

Article

## Simulating Urban Growth Using a Random Forest-Cellular Automata (RF-CA) Model

Courage Kamusoko <sup>1,\*</sup> and Jonah Gamba <sup>2</sup>

<sup>1</sup> Asia Air Survey (AAS) Co., Ltd., Kanagawa 215-0004, Japan

<sup>2</sup> Department of Computer and Information Science, Faculty of Science and Technology, Seikei University, Tokyo 180-8633, Japan; E-Mail: jmgamba@gmail.com

\* Author to whom correspondence should be addressed; E-Mail: cou.kamusoko@ajiko.co.jp; Tel.: +81-44-969-7510; Fax: +81-44-965-0917.

Academic Editor: Wolfgang Kainz

Received: 3 October 2014 / Accepted: 24 March 2015 / Published: 1 April 2015

---

**Abstract:** Sustainable urban planning and management require reliable land change models, which can be used to improve decision making. The objective of this study was to test a random forest-cellular automata (RF-CA) model, which combines random forest (RF) and cellular automata (CA) models. The Kappa simulation (KSimulation), figure of merit, and components of agreement and disagreement statistics were used to validate the RF-CA model. Furthermore, the RF-CA model was compared with support vector machine cellular automata (SVM-CA) and logistic regression cellular automata (LR-CA) models. Results show that the RF-CA model outperformed the SVM-CA and LR-CA models. The RF-CA model had a Kappa simulation (KSimulation) accuracy of 0.51 (with a figure of merit statistic of 47%), while SVM-CA and LR-CA models had a KSimulation accuracy of 0.39 and -0.22 (with figure of merit statistics of 39% and 6%), respectively. Generally, the RF-CA model was relatively accurate at allocating “non-built-up to built-up” changes as reflected by the correct “non-built-up to built-up” components of agreement of 15%. The performance of the RF-CA model was attributed to the relatively accurate RF transition potential maps. Therefore, this study highlights the potential of the RF-CA model for simulating urban growth.

**Keywords:** urban growth; land change models; random forest; cellular automata; kappa simulation

---

## 1. Introduction

Urban land change models are important for analysing the driving forces of land use/cover changes, and simulating “what if” urban growth scenarios [1–3]. This is particularly important in developing countries experiencing rapid urban growth [4–6]. It is estimated that more than three billion people will be living in urban areas by 2050, of which 80% will be inhabitants of cities in developing countries [7,8]. According to the United Nations [7], the urban population in Asia is expected to increase from 1.8 billion in 2010 to 3.4 billion in 2050, while the urban population in Africa is projected to rise from 0.8 billion in 2010 to 1.2 billion in 2050. Rapid urbanisation is expected to increase informal settlements, epidemics and environmental degradation [9,10]. Therefore, urban planners and policy makers require reliable land change models, which can be used to simulate different urban growth or development scenarios [3,11].

The past decades have witnessed the development and application of many urban land change models based on cellular automata (CA) [12–20]. Cellular automata (CA) are bottom-up and discrete dynamic models that were originally conceptualised by Ulam and Von Neumann in the 1940s in order to understand the behaviour of complex systems [21]. The CA model consists of cell space, cell states, neighbourhoods, time steps and transition rules [22]. Space can be represented as a grid of cells, while a neighbourhood is defined as a collection of cells based on adjacency [21,22]. Each cell can assume one of  $i$  discrete states at any one time [23,24]. Time progresses in discrete steps and all cells change their state simultaneously as a function of their own state, together with the state of the adjacent cells according to specified transition rules [25]. The transition rules are key components of CA since they represent the processes of the system being modelled [26]. Distance functions are applied within a neighbourhood to take into account the spatial dependent attractiveness or repulsiveness of one cell state over another [27]. The CA model simulates future land use/cover changes based on the extrapolation of past land use/cover.

Cellular automata (CA) models have significantly contributed to urban growth modelling [2,12,22,24,28,29]. However, previous studies have highlighted limitations regarding the definition of transition rules or transition potential [1,30–32]. In a comparative analysis of twelve empirical transition potential models, Eastman *et al.* [1] revealed that eleven models, including the commonly used multicriteria evaluation (MCE), logistic regression (LR) and weights of evidence (WoE), performed poorly. This is because most of the transition potential models are defined in linear form [33]. As a result, the transition potential models fail to capture the underlying land use/cover change patterns and processes that are often characterised by nonlinearity, complexity, emergence and self-organisation [31,34]. In order to overcome the limitations of linear models, Li and Yeh [35] developed a neural-network CA model to handle complex relationships in urban systems. Although neural networks have been reported to improve land change modelling, they are difficult to calibrate and tend to overfit [35]. Furthermore, Yang *et al.* [33] applied a support vector machine-cellular automata (SVM-CA) model in Shenzhen city. The authors reported that the SVM-CA model achieved higher accuracy and overcame the limitations of neural networks. However, SVMs are sensitive to outliers [36] and generally require more training time, especially if the dataset has many features.

Reliable urban land change models are a key requirement for sustainable urban growth planning [11,37]. While other researchers have recently improved land change models using

auto-logistic regression and multivariate adaptive regression splines models [38–40], there is still need to test other nonlinear models. Random forest (RF) is an ensemble (collection) model [41], which uses bagging (bootstrap aggregated sampling) to build many individual decision trees for a final prediction or classification [42]. The algorithm uses a random subset of predictor variables to split observation data into homogenous subsets [42]. In addition, the RF model uses out-of-bag (OOB) sample data, which are derived from the data that are not in the bootstrap sample to evaluate performance [42]. The advantages of RF models are: (i) they can handle a large database (e.g., thousands of input numerical and categorical variables); (ii) they require less training time compared to other machine learning classifiers (e.g., artificial neural network, SVM, boosting); (iii) they are free of normal distribution assumptions; (iv) they are robust in dealing with outliers and noise; and (v) they quantify each input variable into an importance measure [43]. While, the RF model has been used successfully for remote sensing image classification, to our knowledge the RF model has not been tested for modelling transition potential and simulating urban growth.

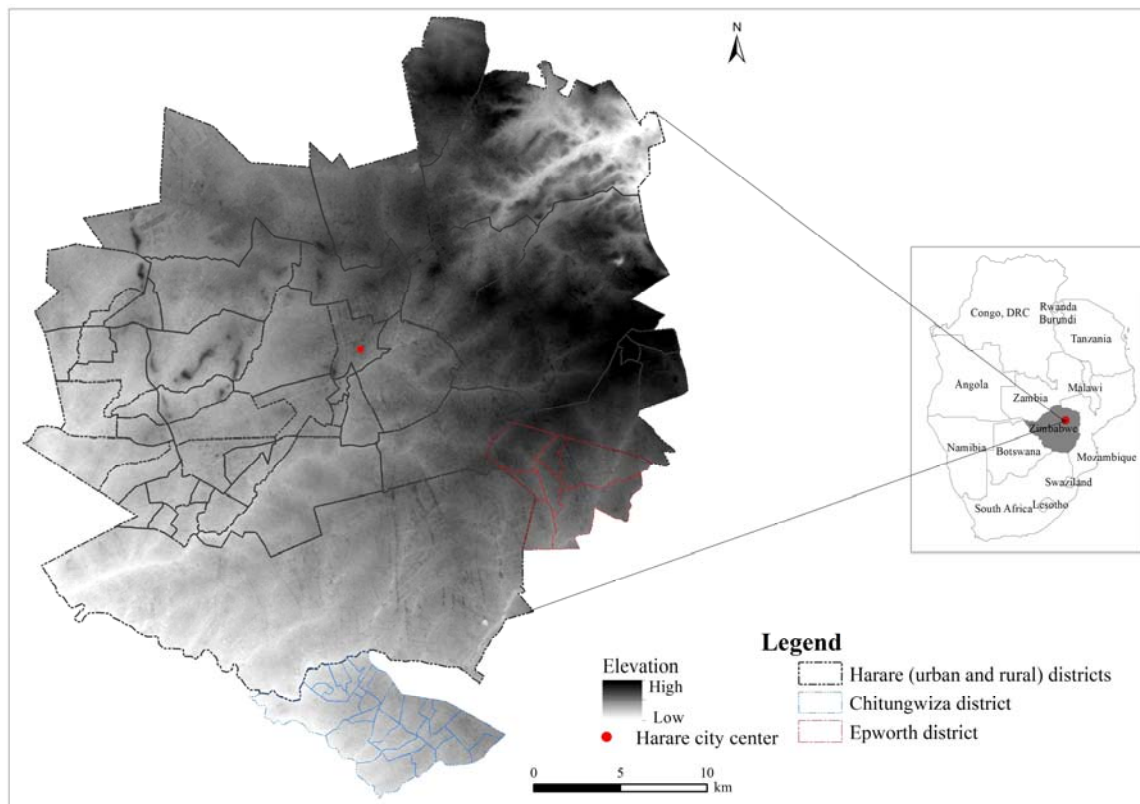
The objective of the study is to test the random forest-cellular automata (RF-CA) urban land change model in Harare Metropolitan Province, Zimbabwe. The RF-CA model applied in this study integrates the RF and CA models in order to test the effectiveness of the RF-CA model for simulating urban growth. First, we calculated multiple-step transition rates from land use/cover maps (1984, 2002 and 2008). Second, the RF model was used to compute transition potential maps. Third, we simulated land use/cover up to 2013 using multiple-step transition rates and a transition potential map based on the CA model. Fourth, the Kappa simulation (KSimulation), figure of merit, and components of agreement and disagreement statistics were used to validate the RF-CA model. In addition, we applied SVM-CA and logistic regression-cellular automata (LR-CA) models in order to compare performance with the RF-CA model. This is because LR-CA and SVM-CA models are some of the commonly used land change models.

## 2. Implementation of the RF-CA Model

### 2.1. Study Area and Data

Harare Metropolitan Province extends between 17°40' and 18°00' south, and between 30°55' and 31°15' east, encompassing an area of about 942 km<sup>2</sup> (Figure 1). The metropolitan province consists of the Harare Urban, Harare Rural, Chitungwiza and Epworth districts. The Harare Urban district incorporates the City of Harare, which is the capital city of Zimbabwe. The spatial structure of the City of Harare is characterised by a radial road network with the central business district (CBD) at its core, and the industrial areas to the east and south [44]. To the north and northeast are low density residential areas on plot sizes of about 1000 m<sup>2</sup> or more, while to the extreme east, south, southwest and west are the high density residential areas on plot sizes of about 300 m<sup>2</sup> [44]. In addition, some medium density residential areas measuring between 800 m<sup>2</sup> and 1000 m<sup>2</sup> are found in the southern part of the City of Harare. Chitungwiza city (in Chitungwiza district) is located approximately 25 km south of the City of Harare. The city was developed by the colonial government in order to allocate residential areas for Africans far from the City of Harare [45]. Although Chitungwiza city has commercial and industrial enterprises, most of its residents work in the City of Harare. Epworth

district, which is located to the south-east of the City of Harare, is an unplanned and informal urban settlement that was formed by war refugees during the liberation struggle in the 1970s [9].



**Figure 1.** Location of Harare Metropolitan Province, Zimbabwe.

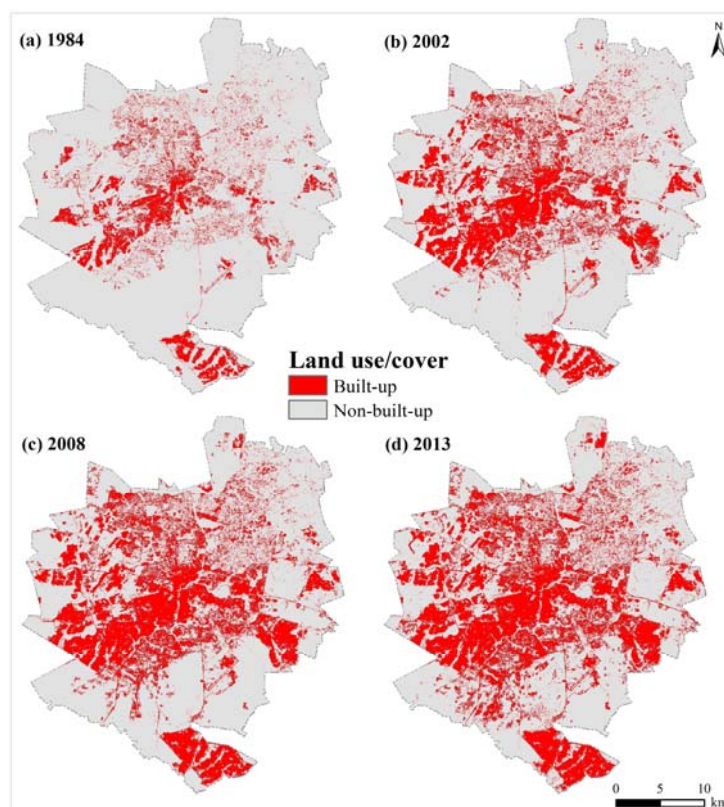
According to Colquhoun [46] and Mutizwa-Mangiza [47], the population of Harare Metropolitan Province increased significantly after independence in 1980, when migration controls were removed. The population in Harare Urban district increased from approximately 642,191 in 1982 to 1,435,784 in 2012, while the population in Harare Rural district increased from 16,173 to 23,023 over the same period [10,48]. However, the population of Chitungwiza City expanded exponentially from approximately 15,000 in 1969 to 354,472 in 2012 [45,48]. The population expansion was mainly driven by people who migrated from rural areas during the liberation struggle in the 1970s [9]. The population of Epworth district also increased rapidly after independence as war refugees were joined by people who could not get accommodation in the City of Harare [45]. Currently, the population of Epworth district is estimated to be 161,840 [48]. Given this rapid population growth and the ensuing urbanisation [49], we selected Harare Metropolitan Province to test the RF-CA model. In addition, Harare Metropolitan Province is characterised by urban growth patterns such as extension, infill and leapfrog developments, which are also observed in other cities in sub-Saharan Africa [17].

We used land use/cover maps and driving factors to develop the RF-CA model (Table 1) for Harare metropolitan province (Table 1 and Figure 2). Land use/cover maps were classified from Landsat imagery for 1984, 2002, 2008, 2013 and validated using anniversary and near-anniversary reference data [49]. Overall accuracy levels for the four dates range from 86% to 93% [49]. Table 2 provides a description of the land use/cover classes. Major roads for the “1984–2002” and “2002–2008” periods were digitised from the 1:30,000 scale Harare Street maps published by the Department of the Surveyor-General

(Zimbabwe) in 1989 and 2005, respectively. In addition, major industrial centers and the city center were also digitised from the 1:30,000 scale Harare Street maps. Elevation was derived from ASTERGDEM, while population density data were acquired from the Zimbabwe Statistical Office [48]. We used built-up areas (extracted from the 1984 and 2002 land cover maps), major roads, major industrial centers, and city center data to compute “distance to built-up areas”, “distance to major roads”, “distance to major industrial areas”, and “distance to city center” using the euclidean distance procedures available in ArcGIS 10.2 (Table 1 and Figure 3). We computed “distance to built-up areas” for 1984 and 2002, and “distance to major roads” for the “1984–2002” and “2002–2008” periods because built-up areas and roads are dynamic driving factors that change over time. Furthermore, we used “distance to built-up areas” as the driving factor because previous urban form influences future urban patterns [26]. Finally, all driving factors were resampled to 30 m × 30 m spatial resolution in order to match the spatial resolution of the Landsat-derived land use/cover maps (Figure 2).

**Table 1.** Input data for calibrating and simulating land use/cover change.

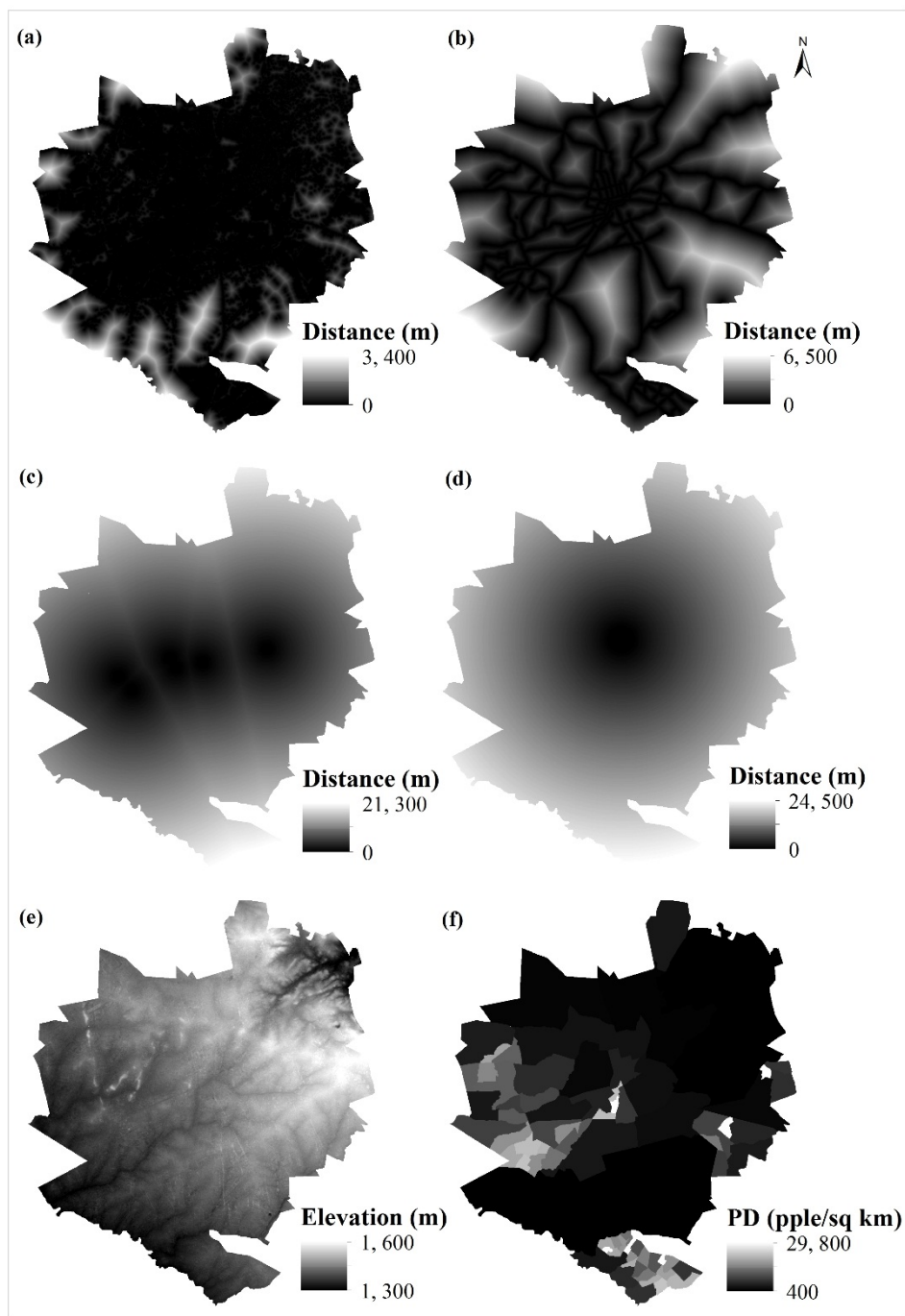
| Variable  | Source   |
|---|--|
| Land use/cover maps (1984, 2002, 2008 and 2013) | Classified from Landsat data                     |
| Distance to built-up areas (1984, 2002)         | Derived from land use/cover maps                 |
| Distance to major roads (1984–2002, 2002–2008)  | Digitised from 1:30,000 scale Harare Street maps |
| Distance to major industrial centers            | Digitised from 1:30,000 scale Harare Street maps |
| Distance to city center                         | Digitised from 1:30,000 scale Harare Street maps |
| Elevation                                       | Derived from ASTERGDEM                           |
| Population density (2002)                       | Derived from Zimbabwe Statistical Office         |



**Figure 2.** Land use/cover for (a) 1984, (b) 2002, (c) 2008 and (d) 2013.

**Table 2.** Land use/cover classes.

| Land Use/Cover Class | Description  |
|----------------------|--|
| Built-up             | Residential, commercial and services, industrial, transportation, communication and utilities, construction sites, and landfills.  |
| Non-built-up         | All wooded areas, riverine vegetation, shrubs and bushes, grass cover, golf courses, parks, cultivated land, fallow land, land under irrigation, bare exposed areas, transitional areas and water. |

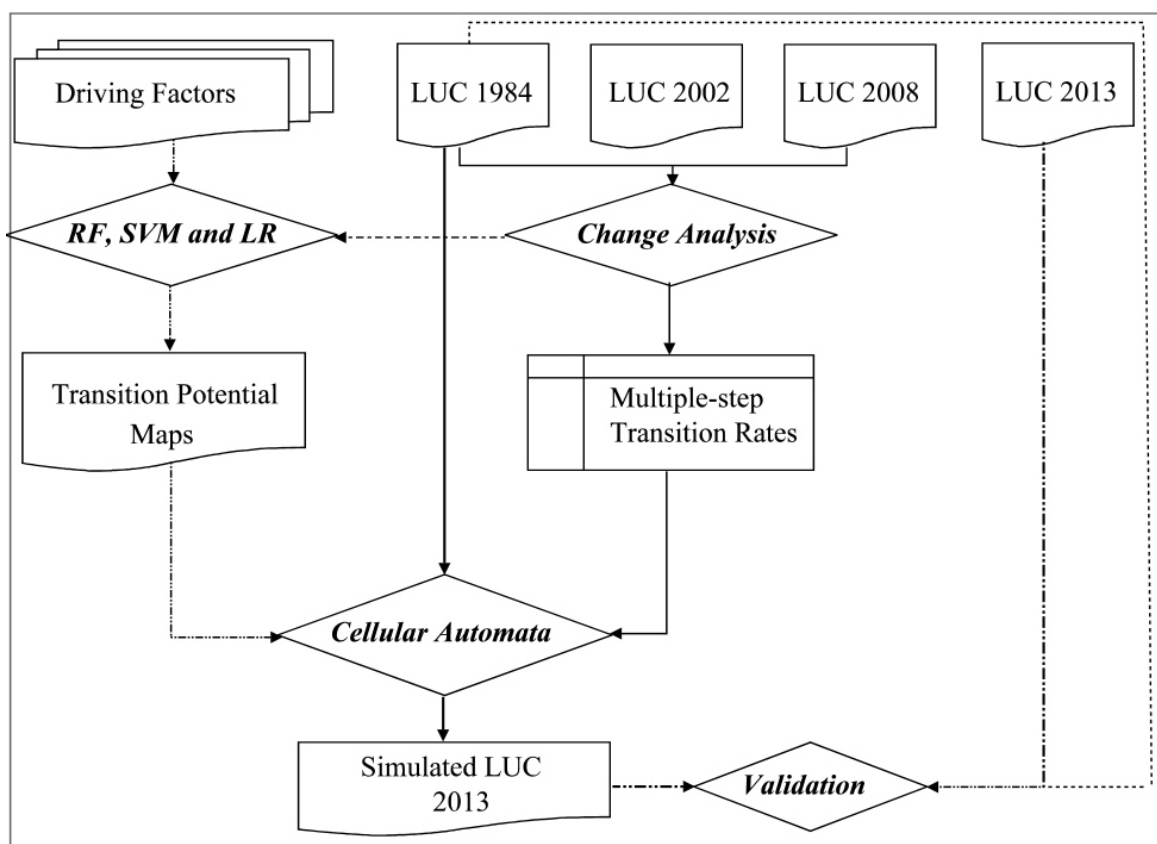


**Figure 3.** Selected driving factors used to compute transition potential maps: (a) distance to built-up areas (2002); (b) distance to major roads (2002); (c) distance to major industrial centers; (d) distance to city center; (e) elevation; and (f) population density (2002).



## 2.2. Model Calibration and Simulation

We used the following procedures to implement the RF-CA model: (1) computing of transition rates; (2) transition potential modelling; and (3) CA simulation, as well as model validation (Figure 4). Machine learning and statistical algorithms available in R were used to model transition potential, while functions available in Dinamica EGO were used to compute transition rates and simulate land use/cover changes. R is a free and open-source statistical and computer graphic software [50], while Dinamica EGO (Environment for Geoprocessing Objects) is freeware that was developed by Soares-Filho *et al.* [51]. Dinamica EGO consists of a sophisticated platform for developing dynamic spatial models, which involve nested iterations, multiple-step transitions, dynamic feedbacks and multi-scale approaches [51].



**Figure 4.** Random forest-cellular automata (RF-CA) model. Note LUC refers to land use/cover.

### 2.2.1. Computation of Transition Rates

We used land use/cover maps for 1984, 2002 and 2008 (Figure 2) to compute single- and multiple-step transition rates in Dinamica EGO. Single-step transition rates refer to absolute aggregate rates computed for a given period (e.g., 16 years), while multiple-step transition rates refer to transition rates that are computed at an annual time step. The single-step transition and multiple-step transition rates are computed according to well-known algorithms available in Dinamica EGO [52]. Table 3a shows that transition rates for the “1984–2002”, “2002–2008” (calibration) and “2008–2013” (validation) periods are different and thus nonstationary. Therefore, we tested the effectiveness of both single and multiple-step transition rates during the CA calibration run. Initial calibration results indicated that the “1984–2008” multiple-step transition rate and the combined “1984–2002”, “2002–2008”, and “1984–2008” multiple-step

transition rates had the best simulation accuracy (Table 3b). However, the combined “1984–2002”, “2002–2008”, and “1984–2008” multiple-step transition rates (Table 3b) produced better spatial allocation accuracy. This is because the “1984–2002” and “2002–2008” multiple-step transition rates allocated the quantity of “non-built-up to built-up” changes, whereas the “1984–2008” multiple-step transition rate regulated or modulated the allocation of “non-built-up to built-up” changes. As a result, overestimation or underestimation was minimised during simulation. Therefore, three multiple-step transition rates from the “1984–2002”, “2002–2008” and “1984–2008” periods were selected for the final CA simulation run (Table 3a). It should be noted that the use of three multiple-step transition rates from the “1984–2002”, “2002–2008” and “1984–2008” periods needs further research in other urban landscapes in order to test its effectiveness.

**Table 3.** (a) Single and multiple-step transition rates (%); (b) Simulation accuracy based on single and multiple-step transition rates (%).

| (a)       |                              |                                |  |
|-----------|------------------------------|--------------------------------|--|
| Period    | Single-Step Transition Rates | Multiple-Step Transition Rates |  |
| 1984–2002 | 14                           | 1                              |  |
| 2002–2008 | 10                           | 2                              |  |
| 1984–2008 | 22                           | 1                              |  |
| 2008–2013 | 11                           | 2                              |  |

| (b)                                    |                                 |                                   |  |
|--|---------------------------------|-----------------------------------|--|
| Period                                 | Single-Step Simulation Accuracy | Multiple-Step Simulation Accuracy |  |
| 1984–2008                              | 6                               | 50                                |  |
| 1984–2002; 2002–2008                   | 6                               | 44                                |  |
| 1984–2002; 2002–2008;<br>and 1984–2008 | 6                               | 50                                |  |

### 2.2.2. Transition Potential Modelling

We used 3000 training points randomly sampled from “non-built-up to built-up” and “no change” (that is, built up and non-built-up persistence) areas between 1984 and 2008 in order to develop the RF model based on the *randomForest* package [53] available in R. All driving factors (Table 1) were used for model development after a multicollinearity test revealed that they were below the threshold value of 0.7 [54], and therefore not redundant. After checking for multicollinearity, we randomly selected 70% of the training points for model development, while 30% were used for cross-validation.

In order to achieve optimum model performance, we adjusted the RF model parameters. The RF algorithm split the input variables into independent groupings based on binary decisions to generate initial large and complex trees. However, large trees tend to overfit the training data, resulting in poor prediction. Therefore, we adjusted the RF model parameters by changing the number of input variables selected at each node split and the total number of trees included in the model (25, 50, 100, and 500). After calibration, 100 trees were used to construct the final RF model and then compute the “non-built-up to built-up” transition potential map.

We also developed SVM and LR models in order to compare performance with the RF model. SVMs are machine-learning techniques based on statistical learning theory [55,56]. The technique was



introduced by Boser *et al.*, [55] and Vapnik [56] to solve classification and regression problems by constructing hyperplanes in a multidimensional space. In general, SVMs select the decision boundary from an infinite number of potential ones, leaving the greatest margin between the closest data points to the hyperplane, which are referred to as “support vectors” [57]. SVM employ a kernel function to transform the training data into higher dimensional feature space for nonlinear classification problems [57]. For the SVM model in this study, we selected a radial basis function as the SVM kernel using the *e071* package available in R [58].

The LR technique models the relationship between a dependent variable and one or more independent variables (which may be categorical or continuous). The LR model can be expressed as:

$$P(Y | X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

where:  $P(Y | X)$  is the probability of the dependent variable  $Y$  given  $X$  (that is, the probability of a cell being urbanised);  $X$  represents independent variables such as distance to roads;  $\beta_0$  is a constant to be estimated; and  $\beta_1$  is a coefficient to be estimated for each independent variable  $X$ . For the LR model in this study, we used the generalized linear model (GLM) available in R [50].

### 2.2.3. Simulation based On the CA Model

Three datasets, (1) the initial land use/cover map (1984); (2) the transition potential maps (1984–2008); and (3) the “1984–2002”, “2002–2008” and “1984–2008” multiple-state transition rates, were used to simulate land use/cover up to 2013 based on the expander and patcher transition CA functions. The expander transition function expands or contracts previous land use/cover class patches, while the patcher transition function forms new patches [51]. The expander and patcher transition functions are composed of an allocation mechanism responsible for identifying cells with the highest transition potential for each transition [51]. For example, the expander transition function performs transitions from state  $i$  (non-built-up) to state  $j$  (built-up) only in the neighbouring cells of state  $j$  in order to expand or contract land use/cover patches. The patcher function then performs transitions from state  $i$  to state  $j$  only in the neighbouring cells with states other than  $j$  [51]. In order to simulate land use/cover changes, both transition functions use a stochastic selecting mechanism [51].

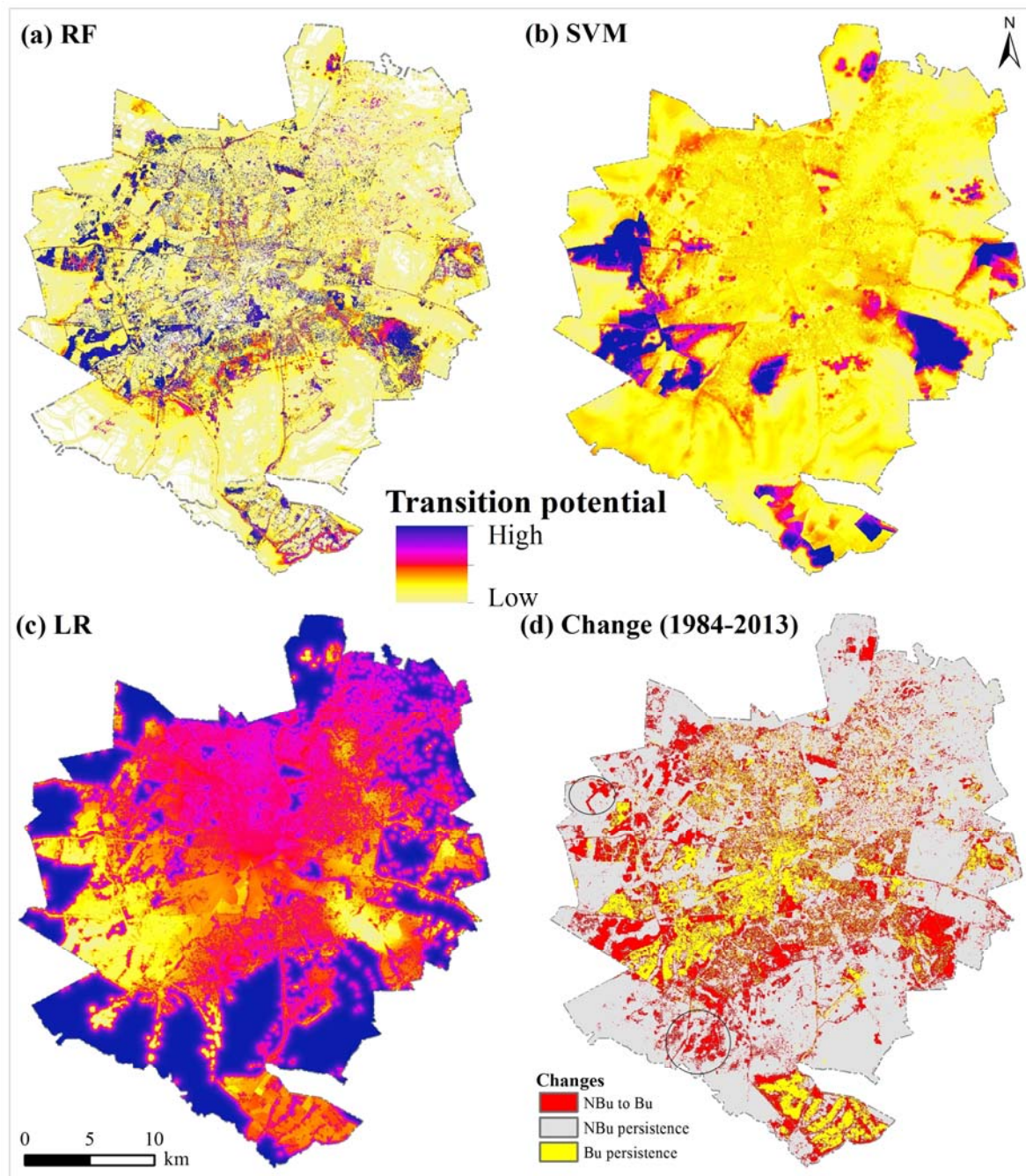
The sizes of new land use/cover patches are set according to a lognormal probability function, whose parameters are defined by the mean patch size (MPS), patch size variance (VAR) and isometry (ISO) [59]. The parameters can be changed to produce various spatial patterns of land use/cover. For this study, we calibrated the CA model by changing the parameters of the expander and patcher transition functions using trial and error. The initial simulation year was 1984, while the final year was 2013 (that is, the observed or reference year). As a result, the CA model had twenty-nine iterations at an annual time-step.

### 3. Results and Discussion

#### 3.1. Evaluating the Goodness-of-Fit of Transition Potential Maps

Figure 5a–c show “non-built-up to built-up” transition potential maps—computed using RF, SVM and LR models—while Figure 5d shows land use/cover changes that occurred between 1984 and 2013. Visual analysis revealed that the RF model produced a relatively accurate transition potential map compared to the SVM and LR models. In particular, the RF model was adept at predicting new infill development and extension built-up areas near previous built-up areas (from 1984 and 2002). Infill development refers to growth of newly developed areas in the urbanised areas of the previous time period (that is, 1984 and 2002), while extension refers to expansion of built-up areas within the urbanised areas [60]. In contrast, the SVM model overestimated the “non-built-up to built-up” changes (Figure 5b). As a result, the SVM transition potential map does not match the observed “non-built-up to built-up” change patterns (Figure 5b,d). This implies that the prediction of newly built-up areas is affected by clumping (that is, correctness bias towards high transition areas) due to overfitting [61]. Figure 5c shows that the LR model performed poorly. This is reflected by the occurrence of high transition potential areas in dominant persistence “non-built-up” areas (Figure 5). Generally, all models fail to predict unplanned leapfrog developments, particularly in the south-western part of the study area (Figure 5d). Leapfrog developments are newly built-up areas that are converted from non-built-up parcels outside of and unconnected with existing urban built-up areas [60]. Previous studies revealed that statistical or machine learning models underpredict the location of new patches that are not connected to existing built-up areas [62] due to spatial or temporal nonstationarity [63].

We first analysed the area under the curve (AUC) for the relative operating characteristic (ROC) statistic to evaluate the goodness-of-fit of transition potential maps [64]. Based on the ROC statistics, a measure with perfect predictive power would yield a value of 1.0, while one with no power (random) would yield a value of 0.5 [1]. Values less than 0.5 (null model) indicate a measure that is systematically incorrect [1]. The AUC ROC statistic—which summarizes the strength of the overall diagnostic availability—was 0.77 for the RF model, 0.75 for the SVM model, and 0.7 for the LR model. However, Figure 5a–c show that the AUC statistic does not provide sufficient information to evaluate model performance in this study. Previous studies revealed that the AUC statistic can be potentially misleading [65,66] because it includes persistence areas in model validation [1]. Therefore, Pontius and Si [67] recommend interpreting the ROC curve as well as using the total operating characteristic (TOC) statistic to evaluate the goodness-of-fit of transition potential maps. The TOC statistic expands on the commonly used ROC statistic [67]. Therefore, TOC statistic provides additional information compared to the ROC statistic, which is useful for evaluating the goodness-of-fit of transition potential maps. For example, the ROC statistic shows only two ratios, hits/(hits plus misses) and false alarms/(false alarms plus correct rejections), while the TOC statistic shows all four entries in the matrix: hits, misses, false alarms and correct rejections [67]. Furthermore, the TOC statistic is more intuitive since it provides results based on the actual units in the contingency table (e.g., square kilometres) instead of a unitless statistic such as AUC [67]. More details about the TOC statistic can be found in Pontius and Si [67].



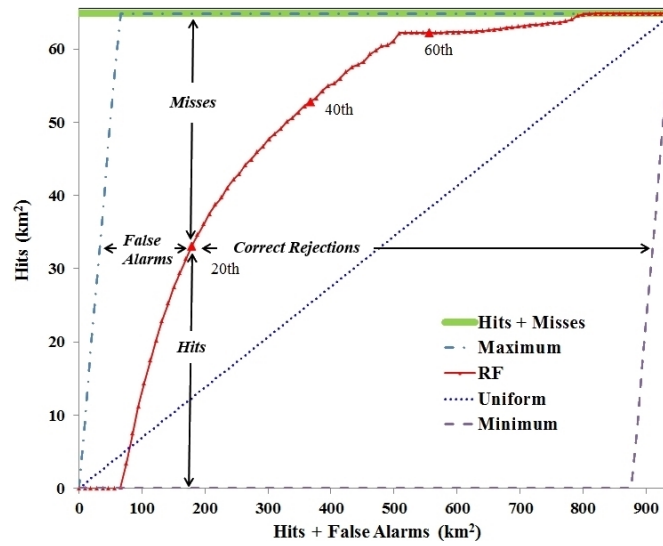
**Figure 5.** (a) RF transition potential map; (b) SVM transition potential map; (c) LR transition potential map; and (d) land use/cover changes between 1984 and 2013 (note black circles show leapfrog developments in the south-western and western parts of the study area). Note NBu represents non-built-up, while Bu represents built-up areas.

Figure 6a–c show the TOC graphs for all models. We focused our model validation on the 20th threshold number, which represents 28.8% or 182 km<sup>2</sup> of the “non-built-up to built-up” changes between 1984 and 2008. Figure 6a shows that the ROC curve for the RF model is above the uniform model at the observed 20th threshold number. This indicates that the RF model is better than the uniform model at predicting the spatial allocation of “non-built-up to built-up” changes. The ROC curve for the SVM model (Figure 6b) is also above the uniform ROC model at the observed 20th threshold

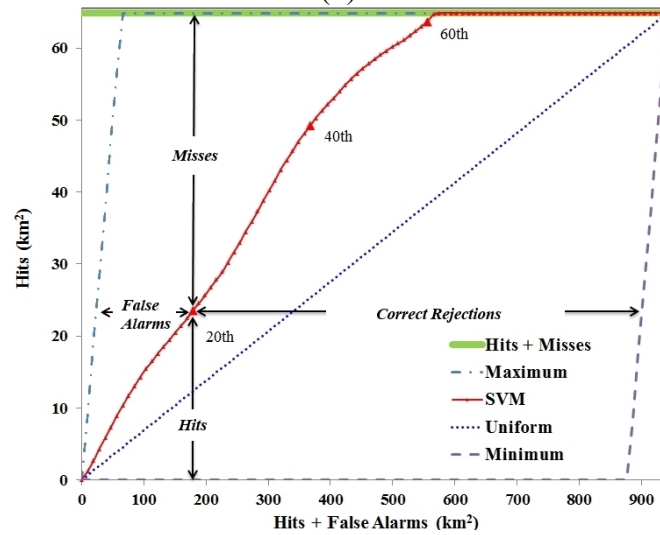
number. However, the ROC curve for the SVM model (Figure 6b) is close to the uniform model, which suggests decreased allocation accuracy for the “non-built-up to built-up” changes. A similar trend is observed with the LR model ROC curve (Figure 6c), which is much closer to the uniform model. This indicates that the LR model is less accurate at predicting the allocation of “non-built-up to built-up” changes. Our results are in agreement with Wang *et al.* [68], who noted that the LR model is less accurate at modelling slow or rapid land use developments. Furthermore, in a study on predictive modelling of potential gold sites, Rodriguez-Galiano *et al.* [69] revealed that LR models overestimated potential gold sites. More importantly, Rodriguez-Galiano *et al.* [69] concluded that RF models performed better than LR models, which is also consistent with our results. Nonetheless, all three models are better than the uniform model at predicting the allocation of non-built-up persistence. This is reflected by the quantity of correct rejections, which is almost similar for all models (Figure 6a–c). Since built-up and non-built-up persistence accounts for approximately 68% of the study area, all models have relatively high AUC. Figure 6a–c shows that the RF model has more hits and fewer misses and false alarms than the SVM and LR models. For example, the RF model had approximately 33.1 km<sup>2</sup> hits (that is, correctly predicted “non-built-up to built-up” changes) compared to 37.5 km<sup>2</sup> of the observed “non-built-up to built-up” changes between 2008 and 2013 (validation period). Contrarily, the SVM and LR models had approximately 23.5 km<sup>2</sup> and 15.7 km<sup>2</sup> hits, respectively. Consequently, the RF model was better at predicting the spatial allocation of “non-built-up to built-up” changes than the SVM and LR models. This is because the RF model can handle the nonlinear relationship between dependent and explanatory driving factors. Therefore, the RF model was well suited to predict urban growth based on both numerical and categorical driving factors used in this study. In addition, the RF model was influenced less by overfitting. As a result, the prediction of new built-up areas was not affected by clumping.

### 3.2. RF-CA Model Validation

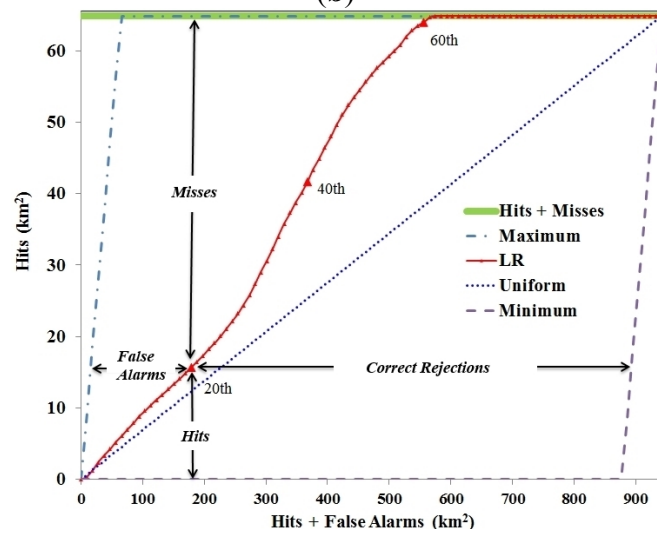
Figure 7 shows the observed and simulated land use/cover maps for the study area. Visual analysis shows that the RF-CA model had the best correspondence between the observed and simulated land use/cover maps for 2013 (Figure 7b). This suggests that the RF-CA model was relatively accurate at allocating “non-built-up to built-up” changes as well simulating infill development and extension built-up patterns in the study area. Figure 7c shows that the simulated built-up patterns do not match the observed built-up patterns. This suggests that while the SVM-CA model has relatively high simulation accuracy in terms of quantity, the spatial allocation of “non-built-up to built-up” changes was poor due to overfitting observed during the calibration of the SVM model (Figure 5b). As a result, the SVM-CA model had difficulty to simulate built-up patterns similar to the observed built-up patterns. Furthermore, the LR-CA model indicates poor correspondence between the observed and simulated land use/cover map (Figure 7d). This shows that the LR-CA model failed to allocate “non-built-up to built-up” changes. Our results are in agreement with some studies which revealed that logistic regression-CA models performed poorly for simulating urban land use changes [33]. Note that all simulation models (RF-CA, SVM-CA and LR-CA) failed to simulate unconnected new built-up areas. This is because all transition potential models (RF, SVM and LR) failed to predict leapfrog developments (Figure 5).



(a)



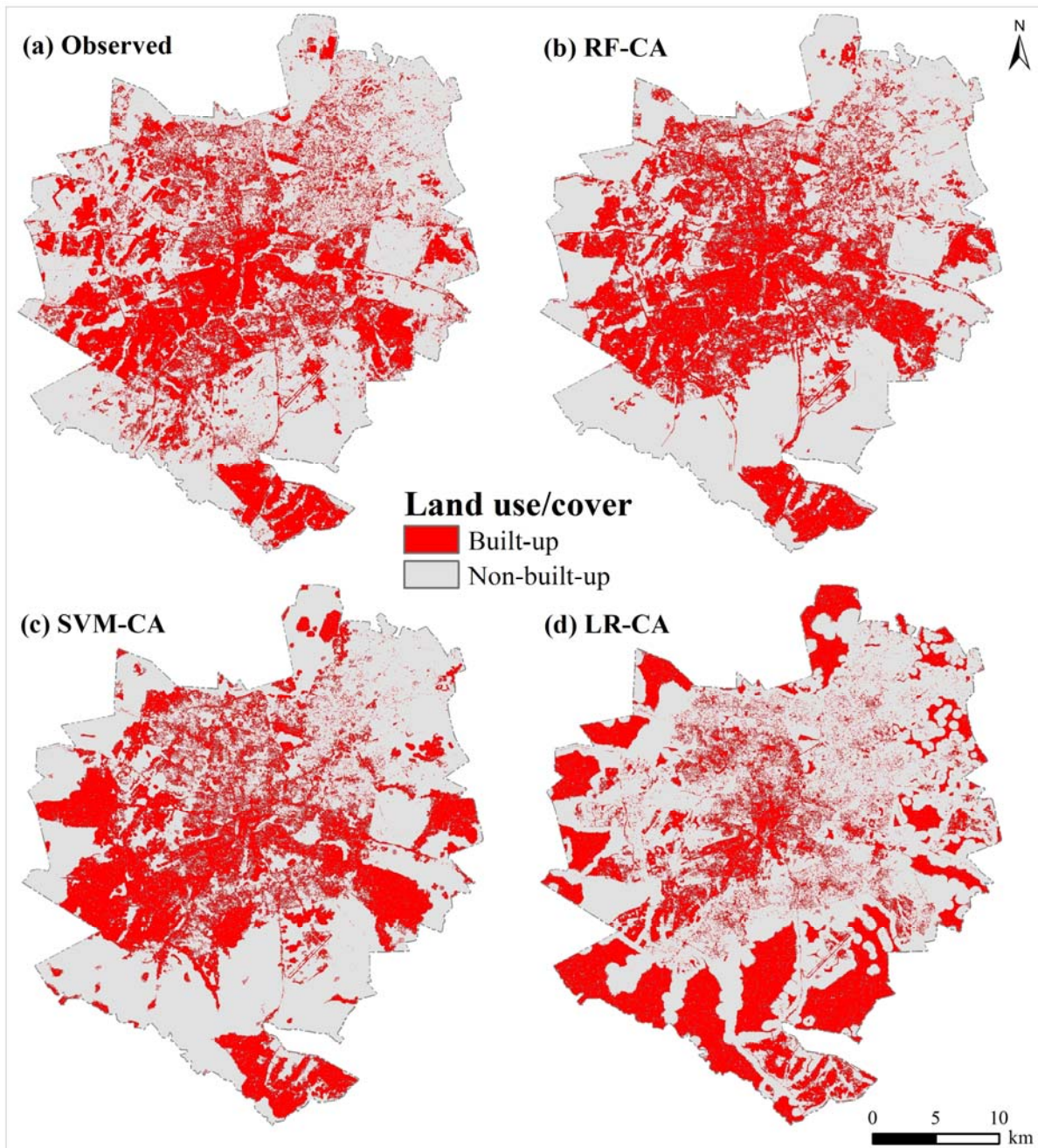
(b)



(c)

**Figure 6.** (a) Total Operating Characteristic (TOC) for the RF model; (b) Total Operating Characteristic (TOC) for the SVM model; (c) Total Operating Characteristic (TOC) for the LR model.





**Figure 7.** Comparison of observed *versus* simulated land use/cover maps for 2013: (a) observed land use/cover map; (b) RF-CA simulated land use/cover map; (c) SVM-CA simulated land use/cover map; and (d) LR-CA simulated land use/cover map.

For quantitative model validation, we used the observed (initial) land use/cover map for 1984, the observed (reference) land use/cover map for 2013, and the simulated land use/cover map for 2013. The Kappa simulation (KSimulation), Kappa transition (KTransition), Kappa translocation (KTranslocation), and the figure of merit statistics were used to validate the RF-CA model [70–73]. The KSimulation expresses the agreement between the simulated land use/cover transitions and reference land use/cover transitions, while KTranslocation measures the degree to which the transitions agree in terms of allocations [70–72]. The KTransition captures the agreement in terms of quantity of built-up and non-built-up transitions, while the figure of merit expresses agreement between the observed and



simulated changes [70–72]. The KSimulation, KTransition and KTranslocation statistics are available in Map Comparison Kit software [70]. More details about the KSimulation, KTransition and KTranslocation statistics can be found in [70,71], while details of the figure of merit are available in [73].

Table 4 shows the validation statistics based on KSimulation, KTranslocation, KTransition and the figure of merit. The overall KSimulation score for the RF-CA model indicates that the “non-built-up to built-up” changes were correctly simulated. Further analysis shows that the RF-CA model correctly simulated allocation and quantity of “non-built-up to built-up” changes. This is reflected by the high KTranslocation and KTransition (Table 4). The overall KSimulation score for the SVM-CA model was 0.12 lower than the RF-CA model. The KTranslocation for the SVM-CA model was lower than the RF-CA model, which indicates that the SVM-CA model poorly simulated the allocation of “non-built-up to built-up” changes (Table 4). In contrast to the RF-CA and SVM-CA models, the overall KSimulation score for the LR-CA model was extremely low. Clearly, the LR-CA model failed to simulate the “non-built-up to built-up” changes. The low KTranslocation of  $-0.22$  indicates that the LR-CA model could not allocate “non-built-up to built-up” changes during the simulation. Most allocation errors for the LR-CA model are attributed to poor performance of the LR model. Generally, the RF-CA model had the highest accuracy as shown by a high figure of merit (approximately 47%). A study by Pontius *et al.* [72] revealed that the figure of merit observed in other land change models ranged from 1% to 59%. Therefore, the accuracy of the RF-CA model is relatively high since the figure of merit is within the upper range of previously observed land change models [73].

It is interesting to note that the KTransition was very high (98% to 99%) for all simulation models because similar multiple-step transition rates were used during CA simulation (Table 3a). A quantitative comparison of the observed and simulated land use/cover maps show that the observed built-up class was  $338.3 \text{ km}^2$ , while the corresponding simulated class was  $340.3 \text{ km}^2$  for the RF-CA model. In contrast, the observed non-built-up class was  $601.9 \text{ km}^2$ , whereas the corresponding simulated class was  $599.9 \text{ km}^2$ . For the SVM-CA model, the observed built-up class was  $338.3 \text{ km}^2$ , while the corresponding simulated class was  $332.9 \text{ km}^2$ . However, the observed non-built-up class was  $601.9 \text{ km}^2$ , while the corresponding simulated class was  $607.4 \text{ km}^2$ . For the LR-CA model, the observed built-up class was  $338.3 \text{ km}^2$ , whereas the corresponding simulated class was  $340.3 \text{ km}^2$ . In contrast, the observed non-built-up class was  $601.9 \text{ km}^2$ , while the corresponding simulated class was  $599.9 \text{ km}^2$ . These results show that all simulation models were relatively accurate for simulating land use/cover quantity.

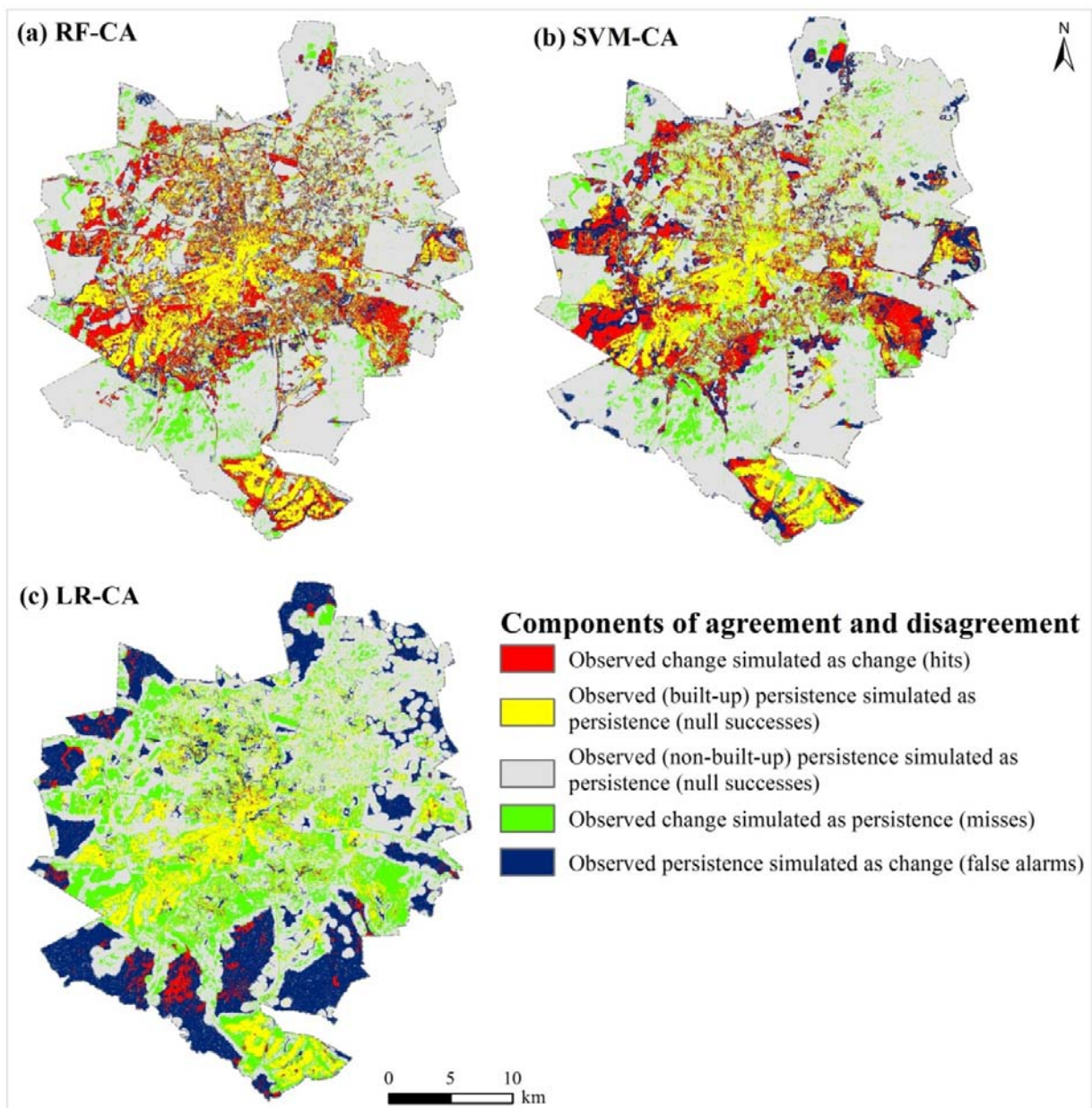
**Table 4.** Validation statistics for all simulation models.

| Model  | KSimulation | KTranslocation | KTransition | Figure of Merit (%) |
|--------|-------------|----------------|-------------|---------------------|
| RF-CA  | 0.51        | 0.51           | 0.99        | 47                  |
| SVM-CA | 0.39        | 0.4            | 0.98        | 39                  |
| LR-CA  | $-0.22$     | $-0.22$        | 0.99        | 6                   |

### 3.3. Analysis of Components of Agreement and Disagreement

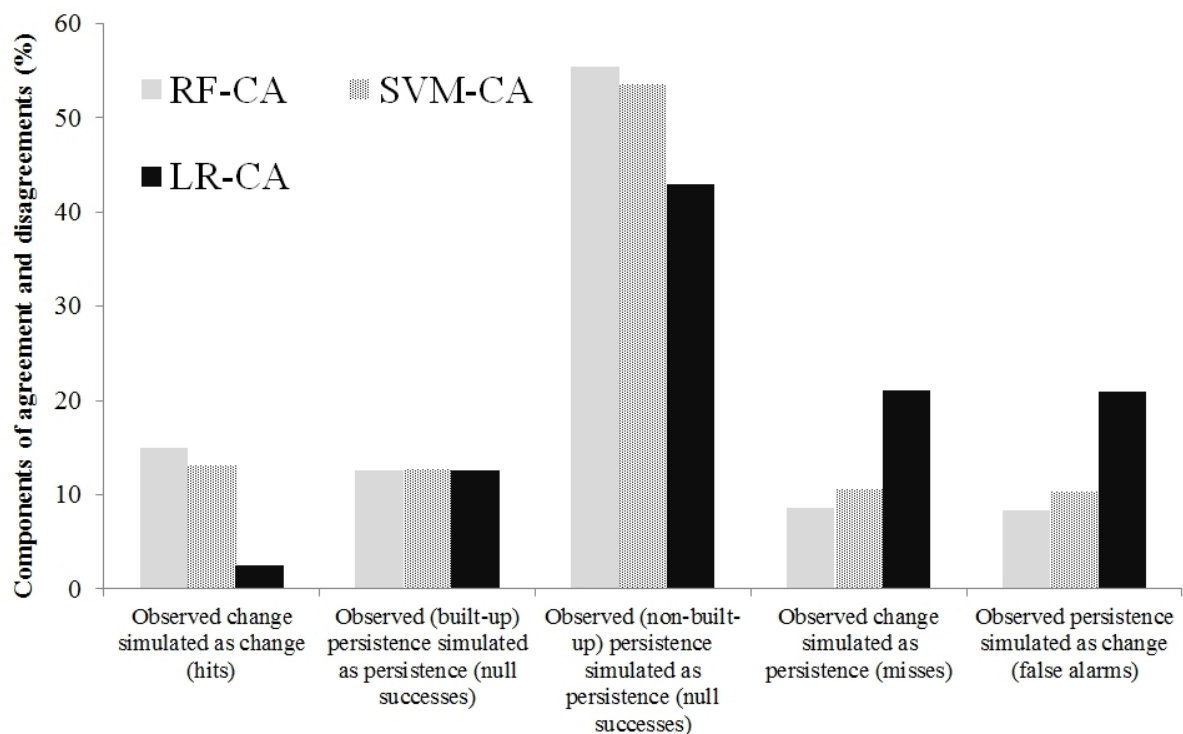
The KSimulation statistic provides a quantitative measure of simulation accuracy. However, KSimulation does not provide the components of agreement and disagreement between the observed

and simulated land use/cover maps. Therefore, we analysed components of agreement and disagreement for the RF-CA, SVM-CA and LR-CA models. Figure 8a–c show the components of agreement and disagreement based on the overlay of the initial (1984), the observed (2013) and simulated land use/cover maps (2013) for all models. The components of agreement and disagreement reveal information such as: (1) observed change simulated correctly as change (hits); (2) observed persistence (that is, built-up and non-built-up) simulated correctly as persistence (null successes); (3) observed change simulated incorrectly as persistence (misses); and (4) observed persistence simulated incorrectly as change (false alarms).



**Figure 8.** Components of agreement and disagreement for: (a) RF-CA simulated land use/cover map; (b) SVM-CA simulated land use/cover map; and (c) LR-CA simulated land use/cover map.

For the RF-CA model, non-built-up persistence had the largest components of agreement, accounting for approximately 55% of the study area (Figures 8a and 9). This is because non-built-up persistence occupied about 68% of the study area between 1984 and 2008. The second largest components of agreement was “non-built-up to built-up” changes accounting for approximately 15% of the study area, while built-up persistence had the smallest components of agreement with approximately 13% (Figures 8a and 9). The largest components of disagreement were the misses (that is, observed change simulated as persistence at 9%) and false alarms (observed persistence simulated as change at 8%). Figure 8a shows that the RF-CA model performed relatively well. However, further analysis indicates (Figure 9) that the combined misses and false alarms (17%) are slightly greater than the hits (15%). For the SVM-CA model, non-built-up persistence had the largest component of agreement with approximately 54%, followed by “non-built-up to built-up” changes and built-up persistence with approximately 13% (Figures 8b and 9). The largest components of disagreement were the misses (that is, observed change simulated as persistence at 10%) and false alarms (observed persistence simulated as change at 10%). The combined misses and false alarms (20%) for the SVM-CA model are greater than the hits (13%). However, for the LR-CA model, non-built-up persistence had the largest component of agreement, with approximately 43% (Figures 8c and 9). The second largest components of agreement was built-up persistence (with approximately 13%), while “non-built-up to built-up” changes had the smallest components of agreement with merely 2% (Figures 8c and 9). The largest components of disagreement were the misses (that is, observed change simulated as persistence at 21%) and false alarms (observed persistence simulated as change at 21%), hence its poor simulation accuracy (Table 4).

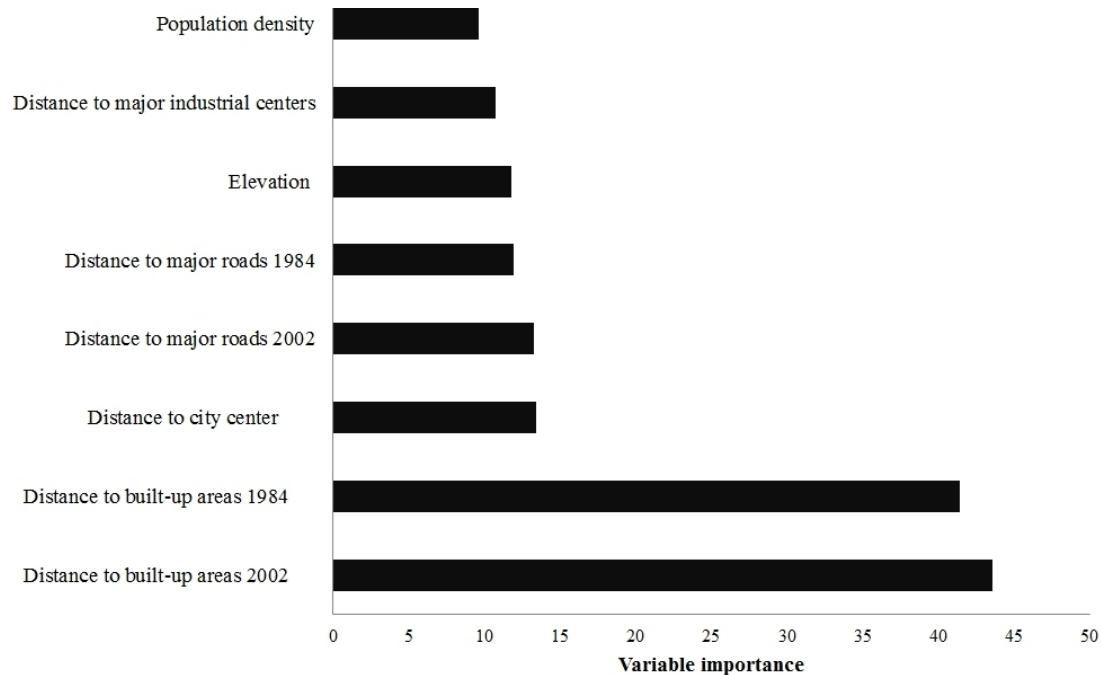


**Figure 9.** Components of agreement and disagreement expressed as a percentage.

The simulation results show that the RF-CA model was substantially more accurate than the SVM-CA and LR-CA models. This is because the RF model was better at modelling the unbalanced land outcomes, namely the combination of rapid and slow urban growth developments, which occurred during the “1984–2002” and “2002–2008” periods. For example, the rate of “non-built-up to built-up” change between 1984 and 2002 was approximately 114.4 km<sup>2</sup>, while the “non-built-up to built-up” change slowed to 69.8 km<sup>2</sup> between 2002 and 2008 [49]. According to Wang *et al.* [68], LR models are not recommended when rapid or slow land change processes result in highly unbalanced land outcomes. It is also important to note that the number of training pixels for the “non-built-up to built-up” change was less than the persistence land use/cover areas (built-up and non-built). As a result of the unbalanced training samples, the LR model failed to generalize resulting in large errors. Our results suggest that the SVM model was also affected by overfitting and hence the SVM-CA model has lower accuracy than the RF-CA model.

This study highlights important insights that may be used to improve land change models. First, the RF-CA model used multiple-step transition rates—from the “1984–2002”, “2002–2008” and “1984–2008” epochs—which were computed from three land use/cover maps (1984, 2002 and 2008). This is important because land use/cover changes, especially over a twenty-nine year time period, follow nonlinear changes that are too complex to be represented by two observation dates only [74,75]. Therefore, the three-epoch multiple-step transition rates improved the spatial allocation of “non-built-up to built-up” changes. Second, we used temporal “distance to previous built-up” (1984 and 2002), and “distance to major roads” (1984–2002 and 2002–2008 periods) driving factors to improve spatial allocation of “non-built-up to built-up” changes in the CA model framework. Third, we employed the RF model, which fits the nonlinear relationship between the “non-built-up to built-up” changes and driving factors based on learning. Results indicate that the RF transition potential map (Figure 5a) shows relatively accurate urban growth patterns such as extension and infill developments. As a result, the RF-CA model was better at allocating “non-built-up to built-up” changes than the SVM-CA and LR-CA models. Fourth, while the RF model cannot analyse the effects of the driving factors on “non-built-up to built-up” changes, variable importance was computed. Figure 10 shows that the “distance to previous built-up” driving factors had the highest importance, followed by “distance to city center” in the study area. Our results are in agreement with a previous study [11], which revealed that urban land in a 1 km neighbourhood and accessibility to the city center were the most influential variables for modelling spatial patterns of urban growth in Africa. Fifth, RF-CA model combines the advantages of both the RF model and spatially explicit dynamic stochastic CA model available in Dinamica EGO. For example, the RF model establishes a nonlinear relationship between land use/cover changes and driving factors in order to produce a transition potential map. The CA model then uses patch and edge expansion functions to allocate change pixels based on the RF transition potential map and multiple-step transition rates [1,59]. In addition, the CA model also incorporates a saturation value parameter, which varies multiple-step transition rates based on dynamic analysis of feedbacks [52,59]. Since the neighbourhood in the CA model is updated during each simulation, spatial allocation of pixels improves given a relatively accurate transition potential map. Last but not least, both the RF transition potential model and the RF-CA simulation model have been validated using validation statistics recommended by land change modelling experts [65–67,70,71]. This is important

because reliable and informative validation statistics provide valuable insights on modelling and simulation errors, which may assist researchers in improving land change models.



**Figure 10.** Variable importance for the “non-built-up to built-up” changes based on mean decrease accuracy.

While this study has highlighted important insights that can be used to improve urban land change models, there are a number of limitations which require further study. First, the combined misses and false alarms are slightly greater than the hits because the RF-CA model failed to simulate unplanned leapfrog developments in the south-western part of the study area (Figure 3d). Second, failure to integrate spatial explicit socioeconomic data (e.g., housing development plans, income levels, *etc.*) due to data unavailability [30] implies that some “non-built-up to built-up” changes will not be predicted and hence cannot be simulated correctly. Furthermore, issues related to nonstationarity need to be addressed by using more temporal land use/cover data (e.g., at five year intervals) or combining RF-CA models with other land change models. Third, we used only built-up and non-built-up classes for simulating urban growth, which simplifies the land use/cover patterns and urban growth processes [76] in the study area. This is because the original input land use/cover data consist of the built-up, non-built-up and water classes only. Therefore, further studies should test the RF-CA model using multiple land use/cover classes.

#### 4. Conclusions

The objective of this study was to test a RF-CA land change model for Harare Metropolitan Province. In order to implement the RF-CA model, we computed multiple-step transition rates, and performed transition potential modelling and CA simulation as well as model validation. In addition, we applied SVM-CA and LR-CA models in order to compare performance with the RF-CA model.

Simulation results show that the RF-CA model performed better than the SVM-CA and LR-CA models. The RF-CA model had a high simulation accuracy, while SVM-CA and LR-CA models had lower simulation accuracies. The performance of RF-CA model was attributed to the relatively accurate RF transition potential maps. Generally, the RF-CA model was relatively accurate at allocating “non-built-up to built-up” changes as well as simulating built-up patterns such as extension and infill developments. For the RF-CA model, the non-built-up persistence had the largest components of agreement, while the second largest components of agreement were “non-built-up to built-up” changes. The modelling and simulation results presented in this paper, however case study specific, provide an initial exploration of the RF-CA model for land change modelling. While some model uncertainties remain, the RF-CA model developed in this study has potential to improve land change modelling in general, and urban growth modelling and simulation in particular.

### Acknowledgments

We would like to thank all the anonymous reviewers who provided useful comments, which were incorporated into this manuscript. The reviewers significantly improved the quality of this paper. We also thank Devena Haggis from the University of Tsukuba in Japan for English editing.

### Author Contributions

Courage Kamusoko was responsible for designing and conducting the study. He calibrated and validated all the simulation models. Jonah Gamba was responsible for revising the manuscript.

### Conflicts of Interest

The authors declare no conflict of interest.

### References

1. Eastman, J.R.; Solorzano, L.A.; van Fossen, M.E. Transition potential modeling for land-cover change. In *GIS, Spatial Analysis, and Modeling*; Maguire, D.J., Batty, M., Goodchild, M.F., Eds.; ESRI Press: California, UK, 2005; pp. 357–385.
2. Torrens, P.M. Simulating sprawl. *Ann. Assoc. Am. Geogr.* **2008**, *96*, 248–275.
3. Cheng, J.; Masser, I. Understanding spatial and temporal processes of urban growth: Cellular automata modelling. *Environ. Plann. B* **2004**, *31*, 167–194.
4. Griffiths, P.; Hostert, P.; Gruebner, O.; van der Linden, S. Mapping megacity growth with multisensory data. *Remote Sens. Environ.* **2010**, *114*, 426–439.
5. Pacione, M. Sustainable urban development in the UK: Rhetoric or reality? *Geography* **2007**, *92*, 246–263.
6. Lopez, E.; Bocco, G.; Mendoza, M.; Duhau, E. Predicting land-cover and land-use in the urban fringe—A case study in Morelia city, Mexico. *Landsc. Urban Plan.* **2001**, *55*, 271–285.
7. United Nations. *World Urbanization Prospects: The 2011 Revision*, 2012. Available online: <http://esa.un.org/unpd/wup/index.htm> (accessed on 25 July 2012).



8. Masser, I. Managing our urban future: The role of remote sensing and geographic information systems. *Habitat Int.* **2001**, *25*, 503–512.
9. Brown, A. Cities for the urban poor in Zimbabwe: urban space as a resource for sustainable development. *Dev. Pract.* **2001**, *11*, 263–281.
10. Rakodi, C. *Harare—Inheriting a Settler-Colonial City: Change or Continuity?*; John Wiley & Sons.: Chichester, UK, 1995.
11. Linard, C.; Tatem, A.J.; Gilbert, M. Modelling spatial patterns of urban growth in Africa. *Appl. Geogr.* **2013**, *44*, 23–32.
12. Clarke, K.C.; Hoppen, S.; Gaydos, L. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay Area. *Environ. Plann. B* **1997**, *24*, 247–261.
13. Batty, M. Urban evolution on the desktop: Simulation with the use of extended cellular automata. *Environ. Plann. B* **1998**, *30*, 1943–1967.
14. Li, X.; Yeh, A.G. Calibration of cellular automata by using neural networks for the simulation of complex urban systems. *Environ. Plann. A* **2001**, *33*, 1445–1462.
15. Pijanowski, B.C.; Pithadia, S.; Shellito, B.A.; Alexandridis, K. Calibrating a neural network-based change model for two metropolitan areas of the Upper Midwest of the United States. *Int. J. Geogr. Inf. Sci.* **2005**, *19*, 197–215.
16. Chunyang, H.; Okada, N.; Zhang, O.; Shia, P.; Zhang, J. Modeling urban expansion scenarios by coupling cellular automata model and system dynamic model in Beijing, China. *Appl. Geogr.* **2006**, *26*, 323–345.
17. Mundia, C.N.; Aniya, M. Modeling urban growth of Nairobi city using cellular automata and geographical information systems. *Geogr. Rev. Jpn.* **2007**, *80*, 777–788.
18. Shan, J.; Alkheder, S.; Wang, J. Genetic algorithm for the calibration of cellular automata urban growth modeling. *Photogramm. Eng. Remote Sens.* **2008**, *74*, 1267–1277.
19. Yeh, A.G.O.; Li, X. Cellular automata and GIS for urban planning. In *Manual of Geographic Information Systems*; Madden, M., Ed.; American Society for Photogrammetry and Remote Sensing: Bethesda, MD, USA, 2009; pp. 591–619.
20. Al-Ahmadi, K.; See, L.; Heppenstall, A.; Hogg, J. Calibration of a fuzzy cellular automata model of urban dynamics in Saudi Arabia. *Ecol. Complex.* **2009**, *6*, 80–101.
21. Moreno, N.; Wang, F.; Marceau, D.J. A geographic object-based approach in cellular automata modeling. *Photogramm. Eng. Remote Sens.* **2010**, *76*, 183–191.
22. White, R.; Engelen, G. Cellular automata as the basis of integrated dynamic regional modeling. *Environ. Plann. B* **1997**, *24*, 235–246.
23. Tobler, W. Cellular geography. In *Philosophy in Geography*; Gale, S., Olsson, G., Eds.; Dordrecht Reidel: Dordrecht, The Netherlands, 1979; pp. 379–386.
24. Couclelis, H. Cellular worlds: A framework for modeling micro-macro dynamics. *Environ. Plann. A* **1985**, *17*, 585–596.
25. Engelen, G.; White, R.; Uljee, I.; Drazan, P. Using cellular automata for integrated modeling of socio-environmental systems. *Environ. Monit. Assess.* **1995**, *34*, 203–214.
26. Liu, Y. *Modelling Urban Development with Geographical Information Systems and Cellular Automata*; CRC Press, Taylor & Francis Group: Boca Raton, FL, USA, 2009; p. 188.

27. Soares-Filho, B.S.; Cerqueira, G.C.; Pennachin, C.L. Modeling the spatial transition probabilities of landscape dynamics in an Amazonian colonization frontier. *BioScience* **2002**, *51*, 1059–1067.
28. Batty, M.; Xie, Y. Urban growth using cellular automata models. In *GIS, Spatial Analysis, and Modeling*; Maguire, D.J., Batty, M., Goodchild, M.F., Eds.; ESRI Press: Redlands, CA, USA, 2005; pp. 151–172.
29. Wu, F.; Webster C.J. Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environ. Plann. B* **1998**, *25*, 103–126.
30. Verburg, P.; de Nijs, T.; Ritsema van Eck, J.; Visser, H.; de Jong, K. A method to analyse neighbourhood characteristics of land use patterns. *Comput. Environ. Urban Syst.* **2004**, *28*, 667–690.
31. Liu, X.; Li, X.; Shi, X.; Wu, S.; Liu, T. Simulating complex urban development using kernel-based non-linear cellular automata. *Ecol. Model.* **2008**, *211*, 169–181.
32. Liu, Y.; Feng, Y.; Pontius, R. Spatially-explicit simulation of urban growth through self-adaptive genetic algorithm and cellular automata modelling. *Land* **2014**, *3*, 719–738.
33. Yang, Q.; Li, X.; Shi, X. Cellular automata for simulating land use changes based on support vector machines. *Comput. Geosci.* **2008**, *34*, 592–602.
34. Torrens, P.M. *How Cellular Models of Urban Systems Work (1. Theory)*; CASA Working Paper Series (28), Centre for Advanced Spatial Analysis (UCL): London, UK, 2000. Available online: [http://www.casa.ucl.ac.uk/working\\_papers/paper28.pdf](http://www.casa.ucl.ac.uk/working_papers/paper28.pdf) (accessed on 17 August 2009).
35. Li, X.; Yeh, A. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *Int. J. Geogr. Inf. Sci.* **2002**, *16*, 323–343.
36. Resler, L.; Shao, Y.; Tomback, D.; Malanson, G. Predicting functional role and occurrence of Whitebark Pine (*Pinus albicaulis*) at Alpine Treelines: Model accuracy and variable importance. *Ann. Assoc. Am. Geogr.* **2014**, *104*, 1–20.
37. Kocabas, V.; Dragicevic, S. Assessing cellular automata model behaviour using sensitivity analysis approach. *Comput. Environ. Urban* **2006**, *30*, 921–953.
38. Lin, Y.P.; Chu, H.J.; Wu, C F.; Verburg, P.H. 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.* **2011**, *25*, 65–87.
39. Tayyebi, A.; Pijanowski, B.C.; Linderman, M.; Gratton, C. Comparing three global parametric and local non-parametric models to simulate land use change in diverse areas of the world. *Environ. Modell. Softw.* **2014**, *59*, 202–221.
40. Tayyebi, A.; Pijanowski, B.C. 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.* **2014**, *28*, 102–116.
41. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
42. Mellor, A.; Haywood, A.; Stone, C.; Jones, S. The performance of random forests in an operational setting for large area Sclerophyll forest classification. *Remote Sens.* **2013**, *5*, 2838–2856.
43. Rodriguez-Galiano, V.F.; Chica-Olmo, M.; Abarca-Hernandez, F.; Atkinson, P.M.; Jeganathan, C. Random forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sens. Environ.* **2012**, *121*, 93–107.
44. Gamanya, R.; de Maeyer, P.; de Dapper, M. Object-oriented change detection for the city of Harare, Zimbabwe. *Expert Syst. Appl.* **2009**, *36*, 571–588.

45. Zinyama, L.; Tevera, D.; Cumming, S. *Harare: The Growth and Problems of the City*; Zinyama, L., Tevera, D., Cumming, S., Eds.; University of Zimbabwe Publications: Harare, Zimbabwe, 1993.
46. Colquhoun, S. Present problems facing the harare city council. In *Harare: The Growth and Problems of the City*; Zinyama, L., Tevera, D., Cumming, S., Eds.; University of Zimbabwe Publications: Harare, Zimbabwe, 1993; pp. 33–41.
47. Mutizwa-Mangiza, N.D. Urban centres in Zimbabwe: Inter-censal changes, 1962–1982. *Geography* **1986**, *71*, 148–151.
48. ZimStats (Zimbabwe National Statistics Agency). *Census 2012: Preliminary Report*; ZimStats (Zimbabwe National Statistics Agency): Harare, Zimbabwe, 2012.
49. Kamusoko, C.; Gamba, J.; Murakami, H. Monitoring urban spatial growth in Harare Metropolitan Province, Zimbabwe. *ARS* **2013**, *2*, 322–331.
50. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing, 2005. Available online: [http://r-project.kr/sites/default/files/2%EA%B0%95%EA%B0%95%EC%A2%8C%EC%86%8C%EA%B0%9C\\_%EC%8B%A0%EC%A2%85%ED%99%94.pdf](http://r-project.kr/sites/default/files/2%EA%B0%95%EA%B0%95%EC%A2%8C%EC%86%8C%EA%B0%9C_%EC%8B%A0%EC%A2%85%ED%99%94.pdf) (accessed on 3 April 2014).
51. Soares-Filho, B.S.; Rodrigues, H.O.; Costa, W.L.S. *Modeling Environmental Dynamics with Dinamica EGO*, 2009. Available online: <http://www.csr.ufmg.br/dinamica/> (accessed on 3 August 2009).
52. Soares-Filho, B.; Alencar, A.; Nepstad, D.; Cerqueira, G.; Vera Diaz, M.; Rivero, S.; Solorzano, L.; Voll, E. Simulating the response of land-cover changes to road paving and governance along a major Amazon highway: The Santarem-Cuiaba corridor. *Glob. Chang. Biol.* **2004**, *10*, 745–764.
53. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
54. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.G.; Gruber, B.; Lafourcade, B.; Leitão, P.J.; *et al.* Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **2013**, *36*, 27–46.
55. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory; ACM, 1992; pp. 144–152. Available online: <http://dl.acm.org/citation.cfm?id=130401> (accessed on 22 February 2014).
56. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 2000.
57. Watanachaturaporn, P.; Arora, M.K.; Varshney, P.K. Multisource classification using support vector machines: An empirical comparison with decision tree and neural network classifiers. *Photogramm. Eng. Remote Sens.* **2008**, *74*, 239–246.
58. Hornik, K.; Meyer, D.; Karatzoglou, A. Support vector machines in R. *J. Stat. Softw.* **2006**, *15*, 1–28.
59. Soares-Filho, B.; Coutinho Cerqueira, G.; Lopes Pennachin, C. DINAMICA: A stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. *Ecol. Model.* **2002**, *154*, 217–235.
60. Yue, W.; Liu, Y.; Fan, P. Measuring urban sprawl and its drivers in large Chinese cities: The case of Hangzhou. *Land Use Policy* **2013**, *31*, 358–370.
61. Meentemeyer, R.; Tang, W.; Dorning, M.; Vogler, J.; Cunniffe, N.; Shoemaker, D. FUTURES: Multilevel simulations of emerging urban-rural landscape structure using a stochastic patch-growing algorithm. *Ann. Assoc. Am. Geogr.* **2013**, *103*, 785–807.

62. Pontius, R.G., Jr.; Malanson, J. Comparison of the structure and accuracy of two land change models. *Int. J. Geogr. Inf. Sci.* **2005**, *19*, 243–265.
63. The State of Land Change Modeling. *Advancing Land Change Modeling: Opportunities and Research Requirements*; The National Academies Press: Washington, DC, USA, 2014.
64. Pontius, R.G., Jr.; Schneider, L. Land-use change model validation by an ROC method. *Agric. Ecosyst. Environ.* **2001**, *85*, 239–248.
65. Mas, J.; Soares Filho, B.; Pontius, R.; Farf'an Guti'errez, M.; Rodrigues, H. A suite of tools for ROC analysis of spatial models. *ISPRS Int. J. Geo-Inf.* **2013**, *2*, 869–887.
66. Pontius, R., Jr.; Parmentier, B. Recommendations for using the relative operating characteristic (ROC). *Landsc. Ecol.* **2014**, *29*, 367–382.
67. Pontius, R., Jr.; Si, K. The total operating characteristic to measure diagnostic ability for multiple thresholds. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 570–583.
68. Wang, N.; Brown, D.G.; An, L.; Yang, S.; Ligmsnn-Zielinsak, A. Comparative performance of logistic regression and survival analysis for detecting spatial predictors of land-use change. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 1960–1982.
69. Rodriguez-Galiano, V.F.; Chica-Olmo, M.; Chica-Rivas, M. Predictive modelling of gold potential with the integration of multisource information based on random forest: A case study on the Rodalquilar area, Southern Spain. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1336–1354.
70. Visser, H.; de Nijs, T. The map comparison kit. *Environ. Modell. Softw.* **2006**, *21*, 346–358.
71. Vliet, J.; Bregt, A.K.; Hagen-Zanker, A. Revisiting Kappa to account for change in the accuracy assessment of land-use change models. *Ecol. Model.* **2011**, *222*, 1367–1375.
72. Pontius, R.G., Jr.; Walker, R.; Yao-Kumah, R.; Arima, E.; Aldrich, S.; Caldas, M.; Vergara, D. Accuracy assessment for a simulation model of Amazonian deforestation. *Ann. Assoc. Am. Geogr.* **2007**, *97*, 677–695.
73. Pontius, R.G., Jr.; Boersma, W.; Castella, J.C.; Clarke, K.; de Nijs, T.; Dietzel, C.; Duan, Z.; Fotsing, E.; Goldstein, N.; Kok, K.; *et al.* Comparing the input, output, and validation maps for several models of land change. *Ann. Regional Sci.* **2008**, *42*, 11–37.
74. Mertens, B.; Lambin, E. Land-cover-change trajectories in Southern Cameroon. *Ann. Assoc. Am. Geogr.* **2000**, *90*, 467–494.
75. Braimoh, A.; Vlek, P. Land-cover change trajectories in Northern Ghana. *Environ. Manag.* **2005**, *36*, 356–373.
76. Dietzel, C.; Clarke, K. The effect of disaggregating land use categories in cellular automata during model calibration and forecasting. *Comput. Environ. Urban Syst.* **2006**, *30*, 78–101.