International Journal of Applied Earth Observation and Geoinformation 35 (2015) 187–198

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Spatiotemporal variability of urban growth factors: A global and local perspective on the megacity of Mumbai

Hossein Shafizadeh-Moghadam[a,b,*], Marco Helbich[a,c,1]

[a] Institute of Geography, University of Heidelberg, Germany
[b] Tarbiat Modares University, Department of GIS and Remote Sensing, Tehran, Iran
[c] Department of Human Geography and Spatial Planning, Utrecht University, Netherlands

ARTICLE INFO

Article history:
Received 6 November 2013
Accepted 19 August 2014
Available online 28 September 2014

Keywords:
Urban growth
Logistic regression
Autologistic regression
Geographically weighted logistic regression
GIS

ABSTRACT

The rapid growth of megacities requires special attention among urban planners worldwide, and particularly in Mumbai, India, where growth is very pronounced. To cope with the planning challenges this will bring, developing a retrospective understanding of urban land-use dynamics and the underlying driving-forces behind urban growth is a key prerequisite. This research uses regression-based land-use change models – and in particular non-spatial logistic regression models (LR) and auto-logistic regression models (ALR) – for the Mumbai region over the period 1973–2010, in order to determine the drivers behind spatiotemporal urban expansion. Both global models are complemented by a local, spatial model, the so-called geographically weighted logistic regression (GWLR) model, one that explicitly permits variations in driving-forces across space. The study comes to two main conclusions. First, both global models suggest similar driving-forces behind urban growth over time, revealing that LRs and ALRs result in estimated coefficients with comparable magnitudes. Second, all the local coefficients show distinctive temporal and spatial variations. It is therefore concluded that GWLR aids our understanding of urban growth processes, and so can assist context-related planning and policymaking activities when seeking to secure a sustainable urban future.

Introduction

Over the past few decades, urban expansion in India, mainly driven by globalization, privatization and liberalization, has resulted in rapid economic, social and environmental change taking place in the country (Van Ginkel, 2008; Tv et al., 2012). Excessive and uncoordinated urban growth has led to a loss of agricultural and forest land, as well as environmental resources, and has led to irreversible land-use conversion (e.g., Bhatta, 2009; Kumar et al., 2011; Punia and Singh, 2012; Munshi et al., 2014). Such rapid urban growth is of particular concern within the megacity[2] of Mumbai. Shafizadeh Moghadam and Helbich (2013) report that Mumbai experienced a 40% decrease in arable and open land in favor of built-up areas between 1973 and 2010. Moreover, additional and unbalanced urban growth is predicted to occur up to 2030, which will put a severe strain on sanitation services and basic infrastructure (e.g., public transportation systems and sewers), and lead to environmental damage and increasing social tensions. In this respect, Van Ginkel (2008) has highlighted how ill-prepared policymakers are for dealing with the outcomes related to this increase of urban living. As a consequence, Mumbai's policymakers face challenges with regard to land-use management, governance and urban planning. A crucial prerequisite for formulating sustainable future planning strategies and policies is to understand the past spatial developments of physical urban structures and the driving-forces behind urban growth (Cheng, 2011; Jokar Arsanjani et al., 2013a; Patino and Duque, 2013), as doing so can support the development of purposeful and goal-oriented planning strategies (Pethe et al., 2014).

Land-use change models (Verburg et al., 2004) based on geospatial technologies, multi-temporal remote sensing and spatial analysis, have proven to be valuable, efficient and technologically sound ways to analyze land conversion activities across space and over time. As well as being able to monitor urban growth (e.g., Bhatta, 2009; Kumar et al., 2011; Basawaraja et al., 2011;

* Corresponding author at: Berliner Straße 48, D-69120 Heidelberg, Germany. Tel.: +49 6221 54 4370.
E-mail addresses: shafeezadeh@uni-heidelberg.de, Hossein.shafeezadeh@geog.uni-heidelberg.de (H. Shafizadeh-Moghadam), m.helbich@uu.nl (M. Helbich).
[1] Heidelberglaan 2, 3584 CS Utrecht, Netherlands.
[2] According to the United Nations (2012), the term 'megacity' refers to an urban area with more than 10 million inhabitants.

http://dx.doi.org/10.1016/j.jag.2014.08.013

Taubenböck et al., 2012), a key advantage of land-use change models is their ability to determine and quantify the driving-forces behind spatiotemporal land-use transitions (e.g., Poelmans and Van Rompaey, 2010; Jokar Arsanjani et al., 2013a; Tayyebi and Pijanowski, 2014). Verburg et al. (2004) split these driving-forces into five categories: (a) environmental characteristics, (b) social factors, (c) economic factors, (d) spatial policies, and (e) spatial neighborhood interactions. Neighborhood variables link urban growth models to economic theories (e.g., the core-periphery model by Krugman, 1991, cited in Dendoncker et al., 2007). In spite of the large number of possible drivers, the majority of empirical studies operationalize environmental (e.g., distance to transportation infrastructures) and, if available, socio-economic determinants (e.g., household incomes), in order to explain urban growth processes (e.g., Hu and Lo, 2007; Poelmans and Van Rompaey, 2010; Cheng, 2011).

A wide-range of advanced methods have been developed to analyze urban expansion processes (Jokar Arsanjani et al., 2014), and these can be divided into descriptive approaches, simulation-based models and statistical urban growth models. Descriptive approaches aim to quantify spatiotemporal urban patterns by means of spatial landscape matrices. For example, Tv et al. (2012) and Punia and Singh (2012) describe the shape, "clumpiness" and "patchiness" of urban growth patterns over time. In contrast to landscape matrices, cellular automata (CA) models focus on the prediction of future urban extents rather than on the characterization of urban patterns (Aljoufie et al., 2013; Jokar Arsanjani et al., 2013a). These bottom-up, cell-based approaches use site-specific rules to simulate urban dynamics over discrete time steps. While both approaches have contributed substantially to our knowledge of urban growth processes, they lack the ability to determine the underlying driving-forces behind such growth (Cheng and Masser, 2003), so instead regression models (e.g., Dubovyk et al., 2011), support vector machines (e.g., Huang et al., 2009), and artificial neural networks (ANN; e.g., Chu et al., 2013) are commonly applied for that purpose. However, the majority of studies have applied non-spatial logistic regression (LR) models (McCullagh and Nelder, 1989) to explain complex urban growth patterns, using a set of predictors (e.g., Hu and Lo, 2007; Poelmans and Van Rompaey, 2010; Dubovyk et al., 2011). While Cheng and Masser (2003) focus on the effectiveness of LR at being able to determine driving-forces, and stress its extensive explanatory power, Hu and Lo (2007) emphasize the multi-scale calibration abilities of LR; so reducing the computational burden. In comparison to the large number of LR applications in existence, spatial autocorrelation (SAC) has received relatively little attention (Dendoncker et al., 2007), and even less research has sought to spatially varying relationships. SAC refers to the coincidence between locational and attribute similarities (Anselin, 2009), the presence of which denotes that non-built-up areas adjacent to existing built-up areas are more likely to themselves become built-up than those areas further away. As SAC violates model assumptions, the parameters estimated may not be reliable (e.g., Augustin et al., 1996; Anselin, 2009).

To account for SAC, spatial sampling can be used, which causes a significant loss of information and makes parameter estimations less reliable (Cheng and Masser, 2003). Alternatively, a spatially explicit model like autologistic regression (ALR) can be applied (Augustin et al., 1996; Huang et al., 2009). ALR assumes that an autocovariate term absorbs SAC and that parts of the variance can be explained through neighborhood effects, by relating a cell's transition to its surroundings (Overmars et al., 2003). While Lin et al. (2011) report a higher level of accuracy for ALR when compared to LR, Dendoncker et al. (2007) highlight the efficiency of ALR when accounting for SAC. In contrast, Dormann (2007) shows that ALR tends to underestimate the true model parameters when compared to aspatial LR, and recommends coupling ALR with alternative models. More recently, Lin et al. (2011) tested the predictive accuracy of LR, ALR and ANN, finding that both the ALR and ANN models perform better than LR, while the difference between ALR and ANN is marginal. In addition, regression-based techniques are easier to interpret than ANNs, which require additional algorithms to be used to derive a variable's importance (Hagenauer and Helbich, 2012).

Despite the fact that ALR appeals for use with land-use modeling, it assumes that a single equation represents land-use transitions and, thus, disregards local variations as potential driving-forces behind spatial characteristics (e.g., Fotheringham et al., 2002). Such a global perspective is too simplistic for Mumbai, and tends to over generalize growth, resulting in invalid policies on a local level. Therefore, it is rational to explore the spatial variability of the driving-forces using geographically weighted logistic regression (GWLR; Brunsdon et al., 1996; Atkinson et al., 2003). When compared to LR and ALR, GWLR has received less attention within urban growth modeling circles, even though the approach examines local land-use conditions based on the spatial variation of drivers (Luo and Wei, 2009), while simultaneously reducing SAC and having appealing visualization capacities for policy-oriented research (Ogneva-Himmelberger et al., 2009).

To conclude, global non-spatial and spatial regression models currently dominate land-use modeling. In this respect, the main objective of this research is to investigate local urban growth drivers of growth in Mumbai. The city's rapid urban expansion makes it an excellent case study for exploring and retrospectively cross-comparing urban growth drivers over the periods 1973–1990, 1990–2001 and 2001–2010. The following questions will be addressed:

- What are the main driving-forces behind Mumbai's urban land-use transition?
- Have these driving-forces been constant over different time periods and across space?
- Does the GWLR model out-perform the global non-spatial LR and spatial ALR models?

Materials

Study area

As the country's commercial and financial center, Mumbai represents one of India's key megacities. The city lies in the state of Maharashtra and is located on the west coast, next to the Arabian Sea (Fig. 1). Mumbai covers an area of 430 km^2 and can be characterized as a poly-nuclear region with emerging sub-centers (Pacione, 2006).

Between 1971 and 2011, Mumbai's population increased steadily, from approximately 5,971,000 to more than 12,478,000 inhabitants. The United Nations (2012) has predicted that the population will continue to increase to 27 million inhabitants by 2025. This growth in population has been accompanied by a massive growth in the number of condominiums and office towers, shopping malls and multiplexes being built, as well as motorways. These large investments have polarized Mumbai; it is becoming both an emerging global city and a place full of informal settlements – a predominant part of the urban landscape (Pethe et al., 2014).

Data and pre-processing

As the world's oldest data archive – available free of charge to users and constantly being updated – Landsat imagery offers efficient mid-resolution remote sensing products, those which can be

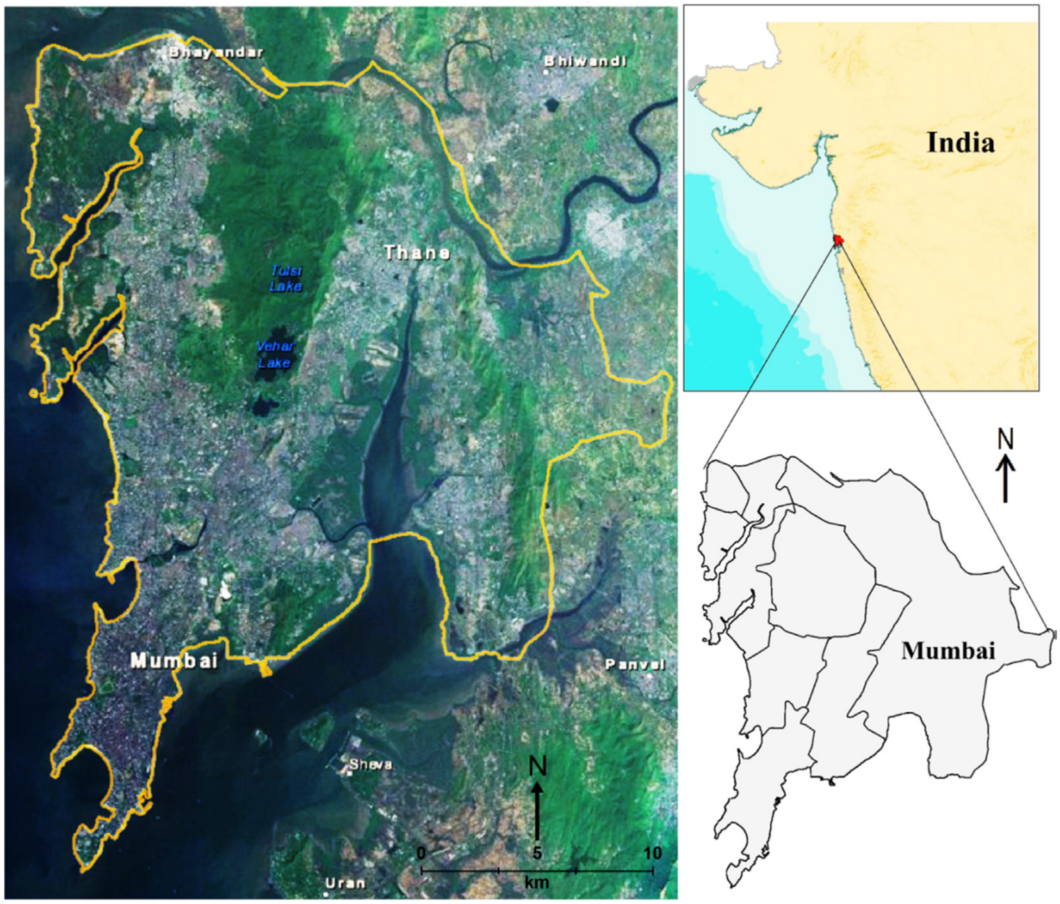

Fig. 1. Study area of Mumbai.

Table 1
Datasets used in the study.

Data	Source	Date	Spatial resolution
Landsat (MSS)	U.S. Geological Survey, German Aerospace Center	1973	79 m
Landsat (ETM, ETM+)	U.S. Geological Survey, German Aerospace Center	1990, 2001, 2010	30 m
Digital elevation model	ASTER (NASA)	2009	30 m
Transportation network	OpenStreetMap	2011	–

used to analyze the megacity Mumbai (Taubenböck et al., 2012; Patino and Duque, 2013). Landsat images were used here to monitor past urban growth and to extract the land-use classes required for the years 1973, 1990, 2001 and 2010. Due to the different spatial resolutions of Landsat products (79/30 m), all data were projected and re-sampled to a common resolution of 30 m. Table 1 gives an overview of all the datasets examined.

Although socio-demographics and zoning plans are utilized in urban growth models (e.g., Hu and Lo, 2007), such precise datasets are not accessible in most developing countries (Luo and Wei, 2009; Jokar Arsanjani et al., 2014), this also being the case for Mumbai. Consequently, we have had to limit the analysis to environmental data, distance-based parameters and neighborhood variables (e.g., Cheng and Masser, 2003; Dendoncker et al., 2007). However, while Dubovyk et al. (2011) have highlighted the practical benefit of producing less complex models, Luo and Wei (2009) demonstrate that these drivers are sufficient when formulating meaningful land-use models, and are able to produce an excellent model fit. Therefore, we decided to use those drivers (Table 2; Fig. 2). Proximity variables are considered in two ways: (a) as proximity to infrastructure factors (e.g., distance to roads), and (b) in the sense of urban economics, where it is assumed that city expansion occurs in more remote areas owing to a larger supply of open land and lower land prices. With the exception of altitude and slope data, which are self-explanatory, other distance-based factors are based on the Euclidean minimum distance to selected features. It must be noted that street data for different timestamps (Cheng, 2011) – that are also suitable for network-based accessibility analysis – was not available here. However, major roads were relatively stable over time and are extracted from OpenStreetMap (Helbich et al., 2012). As indicated by Dendoncker et al. (2007), neighborhood variables modeled by means of focal GIS computations are essential when wishing to operationalize the fact that developments in one area have an impact on adjacent areas. After testing alternative window sizes and considering literature suggestions, a 7 × 7 window was applied.

Methods

Fig. 3 summarizes the methodological aspects of the analysis. While "Determination of an appropriate cell size" section discusses how the optimal cell size for subsequent models is determined, "Non-spatial logistic regression" and "Autologistic regression" sections present the LR and the ALR models. Section "Geographically weighted logistic regression" outlines the principles behind GWLR.

Table 2
Exploratory variables and summary statistics for the year 2001.

Covariates	Literature examples	Range	Mean	SD
Distance to the CBD* (m)	Poelmans and Van Rompaey (2010)	0–33,000	18,069	7482
Distance to main roads* (m)	Cheng (2011)	0–4330	597	748
Distance to built-up areas (m)	Huang et al. (2009)	0–4300	536	764
Distance to water bodies (m)	Luo and Wei (2009)	0–23,000	2636	2150
Distance to wetlands (m)	Tayyebi and Pijanowski (2014)	0–22,000	4188	3775
Distance to forests (m)	Tayyebi and Pijanowski (2014)	0–16,000	2793	2698
Density of built-up areas (cells per area)	Cheng and Masser (2003)	0–49	29	15
Density of avail. open/arable land (cells per area)	Dubovyk et al. (2011)	0–49	24.8	16.8
Altitude* (m)	Jokar Arsanjani et al. (2013a)	0–487	39	60
Slope* (degree)	Hu and Lo (2007)	0–60	5.86	6.01

Note: Covariates marked with a '*' are considered as constant across time. Otherwise $t-1$ is used in subsequent regression. E.g., built-up changes 2001–2010 are regressed on covariates of the year 2001.

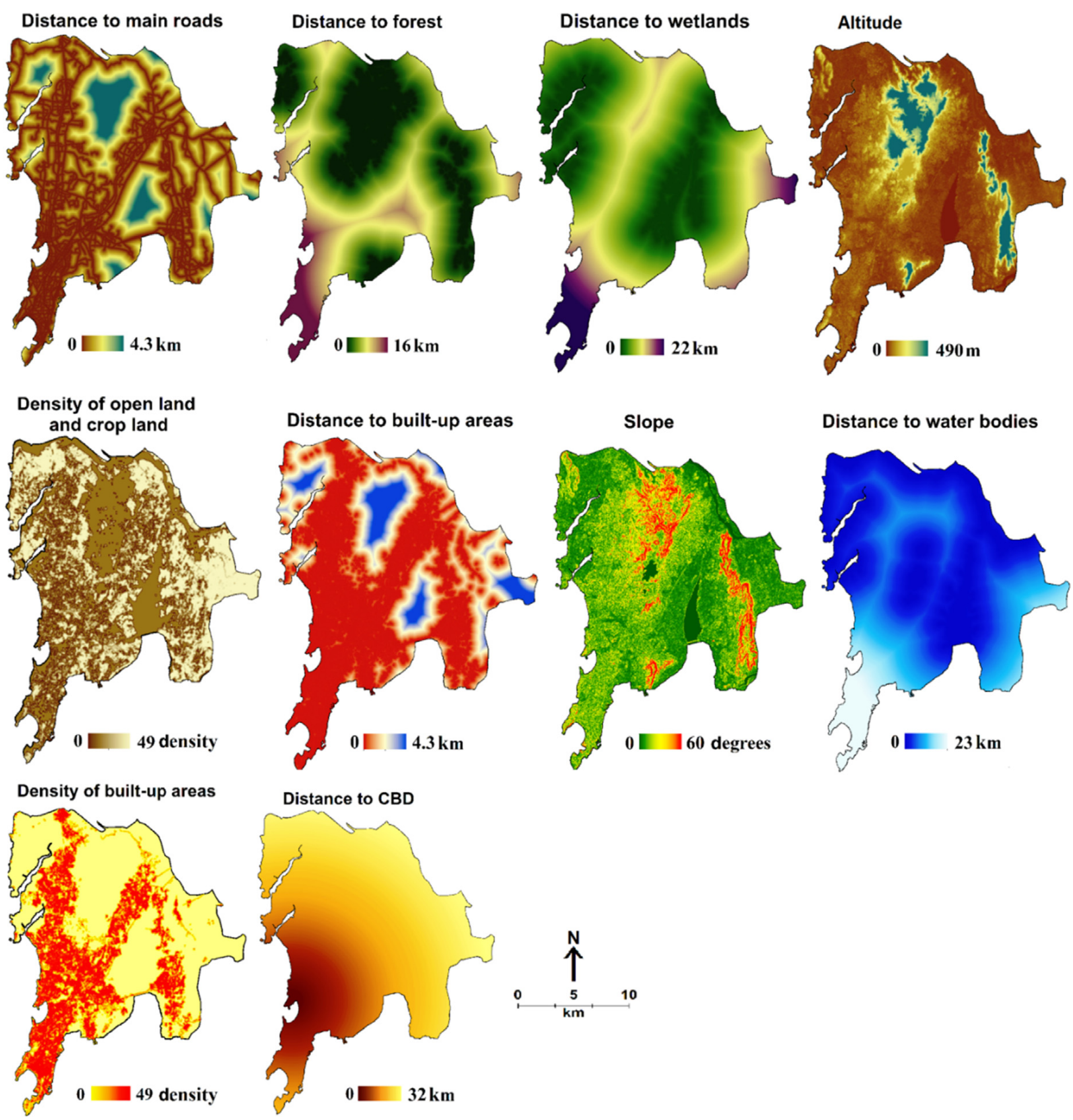

Fig. 2. Driving-forces for 2001.

Determination of an appropriate cell size

Massive amounts of data still present a challenge to spatial statistical modeling. The pre-processed data for this study was available at a spatial resolution of 30 m. While global models are computationally less intensive, GWLR cannot handle one million cells owing to, for example, insufficient memory capacity. For reasons of consistency between global and local models, data was

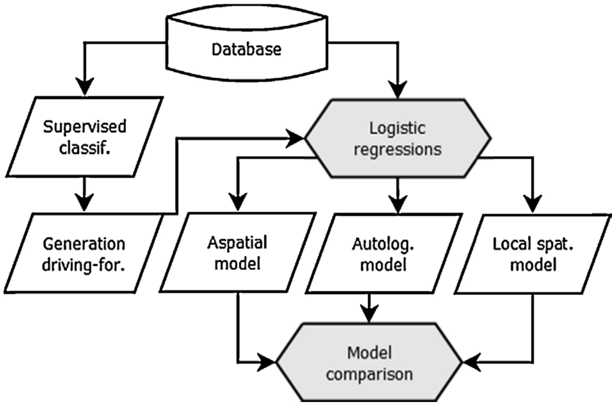

Fig. 3. Research design.

therefore reduced. Following Hu and Lo (2007), we determined a suitable modeling scale through regression-based performance comparisons, using the Nagelkerke R^2 goodness-of-fit measure.

Non-spatial logistic regression

Multiple LR is part of the generalized linear model family (GLM; McCullagh and Nelder, 1989) and links a binary response to a set of continuous and/or categorical covariates. In this study, $Y=1$ if a cell has changed into a built-up class between two timestamps, otherwise $Y=0$. LR allows us to determine the probability that a cell will change its state to built-up, identify significant drivers, and measure the magnitude of their impact on land-use transitions. The LR model used here can be specified as follows:

$$P_{LR} = \log\left(\frac{P}{1-P}\right) = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k \quad (1)$$

where P_{LR} indicates the probability of a cell being transformed into a built-up class, x_k represent the k covariates, α represents the intercept, and β_k are the estimated coefficients for x_k, indicating the impact of a covariate. The LR model assumes that Y is independently distributed across space, which is rather a strong assumption to make in land-use modeling (Lin et al., 2011), and, thus, spatially explicit models such as the ALR tend to be more reasonable.

Autologistic regression

The ALR represents a spatial variant of the LR (Augustin et al., 1996; Lin et al., 2011). By considering an autocovariate term, this model explicitly considers SAC. The ALR can be written as:

$$P_{ALR} = \log\left(\frac{P}{1-P}\right) = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k + \beta_{k+1} autocov_{k+1} \quad (2)$$

The autocovariate term models neighborhood effects by means of an inverse distance-weighted average of j cells in the neighborhood of cell i

$$autocov_i = \frac{\sum d_{ij}^{-1} y_i}{\sum d_{ij}^{-1}}$$

Both global models are limited by the assumption that the behavior of the parameters across space is homogeneous, which may be a model misspecification. Therefore, local models such as GWLR can supplement global analyses.

Geographically weighted logistic regression

GWLR extends the GLM framework by permitting local spatial variations of the parameters (Atkinson et al., 2003). In brief, GWLR performs a set of weighted regressions on a spatial data subset based on a moving-window approach (Brunsdon et al., 1996). The GWLR is written as:

$$P_{GWLR_{(u_i v_i)}} = \log\left(\frac{P_{(u_i,v_i)}}{1-P_{(u_i,v_i)}}\right) = \alpha_{(u_i,v_i)} + \beta_{1(u_i,v_i)} x_1 + \cdots + \beta_{k(u_i,v_i)} x_k \quad (3)$$

where $P_{GWLR_{(u_i v_i)}}$ is the probability of a cell i at the coordinates (u_i, v_i) for being transformed into a built-up cell, and $\beta_{k(u_i,v_i)}$ is a continuous function for the location i. The weighting process involved is based on a spatial kernel function where the weights depend on the Euclidean distance between regression point i and the data points j within a moving-window. Points further away from i receive less weight than points in close proximity, while points outside the kernel are not taken into consideration at all. Following Luo and Wei (2009), the bi-square kernel function is utilized as follows:

$$w_{ij} = \begin{cases} \left[1-\left(\dfrac{d_{ij}}{b}\right)^2\right]^2 & \text{if } d_{ij} < b \\ 0 & \text{if } d_{ij} > b \end{cases} \quad (4)$$

where b is the bandwidth, and d_{ij} represents the distance between the points i and j. More important than the kernel type is the bandwidth b, which describes how many nearest-neighbor points are taken into consideration per local regression. An adaptive bandwidth is appealing, because in sub-regions where the data points are densely distributed across space, a smaller bandwidth is required to model changes over small distances, those which might be missed by overly large, fixed threshold settings. In turn, to circumvent the lack of data caused by an overly narrow bandwidth – which may result in less reliable estimates – regions with a more widely dispersed distribution of data points require larger bandwidths. To determine an optimal bandwidth, the Akaike information criterion (AIC) is minimized (Fotheringham et al., 2002). Even though GWLR has matured, Wheeler and Tiefelsdorf (2005), and Cho et al. (2009), among others, note potential methodological discrepancies, including the multicollinearity of local coefficients and the extreme coefficients which emerge from local distributional irregularities. Thus, it is recommended that GWLR parameter surfaces should not be interpreted uncritically.

Results and discussion

Descriptive and exploratory analysis of urban growth

After the remote sensing images were geo-referenced, rectified, and cropped to the study area, a supervised maximum likelihood classification (Feizizadeh and Helali, 2010) was applied to each Landsat scene. This resulted in a time series of four maps (see the supplementary material) which differentiate between five land-use categories: (a) built-up areas, (b) forests and green space, (c) open land and arable land, (d) wetlands, and (e) water bodies. Among these land-use categories, the built-up areas, subsuming commercial and industrial buildings, and residential areas, among others, are of major relevance. Classification accuracy was evaluated on the basis of 250 randomly selected points and the kappa coefficient. Kappa values between 0.84 and 0.86 provide sufficient classification accuracy for spatial analysis to be carried out (Pontius et al., 2004). Subsequent to the extraction of built-up areas, the detection of changes over the time periods 1973–1990, 1990–2001 and 2001–2010 were computed using GIS-based pixel-by-pixel overlay analysis (Fig. 4). The resulting map shows a considerable increase in built-up areas since 1973. Built-up areas increased from

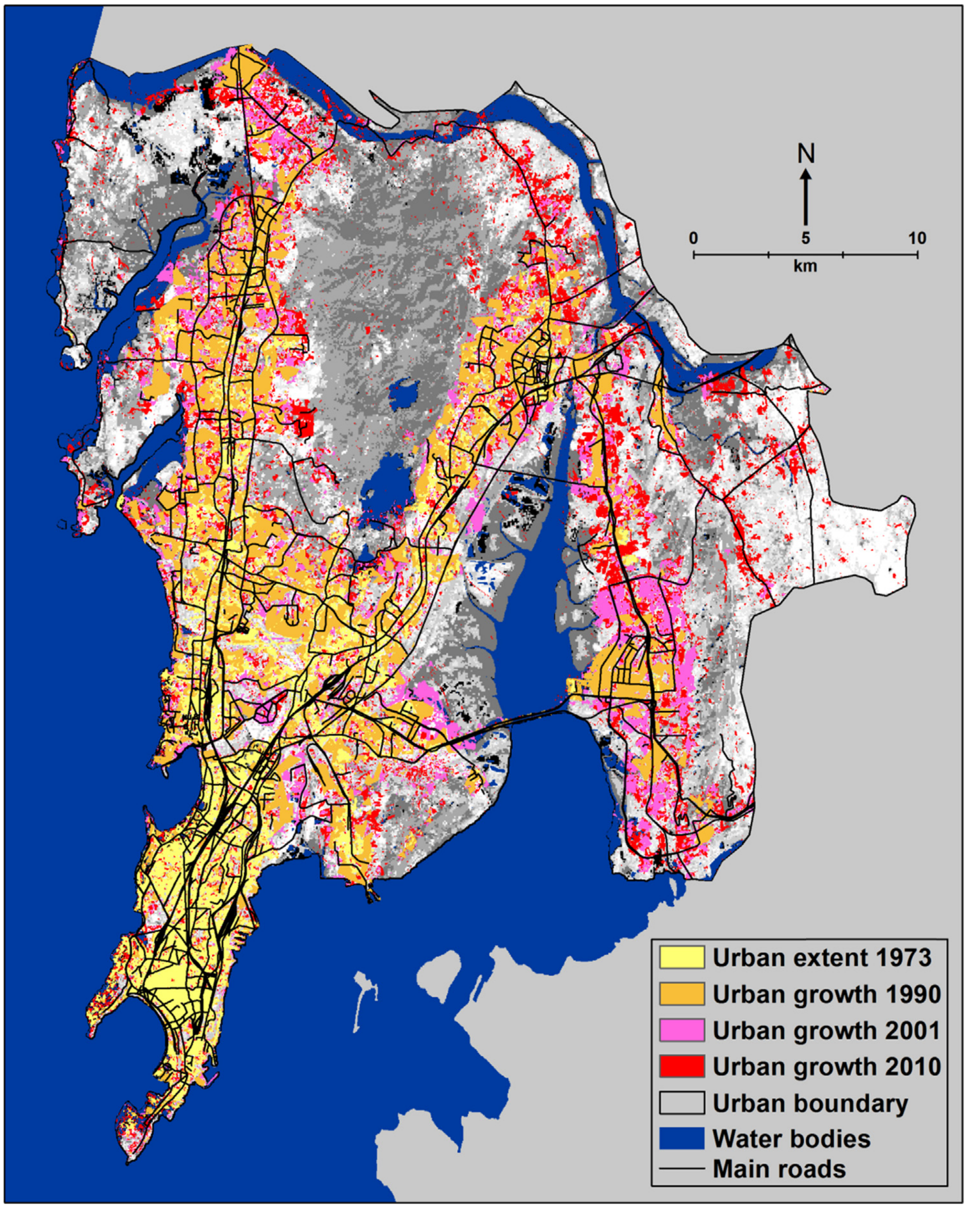

Fig. 4. Urban growth in Mumbai from 1973 to 2010.

approximately 7630 ha in 1973 to 35,607 ha in 2011 (+367%), predominantly in the northern and eastern areas along the main traffic routes.

Next, the spatial distribution of urban expansion was analyzed to determine whether growth had occurred evenly over the study area. With the presence of binary-scaled data, the join count statistic (Cliff and Ord, 1981) measure is a proper exploratory tool to use. The statistic indicated significant clustering of built-up areas ($p < 0.001$), potentially impacting on the subsequent regression estimates.

Global modeling results

As outlined in “Determination of an appropriate cell size” section, the analysis scale was determined in a data-driven manner using LR performance analysis across several cell sizes. Fig. 5 shows that models based on a cell size of 270 m do not differ significantly from those using a resolution of 30 m. Beyond this 270 m threshold, the Nagelkerke R^2 declines rapidly. Thus, 270 m represents a trade-off between computation feasibility, temporal and spatial generality, and model performance. Due to stable global model parameters across different aggregations, it seems that the results are not affected by the modifiable areal unit problem (Jelinski and Wu, 1996).

For each timestamp, an independent, full LR model was estimated. To assure more parsimonious models, stepwise model selection was conducted, where covariates, insignificant at the 0.05 level, were removed. Variance inflation factors indicated no problems with multicollinearity. For all the time periods, four covariates, including the distance to roads, the density of built-up areas, the distance to built-up areas, and the density of available open and arable land were shown to be significant. However, the altitude was only significant for the periods 1990–2001 and 2001–2010. Exploring the LR residuals by means of Moran’s *I*, a significant SAC was revealed ($p < 0.050$), which violates the independence assumption. Hence, the models were re-estimated in a spatial setting using ALR. The significant autocovariate terms reduce residual SAC not explained by the covariates considerably, even though

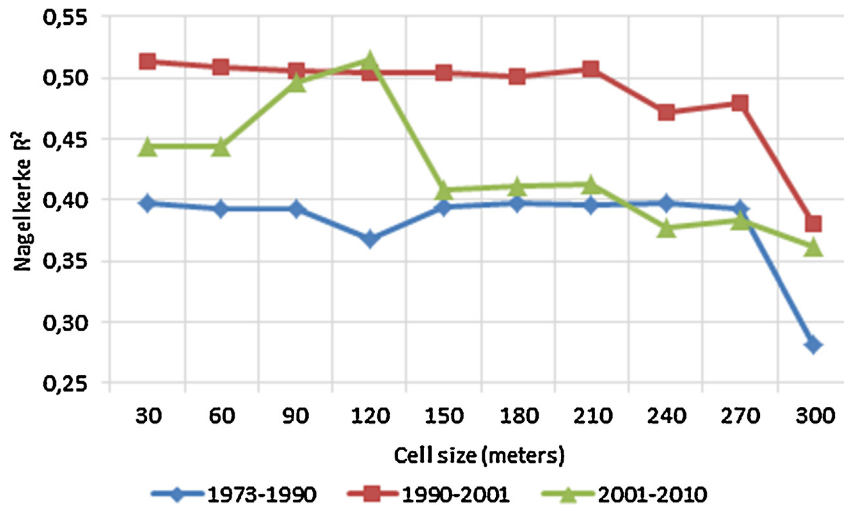

Fig. 5. Regression performance for different aggregation levels.

it remained slightly significant, indicating that neighborhood effects matter. Sensitivity analyses with different autocovariate specifications confirmed the robustness of our results (Table 3).

Overall, the ALRs' signs are intuitive, and while they are in-line with Luo and Wei's findings (2009), the magnitudes of the estimated coefficients are less pronounced. When all other covariates remain constant, the 2001–2010 ALR coefficient for distance to roads implies a negative impact on the odds ratio. As in Cheng and Masser (2003) and Hu and Lo (2007), this driver has a strong impact. Each additional 1000 m distance from the main roads makes it three times less likely that new built-up cells will be found (odds ratio: $e^{1001\times -0.001} = 0.367 \approx 1/3$; Poelmans and Van Rompaey, 2010). Although this driver is less pronounced for the period 1973–1990, its importance increases over time, and since the mid-1970s, road accessibility has been a major driving-force behind urban expansion in Mumbai. As Shafizadeh Moghadam and Helbich (2013) point out, traffic connections will also become increasingly important for the future urban development of Mumbai, predominantly along the two main traffic arteries in the north-east and north-west (Fig. 4). It is expected that these extensions will further increase land transformation pressure in those areas. The estimated sign for the variable density of built-up areas contradicts the findings of Luo and Wei (2009). In our case a higher density of built-up areas has a restrictive influence on urban development. The effect of the covariate distance to built-up areas has a similar sign and magnitude as distance to roads. Once again, areas close to existing built-up areas are more likely to be faced with urban developments. Thus, households look for locations next to existing built-up areas,

Table 3
LR and ALR estimation results.

	LR		S.E.	*p* value	ALR		S.E.	*p* value
	β	Exp(β)			β	Exp(β)		
1973–1990								
Distance to roads	−0.003	0.997	<0.001	<0.001	−0.003	0.997	<0.001	<0.001
Density of built-up areas	−0.014	0.986	0.003	<0.001	−0.017	0.983	0.003	<0.001
Distance to built-up areas	-0.001	0.999	<0.001	<0.001	-0.001	0.999	<0.001	<0.001
Density avail. open/arable land	0.001	1.001	<0.001	<0.005	0.002	1.002	<0.001	<0.001
Intercept	−1.887	0.151	0.093	<0.015	−4.877	0.008	0.31	<0.001
Autocovariate	–				14.438			<0.001
AIC	4881				4781			
Moran's *I* residuals	0.151			<0.001	0.048			<0.001
1990–2001								
Distance to roads	−0.002	0.998	<0.001	<0.001	−0.001	0.999	<0.001	<0.001
Density of built-up areas	−0.01	0.998	0.004	<0.012	−0.003	0.997	0.004	<0.003
Distance to built-up areas	−0.002	0.99	<0.001	<0.001	−0.001	0.999	<0.001	<0.001
Density avail. open/arable land	0.008	1.008	<0.001	<0.001	0.008	1.008	<0.001	<0.001
Altitude	−0.004	0.996	0.002	<0.050	−0.002	0.998	0.002	0.427
Intercept	−4.678	0.009	0.166	<0.001	−9.108	<0.001	0.402	<0.001
Autocovariate	–				41.724			<0.001
AIC	3424				3266			
Moran's *I* residuals	0.087			<0.001	0.055			<0.001
2001–2010								
Distance to roads	−0.002	0.998	<0.001	<0.001	−0.001	0.999	<0.001	<0.001
Density of built-up areas	−0.010	0.99	0.004	<0.012	−0.003	0.997	0.004	<0.003
Distance to built-up areas	−0.002	0.998	<0.001	<0.001	−0.001	0.999	<0.001	<0.001
Density avail. open/arable land	0.008	1.008	<0.001	<0.001	0.008	1.008	<0.001	<0.001
Altitude	−0.002	0.998	0.001	<0.037	−0.002	0.998	0.001	<0.046
Intercept	−4.678	0.009	0.166	<0.001	−9.108	<0.001	0.402	<0.001
Autocovariate	–				29.652			<0.001
AIC	5125				5065			
Moran's *I* residuals	0.131			<0.001	0.019			0.231

Table 4
GWLR results.

1973–1990	Min.	1st QT	Median	3rd QT	Max.	Non-stationarity test[a]
Distance to roads	−0.008	−0.004	−0.003	−0.001	0.003	*
Density of built-up areas	−0.074	−0.020	0.008	0.046	0.305	*
Distance to built-up areas	−0.021	−0.003	−0.001	−0.001	0.001	*
Density avail. open/arable land	−0.008	−0.002	0.000	0.002	0.007	*
Intercept	−3.028	−0.477	0.000	0.528	2.471	*
AIC	2635					
Adj. R^2	0.330					
Moran's *I* error	0.013	$p < 0.001$				
1990–2001						
Distance to roads	−0.005	−0.002	−0.001	0.000	0.002	*
Density of built-up areas	−0.083	−0.020	0.002	0.021	0.100	*
Distance to built-up areas	−0.015	−0.005	−0.003	−0.001	0.004	*
Density avail. open-/arable land	0.000	0.005	0.008	0.011	0.020	*
Altitude	−0.051	−0.009	0.001	0.013	0.098	*
Intercept	−6.843	−3.375	−2.436	−1.702	0.250	*
AIC	2060					
Adj. R^2	0.340					
Moran's *I* error	0.021	$p < 0.001$				
2001–2010						
Distance to roads	−0.008	−0.002	−0.001	−0.001	0.001	*
Density of built-up areas	−0.151	−0.049	−0.033	−0.014	0.060	*
Distance to built-up areas	−0.007	−0.002	−0.001	0.000	0.032	*
Density avail. open-/arable land	−0.019	−0.002	0.001	0.004	0.015	*
Altitude	−0.044	−0.006	0.001	0.012	0.112	*
Intercept	−4.662	−1.268	−0.677	0.016	4.176	*
AIC	3081					
Adj. R^2	0.270					
Moran's *I* error	0.001	$p = 0.439$				

* = non-stationary patterns; QT = quartile.

[a] To get a rough idea of the significance of parameter non-stationarity, a comparative approach between global parameters and the interquartiles of local ones can be conducted. If the local coefficient's interquartile range is larger than two standard errors of the global coefficient, this might be an indication of parameter variation.

with spatial proximity as a key factor. In contrast to Hu and Lo (2007), distance to the CBD turns out to be insignificant, implying that the historic role of the CBD is irrelevant for urban development in Mumbai. When compared to the other driving-forces, the density of available open land and arable land has the most pronounced effect, and is significantly related to urban growth. This relationship indicates that built-up areas tend to appear in more undeveloped areas with more available open and arable land in their neighborhoods. Finally, in contrast to the findings of Lin et al. (2011), who reported no significant relationship, in Mumbai altitude has a significant impact on urban growth.

To sum up, both the LR and ALR models produced similar outputs here, with neither sign reversals nor distinct variations in the magnitude of the estimates appearing. Such marginal distinctions are somehow surprising, because the LR residuals show distinct SAC patterns. The ALR successfully reduces SAC and yields better model fits than LR; nevertheless, the limitation of spatially uniform land-use transition processes still remains.

Local modeling results

The final ALR models were re-estimated utilizing a GWLR specification with bi-square kernel functions. To receive an optimal adaptive bandwidth, the AIC-score was minimized. The resulting, estimated GWLR models are, on average, based on local subsamples of 99 data points for each regression. Moran's *I* statistics for the periods 1973–1990 and 1990–2001 indicate marginal residual correlation, although it is a significant one, possibly caused by a lack of socio-economic drivers or missing temporal drivers for those periods. In contrast, the GWLR model for 2001–2010 is well-behaved ($I = 0.001$, $p = 0.439$). Detailed results of the GWLR modeling process are summarized in Table 4 and mapped in Figs. 6 and 7.

Fig. 6a depicts the spatiotemporal relationship between the distance to main roads and spatial expansion. In accordance with the global models, in large parts of Mumbai the distance to roads has the expected negative impact, meaning that expansion occurs near main roads. This result matches, for instance, Hu and Lo's findings (2007). It is noticeable that areas where this covariate shows a positive impact move from the historic center during the period 1973–1990, northwards between 1990 and 2001, and then reduce in size and magnitude during the final period. Furthermore, the first two periods show strong negative impacts on urban expansion, especially in the northern and eastern areas. As a result, the transportation network seems to determine city growth toward the northern parts of Mumbai, and stretches the primary urban areas on both sides of the national park, so that nowadays it is surrounded by built-up areas. This conclusion corresponds to Shafizadeh Moghadam and Helbich (2013).

Comparing Fig. 5 and Fig. 6b, it is evident that urban development occurs where sufficient space is available, at the expense of open land and arable land. From a temporal perspective, this process is most pronounced during the first two periods. For the period 1973–1990, the growth processes occur in the vicinity of the historic center, as well as in three locations in the northern and eastern hinterlands. From 1973 to 1990 some signs refer to a negative impact, while during the subsequent period (1990–2001), all coefficients are positively related to urban expansion, ranging from nearly 0 to approximately 0.02. When compared to the ALR parameter of 0.008, Fig. 6b shows parameter heterogeneity across space. Contrasting with previous developments, the timeframe 2001–2010 shows a noticeable mix of positive and negative signs. The strongest negative impacts can be found in areas which previously experienced a positive impact (e.g., the historic center), and in the growth poles of northern Mumbai. A possible explanation for such a pattern is the presence of more compact in-fill developments.

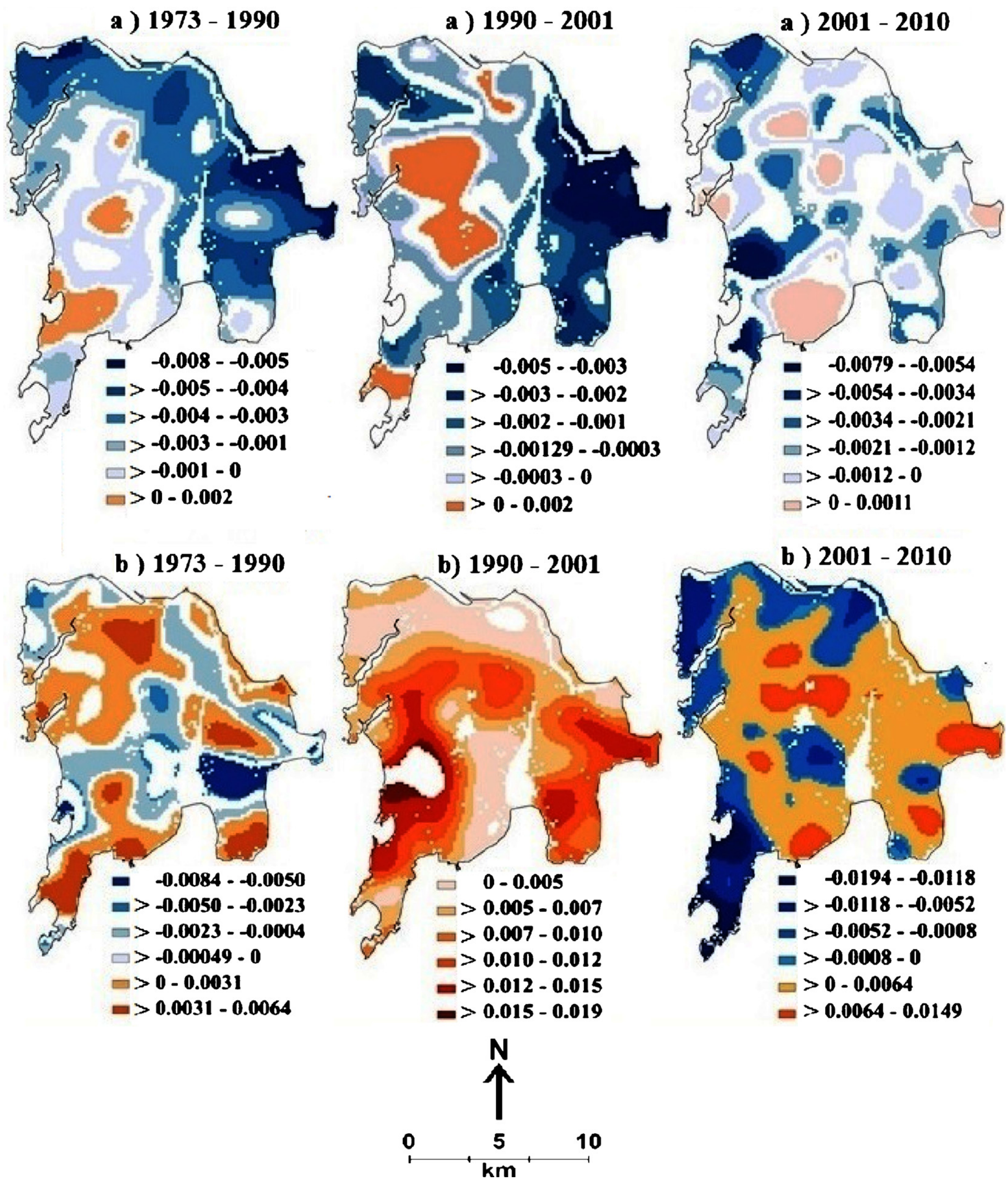

Fig. 6. GWLR parameters (βs) for distance to roads (upper panel-a) and density of open land and arable land (lower panel-b). White areas refer to insignificant areas (5% level) and areas with no data (i.e., water bodies).

Another essential covariate is the density of built-up areas (Fig. 7a). As reported in Luo and Wei (2009), the positive and negative impacts of density on urban growth patterns are distinctive. The ALR shows a sign transition, from a positive impact (1973–1990) to a negative impact over the subsequent periods, which might be due to a change in the proportion of local positive and negative signs. For the period 1973–1990, the negative impacts are noticeable in the south, but then shift eastwards for the period 1990–2001 and later. In addition, areas with a positive influence on urban growth emerge in the center of the region. Comparing these developments with Fig. 4, it is apparent that during the period 2001–2010, the patterns are dominated by negative impacts, meaning that less dense built-up areas are affected by growth processes in the hinterland.

In Mumbai, the distance to built-up areas is negatively related to urban expansion (Fig. 7b), particularly during the 1970s when the pattern is dominated by negative values. Moreover, a decreasing trend from south to north is observable. While between 1973 and 1990 the negative relationship is most pronounced in the historic center, from 1990 to 2001 onwards this area becomes less relevant, with strong negative values emerging in the central part of Mumbai, meaning that areas close to existing built-up areas have a higher likelihood of being converted into built-up areas themselves. Owing to an increase in positive relations, this is only partially true for the period 2001–2010; notably, this covariate has a pronounced positive impact on urban growth in the historic center. Presumably, this is caused by a lack of potential construction areas drawing built-up areas toward the hinterland.

Finally, altitude serves as a growth driver for the periods 1990–2001 and 2001–2010 (Fig. 7c), but not for the initial period. Northern and central areas have positive coefficient signs, meaning that areas of a lower altitude have a higher probability of changing after the period 1990–2001, while during the period 2001–2010, the whole west coast shows an increased probability of urban expansion.

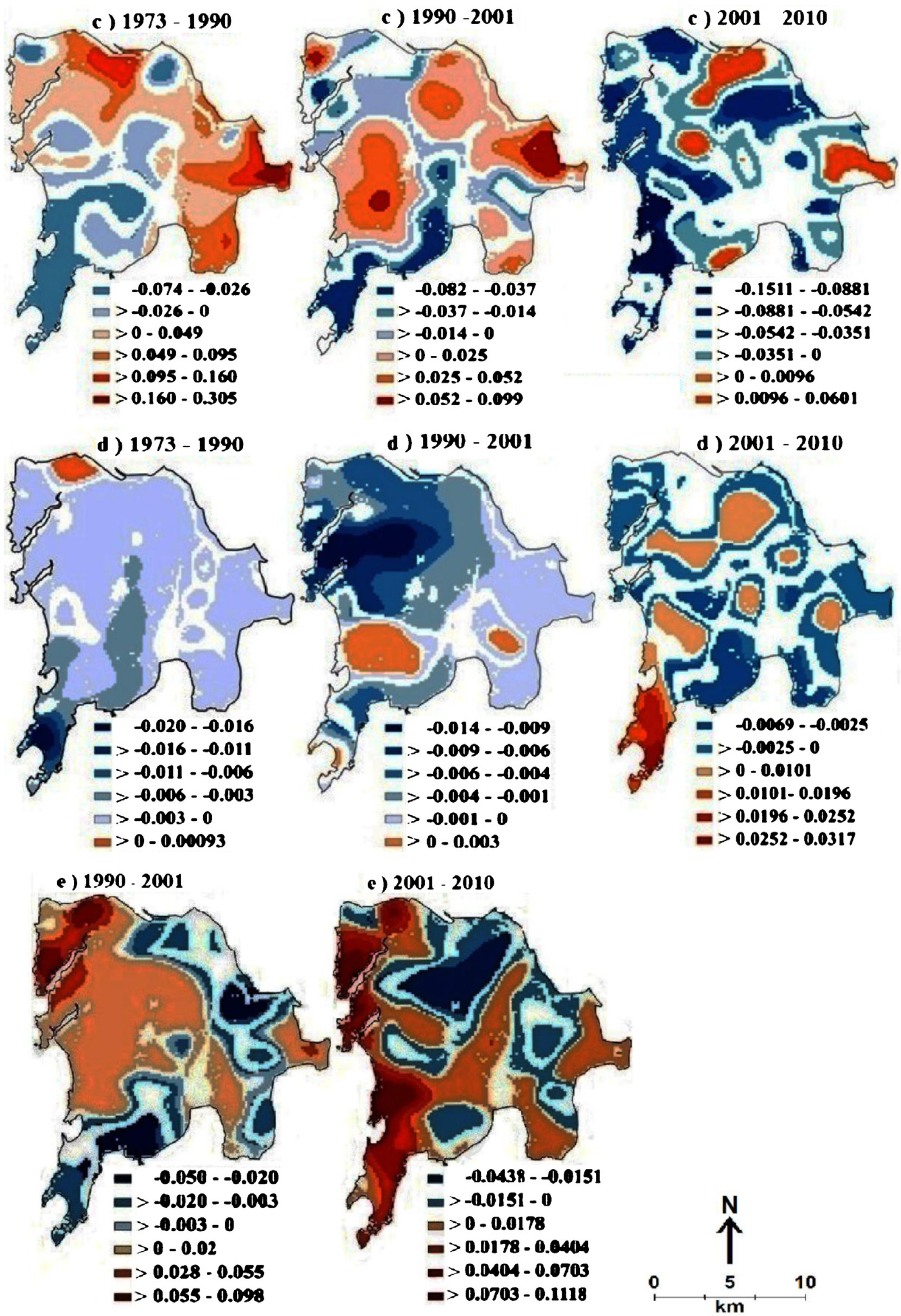

Fig. 7. GWLR parameters (βs) for density of built-up areas (upper panel-c), distance to built-up areas (mid panel-d), and altitude (lower panel-e). White areas refer to insignificant areas (5% level) and areas with no data (i.e., water bodies).

Conclusions

This paper has explored the driving-forces behind urban expansion in Mumbai, India, one of the most vibrant and fastest-growing megacities in the world. While several studies have dealt with the spatiotemporal mapping of urban growth (e.g., Kumar et al., 2011; Taubenböck et al., 2012), no research to date has investigated the underlying driving-forces behind urban development from both a global and local perspective. However, this study fills this gap by utilizing global LR and ALR models, as well as local GWLR.

Although global models are important for planning and decision-making purposes, and have greatly improved our understanding of urban expansion in general, models on a smaller, more detailed scale can support more localized and context-related planning. Previous studies by Poelmans and Van Rompaey (2010), Jokar Arsanjani et al. (2013a) and Munshi et al. (2014), among others, calibrate global LR models but disregard the inherent complexity of land-use data, namely SAC and spatial heterogeneity, which increase the risk of bias and the development of inappropriate urban policies. Although SAC has received some attention in land-use modeling (e.g., Huang et al., 2009; Lin et al., 2011; Chu et al., 2013), spatial heterogeneity has received less attention. As well as the concerns highlighted, it is crucial not to limit any analysis to the latest available time period (Dubovyk et al., 2011). For a more holistic understanding of growth processes, and to facilitate temporal cross-comparisons of growth drivers, several time periods should be investigated. Based on such an approach, this paper has added to the existing knowledge on land-use modeling.

In accordance with previous studies (e.g., Cheng and Masser, 2003; Hu and Lo, 2007), the global LR and ALR show that 5 out of 10 driving-forces, including the distance to roads, the density of built-up areas, the distance to built-up areas, the density and availability of open and arable land, and altitude, are significantly related to urban expansion. While these factors produce the expected signs, of greater importance is the fact that they seem to be largely constant over time on a global scale (as shown for the first time period: 1971–2010). Assuming that this behavior remains constant over the coming years, the results here are important for policymakers, as they will help evaluate and possibly adopt the planning strategies suggested, such as developing growth boundaries. While basic LR shows residual patterns which are considerably reduced by ALR, the coefficient magnitudes exhibit no significant differences. However, the performance of the models favors ALR. As Verburg et al. (2004) pointed out, and as empirically proven by Luo and Wei (2009) later on, our analysis confirms that GWLR further supports our understanding of the complex relationships between urban expansion and the underlying driving-forces. Moreover, and in addition to spatial variation, this research provides evidence that these driving-forces also change their magnitude over time, which agrees with the findings of Ogneva-Himmelberger et al. (2009). Interpreting the results from a planning perspective, it is recommended that policymakers continuously update and adapt their planning strategies to ensure location-based, context-related, and pro-active planning. It is expected that such local models have the potential to improve urban development plans, those recently criticized as being inefficient (Munshi et al., 2014), and so help to reduce the gap between Mumbai's actual and planned urban reality, as caused by recent planning strategies in particular (Pethe et al., 2014).

However, this analysis also has some limitations. Even though detailed insights into land-use transition are provided, the demands placed on decision-makers – to interpret the model outputs appropriately – are high. A potential solution is provided by Helbich et al. (2013), through clustering the parameter surfaces, and so reducing the complexity of the model outputs. Moreover, this study lacks additional socio-economic drivers used over time, a problem already noted by Huang et al. (2009). To produce more comprehensive models, future research should follow Jokar Arsanjani et al. (2013b) and explore whether volunteered geographic information (i.e., OpenStreetMap) permits us to derive reliable variables which describe the urban form, even when official data is lacking.

Acknowledgments

We would like to thank the reviewers and the editor for their informed and constructive comments.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jag.2014.08.013.

References

Aljoufie, M., Zuidgeest, M., Brussel, M., van Vliet, J., van Maarseveen, M., 2013. A cellular automata-based land use and transport interaction model applied to Jeddah, Saudi Arabia. Landsc. Urban Plan. 112, 89–99.
Anselin, L., 2009. Spatial regression. In: Fotheringham, S., Rogerson, P. (Eds.), The SAGE Handbook of Spatial Analysis, Los Angeles. Sage Publications Inc, pp. 255–276.
Atkinson, P., German, S., Sear, D., Clark, M., 2003. Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression. Geogr. Anal. 35, 58–82.
Augustin, N., Mugglestone, M., Buckland, S., 1996. An autologistic model for the spatial distribution of wildlife. J. Appl. Ecol., 339–347.
Basawaraja, R., Chari, K., Mise, S., Chetti, S., 2011. Analysis of the impact of urban sprawl in altering the land use, land-cover pattern of Raichur City, India, using geospatial technologies. J. Geogr. Reg. Plan. 4, 455–462.
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.
Brunsdon, C., Fotheringham, A.S., Charlton, M., 1996. Geographically weighted regression: a method for exploring spatial non-stationarity. Geogr. Anal. 28 (4), 281–298.
Cheng, J., Masser, I., 2003. Urban growth pattern modeling: a case study of Wuhan city, PR China. Landsc. Urban Plan. 62, 199–217.
Cheng, J., 2011. Exploring urban morphology using multi-temporal urban growth data: a case study of Wuhan, China. Asian Geogr. 28, 85–103.
Cho, S., Lambert, D., Kim, S., Jung, S., 2009. Extreme coefficients in geographically weighted regression and their effects on mapping. GIS Remote Sens. 46, 273–288.
Chu, H.-J., Wu, C.-F., Lin, Y.-P., 2013. Incorporating spatial autocorrelation with neural networks in empirical land use change models. Environ. Plan. B: Plan. Des. 40, 384–404.
Cliff, A., Ord, J., 1981. Spatial Processes, Models and Applications. Pion, London.
Dendoncker, N., Rounsevell, M., Bogaert, P., 2007. Spatial analysis and modelling of land use distributions in Belgium. Comput. Environ. Urban Syst. 31, 188–205.
Dormann, C., 2007. Assessing the validity of autologistic regression. Ecol. Model. 207, 234–242.
Dubovyk, O., Sliuzas, R., Flacke, J., 2011. Spatio-temporal modelling of informal settlement development in Sancaktepe district, Istanbul, Turkey. J. Photogram. Remote Sens. 66, 235–246.
Feizizadeh, B., Helali, H., 2010. Comparison pixel-based, object based image analysis and effective parameters in classification land cover/land use of west province Azerbaijan. J. Geogr. Res. 47, 73–84.
Fotheringham, S., Brunsdon, C., Charlton, M., 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley, Los Angeles, SAGE.
Hagenauer, J., Helbich, M., 2012. Mining urban land use patterns from volunteered geographic information by means of genetic algorithms and artificial neural networks. Int. J. Geogr. Inf. Sci. 26, 963–982.
Helbich, M., Amelunxen, C., Neis, P., Zipf, A., 2012. Comparative spatial analysis of positional accuracy of OpenStreetMap and proprietary geodata. In: Jekel, T., Car, A., Strobl, J., Griesebner, G. (Eds.), Geospatial Crossroads @ GI_Forum 2012. Proceedings of the Geoinformatics Forum Salzburg. Wichmann, Heidelberg, pp. 24–33.
Helbich, M., Brunauer, W., Hagenauer, J., Leitner, M., 2013. Data-driven regionalization of housing markets. Ann. Assoc. Am. Geogr. 103, 871–889.
Hu, Z., Lo, C.P., 2007. Modeling urban growth in Atlanta using logistic regression. Comput. Environ. Urban Syst. 31, 667–688.
Huang, B., Zhang, L., Wu, B., 2009. Spatiotemporal analysis of rural–urban land conversion. Int. J. Geogr. Inf. Sci. 23, 379–398.
Jelinski, D., Wu, J., 1996. The modifiable areal unit problem and implications for landscape ecology. Landsc. Ecol. 11, 129–140.
Jokar Arsanjani, J., Helbich, M., Kainz, W., Darvishi, A., 2013a. Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion – the case of Tehran. Int. J. Appl. Earth Obs. Geoinf. 21, 265–275.
Jokar Arsanjani, J., Helbich, M., Bakillah, M., Hagenauer, J., Zipf, A., 2013b. Toward mapping land-use patterns from volunteered geographic information. Int. J. Geogr. Inf. Sci. 27, 2264–2278.
Jokar Arsanjani, A., Helbich, M., Mousivand, A.J., 2014. A morphological approach to predicting urban expansion. Trans. GIS 18, 219–233.
Kumar, A., Pandey, A., Hoda, N., Jeyaseelan, A., 2011. Evaluation of urban sprawl pattern in the tribal-dominated cities of Jharkhand state, India. Int. J. Remote Sens. 32, 7651–7675.
Lin, Y.P., Chu, H., Wu, C., Verburg, P., 2011. Predictive ability of logistic regression, auto-logistic regression and neural network models in empirical land use change modeling – a case study. Int. J. Geogr. Inf. Sci. 25, 65–87.
Luo, J., Wei, Y., 2009. Modeling spatial variations of urban growth patterns in Chinese cities: the case of Nanjing. Landsc. Urban Plan. 91, 51–64.
McCullagh, P., Nelder, J., 1989. Generalized Linear Models. Chapman Hall/CRC, London.

Munshi, T., Zuidgeest, M., Brussel, M., van Maarseveen, M., 2014. Logistic regression and cellular automata-based modelling of retail, commercial and residential development in the city of Ahmedabad, India. Cities 39, 68–86.
Ogneva-Himmelberger, Y., Pearsall, H., Rakshit, R., 2009. Concrete evidence and geographically weighted regression: a regional analysis of wealth and the land cover in Massachusetts. Appl. Geogr. 29, 478–548.
Overmars, K., De Koning, G., Veldkamp, A., 2003. Spatial autocorrelation in multi-scale land use models. Ecol. Model. 164, 257–270.
Pacione, M., 2006. City profile Mumbai. Cities 23, 229–238.
Patino, J., Duque, J., 2013. A review of regional science applications of satellite remote sensing in urban settings. Comput. Environ. Urban Syst. 37, 1–17.
Pethe, A., Nallathiga, R., Gandhi, S., Tandel, V., 2014. Re-thinking urban planning in India: learning from the wedge between the de jure and de facto development in Mumbai. Cities 39, 120–132.
Poelmans, L., Van Rompaey, A., 2010. Complexity and performance of urban expansion models. Comput. Environ. Urban Syst. 34, 17–27.
Pontius, R., Huffaker, D., Denman, K., 2004. Useful techniques of validation for spatially explicit land-change models. Ecol. Model. 179, 445–461.
Punia, M., Singh, L., 2012. Entropy approach for assessment of urban growth: a case study of Jaipur. J. Indian Soc. Remote Sens. 40, 231–244.
ShafizadehMoghadam, H., Helbich, M., 2013. Spatiotemporal urbanization processes in the megacity of Mumbai, India: a Markov chains-cellular automata urban growth model. Appl. Geogr. 40, 140–149.
Taubenböck, H., Esch, T., Felbier, A., Wiesner, M., Roth, A., Dech, A., 2012. Monitoring urbanization in mega cities from space. Remote Sens. Environ. 117, 162–176.
Tayyebi, A., Pijanowski, B., 2014. Modeling multiple land use changes using ANN, CART and MARS: comparing tradeoffs in goodness of fit and explanatory power of data mining tools. Int. J. Appl. Earth Obs. Geoinf. 28, 102–116.
Tv, R., Aithala, B., Sanna, D., 2012. Insights to urban dynamics through landscape spatial pattern analysis. Int. J. Appl. Earth Obs. Geoinf. 18, 329–343.
United Nations, The 2011 Revision, New York 2012. World Urbanization Prospects. United Nations Publication.
Van Ginkel, H., 2008. Urban future. Nature 456, 32–33.
Verburg, P., van Eck, J., de Nijs, T., Dijst, M., Schot, P., 2004. Determinants of land use change patterns in the Netherlands. Environ. Plan. B: Plan. Des. 31, 125–150.
Wheeler, D., Tiefelsdorf, M., 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. J. Geogr. Syst. 7, 161–187.