Published in final edited form as:

Appl Geogr. 2013 October ; 44: 23–32. doi:10.1016/j.apgeog.2013.07.009.

Modelling spatial patterns of urban growth in Africa

Catherine Linard^{a,b,*}, Andrew J. Tatem^{c,d}, and Marius Gilbert^{a,b}

Catherine Linard: linard.catherine@gmail.com; Andrew J. Tatem: andy.tatem@gmail.com; Marius Gilbert: Marius.Gilbert@ulb.ac.be

^aBiological Control and Spatial Ecology, Université Libre de Bruxelles CP 160/12, Av. F.D. Roosevelt 50, B-1050 Brussels, Belgium

^bFonds National de la Recherche Scientifique (F.R.S-FNRS), Rue d'Egmont 5, B-1000 Brussels, Belgium

^cDepartment of Geography and Environment, University of Southampton, Highfield, Southampton, SO17 1BJ, UK

^dFogarty International Center, National Institutes of Health, Bethesda, Maryland, United States of America

Abstract

The population of Africa is predicted to double over the next 40 years, driving exceptionally high urban expansion rates that will induce significant socio-economic, environmental and health changes. In order to prepare for these changes, it is important to better understand urban growth dynamics in Africa and better predict the spatial pattern of rural-urban conversions. Previous work on urban expansion has been carried out at the city level or at the global level with a relatively coarse 5–10 km resolution. The main objective of the present paper was to develop a modelling approach at an intermediate scale in order to identify factors that influence spatial patterns of urban expansion in Africa. Boosted Regression Tree models were developed to predict the spatial pattern of rural-urban conversions in every large African city. Urban change data between circa 1990 and circa 2000 available for 20 large cities across Africa were used as training data. Results showed that the urban land in a 1 km neighbourhood and the accessibility to the city centre were the most influential variables. Results obtained were generally more accurate than results obtained using a distance-based urban expansion model and showed that the spatial pattern of small, compact and fast growing cities were easier to simulate than cities with lower population densities and a lower growth rate. The simulation method developed here will allow the production of spatially detailed urban expansion forecasts for 2020 and 2025 for Africa, data that are increasingly required by global change modellers.

© 2013 The Authors. Published by Elsevier Ltd. All rights reserved.

*Corresponding author: Dr. Catherine Linard, Biological Control and Spatial Ecology, Université Libre de Bruxelles CP 160/12, Av. F.D. Roosevelt 50, B-1050 Brussels, Belgium, Tel: +32 2 650 3781, linard.catherine@gmail.com.

8. Competing Interests

The authors have declared that no competing interests exist.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Africa; urban growth; modelling; spatial pattern; boosted regression trees

---

1. Introduction

According to United Nations forecasts, the population of Africa will double over the next 40 years: from around 1 billion in 2010, Africa is expected to reach 2 billion inhabitants around 2045 (United Nations Population Division, 2010). Much of this anticipated population growth is expected to be concentrated in urban areas, with almost 720 million more people living in cities over the same period (Cohen, 2004; United Nations Population Division, 2011). Such an increase in urban population combined with an overall decrease in average urban population density due to the expansion of periurban areas – which is today no more specific to the developed world but also observed in many cities of the developing world (Angel, 2011) – will lead to higher urban expansion rates in Africa than other regions of the world, with urban land cover growth of more than 12-fold between 2000 and 2050 in sub-Saharan Africa projected in some scenarios (Angel, 2011).

Urbanization in sub-Saharan Africa is characterized by an increasing proportion of poor people living in urban areas (Campbell & Campbell, 2007; Dye, 2008; United Nations Population Fund, 2007). The number of people living under slum conditions doubled in 15 years to reach nearly 200 million people (72% of the urban population) in 2005 in sub-Saharan Africa (United Nations Population Fund, 2007). Even if urban health and well-being have the potential to be much better than that of rural areas – thanks to improved access to healthcare and sanitation – this is usually only true for the urban rich and increases the gap between the urban rich and the urban poor (Dye, 2008). In addition, the poor often settle in marginal land, making them particularly vulnerable to natural disasters such as floods and landslides (Campbell & Campbell, 2007; Hanson et al., 2011; McGranahan et al., 2007). Urban growth exerts pressure on the environment, by reducing the supply of ecosystem services (e.g. agricultural production, freshwater availability or waste absorption) from urbanized regions and increasing the demand for ecosystem services from nearby regions (Alcamo et al., 2005; Eigenbrod et al., 2011; Grimm et al., 2008; McDonald et al., 2011). Urban expansion will also induce global environmental changes, including land use and land cover changes, changes in biogeochemical and hydrological cycles, climate and biodiversity (Grimm et al., 2008; McDonald et al., 2008; Seto et al., 2012). The list of potential changes accompanying urban growth – either positive or negative – is endless and abundantly debated in the literature, but little empirical work has been undertaken to evaluate the costs and benefits of urban growth (Chin, 2002).

The spatial detail and accuracy of urban expansion forecasting methods currently in use must be improved in order to help policy-makers to prepare for future socio-economic, environmental and health changes associated with urban growth and expansion (Montgomery, 2008). The main objective of the present paper is to develop an intermediate approach between high-resolution city-scale studies and low-resolution worldwide-level studies in order to identify, at an intermediate scale, factors that influence spatial patterns of urban expansion at the continental level. By using spatial statistic modelling, we try to

identify factors that commonly drive the spatial pattern of urban expansion in Africa and we evaluate the predictability of the urban spatial pattern only using a parsimonious set of generic factors. Satellite-derived urban change data for 20 large cities across Africa and boosted regression trees (BRT, also called stochastic gradient boosting) models were used to predict the spatial pattern of rural-urban conversions. BRT emerged from extensive research in machine learning and statistics and is one of these relatively new statistical modelling method able to capture complex responses, even with noisy input data (Elith et al., 2006; Friedman, 2002). BRTs were shown very efficient in spatial modelling, e.g. to predict species' spatial distributions based on a set of environmental variables (Elith et al., 2006; Leathwick et al., 2006) or to produce soil predictive maps (Grinand et al., 2008), but to our knowledge was never applied to land cover change modelling. The simulation method developed here will allow the production of spatially detailed urban expansion forecasts for 2020 and 2025 for Africa.

2. Urban growth models

The spatial and temporal dynamics of cities are complex and controlled by various driving forces, such as the regional economy, population movements, the policy environment and socio-cultural processes. However, even though cities have been traditionally seen as places of chaos and disorder requiring control, numerous studies of the last decades have shown that cities are surprisingly structured, with particular spatial patterns resulting from a multitude of decisions and processes in a context of intense competition for space (Batty, 2008). Whereas spatio-temporal urban patterns were traditionally seen from a top-down perspective (i.e. from process to structure), recent studies adopted a new bottom-up approach, which considers cities as complex systems that emerge from individual behaviours. Empirical observations of urban structures and changes over time allowed to quantitatively describe typical urban patterns. Cities became a classic example of fractals (Batty et al., 1989) as they show a hierarchic organization across scales, i.e. from one main center, cities often develop by including a hierarchy of subcenters on their periphery. Urban growth or expansion is generally considered to be a land use change from a non-urban category to an urban category (Batty et al., 1999). To simulate land use changes quantitatively and spatially, cellular automata (CA) have been typically applied (Batty et al., 1999; Batty, 2007; Cheng & Masser, 2004; Santé et al., 2010; White et al., 1997; White & Engelen, 1993). In such models, the landscape is represented by cellular land parcels that can develop into different land use types according to local transition rules. Global patterns of urban growth then emerge from local rural to urban conversions.

The modelling of urban systems has occupied researchers for decades and an abundant literature has emerged. Studies have focused primarily on industrialized countries (e.g. Burchfield et al., 2006; Caruso et al., 2007; Caruso et al., 2005; Dietzel et al., 2005; Hasse & Lathrop, 2003; Kasanko et al., 2006; White & Engelen, 1993, 2000). Only a few studies analyzed and simulated the urban expansion of African cities, including Accra (Yeboah, 2000, 2003), Lagos (Barredo & Demicheli, 2003; Braimoh & Onishi, 2007), Nairobi (Mundia & Aniya, 2005) and Tunis (Weber & Puissant, 2003). It is however the region of the World where the highest urbanization rates are expected in the near future, and the impacts of these changes may be greatest.

Besides the models that focus only on individual cities, some studies have tried to model urban expansion nationally (Eigenbrod et al., 2011; Nowak & Walton, 2005) or even globally (Klein Goldewijk et al., 2010; Seto et al., 2012). The History Database of the Global Environment (HYDE) modelled long-term past land use dynamics and provide maps of population distribution and built-up areas for the period 10 000 BC to AD 2000 at a spatial resolution of 5 minutes (10 km at the equator) (Klein Goldewijk et al., 2010, 2011). Further, Seto et al. (2012) recently simulated global urban expansion to 2030 at a spatial resolution of 5 km using a spatially-explicit land-change model.

The performance of urban expansion models largely depends on the quality and type of data available. Remote sensing data are particularly useful for modelling urban growth because they provide accurate, spatially detailed and consistent urban mapping capabilities at different spatial and temporal scales and allow to link empirical observations with urban theory. The built environment (i.e. non-vegetative, human constructed elements, including buildings, roads, etc.) is easily detectable on satellite imagery of different spatial resolutions (Angel, 2011; Potere et al., 2009; Small, 2005; Tatem, Noor, & Hay, 2005) and is a more precise, consistent and comparable definition of an urban area than notions such as population thresholds or administrative boundaries (Potere et al., 2009; Small, 2005). Here, we used this physical definition to distinguish urban from rural pixels in the vicinity of urban centres of more than 100,000 inhabitants.

3. Data and methods

3.1. Urban change data

The Atlas of Urban Expansion of the Lincoln Institute of Land Policy provides spatial datasets that document urban land cover change between circa 1990 (T_0) and circa 2000 (T_1) for a global sample of 120 large cities of more than 100,000 inhabitants distributed in nine world regions (Angel et al., 2013). Landsat images were acquired for the two time periods and classified into built-up and non-built-up 30×30 meter pixels, using a thematic extraction algorithm (Angel et al., 2005). The accuracy of the Landsat classification was assessed using 10,000 Google Earth validation sites, and results reported user and producer's accuracies of 0.91 and 0.89, respectively (Potere et al., 2009). We acquired the datasets available for the 20 cities located in Africa (8 in Northern Africa and 12 in sub-Saharan Africa) (Table 1).

3.2. Covariates

Several factors influencing the spatial pattern of urban growth have been identified and typically used in existing urban growth models (Santé et al., 2010): (i) the transport network is often identified as the primary factor that shapes urban patterns (Barredo & Demicheli, 2003; Braimoh & Onishi, 2007; Burchfield et al., 2006; Caruso et al., 2005; Li & Yeh, 2000; White & Engelen, 2000), (ii) spatial interactions in the form of neighbourhood indices (e.g. the presence of industrial, commercial or non-urban land in surrounding areas) may influence the attractivity of a pixel by influencing the land rent, the labour market or cultural preferences (Braimoh & Onishi, 2007; Li & Yeh, 2000; White & Engelen, 2000), (iii) policy variables, typically the land-use zoning, are other significant factors as they determine regulations for future land use (Barredo & Demicheli, 2003; Braimoh & Onishi, 2007), and

(iv) the topography may also act as a physical barrier and then influence the urban expansion pattern (Braimoh & Onishi, 2007; Burchfield et al., 2006). A set of spatial datasets representing factors that could influence the spatial pattern of urban expansion in Africa was generated (Table 2). As the objective of the model is to predict urban expansion of every African large city ($n = 373$) at a high resolution, covariates needed to be available – or easily calculated – for the whole of Africa and at a sufficient spatial detail to be tested in the model. The land-use zoning was therefore not taken into account here because such data were not available for most of African cities. Figure 1 shows the set of covariates generated for the city of Tebessa, Algeria.

3.2.1. Accessibility—The accessibility, expressed here as the travel time to the Central Business District (CBD) was calculated using a friction surface produced using a similar method as developed by Nelson for the global map of accessibility (Nelson, 2008). The method was adapted to produce a friction surface (i.e. a gridded surface where the value of each pixel corresponds to the time required to cross that pixel) for each city at a spatial resolution of 100 m. We used the road and railway networks from the US National Imagery and Mapping Agency's (NIMA) Vector Map Level 0 (VMap0) dataset and land cover data from GlobCover (Arino et al., 2008). GlobCover land cover data were aggregated to 10 classes and refined with detailed settlement extents, as described in Linard et al. (2010). Travel speeds were attributed to each 100m pixel following values presented in Table 3 (Nelson, 2008). Accessibility values were scaled and centred for each city, in order to have zero mean and a standard deviation of 1. The CBD location points provided with the Atlas of Urban Expansion were used (Angel et al., 2013).

3.2.2. Urban development in the neighbourhood—Neighbourhood indices were calculated in order to account for the spatial dependence in urban development. The spatial dependence (i.e. the relationship between what happens at one point in space and what happens at a neighbouring point) may vary with the neighbourhood size and therefore three different buffer sizes were used here. The proportion of urbanized land within buffers of 150, 1000 and 5000 meters were computed using Landsat-derived urban extents in T_0 . In addition, we computed the distance to the nearest urban pixel.

3.2.3. Slope—A digital elevation model (DEM) at a spatial resolution of 90 m was downloaded from the CGIAR-CSI website (<http://www.cgiar-csi.org/>). This dataset is an improved version of the global DEM acquired by the Shuttle Radar Topography Mission (SRTM) and distributed by NASA/USGS in which the “no-data” holes of the original dataset are filled using interpolation methods (Jarvis et al., 2008). The slope gradient was derived from this DEM.

3.3. Urban expansion model

The urban expansion model developed here involves two steps. In a first step, Boosted Regression Trees (BRT) were used to predict the rural-urban conversion probability for each 100 m pixel within city study areas. We decided to use BRT as the modelling technique, which combine the strengths of two algorithms: regression trees and boosting (Elith et al., 2008). Whereas conventional regression techniques try to fit one parsimonious model that

best describes the relationship between a response variable and a set of covariates, the boosting algorithm allows the fitting a large number of relatively simple models whose predictions are then combined to produce more robust estimates. It was shown that the predictive performance of BRT was superior to conventional regression techniques (Elith et al., 2006; Leathwick et al., 2006), mainly because of their capability for fitting interactions among predictor variables. BRT present several other advantages, the most useful ones being that (i) they can fit complex non-linear relationships, (ii) there is no need for prior data transformation and (iii) they support different types of predictor variables.

BRT models were developed using the R package 'gbm' version 1.6–3.2 (Ridgeway, 2007) and the additional functions provided by Elith et al. (Elith et al., 2008). In a second step, pixels with the highest conversion probability were converted from rural to urban in order to produce a predicted urban extent map. All analyses have been performed using the R statistical program version 2.15 (R Development Core Team, <http://www.r-project.org>).

3.3.1. Probability raster—Urban expansion changes between circa 1990 and circa 2000 for 20 African large cities (Table 1) were modelled against the 6 predictor variables presented in Table 2. The study area of each city was first defined by drawing a two kilometre buffer around the main city extent. All datasets were resampled and cropped in order to match the extent of the defined study area and the spatial resolution of 100 metres. The dependent variable follows a Bernoulli distribution (i.e. value 1 for new urban development and value 0 for no development). Cells that were already urban in 1990 were removed from the database. For each city, we developed a BRT model based on data from the 19 other cities and produced a predicted raster showing the probability of rural to urban conversion between 1990 and 2000.

The value of three parameters needs to be estimated in order to optimize BRT predictive performance: the learning rate (lr), the tree complexity (tc) and the number of trees (nt). lr determines the contribution of each tree to the final model, tc is the number of tree nodes and represents the level of interactions between variables that are fitted and nt is the total number of trees incorporated in the final model. We determined values for lr and tc that optimize predictive performance while keeping a manageable computing time. The optimal number of trees nt was then determined by the function *gbm.step* (Elith et al., 2008). Given the different number of observations available for each city, we attributed a weight to each city equal to 1 divided by the city size in order to make them contributing equally to the BRT model. We also forced a minimum of 20% of positives (new development cells) compared to negatives (no-development cells) by randomly sampling the negatives. The contribution of each predictive variable in the final BRT model was measured based on the number of times a variable was selected for splitting, weighted by the improvement of the model following each split (Friedman & Meulman, 2003). The relationships between each predictor variables and the response (the probability of rural-urban conversion) were analyzed using partial dependence plots.

3.3.2. Predicted urban extent—In the previous step, a probability of rural-urban conversion was assigned to each cell. This probability raster was then used to decide which rural cells change their state and become urban between T_0 (circa 1990) and T_1 (circa 2000).

The urban land use demand (or city growth) is generated exogenously, as the objective here is to predict the spatial location of rural-urban conversions, and not the rate of change. For each city, we calculated the number of cells that were converted from rural to urban between T_0 and T_1 based on Landsat-derived urban extents (Angel et al., 2013). We then produced a predicted urban extent map for 2000 by selecting the same number of cells with the highest probability of rural-urban conversion from the probability raster. New urban development was restricted to realistic locations, by setting a null conversion probability to national parks and wetlands.

3.3.3. Accuracy assessment—Predicted urban extent maps were produced based on data from 19 independent cities. Predicted and observed urban expansions were then compared for each of the 20 cities. The predictive power of the model was measured using the area under the curve (AUC) and Cohen's kappa. We compared the predictive power of the BRT model with the predictive power of a uniform urban expansion model (hereafter 'distance-based model') in order to evaluate the added-value of the BRT model described here. The distance-based model was built by assigning a conversion probability inversely proportional to the distance to the nearest urban pixel, resulting in a simple city expansion around existing urban pixels. The urban growth was determined exogenously, as for the BRT model. If there were more cells with the highest probability level than the number of cell conversions needed, rural-urban conversions were randomly selected among the cells with the highest probability.

3.4. Predictive power and city characteristics

Given the high heterogeneity in terms of predictive power observed between the 20 cities used as training data, we carried out a Principal Component Analysis (PCA) in order to classify the 20 cities into different clusters and identify the factors that could influence the predictive power. The PCA was carried out using four variables: the urban area in 1990, the average urban population density in 1990 (Angel et al., 2013), the average national urban population growth rate between 1990 and 2000 as estimated by the UN (United Nations Population Division, 2011) and a spatial cohesion index. The cohesion index describes the compactness of the city and is calculated as the ratio of the average distance among all points in an equal-area circle and the average distance among all urban cells of the city (Angel, 2011). A cohesion index close to 1 indicates a city shape close to the circle.

4. Results

4.1. Urban expansion model

We produced 20 different BRT models – one for each of the 20 cities – based on the 19 other cities. BRT parameters were fixed to the following values: $tc = 1$ and $lr = 0.1$. Variations in predictive variable contributions were limited, with the proportion of urban cells in a 1 km neighbourhood and the accessibility always being the most influential variables (Figure 2). The slope and the proportion of urban cells in the close neighbourhood were the least important variables, with contributions always below 3%. The 20 fitted models had a number of trees between 4,300 and 5,250 and an average percentage of explained deviance of $35.66 (\pm 1.90)$.

Partial dependence plots show the relationships between each predictor variable and the response (the probability of rural-urban conversion), after accounting for the average effect of all other variables (Elith et al., 2008) (Figure 3). The partial responses indicate an urban expansion occurring in areas where the percentage urbanized in a 1 km vicinity is higher, in more accessible areas (or low travel time to CBD), and in areas where the distance to the nearest urban pixel is low. The probability of rural-urban conversion is more important in areas where the percentage urbanized in a 5 km vicinity is between 0.2 and 0.3 but decreases where this percentage increases beyond 0.3.

Rural-urban conversion probability values were then used to predict the spatial pattern of urban expansion between 1990 and 2000 for each city (Figure 4). Table 4 and Figure 5 show comparisons of the accuracy assessment statistics (AUC and Cohen's kappa) obtained by the BRT models presented here and the distance-based models. Results indicate that the predictive power of BRT models was significantly higher than the predictive power of distance-based models, with paired Student's *t*-test confirming the hypothesis of improvement for both AUC ($t = -3.759$, $df = 19$, $p\text{-value} < 0.001$) and Cohen's kappa ($t = -3.759$, $df = 19$, $p\text{-value} < 0.001$). For 4 cities (Aswan, Marrakech, Port Sudan, Tebessa), the Cohen's kappa increased by more than 0.1 between the distance-based and the BRT model, showing that the BRT model considerably improved the prediction in these cases. For 5 cities (Algiers, Cairo, Harare, Kigali and Ouagadougou) there was no improvement in the predictions between the distance-based and the BRT model, though predictive powers were very close (differences in Cohen's kappa below 0.03). Figure 5 also shows an important variability in predictive power among cities. The spatial pattern of urban growth was well predicted for some cities (e.g. Ibadan, Kigali, Ouagadougou, Tebessa), with Cohen's kappa greater than 0.4, and very poorly estimated for other cities (e.g. Harare, Johannesburg, Ndola, Pretoria), with Cohen's kappa below 0.1. Predicted maps for the 20 cities can be found in Supplementary Figure SI-1.

4.2. Predictive power and city characteristics

Given the large heterogeneities observed in predictive power, we did some supplementary analyses in order to determine if a low/high predictive power could be due to particular city characteristics. PCA and cluster analysis results are presented in Table 5 and Figure 6. Table 5 shows that 2 components of the PCA explained 66% of the variance. The second component, which mainly distinguishes small cities characterized by a high urban growth rate from larger cities characterized by a low urban growth rate, was significantly correlated to the Cohen's kappa, with a Pearson correlation coefficient of 0.63 ($p\text{-value} = 0.003$, $n = 20$), i.e. the BRT models better predict the urban growth pattern of relatively small, compact and fast growing cities. The other components were not significantly correlated to the Cohen's kappa. Figure 6 suggests that Kigali is an outlier in the plots, probably mainly due to its extremely high urban growth rate of more than 10% per year. The predictive power was very good for Kigali, using both the BRT and the distance-based models (with Cohen's kappa above 0.43). At the other extreme in terms of predictive power, Johannesburg appears as a second outlier with its large urban area and Cohen's kappa of 0.023 for the BRT and even negative for the distance-based model. These two cities are represented in Figure 4, along with Banjul and Tebessa, which were classified in a different cluster (Figure 6).

5. Discussion

The national proportions of urban residents is steadily increasing and each of the major developing regions is expected to become majority urban by 2030 (United Nations Population Division, 2011). Globally, the largest urban expansion rates over the coming decades are expected to be observed in sub-Saharan Africa, with a predicted 12-fold increase in urban land cover surface between 2000 and 2050 according to some current projections (Angel, 2011). This huge demand for urban land cover will probably induce important rates of spontaneous and unplanned expansion, as is already observed today in large and rapidly growing cities (Barredo & Demicheli, 2003), which is generally associated with poverty, land degradation and lack of service capacity (United Nations Population Fund, 2007). Spatial and quantitative projections of urban expansion are important for providing guidance to policy-makers for assessing the potential impact of urban expansion on well-being and environment and for supporting development strategies (Montgomery, 2008; Seto et al., 2012).

Results obtained using our simulation model were generally more accurate than results obtained using a distance-based urban expansion model. However, the predictive power was fairly low and varied considerably according to the city (Cohen's kappa values ranging from 0.023 to 0.506). The 20 cities used as training data are distributed over the whole of Africa and are very heterogeneous in terms of size, spatial pattern, economic level and geographical environment. Predicting the urban expansion of one city based on data from only 19 independent cities is therefore a tricky task, as all heterogeneities cannot be taken into account in one statistical model. Our results showed that the spatial pattern of small, compact and fast growing cities were easier to predict and simulate. These cities are probably in the second phase of urban growth described by Chin (2002), i.e. a phase of suburbanization with fastest growth just outside of the city core. According to Leao et al. (2004), the current urban growth observed in developing countries fits better in this second phase because of the limited mobility and lower standard of living of their populations. On the contrary, our results show that the spatial growth of a large city with lower population density and lower growth rate (e.g. Johannesburg and Pretoria) was more difficult to predict. The spatial pattern of urban growth of these cities appears to be more similar to the more fragmented and sparse patterns typically observed in developed countries (Leao et al., 2004), which corresponds to a phase of counter urbanization following technological and economic progress, with population from the core and suburbs moving out to more rural areas (Chin, 2002; Leao et al., 2004). Such fragmented urban patterns are driven by an improvement in transport allowing the development of industries, shops and housing outside the main city centre and transforming monocentric cities to multi centred scattered urban areas (Chin, 2002; Leao et al., 2004). This suggests that African cities are in different urbanization phases and that future work is needed to develop models more adapted to the different phases. The scarcity of urban expansion data currently available for African cities prevented us here from developing different models.

A second factor that could explain the low predictive power is the temporal dependence of the urban growth process, defined as the relationship between what happens at one time period and what happens in the following time period. In urban growth, initial urban

development – which results from human behaviours that are inevitably partly subjective and stochastic – has an impact on future development, and this adds important complexity to the modelling of urban growth. Even if cellular automata-like models were shown to effectively deal with spatial and temporal interactions (Cheng & Masser, 2004), it is not surprising that modelling urban growth over a period of 10 years results in relatively low predictive power. Higher prediction accuracies could probably be obtained if the training set was made of urban land cover data separated by fewer years.

When modelling urban expansion at a high resolution and for a large area such as continental Africa, we inevitably face important challenges. A first challenge is the urban definition that varies according to countries. Even if remote sensing techniques simplified the problem of comparability and consistency between countries through defining urban areas as built surfaces, discrepancies remain in the definition of urban populations. The classification of populations as urban or rural depends on the country-specific definition used in censuses by national statistical offices (United Nations Population Division, 2011). A population is generally defined as urban when located in an urban locality. However, in an urban locality, all the pixels are not necessarily classified as built-up and can include rural areas according to a remotely-sensed definition. On the contrary, built-up areas can expand beyond the administratively defined city, especially in an urban growth context. Here we have linked urban population counts to built-up areas, which inevitably leads to uncertainties, as both are defined differently. In addition, urbanization is represented here as a dichotomous variable (urban or rural), but in reality, different degrees of urbanization can be observed. Some authors have used more than two land use classes (Barredo & Demicheli, 2003) or continuous variables such as percent built-up (Klein Goldewijk et al., 2010; Seto et al., 2012) or population density (Wu, 1998) in order to include different urbanization degrees. Here we used a dichotomous variable for model simplicity reasons and because the high spatial resolution used (100 meters) limits the heterogeneities that can be observed in one single pixel, compared to datasets with spatial resolutions of several kilometres.

The complexity of cities is another important challenge when modelling urban expansion, and particularly when trying to generalize the process of urban growth at the continental level. African cities are very complex structures emerging from a diversity of political, economic, social and cultural processes. The objective of the present study was not to take all these heterogeneities into account, but to find a way to simulate urban expansion at the continental level using a parsimonious set of variables that are available for the whole continent. All city-specific characteristics such as city planning policies, population movements or economic factors were therefore not considered here.

6. Conclusions

The objective of this work was to explore the possibilities for the production of a simple and generalizable urban expansion model that could be applied to every African large city. Future work will use a version of the model presented here to simulate the urban expansion of every large African city for 2020 and 2025 and to produce projected population distribution datasets under a range of growth scenarios following AfriPop methods (Linard et al., 2012). These upcoming datasets will be more spatially detailed and likely more

accurate than existing large area urban expansion datasets (Klein Goldewijk et al., 2010; Seto et al., 2012) in terms of spatial resolution (100 m vs. 5 or 10 km) and input data, as the datasets will be based on detailed Landsat-derived settlement extent data from AfriPop (Linard et al., 2012; Tatem et al., 2005, 2007). Despite the significant uncertainties in urban pattern predictions, the urban expansion datasets presented here can be seen as realistic scenarios for the future, and the most detailed Africa-wide urban expansion predictions ever undertaken. Such spatial datasets are increasingly in demand by global change modellers, but are currently lacking.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

CL and MG are supported by the Fonds National de la Recherche Scientifique (F.R.S./FNRS). AJT is supported by funding from NIH/NIAID (U19AI089674), the Bill & Melinda Gates Foundation (49446, 1032350), and the RAPIDD program of the Science and Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health. This work forms part of the output of the AfriPop Project (www.afripop.org), principally funded by the Fondation Philippe Wiener - Maurice Anspach, and the Malaria Atlas Project (MAP, www.map.ox.ac.uk), principally funded by the Wellcome Trust, U.K. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Alcamo J, van Vuuren D, Ringler C, Cramer W, Masui T, Alder J, Schulze K. Changes in Nature's Balance Sheet: Model-based Estimates of Future Worldwide Ecosystem Services. *Ecology and Society*. 2005; 10(2):19.
- Angel, S. Making room for a planet of cities. Cambridge, MA: Lincoln Institute of Land Policy; 2011.
- Angel, S.; Parent, J.; Civco, DL.; Blei, AM. Atlas of Urban Expansion. Cambridge MA, USA: Lincoln Institute of Land Policy; 2013. Retrieved from <http://www.lincolnst.edu/subcenters/atlas-urban-expansion/>
- Angel, S.; Sheppard, SC.; Civco, DL.; Buckley, R.; Chabaeva, A.; Gitlin, L.; Krale, A.; Parent, J.; Perlin, M. The dynamics of global urban expansion. Transport and Urban Development Department, The World Bank; 2005.
- Arino, O.; Bicheron, P.; Achard, F.; Latham, J.; Witt, R.; Weber, JL. ESA bulletin. Bulletin ASE. European Space Agency; 2008. GLOBCOVER: The most detailed portrait of Earth; p. 25
- Barredo JJ, Demicheli L. Urban sustainability in developing countries' megacities: modelling and predicting future urban growth in Lagos. *Cities*. 2003; 20(5):297–310.
- Batty, M. Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals. Cambridge, Massachusetts: The MIT press; 2007.
- Batty M. The size, scale, and shape of cities. *Science*. 2008; 319(5864):769. [PubMed: 18258906]
- Batty M, Longley P, Fotheringham S. Urban growth and form: scaling, fractal geometry, and diffusion-limited aggregation. *Environment and Planning A*. 1989; 21(11):1447–1472.
- Batty M, Xie Y, Sun Z. Modeling urban dynamics through GIS-based cellular automata. *Computers, Environment and Urban Systems*. 1999; 23(3):205–233.
- Braimah AK, Onishi T. Spatial determinants of urban land use change in Lagos, Nigeria. *Land Use Policy*. 2007; 24(2):502–515.
- Burchfield M, Overman HG, Puga D, Turner MA. Causes of Sprawl: A Portrait from Space. *The Quarterly Journal of Economics*. 2006; 121(2):587–633.
- Campbell T, Campbell A. Emerging Disease Burdens and the Poor in Cities of the Developing World. *Journal of Urban Health*. 2007; 84(S1):54–64.

- Caruso G, Peeters D, Cavailhès J, Rounsevell M. Spatial configurations in a periurban city. A cellular automata-based microeconomic model. *Regional Science and Urban Economics*. 2007; 37(5):542–567.
- Caruso G, Rounsevell M, Cojocaru G. Exploring a spatio-dynamic neighbourhood-based model of residential behaviour in the Brussels periurban area. *International Journal of Geographical Information Science*. 2005; 19(2):103–123.
- Cheng J, Masser I. Understanding spatial and temporal processes of urban growth: Cellular automata modelling. *Environment and Planning B: Planning and Design*. 2004; 31(2):167–194.
- Chin, N. Unearthing the roots of urban sprawl: a critical analysis of form, function and methodology. Working/discussion paper. 2002. Retrieved 12 February 2013, from http://www.casa.ucl.ac.uk/working_papers/paper47.pdf
- Cohen B. Urban growth in developing countries: a review of current trends and a caution regarding existing forecasts. *World Development*. 2004; 32(1):23–51.
- Dietzel C, Herold M, Hemphill JJ, Clarke KC. Spatio-temporal dynamics in California's Central Valley: Empirical links to urban theory. *International Journal of Geographical Information Science*. 2005; 19(2):175–195.
- Dye C. Health and urban living. *Science*. 2008; 319(5864):766. [PubMed: 18258905]
- Eigenbrod F, Bell VA, Davies HN, Heinemeyer A, Armsworth PR, Gaston KJ. The impact of projected increases in urbanization on ecosystem services. *Proceedings. Biological sciences/The Royal Society*. 2011; 278(1722):3201–3208. [PubMed: 21389035]
- Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. *Journal of Animal Ecology*. 2008; 77(4):802–813. [PubMed: 18397250]
- Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*. 2006; 29(2):129–151.
- Friedman JH. Stochastic gradient boosting. *Computational Statistics & Data Analysis*. 2002; 38(4):367–378.
- Friedman JH, Meulman JJ. Multiple additive regression trees with application in epidemiology. *Statistics in Medicine*. 2003; 22(9):1365–1381. [PubMed: 12704603]
- Grimm NB, Faeth SH, Golubiewski NE, Redman CL, Wu J, Bai X, Briggs JM. Global Change and the Ecology of Cities. *Science*. 2008; 319(5864):756–760. [PubMed: 18258902]
- Grinand C, Arrouays D, Laroche B, Martin MP. Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. *Geoderma*. 2008; 143(1–2):180–190.
- Hanson S, Nicholls R, Ranger N, Hallegatte S, Corfee-Morlot J, Herweijer C, Chateau J. A global ranking of port cities with high exposure to climate extremes. *Climatic Change*. 2011; 104(1):89–111.
- Hasse JE, Lathrop RG. Land resource impact indicators of urban sprawl. *Applied Geography*. 2003; 23(2–3):159–175.
- Jarvis, A.; Reuter, HI.; Nelson, A.; Guevara, E. Hole-filled SRTM for the globe Version 4. 2008. available from the CGIAR-CSI SRTM 90m Database (<http://srtm.csi.cgiar.org>)
- Kasanko M, Barredo JJ, Lavalle C, McCormick N, Demicheli L, Sagris V, Brezger A. Are European cities becoming dispersed?: A comparative analysis of 15 European urban areas. *Landscape and Urban Planning*. 2006; 77(1–2):111–130.
- Klein Goldewijk K, Beusen A, Janssen P. Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1. The Holocene. 2010; 20(4):565–573.
- Klein Goldewijk K, Beusen A, Van Drecht G, De Vos M. The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Global Ecology and Biogeography*. 2011; 20(1):73–86.
- Leao S, Bishop I, Evans D. Simulating Urban Growth in a Developing Nation's Region Using a Cellular Automata-Based Model. *Journal of Urban Planning and Development*. 2004; 130(3):145–158.

- Leathwick JR, Elith J, Francis MP, Hastie T, Taylor P. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series*. 2006; 321:267–281.
- Li X, Yeh AGO. Modelling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science*. 2000; 14(2):131–152.
- Linard C, Gilbert M, Snow RW, Noor AM, Tatem AJ. Population Distribution, Settlement Patterns and Accessibility across Africa in 2010. *PLoS ONE*. 2012; 7(2):e31743. [PubMed: 22363717]
- Linard C, Gilbert M, Tatem AJ. Assessing the use of global land cover data for guiding large area population distribution modelling. *GeoJournal*. 2010; 76:525–538. [PubMed: 23576839]
- McDonald RI, Green P, Balk D, Fekete BM, Revenga C, Todd M, Montgomery M. Urban growth, climate change, and freshwater availability. *Proceedings of the National Academy of Sciences*. 2011; 108(15):6312–6317.
- McDonald RI, Kareiva P, Forman RTT. The implications of current and future urbanization for global protected areas and biodiversity conservation. *Biological Conservation*. 2008; 141(6):1695–1703.
- McGranahan G, Balk D, Anderson B. The rising tide: assessing the risks of climate change and human settlements in low elevation coastal zones. *Environment and Urbanization*. 2007; 19(1):17.
- Montgomery MR. The urban transformation of the developing world. *Science*. 2008; 319(5864):761. [PubMed: 18258903]
- Mundia CN, Aniya M. Analysis of land use/cover changes and urban expansion of Nairobi city using remote sensing and GIS. *International Journal of Remote Sensing*. 2005; 26(13):2831–2849.
- Nelson, A. Estimated travel time to the nearest city of 50,000 or more people in year 2000. Ispra, Italy: Global Environment Monitoring Unit - Joint Research Centre of the European Commission; 2008. Retrieved from <http://bioval.jrc.ec.europa.eu/products/gam>
- Nowak DJ, Walton JT. Projected Urban Growth (2000–2050) and Its Estimated Impact on the US Forest Resource. *Journal of Forestry*. 2005; 103(8):383–389.
- Potere D, Schneider A, Angel S, Civco D. Mapping urban areas on a global scale: which of the eight maps now available is more accurate? *International Journal of Remote Sensing*. 2009; 30(24): 6531–6558.
- Ridgeway, G. Generalized Boosted Models: A guide to the gbm package. 2007.
- Santé I, García AM, Miranda D, Crecente R. Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landscape and Urban Planning*. 2010; 96(2):108–122.
- Seto KC, Güneralp B, Hutyrá LR. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences*. 2012; 109(10):10731–10736. [PubMed: 22116581]
- Small C. A global analysis of urban reflectance. *International Journal of Remote Sensing*. 2005; 26(4): 661–681.
- Tatem AJ, Noor AM, Hay SI. Assessing the accuracy of satellite derived global and national urban maps in Kenya. *Remote Sensing of Environment*. 2005; 96(1):87–97. [PubMed: 22581985]
- Tatem AJ, Noor AM, von Hagen C, Di Gregorio A, Hay SI. High resolution population maps for low income nations: combining land cover and census in East Africa. *PLoS One*. 2007; 2(12):e1298. [PubMed: 18074022]
- United Nations Population Division. World Population Prospects: The 2010 Revision. New York: United Nations; 2010. Retrieved from <http://esa.un.org/wpp/>
- United Nations Population Division. World Urbanization Prospects: The 2011 Revision. New York: United Nations; 2011. Retrieved from <http://esa.un.org/unup/>
- United Nations Population Fund. State of world population 2007: unleashing the potential of urban growth. New York: United Nations Population Fund; 2007.
- Weber C, Puissant A. Urbanization pressure and modeling of urban growth: example of the Tunis Metropolitan Area. *Remote Sensing of Environment*. 2003; 86(3):341–352.
- White &, Engelen G. High-resolution integrated modelling of the spatial dynamics of urban and regional systems. *Computers Environment and Urban Systems*. 2000; 24(5):383–400.

- White R, Engelen G. Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. *Environment and Planning A*. 1993; 25(8):1175–1199.
- White R, Engelen G, Uljee I. The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B: Planning and Design*. 1997; 24(3):323–343.
- Wu F. An experiment on the generic polycentricity of urban growth in a cellular automatic city. *Environment and Planning B: Planning and Design*. 1998; 25(5):731–752.
- Yeboah IEA. Structural Adjustment and Emerging Urban Form in Accra, Ghana. *Africa Today*. 2000; 47(2):61.
- Yeboah IEA. Demographic and Housing Aspects of Structural Adjustment and Emerging Urban Form in Accra, Ghana. *Africa Today*. 2003; 50(1):107–119.

Highlights

- The spatial pattern of rural-urban conversions in African cities was simulated.
- Boosted regression trees performed better than distance-based urban growth models.
- Urban expansion of small, compact and fast growing cities was easier to predict.
- The 1 km neighbourhood and the accessibility were the most influential variables.
- The method allows the production of spatially detailed urban expansion forecasts, with low to moderate accuracy levels.

Fig. 1. Spatial datasets used as covariates for modelling the spatial pattern of urban growth showed for the city of Tebessa, Algeria. The black line shows the urban extent in T₀ (circa 1990). For an explanation of variables and units, see Table 2.

Fig. 2.

Summary of the relative contributions (%) of predictor variables for 20 BRT models developed with cross-validation. For an explanation of variables see Table 2.

Fig. 3.

Partial dependence plots for the six variables included in the models for rural-urban conversion (the most influential being ranked first). Plots from 20 different models were overlaid. For an explanation of variables see Table 2. Y axes are on the logit scale and are centred to have zero mean over the data distribution. Dashes at inside top of plots show distribution of observations (cells) across that variable, in deciles.

Fig. 4.

Examples of urban expansion pattern for 4 cities classified in 4 different clusters:

Johannesburg (South Africa), Kigali (Rwanda), Banjul (Gambia) and Tebessa (Algeria). The figure shows urban change between T_0 (circa 1990) and T_1 (circa 2000) as observed in reality, predicted by the BRT and predicted by the distance-based model. Maroon = urban in T_0 ; red = urban in T_1 ; blue = water; grey = rural; K = Cohen's kappa index.

Fig. 5.

Boxplots of the accuracy assessment statistics (AUC and Cohen's kappa) obtained using distance-based models and BRT models.

Fig. 6.

PCA and cluster analysis results. A. Biplot showing the position of each city related to variables used in the PCA. B. Cluster dendrogram resulting from the PCA. Abbreviation of city names used in A are shown in Table 1.

Table 1

20 large cities used for modelling urban growth and their evolution between T_0 (circa 1990) and T_1 (circa 2000), following the Atlas of urban expansion (Angel et al., 2013)

City name	Abbreviation	Country	T_0	T_1	Population		Built-up area (ha)		Population growth (%)	Urban area growth (%)
					T_0	T_1	T_0	T_1		
Accra	Accr	Ghana	1985	2000	1,818,178	2,693,371	12,973	32,834	48.14	153.10
Addis Ababa	AddA	Ethiopia	1986	2000	1,653,379	2,510,904	8,193	11,865	51.86	44.82
Alexandria	Alxn	Egypt	1984	1999	3,042,907	3,378,392	11,342	18,780	11.03	65.58
Algiers	Algr	Algeria	1987	2000	2,671,427	3,627,912	13,946	22,913	35.80	64.29
Aswan	Aswn	Egypt	1986	2000	225,969	310,069	1,260	1,579	37.22	25.31
Bamako	Bamk	Mali	1986	1999	829,438	1,239,935	6,884	12,992	49.49	88.73
Banjul	Bnjl	Gambia	1986	2000	235,692	447,985	3,124	4,949	90.07	58.44
Cairo	Cair	Egypt	1984	2000	10,161,703	13,083,621	36,605	56,917	28.75	55.49
Casablanca	Csbl	Morocco	1987	2001	2,291,087	3,004,505	8,054	11,431	31.14	41.92
Harare	Harr	Zimbabwe	1989	2000	838,775	889,421	16,506	24,310	6.04	47.27
Ibadan	Ibdn	Nigeria	1984	2000	1,565,805	2,421,369	20,921	29,996	54.64	43.38
Johannesburg	Jhnn	South Africa	1991	2000	3,521,614	4,695,165	87,163	99,279	33.32	13.90
Kampala	Kmpl	Uganda	1995	2001	1,314,603	1,761,733	13,156	20,169	34.01	53.30
Kigali	Kigl	Rwanda	1984	1999	296,879	354,273	1,517	4,502	19.33	196.78
Marrakech	Mrrk	Morocco	1987	2000	1,328,537	1,722,999	9,706	16,475	29.69	69.74
Ndola	Ndol	Zambia	1989	2002	350,512	326,119	4,133	5,390	-6.96	30.41
Ougadougou	Ogdg	Burkina Faso	1986	2001	615,293	874,623	6,027	13,746	42.15	128.08
Port Sudan	Prts	Sudan	1984	2001	155,821	209,154	2,673	4,104	34.23	53.55
Pretoria	Pptr	South Africa	1991	2000	432,031	659,834	15,011	17,004	52.73	13.28
Tebessa	Tbss	Algeria	1987	2001	366,628	457,364	5,091	7,537	24.75	48.04

Table 2

Covariates used for modelling urban growth in Africa

Variable	Description	Resolution	Data source
ACCESS	Travel time to CBD (scaled and centered)	3 arc-seconds (~100 m)	Refined GlobCover, road and railway networks from VMAP0; method adapted from (Nelson, 2008)
N150	Proportion of urban pixels in a 150 m buffer	3 arc-seconds (~100 m)	Atlas of urban expansion (Angel et al., 2013)
N1000	Proportion of urban pixels in a 1 km buffer	3 arc-seconds (~100 m)	Atlas of urban expansion (Angel et al., 2013)
N5000	Proportion of urban pixels in a 5 km buffer	3 arc-seconds (~100 m)	Atlas of urban expansion (Angel et al., 2013)
DISTURB	Distance to the nearest urban pixel (meters)	3 arc-seconds (~100 m)	Atlas of urban expansion (Angel et al., 2013)
SLOPE	Slope (%)	1 arc-second (~30 m)	SRTM (Jarvis et al., 2008)

Table 3

Travel speeds assigned to each transport or land cover class in order to produce a friction surface following (Nelson, 2008)

		Travel speed (km/h)
Transport network (VMAP0)	Motorways	120
	Major roads	60
	Tracks	10
	Railways	40
Land cover (refined GlobCover)	Cultivated Terrestrial Areas and Managed Lands	1.66
	Natural and Semi-natural Terrestrial Vegetation – Woody/Trees	1.25
	Natural and Seminal Terrestrial Vegetation – shrubs and herbaceous	1.66
	Natural and Semi-natural Terrestrial Vegetation – Sparse	2.5
	Natural and Seminal Aquatic Vegetation	1
	Urban area, industrial area and rural settlements	30
	Bare areas	2.5
	Water bodies	1

Table 4

Accuracy assessment statistics (Area Under the Curve and Cohen's kappa) obtained by the distance-based model and BRT models

City	Distance-based model		BRT models	
	AUC	Cohen's kappa	AUC	Cohen's kappa
Accra	0.648	0.295	0.680	0.360
Addis Ababa	0.549	0.098	0.571	0.141
Alexandria	0.650	0.301	0.659	0.319
Algiers	0.603	0.206	0.600	0.201
Aswan	0.536	0.072	0.616	0.232
Bamako	0.666	0.331	0.681	0.361
Banjul	0.620	0.240	0.650	0.300
Cairo	0.628	0.256	0.623	0.245
Casablanca	0.598	0.197	0.632	0.264
Harare	0.528	0.055	0.517	0.034
Ibadan	0.679	0.359	0.726	0.452
Johannesburg	0.499	-0.002	0.512	0.023
Kampala	0.582	0.164	0.614	0.229
Kigali	0.728	0.455	0.715	0.430
Marrakech	0.583	0.165	0.645	0.290
Ndola	0.512	0.024	0.517	0.034
Ougadougou	0.720	0.441	0.710	0.421
Port Sudan	0.637	0.275	0.700	0.400
Pretoria	0.507	0.013	0.514	0.028
Tebessa	0.679	0.358	0.753	0.506
Average	0.608	0.215	0.632	0.264

Table 5

PCA and cluster analysis results

	Comp.1	Comp.2	Comp.3	Comp.4
Loadings				
Spatial cohesion	−0.640	0.147	0.565	−0.499
Urban area	−0.241	−0.714	−0.482	−0.447
Urban population density	0.683	0.140	0.110	−0.709
Urban population growth	−0.258	0.670	−0.660	−0.219
Importance of components				
Standard deviation	1.192	1.105	0.867	0.778
Proportion of Variance	0.355	0.305	0.188	0.152
Cumulative Proportion	0.355	0.660	0.848	1.000