ASIAN CONFERENCE ON REMOTE SENSING ACRS 2017

IDENTIFICATION OF URBAN GROWTH DRIVING FACTORS IN PUNE USING REMOTE SENSING AND LOGISTIC REGRESSION

Lakshmi N. Kantakumar[1], Shamita Kumar[1] and Karl Schneider[2]

1: Bharati Vidyapeeth Deemed University, Institute of Environment Education and Research, Pune-Satara Road, Pune-411043, India
email: Lakshmikanth@bvieer.edu.in

2: Institute of Geography, University of Cologne, Zulpicher Str.45, 50674, Cologne, Germany

KEY WORDS: Urban growth, Driving factors, Remote sensing, Logistic regression, Pune.

ABSTRACT

Urban growth of Indian cities has recently received wide attention since the announcement of the smart city initiative by the Government of India. Understanding the spatial extent of urban growth along with the driving factors is crucial to plan and make a city more livable, sustainable or smart. Many Indian and international scholars have used remote sensing to quantify the urban sprawl of Indian cities, while paying less attention to understand the driving factors responsible for urban growth. In this study, the process of urban growth based upon driving factors identified and derived from remote sensing data were analyzed utilizing a multiple logistic regression technique. Fifteen driving factors were considered for the analysis pertaining to physical, proximal, socio-economic and legal characteristics. Using these driving factors as independent variables, a series of multiple logistic regression models were fitted to model the urban growth of Pune. All models were further carefully scrutinized to identify the best model based on the Akaike Information Criterion (AIC) value. The results reveal that the proximal characteristics (esp., distance to built-up, highways and industries) are the most important driving factors. The proxy socio-economic variable derived from DMSP OLS sensor showed a positive relationship to urban growth. Legal constraints such as reserved forest, red zone (buffer zone around deference establishment) and protected areas around water bodies showed the expected suppressive effect on urban development.

1. INTRODUCTION

Urban growth of Indian cities has recently received wide attention since the announcement of the smart city initiative by the Government of India. Many Indian and international scholars have used remote sensing to quantify the urban sprawl of Indian cities (Bhatta, 2009; Jat et al., 2008; Kantakumar et al., 2016; Taubenböck et al., 2009). While less attention was paid to understand the driving factors responsible for urban growth. Understanding driving factor and their effect is necessary to devise better and sustainable development plans to develop a city as smarter. Urban growth is typically driven by a variety of forces that relate one another based on different spatial and temporal scales (Thapa and Murayama, 2010). Understanding the spatial extent of urban growth along with the driving factors is crucial to plan and make a city more livable, sustainable or smart. In practice, three methods are often used to quantify the relationship of driving forces to land use (Verburg et al., 2004). The first approach is purely based on theoretical and physical laws such as economic and utility optimization models. The second approach is purely on the basis of empirical methods such as regression. The third approach is purely based on expert knowledge. Out of these three models, empirical model based on regression is often used to model the relationship between driving factors and urban growth (Cheng and Masser, 2003; Hu et al., 2007; Li et al., 2013; Ma and Xu, 2010; Mondal et al., 2015; Shu et al., 2014). Hu et al., (2007) used simple stepwise linear regression method to assess the driving factors of a mining based Xuzhou city of China. A similar study was performed on Guangzhou city of china by Ma and Xu, (2010) using curve fitting optimization technique. However, most of the studies around world used logistic regression to model the urban growth (Alsharif and Pradhan, 2013; Hu and Lo, 2007; Shu et al., 2014). Since the dependent variable urban growth is in binary nature, logistic regression can serve as best suitable method to model urban growth.

In this study remotely sensed data was used to study the driving factors responsible for urban growth in Pune city. Further, chi-square test was used to test the dependency of urban growth on driving factor derived from remote sensing data. The driving factors further used to model the urban growth of Pune by using logistic regression technique.

2. MATERIALS

2.1. Study Area

The Pune metropolis is one of the fastest growing metropolis in the Asia-pacific region. It is eight largest city in India and second largest city in Maharashtra state with population more than six million people. The study area is encompassing about 1643 sq. km area and bounded between the latitudes of 18.32 to 18.75 degrees north of equator and longitudes of 73.58 to 74.2 degrees east of Greenwich (Fig.1). The study area is home to many national and international automobile, information and bio-technology firms.

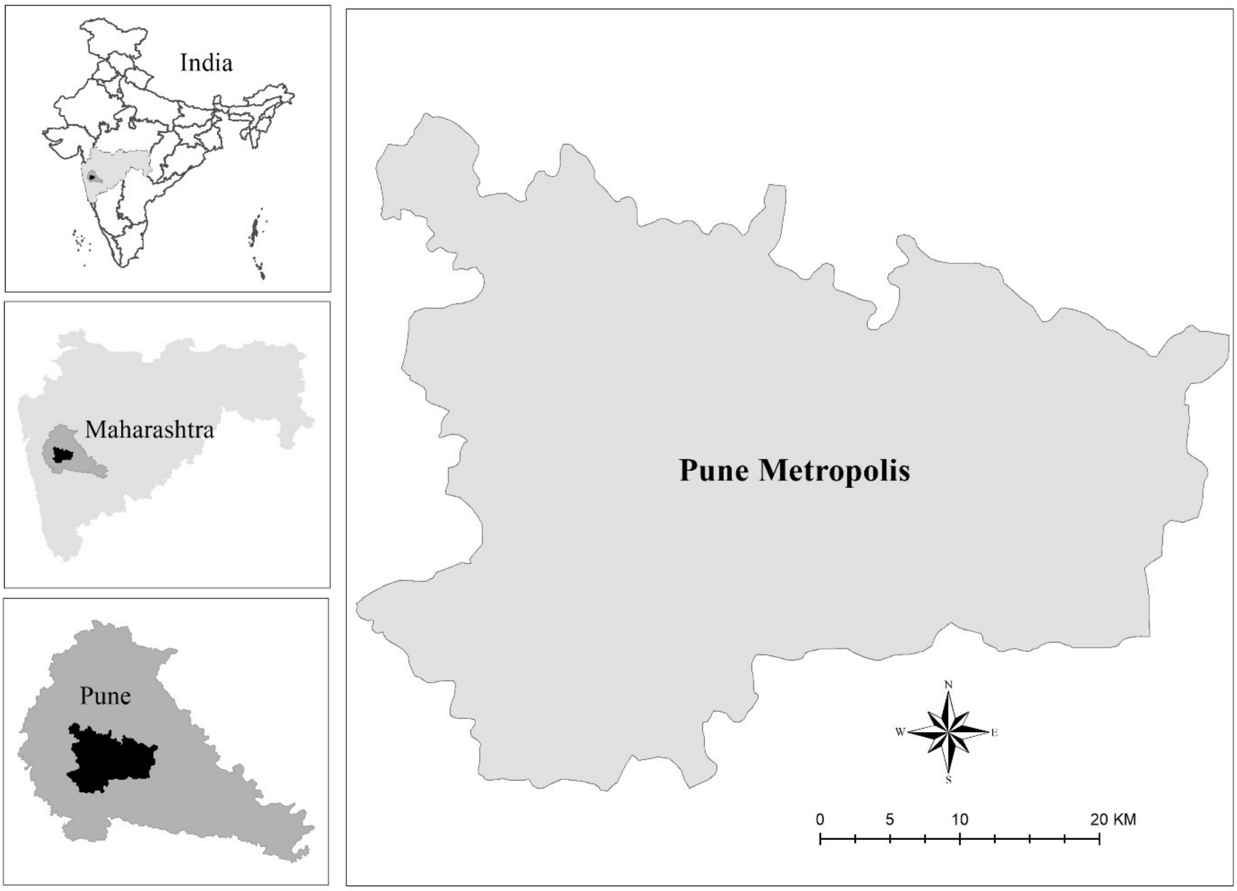

Figure 1: Study area map, i.e., Pune metropolis

2.2. Datasets

In this study, geometrically corrected cloud free multi-temporal datasets of Landsat-5 and 8 dated 04-Dec-1992 and 14-Dec-2013 respectively were downloaded from the public domain service of United States Geological Survey's Earth Resource Observation and Science (USGS EROS) data center, Sioux Falls, USA. The geometrically corrected satellite images were atmospherically corrected by using QUick Atmospheric Correction (QUAC). A knowledge based maximum likelihood supervised classification (Kantakumar and Neelamsetti, 2015) approach was used to classify the atmospherically corrected satellite images. The overall accuracy of land use classifications of 1992 and 2013 images are 89.3% and 89% respectively. Survey of India toposheet used to extract the road network, center of outgrowth villages and reserved forest maps. ASTER DEM version 2 used to prepare the slope map of the study area in degrees. The municipal corporation land use maps are used to prepare the location of industries, education centers, red zone and special economic zone maps of the study area.

2.3. Driving Factors

Fifteen driving factors were considered for the analysis pertaining to physical, proximal, socio-economic and legal characteristics (Fig. 2). The physical factor slope in degree is fundamental determent that shapes the urban form. The steeper slopes are less likely to be urbanized. The change in stable light intensity used as proxy measure for population density and Gross domestic product, because of population density and Gross domestic product were not available with sufficient spatial resolution during the study period. Eight proximity factors, viz., distance to city center, distance to outgrowth villages, distance to road, rail and airport, distance to education and industrial centers, distance to existing built-up land and percentage of built-up in one square kilometer neighborhood were included in logistic regression by assuming increasing distance from facilities will reduce the likely hood of the urbanization. The areas around defense vital installations, known as red zones, protected banks of water bodies and reserved forest and special economic zones were included in the regression under legal category.

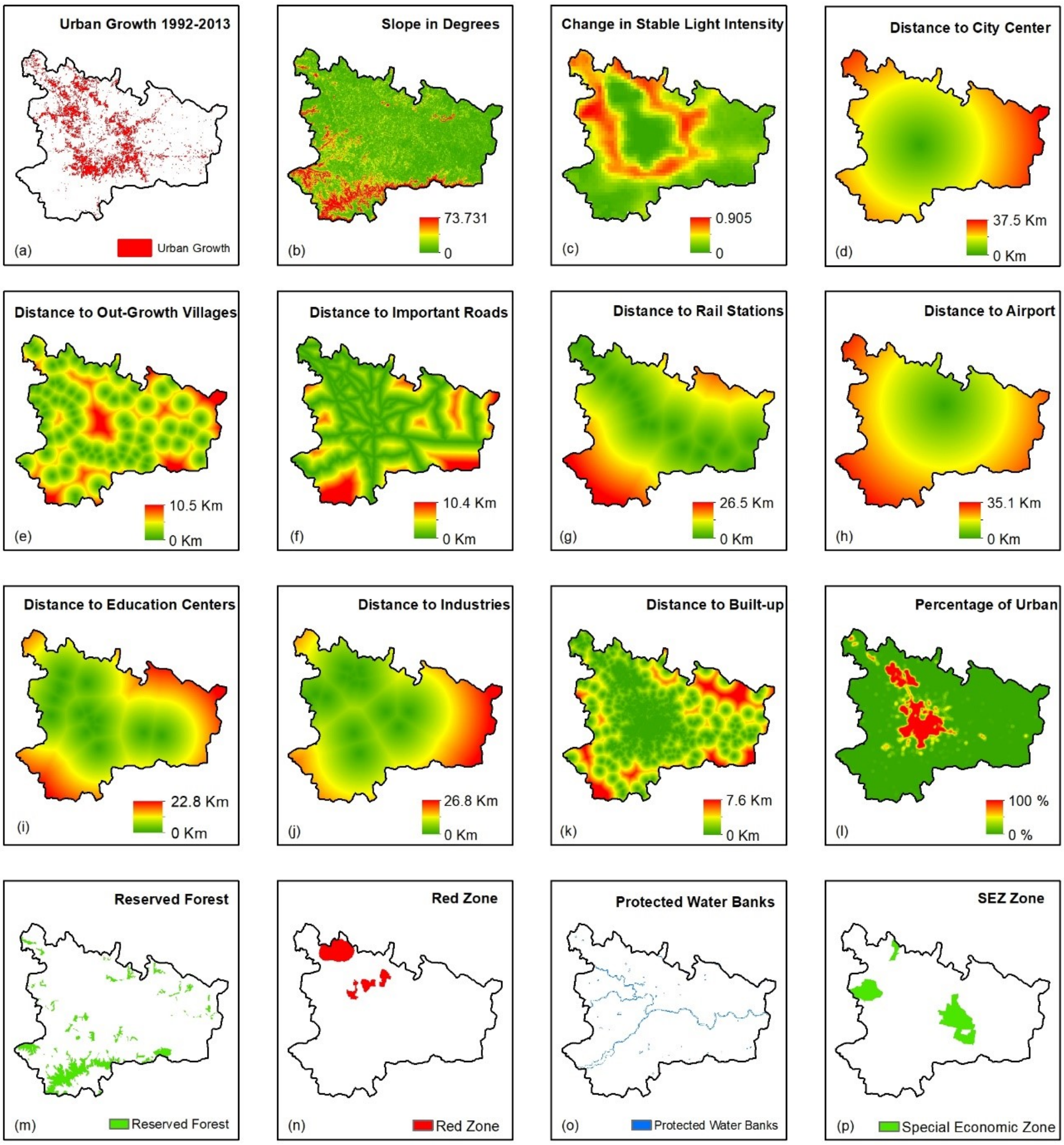

Figure 2: Dependent variable (a) Urban growth from 1992 to 2013. Independent variables Physical variable (b); Proxy Socio-economic variable (c); Proximity variables (d) to (l); Legal variables (m) to (p).

3. METHODS
3.1. Chi-square test of Independency

The Chi-square test of independency used to determine dependency of two nominal variables based on their proportion of distributions (McDonald, 2009). It is one of the most useful non-parametric test for hypothesis testing and also known as Pearson Chi-square test (McHugh, 2013). In this study χ^2 statistic was used to test the following null hypothesis.

Null hypothesis H_0: Urban growth and the driving factor understudy are independent
Alternative hypothesis H_1: Urban growth and the driving factor understudy are dependent.

The χ^2 statistic was calculated by using following formula (McHugh, 2013):

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Where, χ^2 = Chi-square value, O = observed values and E = expected values.

3.2. Multiple Logistic Regression

Multiple logistic regression is capable to model the relationship of a dichotomous variable with a set of continuous and categorical variables. The depended variable in this study is urban growth. It is binary in nature i.e., the state 1: represent a non-urban pixel in time t_1 converted into urban t_2 and state 0: represent no change in the state of non-urban pixel. The selection of sampling scheme is crucial to avoid spatial autocorrelation effect in model residuals, because logistic regression does not account spatial dependency. To overcome the spatial dependency in model residuals, the sampling scheme proposed by Cheng & Masser (2003) was used. As a first step the samples were placed systematically at 300m apart. Further random sampling approach used to collect 1278 samples where each state 0 and 1 contains 639 samples each. The general formula of multiple logistic regression used in this study is (Hosmer and Lemeshow, 2004):

$$P(X) = \frac{e^{f(X)}}{1 + e^{f(X)}}$$

Where, $P(X)$ is the conditional probability of a non-urban pixel to become urban and 'X' is the set of urban growth driving factors.

4. RESULTS AND DISCUSSION

4.1. Test of Dependency

The Chi-square test was used to test the dependency of urban growth on each driving factor under study. Since Chi-square test is in demand of nominal data, all the samples were categorized. Slope was categorized by using five degree interval, 10% interval was used for change in light intensity and 5% interval was used in percentage of urban in one square kilometer neighborhood variable. One kilometer used as interval in all proximity factors. Based on the intervals the observed urban pixels were counted in each category and then these counts are used to calculate the expected distribution. The observed and expected distributions are used to calculate the test statistic χ^2 at alpha level 0.05 to test the dependency of urban growth on each driver. The following table 1 shows the χ^2 statistic and accepted hypothesis.

Table 1: Results of Chi-square test at $\alpha = 0.05$

Driving Factor	Chi-Square statistic	Critical value	Degrees of Freedom	Accepted Hypothesis
Physical Factor				
Slope	38.314	14.067	7	H_1
Socio-economic factor				
Change in light intensity	41.308	16.919	9	H_1
Proximity Factors				
Dist. City centre	147.720	48.602	34	H_1
Dist. Out growth villages	42.946	15.507	8	H_1
Dist. Roads	134.269	18.307	10	H_1
Dist. Railway stations	111.771	37.652	25	H_1
Dist. Airport	112.424	47.400	33	H_1
Dist. Education centres	139.926	31.410	20	H_1
Dist. Industries	169.923	36.415	24	H_1
Dist. Built-up	183.660	12.592	6	H_1
Percentage of urban	92.560	26.296	16	H_1
Legal factors				
Red Zone	2.312	3.841	1	H_0
Reserved forest	25.253	3.841	1	H_1
Protected banks	0.514	3.841	1	H_0
SEZ	17.471	3.841	1	H_1

The results of chi-square test showed urban growth is depend on slope, change in light intensity and all proximity factors. However, red zone and protected banks were observed statistically independent with urban growth.

4.2. Analysis of Driving Factors

The Pune metropolis experienced 215.4 square kilometer urban growth from 107.5 sq.km in 1992 to 322.9 sq.km in 2013. A series of logistic regression models were fitted to model urban growth of the study areas by using several combinations of driving factors. As a first step, all the variables are used to model urban growth of the study area and referred as full model. The full model further used to find the best simplified model by stepwise elimination of driving factors with p-value greater than 0.10. The model with lowest Akaike Information Criterion (AIC) value was selected as the best simplified model. The following table 2 shows the results of fitted logistic regression models i.e., full model with all variable and best simplified model.

Table 2: Results of multiple logistic regression

	Full Model				Best Model			
Variable	β		Odds Ratio	S.E.	β		Odds Ratio	S.E.
Intercept	2.739	***	--	7.611	2.604	***	--	8.078
Physical Factor								
Slope	-0.015		0.985	-0.609	--		--	--
Socio-economic factor								
Change in light intensity	0.903	**	2.466	2.304	1.073	***	2.925	2.848
Proximity Factors								
Dist. City centre	-0.058	**	0.944	-2.279	--		--	--
Dist. Out growth villages	-0.164	**	0.849	-2.003	-0.152	*	0.859	-1.914
Dist. Roads	-0.491	***	0.612	-5.301	-0.481	***	0.618	-5.334
Dist. Railway stations	-0.116	***	0.891	-4.264	-0.086	***	0.918	-4.279
Dist. Airport	0.034	*	1.035	1.765	--		--	--
Dist. Education centres	0.051		1.052	1.377	--		--	--
Dist. industries	-0.114	***	0.892	-3.891	-0.128	***	0.880	-6.714
Dist. Built-up	-0.886	***	0.412	-5.491	-0.982	***	0.375	-6.575
Percentage of urban	0.021	**	1.021	2.289	0.023	***	1.023	2.609
Legal factors								
Red Zone	-0.703	**	0.495	-2.147	-0.828	***	0.437	-2.675
Reserved forest	-1.644	**	0.193	-2.571	-1.623	***	--	-2.644
Protected banks	-0.794	*	0.452	-1.823	-0.741	*	0.476	-1.732
SEZ	0.373		1.452	1.482	--		--	--
Nagelkerke's R2	0.55				0.55			
AIC	1124.18				1122.70			
PCP	81.30%				81.10%			
AUC	0.89				0.88			
Moran's I of Residuals	0.0489***				0.0497***			

****p*<.01; ***p*<.05; **p*<0.1

From the results of logistic regression (Tab. 2), it was found that slope, distance to education centers and special economic zones are found statistically insignificant during the study period from 1992 to 2013. The change in stable light intensity showed positive effect on urbanization. One percent increase in stable light intensity from DMSP-OLS will boost the chances of urbanization by 147% (2.466 – 1 ≈1.47). All the significant proximity factors excluding distance to airport and percentage of urban showed negative impact on urbanization. The increasing the distance from city center, out growth villages, road, rail stations, industries and built-up will decrease the chance of urbanization. Out of these proximity factors, the increasing distance to the built-up showed strong negative effect on the urbanization. According to the results of full model, one kilometer increase in distance to the built-up while keeping all other driving factors unchanged will decrease the odds of urbanization by 2.4 times (0.412:1 ≈ 1:2.4). The results show that urbanization mostly occurs near already existed built-up land (Shafizadeh-Moghadam and Helbich, 2015). One kilometer increase in distance from highway roads will decrease the odds of urbanization by 1.6 times. The results suggesting highways play an important role in urbanization. Stanilov & Batty (2011) has shown accessibility factors shape the urban form rather than the socio-economic factors in London and similar results were found in Beijing (Li et al., 2013). All the legal factors excluding SEZ areas showed negative impact on urbanization.

The performance of the full and simplified models assessed by using percentage of correct prediction (PCP) using 0.5 as threshold and Area Under Curve (AUC). An AUC value '1' represents perfect fit while a value '0.5' represents agreement due to chance. The PCP values are above 81% and AUC values are above '0.88' for both full and simplified model showing better model fit. It was found that the residuals of both full and simplified models showed significantly less positive spatial autocorrelation.

5. CONCLUSION

The study showed that the proximity factors and socio-economic factors are more important urban growth driving factors in Pune metropolis. Among the proximity factors, distance to the existing built-up land and roads are ranked in top of the list of driving factors along with the change in stable light intensity variable.

6. ACKNOWLEDGEMENT

The authors greatly acknowledge Landsat data from EROS data center, USGS, Sioux Falls, ASTER DEM data from Japan space system. *The first author gratefully acknowledge the support of* the director and other staff of Maharashtra and Goa Geospatial Data Center, Pune.

7. REFERENCES

Alsharif, A.A.A., Pradhan, B., 2013. Urban Sprawl Analysis of Tripoli Metropolitan City (Libya) Using Remote Sensing Data and Multivariate Logistic Regression Model. J. Indian Soc. Remote Sens. 42, 149–163. doi:10.1007/s12524-013-0299-7

Bhatta, B., 2009. Analysis of urban growth pattern using remote sensing and GIS: a case study of Kolkata, India. Int. J. Remote Sens. 30, 4733–4746. doi:10.1080/01431160802651967

Cheng, J., Masser, I., 2003. Urban growth pattern modeling: a case study of Wuhan city, PR China. Landsc. Urban Plan. 62, 199–217. doi:10.1016/S0169-2046(02)00150-0

Hosmer, D.W., Lemeshow, S., 2004. Applied Logistic Regression. John Wiley & Sons.

Hu, Z., Du, P., Guo, D., 2007. Analysis of Urban Expansion and Driving Forces in Xuzhou City Based on Remote Sensing. J. China Univ. Min. Technol. 17, 267–271. doi:10.1016/S1006-1266(07)60086-8

Hu, Z., Lo, C.P., 2007. Modeling urban growth in Atlanta using logistic regression. Comput. Environ. Urban Syst. 31, 667–688.

Jat, M.K., Garg, P.K., Khare, D., 2008. Monitoring and modelling of urban sprawl using remote sensing and GIS techniques. Int. J. Appl. Earth Obs. Geoinformation 10, 26–43. doi:10.1016/j.jag.2007.04.002

Kantakumar, L.N., Kumar, S., Schneider, K., 2016. Spatiotemporal urban expansion in Pune metropolis, India using remote sensing. Habitat Int. 51, 11–22. doi:10.1016/j.habitatint.2015.10.007

Kantakumar, L.N., Neelamsetti, P., 2015. Multi-temporal land use classification using hybrid approach. Egypt. J. Remote Sens. Space Sci. 18, 289–295. doi:10.1016/j.ejrs.2015.09.003

Li, X., Zhou, W., Ouyang, Z., 2013. Forty years of urban expansion in Beijing: What is the relative importance of physical, socioeconomic, and neighborhood factors? Appl. Geogr. 38, 1–10. doi:10.1016/j.apgeog.2012.11.004

Ma, Y., Xu, R., 2010. Remote sensing monitoring and driving force analysis of urban expansion in Guangzhou City, China. Habitat Int. 34, 228–235. doi:10.1016/j.habitatint.2009.09.007

McDonald, J.H., 2009. Handbook of biological statistics. Sparky House Publishing Baltimore, MD.

McHugh, M.L., 2013. The Chi-square test of independence. Biochem. Medica 23, 143–149. doi:10.11613/BM.2013.018

Mondal, B., Das, D.N., Dolui, G., 2015. Modeling spatial variation of explanatory factors of urban expansion of Kolkata: a geographically weighted regression approach. Model. Earth Syst. Environ. 1, 1–13. doi:10.1007/s40808-015-0026-1

Shafizadeh-Moghadam, H., Helbich, M., 2015. Spatiotemporal variability of urban growth factors: A global and local perspective on the megacity of Mumbai. Int. J. Appl. Earth Obs. Geoinformation 35, Part B, 187–198. doi:10.1016/j.jag.2014.08.013

Shu, B., Zhang, H., Li, Y., Qu, Y., Chen, L., 2014. Spatiotemporal variation analysis of driving forces of urban land spatial expansion using logistic regression: A case study of port towns in Taicang City, China. Habitat Int. 43, 181–190. doi:10.1016/j.habitatint.2014.02.004

Taubenböck, H., Wegmann, M., Roth, A., Mehl, H., Dech, S., 2009. Urbanization in India – Spatiotemporal analysis using remote sensing data. Comput. Environ. Urban Syst. 33, 179–188. doi:10.1016/j.compenvurbsys.2008.09.003
Thapa, R.B., Murayama, Y., 2010. Drivers of urban growth in the Kathmandu valley, Nepal: Examining the efficacy of the analytic hierarchy process. Appl. Geogr. 30, 70–83. doi:10.1016/j.apgeog.2009.10.002
Verburg, P.H., Schot, P.P., Dijst, M.J., Veldkamp, A., 2004. Land use change modelling: current practice and research priorities. GeoJournal 61, 309–324. doi:10.1007/s10708-004-4946-y