# Past Expansion and Future Prediction of Land Use and Land Cover of Sofia City

Darin Pavlov[1, 2], Lidia Lazarova Vitanova [1], Dessislava Petrova-Antonova [1]

[1] GATE Institute, Sofia University "St. Kliment Ohridski", 1164 Sofia, Bulgaria
{darin.pavlov, lidia.vitanova, dessislava.petrova}@gate-ai.eu

[2] Faculty of Technology, Policy and Management, TU Delft, Delft, the Netherlands
{pavlovdarin01}@gmail.com

**Keywords:** Land Use and Land Cover, Urbanization, Urban Change Prediction, Artificial Neural Network, Cellular Automata

**Abstract**

Rapid urbanisation impacts land use and land cover (LULC), contributing to environmental challenges such as urban heat islands (UHIs) and air pollution. This study investigates the morphological changes in Sofia's LULC from 1990 to 2018, predicts future LULC patterns, and evaluates the effectiveness of Cellular Automata-Artificial Neural Network (CA-ANN) models for LULC forecasting. Historical LULC data from Urban Atlas (UA) and Corine Land Cover (CLC) and predictive variable data such as population density, road networks, water bodies, and elevation are used to model transition potential and simulate future land cover. The analysis revealed a slight increase in urban areas, primarily at the expense of cropland, between 1990-2018. Simulations for 2074 suggest a continued urban expansion, with a significant cropland decline. Validation of CA-ANN models showed high accuracy but limited ability to predict small-scale transitions due to low transition potential. This study highlights the importance of input data quality and temporal range in predictive accuracy. Furthermore, it provides valuable insights for urban planning, sustainable development, and climate adaptation strategies by offering a data-driven approach to forecasting LULC changes. Future research should integrate additional socio-economic factors and alternative approaches to enhance prediction reliability.

## 1. Introduction

Rapid urbanisation leads to changes in land use and land cover (LULC), adversely affecting the urban environment (Nuissl & Siedentop, 2021; Orlov et al., 2023). In the context of global climate anomalies, cities around the world face significant challenges, leading to phenomena such as urban heat islands (UHIs) (Vitanova et al., 2021; Vitanova & Kusaka, 2018), heatwaves and the prevalence of air pollution (Shen et al., 2022; Wang, 2018). To address these issues, it is essential to develop informed urban planning processes accounting for potential risks.

LULC analysis is an essential field in environmental and urban studies (Solomon & Lukas, 2024). It is used to categorise the Earth's surface by how it is utilised (land use) and the natural or modified physical state of that surface (land cover) (Sudhakar & Rao, 2010). LULC information is used in diverse applications, including urban planning (Xie & Sun, 2021), ecological conservation (Cunha et al., 2021), resource management, and environmental impact assessment (Solomon & Lukas, 2024). By monitoring changes in LULC, researchers and policymakers can understand urban expansion, agricultural growth, and ecosystem shifts - insights critical for sustainable management and development (Cunha et al., 2021). Predictive models, like Cellular Automata (CA), are successfully applied to LULC forecasting by capturing temporal dynamics and spatial patterns, though often constrained by historical data and computational challenges (Ghosh et al., 2017; Li & Li, 2015; Santé et al., 2010).

Although the dynamics of LULC are analysed on a European level (European Environment Agency, 2019), research on the application of LULC analysis remains underexplored in the context of the city of Sofia, Bulgaria. The General Urban Development Plan of Sofia, for example, considers the importance of sustainability but lacks an explicit discussion of how urban growth affects it (Slaev & Nedovic-Budic, 2017). The absence of such an analysis requires categorising the Earth's surface according to its utilisation or physical state to understand how the morphological structure has changed over the past and how it will change in the future (Sudhakar & Rao, 2010).

The goal of this study is firstly to provide insights into the patterns of morphological change of Sofia's LULC between 1990 and 2018, secondly to predict future LULC, and thirdly to assess the effectiveness of the CA-ANN models in predicting future development.

The first aim is achieved by analysing the change in LULC classes between 1990 and 2018 and investigating the transitions between classes. The second goal is realised by predicting the transition potential of the LULC of Sofia. A standard method for such problems is to apply an algorithm to predict the transition potential of each point of the region of interest in combination with an algorithm that also considers the relative position of the fact, such as Cellular Automata (Santé et al., 2010). Using Artificial Neural Networks (ANN) has shown promising results in estimating the transition potential (Li & Li, 2015). The simulation results are finally validated, which provides results for the third aim.

Hence, the following research questions are tackled:
*1) How did the LULC of Sofia City change between 1990 and 2018?*
*2) What will be the impact of urbanisation on the LULC of Sofia City in 2078?*
*3)How practical is the CA-ANN approach in predicting future LULC in Sofia City?*

The findings of this research could contribute to more informed decision-making in urban planning and prevention of climate change effects on human well-being and the environment. Moreover, they can serve as a stepping stone for further research on the urban climate, UHI, thermal environment and mobility.

The paper is structured as follows. Section 2 outlines the methodology used. The experimental results are described in section 3. A discussion and conclusion are given in sections 4 and 5, respectively.

## 2. Methodology

The structure of the methodology is as follows: the first step is data collection, which involves obtaining historical LULC and predictive variable data. Then, the data is prepared into maps used in subsequent analysis. This is achieved by exploring the accuracy of the LULC data, reclassifying it into broad categories and transforming the four supporting datasets. In addition, all maps are standardised to ensure consistency. The data processing included correlation analysis and area change detection to understand past and current LULC developments. This is followed by transition potential modelling using an ANN, the output of which serves as an input to a CA algorithm. First, the CA generates a simulation of a LULC map that can be compared to an actual LULC map of the same year for validation. If the results are satisfactory (< 85 %), the algorithm is used to simulate a LULC map in the far future (the 2070s). This methodology is applied in 4 different experiments. The QGIS environment (QGIS Development Team, 2024) is used for data preparation and processing. Simulations are produced through the MOLUSCE plug-in (NextGIS, 2024).

### 2.1 Study Area

This study examines Sofia, Bulgaria's capital, and its surrounding region. Since 2000, the city's population has experienced a steady increase, reaching nearly 1.3 million inhabitants as of 2024 (World Population Review, 2024), which accounts for approximately one-fifth of the country's total population. Geographically, Sofia is situated within the Sofia Valley, which supports agricultural activities in the adjacent areas. The city is encircled by mountainous terrain, where forests constitute the predominant land use and land cover (LULC) category. The diverse landscape, encompassing variations in land use, natural cover, elevation, and population distribution, renders the region well-suited for the objectives of this study.

### 2.2 Data Collection

The study utilises historical LULC data alongside predictive variables encompassing population density, road networks, water bodies, and elevation metrics. More information for each type of data is provided below.

**Corine Land Cover (CLC):** The CLC is raster data with 100m resolution in the EPSG:4326 projection, derived by classifying satellite images (Copernicus Land Monitoring Service, 2024a). Each data contains 48 classes of LULC. All available years (1990, 2000, 2006, 2012 and 2018) are analysed.

**Urban Atlas (UA):** The UA is vector data, which provides a high level of detail (Copernicus Land Monitoring Service, 2024b). The available data spans only three years (2006, 2012, and 2018), constraining its applicability for change analysis. The UA dataset assesses the input data's influence on the simulation and evaluates its accuracy compared to the CLC. Preliminary validation indicates that the classification accuracy is superior.

**Water bodies:** Water bodies and wetness data are provided by Copernicus (Copernicus Land Monitoring Service, 2024c). Similarly to CLC ones, the data is in raster format with 100m resolution in the EPSG:4326 projection. It derives proximity to water bodies as a supporting predictor of LULC change.

**Population density:** Population density data is derived from the Google Earth Engine using a script that shows the population density for a specific year and region (Schiavina M. et al., 2023). For this case, the 2018 data is taken as a raster file in the EPSG: 32634 projections with the resolution set to 100m.

**Road network:** The road network data is obtained by the TomTom Traffic Stats (TomTom, 2024). The data includes a road network dataset of eight functional road classes to categorise segments based on their functional importance.

**Elevation:** The DEM is a tool for generating slope maps within the specified area of interest (Copernicus Land Monitoring Service, 2024c). It is provided as a raster dataset with a resolution of 90 meters, adhering to the EPSG:4326 projection standard.

### 2.3 Data preparation

The accuracy of historical LULC data is explored before being transformed. A few instances of misclassification are identified during the initial validation process in the CLC data. The misclassifications are considered during the reclassification step. Additionally, CLC is simplified into five broad classes: urban, cropland, forests, bare land, and water (Figure 1).

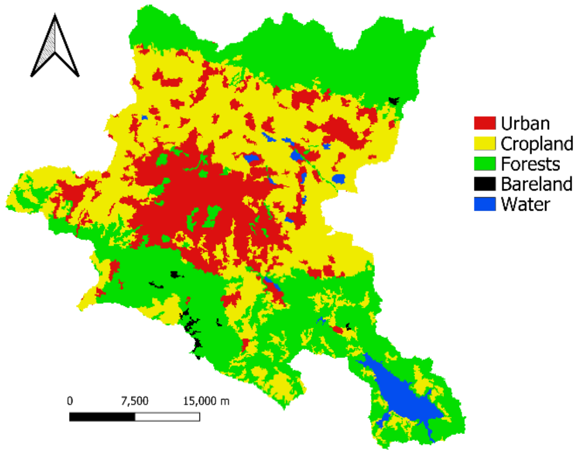


Figure 1. Reclassified 2018 CLC LULC map

The aggregation of UA data is conducted similarly to that of CLC. Notably, the bare lands classification is underrepresented in this context (Figure 2). This dataset is transformed to enhance the efficacy of the algorithms and the ensuing analysis. Furthermore, multiple iterations of the maps are developed to serve various objectives.

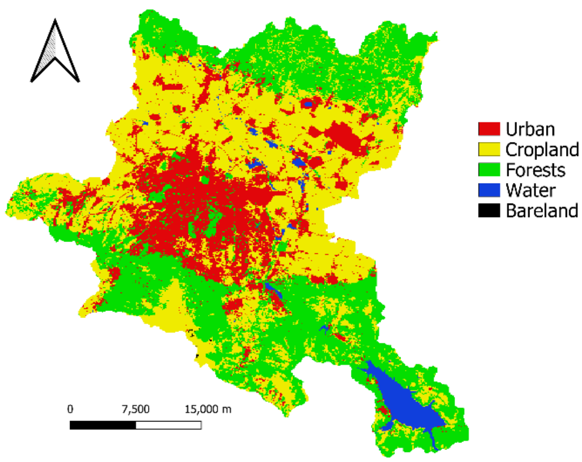


Figure 2. Reclassified 2018 UA LULC map

The preprocessing of the road network involves reducing the network to the main roads by filtering the attributes of the vector. Then, the vector is rasterised using the "Rasterize" function in QGIS. The resulting raster layer is transformed into a proximity map through the "Proximity" function, where black indicates high proximity (Figure 3).

Figure 3. Road Proximity Map of Sofia City in 2022

The water bodies and wetness data are cleaned by removing wetness areas and leaving only the water bodies. Like the road data, it is transformed into a water proximity map (black indicates high proximity) (Figure 4).

Figure 4. Water Proximity Map of Sofia City in 2018

The population density map has already been prepared during the data collection phase, as it is extracted from the Google Earth engine in the required format (Figure 5).

Figure 5. Population Density Map of Sofia City in 2018

The DEM is transformed into a slope map by assessing the terrain's angle of inclination (white indicates high elevation) (Figure 6). This enhancement introduces an additional dimension to the input data.

Figure 6. Elevation map of Sofia City between 2011-2015

All raster layers are standardised to have a matching extent of 567:586 and pixel values of 100.565x100.565 meters. In addition, the datasets are reprojected to the EPSG: 32634 - WGS 84 / UTM 34N Coordinate Reference System.

### 2.4 Data processing

This section presents the correlation evaluation and area change detection.

**Evaluating Correlation:** The correlation between the predicting variables is explored once the data is consistent. This step is required to understand the relationship between the predictive variables, which can bring additional insight into the patterns of change in the area. The findings are presented in section 3.1.

**Area Change Analysis:** The changes in the area of interest from 1990 to 2018 are quantified by calculating the differences in LULC proportions over this period and developing transition matrices to monitor changes among the various classes. This timeframe is chosen based on data availability, representing the longest span identified for the region. The total area and the extent of change for each of the five LULC classes are estimated. Additionally, a transition matrix is constructed to illustrate the proportional transitions of each LULC type to the others. The CLC maps are used for this analysis, as the UA maps do not cover the entire timespan. The findings are presented in section 3.2.

### 2.5 Transition Potential Modelling

The transition potential modelling is the first part of the future LULC simulation, achieved with a single-layer ANN. The input consists of two historical LULC maps and the four predictive variable maps (Water proximity, road proximity, elevation and population density maps). The following hyperparameters are tuned to achieve optimal performance of the model for each experiment:

**Sampling:** The sampling method can be random or stratified. The latter option allows for a more balanced class distribution in the input sample, but the resulting simulation is unrealistic. In addition, the number of samples can be chosen. A larger sample size would take more time and could lead to overfitting.

**Neighbourhood:** This parameter defines the number of surrounding pixels to be considered when estimating the transition potential, as the value of each pixel is determined by its relative position. Therefore, 1 pixel means a 3x3 region or 9 pixels. Given that our resolution is 100 meters, it is assumed that 1 pixel is an adequate value for this parameter; otherwise, the estimation would be based on a too-broad area.

**Learning rate and momentum:** The learning rate and the momentum are subject to a trade-off between the learning process's stability and the training speed. Large values for these parameters result in unstable but fast learning, while smaller values provide a stable but slow learning process.

The performance of the model's training is measured in minimum error and validation Kappa.

### 2.6 CA-ANN Validation

The underlying model's ability to predict correctly is tested before generating a simulation of future land use. This is done by feeding the results of the transition potential modelling to the CA algorithm and creating a simulation of a LULC map, which can be compared to its real counterpart (also called a reference map). Therefore, a simulation of the LULC of Sofia City from 2018 would be compared to an actual one from 2018. The model's predictive performance is estimated using two metrics: percentage of correctness and Kappa coefficient. The threshold for satisfactory results is >85% on the percentage of correctness and >0.75. In addition, the transition potential across the region is inspected for further insight. The transition potential can range between 0 and 100. The validation step helps understand the predictive performance of the algorithms and tune the hyperparameters if needed.

### 2.7 Future Cellular Automata Simulation

The CA algorithm is applied twice in this methodology: firstly, in the validation step, and secondly, if the validation results are satisfactory, to derive a simulation map of the future LULC. This algorithm has gained popularity in the spatial domain in the past two decades, as, in combination with advancements in remote sensing technologies, it offers many applications (Ghosh et al., 2017). One notable application is the prediction of LULC changes. The algorithm simulates spatial transformations over time by utilising established transition rules and considering neighbourhood effects (Ghosh et al., 2017; Santé et al., 2010). The study area is a grid where each cell's future state depends on its current state, surrounding cells, and transition probabilities derived from historical data. The model updates the landscape through multiple iterations, capturing urban expansion, land conversions, and environmental shifts, making it a powerful tool for LULC forecasting and urban planning. A single iteration equals the timespan between the years of the LULC input.

### 2.8 Simulation experiments

The study uses four simulation experiments as follows. The first experiment (Experiment 1) takes the CLC maps from 2006 and 2012 as input, as the magnitude of change between these two years is the highest. Thus, it is assumed that the model would yield simulations with higher confidence. The experiment is conducted using a random sampling method of 10000 samples. The neighbourhood is set to 1 pixel and the learning rate to 0.005 with a momentum of 0.010.

The second experiment (Experiment 2) repeats the methodology of the first one - the input includes the predictive variable maps. The UA maps of 2006 and 2012 are used instead of the CLC ones to see if the lack of change is attributed to the source. It uses a random sampling method with 10000 samples. Neighbourhood is set to 1 pixel, the learning rate to 0.001 with a momentum of 0.001.

The third experiment (Experiment 3) aims to understand if the map's level of detail, measured in a number of classes, could influence the output. Therefore, a different classification is used. This experiment divides the UA LULC maps into six classes: high urban, low urban, cropland, forests, water and urban green areas. The input includes the four predictive variable maps and the newly reclassified LULC maps of 2006 and 2012. It uses a random sampling method with 20000 samples. Neighbourhood is set to 1 pixel, the learning rate to 0.001 with a momentum of 0.001.

Finally, the fourth experiment (Experiment 4) focuses on the importance of the input timespan. In the previous experiments, the timespan is 6 years. Therefore, this experiment uses the four predictive variables and the CLC data from 1990 and 2018, the largest available timespan. A random sampling method is applied, sampling 10000 instances. The neighbourhood is set to 1 pixel, and the learning rate is 0.005 with a momentum of 0.01.

## 3. Results

This section presents results from the correlation evaluation, area change analysis, CA-ANN validation and future simulations, including four experiments.

### 3.1 Correlation evaluation

Table 1 shows Pearson's correlation between the predictive variables. Population density has a weak negative correlation with road proximity (-0.24) and elevation (-0.22), indicating areas with higher population density are closer to roads and flatter. Road proximity and elevation have a moderate positive correlation (0.46), suggesting roads are more likely to be in less steep areas. Water proximity has a weak correlation with population density (0.01) but moderate correlations with road proximity (0.49) and elevation (0.39), suggesting that water areas are often near roads and less steep land.

| | Population Density | Road Proximity | Water Proximity | Elevation |
|---|---|---|---|---|
| **Population Density** | - | -0.24 | 0.01 | -0.22 |
| **Road Proximity** | -0.24 | - | 0.49 | 0.46 |
| **Water Proximity** | 0.01 | 0.49 | - | 0.39 |
| **Elevation** | -0.22 | 0.46 | 0.39 | - |

Table 1. Correlation Between Variables.

### 3.2 Area Change Analysis

Table 2 summarises the area change analysis results in km$^2$ by showing the proportions of each LULC type based on all available years from the CLC. Croplands and forests take up more than 75% of the total area.

| | 1990 | 2000 | 2006 | 2012 | 2018 |
|---|---|---|---|---|---|
| **Urban** | 253.26 | 254.04 | 258.60 | 264.8 | 266.73 |
| **Cropland** | 523.79 | 522.75 | 519.18 | 513.33 | 510.15 |


https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-1159-2025 | © Author(s) 2025. CC BY 4.0 License.

| | | | | | |
|---|---|---|---|---|---|
| **Forests** | 544.72 | 545.3 | 544.64 | 543.02 | 544.2 |
| **Bare lands** | 8.76 | 8.05 | 4.88 | 5.14 | 5.14 |
| **Water** | 33.57 | 33.95 | 36.78 | 37.80 | 37.87 |

Table 2. Area Change 1990-2018

The transition matrix for the period 1990-2018 is shown in Table 3:

| | **Urban** | **Cropland** | **Forests** | **Bare lands** | **Water** |
|---|---|---|---|---|---|
| **Urban** | 0.9155 | 0.0549 | 0.0118 | 0 | 0.0179 |
| **Cropland** | 0.0566 | 0.9074 | 0.0340 | 0 | 0.0020 |
| **Forests** | 0.0084 | 0.0356 | 0.9545 | 0.0007 | 0.0007 |
| **Bare lands** | 0 | 0.0704 | 0.3880 | 0.5416 | 0.0000 |
| **Water** | 0.0190 | 0.0283 | 0.0024 | 0 | 0.9503 |

Table 3. Transition Matrix 1990-2018

Tables 2 and 3 reveal the following trends: firstly, the urban area increased by 13.47 km$^2$, by 5.32% growth. The most significant development occurred between 2006 and 2012, by 6.2 km$^2$. The transition to other classes is limited, with 91% of the urban class remaining unchanged. Furthermore, the cropland class continuously decreases, which aligns with urban expansion. The area is reduced by 13.64 km$^2$ or 2.6%. The most significant loss occurred between 2006 and 2012. On the other hand, forests have remained relatively stable with minor fluctuations over the 28 years. The class showed a decreasing trend between 2000 and 2012 but grew afterwards. Total loss is 0.52 km$^2$, or 0.1%. Moreover, the bare lands experienced significant reductions until 2006, after which they stabilised. The class lost 41.32% of its area and is the least represented among the five broad classes. The transition matrix suggests that most losses are due to conversion into forests. Finally, the water area gradually rises, with minor losses to other area types. The most significant gains occurred between 2000 and 2006 when the area grew by 2.83 km$^2$. By 2018, the water class has increased by 12.81%.

The area changes in terms of the difference between two periods of the year in km$^2$ (expressed as delta) can be observed in Table 4. The delta is calculated per LULC class for each period, and the overall change is calculated as the sum of their absolute values. The overall delta shows the magnitude of change over the years. The transition between classes varies over the years. The most dynamic period is between 2000 and 2012, specifically between 2006-2012, mainly due to significant urbanisation and reduction of cropland. The results are supported by Figure 7 showing the total area change over the period between 1990-2018.

| | **1990-2000** | **2000- 2006** | **2006-2012** | **2012-2018** |
|---|---|---|---|---|
| **Urban** | 0.78 | 4.56 | 6.20 | 1.93 |
| **Cropland** | -1.04 | -3.57 | -5.85 | -3.18 |
| **Forests** | 0.58 | -0.66 | -1.62 | 1.18 |
| **Bare lands** | -0.71 | -3.17 | 0.26 | 0 |
| **Water** | 0.38 | 2.83 | 1.02 | 0.07 |
| **Total Delta** | **3.49** | **14.79** | **14.95** | **1.93** |

Table 4. Area Change Magnitudes


Figure 7. Area Change Map 1990-2018

### 3.3 CA-ANN Validation

The validation results of Experiment 1 demonstrate an exceptionally high level of accuracy, with a correctness of 99% and an overall Kappa coefficient of 0.99, indicating near-perfect agreement (Figure 8).

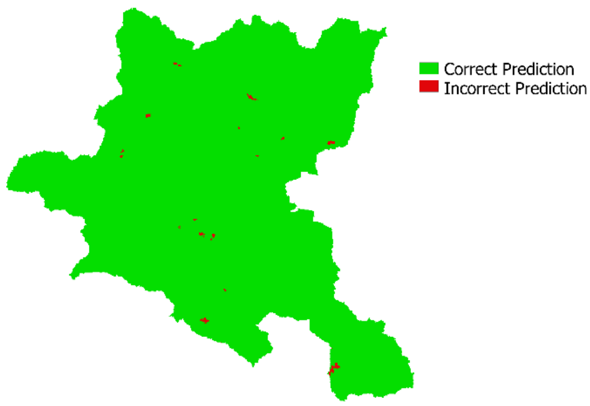


Figure 8. Results of Validation of Experiment 1.

The results indicate that the map is identical to the area change map between 2012 and 2018 (Figure 9). The identified errors are localised precisely in the regions that experienced transitions during this period.

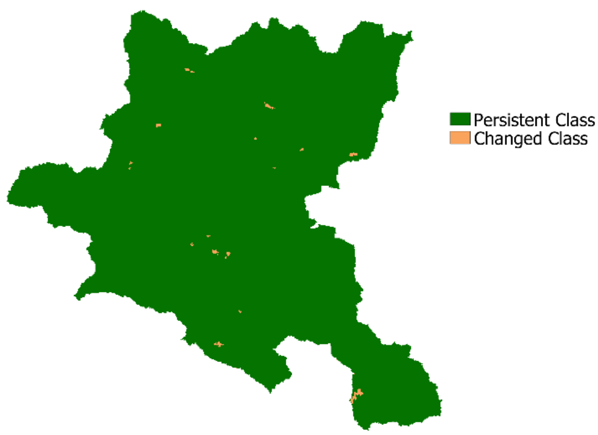


Figure 9. Area Change Map 2012-2018.

Furthermore, the transition potential analysis shows that there are instances where the model predicts a potential up to 11, although the majority remains close to 0 (Figure 10). Only Experiment 1's validation is visualised as Experiment 2 and Experiment 3 have similar outcomes.


Figure 10. Transition Potential Map Experiment 1.

Regarding Experiment 2, the simulated 2018 LULC map reached 99.5% correctness and 0.99 overall Kappa. However, no change is predicted between 2012 and 2018. Similar to Experiment 1, the validation map errors cover the areas that have transitioned between 2012 and 2018. The highest potential in experiment 2 is 21, again in very few cases.

The validation results of the 2018 simulation within Experiment 3 are 99.4% correctness and 0.99 overall Kappa for the latter. Again, the errors on the validation map cover the areas that have transitioned between 2012 and 2018. Similar to previous experiments, the highest transition potential is low (17), and the instances where it can be observed are outliers.

Finally, the output of Experiment 4 cannot be validated as the simulation cannot be compared to any existing map. A single iteration of the CA algorithm, in this case, would equal 28 years, which is too far into the future to be able to gather such data. Therefore, the only way to assess the output is to compare it to the trends from the period on which it is based (1990-2018) and determine if it follows them. In addition, the transition potential reaches 65, and, unlike the previous experiments, it is high throughout most of the region, especially around the urban area (Figure 11).


Figure 11. Transition Potential Map for Experiment 4.

### 3.4 Future Simulations

This section describes the future simulations, and area changes for the experiments. Regarding Experiment 1, the obtained validation Kappa for the ANN model is 0.971 with a minimum error of 0.007. The simulation of the 2078 LULC has the same properties as the 2012 LULC and does not predict any change. The obtained validation Kappa of the ANN model training within Experiment 2 is 0.988 with a minimum error of 0.004. Like Experiment 1, the 2078 simulation does not expect any changes. The obtained validation Kappa of the ANN model training within Experiment 3 is 0.89 with a minimum error of 0.007. However, the results did not differ from the previous experiments. Finally, the obtained validation Kappa of the ANN model training within Experiment 4 is 0.87 with a minimum validation overall error of 0.02. The output of the simulation of the LULC in 2074 is shown in
Figure **12**.

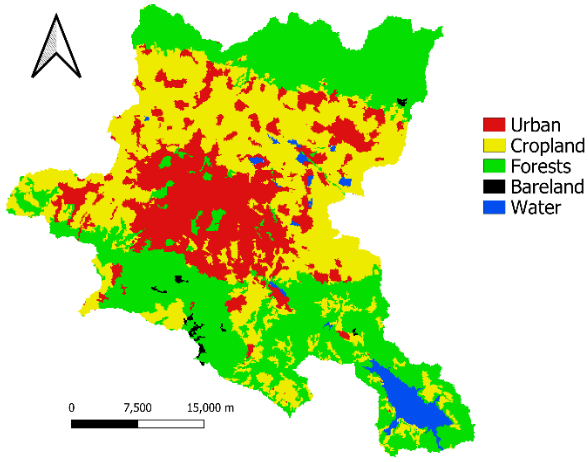


Figure 12. Simulation of the 2074 LULC Map.

According to the results, the urban area is expected to grow by 21.35 $km^2$, an increase of 8.01%, while the cropland will decline by 17.88 $km^2$. The forest and the water area will drop by 2.31 and 1.16 $km^{2,}$ respectively, and no change is expected for the bare lands. Most of the urban growth can be expected in the southern part of the city, as shown in Figure 13.

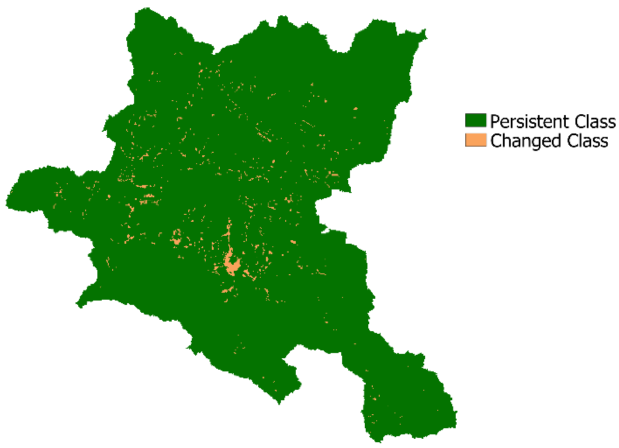


Figure 13. Area Change between 2018-2074.

# 4. Discussion

The area changes analysis shows that the urban area experienced an increase of 13.47 $km^2$ (5.32%) over the study period, primarily driven by growth between 2006 and 2012, when 6.2 km2 of land was urbanised. This period represents the most dynamic phase, likely caused by infrastructure development and population growth. Cropland, however, decreased, losing 13.64 $km^2$ (2.60%) over the 28 years. The most pronounced reduction occurred between 2006 and 2012, coinciding with the peak of urban growth. This relationship could be explained by the direct conversion of agricultural land into urban areas, a common trend in urbanising regions, especially in Eastern Europe (Kuemmerle et al., 2016). The forest class remained relatively stable, with a total loss of 0.52 $km^2$ (0.1%). Although there was a temporary decline between 2000 and 2006, forests showed signs of recovery in subsequent years. Bare lands experienced a substantial reduction of 41.32% in their total area, with the most significant losses occurring before 2006. The transition matrix reveals that a significant portion of bare lands transitioned into forests, suggesting a process of natural or deliberate reforestation efforts. Finally, water bodies increased by 12.81%. This growth could be attributed mainly to water management consequences such as reservoir construction. The relative stability observed in forest, water, and bare lands after 2006 indicates a slowdown in land cover changes, with slow but steady urban growth becoming the dominant trend. The findings of SOER 2020 (European Environment Agency, 2019) align with the observed LULC changes in Sofia, confirming that urban expansion is a steady but not dominant land transformation trend in Europe. Between 2000 and 2018, artificial surfaces in Europe increased by 7.1%, primarily at the expense of cropland, which mirrors the 5.32% urban growth in Sofia. Additionally, the report highlights stable forest areas across Europe, consistent with Sofia's minimal forest loss and signs of regeneration. These similarities suggest that Sofia's urbanisation patterns reflect broader European trends of moderate urban growth, cropland decline, and relative forest stability.

Based on the metrics, Experiments 1 and 2 accurately predict the 2018 LULC of the study area. However, upon closer investigation, the output imitates perfectly the map of 2012. The simulation does not predict any changes in the future. The low transition potential explains this phenomenon, which indicates that the model performance can be improved. These experiments showed that the input data source does not affect the transition potential in the absence of a large timespan. However, the classification accuracy during the generation of historical LULC maps matters, as it affects both the area change analysis and the prediction of future LULC. For example, some water bodies are incorrectly classified as landfills in 2006 and then correctly classified as water bodies in 2012, which is considered an area change. Therefore, the quality of the input determines the quality of the output. Although Experiment 3 failed to produce better results, the increase in the level of detail of the input map gives more clarity and makes the output more explainable. Real-life application of such methodology would benefit from a higher level of detail. Finally, the results of Experiment 4 prove that the large timespan is an essential factor when projecting the future LULC. Even without validation, the simulation aligns with current and past trends in the region's LULC development, making it a plausible scenario for future studies.

An essential point of consideration is that landscape change is a subject not only to natural causes but also urban planning decisions, which can hardly be predicted (Santé et al., 2010). However, if these decisions can be represented as variables, they could increase the accuracy of the prediction, thus making the results more reliable.

# 5. Conclusion and Future Work

## 5.1 Main Conclusions

This study demonstrates how LULC analysis and predictive modelling can provide valuable insights into urban expansion and environmental changes. The following conclusions are considered:

- The urban areas in Sofia grew consistently between 1990 and 2018, increasing by 5.32%, mainly at the expense of croplands, which decreased by 2.6%. The change is in line with the general trends in Europe.
- Overall, the simulations of the future land use in 2074 show that the urban area is expected to grow by 21.35 $km^2$, which is an increase of 8.01%, while the cropland will decline by 17.88 $km^2$. The forest and the water area will drop by 2.31 and 1.16 $km^{2,}$ respectively, and no change is expected for the bare lands.
- The simulation experiments found that two aspects of the input play a key role in the reliability of the output – firstly, the quality of the data, which affects both the area change analysis and the subsequent predictions, and secondly, the timespan between input maps.
- A short input timespan (6 years) might not capture the magnitude of change and, therefore, produce weaker simulations. Thus, a combination of the CLC data's timespan (28 years) and the UA's accuracy is required to simulate the future LULC successfully.

## 5.2 Limitations and Future Work

More steps can further improve the findings of this research. Firstly, using LULC data from a more distant past can increase the input timespan, thus leading to better results. This might require gathering historical satellite images and transforming them into LULC maps through classification algorithms. Secondly, the study could include more predictive variable maps, as more factors might influence an area's development. As the development of a region is a consequence of economic and natural changes, both need to be captured to achieve reliability.

In addition, a more detailed classification of the input data could increase the explainability of the changes. The complexity of the input data would contribute to both the area change analysis and the simulation output. Another future step would be to use alternative algorithms and compare their performance to find an optimal one. Logistic regression is one potential candidate. (Salem et al., 2021). Finally, the findings of this study will offer adaptable methodologies and insights that can be applied to cities facing comparable developmental challenges.

Understanding a region's past and current state helps project the consequences of the LULC development in future scenarios. This approach supports the development of scenario-based planning, where stakeholders assess various "what-if" situations based on different inputs and determine the optimal outcome. By applying the methodology suggested in this study, policymakers and urban designers can leverage the power of remote sensing and machine learning to adapt their strategies to account for economic growth, environmental sustainability, and societal well-being.

## Acknowledgements

This research is supported by the GATE project, funded by the H2020 WIDESPREAD-2018-2020 TEAMING Phase 2 programme, under grant agreement no. 857155, FLEdge project, funded by the Driving Urban Transitions (DUT) Partnership programme, under agreement no. KP-06-D002/5 and deployEMDS project, co-funded by Digital Europe Programme DIGITAL-2022-CLOUD-AI-03 under grant agreement no. 101123520.

## References

Copernicus Land Monitoring Service, 2024a. CORINE Land Cover [WWW Document]. URL https://land.copernicus.eu/en/products/corine-land-cover (accessed 5.18.24).

Copernicus Land Monitoring Service, 2024b. Urban Atlas [WWW Document]. URL https://land.copernicus.eu/en/products/urban-atlas (accessed 5.18.24).

Copernicus Land Monitoring Service, 2024c. High-resolution spatial dataset [WWW Document]. URL https://land.copernicus.eu/en/dataset-catalog (accessed 5.18.24).

Cunha, E.R. da, Santos, C.A.G., Silva, R.M. da, Bacani, V.M., Pott, A., 2021. Future scenarios based on a CA-Markov land use and land cover simulation model for a tropical humid basin in the Cerrado/Atlantic forest ecotone of Brazil. Land use policy 101. https://doi.org/10.1016/j.landusepol.2020.105141

European Environment Agency, 2019. The European environment – state and outlook 2020: Knowledge for transition to a sustainable Europe [WWW Document]. URL https://www.eea.europa.eu/soer/publications/soer-2020 (accessed 2.9.25).

Ghosh, P., Mukhopadhyay, A., Chanda, A., Mondal, P., Akhand, A., Mukherjee, S., Nayak, S.K., Ghosh, S., Mitra, D., Ghosh, T., Hazra, S., 2017. Application of Cellular automata and Markov-chain model in geospatial environmental modeling- A review. Remote Sens Appl. https://doi.org/10.1016/j.rsase.2017.01.005

Kuemmerle, T., Levers, C., Erb, K., Estel, S., Jepsen, M.R., Müller, D., Plutzar, C., Stürck, J., Verkerk, P.J., Verburg, P.H., Reenberg, A., 2016. Hotspots of land use change in Europe. Environmental Research Letters 11. https://doi.org/10.1088/1748-9326/11/6/064020

Li, T., Li, W., 2015. Multiple land use change simulation with Monte Carlo approach and CA-ANN model, a case study in Shenzhen, China. Environmental Systems Research 4. https://doi.org/10.1186/s40068-014-0026-6

NextGIS, 2024. QGIS Python Plugins Repository MOLUSCE [WWW Document]. URL https://plugins.qgis.org/plugins/molusce/ (accessed 2.9.25).

Nuissl, H., Siedentop, S., 2021. Urbanisation and Land Use Change. https://doi.org/10.1007/978-3-030-50841-8_5

Orlov, A., Aunan, K., Mistry, M.N., Lejeune, Q., Pongratz, J., Thiery, W., Gasparrini, A., Reed, E.U., Schleussner, C.F., 2023. Neglected implications of land-use and land-cover changes on the climate-health nexus. Environmental Research Letters 18. https://doi.org/10.1088/1748-9326/acd799

QGIS Development Team, 2024. QGIS Geographic Information System [WWW Document]. URL https://qgis.org (accessed 2.9.25).

Salem, M., Bose, A., Bashir, B., Basak, D., Roy, S., Chowdhury, I.R., Alsalman, A., Tsurusaki, N., 2021. Urban expansion simulation based on various driving factors using a logistic regression model: Delhi as a case study. Sustainability (Switzerland) 13. https://doi.org/10.3390/su131910805

Santé, I., García, A.M., Miranda, D., Crecente, R., 2010. Cellular automata models for the simulation of real-world urban processes: A review and analysis. Landsc Urban Plan 96. https://doi.org/10.1016/j.landurbplan.2010.03.001

Schiavina M., Freire S., Carioli A., MacManus K., 2023. GHS-POP R2023A - GHS population grid multitemporal (1975-2030) [WWW Document]. https://doi.org/10.2905/2FF68A52-5B5B-4A22-8F40-C41DA8332CFE

Shen, C., Hou, H., Zheng, Y., Murayama, Y., Wang, R., Hu, T., 2022. Prediction of the future urban heat island intensity and distribution based on landscape composition and configuration: A case study in Hangzhou. Sustain Cities Soc 83. https://doi.org/10.1016/j.scs.2022.103992

Slaev, A.D., Nedovic-Budic, Z., 2017. The challenges of implementing sustainable development: The case of Sofia's master plan. Sustainability (Switzerland) 9. https://doi.org/10.3390/su9010015

Solomon, T., Lukas, P., 2024. Land use land cover analysis of the Great Ethiopian Renaissance Dam (GERD) catchment using remote sensing and GIS techniques. Geology, Ecology, and Landscapes 8. https://doi.org/10.1080/24749508.2022.2138027

Sudhakar, S., Rao, K.S., 2010. Land Use and Land Cover Analysis. Remote Sensing Applications.

TomTom, 2024. TomTom Traffic Stats [WWW Document]. URL https://www.tomtom.com/products/traffic-stats/ (accessed 10.5.24).

Vitanova, L.L., Kusaka, H., 2018. Study on the urban heat island in Sofia City: Numerical simulations with potential natural vegetation and present land use data. Sustain Cities Soc 40, 110–125. https://doi.org/10.1016/J.SCS.2018.03.012

Vitanova, L.L., Kusaka, H., Doan, Q., Subasinghe, S., 2021. How urban growth changes the heat island effect and human thermal sensations over the last 100 years and towards the future in a European city? Meteorological Applications 28. https://doi.org/10.1002/met.2019

Wang, Q., 2018. Urbanization and global health: The role of air pollution. Iran J Public Health 47.

World Population Review, 2024. Sofia, Bulgaria Population .

Wu, X., Wang, Lunche, Yao, R., Luo, M., Wang, S., Wang, Lizhe, 2020. Quantitatively evaluating the effect of urbanization on heat waves in China. Science of The Total Environment 731, 138857. https://doi.org/10.1016/J.SCITOTENV.2020.138857

Xie, Q., Sun, Q., 2021. Monitoring the spatial variation of aerosol optical depth and its correlation with land use/land cover in Wuhan, China: A perspective of urban planning. Int J Environ Res Public Health 18. https://doi.org/10.3390/ijerph18031132