Modeling Urban Growth using Suitability based Cellular Automata: Case Study for Raipur, Chhattisgarh, India

Breny Kurien
Centre for Environment, Institute of Science and Technology,
Jawaharlal Nehru Technological University,
Hyderabad, Telangana, India
brenykurien007@gmail.com

Keywords: GIS, Cellular Automata, Multi Criteria Decision Making, Urban Growth Model.

Abstract: This paper talks about developing a model which integrates GIS with Cellular Automata and Multi-criteria evaluation process for predicting spatially dynamic phenomena like, urban development, forest fire propagation, deforestation, land degradation, etc. The model is developed as a windows application on the visual studio platform and is programmed in C#. Input data for the model were generated using open source software like QGIS, GRASS. The model developed is used to evaluate urban growth patterns for Raipur city, Chhattisgarh, India by identifying criteria for urban growth and assigning weights according to their importance. Finally, urban growth predictions generated from the model can thus be utilized in planning and mitigating problems related to urban growth. Furthermore, issue related to spatial data availability for the city, limitation of cellular automata based models and observations on results from the case study are discussed.

1 INTRODUCTION

With the rapid population explosion, unplanned urban sprawl has increasingly become a major challenge in many parts of India. There is a need to understand how these areas expand across space and time, with growth in population and socioeconomic development which are often dominated by physical constraints.

Several models are available for urban growth phenomena such as cellular automata (Toffoli, 1987 and Itami, 1988), agent based, spatial-statistics, artificial neural network (Li, X., 2001) to name a few. In this paper, we use a combination cellular automaton(CA)(John V.N.,1966 and John, C.,1982) with the multi-criteria (Jose, M.C.P, 1993) evaluation process to develop a general purpose model to simulate land use change dynamics, depending on the suitability of land and its probability to change. This model can act as a generic framework which can handle spatially dynamic phenomenon's and predict land use changes. The suitability-based cellular automata model is then used to evaluate growth patterns for Raipur city which can, in turn, be helpful in carrying out urban planning activities.

2 STUDY AREA

Raipur, the capital of Chhattisgarh, is a growing city with a population of about 1.4 million, according to 2011 censuses. The figure below shows Chhattisgarh in yellow and Raipur in orange.

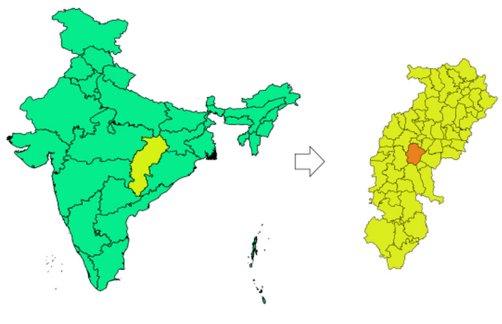

Figure 1: Indian map with Chhattisgarh in yellow and Raipur in orange.

The study area extends over longitude of 81°26′24″E to 81°44′59″E and latitude of 21°09′21″N to 21°22′11″N covering an area of 317.11 square miles.

Three different temporal datasets for Raipur, 2005-06, 2011-12, and 2017 was taken from Bhuvan Indian earth observation website portal. The following are the datasets utilized.

1. Land Use Land Cover (1:50K): 2005-06
2. Land Use Land Cover (1:50K): 2011-12
3. Bhuvan satellite image with 2.5m resolution, 2017.
4. CartoDEM for generating slope with 30m resolution.
5. Open-street map for identifying major roads and railway station.

Using these datasets inputs for the model was generated using QGIS and GRASS. All the inputs to the model were of the size 1116 x 771 pixels with 30m resolution.

3 METHODOLOGY

This section starts out with a basic introduction of cellular automata and works into the concepts of Suitability based Cellular Automata. Next, we discuss the driving factors for urban growth and weights assigned to them. The flow diagram of the model's process flow for computing suitability is shown and ends with a brief description of models development, image libraries used and computation methods.

3.1 Cellular Automata

A cellular automaton model can be described as a system of "cell" objects having the following characteristics.

1. The cells live on a grid. The simplest grid would be a one-dimensional line of cells or a matrix of cells for higher dimensions.
2. Each cell has a state. The simplest example has the two possibilities of 1 and 0 (otherwise referred to as "on" and "off").
3. Each cell has a neighbourhood. This can be defined in any number of ways, but it is typically a list of adjacent cells.
4. The state of the cell changes according to the transition rule matrix we define and local neighbourhood conditions.

Beginning with "The Nature of Code" (by Daniel Shiffmann, 2012) helped to understand the programming part of CA, by explaining 1D & 2D Cellular Automata for 2 states of 0's & 1's. This gave a better understanding of CA and as a base for developing the model.

The model developed carries out CA on three states, "Urbanized", "Can-be urbanized" and "Protected" state. The cell is on a two-dimensional grid with neighbourhood defined as a 3 x 3 matrix surrounding the cell of interest. Only "can-be urbanized" cells can change its state depending on the neighbourhood cell's state and other suitability criteria. The other two cells do not change their state, i.e. an urbanized cell remains urbanized same is the case with the protected state. Protected state includes land under forest, water bodies, and rivers, etc. Transition rules are explained in the next section.

3.2 Suitability-based Cellular Automata

Simulating urban growth patterns is based on the concepts of "development probability" and "development suitability". In this, we assume a relationship between the state of the cell (whether urbanized or not), development probability and development suitability.

$$S^{t+1}\{x, y\} = f\,(P^{t}\{x, y\})$$
$$P\{x, y\} = f'\,(DS^{t}\{x, y\})$$

Where $S\{x, y\}$ is the state at location $\{x, y\}$,

$P\{x, y\}$ is the probability of transition to the state S at the location. $DS\{x, y\}$ is the suitability of conversion to the state S. f and f' are transition functions.

This logic has been extended in the present study to model urban growth dynamics. The CA simulation depends on the calculation of the suitability score based on the state of the cell and neighbourhood configuration. The suitability of a cell to change its state as urbanized is usually evaluated by location factors and site properties taken into consideration. The conversion criterion is that cells with a high degree of suitability will be first selected for urbanization. Land suitability has been computed using multi-criteria evaluation process.

Land suitability describes the potential of a cell for a specific type of land use to urbanize, is an important constraint in the CA model. For example, we may allow faster urbanization of a cell if it's close to a road network and more restricted near to water bodies or near a forest area. Therefore, suitability plays an important role in the transfer of the state of a cell. Suitability scores are re-computed for each year (or iteration) and conversion criteria's are simulated on it to achieve compatible land use.

The transition to the next state and its suitability score expressed as a function of present states S(t) and suitability DS(t):

$$(S^{t+1}, DS^{t+1}) = f\,(S^t, DS^t, N)$$

Where N is the neighbourhood and f is the transition function.

3.3 Land Suitability Analysis

An important problem in GIS is how to efficiently integrate data from various sources. For example, slope and land use are important criteria for land use evaluation, but how to integrate these two different driving factors to generate land suitability. In this case, a weighted linear additive model for data integration was used. If there is a decrease of one unit on one criterion, it can be totally compensated by an equivalent gain on any other criteria i.e. a total compensation between criteria is assumed. To avoid this Compromise Programming technique, Ideal Point Analysis is used to arrive at a non-compensatory solution. This is done by measuring deviation from the ideal point for each layer and applying the min-max rule on the weighted deviation. "The best compromise solution is defined as that which is at the minimum distance from the theoretical ideal."

Moving on we discuss driving factors for urban growth given as different layers to the model. In the paper by Mahar, 2015 is a list of driving factors considered for urban growth for different cities. Broadly speaking the driving factors can be classified into two levels as shown in the table below.

Table 1: Driving factors for urbanization.

Level I	***Level II***
Physical factors	*Slope, elevation etc*
Accessibility factors	*Distance from transport systems*
Social factors	*Population density*
Economic factors	*Distance from economic centres, land value, etc*
Environmental factors	*Vegetation lands, soil types, land cover , etc.*

The factors to be considered as driving layers changes from one city to another based on its spatial data availability and importance. For example, for one city, land value may be an important driving factor and distance from transportation for another.

Depending on availability of spatial data for Raipur, the following layers were considered as driving factors and weights were assigned to them according to their importance. The table below shows the layers along with weights assigned to them for simulation. The weights assigned were in accordance with their relative importance and the ones generating best calibration and validation results were used.

Table 2: Weights given to available layer.

Layer	***Weight given***
Urban seed layer (3 states)	*100*
Land use Land cover(LULC)	*95*
Accessibility	*80*
Slope	*70*

Seed layer was given higher preference because in a cellular automata model the state of the cell would depend on the state of its neighbourhood, i.e. whether there are any urbanized cells in its neighbourhood. Next preference was to Land use land cover layer with crop, scrub and barren land identified. These were likely to get converted for urbanization. The conclusion was made by comparing LULC for 2005 and 2011. Accessibility layer includes major road networks with a buffer length of 50m and railway station with a circular buffer of 1000m in diameter. Last driving factor considered was slope and divided into 5 classes based on degree of slope.

Suitability score DS is computed using the distance metric given below as described in the paper by Novaline, 2008.

$$DS = [\sum_{i=1}^{n} \beta^p (x_i^* - x_{ik})^p]^{\frac{1}{p}} \qquad (1)$$

Where i is the map layer β is the criterion preference or weight, x_i* is the ideal point, x_{ik} is the cell value in k^{th} cell for i^{th} parameter and p is the factor which leads to a non-compromising solution. The value of p can vary from 1 to infinity. In this case, p set at 4. This suitability score is recomputed every year.

Usually, CA simulation is done on binary value's of the state i.e. 0's & 1's. The conversion is based on the calculation of probability, which is given by a transition rule depending on the neighbourhood states. This model works on 3 states, 100's, 50's and 0's i.e. "urbanized" "can-be urbanized" and "protected" states respectively. Only 50's are chosen to change its state to 100's after computing suitability value each year.

In real-world conditions, the best suitable land is not always chosen to urbanize because of time limits and information barriers. Less desirable sites sometimes still have a chance of being urbanized. Thus an unknown error element, a stochastic disturbance term is added during simulation. The suitability values are converted to probability values using this parameter. Thus, it gives the probability of a site selected to change its state. This means land suitability is dynamic in nature as it changes over simulation time. To transform suitability score to probability, maximum score during a simulation is used as a benchmark to represent its relative availability at that given time. The probability is defined in a nonlinear fashion with respect to the evaluation score.

$$P_{xy}^{t} = \exp[\alpha((DS_{xy}^{t} / DS_{\max}^{t}) - 1)] \quad \text{if } DS^{t}xy \neq 0 \qquad (2)$$
$$0 \quad \text{if } DS^{t}xy = 0$$

Where $P^t xy$ is the probability of land conversion from "can-be urbanized" to "urbanized" state at the location x,y at time t, $DS^t xy$ is the land suitability score at the same location at time t, $DS^t max$ is the maximum score of land suitability at the simulation time t which is calculated from equation(1). The higher the value of stochastic disturbance term "α" more stringent is the site selection process. The exponent term in the equation (2) makes α to behave in the required form.

The below example shows how the probability value changes for a suitability value of 0, DS=0.

When α=4, exp (-4) = 0.018
When α=1, exp (-1) = 0.3678
When α=10, exp (-10) = 0.000045(almost 0)

If $DS^t xy = DS^t max$, then exp (0) = 1 i.e. if suitability is score is equal to the highest, then probability of conversion is 1 irrespective of α value.

Therefore from the example above we see that α can take a value between 0 and 10. The model computes changes in the input each year using a growth rate formula. The formula for calculating growth rate is given as shown below.

$$g_t = ((x_t / x_{t-n}) - 1) * 100 \qquad (3)$$

Where g_t is the growth rate in period t, x is the variable being examined and n is the time period of interest was used.

A flowchart describing the model's computation process is given in figure 3.1. The modeling process starts with driver variables and weights given to them. Starting with 2005-06 data we try to simulate the model for next 6 years, and then calibrate it with 2011-12 data. Next, we simulate the model for 12 years and calibrate it with 2017 conditions. The number of urbanized cells generated depends on the weights given to each layer, cut-off probability set and value of stochastic disturbance chosen.

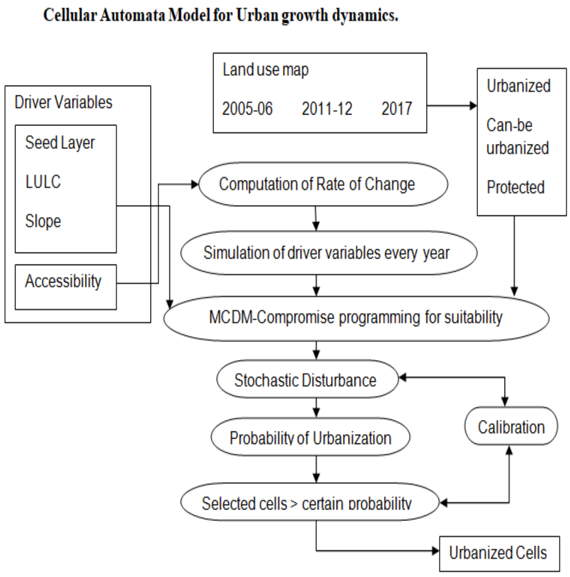

Figure 2: Flow diagram of the computational process.

3.4 Model Development

A windows application (the model) was developed on visual studio framework and programmed in C# language. It uses libtiff.NET libraries to read and write tiff images. Geotiff images from QGIS was converted to Tiff and given as input to the model. The final prediction can also be converted to Geotiff using projection extracted from the original input.

To implement cellular automata lookup table concept was used. Generally, a cell takes on a new value which is computed on the basis of information from the cell's neighbourhood. In this case, we pack the information in the neighbourhood into an integer and use this integer to index a look-up table. This way computation corresponding to the update rule only needs to be done once, at the start of the run and these results can be cached to be used repeatedly while updating the grid. This method has a tendency to use up memory space and thus must be used cautiously.

Parallel for loops was implemented to read the cells line by line, instead of reading them one cell at a time, thus significantly improving the simulation time of the model.

A brief description of the model and its tabs to work with are given below.

- **Driving Variables:**

  This tab is used to load all the driving factors for urban growth phenomena along with their weights, growth rate per year and physical location to the tiff images. The first step is to decide on the total number of layers and click on "Set" button, the model prompts to load all the layers along with its details. "Show all layers" button shows all the layer details in a message box. "Clear all" button removes all the layers entered.

- **Simulation:**

  This is the simulation stage where the user is asked to set the number of years for simulation, cut-off probability, and α value. Different values for cut-off values and α value can be tried here for simulation.

- **Calibration & Validation:**

  This tab has two parts for calibrating and validating the model. The outputs from the simulation are compared with actual data for that year. The result is shown as an image in its respective picture boxes, along with details of correctly predicted, commission and omission errors which are described in the next chapter. Change values of weights assigned to driving factors, cut-off probability, and α value to get best calibration and validation results.

- **Prediction:**

  Once the model is calibrated and validated, all the weights, cut-off probability, and α value are stored and data for the present year condition is loaded into the model. Set the number of years up to predict the future urban growth patterns.

4 MODELING PROCESS AND RESULTS

4.1 Driving Factors and Weights

Before firing up the model decide on the number of driving factors and sequence of the layers in mind. The first layer given to the model is called the seed layer, it has 3 states which are Urbanized (100-pixel value), Can be urbanized (50-pixel value) and Protected (0-pixel value). Protected are the regions where urbanization is not allowed which include water bodies, forest area etc. Can be urbanized cells convert to urbanized state if its probability of transition calculated is more than the set threshold. The figure below shows 4 as the number of layers given to the model and seed layer given with 100 as its weight and growth percentage of 0.

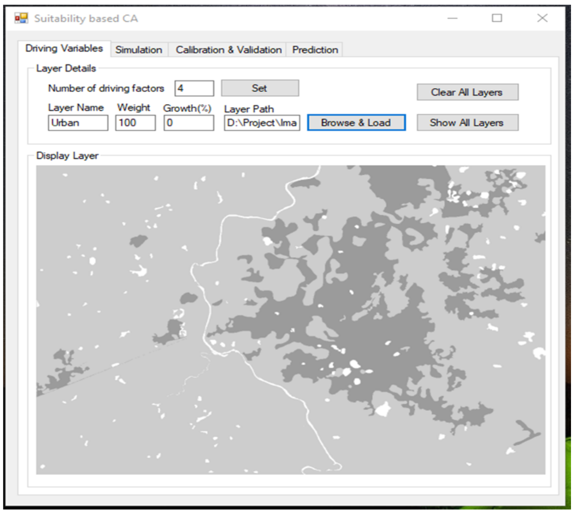

Figure 3: Urban seed layer.

The next layer given is accessibility. It contains all the major road networks and railway stations of Raipur City. The roads were given a buffer of 50m either side and railway stations a circular buffer of diameter 1000m with a pixel value of 70. It is assumed that the transport facilities improve every year with a growth of 5% every year. The figure below shows the layer along with its name, weight of 80 been given to the model.

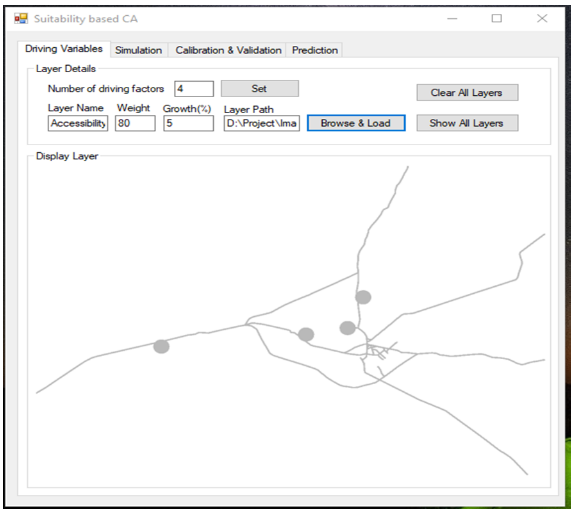

Figure 4: Accessibility layer.

The third layer is labelled as LULC. Three classed have been identified, Crop lands, Barren lands, and Scrub land. These have been given a weight of 50, 30 and 40 respectively. From visual inspection, it was seen that mostly crop lands were being used for urbanization. This layer was given a weight of 95 with 0% growth rate.

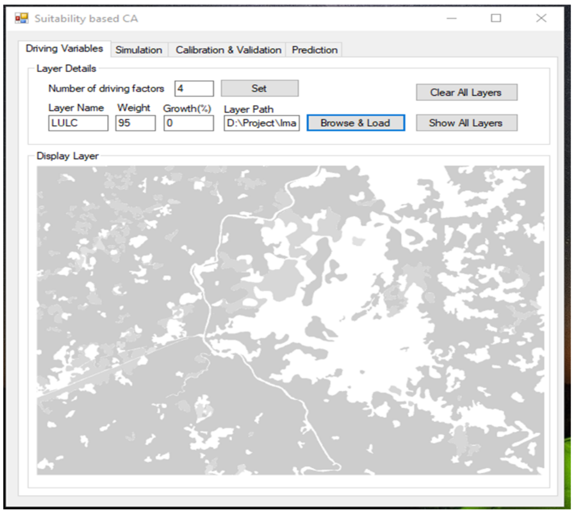

Figure 5: Land use and Land Cover.

The fourth layer is the slope. The slope in degree was generated from DEM for the area of interest and categorized into 5 classes, greater than 3, between 3-6, between 6-9, between 9-12 and greater than 12. Lower the slope more suitable is the land for urbanization. The figure below shows the slope layer, clicking on "Show All Layers" a list of the entire layers name, weight and along with their growth percentage.

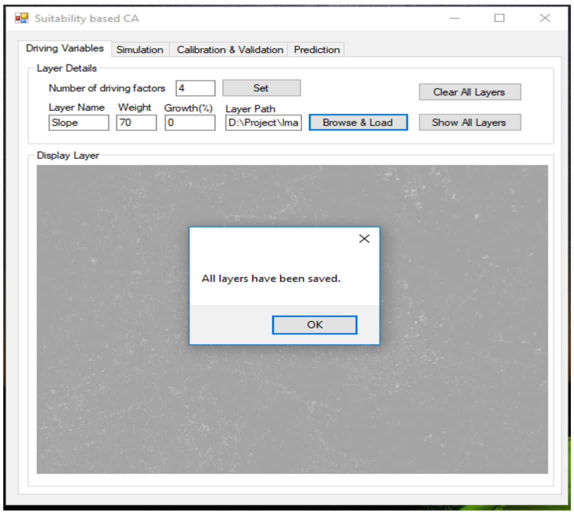

Figure 6: Slope layer.

For each iteration (corresponding to one year), cells beyond certain probability are selected. The threshold value is learned through the calibration process. During the calibration phase, the model was also tested with different values of α and was finally set at 8. The preference value given for the driver variables was also changed and the corresponding results were checked during the calibration phase.

4.2 Calibration and Validation

Simulation is first to run for 6 years with threshold probability set as 0.5 and α as 8 and compared with 2011 actual data.

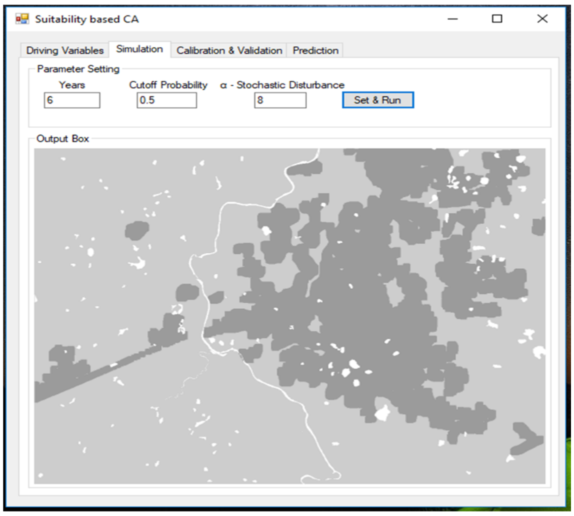

Figure 7: Simulation for 6years.

Next simulation is for 12 years with the same settings and compared with actual 2017 data.

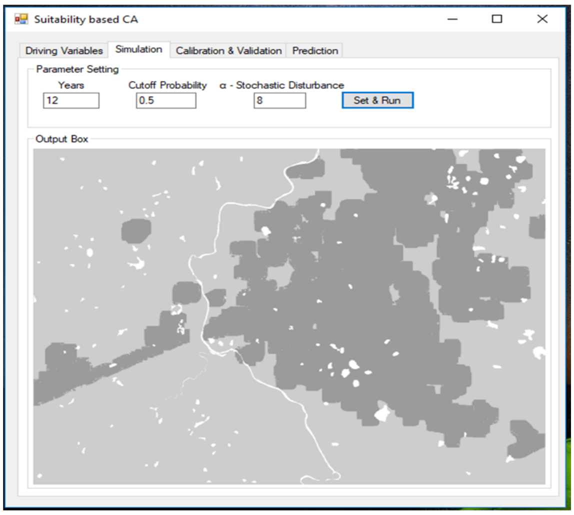

Figure 8: Simulation for 12 years.

The calibration and validation results are shown in the figure below.

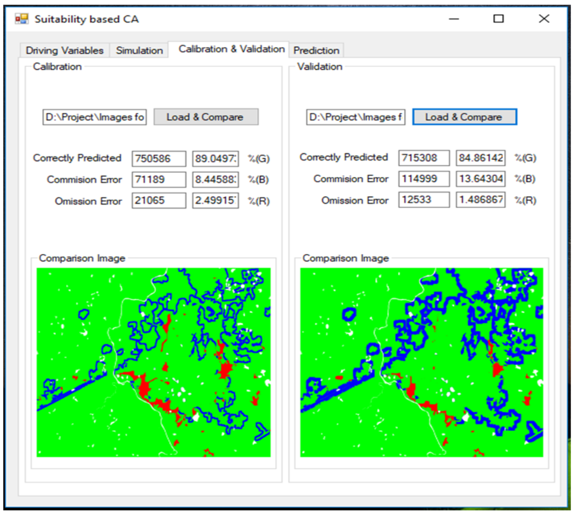

Figure 9: Calibration and Validation results.

For the calibration data set corresponding to the year 2011-12, the percentage of correctly predicted cells is 89.04%. For the validation data set corresponding to the year 2017, the percentage of correctly predicted cells is 84.86% figure-4.9.

Table 3: Calibration & Validation results.

Calibration report for 2011		***Validation report for 2017***	
Correctly predicted cells:	*750586*	*Correctly predicted cells:*	*715308*
Commission error cells:	*71189*	*Commission error cells:*	*114999*
Omission error cells:	*21065*	*Omission error cells:*	*12533*

Correctly predicted cells include urbanized and can-be urbanized cells present in both original land map and the predicted layer denoted by green colour. Commission error indicates cells, which were not found as urbanized in the original land map, but has been predicted as urbanized denoted by blue colour. Omission error indicates cells, which were found as urbanized in the original land map, but has not been predicted as urbanized denoted by red colour.

From the calibration and validation results, we can see that blue portion is more predominant towards the north direction, this is because of cellular automata growth happening in the outwards direction and absence of growth in the actual scenario. The red portion is more predominant towards the southwest direction closer towards the river the model fails to predict this growth. This may be because land closer to water body has higher tendencies to urbanize.

4.3 Prediction

With the above-given weights, the model has been calibrated and validated for Raipur city. Finally using these weights prediction using 2017 as the base year is made for the next 12 years. The seed layer, accessibility layer, LULC and slope for 2017 are given as inputs to the model for prediction. Result for 12-year simulation is shown below on clicking Predict Output button.

Figure 10: Prediction for the year 2029.

Where red is urbanized part, black includes different water bodies like lakes, rivers, and protected forest areas. This prediction is useful in locating most likely areas to be urbanized and hence helpful in allowing urban growth in a planned manner.

5 ISSUES AND CONCLUSION

One of the important observations derived from the model is that it is sensitive to cell size and neighbourhood configurations. This is because CA is represented spatially using regular grids. Though using grid notation has its advantage of saving computation time and simple in terms of implementation, this tends to be a good method when studying complex systems with an aim of understanding its global behaviour. However, for urban planning, we need to work on a local scale and thus grid representation of CA does not represent urban morphology properly.

To address this issue studies on a new kind of prototype for CA-based models for urban growth have started coming up. One such model uses cadastral parcels as cellular space in the paper by Pablo, 2015 and 2017. Using cadastral parcels as basic unit represents urban morphology in a more realistic manner. Thus instead of working on regular cells, the approach is to work on irregular elements or another kind of geographical entities with its representation much closer to reality. Vector-based approach gives a more flexible way of addressing this issue with cells and its neighbourhood definition.

Concluding it can be said that suitability based automata model framework provides a promising environment with the capability to model a variety of spatially dynamic phenomena. This model can be used as a planning tool to test effects of land use change scenarios and its behaviour on a global scale.

ACKNOWLEDGEMENTS

I would like to convey my sincere thanks to my guide **Dr Novaline Jacob**, Scientist "G" & Head DKD ADRIN, Department of Space, Secunderabad and **Dr Vijaya Lakshmi Thatiparthi**, Assistant Professor, Center for Environment, IST and JNTU Hyderabad for their extraneous support and guidance throughout this project.

REFERENCES

Toffoli, T. and Margolus, N., 1987, Cellular Automata: A New Environment for Modeling. *Cambridge, MA: MIT Press.*

Itami, R.M., 1988, Cellular worlds: models for dynamic conceptions of landscape. *Landscape Architecture,* July/August, 52-57.

Li, X., and Yeh, A.G.O., 2001, Calibration of cellular automata by using neural networks for the simulation of complex urban systems. *Environment and planning A, 33, 1445-1462.*

John, V.N., 1966, Theory of Self-Reproducing Automata, *Burks, A.W. (Ed.), Urbana, Illinois: University of Illinois Press*.

John, C., Berlekamp, E., and Guy, R., 1982, Winning Ways, Vol. 2, *New York: Academic Press.*

Gardener, M., 1974, The Game of Life, Time, January 21.

Jose, M.C.P., & Lucien, D., 1993, A Multiple Criteria Decision-Making approach to GIS-based Land Suitability Evaluation. *International Journal of Geographic Information Systems, Vol. 7, No. 5, 407 – 424.*

The Nature of Code by Daniel Shiffmann, Chapter 7. *Cellular Automata. (2012).*

Maher Milad Aburas, Sabrina Ho Abdullah , Mohammad Firuz Ramli, Zulfa Hanan Ash'aari,"A Review of Land Suitability Analysis for Urban Growth by using the GIS-Based Analytic Hierarchy Process", *Asian Journal of Applied Sciences ,Volume 03 – Issue 06, December 2015.*

Novaline Jacob, Krishnan R., Prasada Raju PVSP, Saibaba J ,"Spatial and Dynamic modeling techniques for Land Use Change dynamic study", *Advanced Data Processing Research Institute (ADRIN), Department of Space, Government of India.*

Pablo Barreira González, Montserrat Gómez-Delgado and Francisco-Aguilera Benavente "Vector-based Cellular Automata: exploring new methods of urban growth simulation with cadastral parcels and graph theory." (2015).

Pablo Barreira-González & Joana Barros (2017) Configuring the neighbourhood effect in irregular cellular automata based models, *International Journal of Geographical Information Science, 31:3, 617-636.*