A tool to predict the growth of urban regions based on QGIS/MOLUSCE using MapBiomas image time series

Mario Arthur Sclafani Pujatti[1], Marconi de Arruda Pereira[1], Leonardo Jose Silvestre[2]

[1]Departamento de Tecnologias (DTECH) – Universidade Federal de São João del Rei (UFSJ)
Ouro Branco – MG – Brasil

[2]Departamento de Computação e Eletrônica (DCE/UFES)
Universidade Federal do Espirito Santo (UFES) – São Mateus, ES – Brasil

mario_arthur_spujatti@hotmail.com, marconi@ufsj.edu.br

leonardo.silvestre@ufes.br

Abstract. *This paper presents a method to generate future scenarios of Land-Use and Land-Cover (LULC) classification images by implementing an artificial neural network that can be used to predict urban growth. In this study, LULC data from 1985 to 2020 with annual intervals, obtained through MapBiomas, were used. These data were inserted into a neural network integrated with the MOLUSCE plugin from QGIS to model the possible spatio-temporal changes to simulate the evolution of LULC. MapBiomas is a powerful tool, that uses data from time series from Landsat Satellites and machine learning algorithms to provide reliable products. Our analysis focused on cities that have expanded greatly over the past two decades according to studies made by IBGE. The results obtained were better than those presented in related works, obtaining a kappa value of 0.74 and an accuracy value of at least 80% in all tests performed.*

1. Introduction

Urbanization is one of the most evident global changes. In the last 200 years, while the world population has increased six times, the urban population has multiplied 100 times [Stalker 2000]. In recent decades, with global urbanization, more than half of the world's population lived in cities in 2018, and this proportion is expected to reach 68% by 2050 [UN 2018]. This type of urban expansion and changes in urban land structure has social, economic, ecological and environmental impacts on urban populations and sustainable urban development [Stone et al. 2010]. One of the biggest challenges for global sustainability is the pressure for natural resources driven by intensive urban expansion and economic growth. In this perspective, accelerated urbanization without proper planning is one of the great catalysts contributing to climate change and is an essential factor that compromises ecosystems and their global functionalities [Yang et al. 2020]. As a result, the planning and management in growing urban areas become more complex and difficult. A better understanding of the process of urban growth and the effects of this growth and land use change is required for more efficient planning and management [Leao et al. 2004].

Since the land use and land cover (LULC) change project was launched by the International Geosphere and Biosphere Program (IGBP) and the International Human Dimensions Program (IHDP) on Global Change [Lambin et al. 2000], land use research programs on a global scale have become central to international climate and environmental change research [Liu et al. 2014]. Thematic maps of LULC serves as a reference for scrutinizing, source administration and forecasting, making it easier to establish plans that balance preservation, competing uses, and growth compressions [Kamaraj and Rangarajan 2022].

On another topic, Artificial Intelligence (AI), has had an increasingly important role in our society in almost every field, including among the territory analysis methods [Gomes et al. 2019]. Different AI systems use different algorithms and are suitable for different purposes [Faceli et al. 2011]. LULC dynamics can be typically modeled by methods that have various implementation complexities and efficiencies [Agarwal 2002] such as artificial neural networks (ANN) [Basse et al. 2014a]. Its applications are suitable for urban and rural environments and it can be applied to different settings. LULC dynamics often result from a complex inter-system combination of factors, a non-trivial collective behavior, that cannot be derived from an individual or a simple collection of systemic analyses [Souza et al. 2022].

Besides being an approach directed towards land usage and cover, we will focus on the urban mesh changes. As so, this work presents a method for predicting and simulating structures of the urban mesh that can contribute to planning and analyzing the structure of the urban mesh. By proposing a way of collecting and reclassifying land usage images, we can build a database of spacial variables that can be used along with data from other sources to simulate the expanse of the urban mesh.

Unlike other approaches, we present a method capable of generating good results using only LULC images that represent the evolution of urban occupation as input. On the other hand, the method also allows the use of the digital model of terrain elevation (MDES) and other parameters to seek to improve the results. Therefore it is possible to achieve results even with only two land cover images. Thus, using the Google Earth Engine (GEE)[1] along with MapBiomas[2] to get LULC, our proposed method can perform simulations for the entire extension of the national territory, even where getting data would be complex otherwise. Using the MOLUSCE[3] interface, it is possible to determine the correlation between the variables used, thus defining their importance for the forecast. Perhaps the most relevant point to justify its use is the replicability of the method, its ready-to-use nature allows users without previous experience to train neural networks to simulate future scenarios.

2. Related Works

Making predictions related to LULC is a frequently revisited area of study. An effective approach has immediate consequences in expanding the use of this type of tool. With the recent advances of studies in the area of machine learning large number of tools that can be applied in the most diverse areas of knowledge are emerging. [Souza et al. 2022] developed a simulation and analysis model of LULC change trends for the year 2036, based on artificial neural networks, in the Chapecó River ecological corridor to provide a robust tool to support decision-making, land use planning and sustainable development.

[Han et al. 2015] shows us an application using a Markov model and a Beijing

[1]https://earthengine.google.com/
[2]https://mapbiomas.org
[3]https://plugins.qgis.org/plugins/molusce/

case study to describe the related driving factors from land-adaptive variables, regional spatial variables and socio-economic variables and then simulate future land use scenarios from 2010 to 2020. Beijing has undergone rapid urbanization and economic growth since the economic reforms of 1978, and in 2013, the population was approximately 21.15 million. Within the administrative region, the urban population is 18.25 million, which accounts for 86.31% of permanent residents. The new population will need new space and so creates a higher demand for residential land, thus encouraging rapid expansion of the region's urbanized area. According to the overall land utilization plan in Beijing (2006 – 2020), the amount of reserved cultivated land cannot be less than 2147 km^2 by 2020, which means that the amount of cultivated land converted to built-up land is very limited. Thus, an effective application of the proposed method can serve as a powerful tool for planning.

Another practical use can be seen on [Brovelli et al. 2020] work, which proposes a way of monitoring forest cover in the Amazon, using Multi-Temporal Remote Sensing Data and Machine Learning Classification. Emerging countries often suffer from problems related to the modification of basic infrastructure and its balance with the environment. Deforestation causes diverse and profound consequences for the environment and species. Direct or indirect effects can be related to climate change, biodiversity loss, soil erosion, floods, landslides, etc. As such a significant process, timely and continuous monitoring of forest dynamics is important, to constantly follow existing policies and develop new mitigation measures. Using the random forest algorithm as well as GEE a machine learning classification of multispectral satellite imagery on cloud computing service was implemented.

[de Brito et al. 2021] had a similar approach to ours but used some different tools. Also using the MapBiomas database, it aimed to implement a cellular automata-based model using MapBiomas data series as input to develop LULC future scenarios focusing on human activities drivers in a vital watershed for Paraiba state's water security. Using a similar process of imaging and reclassification, our study differs mainly in using MOLUSCE compared to the SIMLANDER chosen by the authors. The effectiveness of our method is evident when the first pre-calibration tests performed by De Brito et al. indicated the accuracy of approximately 56% between the 2015 simulated data and the MapBiomas reference date and after calibration reached an average accuracy of 84.41% for the year 2004, both results lower than those achieved by our process.

3. Methodology

Designing our method required using open source or free services and tools for research purposes. The Brazilian Annual LULC Mapping Project (MapBiomas) is an initiative that involves a collaborative network of biomes, land use, remote sensing, GIS, and computer science experts. The GEE was used together with the MapBiomas project toolkit as a LULC database tool, as well as a way to download them.

We can divide the proposed model for predicting urban growth into three steps. First, the pre-processing step prepares and processes the raw data obtained through MapBiomas, undergoing a reclassification process that is proven to be more effective for the final simulation. The pre-processed data is the model's input. The second step, calibration, has the main goal of establishing the spatial variables for the network training. It means interpreting which raster types besides LULC have relevant influence to improve the final simulation, as well as discover the optimal values for each parameter used. Finally, the third step intends to validate the results. Thus, Cohen's kappa method is used to compare the generated maps in simulations/predictions with real data from MapBiomas. Figure 1 shows the process.

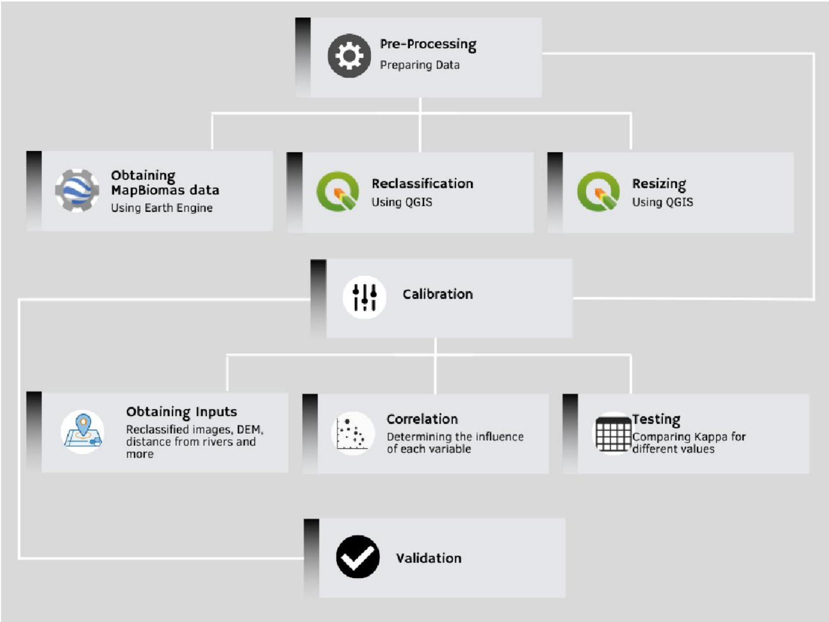

Figure 1. Steps of the proposed method for predicting urban growth

3.1. Pre-processing

Initially, we explored the MapBiomas toolkit for GEE, because it allows visualizing a variety of different land use classifications. The MapBiomas project relies on the GEE platform and its cloud processing and automated classifiers capabilities to generate Brazil's annual land use and land cover time series from 1985 to 2020 (last available collection). GEE combines a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities. Scientists, researchers and developers use GEE to detect changes, map trends, and quantify differences on the Earth's surface. GEE is now available for commercial use, and remains free for academic and research use. The MapBiomas data uses the time series from Landsat[4] satellites. The entire process is based on machine learning algorithms providing highly reliable products for the entire territorial extension of the country in a free and accessible way. The spatial resolution of the available satellite images mosaics for each Brazilian biome is 30 m. The mosaics are a composition of pixels in each set of images for a specific period (e.g. filtering the clouds). The periods of the year in which the images are selected vary by region (e.g. wet season in the caatinga biome). We focused on the Brazil subset 6.0 collection. However, MapBiomas' classification method counts almost 50 different classes to represent real-world data as truthfully as possible (Figure 2). As [Congalton 1991] assert, the use of more general categories can be essential when trying to meet predetermined accuracy standards,

[4]https://landsat.gsfc.nasa.gov/

and after the first batch of testing, it was confirmed that using LULC with a large number of classes can be harmful to the network learning, and hence to the final simulations.

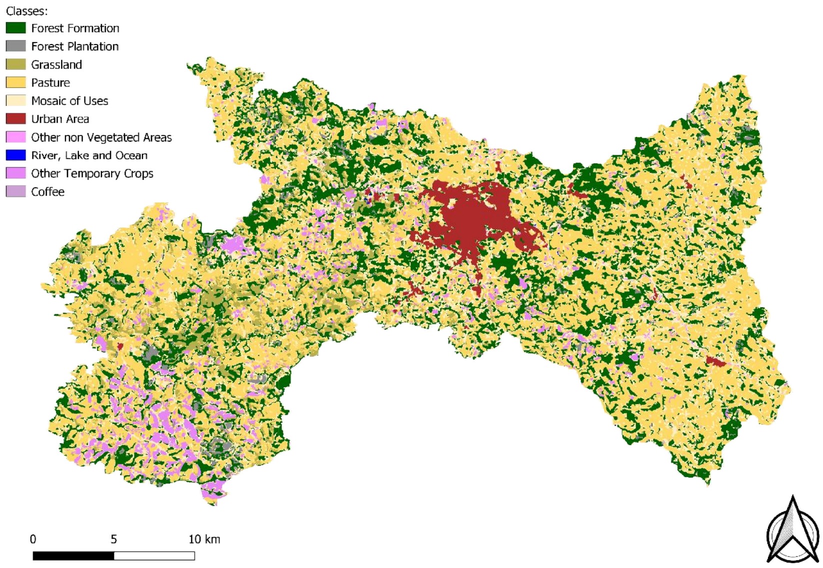

Figure 2. Original Class Distribution

Thereby we recognized the need for a LULC reclassification. Initially, it was considered using only two classes (urban and non-urban), seeing that our analysis is only interested in urban mesh changes, but after the second round of experimentation, it was proven that a too simple reclassification is also damaging on the final simulations.

After the experiments, we concluded that the best option would be to organize the 50 classes into a set of 6 superclasses that successfully encompass the characteristics represented in each of the divisions previously presented. And again, tests were performed showing significant improvement in the proposed simulations. The reclassification processs is shown at Figure 3. All forest subtypes were reclassified as forest (ID 1). The same procedure was adopted for the following classes, except for the non-vegetated area: here the urban area continued to be detached from the remaining non-vegetated area. The result of the reclassification process can be seen in Figure 4.

According to Exame[5] magazine, between 2003 and 2013, medium-sized cities grew faster than any metropolis in the country, so we directed our study to those cities that had the highest population growth in Brazil in this interval, as shown on Table 1. We understand that there is a correlation between population growth and expansion of the urban fabric, and according to studies carried out by [Seto et al. 2011], urban land expansion rates are higher than or equal to urban population growth rates, suggesting that urban growth is becoming more expansive than compact. It allows us to conclude that we made an adequate choice of cities to analyze. It was necessary to use GEE to download the LULC of each city because it allows us to filter the variables according to the area of interest in addition to having methods that can be used to download any

[5]https://exame.com/brasil/25-cidades-que-sofreram-um-boom-populacional-no-brasil/

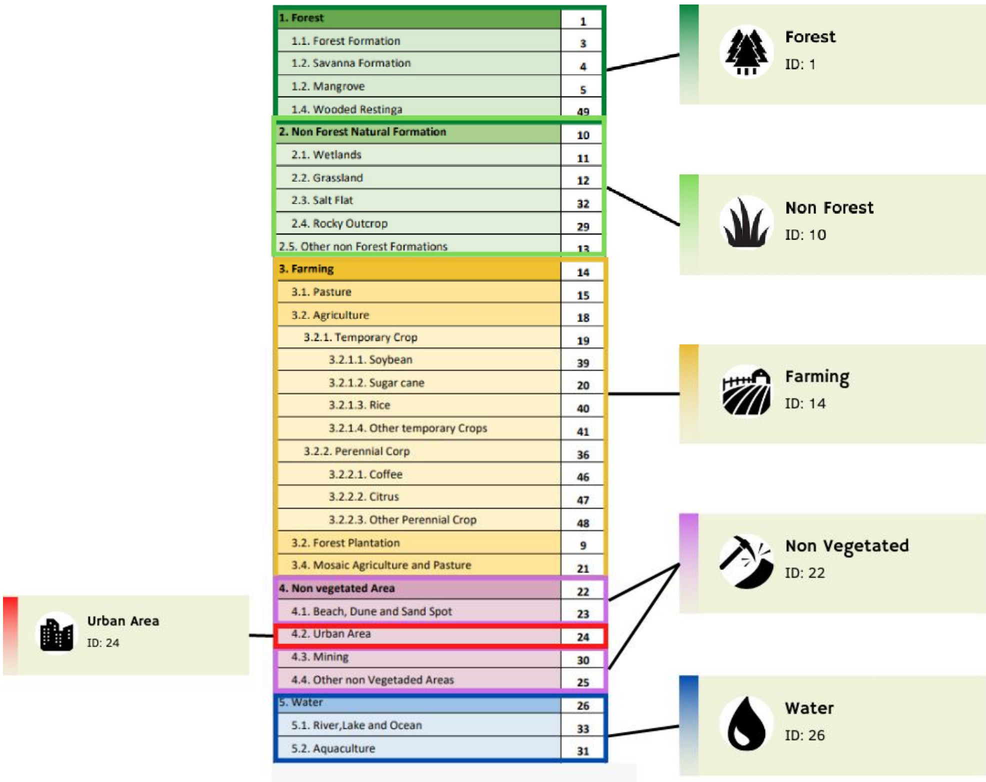

Figure 3. Reclassification Process

variable individually when using the MapBiomas toolkit. So we were able to use these methods to import variables already classified and cropped to Google Drive, enabling the download. Therefore we obtained a vast collection that by itself can potentially be employed for simulating and predicting.

For constructing the land use image database, each of the selected cities had the 36 raw annual images obtained through the MapBiomas toolkit, that were reclassified afterward. However, it was observed in certain cases, a large difference between the extension of the urban mesh and the municipal perimeter, where the perimeter has significantly higher proportions than the urban extension itself. That is, it would be included relatively insignificant data for the study in the analysis. Thus, in these cases, it was necessary to resize the study area, finishing the pre-processing step.

3.2. Calibration

The calibration step consists of analyzing and selecting the variables that will be used for network training. We used QGIS and an add-on called MOLUSCE (Modules for Land Use Change Simulations) to carry out the predictions and simulations. QGIS is open-source, free software. It is a system of geographic information that permits visualizing, editing, and analyzing geo-referenced data, but also executing a variety of helpful algorithms on geoprocessing [Guimarães et al. 2021]. QGIS tools, especially its plugins, are in constant development by a great range of geotechnological sectors. MOLUSCE is one of the community-made plugins available from QGIS menus and provides a set of algo-

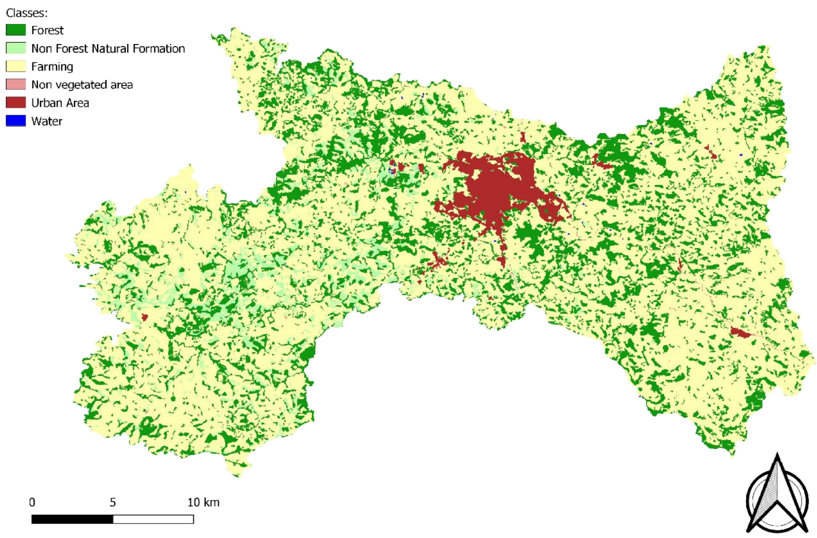

Figure 4. Reclassification Result

	Population Growth	Inicial Population	Final Population
Rio das Ostras	190%	42.024	122.196
Parauapebas	116%	81.428	176.582
Maricá	62%	86.038	139.552
Barcarena	60%	68.604	109.975
Parnamirim	59%	143.598	229.414
Rio Verde	57%	124.753	197.048
Macaé	55%	144.207	275.575
Palmas	49%	172.176	257.903
Lauro de Freitas	45%	127.182	184.383
São José de Ribamar	43%	118.725	170.423

Table 1. Population Growth between 2003 and 2013

rithms for land use change simulations such as neural network, linear regression, among others. MOLUSCE presents a user-friendly interface and has a ready-to-use proposal. It was developed to investigate a range of applications, including studying temporal LULC shifts and projecting future land use, anticipating prospective shifts in land cover and forest cover, and detecting deforestation in sensitive locations[Aneesha Satya et al. 2020].

To choose the model that we would use to perform the simulations, several options were considered. It can be said that every model has its own specialty for addressing the composite issues of LULC. Among these models, cellular automata (CA) are common approaches to simulate LULC and can effectively represent nonlinear spatially stochastic land-use change processes [Batty et al. 1997]. CA are powerful approaches for understanding land-use systems and their integral dynamics [Wu 2002], especially when integrated with other tools, such as ANNs [Basse et al. 2014b]. The Cellular Automata-Artificial Neural Network (CA-ANN) works on what-if scenarios. Therefore, it can

	Year: 1986	Year: 1987	Year: 1988	Year: 1989
Year: 1986	–	0.97522	0.95814	0.94643
Year: 1987	–	–	0.97861	0.95761
Year: 1988	–	–	–	0.97498
Year: 1989	–	–	–	–

Table 2. Pearson's Coeficient

be useful for planning [Araya and Cabral 2010] and land-use change simulation studies [Pahlavani et al. 2017]. The Cellular Automata model employs the transition probabilities from the ANN learning process to describe the LULC changes.

The first step for modeling is to insert the LULCs for the starting date and end of the analysis, in addition to any other spatial variables that you want to use, such as MDES, distance from streets, and other intermediate LULC variables. The plugin compares the geometries of the inserted rasters to get a direct relationship between everything that was entered. Besides correlating the extension and pixel size data, the plugin relies on Pearson's coefficient to simulate a linear correlation between the variables chosen. According to its definition, Pearson's correlation coefficient is the ratio between the covariance of two variables and the product of their standard deviations. Thus, it is essentially a normalized measurement of the covariance, such that the result always has a value between -1 and 1. As the covariance itself, the measure can only reflect a linear correlation of variables and ignores many other types of relationships or correlations. Thus, as illustrated on Table 2, we have a point value that indicates how theoretically relevant each of the variables represents for training and simulation.

After the preliminary tests, we could see that the intermediate LULCs by themselves have a great impact on training, as well as the MDE, the distance from streets, and the distance from rivers. Having calculated Pearson's coefficient, we can proceed to the next step of the process, which is to generate the transition matrix. The transition matrix plays an essential role in analyzing temporal changes within a set of LULC categories. The matrix represents the proportions of pixels changing from one land use category to another. Its rows represent the categories in the initial year, while the columns indicate the same order of LULC categories in the final year. The diagonal entries indicate the size of class stability, and each off-diagonal entry represents the size of the transition from one class to a different class. Values close to 1 in diagonal entries represent the stability of a category. Researchers mostly use transition matrices to compare the temporal changes in different regions.

After creating the matrices, we can generate a map that represents the changes identified in each of the classes from the transition matrix. For each original LULC class, a new class will be generated, representing the change from one state to another. Therefore, we have that in an example of 4 classes we will have 4×4 possible transitions. Each of these transitions receives a code that identifies the change made. The output from the area analysis stage is the change map, indicated on Figure 5, where each pixel is an integer value that indicates its transition.

The next step is to choose the model that will be used to make the forecasts. A CA-ANN model in MOLUSCE is a reliable tool for predicting future LULC that may

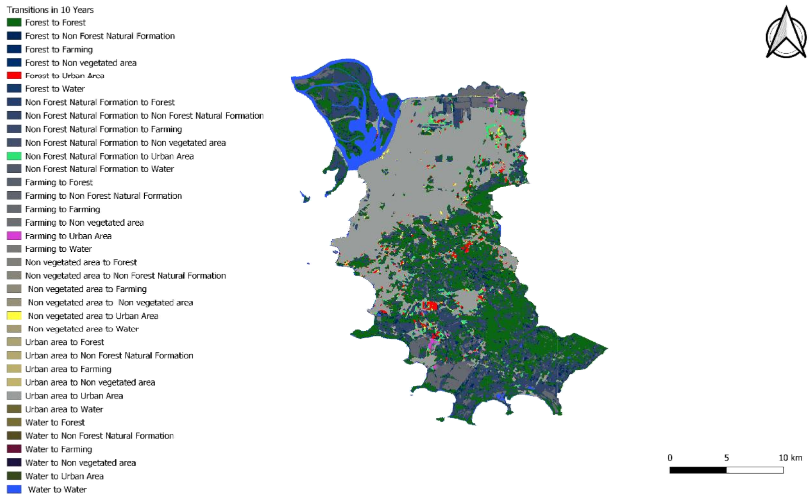

Figure 5. Changes Map

be used in land use planning and management. This approach is being used for predicting the spatial LULC shift because it estimates the pixel's current condition based on its initial situation, adjacent neighborhood eventuality and changeover laws. Moreover, this accurately depicts nonlinear spatial stochastic LULC change processes and produces complex patterns [Saputra and Lee 2019]. The ANN assumes 5 input parameters in addition to the constants that can only be accessed directly in the source code, and their calibration represents an important step to ensure that we can get the best possible result from the simulations. The Neighbourhood variable sets the count of neighboring pixels around the current pixel. Size 1 means 9 pixels (3x3). Learning rate, momentum and max iterations number define learning parameters. The higher the learning rate and momentum, the faster the learning, but the process can be unstable. Hidden Layers takes a list of numbers: N^1 N^2 .... N^k, where N^1 is the number of neurons in the first hidden layer, N^2 in the second, and so on. After several tests, we used the following parameters: Neighbourhood: 5px; Learning Rate: 0.100; Maximum Iterations: 1000; Momentum: 0.050; Hidden Layers: 10.

3.3. Validation

Finally, after training the network, we can generate the simulation. The time interval for the generated simulations is the same interval defined when selecting the inicial and final LULC.. That is, if we use a 10 years interval (Initial = 2000 and final = 2010) each iteration of the simulation will also represent an interval of 10 years. The generated LULC maintains the same classes previously represented in the same representation scheme. It is important that we can validate the results of the simulations, and MOLUSCE has the last step that applies the concept of Cohen's kappa, which serves as a point indicator of performance. Cohen's kappa coefficient is a statistic used to measure inter-rater reliability for qualitative items. Is generally considered a more robust measure than the simple percentage agreement calculation, as it takes into account the possibility of agreement occurring by chance. Therefore, a simple and effective way to measure the reliability of the method was to perform the simulations for the year 2020 and compare them with the

	Kappa Coeficient
Barbacena	0.82748
Barcarena	0.83193
Lauro de Freitas	0.78486
Palmas	0.84707
Parauapebas	0.81884
Parnamirim	0.75129
Rio das Ostras	0.86332
Rio Verde	0.85729

Table 3. Simulation Results Validation

maps already produced by MapBiomas, calculating the kappa value for the comparison.

We used an extensive testing method to evaluate what would be the best way to generate the predictions. So, we iterate through pre-defined values for each of the variables, testing all combinations for them. In this way, we evaluated the values that generated a higher kappa with a reasonable training time. Another important point was to repeat the same process for all the cities previously selected, showing that the method is not biased for a specific case.

The proposed method obtained good results in the predictions, as can be seen in Table 3. The lowest kappa achieved by all tests in the 10 cities was 0.75. With the model well established, we can finally generate simulated maps for any future year that is interesting, showing a reliable way to create a basis for geo-economic studies.

To reach one more stage of validation, we applied the proposed method in situations already analyzed by other authors. Then, we compared the results directly, according to the kappa achieved in the validation, which shows similar or better results even with the use of a smaller number of spatial variables. It is evident that the proposed model is effective in simulating future scenarios. Thus, together with the calibration of the algorithm used, we achieved the aim of presenting an easily replicable method of great informational value. A clear example can be observed when we analyze the study carried out by [de Brito et al. 2021] which shows lower Kappa values in all iterations performed, as seen in Table 4.

4. Conclusions

It is undeniable that studies in geo-computing depend on countless variables for an authentic representation of real situations. Even though each forecast relates directly to the spatial variables used, the implementation we used proposes a simple way of obtaining and preparing spatial variables for the entire extension of Brazil.

The proposed process went through several stages of improvement so that the results were the best, considering an implementation simple to be replicated. First, the use of over ten cities in all test steps contributed to the notion that it did not bias the method for certain regions. The experiments carried out concerning the clustering and reclassification of land use were a step to ensure a distribution with enough classes to achieve a satisfactory result. Testing three different classification types, the original with almost 50 classes, a simplistic reclassification with only two classes, and finally, the set

Upper Paraíba River Watershed	Debrito's Kappa	Achieved Kappa
Year: 2005	0.748	0.759
Year: 2006	0.709	0.773
Year: 2007	0.666	0.775
Year: 2008	0.623	0.766
Year: 2009	0.597	0.742
Year: 2010	0.562	0.786
Year: 2011	0.546	0.808
Year: 2012	0.523	0.801
Year: 2013	0.513	0.829
Year: 2014	0.518	0.796
Year: 2015	0.515	0.783
Year: 2016	0.533	0.796
Year: 2017	0.535	0.812
Year: 2018	0.544	0.786

Table 4. Methods Comparison

of 6 superclasses, for all the proposed cities produced a range of results large enough to decide on the best representation to use. A similar method was also used to define the values of the variables used by MOLUSCE to produce a realistic simulation with a viable network training time. In a trial-and-error-like approach, we tested several values for each of the individual variables for later exchanging with each other, thus testing all combinations of pre-selected values.

The tool is available at: `https://github.com/marioasp/sim_expansao_urb`

Acknowledgements

The authors thank CNPq and FAPEMIG (project APQ-00718-21) for the financial support.

References

Agarwal, C. (2002). A review and assessment of land-use change models: dynamics of space, time, and human choice.

Aneesha Satya, B., Shashi, M., and Deva, P. (2020). Future land use land cover scenario simulation using open source gis for the city of warangal, telangana, india. *Applied Geomatics*, 12(3):281–290.

Araya, Y. H. and Cabral, P. (2010). Analysis and modeling of urban land cover change in setúbal and sesimbra, portugal. *Remote Sensing*, 2(6):1549–1563.

Basse, R. M., Omrani, H., Charif, O., Gerber, P., and Bodis, K. (2014a). Land use changes modelling using advanced methods: Cellular automata and artificial neural networks. the spatial and explicit representation of land cover dynamics at the cross-border region scale. *Applied Geography*, 53:160 â 171.

Basse, R. M., Omrani, H., Charif, O., Gerber, P., and Bódis, K. (2014b). Land use changes modelling using advanced methods: Cellular automata and artificial neural networks. the spatial and explicit representation of land cover dynamics at the cross-border region scale. *Applied Geography*, 53:160–171.

Batty, M., Couclelis, H., and Eichen, M. (1997). Urban systems as cellular automata.

Brovelli, M. A., Sun, Y., and Yordanov, V. (2020). Monitoring forest change in the amazon using multi-temporal remote sensing data and machine learning classification on google earth engine. *ISPRS International Journal of Geo-Information*, 9(10):580.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of environment*, 37(1):35–46.

de Brito, H. C., Rufino, I. A. A., and Djordjević, S. (2021). Cellular automata predictive model for man-made environment growth in a brazilian semi-arid watershed. *Environmental Monitoring and Assessment*, 193(6):1–19.

Faceli, K., Lorena, A. C., Gama, J., and Carvalho, A. C. P. d. L. F. d. (2011). Inteligência artificial: uma abordagem de aprendizado de máquina.

Gomes, E., Banos, A., Abrantes, P., Rocha, J., Kristensen, S. B. P., and Busck, A. (2019). Agricultural land fragmentation analysis in a peri-urban context: From the past into the future. *Ecological Indicators*, 97:380–388.

Guimarães, P. V. D., de Arruda Pereira, M., da Costa Teixeira, E. K., Júnior, C. A. D., Arthur, M., and Pujatti, S. (2021). A framework for the generation of the rainwater flow model in streets. XXII Brazilian Symposium on Geoinformatics, GEOINFO2021.

Han, H., Yang, C., and Song, J. (2015). Scenario simulation and the prediction of land use and land cover change in beijing, china. *Sustainability*, 7(4):4260–4279.

Kamaraj, M. and Rangarajan, S. (2022). Predicting the future land use and land cover changes for bhavani basin, tamil nadu, india, using qgis molusce plugin. *Environmental Science and Pollution Research*, pages 1–12.

Lambin, E., Baulies, X., Bockstael, N., et al. (2000). Land-useand land-cover change, implementation-strategy. *IGBP Report No*, 48.

Leao, S., Bishop, I., and Evans, D. (2004). Simulating urban growth in a developing nationâs region using a cellular automata-based model. *Journal of urban planning and development*, 130(3):145–158.

Liu, J., Kuang, W., Zhang, Z., Xu, X., Qin, Y., Ning, J., Zhou, W., Zhang, S., Li, R., Yan, C., et al. (2014). Spatiotemporal characteristics, patterns, and causes of land-use changes in china since the late 1980s. *Journal of Geographical sciences*, 24(2):195–210.

Pahlavani, P., Askarian Omran, H., and Bigdeli, B. (2017). A multiple land use change model based on artificial neural network, markov chain, and multi objective land allocation. *Earth Observation and Geomatics Engineering*, 1(2):82–99.

Saputra, M. H. and Lee, H. S. (2019). Prediction of land use and land cover changes for north sumatra, indonesia, using an artificial-neural-network-based cellular automaton. *Sustainability*, 11(11):3024.

Seto, K. C., Fragkias, M., Güneralp, B., and Reilly, M. K. (2011). A meta-analysis of global urban land expansion. *PloS one*, 6(8):e23777.

Souza, J. M. d., Morgado, P., Costa, E. M. d., and Vianna, L. F. d. N. (2022). Modeling of land use and land cover (lulc) change based on artificial neural networks for the chapecó river ecological corridor, santa catarina/brazil. *Sustainability*, 14(7):4038.

Stalker, P. (2000). *Handbook of the World*. Oxford University Press, USA.

Stone, B., Hess, J. J., and Frumkin, H. (2010). Urban form and extreme heat events: are sprawling cities more vulnerable to climate change than compact cities? *Environmental health perspectives*, 118(10):1425–1428.

UN (2018). 68% of the world population projected to live in urban areas by 2050. pages https://www.un.org/development/desa/en/news/population/2018–revision–of–world–urbanization–prospects.html. Accessed: 2022-06-26.

Wu, F. (2002). Calibration of stochastic cellular automata: the application to rural-urban land conversions. *International journal of geographical information science*, 16(8):795–818.

Yang, J., Gong, J., Tang, W., and Liu, C. (2020). Patch-based cellular automata model of urban growth simulation: Integrating feedback between quantitative composition and spatial configuration. *Computers, Environment and Urban Systems*, 79:101402.