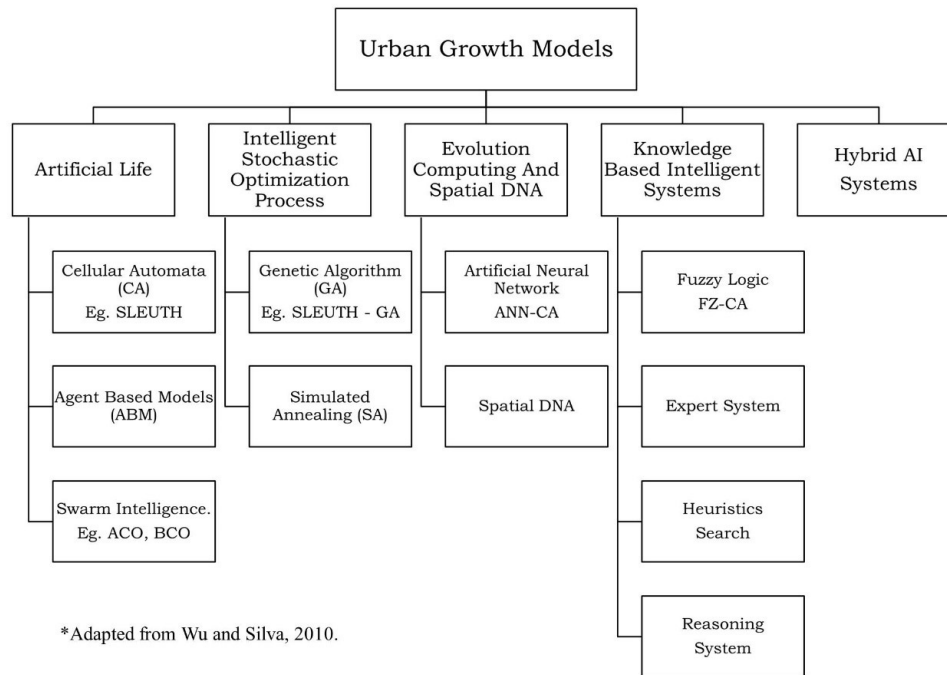
Figure 2 Types of urban growth models

1.1.   Cellular automata and urban growth models

Wolfram, 2002, defined CA as discrete two dimensional dynamic systems in which local interactions among components generate global changes in space and time. Torrens, 2000 explains CA is composed of the below four components:

1. Cell space represented by an array of cells

2. Several finite states that qualifies the state of each cell

3. Neighborhood

4. Transition functions, define the next state of the cell in the next period, based on the given state of the cell itself and its neighborhood cells

Standard cellular automata apply a "bottom-up" approach. The approach argues that local rules can create complex patterns by running the models in iterations (Batty & Xie, 1994). Itami, 1994 states that CA are seen not only as a framework for dynamic spatial modeling but as a paradigm for thinking about complex spatial-temporal phenomena and an experimental laboratory for testing ideas. Researchers have successfully used CA in urban modeling over the decades and now have tried to address various related research questions with the help of CA-based models (Alves & Albrecht, 2010; Yang et al., 2012; Torrens & Benenson, 2005).

SLEUTH is a CA-based urban growth model, started as open source, has witnessed numerous applications over different parts of the world. The acronym of SLEUTH stands for slope, land use, excluded, urban, transportation and hillshade, the layers used as input for the model, consists of C-language, using UNIX or UNIX based operating systems (Nimish et al., 2018; Chandan et al., 2019). It has an ability to predict urban/non-urban land use dynamics based on two sub models: urban growth model (UGM) and Deltron land model (DLM) (Dietzel & Clarke, 2004). Model is dependent on five factors controlling the behavior of growth: diffusion, breed, spread, slope resistance, and road gravity. It also considers four growth rules: spontaneous growth, new spreading center growth, edge growth, and road-influenced growth. Best set of coefficients are derived for the five factors using these four rules. Traditional brute force method (BFM) takes days together to arrive at three phases of calibration i.e., coarse, fine, full and shortlisted values for prediction. Most researchers have made successful attempt to reduce SLEUTH computation time and therefore to increase its efficiency. The original version includes SLEUTH UGM and DLM. To reduce computation time, Dietzel and Clarke, 2007 worked on the development of optimization of SLEUTH metric called optimal SLEUTH metric (OSM). OSM narrows down the parameter range and returns a single goodness of fit metric. OSM is the product of various individual metrics like compare, population, edges, clusters, slope, X-mean and Y-mean providing reliable calibration results. The source code for OSM is also available in Project Gigalopolis website and one can readily implement OSM technique. Sakieh & Salmanmahiny, 2016 explored OSM and implemented to Gorgan area of Iran. Authors have mentioned the term "cancer-treating" to understand and predict the urban spread in three different scenarios: afforest protection, rangeland protection and historical growth (without any restrictions). They have also adopted a comparative assessment to measure association between landscape metrics and land suitability values using Spearman correlation to address cancer-treating urban growth pattern. Other improvements of SLEUTH include pSLEUTH, SLEUTH-3r and SLEUTH-Genetic Algorithm (GA) (Guan & Clarke, 2010; Jantz et al., 2010; Clarke-Lauer & Clarke, 2011). GA returns a single best fit statistic for the entire dynamic range consisting of several Monte Carlo iterations. Research in the area of improving run time and various variable optimisation in SLEUTH model is yet unexplored by scientists and this paper presents this novel technique of integrating PSO rules with existing sleuth model to improve its run time by optimising the initial set of variables.

1.2.   Integration of genetic algorithm and particle Swarm optimization techniques with SLEUTH

GA includes heuristic search algorithms based on natural selection and natural genetics using the key concept of simulating, imitating or mimicking evolution process to obtain optimized results (Holland, 1975; Goldberg, 1989). It adopts Darwinian "Survival of the fittest" theory, only the fittest shall survive, reproduce and successive generations shall become superior compared to older generations (Mayer et al., 1999). Developed by Holland in early 1970s, the central theme of GA is robust optimization technique which considers independent variables of the model and converts them into genetic pairs (binary).

SLEUTH-GA is adopted to improve calibration process and to obtain optimized values in less computation time. Figure 3 shows SLEUTH-GA model calibration approach. Coarse, fine and final calibration procedure employed in BFM is replaced by GA calibration, apart from which other procedure remains same. Strength of GA comes from its ability to explore search space with improved results after each iteration. GA applied to SLEUTH follows a four-step process:

Population initialization, Selection strategies, Crossover breeding and finally Mutation. In SLEUTH procedure, the minimum and maximum value for a gene is fixed as 1 and 100 respectively, defining the range of each coefficient. Overall GA analysis involves calibration of metric values, number of iteration needed for calibration and assessing forecasting behaviour (Jafarnezhad et al., 2015).
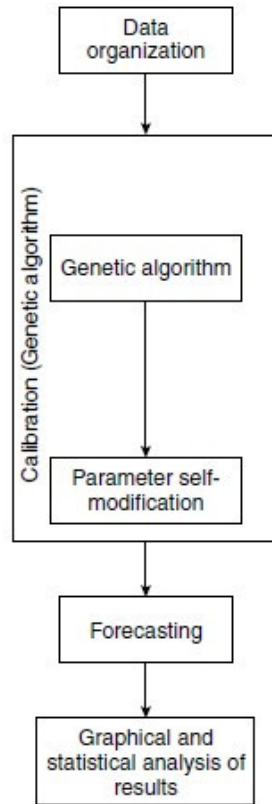
Figure 3 Genetic algorithm and SLEUTH model calibration approach (a Proposed approach based on Goldstein, 2004)

PSO is a population-based stochastic optimization technique inspired by social behaviour of birds flocking and fish schooling (Eberhart & Kennedy, 1995). Each solution to a problem is considered as a bird called particle. All particles have fitness values, and these values are calculated from objective function. Each particle has personal best and global best performance; velocity of the group is adjusted based on best performer in the group. While the advantage of PSO is that complex problem can be solved cost effectively with improved performance whereas unlike several Swarm intelligence-based optimization techniques, PSO also has limitation of particles escaping local optimum and decreased chances of finding global optimum (Shi et al., 2011). Feng et al., 2011 conducted a study to explore advantages of integrated CA and PSO approach. Objective of the study was to stochastically optimize transition rules and therefore reduce computation time, prediction uncertainties as well as improving location accuracy, a case of Fengxian District, Shanghai, China. Authors claim similarity between PSO and CA drives to search global optimum parameters of CA rules. Further, this concept can be applied to SLEUTH-PSO spread optimization to reduce computation time and predict accurate transitions. PSO optimizes allocation of pixels by

considering 3 x 3 kernel and returns a set of eight values that represent Moore neighborhood as illustrated later in figure 10.

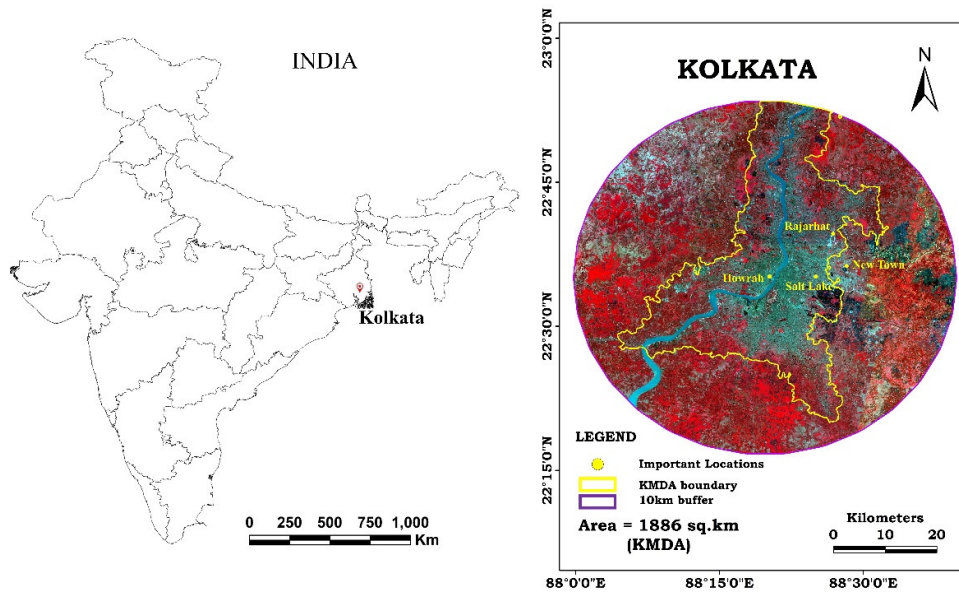## 1.3.   Urban building structure extraction

Machine learning techniques such as Random Forest (RF) and Support Vector Machine (SVM) are employed by scientists to extract buildings in urban areas. Random forest classifier is made up of an ensemble of multiple decision trees, here majority voting from the decision tree classifiers with 'yes' or 'no' outputs are used to classify the data (Breiman, 2001). The outlier in training is well handled by RF is advantageous in classifying spatial data efficiently (Horning, 2010). SVM works on the principle of identifying right hyperplanes that maximizes the distance to the closest data points on both sides of the classes (Vapnik, 1995; Dixon, 2008; Das et al., 2018). SVM is a supervised learning technique and it can be applied to a multiclass classification. In SVM classification, four kernel functions could be used based on the transformation needed, namely sigmoid, radial basis function, linear and polynomial. Previous studies showed that radial basis function performs best for binary classification from satellite images (Prakash et al., 2018).

There currently exists a gap in planning and visualization of urban structures at metropolitan scale and beyond, that modeling can fulfill with agenda of developing sustainable smart cities. It is also necessary to have insights through dynamic integrated urban growth models. Based on above discussion and detailed literature survey, we have identified the research gaps, aiding in forming major objectives of this research communication as: a) Extracting building from high-resolution satellite imagery, b) Analyzing spatio-temporal changes in urban land use for Kolkata Metropolitan Area, c) Modeling future trends of urban growth using an integrated CA-based SLEUTH-PSO approach.

2. Study area and data

Located at the eastern part of India, Kolkata has been considered for study is also famously known as "The city of Joy" or "Calcutta" during the British rule in India. The city is built on the banks of River Hooghly. The geographical extents of Kolkata metropolitan area (KMA) are (depicted in figure 4) 22°19'59" - 23°0'20" N and 88°4'23" - 88°32'37" E. Kolkata lies about an average of 3-9m above the MSL and situated very close to Ganges delta and Bay of Bengal. The city served as former capital of British India from the year 1772-1911 and presently it is the capital of West Bengal state. The region is also famous for numerous number of wetland ecosystem, natural tanks and lakes. Taking advantage of all four transport modes i.e., road, railway, airway and waterway, Kolkata has emerged as one of the dominant urban centers of Eastern India. Metropolitan urban agglomeration has an area about 1886 km^2 (KMA) with population of 14,035,959 (Census of India, 2011), makes it to the 3rd rank in metro cities of India after Mumbai and Delhi. Currently the physical infrastructure, services, planning and development of the region is administered by Kolkata Metropolitan Development Authority (KMDA).

![ISOCARP logo] ISOCARP — Knowledge for Better Cities

55th ISOCARP World Planning Congress Jakarta-Bogor, Indonesia
International Society of City and Regional Planners

Figure 4 Location map of Kolkata with false color composite as background derived from IRS LISS4 satellite data (Bands: 2,3 and 4)

Stable and floating population can be attributed to a majority of people commuting as tourists, state and central governmental office affairs, academia, healthcare, economy-oriented conferences and so on. Apart from these, region is a commercial hub of East and North-East India with Rs 1.84 trillion GDP in 2014. Figure 4 illustrates the KMDA with 10 km buffer to understand the influence of expanding urban trends, its influence on environment, and future scenario on the periphery. Data required for the study are listed in table 1.

Landsat series (5 and 8), ASTER, Indian Remote Sensing LISS 4 and Cartosat satellite data were obtained from United States Geological Survey (USGS) and National Remote Sensing Center (NRSC). During field visit (figure 5), GPS was used to collect ground truth data of various points and polygons, geo-tagged images were also obtained, aiding in validating land use maps. Toposheets from Survey of India online repository, of scale 1:25000 and 1:50000 were also considered to delineate base layers of administration and other boundaries. City development plan maps obtained from governmental agencies helped in understanding and digitizing restricted areas of future urban expansion for instance: lakes, military area, reserve forest, parks, wetland regions etc. Google earth platform was used as ancillary data in places which were not accessible for GPS data collection. Road network was digitized directly from Google earth and open street maps (OSM).

![ISOCARP logo] **ISOCARP** Knowledge for Better Cities

55th **ISOCARP World Planning Congress Jakarta-Bogor, Indonesia**
International Society of City and Regional Planners

3. Method

An integrated method adopted for building extraction and urban growth visualization is depicted in figure 6.

Table 1 Data and layers used for the analysis

Data/Layers		Purpose/Source		
Google earth		Geo-correction, classification and validation. Collection of point, line and polygon data.		
Boundary maps and raster layers		To create excluded maps or regions restricted from future development data sets based on City Development Plans.		
Survey of India Topographic maps, online portals (Google and Bhuvan data).		Base layers of the administrative boundary.		
Field data – Using GPS		Geo-correction, training data and validation data, Agents Extraction		
Slope and Hillshade		Processed from ASTER DEM (raster)		
Transportation (Roads)		Street data (OSM, Bhuvan and Google maps) updated with classified images (originally vector, rasterized).		
Satellite data		Land use analysis, slope, hillshade maps, building and urban data extraction		
Satellite/Sensor	Resolution (m)	Date	Path/Row	Remarks
Landsat 5 TM	30	06-Mar-1991 28-Mar-1999 07-Mar-2009	138/44	--
Landsat 8 OLI_TIRS	30	13-Mar-2017	138/44	--
ASTER GDEM	30	17-Oct-2011	N22E088	Reprojected to UTM
IRS LISS 4	5.8	17-Mar-2018	108/56	Reprojected to UTM
IRS Cartosat 2	1	02-Jan-2018	N. A	--

3.1. Pre-processing

To avoid inconsistency between different datasets, pre-processing of data along with geometric correction, resampling wherever necessary (to maintain similar resolution) and

image enhancement were conducted. Satellite images were geo-corrected using ground control points (GCP) collected during field visit and data pertaining to study regions were cropped according to buffer boundary. Contrast stretching was performed wherever necessary to maintain the dynamic range. Satellite data were registered to world geodetic system (WGS) 84, universal transverse Mercator (UTM) Zone number 45.

Figure 5 Data collection during field visit to Kolkata metropolitan area. A) Waypoints collected using GPS B) Wetland area and surroundings, Kamduni C) Hindustan motors industry (abandoned), Konnagar, Uttarapara and surrounding environs as seen from IRS LISS 4 FCC image (D) Panoramic view of Hindustan motors lake. (Date taken: 25 Aug, 2017, Source: Author)

3.2.   Application of machine learning to extract buildings

The resourcesat LISS IV imagery of 5.8 meters multispectral is merged with the Cartosat 1.0-meter panchromatic image using Hyperspherical Color Sphere (HCS) method (Ghosh & Joshi, 2013). The resultant high-resolution color image is clipped for the study area KMDA with 10 km buffer. The training dataset is prepared for two classes; building and non-building category with the help of ground truth and observation from the spatial imagery. Machine learning classifiers (SVM and DT) are trained using the training data obtained from the multispectral image. The training set contains data from three bands that are labeled as '0' and '1' for non-building and building classes, respectively. The predicted output from the classifiers is divided into classes by judiciously identified threshold value to obtain a binary map containing building and non-buildings.

3.3.   Land use analysis

Primary step of land use map generation is image classification. It involves various steps such as creation of FCC (false color composite) image directly helps in identifying heterogeneous patches in the landscape (Bharath et al., 2019). Bands green, red and near infrared are chosen to generate FCC image. Training polygons are digitized based on the spectral reflectance of

various features in FCC, covering the study area ensuring that polygons are well distributed throughout study region. These polygons and its coordinates are entered into GPS and ground truth data is verified during field visit to identify various land use types. A supervised classification technique famously known as Gaussian maximum likelihood classifier (GMLC) is adopted for the study. The classifier proves best amongst others because it takes into account variance, covariance and probability density contours to evaluate the land use class of every pixel under consideration (Duda et al., 2000; Ramachandra et al., 2012). Four land use classes were chosen for the study based on literature, they are: urban, vegetation, water and others (Anderson, 1976; Nimish et al., 2017).
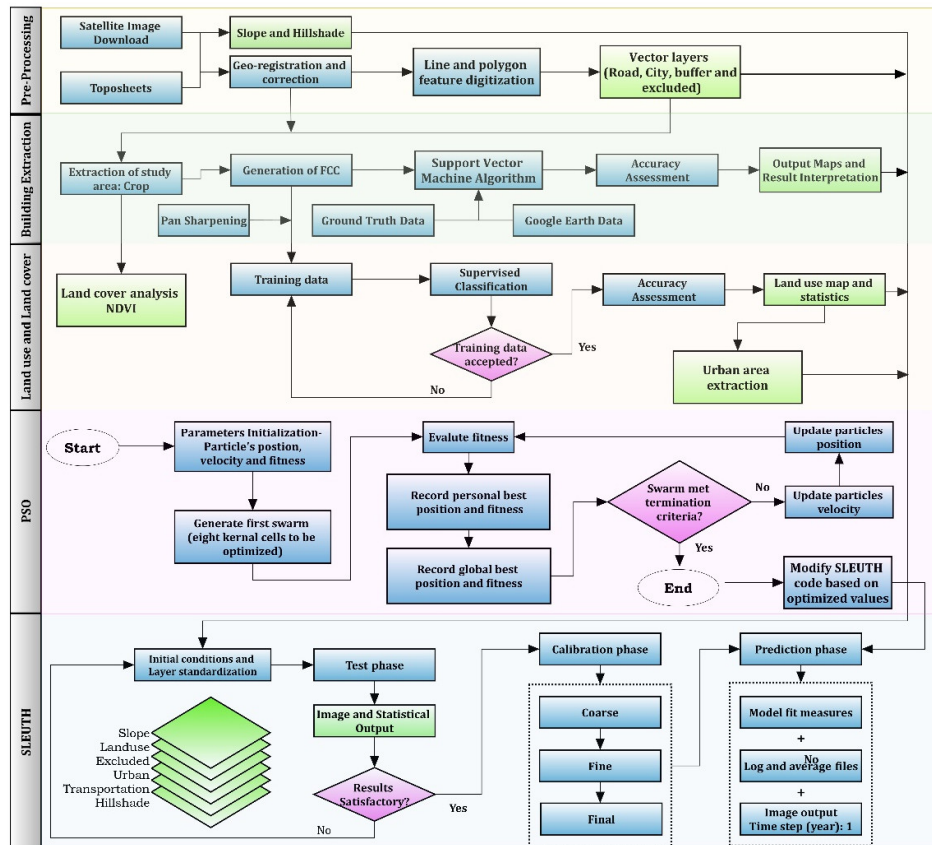

Figure 6 Integrated method adopted for the study

Accuracy assessment was performed in terms of overall accuracy and kappa statistic by comparing the classified map with the validation map. Error/confusion matrix generated helps to estimate producer's accuracy, user's accuracy, overall accuracy and kappa coefficient (Bharath et al., 2016).

3.4.   Implementation of PSO and SLEUTH

PSO is popular among various nature-inspired techniques for its simplicity, minimum mathematical computation, achieving improved optimization in less computation time and ease of application to general engineering problems (Rabbani et al., 2011; Yao et al., 2015; Hu et al., 2018). PSO resembles the behaviour of a bird flock or fish school. PSO, in terms of bird flock can be understood as a) Search space of optimization can be treated as birds flying space

b) Every bird is treated as particle and has its own solution or its own space closer to food c) Optimum solution to the problem is considered as food to all birds d) Each bird will search for food, during this process one of the important strategy adopted is search for peripheral region very close to food e) All particles has fitness values f) Estimating position of food by personal flying experience, regarded as Pbest or pid g) All particles preserve their independent best performance h) Group adjusts their velocity based on best performer in the entire group, i.e. best value obtained at any time interval in the entire population is called as global best, gbest or pgd (Stephen et al., 2013). An assumption is made that in a D-dimensional search space with n particles includes possible personal best solutions and global best solutions. These are represented as given in equation 1.

$$V_{id}^{(t+1)} = WV_{id}^{(t)} + c_1 r_1 \left[p_{id} - x_{id}^{\;(t)} \right] + c_2 r_2 \left[p_{gd} - x_{id}^{\;(t)} \right] \qquad (1)$$

Where,

   i = Particle index; d = dimension/iteration

c_1 and c_2 are cognitive/social constants between 0 and 2

   r_1 and r_2 are random numbers between 0 and 1, regenerated after every velocity update

   p_{id} = population best or present best solution

   p_{gd} = globally best solution

   $V_{id}^{(t)}$ = particle's velocity at time t

   $x_{id}^{(t)}$ = particle's position at time t;

Each particle keeps track of its coordinates in hyperspace which are associated with the best solution (fitness) it has achieved so far. The value of that fitness is also stored after every iteration. Each particle's position is updated using the equation 2.

$$X_{id}^{(t+1)} = X_{id}^{(t)} + V_{id}^{(t+1)} \qquad (2)$$

Integrating concepts of PSO with SLEUTH includes creation of a new script (Script.sh) file containing eight Moore neighborhood cell values (as illustrated in figure 10) that are to be automatically updated after each swarm iteration and storing best values. PSO provides optimum spread solution ranging between 1 and 99 for these cell values. 1 indicating least chance of urban spread in that direction and 99 indicates highest chance. These best values obtained are then adopted in spread file of SLEUTH3.0 code with p01 patch. Standardized input layers (select) used for model is illustrated in figure 7. Routine full calibration was performed using brute force method (BFM), output controlstat.log file was carefully examined to obtain model fit statistics and optimum SLEUTH metric (OSM), calculated by multiplying metrics such as compare, population, edges, cluster, slope, X-mean and Y-mean. Best OSM was chosen and corresponding coefficient values were replaced in the scenario file during prediction phase.
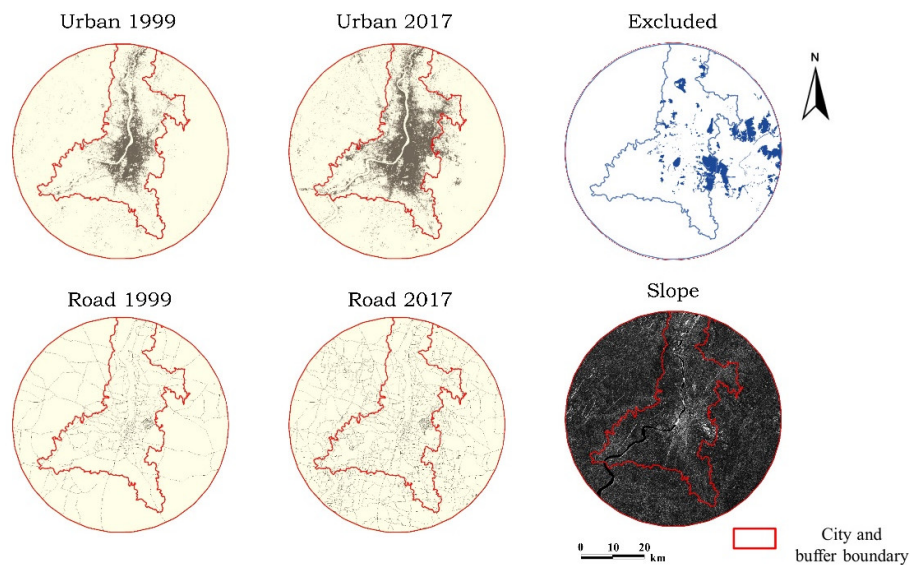
Figure 7 Standardized input layers used in PSO-SLEUTH model

4. Result and discussion

4.1. Building extraction

The data preparation for building extraction involves pan-sharpening of multispectral image using higher resolution panchromatic image. Figure 8, shows the subset of study area, Salt Lake, Bidhannagar, Kolkata, containing the dataset and extracted buildings. It is evident that from a closer look shows improvement in the quality due to image fusion in the picture (8C) and hence to obtain more detailed building information. In the binary output of reclassified image, white pixels represent the buildings and dark area corresponds to non-building class. Artifacts present in the output needs to be removed by performing morphological clean-up operations for further improvement of results. Outcome of this exercise is directly used in land use analysis and modeling as an ancillary urban dataset to describe the pace of urban growth as discussed over the next sections.

4.2. Land Use Analysis

For better understanding of the region, land use analysis was performed and results are as shown in Figure 9, corresponding statistics are tabulated in table 2. There has been significant increase in the urban areas consistently from 1990-2017. Strong economic and administrative base, high profile living, efficient service delivery mechanism, ease of access to public transport system and varied employment opportunities have been the major stimulators of urban growth in the study area. Nearby smaller villages have evolved into larger ones and many new settlements can be recognized along the transport corridors. Urban area saw an increase of 181% in just 27 years. Building extraction from high-resolution data also helps in quantifying urban growth on a microscopic scale. Water body has also seemingly increased in 2017, this can be understood on basis of visualization (temporal scale on Google earth) as well. Increase in water body, especially towards NE and SE directions are attributed to evolving aquaculture and pond ecosystem in the periphery of KMDA. However, significant waterbodies

in the form of ponds, lakes, wetlands, swamps, marshland etc. have been converted into other land use types or encroached over the period. Vegetation category shows consistent statistics, whereas other categories have decreased. Table 2 shows the kappa coefficient and overall accuracy of the classification.
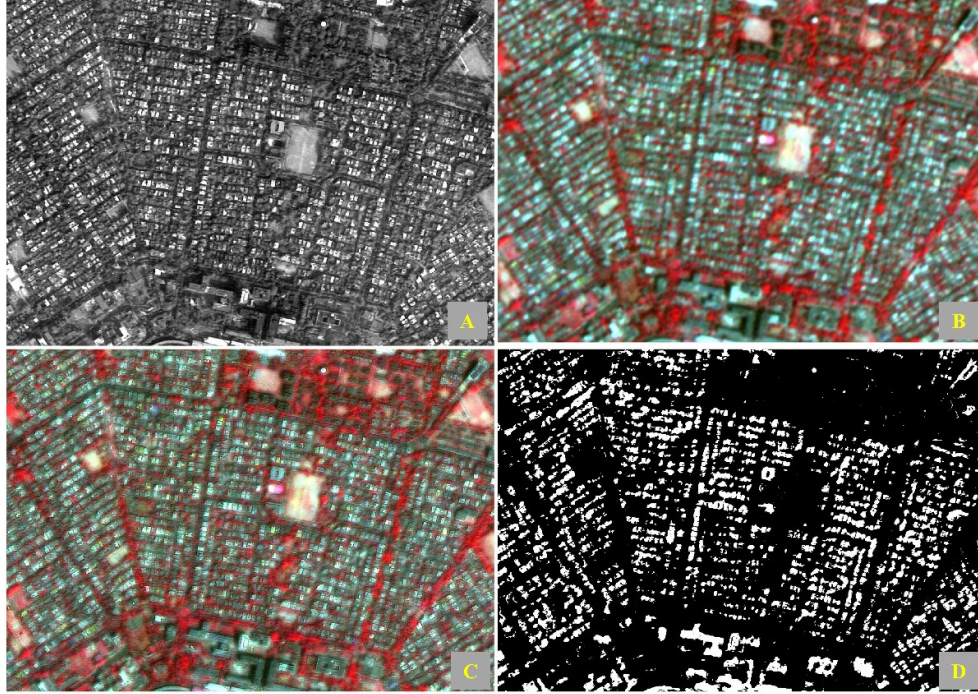

Figure 8 Data used and building extraction results for subset of study area, Part of Salt Lake, Bidhannagar, Kolkata: (A) Cartosat – 2, panchromatic, (B) LISS-IV multispectral, (C) Pan-sharpened image, (D) Extracted buildings

4.3.   PSO-SLEUTH model

SLEUTH codes were modified with respect to spread optimization obtained from PSO. Spread optimization gave eight best fit values indicating eight directions of optimum growth as shown in Moore neighborhood network figure 10C. Best PSO results were achieved at run number 83 with an OSM value of 0.824. After adopting coefficients from BFM, the spread.c, script.c and scenario file were modified accordingly and routine run to obtain control stats and therefore model fit measures.
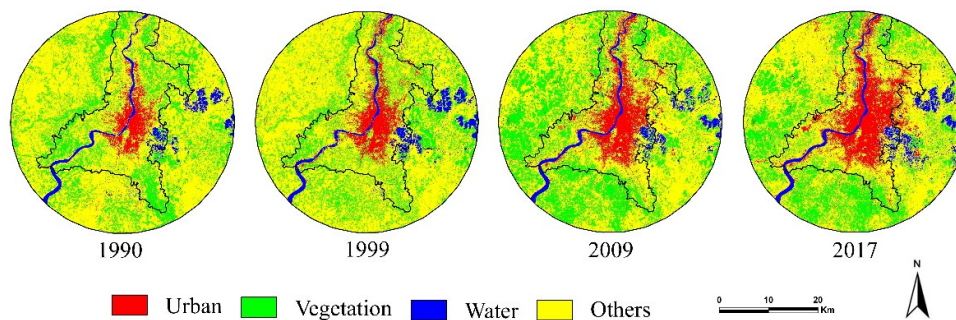

Figure 9 Land use results for KMDA region 1990-2017

Results obtained from PSO were compared with traditional brute-force method (BFM). Corresponding statistics are pictorially shown in figure 11. Substantial enhancement in terms of metrics such as Product, Compare, Edges, Clusters, Size, X-Mean and Y-Mean were observed by adopting PSO technique, while Population, Slope, % Urban and Rad showed minimum difference between these two methods. Leesalee metric was higher in BFM method with a value of 0.519 and lesser in PSO with a value of 0.480. OSM was calculated, achieved OSM was 0.0501 during BFM and 0.600 during PSO-SLEUTH. This clearly shows significant improvement in calibration of coefficients with PSO improvised spread optimized pixel values. Another important development was observed in computation time. Post full calibration process, PSO took 4hours, 12mins and 32 seconds to yield model fit statistics. These results not only indicate the model robustness by integrating PSO, but also it can deliver excellent fit statistics without compromising optimum coefficients.

Table 2 Land use and accuracy statistics

	Land use categories				Accuracy assessment	
Year	Urban	Vegetation	Water	Others	Overall accuracy (%)	Kappa statistic
1990	4.12	23.08	4.2	68.61	88	0.94
1999	6.45	21.17	4.7	67.68	93	0.96
2009	9.04	28.23	4.73	58	91	0.85
2017	11.58	22.18	6.24	60	91	0.89

Further, prediction for KMDA was performed using best-fit coefficients. After three phase calibration, optimum growth coefficients were derived. Of all the coefficients, breed (52) and spread (64) showed medium values whereas road gravity was the highest (80). This indicates Kolkata region has a significant influence of breeding a new urban site in the periphery or spreading type of urban growth from already existing core urban area, can also be referred to as sprawling at the outskirts. High values of road gravity reflect road acting as key urbanization factor. To name a few, Kolkata region has major connecting roads (AH1) towards NW direction, Bardhaman, Durgapur and Asansol; SW direction Kharagpur and further this road (AH45) connects to Orissa; NE direction, AH1 connects to Jessore (Bangladesh) and other important road network such as: Grand trunk road, NH 34, 35, 117, SH 1, 2, 3, 13 and 15. Diffusion and slope resistance showed the lower values of 8 and 16 respectively. Prediction was performed using PSO-SLEUTH to potentially visualize the scenario of urban growth and it was estimated to reach 757.91km^2 in comparison with 494.73km^2 in 2017, can be observed in figure 12.

ISOCARP Knowledge for Better Cities

55th **ISOCARP World Planning Congress Jakarta-Bogor, Indonesia**
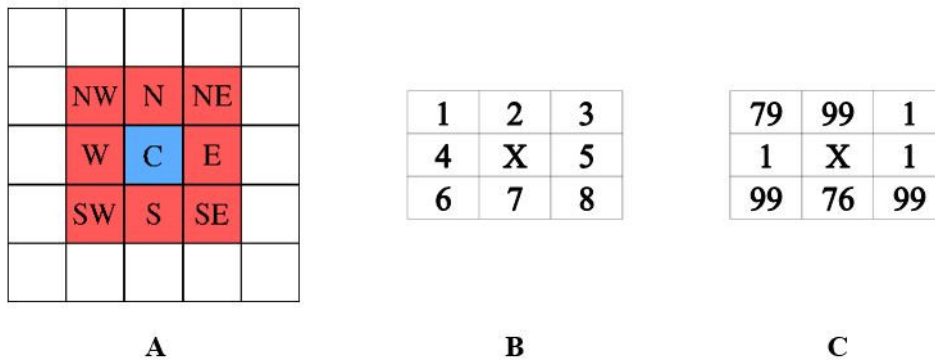International Society of City and Regional Planners

Figure 10 Optimization of pixel values based on Moore neighborhood. A) 3 x 3 kernel and directions B) Allotment pattern of pixel values in code C) Values achieved after running PSO
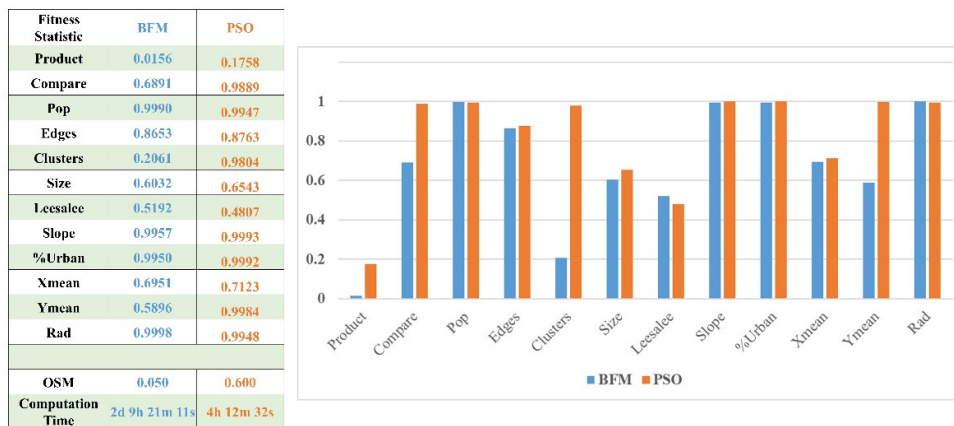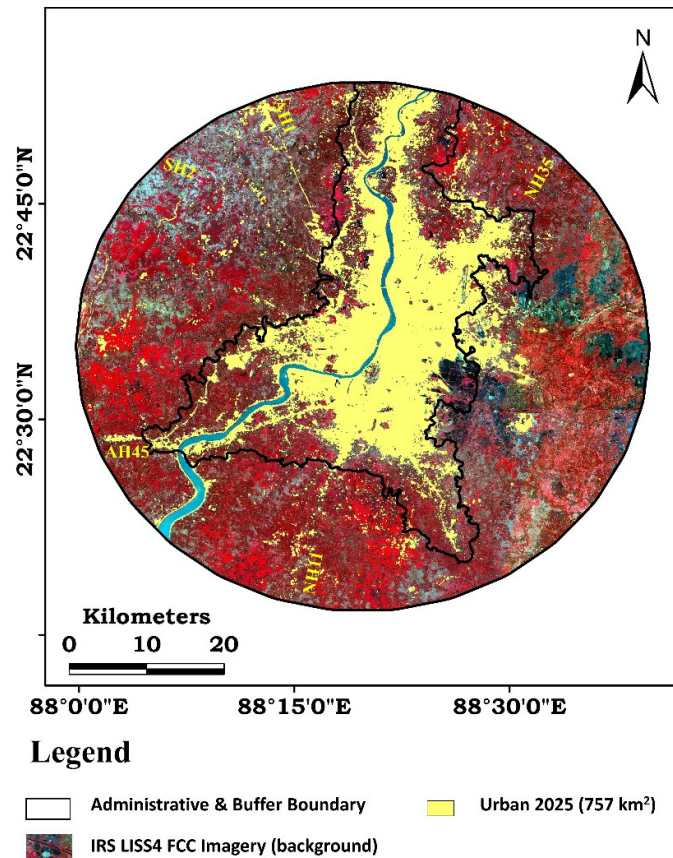

Figure 11 Summary of model fit statistics, comparison of metric values between BFM and PSO

5. Conclusion

Urban land use change is widespread, rapid and has direct effect on ecosystem and environment. It is mainly driven by human-induced activities carried out beyond limits and exploring natural resources to the fullest without any regard to concepts of sustainability. Extraction of building data from satellite imagery has always helped government agencies and planners to assess critical impacts of urban expansion on other land use categories. This research article highlights the significance of building extraction using high-resolution imagery and also discusses the potential of data merging, i.e., panchromatic image and multispectral image using pan-sharpening method to obtain binary classification of buildings, further this data can be used for modelling purpose as an ancillary dataset. This study also demonstrates the successful implementation and monitoring spatio-temporal urban growth using satellite data over a period of 27 years. Accuracies ranging from 88%-93% and kappa between 0.85-0.96 indicated excellent agreement between generated land use image and real-time dynamics.

Legend

☐ **Administrative & Buffer Boundary**   ☐ **Urban 2025 (757 km²)**

☐ **IRS LISS4 FCC Imagery (background)**

Figure 12 PSO-SLEUTH modeled output for the year 2025, KMDA region

Modeling and visualizing these changes have always been regarded, helps in formulating effective local, regional and national level sustainable environmental policies and management strategies. CA-based models, SLEUTH model in particular, its origins, theory behind, application and its strength when integrated with multiple tools, multi-disciplinary approach and recent advanced techniques such as genetic algorithm, particle swarm optimization, ant colony optimization etc. proves to be worthy urban models. In this study, we perform an integrated novel method of adopting PSO-SLEUTH to optimize spread using Moore neighborhood, calibration and prediction of urban expansion for the year 2025. Unlike other optimization techniques, PSO is faster in terms of computation time, efficient, reliable and has lesser parameters for human intervention. These advantages are explored in this paper, BFM based PSO-SLEUTH calibration saw excellent improvement in results and it can be realized by obtaining OSM of 0.6 using PSO in comparison with 0.05 using traditional BFM. Apart from OSM, other model fit statistics also showed satisfactory values suggesting the robustness of SLEUTH model to integrate with any other improvised algorithm. Visuals from model output suggests KMDA to have at least 757.91km² of urban land use by the year 2025, which is 53% increment from 2017. There appears to be a considerable gap in research when it comes to validation and accuracy of models. Adopting optimization techniques along with agent-based models seems to be governing aspects of future efforts. These statistics and findings insists administrators, planners and concerned government officials to focus on emerging land use issues and sustainable urban growth concepts.

6. Acknowledgment

7. References

Alves, R.M. and Albrecht, C.H. 2010. The Use of Simulation in Urban Modelling. Methods and Techniques in Urban Engineering, 109.

Anderson, J.R., 1976. A land use and land cover classification system for use with remote sensor data (Vol. 964). US Government Printing Office.

Batty, M. and Xie, Y., 1994. From cells to cities. Environment and planning B: Planning and design, 21(7), 31-48.

Batty, M., 2009. Urban modeling. International encyclopedia of human geography. Elsevier, Oxford.

Bharath, H. A., Chandan, M.C. and Nimish, G., 2019. Assessing land surface temperature and land use change through spatio-temporal analysis: a case study of select major cities of India. Arabian Journal of Geosciences, 12(11), 367.

Bharath, H. A., Vinay, S., Chandan, M. C., Gouri, B. A. and Ramachandra, T. V., 2018. Green to gray: Silicon Valley of India. Journal of Environment Management, 206, 1287-1295.

Bharath, H. A. and Ramachandra, T.V., 2016. Visualization of urban growth pattern in Chennai using geoinformatics and spatial metrics. Journal of Indian Society of Remote Sensing, 44, 617-633.

Bharath, H. A., Vinay, S., Ramachandra, T.V., 2014. Landscape dynamics modelling through integrated Markov, Fuzzy-AHP and cellular automata. In: 2014 IEEE Geoscience and Remote Sensing Symposium, 3160-3163.

Boori, M.S., Voženílek, V. and Choudhary, K., 2015. Land use/cover disturbance due to tourism in Jeseníky Mountain, Czech Republic: A remote sensing and GIS based approach. The Egyptian Journal of Remote Sensing and Space Science, 18(1), 17-26.

Breiman, L., 2001. Random forests. Machine learning, 45(1), 5-32.

Chandan, M.C., Nimish, G. and Bharath, H.A., 2019. Analysing spatial patterns and trend of future urban expansion using SLEUTH. Spatial Information Research, 1-13.

Chen, S., Zeng, S. and Xie, C., 2000. Remote sensing and GIS for urban growth analysis in China. Photogrammetric Engineering and Remote Sensing, 66(5), 593-598.

Clarke-Lauer, M.D. and Clarke, K.C., 2011, July. Evolving simulation modeling: Calibrating SLEUTH using a genetic algorithm. In Proceedings of the 11th International Conference on GeoComputation, London, UK.

Cortes, C. and Vapnik, V., 1995. Support-vector networks. Machine learning, 20(3), 273-297.

Crooks, A.T., 2010. Constructing and implementing an agent-based model of residential segregation through vector GIS. International Journal of Geographical Information Science, 24(5), 661-675.

Das, S. K., Bharath, H.A. and Prakash, P.S., 2018. Automated building extraction using high-resolution satellite imagery though Ensemble modelling and Machine learning. Remote Sensing of Land, 1–14.

Dietzel, C. and Clarke, K.C., 2007. Toward optimal calibration of the SLEUTH land use change model. Transactions in GIS, 11(1), 29-45.

Dixon, B. and Candade, N., 2008. Multispectral landuse classification using neural networks and support vector machines: one or the other, or both? International Journal of Remote Sensing, 29(4), 1185-1206.

Duda, R.O., Hart, P.E. and Stork, D.G., 2012. Pattern classification. John Wiley & Sons.

Eberhart, R. and Kennedy, J., 1995, October. A new optimizer using particle swarm theory. In MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, IEEE, 39-43.

Feng, Y., Liu, Y., Tong, X., Liu, M. and Deng, S., 2011. Modeling dynamic urban growth using cellular automata and particle swarm optimization rules. Landscape and Urban Planning, 102(3), 188-196.

Ghosh, A. and Joshi, P.K., 2013. Assessment of pan-sharpened very high-resolution WorldView-2 images. International journal of remote sensing, 34(23), 8336-8359.

Goldberg, D.E., 1989. Genetic algorithms in search. Optimization, and MachineLearning. Addison Wesley.

Goldstein, N.C., 2004. Brains versus brawn-comparative strategies for the calibration of a cellular automata-based urban growth model. GeoDynamics, 249-272.

Guan, Q. and Clarke, K.C., 2010. A general-purpose parallel raster processing programming library test application using a geographic cellular automata model. International Journal of Geographical Information Science, 24(5), 695-722.

Holland, J.H., 1975. Adaptation in natural and artificial systems Ann Arbor. The University of Michigan Press, 1, p.975.

Horning, N., 2010, December. Random Forests: An algorithm for image classification and generation of continuous fields data sets. In Proceedings of the International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences, Osaka, Japan (Vol. 911).

Hu, W., Wang, H., Yan, L. and Du, B., 2018. A Hybrid Cellular Swarm Optimization Method for Traffic-Light Scheduling. Chinese Journal of Electronics, 27(3), 611-616.

Jafarnezhad, J., Salmanmahiny, A. and Sakieh, Y., 2015. Subjectivity versus objectivity: comparative study between brute force method and genetic algorithm for calibrating the SLEUTH urban growth model. Journal of Urban Planning and Development, 142(3), 05015015.

Jantz, C.A., Goetz, S.J., Donato, D. and Claggett, P., 2010. Designing and implementing a regional urban modeling system using the SLEUTH cellular urban model. Computers, Environment and Urban Systems, 34(1), 1-16.

Ji, C.Y., Liu, Q., Sun, D., Wang, S., Lin, P. and Li, X., 2001. Monitoring urban expansion with remote sensing in China. International Journal of Remote Sensing, 22(8), 1441-1455.

Li, X. and Liu, X., 2006. An extended cellular automaton using case-based reasoning for simulating urban development in a large complex region. International Journal of Geographical Information Science, 20(10), 1109-1136.

NASA, 2001. Satellite maps provide better urban sprawl insight. Science Daily. Accessed on May 3, 2019. Retrieved from https://www.sciencedaily.com/

Nimish, G., Chandan, M.C. and Bharath, H.A., 2018. Understanding current and future landuse dynamics with land surface temperature alterations: a case study of Chandigarh. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 4(5).

Nimish, G., Setturu, B., and Ramachandra, T.V., 2017. Modelling Landscape Dynamics in Protected Areas, Karnataka. Paper presented at Indian Biodiversity Congress - Vol 4, Puducherry, India.

Prakash, P.S., Soumya, K.D. and Bharath, H.A., 2018, November. Urban building extraction using satellite imagery through Machine Learning. In 2018 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 1670-1675.

Rabbani, A., Aghababaee, H. and Rajabi, M.A., 2012. Modeling dynamic urban growth using hybrid cellular automata and particle swarm optimization. Journal of Applied Remote Sensing, 6(1), 063582.

Ramachandra, T. V., Aithal, B. H., and Sanna, D. D. (2012). Insights to urban dynamics through landscape spatial pattern analysis. International Journal of Applied Earth Observation and Geoinformation, 18(1), 329–343.

Sakieh, Y. and Salmanmahiny, A., 2016. Treating a cancerous landscape: implications from medical sciences for urban and landscape planning in a developing region. Habitat International, 55, 180-191.

Shaw, R. and Das, A., 2018. Identifying peri-urban growth in small and medium towns using GIS and remote sensing technique: A case study of English Bazar Urban Agglomeration, West Bengal, India. The Egyptian Journal of Remote Sensing and Space Science, 21(2), 159-172.

Shi, Y., Liu, H., Gao, L. and Zhang, G., 2011. Cellular particle swarm optimization. Information Sciences, 181(20), 4460-4493.

Stephen, M.J., Reddy, P. and Vasavi, V., 2013. Fingerprint image enhancement through particle swarm optimization. International Journal of Computer Applications, 66(21), 34-40.

Thanh Noi, P. and Kappas, M., 2018. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. Sensors, 18(1), 18.

Torrens, P.M. and Benenson, I., 2005. Geographic automata systems. International Journal of Geographical Information Science, 19(4), 385-412.

Torrens, P.M., 2000. How cellular models of urban systems work (1. Theory).

Tran, H., Tran, T. and Kervyn, M., 2015. Dynamics of land cover/land use changes in the Mekong Delta, 1973–2011: A remote sensing analysis of the Tran Van Thoi District, Ca Mau Province, Vietnam. Remote Sensing, 7(3), 2899-2925.

UN DESA, 2015. World urbanization prospects: The 2014 revision. Accessed on February 25, 2019. Retrieved from URL of United Nations Department of Economics and Social Affairs, Population Division: New York, NY, USA.

UNSDS, 2015. New York Summit: Transforming our world: the 2030 Agenda for Sustainable Development. Accessed on April 7, 2019. Retrieved from https://www.un.org/

Verburg, P.H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V. and Mastura, S.S., 2002. Modeling the spatial dynamics of regional land use: the CLUE-S model. Environmental management, 30(3), 391-405.

Wolfram, S., 2002. A new kind of science (Vol. 5, 130). Champaign, IL: Wolfram media.

Wu, N. and Silva, E.A., 2010. Artificial intelligence solutions for urban land dynamics: a review. Journal of Planning Literature, 24(3), 246-265.

Yang, X., Zheng, X.Q. and Lv, L.N., 2012. A spatiotemporal model of land use change based on ant colony optimization, Markov chain and cellular automata. Ecological Modelling, 233, 11-19.

Yao, F., Hao, C. and Zhang, J., 2016. Simulating urban growth processes by integrating cellular automata model and artificial optimization in Binhai New Area of Tianjin, China. Geocarto International, 31(6), 612-627.