UNIVERSITY OF BELGRADE

FACULTY OF CIVIL ENGINEERING

Mileva S. Samardžić-Petrović

PREDICTING LAND USE CHANGE WITH DATA-DRIVEN MODELS

DOCTORAL DISSERTATION

Belgrade, 2014.

УНИВЕРЗИТЕТ У БЕОГРАДУ

ГРАЂЕВИНСКИ ФАКУЛТЕТ

Милева С. Самарџић-Петровић

ПРЕДВИЂАЊЕ ПРОМЕНА У КОРИШЋЕЊУ ЗЕМЉИШТА ПРИМЕНОМ МОДЕЛА ВОЂЕНИХ ПОДАЦИМА (DATA-DRIVEN MODELS)

Докторска дисертација

Београд, 2014. година

Ментор:

В. проф. др Бранислав Бајат, дипл. геод. инж.
Универзитет у Београду, Грађевински факултет

Коментор:

В. проф. др Милош Ковачевић, дипл. електр. инж.
Универзитет у Београду, Грађевински факултет

Чланови комисије:

Доц. др Жељко Цвијетиновић, дипл. геод. инж.
Универзитет у Београду, Грађевински факултет

Проф. др Сузана Драгићевић, дипл. геод. инж.
Simon Fraser University, Department of Geography,

Проф. др Дејан Ђорђевић, дипл. пр. планер
Универзитет у Београду, Географски факултет

Датум одбране:

„Најбољи људи носе своје злато не у џепу него у срцу. Једно унутрашње сунце целог њиховог живота обасјава и позлаћује све на свету чега се дотакне њихова рука и њихова мисао.“

Јован Дучић

Dedicated to memory of my beloved father,

He will always be my inspiration in life to be honourable, brave, compassionate, grateful, hardworking and loving.

Acknowledgments

I express my deepest gratitude to all people who supported and helped me, professionally and personally, during the long and not an easy road of this research work.

First of all, I would like to thank my supervisor, Associate professor Branislav Bajat and co- supervisor, Associate professor Miloš Kovačević for their guidance during the entire research work. I am very grateful to them for generously giving me valuable knowledge through intensive and usefully discussions and for helping me successively finish dissertation.

Secondly, I am also grateful to my comity members. I would like to express my deepest gratitude to Professor Suzana Dragićević for her valuable guidelines and for providing me resources which have led to the completion of this research. It has been an honour to be visiting researcher in Spatial Analyzing and Modeling lab at Simon Fraser University, under her supervision, for five months during which I have gained tremendous experience. I would also like to thank Assistant professor Željko Cvijetinović and Professor Dejan Đorđević for their helpful feedback on my research which have led to a more clear and comprehensive expression of the dissertation.

Special acknowledgments go to Urban Planning Institute of Belgrade and companies MapSoft and Geo Info Strategies, especially to Milica Joksić, Vladimir Vasiljev and once again to Assistant professor Željko Cvijetinović for providing me with spatial data, which made this research possible.

I would like to offer my special thanks to Dr. Nikola Krunić for his numerous advice, suggestions and help in perceiving the problems of land use change and urban development in more inclusive sense.

Special thanks also go to Professor Dragan Blagojević, my supervisor at graduate thesis work, whose knowledge and approach to research work inspired me to enrol PhD studies and conduct further research.

Above all, I would like to thank my beloved family. Thank my mom for her infinite care, tenderness and for teaching me to appreciate the real values in life. She, my sister and brother have always been my pillar that I can lean on and made my life much happier and easier. Thank you for unwavering support, encouragement, understanding and constant love. Thank my husband, for his love, support and infinite patience at all time throughout my PhD. For peer-reviewing the text and giving me valuable comments and suggestions. To my adorable nieces Vanja and Nadja and nephew Ognjen for their unconditional love and for being the smile in my day and always cheering me up. I would also like to thank to all other special people in my life that were always there for me.

Mileva Samardžić-Petrović

Predicting Land use change with Data-Driven Models

Abstract

One of the main tasks of data-driven modelling methods is to induce a representative model of underlying spatial - temporal processes using past data and data mining and machine learning approach. As relatively new methods, known to be capable of solving complex nonlinear problems, data-driven methods are insufficiently researched in the field of land use. The main objective of this dissertation is to develop a methodology for predictive urban land use change models using data-driven approach together with evaluation of the performance of different data-driven methods, which in the stage of finding patterns of land use changes use three different machine learning techniques: Decision Trees, Neural Networks and Support Vector Machines. The proposed methodology of data-driven methods was presented and special attention was paid to different data representation, data sampling and the selection of attributes by four methods (χ^2, Info Gain, Gain Ratio and Correlation-based Feature Subset) that best describe the process of land use change. Additionally, a sensitivity analysis of the Support Vector Machines -based models was performed with regards to attribute selection and parameter changes. Development and evaluation of the methodology was performed using data on three Belgrade municipalities (Zemun, New Belgrade and Surčin), which are represented as 10×10 m grid cells in four different moments in time (2001, 2003, 2007 and 2010).

The obtained results indicate that the proposed data-driven methodology provides predictive models which could be successfully used for creation of possible scenarios of urban land use changes in the future. All three examined machine learning techniques are suitable for modeling land use change. Accuracy and performance of models can be improved using proposed balanced data sampling, including the information about neighbourhood and history in data representations and relevant attribute selections. Additionally, using selected subset of attributes resulted in a simple model and with less possibility to be overfitted with higher values of Support Vector Machines parameters.

Key words: data-driven modeling, data mining, machine learning, spatial-temporal modeling, land use changes, Geographic Information Systems

Scientific area: Geodesy

Scientific sub-area: Land Information System

UDC number: 007:528.9]:004(043.3)

Предвиђање промена у коришћењу земљишта применом модела вођених подацима (Data-driven models)

Резиме

Један од главних задатака моделирања метода вођених подацима (*Data-driven methods*) је проналажење репрезентативног модела испитивног просторно временског процеса, применом податка из прошлости и *Data Mining* и *Machine Learning* приступа. Попут других релативно нових метода, које решавају комплексне нелинеарне проблеме, методе вођене подацима су недовољно истражене у области коришћења земљишта. Главни циљ ове дисертације је развој методологије за моделе предвиђања промена у коришћењу земљишта употребом раличитих модела вођених подацима, оцене њихових перформанси у фази проналажења образаца у промени коришћења земљишта применом три различите технике машинског учења (*Machine Learning Techniques*): 1) Стабла одлуке (*Decision Trees*), 2) Неуронске мреже (*Neural Networks*) и 3) Метод вектора подршке (*Support Vector Machines*). У дисертацији је предложена методологија примене метода вођених подацима са посебним освртом на репрезентацију података, узорковању података и избору атрибута који најбоље описују процес промене коришћења земљишта, употребом четири методе: χ^2, *Info Gain*, *Gain Ratio* и *Correlation-based Feature Subset*. Извршена је анализа осетљивости модела заснованог на Методи вектора подршке у зависности од избора атрибута и промене параметара. Развој и оцена методологије је урађена коришћењем података са три Београдске општине (Земун, Нови Београд и Сурчин), које су представљене растерским ћелијама величине 10×10 m у четири различите временске епохе (2001, 2003, 2007 и 2010).

Добијени резултати указују да предложена методологија даје моделе предвиђања који се могу успешно користити за креирање могућих сценарија за будуће промене у коришћењу урбаног земљишта. Све три испитиване технике машинског учења су погодне за моделирање промена коришћења земљишта. Тачност и перформансе модела се могу побољшати употребом предложеног балансираног узорковања података, укључивањем информација о суседству и историји коришћења земљишта у репрезентацији података и употребом релевантног избора атрибута. Употреба изабраних подскупова атрибута резултује једноставнијим моделима са мањом могућности да буду претренирани (да имају мању могућност генерализовања промена) високим вредностима параметара Метода вектора подршке.

Кључне речи: модели вођени подацима (data-driven models), машинско учење, просторно-временско моделирање, промена коришћења земљишта, географски информациони системи

Научна област: Геодезија

Ужа научна област: Земљишни информациони системи

УДК број: 007:528.9]:004(043.3)

Table of Contents

List of abbreviations:

ABM – Agent-based model

ANN – Artificial Neural Network

BHT – Building height typology

CA – Cellular Automata

CFS – Square and Correlation-based Feature Subset

CORINE – COoRdinate INformation on the Environment

DD – Data-Driven

DM – Data Mining

DT – Decision Trees

EEA – European Environment Agency

GIS – Geographic Information Systems

GMES – Global Monitoring for Environment and Security

GR – Gain Ratio

IG – Info Gain

IGBP – International Geosphere - Biosphere Programme

IHDP – International Human Dimension Programme on Global Environmental Change

IT – Information Technology

LUC – Land use changes

LULC – Land use and land cover changes

MCK – Map Comparison Kit

ML – Machine Learning

MLP – Multy-layer Perceptron

NASA – National Aeronautics and Space Administration

NN – Neural Networks

PCI – Population Change Index

RBF – Radial Basis Function

SVM – Support Vector Machines

List of Figures:

List of Tables:

Chapter 1:

Introduction

Land use and land cover changes (LULC) play important role in human and physical system and have significant impact on environment at local, regional and global scale. Land cover refers to the physical and biological cover over the surface, while land use refers to the purpose of the land in the sense of its exploitation. They are connected by the proximate sources of changes; human action that directly alter the physical environment (Meyer and Turner, 1994). Therefore, issues regarding land use changes over large areas are increasingly important for many studies related to environment in general and global change in particular (Cihlara and Jansenb, 2001).

In order to use land more efficiently in the future, one of the prime prerequisites is information on existing land use patterns and changes in land use through time (Anderson, 1976). Understanding the complexity of land use change and the evaluation of its impact on the environment comprises the procedures of detection and modeling of those changes (Huang et al., 2010).

Over the last few decades, a wide range of different types of land use change models have been developed in attempt to assess and project the future role of LULC in the functioning of the earth system (Veldkamp and Lambin, 2001). Literature provides an overview of many operation models for land use changes and urban growth (Wegener, 1994, U.S. EPA, 2000, Jones, 2005). However, clients for LULC models, such as urban planners and environmental agencies, have constant need for models that would be more adequate for their specific needs. Along with proper response to propagation of specific phenomena, a decisive requirement to support sustainable growth and planning is to improve the reliability of predictive models (Kocabas and Dragicevic, 2006).

The development of models for the analysis and prediction of dynamic geographic phenomena such as the movement of the Earth's crust, landslides, climate change or land use and land cover change among others, have been spurred recently by the vast availability of digital data and geospatial datasets, new tools for data acquisition and storage in large databases, geographic information systems (GIS) and related technologies. Due to a large amount of available data and rapid advances in computer technology, a need for new modeling methods, data-driven methods, appeared.

The data-driven (DD) methods provide the capability to develop modelling procedures for representation of the underlying processes from historical datasets. The focus of data-driven modelling methods is to find patterns and trends or to induce a representative model of underlying processes using past data and data mining and machine learning approach. As a class of methods known to be capable of solving nonlinear problems, they have been successfully applied to different dynamic geographic phenomena. However, as a relatively new approach, data-driven methods are insufficiently researched in the field of land use, particularly for building prediction models of land use change (LUC).

Considering the significance of LUC modeling and the promising potential of data-driven methods which was not sufficiently researched, the primary question that this dissertation answers is to what extent can certain DD models can be used for modelling of land use change in case of high thematic resolution land use data. Starting hypothesis of this dissertation was:

"Based on the available data on land use (land cover) from two or more time horizons and on the choice of appropriate auxiliary predictors, modeling of land use change on the area of interest can be carried out by using data-driven methods."

Therefore, the main goals of this dissertation are:

- Design and development of land use change models based on different machine learning techniques,
- Performance testing of developed models,
- Appropriate validation of developed models,
- Discussion of obtained results in regard to actual land use.

The dissertation is organised in six chapters, including introduction. The second and third chapter describe the theoretical backgrounds of land use modeling and data-driven methods; fourth chapter presents study area; the last two chapters are devoted to the presentation of the conducted experiments, analysis and discussions on the obtained results and conclusions.

In second chapter **Land use modeling** is presented. Basics of predictive modeling - the terms and objectives of modeling are introduced along with a brief review of the literature and different approaches for modeling of land use changes.

Theoretical background and outline of the proposed methodology is presented in chapter **Data-driven methods for land use change modelling** which includes: defining the problem and data representation; three machine learning techniques - Decision Trees (DT), Neural Networks (NN) and Support Vector Machines (SVM); methods for attribute selection - Info Gain (IG), Gain Ratio (GR), Chi-Square and Correlation-based Feature Subset (CFS); appropriate data sampling and various forms of the Kappa statistics. Detailed review of the literature regarding previous application of data-driven methods in the land use field is presented.

The detailed analysis of social-economic aspects of study area and preparation and analysis of used data are presented in chapter **Study area, preparation and data analysis**. In addition, assessment of similarity between planned and actual land use maps was performed and presented in this chapter.

Results and discussions are presented in two sections. First section covers four conducted experiments designed in order to evaluate the performance of proposed data-driven methodology, focusing on data sampling, datasets representation and attributes selection. Model outcomes were analysed and discussed. Second section is focused on sensitivity analysis of SVM techniques and appropriate selection of parameters.

Chapter 2:

Land use change modeling

2.1 Introduction to modelling

The intention of the developed data-driven models in this dissertation is to model a spatial - temporal process that can be particularly used for prediction of land use changes. Consequently, it is necessary to define the basic theoretical background behind the spatial - temporal model and the predictive modelling used herein.

There are various definitions from different authors found in literature referring to the term "spatial - temporal model" and they can be summarized as (Dragicevic, 2013a):

Model is an abstract, simplified or partial representation or description of some or several aspects of the real world, phenomenon, process or system.

Spatial model is an abstract, simplified or partial representation or description of some or several aspects of the geographic phenomenon that manifest in two or three dimensional space.

Spatial – temporal model refers when the simplified representation is for dynamic geographic phenomena and uses input information on the spatial and temporal dimension.

Since the main task of this research is to create predictions of land use changes, it is also necessary to define the term of ***predictive modeling***. One of the most cited definitions of that term in the domain of land use modelling is:

Predictive modelling is the process by which a model is created or chosen to try to best predict the probability of an outcome (Geisser, 1993, page 31).

Joshua Epstein (2008) explained that prediction can be a goal, and it is feasible, particularly “if one admits statistical prediction in which stationary distributions (of wealth or epidemic sizes, for instance) are the regularities of interest”. In addition, Epstein has presented sixteen reasons, beside prediction, to build models:

1. Explain,
2. Guide data collection,
3. Illuminate core dynamics,
4. Discover new questions,
5. Illuminate core uncertainties,
6. Suggest dynamical analogies,
7. Promote a scientific habit of mind,
8. Bound outcomes to plausible ranges,
9. Offer crisis options in near-real time,
10. Demonstrate tradeoffs / suggest efficiencies,
11. Challenge the robustness of prevailing theory through perturbations,
12. Expose prevailing wisdom as incompatible with available data,
13. Reveal the apparently simple (complex) to be complex (simple),
14. Train practitioners,
15. Discipline the policy dialogue, and
16. Educate the general public.

Some definitions of a predictive model describe it as: “any device or mechanism which generates a prediction” (Haines-Young and Petch 1986, page 144) and “a simplified representation or description of a system or complex entity, especially one designed to facilitate calculations and predictions” (Makins 1995, page1003); thereby, the prediction presents the main goal of modelling.

However, the most appropriate definition considering the aim of this research can be found in the IT Glossary (https://www.gartner.com/it-glossary) where predictive modelling is defined as solutions that "have a form of data-mining technology that works by analyzing historical and current data and generate a model to help predict future outcome".

Models can be classified in various ways. Brimicombe (2010) defined four classes of models loosely based on classification defined by Chorley and Haggett (1967):

1. Natural analogues – these are descriptive models that can be historical (using events from past to explain present events) or spatial (using events from one place to explain events on another place),
2. Hardware – these models use physical miniaturization of a phenomenon in order to examine general behaviour, changes in state, influence of variables and etc.,
3. Mathematical - phenomenon are described using equations, functions, or statistics. It can be deterministic (providing single solution) or stochastic (providing probabilistic solution taking into account random behaviour),
4. Computational – using code and data to express a phenomenon and its behaviour. They include deterministic and/or stochastic elements alongside heuristics, logical operators, set operators, etc. Since most of the investigated phenomena in the real world are complex dynamic processes, it is unlikely that they can be reduced to a formal mathematical model. Computational models perhaps offer less precision and clarity, but they tend to offer a greater level of realism and flexibility.

Furthermore, Brimicombe adds additional descriptors for computational models depending on:

1. Role of time:
   - Static - elements of the model are fixed over time, and
   - Dynamic - variables in a model are allowed to vary in time.
2. Degree of specification of the model as a system:
   - White box – model representing a system is fully specified,
   - Grey box - model representing a system is partially specified, and

- Black box - model representing a system is not specified.

3. Way in which the model is being used:
   - Exploratory - models seek to reveal the mechanism of some phenomenon, and
   - Prescriptive - models are used to provide answers (resulting outputs based on given inputs).

Based on this classification, computational, dynamic, prescriptive and white (DT) or black box (NN and SVM) models were used in this research.

2.2 Land use change models

Land use changes ultimately affect future changes in the Earth's climate and consequently have great implications for subsequent land use change (Agarwal et al., 2002). Land use has been often considered as a local environmental issue but it is now known that it is one of the main contributors related to environmental degradation and climate change at global levels (Foley et al., 2005, Lambin and Geist, 2006). The issue of LUC have become an important part of several international programs such as the International Geosphere - Biosphere Programme (IGBP), International Human Dimension Programme on Global Environmental Change (IHDP) and NASA's Land Cover and Land Use Change Program (Cheng, 2003, Zhao et al., 2011). Modelling of LUC is conducted at different time and space scale levels and was the subject of study in many scientific fields including: geography, urban planning, geo-information science, ecology and land use science (Verburg et al., 2004, Agarwal et al., 2002, Turner B.L. et al., 2007).

Land use is determined by the spatial – temporal interaction of human and biophysical factors (Veldkamp and Fresco, 1996). Agarwal et al. (2002) proposed an analytical framework for the categorization and summary of land use change models based on scale and complexity of the model by taking into consideration space, time, human and biophysical factors. The scale of a model is defined by the temporal scale (time steps and duration), the spatial scale (spatial resolution and extent) and scales of human decision-making (agent and domain). Respectively, model complexity can be represented with an index that considers the complexity of time (number of used time steps), space (spatial dimension and neighbourhood) or human decision-making (level of influence on human decision-making processes).

Therefore, land use change is a very complex class of dynamic nonlinear geographic processes that is dependent on many factors. Based on the conclusions of Lambin and Geist (2006), the causes of land use change can be divided into two categories:

- Proximate (direct, or local) causes explain how and why local land cover and ecosystem processes are modified directly by humans, and
- Underlying (indirect or root) causes explain the broader context and fundamental forces underpinning these local actions.

Some of the commonly associated causal factors (attributes) used in the modelling of land use change are: demography (population size, growth or density), accessibility (distance to city center, road, markets), economic (housing/land prices, job growth), social (affluence, human attitudes and values), physical characteristics of terrain (slope, elevation and aspect), biological characteristics of terrain (soil quality) and many others.

Being complex and diverse in nature, the modelling of LUC is a difficult task and can be solved using different approaches ranging from the Markov model (Turner M.G., 1988, Muller and Middleton, 1994, López et al., 2001), logistic and multiple regression (Wu and Yeh, 1997, Theobald and Hobbs, 1998, Schneider and Pontius, 2001, Hu and Lo, 2007), fractal (White and Engelen, 1993; Shen, 2002, Triantakonstantis, 2012), cellular automata (Clarke et al., 1997, White et al., 1997, Stevens and Dragicevic, 2007) to more recently, agent-based models (Castella et al., 2005, Brown et al., 2005, Xie et al., 2007, Kocabas and Dragicevic, 2009). The brief theoretical backgrounds of the two most popular approaches are presented below.

Cellular automata (CA) has a long tradition in land use change modeling. In the late 1940’s John von Neumann and Stanislawa Ulam were the first to develop CA. Several decades later, Waldo Tobler (1970) presented research in the simulation of population increase based on the cellular model. He released another very important study in which he used the CA for the consideration of physical phenomena (Tobler, 1979a). Helen Couclelis (1985) first referred to Tobler's work in the context of linking raster models and CA, claiming that the achievements of CA and systems theory could be combined and applied in urban and other geographic systems. From that time CA was used in a range of land use change issues.

CA presents an effective bottom-up simulation tool for modeling dynamic processes. It is defined through five components: the grid space, the neighborhood, the finite set of states of each individual automaton, the transition rules, and the time step (Lai and Dragićević, 2011). The study area is commonly presented as a lattice of cells (individual automata). Each cell exists in one of a finite set of states, and its future states depend on transition rules considering the local neighborhood. There are various approaches for transition rules (“if-then”, Markov chain, fuzzy logic, artificial neural network, etc.).

Agent - based model (ABM), with CA, presents a bottom-up approach for modelling processes and it is based on complex system theory. The process of modelling is carried out by agents (O'Sullivan and Haklay, 2000). Agents are used to represent entities and to make them communicate and interact with each other and/or with their environment. One of the main characteristics of the agents is autonomy, which can be defined as a control over their behaviour and internal states in order to make decisions and achieve goals (Dragicevic, 2013b). The formal definition of agent-based-modelling applied by Gilbert (2008, page 2) is that ABM is a “computational method that enables a researcher to create, analyze, and experiment with models composed of agents that interact within an environment”. ABMs were successfully applied in the field of land use for various problems including land use planning and urban growth (Matthews et al., 2007). A new approach was recently developed by linking CA and ABM in order to better explore urban growth (Sudhira, 2004, Torrens, 2006).

The DD methods represent a relatively new approach in the field of land use change. The detailed review of the literature with regards to application of DD methods in the land use field is presented in the “Data-driven methods for land use change modelling” chapter.

Chapter 3:

Data-driven methods for land use change modelling

The modern world can be considered as ”data-driven” due to the availability of large amounts of data, numerical figures and other bits of information in the digital format. Data must be analysed and processed into a form that informs, instructs, answers and aids the understanding of the real world and decision making processes (Kantardzic, 2011). Data-driven methods offer the ability to develop modelling procedures that are based on historical datasets and are analysed for representation of change processes. Therefore, the main task of data-driven modelling is to find patterns, trends or to induce a representation of natural phenomenon. Additionally, according to Shahab Araghinejad (2014), some of the purposes of data-driven modelling are:

- Data classification and clustering,
- Function approximation,
- Forecasting,
- Data generation,
- General simulation.

Data-driven methods present entire discovering procedures for the processes that are being investigated: defining the problem, collecting and preparation of data, analysing data, finding /extracting/ learning patterns or building a model, validation and analysis of results and implementation of the results (Figure 3.1).

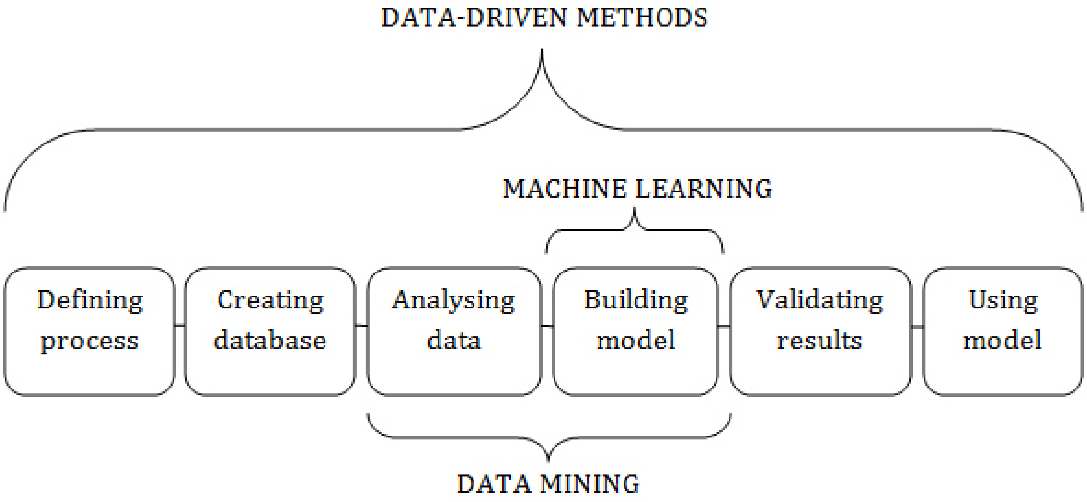


Figure 3.1 Data-driven methods as discovering procedure.

Data mining methods are often used to analyse and learn patterns that can be retrieved from data already present in extensive databases (Fayyad et al., 1996).

Furthermore, Data mining (DM) uses Machine learning (ML), pattern recognition or statistical techniques to learn these patterns.

It can be seen that DD methods encompass ML and DM approaches (Solomatine, 2002). ML and DM use existing data which describe the phenomena of interest to learn the unknown relations between input and output variables. These continual and/or categorical variables can be processed often without explicit knowledge of mutual interactions. ML techniques include many methods with different kinds of learning algorithms. In this dissertation, ML techniques, such as Decision Trees, Neural Networks and Support Vector Machines are used for LUC modeling.

Data-driven methods have been successfully used in many fields such as medicine (Khan et al., 2001), biological engineering (Benedict and Lauffenburger, 2013), biology (Knudby et al., 2010), chemistry (Zhou, 2004), economy (Huang Z. et al., 2004), engineering and manufacturing (Paliwal and Kumar, 2009). Moreover, they have found application in geo-sciences such as hydrology for predicting runoff and management of river basins (Solomatine and Ostfeld, 2008), geology for prediction of landslide hazards (Tien Bui et al., 2012, Marjanović et al., 2011) and remote sensing for image classification (Bischof et al., 1992, Friedl and Brodley, 1997).

One of the first DD methods used for modelling changes in land use were based on NN that was often coupled with cellular automata approaches (Yeh and Li 2003, Liu et al., 2007, Almeida et al. 2008, Thekkudan, 2008, Mahajan and Venkatachalam, 2009) or GIS (Pijanowski et al. 2002). Applications of DT (Li and Yeh, 2004) and SVM (Yang et al., 2008) were used as methods for constructing transition rules for cellular automata models of land use change. The use of DT and SVM methods in the field of LUC are starting to draw more attention in the research community (Okwuashi et al, 2012, Charif et al., 2012, Triantakonstantis et al., 2011). Although interest in the application of these methods in the field of land use change has been growing over the last few years, the capability of DT, NN and SVM to predict land use changes have not been significantly explored.

3.1 Outline of the proposed methodology

The main objective of this dissertation is to develop a methodology for building predictive models in urban LUC environment using DD approach. Performance evaluation of the proposed DD methods was accomplished using different data representations, data sampling, different ML techniques (Decision Trees, Neural Networks and Support Vector Machines) and different attribute selection methods (χ2, Info Gain, Gain Ratio and Correlation-based Feature Subset).

The methodology itself is finally shaped and specified after thorough analysis of the results presented in the fifth chapter. The results provided basis for the verification of the proposed methodology and each of the DD methods. The results also enabled drawing conclusions that are vital for successful application of each method in terms of selection of appropriate data representation, data sampling, ML techniques and different attribute selection methods.

Proposed methodology is presented in Figure 3.2 and it can be applied on many spatial - temporal phenomena.

Each DD method contains seven stages:

- (1) Defining the problem/objective,
- (2) Collection and preparation of data - creating a database,
- (3) Data sampling – selecting appropriate datasets for building and validating the models,
- (4) Data analysis - finding best describing LUC attributes,
- (5) Building the models – application of ML techniques
- (6) Validation of the built models,
- (7) Using the best performing model.

The objective of a DD model is to predict urban land use changes. The problem of LUC prediction can be formulated as a classification task, in which land use classes represent the output of the model (section 3.2).

Spatial - temporal processes, such as LUC, present very complex non linear problems and they depend on many different factors (attributes). Therefore, in order to define the distribution of those factors on a study area for several different moments in time, it is necessary to create a database in GIS environment. GIS database creation, which represents the second stage of the methodology, is represented in chapter 4. Data from this database are used to create training and test datasets which are necessary to built and validate model.

As DD methods deal with very large data volumes, it is necessary to sample the data in order to obtain smaller, independent data frames, taking into consideration both training and test datasets for creation of representative samples for model building and validation. Proposed data sampling, which represents the third stage of the methodology, was presented in section 3.4.

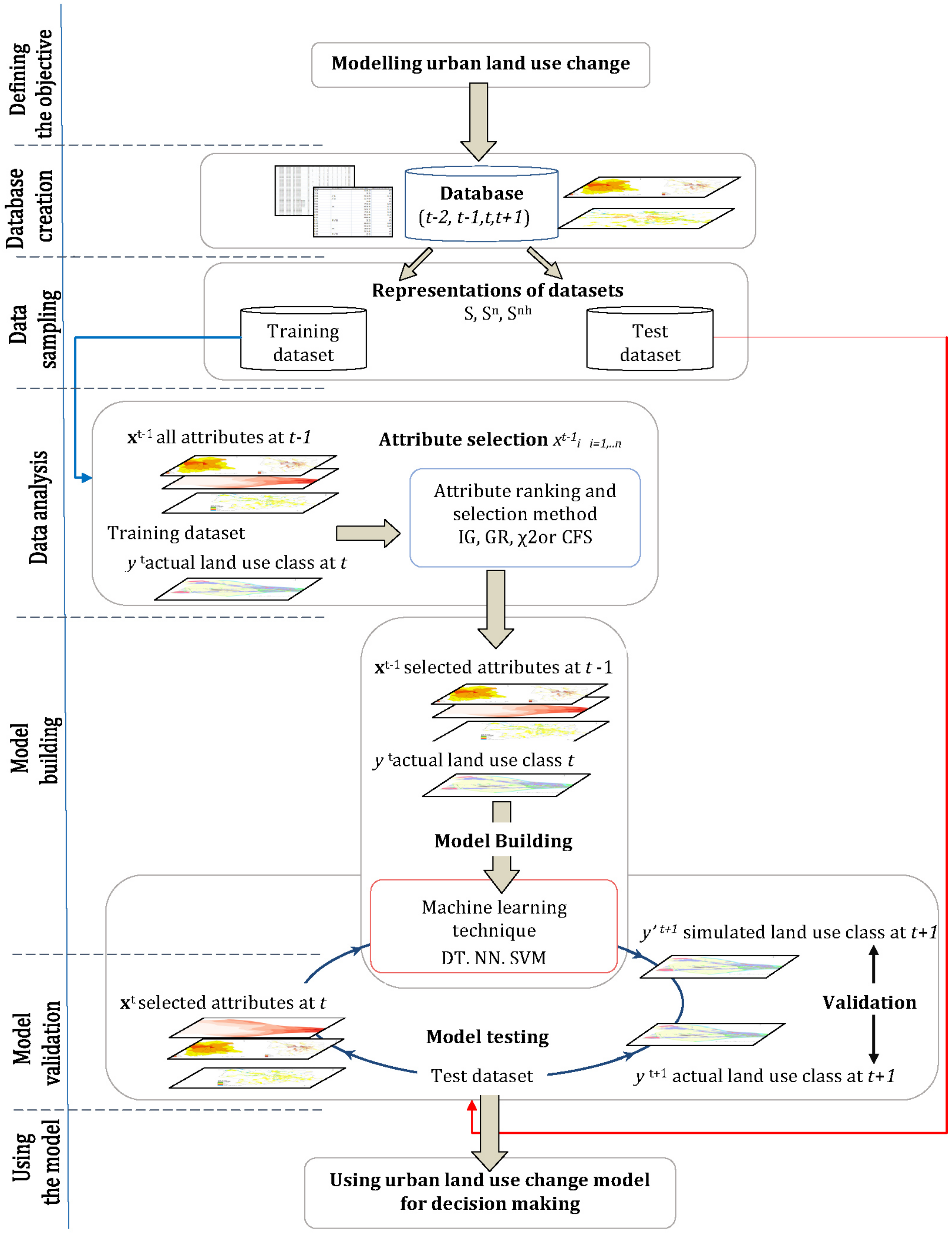

Defining the objective
Database creation
Data sampling
Data analysis
Model building
Model validation
Using the model
Modelling urban land use change
Database
(t-2, t-1,t,t+1)
Representations of datasets
S, S^{n}, S^{nh}
Training dataset
Test dataset
$\mathbf{x}^{t-1}$ all attributes at t-1
Attribute selection x^{t-1}_{i} $_{i=1,..n}$
Attribute ranking and selection method
IG, GR, χ2or CFS
Training dataset
y^{t}actual land use class at t
$\mathbf{x}^{t-1}$ selected attributes at t -1
y^{t}actual land use class t
Model Building
Machine learning technique
DT, NN, SVM
y'^{t+1} simulated land use class at t+1
Validation
$\mathbf{x}^{t}$ selected attributes at t
Model testing
Test dataset
y^{t+1} actual land use class at t+1
Using urban land use change model for decision making

Figure 3.2 Proposed methodology.

One of the key characteristics of DD methods is the choice of dataset representation and the choice of appropriate attributes for the most informative representation of real-world entities and the problem situation (section 3.5). There are a lot of attribute ranking and selection methods and techniques which can be used in the fourth stage. Four attribute ranking methods were used and compared in this research.

As explained at the beginning of this chapter, in the "learning" phase of spatial - temporal process (model building stage), DD methods use one of many ML techniques. Three ML techniques (section 3.3) were used and compared in the fifth stage of the proposed methodology.

The sixth stage implies validation of the built model. In order to validate built models, the actual and predicted (outcomes of the model) land use maps were compared and various map comparison measures were used (section 3. 6).

The last stage of the proposed methodology assumes the use of the best performing model. As shown in chapter 1 and 2, the purpose of the land use change prediction modeling can be various and it depends on attributes used for modelling, its scale and complexity.

Theoretical background of the proposed methodology is completely presented in following sections of this chapter.

3.2 Defining the problem and data representation

The modelling of LUC using DD methods assumes the transformation of physical, socio-economical, neighbouring and other related data for each unit cell into an appropriate data representation for the area under consideration. The study area is represented as a grid of cells, often registered as raster-based GIS data layers, where each cell (pixel) has a rectangular shape and is uniquely identified with its accompanied attributes and land use class. Several GIS data layers of the study area at different moments in time are required for building the prediction model.

When the data is organized in the previously described fashion, the function of the land use change can be derived based on the various ML techniques.

The corresponding learning problem could be formulated as follows: each cell (*instance*) is represented as an *n*-dimensional vector $\mathbf{x}^t$, where coordinate x^t_i represents the value of the i^{th} urban attribute associated with the cell $\mathbf{x}^t$ ($\mathbf{x}^t$=<x^t_1, x^t_2,..., x^t_n>). Further, let C = {c_1, c_2,...c_k } be the set of *k* predefined land use classes and $y^{t+1} \in$ C land use class of $\mathbf{x}^t$ in time *t+1*. A function applied over each $\mathbf{x}^t$ from the grid representing the study area, f_p: $\mathbf{x}^t \rightarrow y^{t+1}$, is called a *prediction* if for each $\mathbf{x}^t$ holds that f_p ($\mathbf{x}^t$) = y^{t+1} whenever a cell $\mathbf{x}^t$ changes its land use to the class $y^{t+1} \in$ C. Values from C are usually mapped into natural numbers with each representing a particular land use class and are commonly referred as target attributes when predicted in the form of y^{t+1}.

The ML techniques try to find a function f_p' that is the best possible approximation of a real unknown function f_p using only the training dataset in which all attribute values and land use classes are known in advance and are applied to a specific learning method.

Finding the function f_p' defines the well-known classification task in which classes represent land use classes to be *predicted* at time *t*+1 and input values represent attributes of grid cells at the previous time *t*. Various ML techniques could be applied to solve the problem at hand.

In order to build the predictive model, one needs a dataset containing grid cells with accompanied attributes at time *t* – 1 (past) and corresponding land use classes at time *t* (present) in the form of ($\mathbf{x}^{t-1}$, y^t)$_k$, *k*=1,2,...,*N*, where *N* is the number of cells. This dataset is called a *training set* S_{TR} for the study area.

The goodness of f_p' as a classification function is measured through its capacity to predict (classify) future changes ($\mathbf{x}^t$, y^{t+1}). In order to evaluate the goodness of f_p', a separate *test set* S_{TE} in the form of ($\mathbf{x}^t$, y^{t+1})$_k$, *k*=1,2,...,*N* is used to compare the predicted and the real land use classes at time *t*+1.

In order to better understand the need for all of the stages in the modeling process, methods of ML will be explained first (section 3.3) followed by the data sampling (section 3.4), selection of attributes (section 3.5) and validation of model (section 3.6).

3.3 Machine learning techniques

In the 1950's, an early pioneer in the field, Arthur Samuel, developed the first self-learning program for the game of checkers (Samuel, 1959). He defined machine learning as a "Field of study that gives computers the ability to learn without being explicitly programmed" (Simon, 2013). Over the past 60 years the study of machine learning has grown and many definitions were created. Some of the most cited are:

- By Herbert Simon in 1983: "Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more effectively the next time." (Egresits et al, 1998, page 323),
- By Tom Mitchell (1997, page 2): "Computer program that improves its performance at some task through experience.", more precisely "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

Machine learning techniques can be used to solve different types of problems including: Classification, Regression, Numeric prediction, Clustering, Association and Concept description. However, all ML techniques can be divided into *supervised* and *unsupervised*. The learning technique is supervised if it "learns" with the actual outcome for each of the training examples (for classification, regression and numeric prediction). Oppositely, the technique is unsupervised (for clustering) if the outcomes are not provided (the training examples do not include the output values).

In this dissertation, following supervised ML techniques have been considered: Decision Trees, Neural Networks and Support Vector Machines. The theoretical backgrounds of the ML methods used within this body of work are described in detail in the following sub sections.

3.3.1 Decision Trees

Decision Tree is a simple but powerful method used for classification and regression. There is a number of different DT learning algorithms for classification: CART (Breiman et al., 1984), ID3 (Quinlan 1986) and C4.5 (Quinlan, 1993). In this dissertation the C4.5 decision-tree classifier was used. It classifies *instances* described with a set of attributes by testing the value of one particular attribute a_i per node, commencing from the root of the tree (attributes are urban parameters $a_i = x^t_i$ observed at time *t*). Testing then follows a certain path in the tree structure, which depends on the tests in previous nodes, and finally reaches one of the leaf nodes labelled with a class label. Each path leading from the root to a certain leaf node (class label) can be interpreted as a conjunction of tests involving attributes on that path. Since there could be more leaf nodes with the same class labels, one could interpret each class as a *disjunction* of *conjunctions* of constraints for the attribute values of *instances* from the dataset.

The tree construction process performs a *greedy* search in the space of all possible trees starting from the empty tree and adding new nodes in order to increase the classification accuracy on the training set. A new node (candidate attribute test) is added below a particular branch if the *instances* following the branch are partitioned after the candidate attribute test in such way that the distinction between the classes becomes more evident (Figure 3.3). A perfect attribute choice is discerned if the test on attribute a_i splits the *instances* into subsets in which all elements have the same class labels (those subsets become leaf nodes), Figure 3.3a. On the other hand, the worst attribute choice can be discerned if the *instances* are distributed into subsets with equal numbers of elements belonging to different classes (Figure 3.3b). Hence, the root node should be tested against the most informative attribute concerning the whole training set.

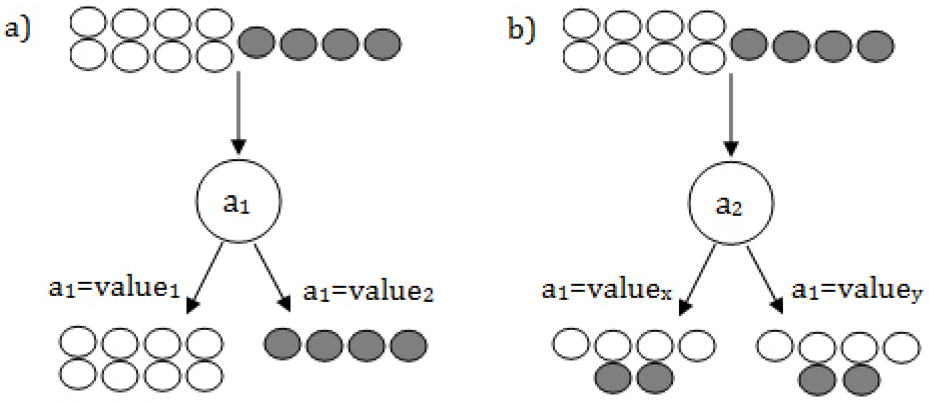


Figure 3.3 Choosing the attribute in the internal node of the growing tree; a) perfect attribute choice, b) the worst attribute choice.

The C4.5 classifier uses the Gain Ratio (GR) measure (Quinlan, 1986) to choose between the available attributes and is heavily dependent on the notion of entropy (Shannon, 1948). Therefore, GR effectively measures the capacity of an attribute to split the input set into sets with lower Entropy concerning class labels of containing *instances* (land use of grid cells $\mathbf{x}^t$=<x^t_1, x^t_2,..., x^t_n>). Figure 3.4 explains the calculation of Gain Ratio.

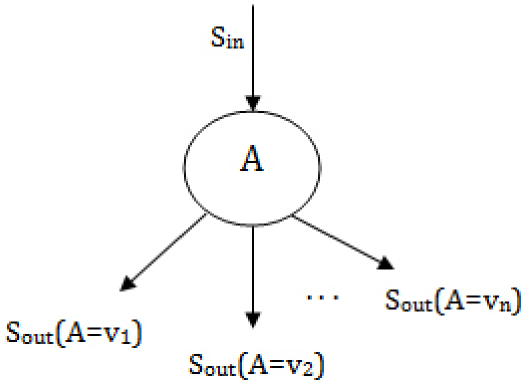


Figure 3.4 Calculating Gain Ratio of an attribute in the internal node of the growing tree.

Let S$_{\text{in}}$ be the set of N instances for which the preceding test in the parent node forwarded them to the current node. Further, let n_i be the number of instances from S$_{\text{in}}$ that belong to class c_{i}, i=1,...,k. The entropy E(S$_{\text{in}}$) is defined as a measure of impurity (with respect to the class label) of the set S$_{\text{in}}$ as:

$$E(\mathrm{S}_{in}) = -\sum_{i=1}^{k} \frac{n_i}{N} \log_2 \frac{\mathrm{n_i}}{\mathrm{N}}. \tag{3.1}$$

The entropy of the system is zero if all instances belong to the same class. On the other hand, if all classes are equally present, the entropy is a maximum (log$_2$k). In the case of this research, the setting A denotes the candidate attribute of an instance **x**. Since assumption A is categorical and can take n different values v_1, v_2,..., v_{n}, there are n branches leading from the current node. Each S$_{\text{out}}$(A=ν$_{\text{i}}$) represents the set of instances for which A takes the value ν$_{\text{i}}$. The informative capacity of A concerning the classification into k predefined classes can be expressed using the notion of *Information Gain* (IG):

$$IG(\mathrm{S}_{in}, \mathrm{A}) = E(\mathrm{S}_{in}) - \sum_{v \in (v_1, \ldots, v_m)} \frac{|\mathrm{S}_{out}(\mathrm{A} = v)|}{N} E(\mathrm{S}_{out}(\mathrm{A} = v)). \tag{3.2}$$

In Equation (3.2) |S$_{\text{out}}$(A=ν)| represents the number of instances in the set S$_{\text{out}}$(A=ν) and E(S$_{\text{out}}$(A=ν)) is the entropy of that set calculated using Equation (3.1). The higher the IG is, the more informative attribute A is for classification in the current node, and vice versa (Mitchell, 1997).

The main disadvantage of the IG measure is that it favours attributes with many values over those with fewer. This leads to wide trees with many branches starting from corresponding nodes. Complex trees with lots of leaf nodes lead to models that are expected to overfit the data (it will learn the anomalies of the training data and its generalization capacity, i.e., the classification accuracy on unseen instances will be decreased). In order to reduce the effect of overfitting, C4.5 further normalizes IG by the entropy calculated with respect to the attribute values instead of class labels (*Split Information*) to obtain the *Gain Ratio* (GR):

$$SI(\mathrm{S}_{in}, \mathrm{A}) = -\sum_{v \in (v_1, \ldots, v_m)} \frac{|\mathrm{S}_{out}(\mathrm{A} = v)|}{N} \log_2 \frac{|\mathrm{S}_{out}(\mathrm{A} = v)|}{N},$$

$$GR(\mathrm{S}_{in}, \mathrm{A}) = \frac{IG(\mathrm{S}_{in}, \mathrm{A})}{SI(\mathrm{S}_{in}, \mathrm{A})}. \tag{3.3}$$

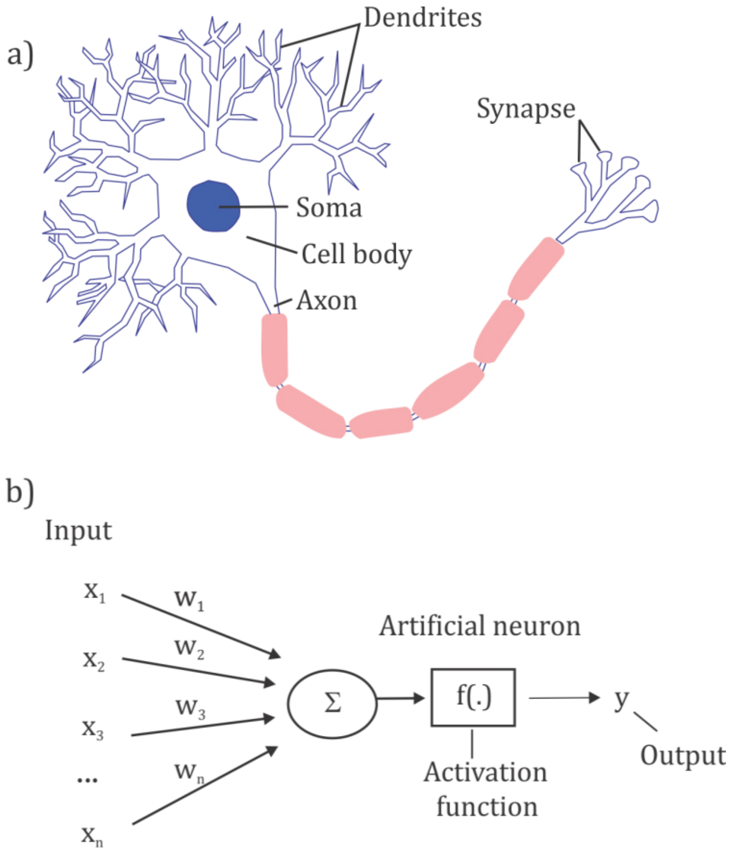
C4.5 uses GR to drive the greedy search over all possible trees. If the attribute is numerical (this is the case for most attributes in our application) C4.5 detects the candidate thresholds that separate the instances into different classes. Let (A, c_i) pairs be (50, 0), (60, 1), (70, 1) (80, 1), (90, 0), (100, 0). C4.5 identifies two thresholds on the boundaries of different classes: A<55 and A<85. The variable A now becomes a binary attribute (true or false) and the same GR procedure is applied to select from among the two thresholds when considering the introduction of this attribute test into the growing tree.

The C4.5 uses the so-called *post-pruning technique* to reduce the size of the tree (i.e. complexity of the model). After growing a tree that classifies all the training examples as well as a possible (overfitted model), a procedure is performed to remove and/or join some nodes yielding a tree that shows good behavior on the training set but is more general for the problem domain. There are many variants of the pruning technique but all of them can be compared with the adjusting parameter *C* in the SVM algorithm (explained in 3.2.3) since both techniques trade-off the training error versus the model complexity in order to increase the generalization power of the induced classification model.

The ability of DT to interpret the derived model as a set of IF – THEN rules enables a domain expert to have a better understanding of the problem and in many cases could be preferable to functional methods such as SVMs and NN.

3.3.2 Neural Networks

Neural Networks, which are also known as “Artificial” Neural Networks (ANN), are based on the logic of biological nervous systems (Figure 3.5). The first to introduce the idea of the mathematical model of biological neurons were neuropsychiatrist



Dendrites=Input; Synapse=Weight (w); Soma=Artificial neuron; Axon=Output

Figure 3.5 Similarity between a) Biological neural cell and b) Artificial neuron.

Warren McCulloch and mathematician Walter Pitts in a publication "A Logical Calculus of the Ideas Immanent in Nervous Activity" in 1943 (Abraham, 2002). Another contribution for development of neural networks was made by Donald Hebb in 1949, who postulated the first rule for self-organized learning (Haykin, 2009). Using the McCulloch-Pitts model and the Hebb rule, Rosenblatt (1958) presented the *perceptron*, the simplest Artificial Neural Network.

The *perceptron* consists of a single *node* (*artificial neuron*) (Figure 3.5b) and presents a binary classifier and can only classify linearly separable cases with a binary target (1, 0). The inputs to the neuron (x_1, x_2, ..., x_n) are multiplied by corresponding connected weights (w_1, w_2, ..., w_n) to form the weighted sum s (Equation 3.4). The weighted sum of inputs is then passed through an *activation* *function*, *f*(s), to produce the neuron's output signal *y* (Equation 3.5). If the sum is above the threshold *θ*, the perceptron is activated (value of function is 1):

$$s = \sum_{i=1}^{n} w_i \cdot x_i \ , \tag{3.4}$$

$$y = f\left(s\right) = \begin{cases} 1 \text{ if } s > \theta \\ 0 \text{ if } s \le \theta \end{cases} \Rightarrow \begin{cases} 1 \text{ if } \sum_{i=1}^{n} w_i \cdot x_i - \theta > 0 \\ 0 \text{ if } \sum_{i=1}^{n} w_i \cdot x_i - \theta \le 0 \end{cases} \tag{3.5}$$

In order to eliminate the threshold *θ* from Equation 3.5 a set of inputs to the neuron is often expanded with the additional constant input attribute x_0 = 1 and the associated weight w_0 (Jain et al., 1996). This additional input is called the *bias*, *b*. Therefore, with bias, Equation 3.4 becomes:

$$s = \sum_{i=1}^{n} w_i \cdot x_i - \theta = \sum_{i=1}^{n} w_i \cdot x_i + b = w_0 \cdot x_0 + w_1 \cdot x_1 + w_2 \cdot x_2 + \ldots + w_n \cdot x_n \, . \tag{3.6}$$

Activation functions that are commonly used are presented in Table 3.1.

Table 3.1 Some types of activation function, *f* (s).

Threshold	Piecewise-Linear	Sigmoid	Gaussian
$f\left(s\right)=\begin{cases} 1 \text{ if } s > 0 \\ 0 \text{ if } s \le 0 \end{cases}$	$f\left(s\right)=\begin{cases} 1 \text{ if } s \ge \frac{1}{2} \\ s \text{ if } \frac{1}{2} > s > -\frac{1}{2} \\ 0 \text{ if } s \le -\frac{1}{2} \end{cases}$	$f\left(s\right)=\frac{1}{1+e^{-s}}$	$f\left(s\right)=\frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(s-\mu)^2}{2\sigma^2}}$

Neural networks can be divided into two groups, based on the connections between the neurons (Jain el al., 1996):

- *Feed-forward*: the data from input to output units is strictly feed-forward. In other words, signals can only travel in one direction.

- *Feed-back (Recurrent)*: the data from input to output can travel in both directions, using loops and all possible connections between neurons.

In this dissertation, the Multi-layer Perceptron (MLP) neural network defined by Rumelhart, Hinton, and Williams (1986a) was used. The MLP is a *feed-forward* neural network and one of the most widely used ANNs (Pijanowski et al, 2002). The MLP is comprised of an input layer, one or more hidden layers and an output layer (Figure 3.6).

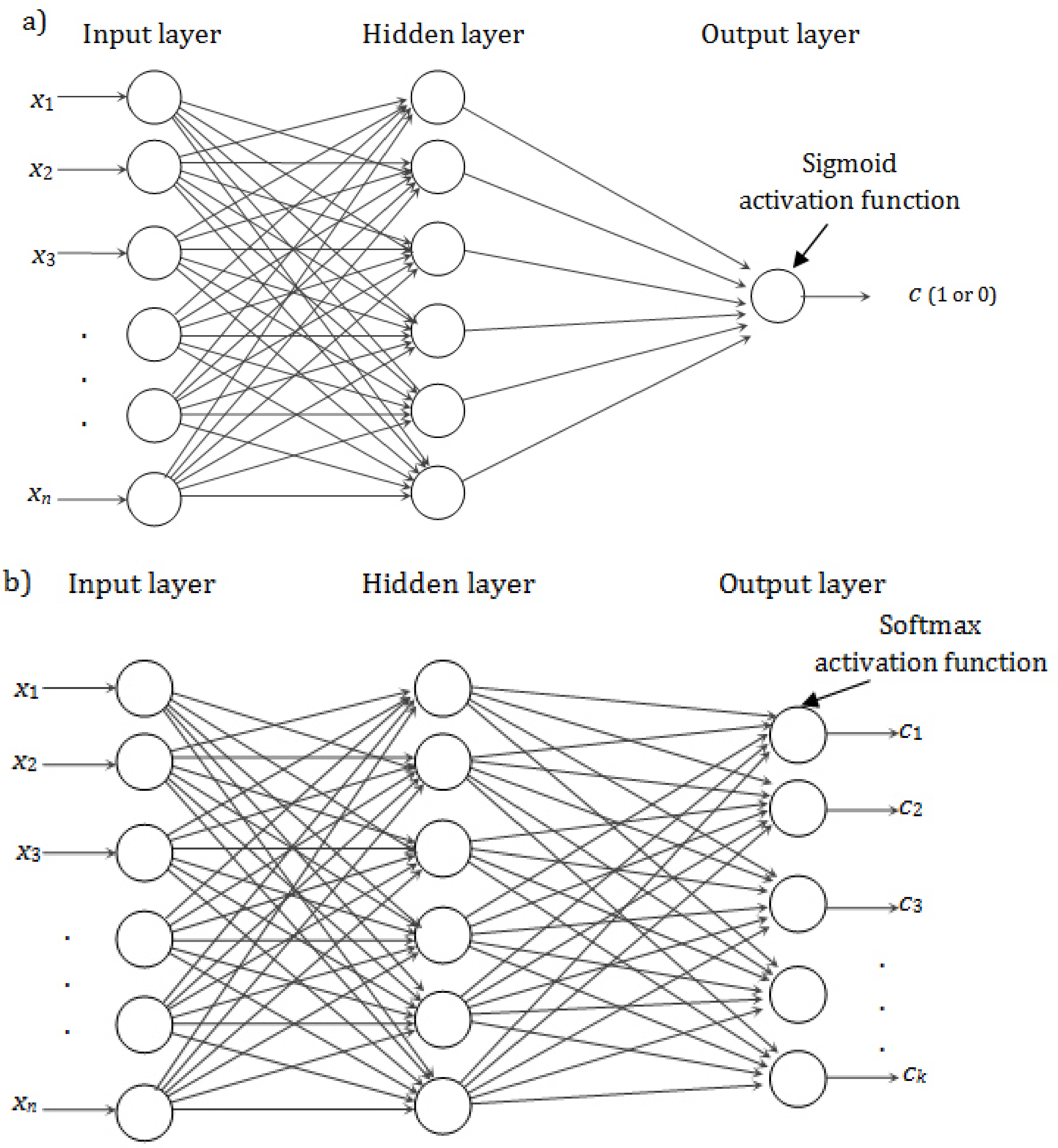


Figure 3.6 Schematic representation of a MLP classification network for a) two classes and b) *k*-classes, *k*>2.

In the simplest two class case (0 and 1) the output layer consists of only one neuron with a sigmoid activation function (Figure 3.6a). The network is trained with labelled examples from a training set using the well-known *back-propagation* algorithm (Rumelhart et al, 1986b). In the first iteration of the algorithm, each initial weight has a randomly chosen value (based on a normal distribution with zero mean). In each iteration every training example (urban grid cell) in the form of $\mathbf{x}_j=<x_1, x_2,..., x_n>_j$ is propagated through the network and the outputs are compared with the desired values (y_j = 0 or 1). For that purpose an error function E is defined to be:

$$E = \frac{1}{2}\sum_j \left(y_j - f\left(s_j\right)\right)^2 . \tag{3.7}$$

Since $f(s_j)$ is a function of all network weights, $E=E(\mathbf{w})$ represents a surface in the space of the weights **w.** The back-propagation algorithm uses gradient descent approach to move on the error surface in the direction of the fastest decrease of the function E. Using gradient descent each weight is updated with increment $\Delta w_i=-\eta(\partial E/\partial w_i)$, where η denotes learning constant (a proportionality parameter which defines the step length of each iteration in the negative gradient direction). The training is finished after a predefined number of iterations or when the error on a separate validation set could not be decreased anymore (Figure 3.7).

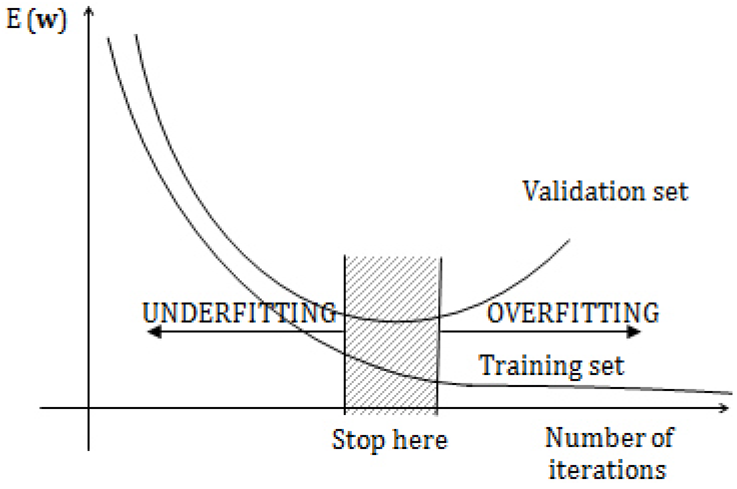


Figure 3.7 Stop criteria for training NN using validation and training sets.

LUC models in this research deal with more than two land use classes (*k*) (Figure 3.6b). For multiclass problems, MLP contains one output neuron for each class. Each training example should be accompanied with a binary vector consisting of all zeros, except on the place that corresponds to the related class. Therefore network outputs are no longer independent of each other and the sigmoid function is no longer appropriate for the output layer neurons. Classification MLP uses *softmax* activation function in which the output of the neuron *k* depends on all network outputs and the sum of the outputs equals to 1:

$$f(s_k) = \frac{e^{s_k}}{\sum_{o} e^{s_o}}, \tag{3.8}$$

Network is trained using the same *back-propagation* method except that the error function is defined to be:

$$E = -\sum_{o} y_o \ln f(s_o). \tag{3.9}$$

3.3.3 Support Vector Machines

Originally, the SVM method (Vapnik 1995, Cristianini and Shawe-Taylor, 2000) was designed as a linear binary classifier that permits *instances* to be classified as only one of the two classes. However, one can easily transform an *n*-classes problem into a sequence of *n* (one-versus-all) or *n*(*n*-1)/2 (one-versus-one) binary classification tasks by using different voting schemes that lead to a final decision (Belousov et al., 2002). Given a binary training set ($\mathbf{x}_i$, y_i), $x_i \in R^n$, $y_i \in \{-1,1\}$, *i*=1,...,*m*, the basic variant of the SVM algorithm attempts to generate a separating hyper-plane in the original space of *n* coordinates (x_i parameters in vector **x**) between two distinct classes (Figure 3.8).

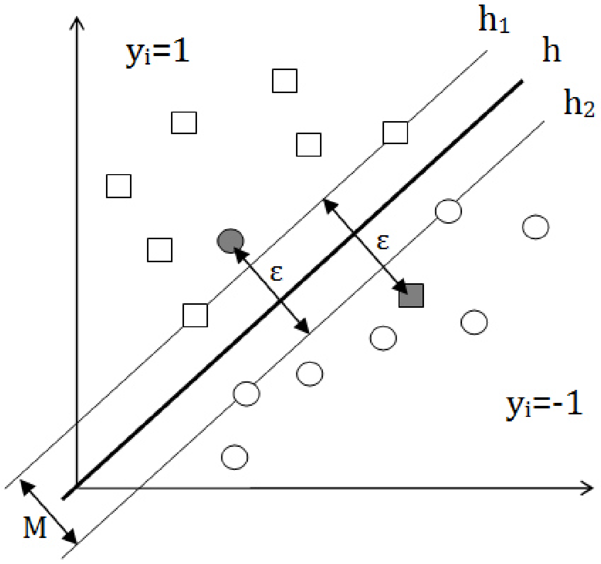


Figure 3.8 General binary classification case (h: $\mathbf{wx}+b=0$; h_1: $\mathbf{wx}+b=1$; h_2: $\mathbf{wx}+b=-1$). Shaded points represent misclassified instances.

During the training phase, the algorithm seeks out a hyper-plane that best separates the samples of binary classes (classes 1 and −1). Let h_1: $\mathbf{wx} + b = 1$ and h_{-1}: $\mathbf{wx} + b = -1$ ($\mathbf{w}, \mathbf{x} \in R^n$, $b \in R$) to be possible hyper-planes such that the majority of class 1 instances lie above h_1 ($\boldsymbol{wx} + b > 1$) and the majority of class −1 fall below h_{-1} ($\mathbf{wx} + b < -1$), whereas the elements belonging to h_1, h_{-1} are defined as Support Vectors. Finding another hyper-plane h: $\mathbf{wx} + b = 0$ as the best separating (lying in the middle of h_1, h_{-1}) involves calculating $\mathbf{w}$ and b, i.e., solving the nonlinear convex programming problem.

The notion of the best separation can be formulated as finding the maximum margin M between the two classes since $M = 2||\mathbf{w}||^{-1}$ maximization of the margin leads to the constrained optimization problem of Equation (3.10):

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \varepsilon_i$$

$$\text{w.r.t}: 1 - \varepsilon_{\text{i}} - y_{\text{i}}(\mathbf{w}\cdot\mathbf{x}+b) \le 0, -\varepsilon_{\text{i}} \le 0, i = 1,\dots,m. \tag{3.10}$$

Despite having some of the instances misclassified (Figure 3.8), it is still possible to balance between the incorrectly classified instances and the width of the separating margin. In this context, the positive slack variables ε_i and the penalty parameter *C* are introduced. Slacks represent the distances between the misclassified points and the initial hyper-plane, whereas parameter *C* models the penalty for misclassified training points that trade-off the margin size for the number of erroneous classifications (the bigger the *C* the smaller the number of misclassifications and smaller the margin). The goal is to find a hyper-plane that minimizes the misclassification errors while maximizing the margin between classes. This optimization problem is usually solved in its dual form (dual space of Lagrange multipliers).

$$\mathbf{w}^* = \sum_{i=1}^{m} \alpha_i y_i \mathbf{x}_i, \mathrm{C} \geq \alpha_i \geq 0, i = 1,...m, \tag{3.11}$$

where **w*** is a linear combination of training examples for an optimal hyper-plane.

However, it can be shown that **w*** represents a linear combination of Support Vectors x_i for which the corresponding α_i Lagrangian multipliers are non-zero values. Support Vectors for which the $C > \alpha_i > 0$ condition holds belong either to h_1 or h_{-1}. Let x_a and x_b be two such Support Vectors ($C > \alpha_a, \alpha_b > 0$) for which $y_a = 1$ and $y_b = -1$. Now *b* could be calculated from $b^* = -0.5w^*(x_a + x_b)$, so that the classification (decision) function finally becomes:

$$f(\mathbf{x}) = \mathrm{sgn} \sum_{i=1}^{m} \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{x}) + b^*. \tag{3.12}$$

In order to cope with non-linearity even further, one can propose the mapping of instances to a feature space of very high dimension: φ: $R^n \rightarrow R^d$, $n << d$, i.e., $\mathbf{x} \rightarrow \varphi(\mathbf{x})$. The basic idea behind mapping into a high dimensional space is to transform the non-linear case into a linear form that can then be applied to the general algorithm already explained in Equations (3.10-3.12). In such space, the dot-product from Equation (3.12) transforms into $\varphi(\mathbf{x}_i) \rightarrow \varphi(\mathbf{x})$. A certain class of functions for which $k(\mathbf{x},\mathbf{y})=\varphi(\mathbf{x})\cdot\varphi(\mathbf{y})$ is true are called *kernels* (Cristianini and Shawe-Taylor, 2000). They represent dot-products in high-dimensional dot-product spaces (feature spaces) and could be easily computed into the original space.

In this dissertation, a Radial Basis Function kernel (Equation 3.13) also known as a Gaussian kernel (Abe 2010), gave encouraging results and was implemented in the experimental procedure.

$$k(\mathbf{x},\mathbf{y}) = \exp\left(-\gamma\|\mathbf{x}-\mathbf{y}\|^2\right). \tag{3.13}$$

Now Equation (3.12) becomes:

$$f(\mathbf{x}) = \operatorname{sgn}\sum_{i=1}^{m}\alpha_i y_i k(\mathbf{x}_i \cdot \mathbf{x}) + b^*. \tag{3.14}$$

In Equation 3.13, γ is used to control the radius of influence of each training point for which α_i is a non-zero value (support vector) to the classification outcome. If γ increases then the number of support vectors for which the related summand in the expansion given with Equation 3.14 has a non-zero value decreases. Therefore, increasing the parameter γ over a certain threshold when C is kept constant, leads to a more complex model (overfitting) since the shape of a decision surface is more influenced by local support vectors. Smaller values for γ produce smoother surfaces (classification outcome depends on many support vectors). Successful SVM models require the optimal combination of C and γ in the process of training. These parameters can be found in the process of cross-validation in which a model is trained on a portion of the training set and validated on the remaining part.

3.4 Data sampling

As DD methods deal with very large volumes of data, it is required to sample the data in smaller sizes as representative samples for model building. It is necessary to create sample data that are large enough to contain significant information required for modelling yet small enough to enable feasible computational processing. There are lots of sampling techniques, however, the most commonly used technique in the LUC field (Yang et al., 2008, Lakes et al., 2010, Okwuashi et al., 2012) is random sampling. A random sample implies that the instances are randomly selected (all instances have an equal chance to be selected).

In many cases land use change modelling deals with a large amount of data in which changes occur within a small percentage of the whole study area. Hence, if random sampling technique is used to create a training set the built models would be biased to predict the majority class (no change). Considering the whole study area containing *N* cells, *n* << *N* cells were sampled in order to build the training set $(x^{t-1}, y^{t})_{k}$, *k*=1,2,…,*n*. The training set contained all changed and an equal number of unchanged cells that were uniformly distributed over the area, thereby preserving original distribution over the classes. This balanced dataset would produce more realistic predictive model with less bias towards the majority class.

In addition, high agreement between the predicted and the actual class measured by a single metric like *kappa* statistics does not necessarily indicate an accurate model in cases when the study area contains a small percentage of changes (van Vliet et al. 2011). Hence, it is necessary to select only the subset of all cells at time *t* to form the test set $(x^{t}, y^{t+1})_{k}$, *k*=1,2,…,*m*, *m* << *N* using the same approach as for the training set.

The proposed sampling approach obtains a more realistic dataset for model creation and its evaluation. The approach was tested in one of the experiments that were carried out in the dissertation.

3.5 Attribute selection

The process of land use changes is complex and influenced by many factors such as physical, social and economical factors (Geurs and Van Wee 2004, Pickett et al. 2001). There are many attributes that describe the process, but usually the main problem is the availability and quality of those data. Commonly used attributes include distances (to transportation networks, schools, industry, commercial and shopping centres and other objects of interest), slope, population and neighbourhood description.

However, irrelevant attributes often confuse the learning process and therefore it is appropriate to perform the attribute selection process. The main idea behind attribute selection is to choose a subset of informative attributes by eliminating those with little or no predictive information (Kim et al. 2003). The four main reasons to perform attribute selection are:

- Improving accuracy of the model,
- Reducing model complexity,
- Reducing overfitting,
- Reducing the time required to train (learn).

The method for attribute (also known as feature) selection can be classified into two types: the *wrapper* and *filter* method. The *Filter* method is independent of the used ML technique and ranks each attribute according to some metric, selecting the highest ranked attributes. The *Wrapper* method, on the other hand, uses the selected method of classification itself. The ML algorithm used is wrapped into the selection procedure (Witten et al., 2011) using cross-validation to calculate the benefits of adding or removing a particular attribute from the used attribute subset (Das, 2001). Both methods have advantages and disadvantages (Das, 2001, Talavera, 2005, Zhu et al., 2007); however, in this dissertation several different filter methods were used and compared.

There are a lot of attribute ranking and selection methods and techniques. Attribute ranking methods rank attributes independently of each other according to their measure of association with the land use class (the nature of the measure is different among the methods, i.e. correlation for χ^2). Since these methods produce a ranking, an additional method must be used to select the appropriate number of attributes (most informative subset of attributes). In this dissertation three ranking methods were used χ^2, Info Gain and Gain Ratio and *recursive attribute elimination* method (Witten et al., 2011) is performed. By using the *recursive attribute elimination* method, a subset of the most informative attribute is obtained in the following manner: build a model using one of the ML techniques (DT, NN or SVM) with all attributes and perform the validation of the obtained model. In the following steps, the lowest ranked attribute is removed and the process is repeated until all attributes have been removed. The obtained results are compared and the first *m* (*m*<*n*) attributes, for which the model obtains the best results, are selected.

In addition to these three ranking methods, a Correlation-based Feature Subset selection method was used. Since this method automatically determines a subset of relevant attributes, it was not necessary to perform a *recursive attribute elimination* method.

In the following text χ^2 and Correlation-based Feature Subset are described, since the Info Gain and Gain Ratio were described in detail in section 3.3.1.

3.5.1 Chi-Square (χ^2)

Chi-Square evaluates attributes based on the χ^2 statistics which tests the independence between an input attribute A and the class attribute C (land use class). Continual attributes are discretized into a several number of intervals after sorting their values.

Let in O_{ij} be the number of observed instances (raster cells) having the value of the attribute A=a_i (or value from the *i*-th interval if A is continuous) and belonging to the class C = c_j. Further let E_{ij} be the corresponding expected number of such instances under the assumption that the attribute A and the class C are mutually independent. The χ^2 statistic is then defined as (Liu and Setiono, 1995):

$$\chi^2 = \sum_{i=1}^{l}\sum_{j=1}^{k}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}}, \tag{3.15}$$

where *l* is the number of A's different values (intervals) and *k* is the number of classes. If A and C are independent then the expected frequency is calculated based on:

$$E_{ij} = \frac{na_i * na_j}{N}, \tag{3.16}$$

where na_j is the number of instances for which A = a_i, and nc_j is the number of instances belonging to class c_j. The statistical degree of freedom for this problem setting is equal to (k-1)·(l-1). After lookup in the table of χ^2 statistics it is possible to accept or reject the independence hypotheses for certain α level.

In principle, higher values of χ^2 indicate stronger dependence between the two attributes.

3.5.2 Correlation – based Feature function (CFS)

The Correlation-based Feature Subset (CFS) (Hall and Smith 1998) evaluates a subset of attributes (features) by considering the individual predictive ability of each attribute along with the degree of redundancy among them. Since the CFS ranks the subset of attributes according to a correlation based heuristic evaluation function, this method favours the subsets of attributes that are highly correlated with the land use class and uncorrelated with each other. After sorting all attributes according to their respective correlation with the class, attributes are added to the subset beginning with the most correlated one. The next attribute is added if it has a higher correlation with the class than with any other attribute already in the subset. The method is therefore capable to automatically determine a subset of relevant attributes.

In order to get information on how predictive a group of attributes is, CFS calculates the heuristic "*merit*" of an attribute subset *S* with *k* attributes:

$$Merit_s = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k-1)\bar{r}_{ff}}}, \tag{3.17}$$

where $\bar{r}_{cf}$ is the mean attribute-class correlation ($f \in S$) and $\bar{r}_{ff}$ is the average attribute-attribute inter-correlation.

Since some attributes have higher and some have lower values, it is necessary to normalize the correlation used in Equation 3.17 to ensure that all attributes have equal effect and that they are comparable. The correlation between two nominal attributes A and B can be measured and normalized using *Symmetrical Uncertainty* compensation (Hall and Smith, 1999):

$$\mathrm{S}U(A,B)=2\times\left[\frac{E(A)+E(B)-E(A,B)}{E(A)+E(B)}\right], \tag{3.18}$$

where E(A) and E(B) are the entropy functions explained in section 3.3.1 (entropy is based on the probability associated with each attribute value) and E(A, B) presents the joint entropy of attributes A and B (calculated from the joint probability of all combinations of values of attributes A and B). Therefore, CFS determines the goodness of a set of attributes using:

$$\frac{\sum_j SU(A_j, C)}{\sqrt{\sum_i \sum_j SU(A_i, A_j)}}, \tag{3.19}$$

where C is the class attribute (target attribute) and the indices *i* and *j* range over all attributes in the set S (Witten et al. 2011)

3.6 Measures for model validation

The predictive performance of the built (trained) models was tested using a test set (x^t, y^{t+1}) that belongs to the future from the perspective of data used to build the models (section 3.2). The model performance on such a “future” test set could be regarded as a realistic assessment of its capability to predict future land use changes. The built models were used in order to predict a land use class for each grid cell at time *t*+1 based on the cell attributes at time *t* ($\mathbf{x}^t \rightarrow y^{t+1}$). The predicted land use classes were compared with the real classes at time *t*+1. Therefore, LUC models are evaluated using validation measures of map agreement.

Since the processes of land use changes are very complex, effective measurement of the predictive performance of each built model requires the use of different validation measures. These measures were analysed in order to compare the real land use state with the predicted scenario: *kappa*, *kappa location*, *kappa histo*, *kappa simulation* and *fuzzy kappa*.

3.6.1 Kappa

Geographical information systems, high-resolution spatial modelling techniques and more accessible remote sensing data offer simple approaches for comparison in which two raster maps can be spatially matched cell-by-cell to estimate the total number of matching cells. This cell-by-cell approach is by far the simplest method, but finer details can be achieved by developing more advanced spatial comparisons using the aforementioned computing tools.

Over the past decade, standard *kappa* statistics was developed, adapted and used frequently to determine the similarity assessment of two raster maps. *Kappa* statistics is not only applied for geographical problems but is used in many other fields including medicine, biostatistics, social sciences etc. There is also a wide number of applications that are related to the validation of ML techniques in spatial modeling (Foody, 2004, Aronoff, 2005, Kovačević et al, 2009).

Kappa statistics can be used to present the level of agreement between two compared maps that, through the preparation of a contingency table, details how the distribution of categories in map A differs from map B. Each element p_{ij} in the contingency table (Table 3.2) indicates the fraction of cells that have category *i* in Map A and category *j* in Map B:

Table 3.2 Generic form of a contingency table.

		Map B					
	Classes	1	2	3	…	N	Σ map A
Map A	1	p_{11}	p_{12}	p_{13}	…	p_{1n}	p_{1+}
	2	p_{21}	p_{22}	p_{22}	…	p_{2n}	p_{2+}
	3	p_{31}	p_{32}	p_{33}	…	p_{3n}	p_{3+}
	…	…	…	…	…	…	…
	N	p_{n1}	p_{n2}	p_{n2}	…	p_{nn}	p_{n+}
	Σ map B	p_{+1}	p_{+2}	p_{+2}	…	p_{+n}	1

The *kappa* index that was introduced by Cohen (1960) presents a measure of agreement adjusted for chance and it is calculated by Equation 3.20.

$$Kappa = \frac{P(O) - P(E)}{1 - P(E)}, \tag{3.20}$$

where *P(O)* presents observed fraction of agreement (Equation 3.21):

$$P(O) = \sum_{i=1}^{n} p_{ii} \,. \tag{3.21}$$

And *P(E)* is the proportion of the fraction of agreement that may be expected to arise by chance (Equation 3.22):

$$P(E) = \sum_{i=1}^{n} p_{i+} \cdot p_{+i} \,. \tag{3.22}$$

Kappa has values between -1, representing no agreement at all, and 1, representing a perfect matching of two maps. The *kappa* index values <0 indicate no agreement and values falling in the ranges of 0-0.20 are categorized as slight, 0.21–0.40 as fair, 0.41–0.60 are categorized as moderate, values between 0.61-0.81 are substantial and values higher than 0.81 are considered as almost perfect as was outlined in similar studies (Landis and Koch, 1977).

3.6.2 Kappa location and Kappa histo

Pontius (2000) was one of the first, who criticized the use of simple kappa statistics for comparison of digital raster maps. He introduced two new statistics indices that include *kappa location* and *kappa quantity* in order to examine the similarity of location and quantity separately. The *kappa location* presents the measurement of similarities in the spatial allocation of categories in the two compared maps. It gives the similarity scaled as the maximum similarity that can be reached with the given quantities and is calculated according to Equation 3.23.

$$K_{location} = \frac{P(O) - P(E)}{P(\max) - P(E)} \tag{3.23}$$

P(max) presents maximum fraction of agreement given the distribution of class sizes and is calculated according to Equation 3.24.

$$P(\max) = \sum_{i=1}^{n} \min(p_{i+}, p_{+i}) \tag{3.24}$$

Unfortunately, the *kappa quantity* introduced by Pontius has proven to be unstable and incomprehensible in various studies (Sousa et al., 2002). In order to overcome the drawbacks of the *kappa quantity* statistics, Hagen (2002) introduced a new statistical index called *kappa histo* (kappa histogram). *Kappa histo* can be calculated directly from the histograms of two maps by Equation (3.25).

$$K_{histo} = \frac{P(\max) - P(E)}{1 - P(E)} \tag{3.25}$$

The mutual relationship between *kappa location*, *kappa histo* and standard *kappa* could be expressed as Equation 3.26.

$$Kappa = K_{histo} \cdot K_{location} \tag{3.26}$$

Same as *kappa, kappa location* can get values from -1 to 1, were -1 represent no agreement at all at specifying location, 1 perfect agreement at specifying location and 0 indicates the agreement as can be expected by chance. *Kappa histo* has values between 0 and 1, where 1 indicates a perfect agreement and 0 indicates that there is no agreement at all in the class sizes (Vliet, 2011).

3.6.3 Kappa simulation

In order to assess the accuracy of land use change models, van Vliet (2009) proposed a new kappa index called *kappa simulation*. The accuracy of the land use change models is mainly achieved by comparing actual land use with the simulation (prediction) result. Van Vliet started from the fact that changes in land use cover only a small percent of the total study area in the most of LUC models. He introduced a statistic similar to kappa statistics with a more appropriate stochastic model of random allocation of class transitions relative to the initial map.

It is found that only *kappa simulation* truly tests models in their capacity to explain LUC over time, but unlike *kappa* it does not inflate results for simulations where little change takes place over time (van Vliet et al., 2011).

Furthermore, by considering the distribution of class transitions (interpreted as conditional probabilities), *kappa statistics* is modified by integrating the amount of land use changes in the expected agreement. Therefore, the chance of finding a certain class at a location will depend on the class that was originally there. It is necessary to express the size of class transitions as a function of the original land use map and the simulated or actual land use map. The fraction of cells that changed from land use *j* in the original map to land use *i* in the simulated land use map S express as p(i^S |j^O) and A as p(i^A | j^O) for the actual land use map A accordingly. Because the original land use map (O) is the same for both the simulated and actual LUC, the expected agreement between the simulated land use map and the actual land use map can be expressed by Equations 3.27, 3.28 and 3.29.

$$P(E)_{simulation} = \sum_{j=1}^{n} P_j^O \cdot \sum_{i=1}^{n} \left(P(i^A \mid j^O) \cdot P(i^S \mid j^O) \right), \tag{3.27}$$

where P(E) simulation defined the expected fraction of agreement, given the sizes of the class transitions.

$$P(\max)_{simulation} = \sum_{j=1}^{n} P_j^O \cdot \sum_{i=1}^{n} \min\left(P(i^A \mid j^O), P(i^S \mid j^O) \right), \tag{3.28}$$

where P(max) presents the maximum accuracy that can be achieved given the sizes of the class transitions.

$K_{Simulation}$ is the coefficient of agreement between the simulated land use transitions and the actual land use transitions and it states as follows:

$$K_{Simulaton} = \frac{P(O) - P(E)_{simulation}}{1 - P(E)_{simulation}}. \tag{3.29}$$

Same as standard *kappa*, *kappa simulation* is the result of two types of similarities, *kappa transloc* and *kappa transition*. *Kappa transition* indicates the similarity in class transitions, while *kappa transloc* indicates the similarity in the allocation of these transitions. These two values can by calculated using Equations 3.30 and 3.31 (van Vliet et al., 2011).

$$K_{transloc} = \frac{P(O) - P(E)_{simulation}}{P(\max)_{simulation} - P(E)_{simulation}} \tag{3.30}$$

$$K_{transition} = \frac{P(\max)_{simulation} - P(E)_{simulation}}{1 - P(E)_{simulation}} \tag{3.31}$$

The values of *kappa simulation, kappa transloc* and *kappa transition* have the same range as *kappa, kappa location* and *kappa histo,* respectively.

3.6.4 Fuzzy Kappa

The latest approach in assessing similarities of raster maps is based on fuzzy set theory (Zadeh, 1965). Geoscientists and GIS professionals adopted this theory (Burrough, 1996; Burrough and McDonnell, 1998) with the purpose of characterizing inexactly defined spatial classes or entities that deal with ambiguity, vagueness and ambivalence in mathematical or conceptual models of spatial phenomena. Based on fuzzy set theory, Hagen (2003) proposed the new approach in assessing spatial similarities and changes between raster maps. The fuzzy-based map-comparison method was primarily developed for the calibration and validation of the cellular automata models for land use dynamics.

Fuzziness can be considered from two aspects; a) locational based on the concept that the fuzzy representation of a raster cell depends on the cell itself and, to a lesser extent, also on the cells in its neighbourhood, and b) categorical which originate from vague distinctions between categories.

The extent of the neighbouring cells or locational fuzziness could be expressed by a distance decay function. The categorical fuzziness can be introduced by setting off-diagonal elements in the Category Similarity Matrix to a number between 0 and 1 that corresponds to membership values of different categories. Since there are no straightforward rules for assigning membership values, choosing values in the matrix is subjective and it could be selected on the basis of a priori experience.

The *kappa fuzzy* index is similar to the traditional kappa statistic in that the expected percentage of agreement between two maps is corrected for the fraction of agreement that is statistically expected from randomly relocating all cells in compared maps:

$$K_{fuzzy} = \frac{P(0) - P(E)_{fuzzy}}{1 - P(E)_{fuzzy}}, \tag{3.32}$$

$$P(E)_{fuzzy} = \sum_{i=0}^{R} E(i) * M(d_i), \tag{3.33}$$

where *R* is the number of the furthest neighbourhood ring, *E* is probability function of *i*-th neighbourhood ring calculated for each combination of categories for matched cells, *M* is the fuzzy membership function and d_i is the radius of the *i*-th ring.

Chapter 4:

Study area, preparation and data analysis

4.1. Analysis of social-economic aspects of study area

Belgrade, the capital city of Republic of Serbia (Figure 4.1), is situated at the confluence of the Sava and Danube rivers, and surrounded by Pannonian Plain, on the north side and Avala (511 m) and Kosmaj (628 m) mountains, on the south side. Belgrade lies on the average elevation of 117 m, at geographic coordinates latitude 44°49’14’’N and longitude 20°27’44’’E. The administrative area covers 3,223 km^2 and the city has around 1.6 million inhabitants. Its territory is divided into 17 municipalities that comprise of 157 settlements (census designated places, CDPs). The urbanized area and the inner part of the city of 775 km^2 include 11 urban municipalities with 32 CDPs. In the period 2002-2011 the Region of City of Belgrade has had an increase of population (approx. 4%) while in other Regions in Serbia population decreased from 5% in the Western and Northern to 11% in the Southern and Eastern Serbia (Petrić et al., 2012).

Belgrade is one of the oldest cities in Europe and has a rich, vivid and long history that dates back from over seven thousand years ago. The Belgrade area was developed under different cultural, social and economic conditions as part of many different reigns. The most intensive demographic, socio-economic and socio-geographic changes of the 20th century in the territory of Serbia took place between the 1960’s and 1980’s, thereby dramatically altering the organization and form of space use. The major causes of these changes were the distinctly planned industrialization of the former Yugoslavia as well as the politically initiated

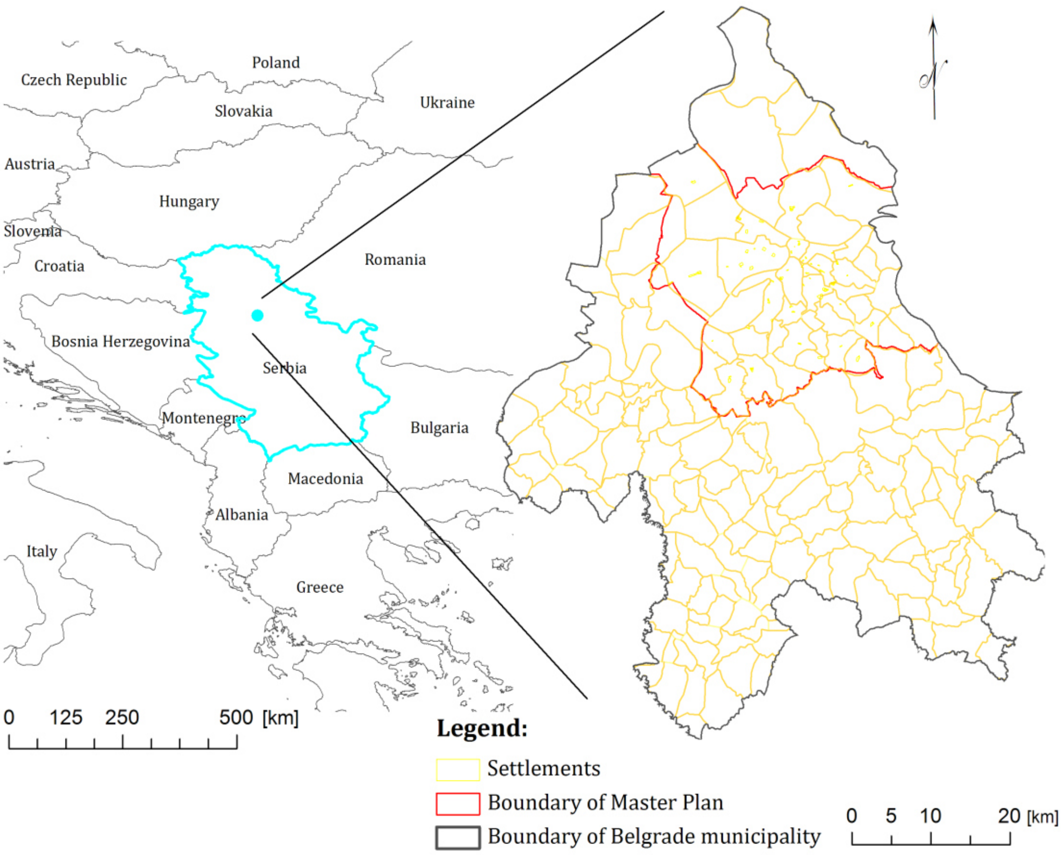


Figure 4.1 Location of Belgrade.

urbanization and deagrarization.

The development of Belgrade and its agglomeration has several stages in spatio-morphological, economic and demographic development (Vojković et al., 2010). At the beginning of the 20th century, the central Belgrade area covered only about 12 km^2 with about 70,000 inhabitants (in the year 1900), while the administrative territory of the Belgrade district, in that time, spread over the area of 2,025 km^2 with about 126,000 inhabitants. Due to accelerated industrialization and abrupt urbanization, the Belgrade area permanently grew in the second half of the 20th century and changed its spatio-functional structure. From the end of World War II to the 2002 census, Belgrade multiplied the number of its inhabitants by 2.5 times. Intensive demographic growth was a result of migration flows and territorial expansion whereby new settlements were included in the administrative town area and immigrant streams were intensive. Powerful and disorganized migration, not only from the territory of Serbia but also from the other republics of former Yugoslavia, proves the significance of Belgrade in broader surroundings.

Until the 1970's, the strict urban area encompassed the majority (90%) of Belgrade's total population growth. In that first period of urban development, right after the World War II, the highest population growth rates were found in the central Belgrade municipalities. As the old central town core had already been urbanized and densely inhabited, higher growth rates were also established in the broader zone of the Belgrade urban area during the inter-census period from 1953-1961. Numerous settlements from the immediate hinterlands and suburban municipalities of the time were losing their population as they kept moving to Belgrade. The intensive industrialization process expanded from the strict urban area towards peripheral zones during the period between the census years 1961 and 1981, resulting in the harmonization of the growing population with employment in industry. The central Belgrade area (consists of parts of 10 town municipalities, Master Plan area) is characterized by specific demographic development and polarization of demographic trends: a) depopulation of the oldest urban core of the town (municipalities Stari Grad, Vračar and Savski Venac); b) dynamic population growth in the municipalities of Voždovac, Zvezdara, Zemun and Palilula; c) intensive concentration of population in the municipalities of New Belgrade, Čukarica and Rakovica. In the economic structure of Belgrade, the predominant activities are those of the tertiary-quaternary sector, with slow modernization of industry. The most important spatial changes caused by deindustrialization are visible in the central zones of the strict town core (desistance of productive activities, but often without formal change of land use due to incomplete company restructuring processes) and in industrial centers in the broader town area. Reindustrialization is a feature of the peri-urban agglomeration zone and is mainly occurring along traffic corridors. The traditional Belgrade and Zemun town centers, which have a distinct concentration of business-related contents from the preceding period, are gradually losing their primacy in comparison with the dynamic development of business centers in the New Belgrade zone (Vojković et al., 2010).

Due to the amount of data preparation, which was time consuming, and limited capacity of computer hardware used, it was necessary to separate only a part of Belgrade area for further consideration. Therefore, the study area used includes four municipalities: Zemun, New Belgrade, Surčin and a part of Dobanovci municipality. However, since only a fraction of Dobanovci municipality is used, it is further, in most cases, considered as a part of Surčin municipality in this dissertation (Figure 4.2). Thus, these three neighbouring municipalities highlighted in the Figure below were selected because they represent completely different urban types.

The Old Core of Zemun constitutes an integrated urban phenomenon expressed in a multiplicity of shapes, contents and meanings, and is designated as a national historic site within the city (Grozdanić, 2010). The Master Plan of Belgrade includes the development of two suburban settlements: Zemun Polje and Batajnica, and the further development of one urban municipality, Zemun. The municipality of Zemun occupies an area of 9,942 ha of the Master Plan and contains a population of about 15,000.

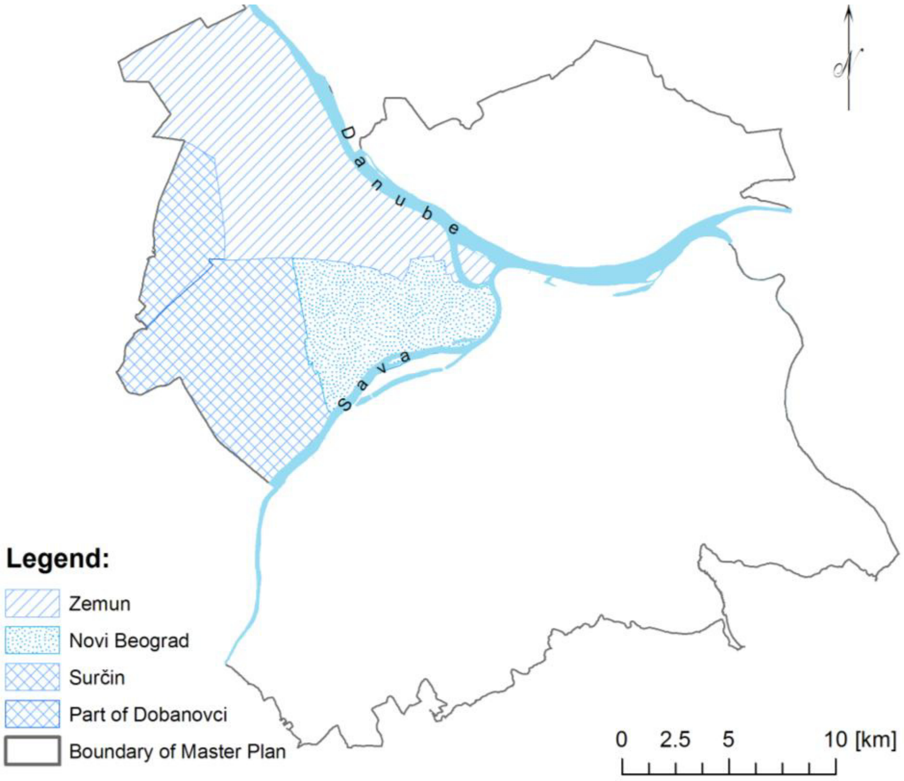


Figure 4.2 Study area.

The construction of New Belgrade began after World War II. It was conceived as a modern city on the left bank of the Sava River and played a key role in transforming the previous capitalist image of the city into socialist one (Marić et al., 2010). The municipality of New Belgrade is divided into large rectangular residential blocks that are separated by wide boulevards. The population density of New Belgrade is the highest in the city and contains 220,000 inhabitants within an area of 4,096 ha. New Belgrade has recently become the commercial center of Belgrade.

Until 2004, Surčin was a part of the territory of Zemun, however it was later formed as a separate municipality consisting of seven settlements (villages). Most of the settlements of Surčin are situated within the boundary of the Master Plan. The airport complex "Nikola Tesla" is situated in the north-eastern part of the municipality and has a significant influence on Surčin's spatial development. As already mentioned, this study included the settlement of Surčin and part of the settlement Dobanovci. The total research area of Surčin covers 6,119 ha with a population of about 41,000.

The main study region covered an area of about 20,157 ha and was buffered by 100m on each side to minimize any potential edge effects.

4.2. Used data and software

Since the land use changes present complex process influenced by many different factors, it was necessary to collect and prepare different types of available data, which represent land use state in several different moments in time. The main problem was to collect different types of data (such as land use map and population) for same years. Consequently, four years were chosen: 2001, 2003, 2007 and 2010. GIS database was created based on the following data:

- Actual land use maps for two years (2003 and 2010),
- Map of Master Plan of Belgrade for 2021,
- Orthophoto for four time years(2001, 2003, 2007 and 2010),

- Available census data (2001, 2002, 2003, 2007, 2010 and 2011),
- Soil Sealing raster map and
- Residential Building Blocks Layer.

The collection, preparation and analysis of data lasted for two years and several various software were used. GIS database was created using ArcGIS (ESRI, 2011) software. ArcGIS software and SAGA (System for Automated Geoscientific Analyses) GIS environment (Böhner, J., et al., 2008) were used in order to create attributes and to analyze the data. The Java programming routines have been developed in order to generate some attributes and all datasets which were used for the model building and its validation. The Map Comparison Kit (MCK) (Visser and de Nijs, 2006) software was used for assessment of similarity between planned and actual land use maps.

4.2.1 Creation of Land use maps

Land use maps were created based on four orthophoto maps from 2001, 2003, 2007 and 2010 and actual land use maps from 2001 and 2010 (vector maps in GIS environment).

Actual land use maps were obtained from Urban Planning Institute of Belgrade as well as map of Master Plan of Belgrade for 2021. Land use classification differs due to different development priorities of local authorities (Đorđević, 1997). Since the classification of those three maps was different, it was necessary to adopt a common classification which will be used for the further research.

Classification of land use was achieved by generalizing the official 13 classes of land use outlined in the 2021 Master Plan of Belgrade into 9 classes, considering classification on actual land use maps from 2001 and 2010 as follows (Table 4.1): *Agricultural*, *Wetlands*, *Traffic areas*, *Infrastructure*, *Residential*, *Commercial*, *Industry*, *Special use* and *Green areas*.

Table 4.1 Classification of land use.

Land use class 2001	Land use class 2010	Land use class Master plan 2021	Used Land use class
1. Agricultural	1. Agricultural	1. Agricultural	1. Agricultural
2. Wetlands	2. Wetlands	2. Wetlands	2. Wetlands
3. Traffic areas	3. Traffic areas	3. Traffic areas	3. Traffic areas
4. Infrastructure	4. Infrastructure	4. Infrastructure	4. Infrastructure
5. Residential	5. Residential	5. Residential	5. Residential
6. Commercial (Centre)	6. Commercial (Centre)	6. Commercial (Centre)	6. Commercial (Centre)
7. Industry	7. Industry	7. Industry	7. Industry
8. Special use (Public service)	8. Public service (Special use)	8. Public service	8. Special use
a. Culture	a. Culture	9. Sports complex	9. Green areas
b. Science	b. Science	10. Green areas	
c. Education	c. Education	11. Forests and forest land	
d. Health services	d. Health services	12. Protective vegetation.	
e. Social protection	e. Social protection	13. Communal areas (Cemetery)	
f. Religious	f. Religious		
g. Special service	g. Special service		
h. Other	h. Other		
i. Sport	9. Sports complex		
9. Green areas	10. Green areas		
a. Parks	Parks		
b. Cemetery	Recreation		
c. Recreation	11. Communal areas (Cemetery)		
10. Not built	12. Not built		
11. Farms			

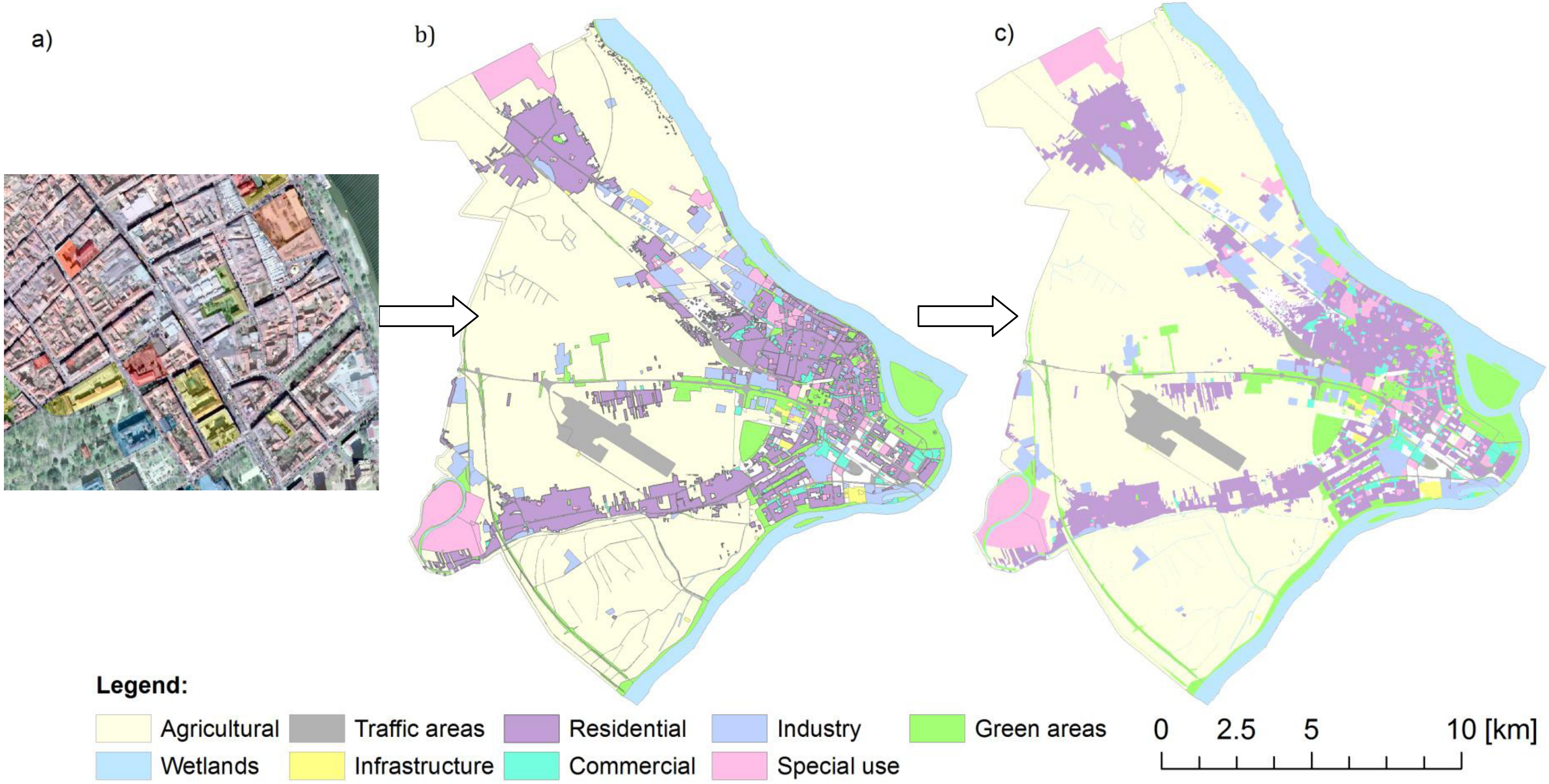


Figure 4.3 a) Digitalization b) Polygons of land use class for year 2001 c) Raster map of land use class for year 2001.

Maps of actual and planed land use class are obtained in vector format (polygons of land use class), so that each class is represented as a single GIS layer. Those layers were reclassified into nine previously defined classes.

As already mentioned, the shapes representing land use classes were induced from actual land use maps and orthophoto for all four years (Figure 4.3a). The main characteristics of used orthophoto maps are present in Table 4.2.

Table 4.2 The main characteristics of orthophoto maps.

Company	Date of Aerophotogrammetric survey	Resolution [m]	Number of Bands
MapSoft	2001	0.30×0.30	1 BV
MapSoft	2003	0.30×0.30	3 RGB
MapSoft	2007	0.25×0.25	3 RGB
Geo Info Strategies	2010	0.20×0.20	3 RGB

The polygons of Not built class (indicated in the Table 4.1) on actual land use maps for years 2001 and 2010 were predefined as *Agriculture* or *Green areas* based on actual state detected on ortophoto maps. Furthermore, after correcting observed irregularities such as overlapping polygons and undefined areas, all individual *.shp files of classes were merged into a single one which represents polygons of land use classes for a given year. The maps of land use classes for 2003 and 2007 were created by digitizing, based on orthophoto maps of the respective years.

In order to represent study area as grid cell (explained in section 3.2) all maps of land use class were converted from vector to raster format with appropriate resolution. Since the scale of Master Plan of Belgrade is 1:20.000 (URBEL, 2003), the corresponding resolution of cells is 20×20 m. However, by using this resolution a lot of information regarding changes will be lost during rasterization. Therefore, the land use polygons (Figure 4.3b) for all four years were rasterized at a 10×10 m resolution, where each grid cell was associated with corresponding land use class (Figure 4.3c).

In order to build the model of land use changes, beside maps of land use class it was necessary to consider additional information that has influence on these changes. Therefore, the auxiliary maps representing additional attributes were created and they contain information on *accessibility*, *population density* and *spatial neighbourhoods*. Due to the relatively flat terrain of the study area, attributes regarding the elevation were not taken into consideration.

4.2.2 Accessibility maps

After the analysis of land use change direction for the period 2001-2010 and consultations with urban planners, the following accessibility raster maps referred to the distance variables were created:

- Euclidean distance of grid cell to city centre,
- Euclidean distance of grid cell to municipality centre,
- Euclidean distance of grid cell to the closest rivers (Danube and Sava),
- Euclidean distance of grid cell to the closest big green areas (only areas greater than 10 ha) at time *t*,
- Euclidean distance of grid cell to the closest railway lines at time *t*,
- Euclidean distance of grid cell to the closest highways at time *t*,
- Euclidean distance of grid cell to the closest main roads at time *t*,
- Euclidean distance of grid cell to the closest streets I category at time *t*,
- Euclidean distance of grid cell to the closest streets II category at time *t*.

Those accessibility maps were created for all four moments in time with 10×10 m resolution, taking care for created grid cells to be spatially overlapped with already established grid cells of land use maps. Some of accessibility maps for year 2001 are presented in Figure 4.4.

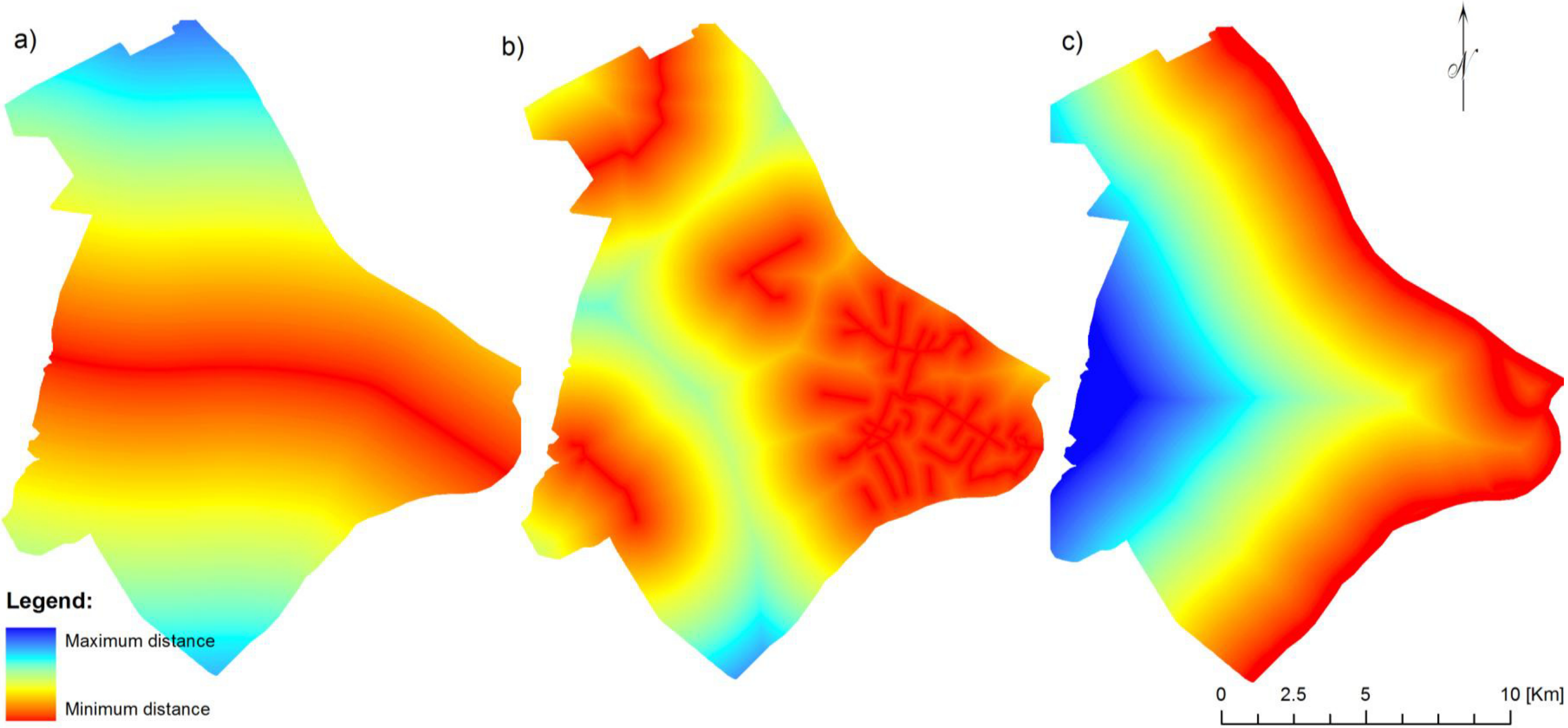


Figure 4.4 Map of Euclidean distance of grid cell to the closest: a) Highway in 2001, b) Street I category in 2001 and c) Rivers.

4.2.3 Population maps (Dasymetric modelling of population)

Precise presentation of population distribution and its dynamics in urban planning is important for several reasons, including:

- Understanding the directions and intensity of population redistribution in order to determine the strategic trends of urban area development and
- To provide an accurate depiction of population density for the purposes of urban planning and restructuring.

Therefore, special attention is given to attribute that describes the population distribution for all four years of interest and as well as dynamics of population between two censuses.

In the Republic of Serbia, publicly available census data are presented on the level of census designation places (settlements - municipality) which are usually graphically presented as choropleth maps. However, a main drawback to presenting population density data in choropleth maps is that uninhabited areas become misrepresented since the aggregation of census data results in the construction of statistical surfaces for inhabited areas only.

Furthermore, using choropleth map in order to created attribute which describes population, in this dissertation, does not make sense because all grid cells located in a municipality have the same value of attribute (all cells in one municipality have uniform distribution)(Figure 4.5).

Hence, in order to aggregate the population data and model population changes between two census years on the level of spatial units (grid cell), using publicly available data, a methodology for dasymetric mapping of population was developed. A detailed description of the proposed methodology and the achieved accuracy is presented in the paper Bajat et al (2013), and only a brief overview will be given in this dissertation.

The proposed methodology for dasymetric mapping is a modified formula for the estimate of population in buildings as defined by Lwin and Murayama (2009).

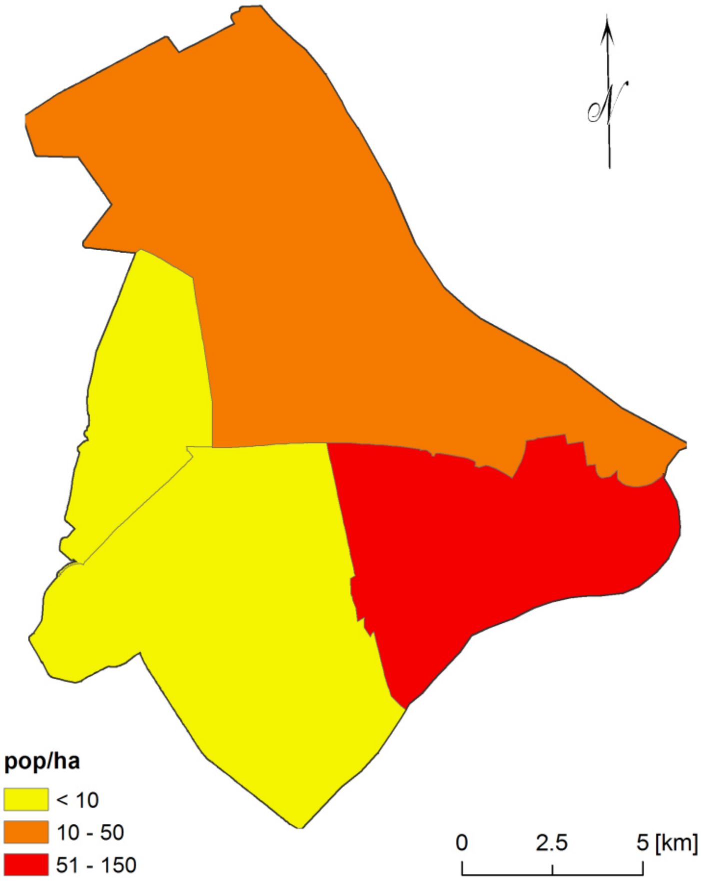


Figure 4.5 Choropleth map of study area for the estimated population for year 2001.

They proposed two methods, the first being areametric and the second being volumetric. Each of these methods is based on footprint layer for each building in the considered area. The proposed formula for the volumetric method reads:

$$\mathrm{BP}_i = \left(\frac{CP}{\sum_{k=1}^{n} BA_k \cdot BF_k} \right) BA_i \cdot BF_i \ , \qquad (4.1)$$

where BP$_i$ is the population of the building *i*, *CP* – the census tract population, BA_i - the footprint area of the building *i*, BF_i- the number of floors in the building *i*.

A modified formula which would substitute the footprint and number of floors by soil-sealing and height typology would read:

$$\mathrm{Bs}_p = \left(\frac{CP}{I_n}\right)\mathrm{Ss}_p \cdot \mathrm{Ts}_b,\quad I_n = \sum\left(\mathrm{Ss}_p \cdot \mathrm{Ts}_b\right), \tag{4.2}$$

where Bs$_p$-number of inhabitants per target grid cell, Ss$_p$-soil sealing value per grid cell, Ts$_b$- building block height typology weights, *CP*-total number of inhabitants within census designation place, I$_n$- census designation place index that corresponds to total sum of multiplication of soil sealing values and building typology.

In this way, the population data are directly disaggregated to the grid cell level. The proposed formula ensures that the total number of people within municipality area remains the same. This is referred to as the pycnophylactic property of dasymetric maps (Tobler, 1979b).

The developed method uses primarily publicly available national statistics data as well as standard data related to land use that is already in the planners' possession. In order to obtain the dasymetric (population) maps for all four years of interest and population changes between two censuses, following data were used:

- Population counts per municipality for 2002 and 2011 from the Serbian Census and official estimate of population for 2001, 2003, 2007 and 2010 (Statistical Office of the Republic of Serbia) (Table 4.3),
- Soil Sealing Database and
- Residential Building Blocks Layer.

Table 4.3 Population per municipality.

Municipality	Census		Official estimate of population			
	2002	2011	2001	2003	2007	2010
Zemun	145751	151811	157240	154129	157021	161531
New Belgrade	217773	212104	225470	217361	219208	218504
Surčin	14292	17356	39260	38422	39615	40974
Dobanovci	162	157	162	160	160	158

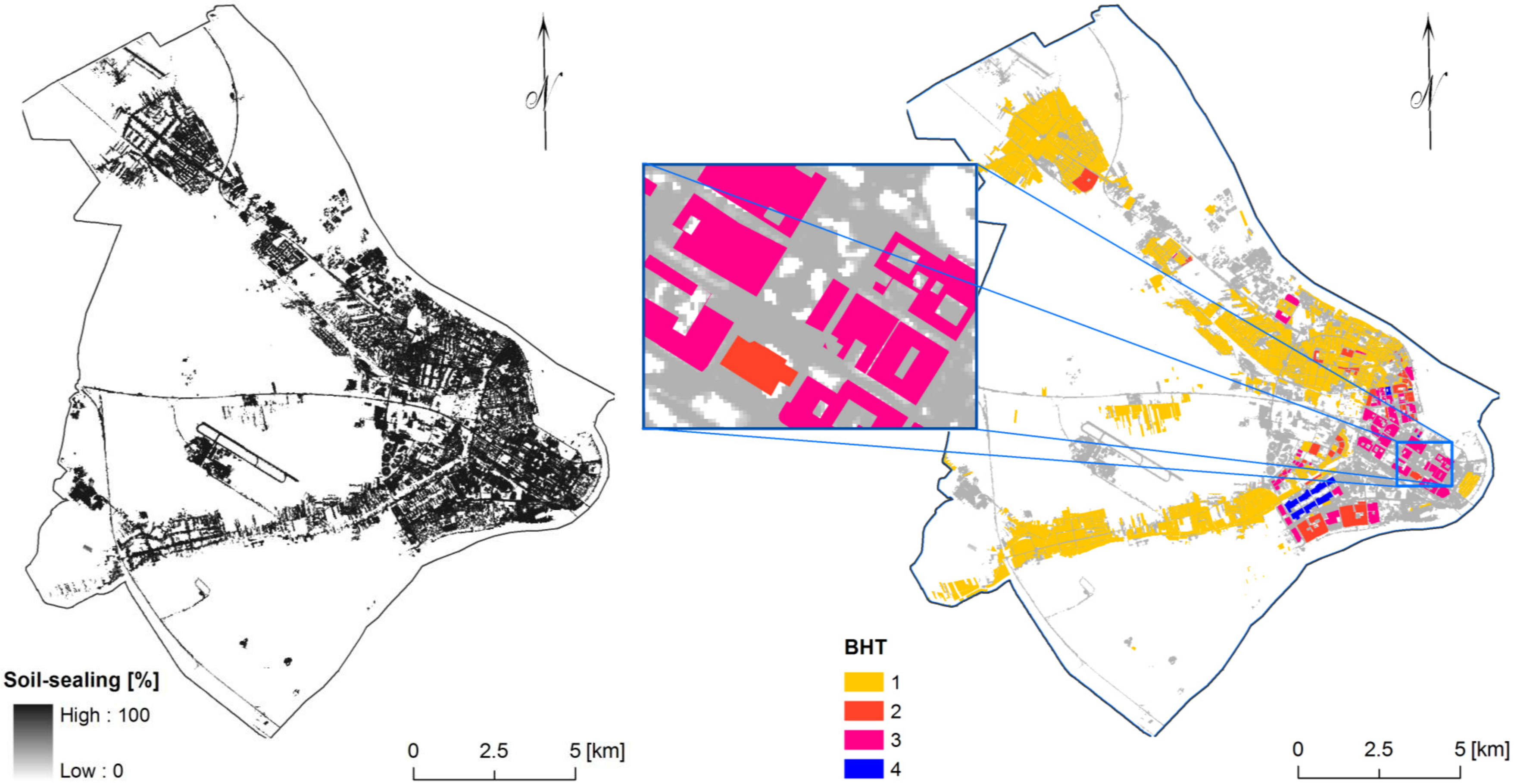


Figure 4.6 a) Soil - Sealing data layer, b) Residential Building Blocks Layer for the year 2001.

A Soil - Sealing (or imperviousness) high-resolution raster layer (Figure 4.6a) was produced during 2006-2008 as part of the Global Monitoring for Environment and Security (GMES) programme with the aim to complement the CORINE (COoRdinate INformation on the Environment) (Nestorov and Protić, 2009) land cover data. The need for production of five high-resolution land cover layers emerged on behalf of the European Environment Agency (EEA): imperviousness, forest, grassland, wetland and water. The database is available in two spatial resolutions of 20 m and 100 m (European Environmental Agency, 2010). For the purpose of this dissertation, the 20 m resolution database has been used after being resampled to 10 m resolution.

The residential blocks are an integral part of planning documents which were made previously for the Belgrade Master Plan in the year 2000 and are in digital form appropriate for the GIS environment, i.e. they are presented in vector format (*.shp files) (Figure 4.6b).

As it can be seen on Figure 4.6b, a building block is designated as a residential area clearly delimited by roads. In this dissertation, Residential Building Blocks were generated separately for each year of interest as a by-product of digitalization of residential class.

Table 4.4 Building height typology of residential blocks.

Number of storeys	Building Height Typology (BHT)	Weights
up to 3 (ground floor [GF]+1+ garret[G])	1	1
4-5 (GF+3+G)	2	3
6-8 (GF+6+G)	3	5
above 9 (GF+6+G)	4	7

An attribute associated to each block is its building height typology (BHT). The typology defines 4 classes in compliance with national regulations (Table 4.4). Moreover, the obtained weight coefficients (Table 4.4) correspond to the mean

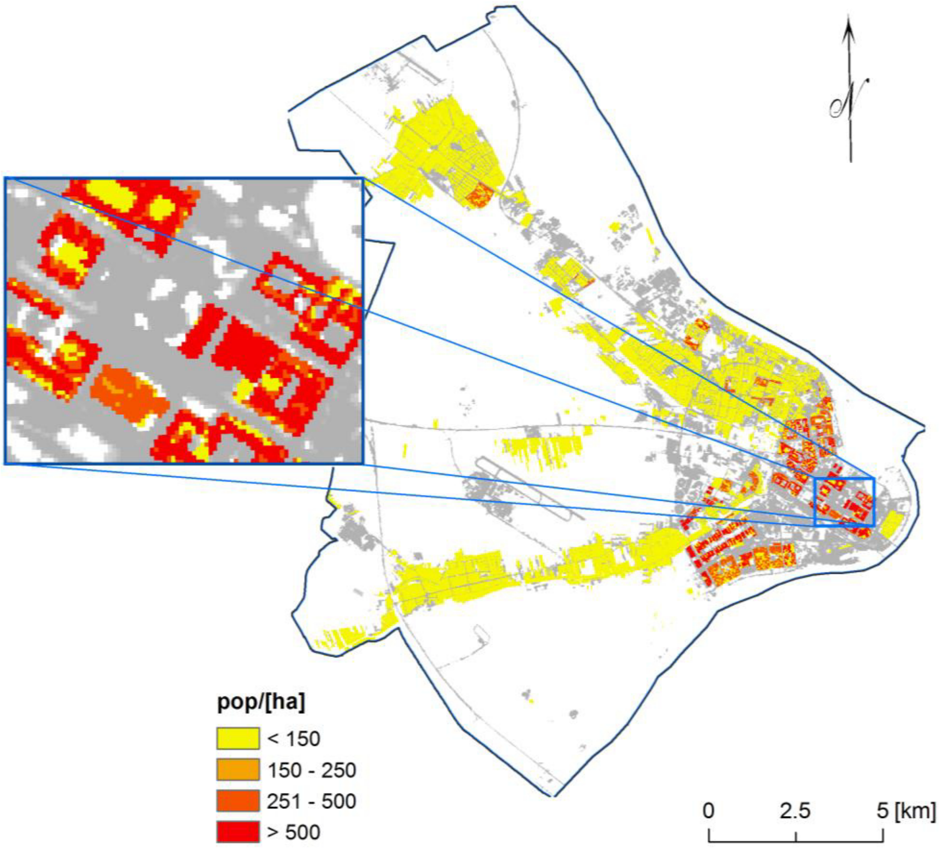


Figure 4.7 Dasymetric map depicting population density for the year 2001.

number of inhabitants who would reside in a vertical line of a building on a 20 m^2 area based on the average number of square meters and structure of housing units which correspond to each class (having in mind the gross area inclusive of staircases and hallways in buildings).

Using the previously explained methodology, for all six years of interest dasymetric maps were created. The dasymetric map (Figure 4.7) depicts the population density using the grid cell as the basic unit. Grey colored areas on the map mark the soil-sealing layer which does not overlap with residential blocks.

Additionally, the Population Change Index (PCI) is also presented as an attribute that provides a standardized measure for comparing population changes over time and across study area. PCI represents the ratio of change in the number of inhabitants per each cell between two censuses, 2002 and 2011 (Bajat et al, 2013).

Dasymetric maps for 2002 and 2011 were created, using previously described methodology, after which the PCI map was created. The PCI map is generated by incorporating map algebra, i.e. the two grids division operation:

$$PCI_p = \left(\frac{Bs_p^{2011}}{Bs_p^{2002}} \right) \cdot 100\% , \qquad (4.3)$$

where PCI_p – population change index per target cell, Bs_p^{2011}-number of inhabitants in year 2011 per target grid cell, Bs_p^{2002}-number of inhabitants in year 2002 per target grid cell. Generated PCI map is presented in Figure 4.8.

In classic studies, PCI is usually represented on the level of administrative units such as in the case of census data. The described methodology obtains data on the grid cell level which enables subsequent aggregation of the data to the level of a spatial unit suitable for specific application.

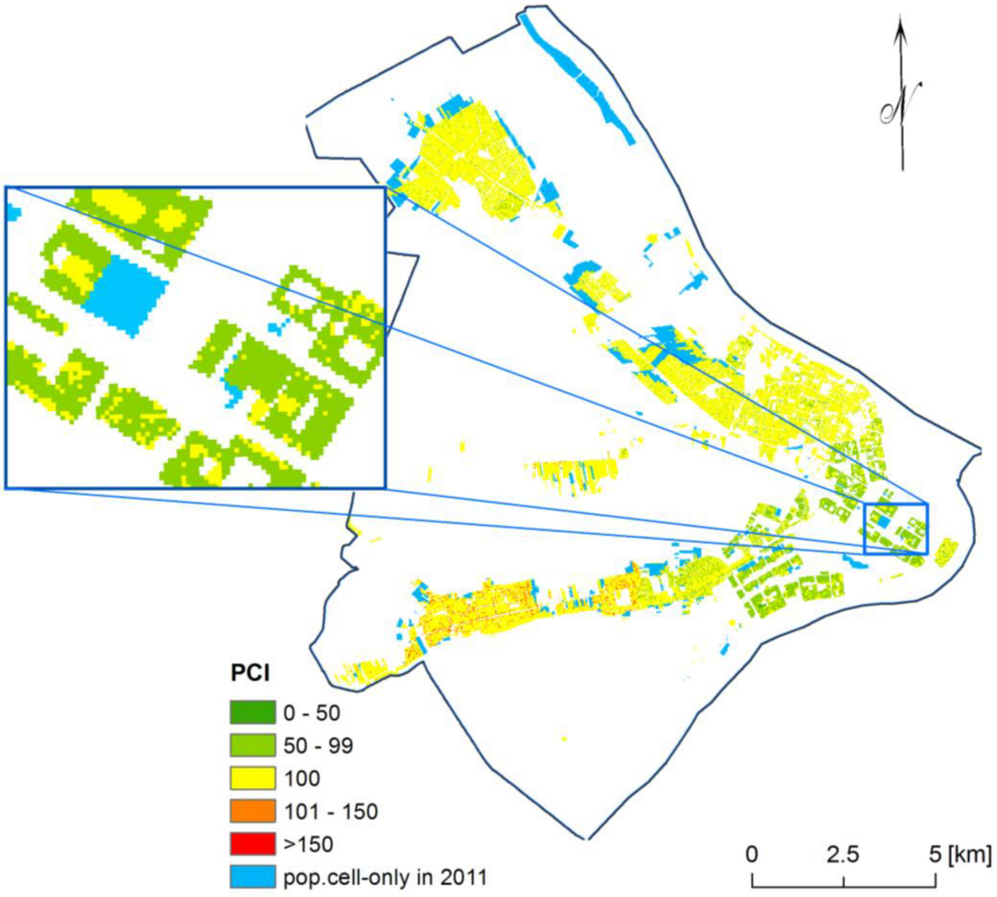


Figure 4.8 Population dynamics modeling between two censuses by PCI map.

4.3. Creation of datasets and attributes used for modeling

For each year of interest (2001, 2003, 2007 and 2010) all grid cells of study area are represented as vectors of attributes $\mathbf{x}^t$ (Table 4.5). Those attributes represent the value of previously created maps (which contain information on land use class, different accessibility, population density and PCI) and are associated with particular cell $\mathbf{x}^t$ ($\mathbf{x}^t$=<$x^t{}_1$, $x^t{}_2$,..., $x^t{}_n$>).

Table 4.5 Basic attributes.

Attributes		Description
x_1	Municipality	Zemun, New Belgrade, Surčin and Dobanovci-
x_2	ed. city centre	Euclidean distance of grid cell to city centre
x_3	ed. Centre municipality	Euclidean distance of grid cell to municipality centre
x_4	ed. River	Euclidean distance of grid cell to the closest rivers (Danube and Sava)
x_5	ed. Green	Euclidean distance of grid cell to the closest big green areas
x_6	ed. Railway	Euclidean distance of grid cell to the closest railway lines at time *t*
x_7	ed. Highway	Euclidean distance of grid cell to the closest highway at time *t*
x_8	ed. Main road	Euclidean distance of grid cell to the closest main road at time *t*
x_9	ed. str. I category (t)	Euclidean distance of grid cell to the closest street I category at time *t*
x_{10}	ed. str. II category (t)	Euclidean distance of grid cell to the closest street II category at time *t*
x_{11}	No. of inhabitants (t)	Number of inhabitants at time *t*
x_{12}	PCI	Population Change Index between two census
x_{13}	Class (t)	Land use class at time *t*

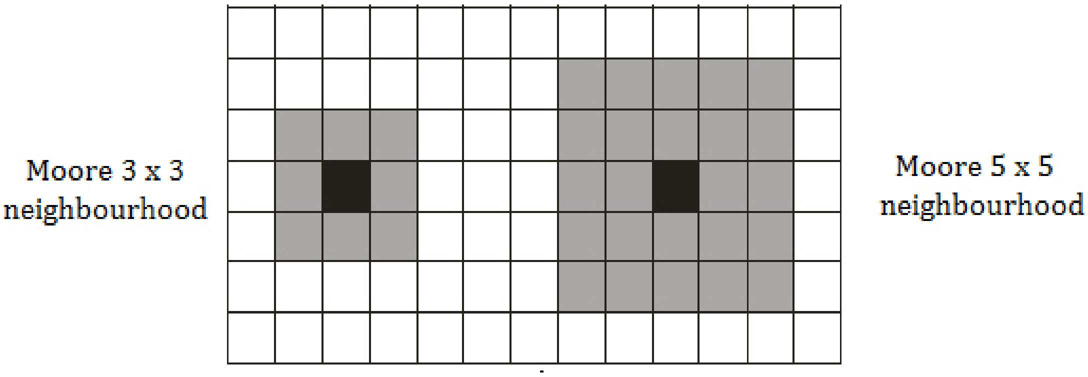


Figure 4.9 Moore neighbourhoods.

Considering that study area covers four municipalities which represent different urban types, an additional attribute was defined containing information on cells location (in which municipality the cell is located).

Those vectors $\mathbf{x}^t$ ($\mathbf{x}^t$=<$x^t{}_1$, $x^t{}_2$,..., $x^t{}_n$>) are created in ArcGIS as *.shp files per each year, where cells are represented as point with corresponding attributes. After that, four *.shp files are exported as *.txt files. These files are to be used for further processing and creation of additional attributes as described in following paragraphs.

Guided by the Waldo Tobler's (Tobler, 1970, page 3) first law of geography "Everything is related with everything else but near things are more related than distant things" additional attributes were defined. They represent spatial neighbourhood within the very local area determined by the Moore's neighbourhood (Figure 4.9).

After the analysis was performed (by testing and comparing several different neighbourhood sizes) two neighbourhood sizes, 7×7 and 21×21 cells, were chosen and used to generate attributes represented in Table 4.6.

Table 4.6 Attributes describing neighbourhood.

Attributes		Description
x_{14}	Neighbours 1	Most frequent land use class in Moore neighbourhood 7x7 at time *t*
x_{15}	Neighbours 2	Second frequent land use class in Moore neighbourhood 7x7 at time *t*
x_{16}- x_{24}	Neighbours 3-11	Frequencies of each class in particular in Moore neighbourhood 7x7at time *t*
x_{25}- x_{33}	Neighbours 12-20	Frequencies of each class in particular in Moore neighbourhood 21x21 at time *t*

Considering that changes in values for some attributes, between two considering years, can provide supplementary useful information for model, additional attributes, represented in Table 4.7, were created. For example attribute x_{35} has been generated to indicate the creation of new street of the category I that happened in period from *t*-1 to *t*. If the new street was not created the value for that attribute is equal to 0.

Table 4.7 Attributes describing changes between two considering years.

Attributes		Description
x_{34}	New inhabitants	New populated cells at time *t* are coded with 1, while cells where number of inhabitants was changed during the period (*t*-1)-(*t*) are coded with 0 (Dummy variable)
x_{35}	Delta str. I category	$\text{Delta st.Icategory} =$ $= \left(\frac{\text{ed. st.Icategory } (t-1) - \text{ed. st.Icategory } (t)}{\text{ed. st.Icategory } (t-1)} \right)$
x_{36}	Delta str. II category	$\text{Delta st.IIcategory} =$ $= \left(\frac{\text{ed. st.IIcategory } (t-1) - \text{ed. st.IIcategory } (t)}{\text{ed. st.IIcategory } (t-1)} \right)$

Therefore, four *.txt files, that represent the state of investigated area for each year of interest, were created by adding 36 attributes for each of 2 263 577 cells. These files are to be used for creating training and test datasets which will be used for further processing and preparation for the purpose of creation and validation of land use change models. Those training and test datasets are to be explained in detail within section 6.

4.4. Assessment of similarity between planned and actual land use maps

Part of this dissertation is the assessment of the extent of realization of Master Plan of Belgrade 2021. A Master Plan represents a long-term concept and spatial organization of settlements. However, adhering to the Master Plan is difficult since urban growth is a complex spatial process and depends on the changing socio-economic conditions, demography, relief, infrastructure and planning constraints. Therefore, in the last hundred years, several urban plans were made and modified.

In 1912 Belgrade got its first Master Plan of Belgrade (Plan varošice Beograd). It was made by French architect Alban Šambon (URBEL). After the World War I, in 1924, Belgrade, as the capital of Kingdom of Serbs, Croats and Slovenes, adopted the new Master Plan made by Djordje Kovaljevski. Then, after the World War II, in what was then Socialist Federal Republic of Yugoslavia (SFRY), Miloš Somborski made Master Plan for Belgrade that was adopted in 1950. Belgrade, as the capital city of SFRY, got its second Master Plan in 1972, made by Aleksandar Djordjević and Milutin Glavički. In 1985 architect Konstantin Kostić conducted Modifications and Supplements to Master Plan of Belgrade 2002. By these modifications, the concept of spatial organization of the city reposes on preservation of certain good parts of the city, with development of new city entireties in agreement with existing values (Vrzić Đ, 2010). Finally, in 2003 Master Plan of Belgrade 2021 was made by Urban Planning Institute of Belgrade.

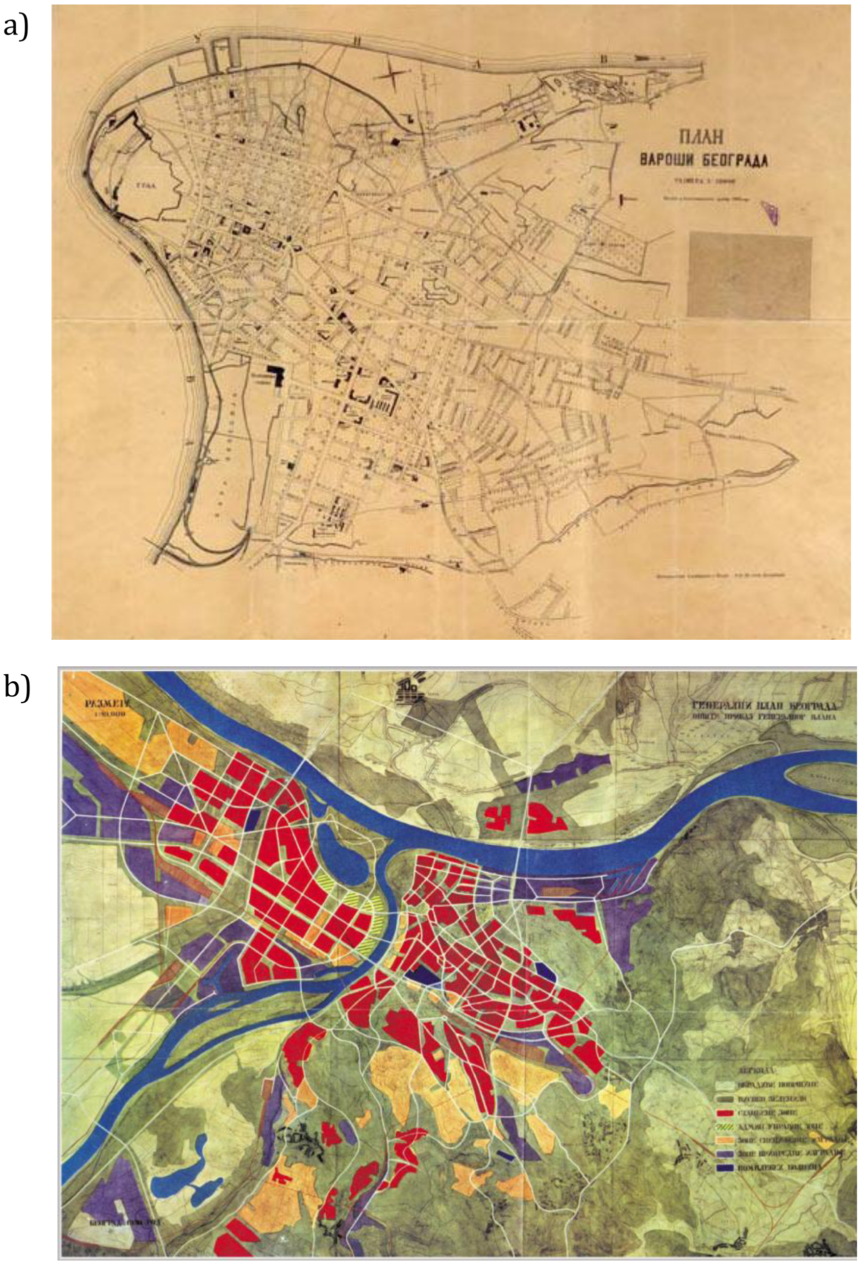


Figure 4.10 a) Master Plan of Belgrade 1915 and b) Master Plan of Belgrade 1950 (http://www.urbel.com/img/ilu/velike/slika-10.jpg).

Assessment of realization of Master Plan of Belgrade 2021 was carried out using all kappa statistics measures (section 3.6). Obtained values can provide adequate analysis of realization degree per class, as shown in following paragraphs.

The data that has been used includes three maps: a map of the Master Plan for 2021 (Figure 4.11a) and maps of actual land use in 2010 (Figure 4.11b) and 2001 (Figure 4.11c).

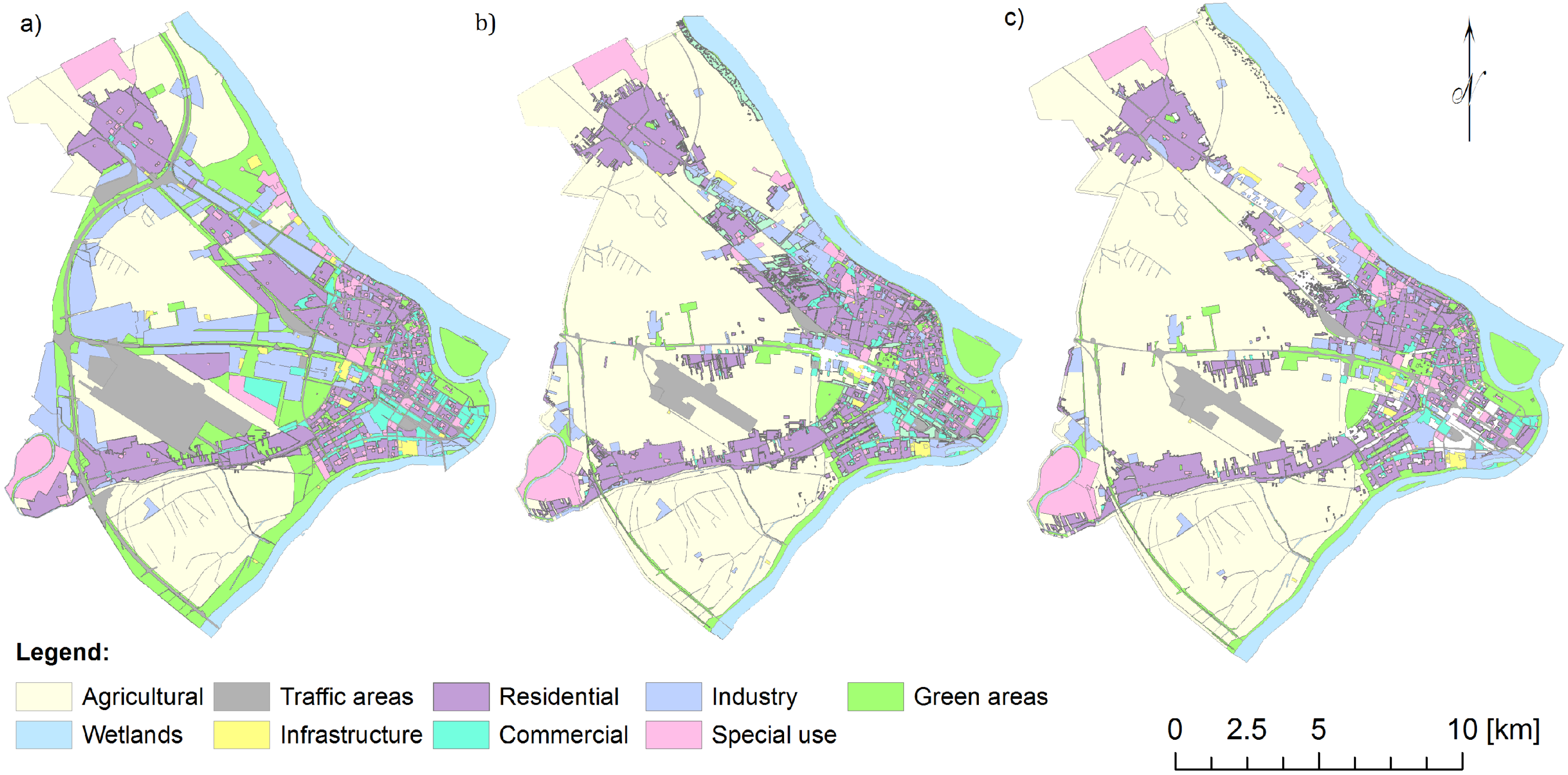


Figure 4.11 a) Map of Master Plan 2021, b) Map of Actual land use 2010 and c) Map of Actual land use 2001.

As explained in section 3.6, in order to obtain a *kappa simulation*, it was necessary to have an initial use map. The Master Plan of Belgrade started to be built using an information database that was available for the Institute of Urbanism of Belgrade and other city offices at the beginning of 2001 (URBEL, 2003). Therefore, the map of actual land use based upon orthophotos taken in 2001 was used as an initial land use map in this experiment.

Assessment of the similarity between planned and actual land use maps, based on the values of previously explained kappa indices (section 3.6), was performed for the total study area as well as for each municipality separately. The obtained results and discussions are presented below.

The Figure 4.12 shows the spatial distribution of agreement, since kappa statistics is based on a straightforward cell-by-cell map comparison.

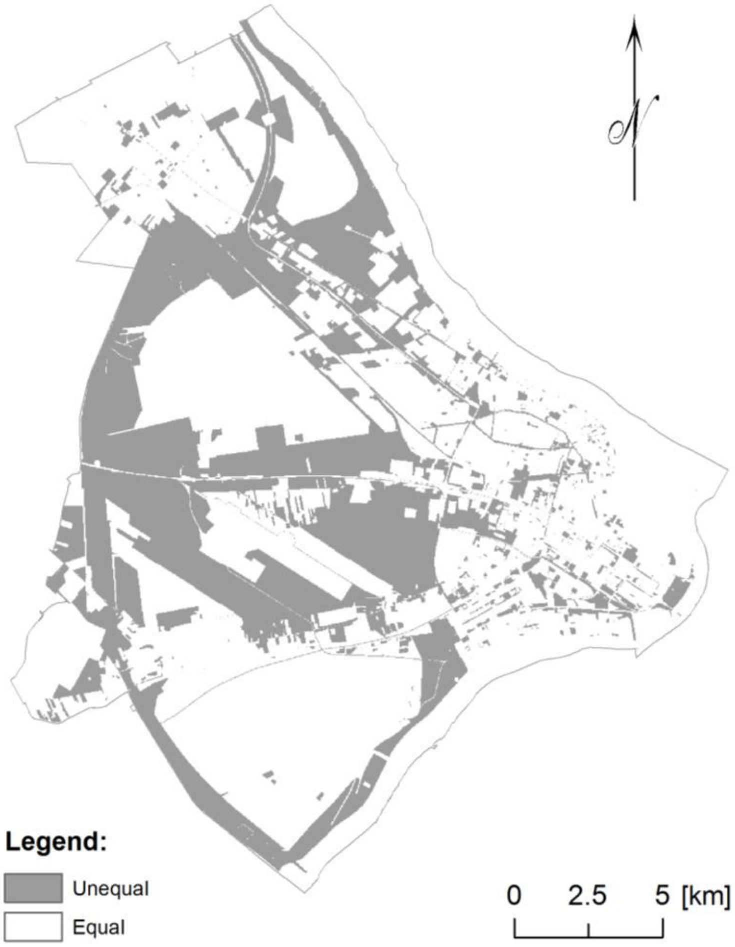


Figure 4.12 Cell by cell comparison.

Table 4.8 Values of *kappa*, *kappa location* and *kappa histo* per municipality.

	Total area	Zemun	New Belgrade	Surčin
Kappa	0.569	0.627	0.596	0.429
$K_{location}$	0.855	0.863	0.800	0.914
K_{histo}	0.662	0.727	0.746	0.469

Based on the results, (Table 4.8) it can be concluded that similarity between the Master Plan and the actual land use map for the year 2010 can be classified in a moderate category based on a standard *kappa* value for the total area. The *kappa location* values indicate that distributions of land use class for these two maps have almost perfect match in location. However, differences between planned and real states (conditions) of land use are more reflected in quantitative dissimilarities. On the other hand, the values per areas differ especially in Surčin. One can conclude that the Surčin municipality has the lowest realization ratio with respect to an urban growth sense.

In order to get better insight in each class behavior, the *kappa* statistics per class were calculated (Table 4.9).

Table 4.9 Values of *kappa*, *kappa location* and *kappa histo* per class for total area.

	Kappa	$K_{location}$	K_{histo}
1. Agriculture	0.474	0.911	0.520
2. Wetland	0.996	1.00	0.996
3. Traffic areas	0.522	0.964	0.542
4. Infrastructure	0.586	0.761	0.769
5. Residential	0.826	0.861	0.959
6. Commercial	0.520	0.708	0.734
7. Industry	0.362	0.777	0.466
8. Special use	0.785	0.855	0.918
9. Green areas	0.361	0.619	0.583

Largest discrepancies between these two maps occur in *Green areas*, *Industrial* and *Agricultural* classes. The green and agriculture areas are classified as unbuilt classes and therefore some of the discrepancies observed in table 4.9 are caused by illegal construction on these two classes. In addition, a large part of area that is used for agricultural purposes in 2010, according to Master Plan, should be transformed into *Green areas*. Astonishingly, only 47% of the total area anticipated by the Master plan for industrial development was used for this purpose by 2010. However, almost perfect location similarities in the class of *Traffic areas* indicate that existing traffic areas are located according to the plan.

Standard *kappa*, *kappa location*, *kappa histo* and *fuzzy kappa* statistics were the main subject of study in a paper published by Samardžić-Petrović et al (2013a). Their publication is a detailed overview regarding results and conclusions that can be made using these statistical values and supports their use in this dissertation.

Since the Master Plan anticipated changes in land use for 35% of the total test area, the values of *kappa*, *kappa location* and *kappa histo* indicate an assessment of similarity for these two maps but do not entirely explain the true extent of realization for land use changes. Furthermore, in order to provide a real assessment of the implementation level of the Master plan up to the year 2010, *kappa simulation* was produced, that refer only to areas subjected to changes with regard to the initial actual use map.

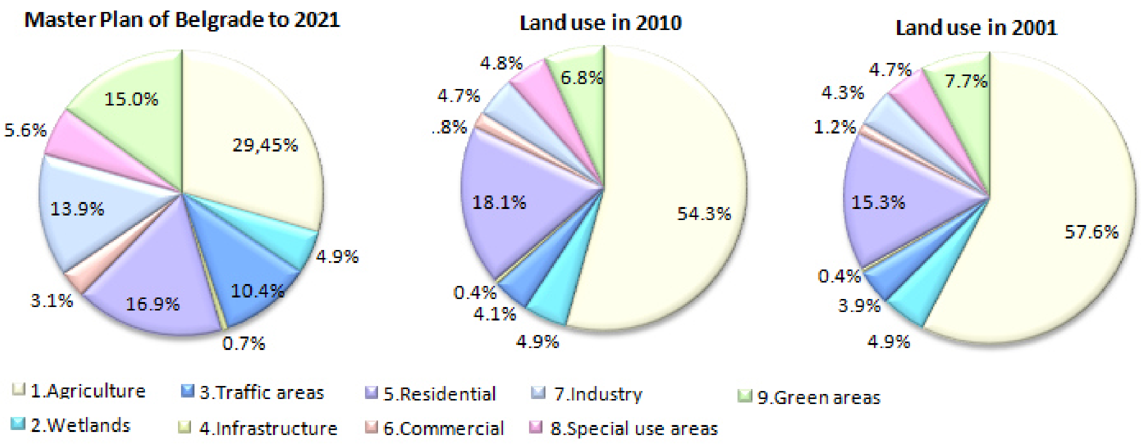


Figure 4.13 Distribution of classes.

As shown in Figure 4.13 the Master Plan anticipated largest changes in land use for agricultural land (50% of total area for that class), whilst the smallest changes are anticipated for water surfaces, which can be caused by bridge construction (less than 1% of total water area) and the rest of the class changes range from 3%-40%. Relatively small changes (a few percent) cannot be considered merely based on standard *kappa*, *kappa location* and *kappa histo* indices values because the small presence in that class is the very reason that they are hidden. That is the additional reason why it is necessary to consider them based on *kappa simulation* values (Table 4.10).

Table 4.10 Values of *kappa simulation*, *kappa transloc* and *kappa transition* per municipality.

	Total area	Zemun	New Belgrade	Surčin
$K_{simulation}$	0.091	0.125	0.151	0.024
$K_{transloc}$	0.529	0.511	0.700	0.513
$K_{transiton}$	0.171	0.245	0.216	0.046

A comparison of the values of standard *kappa* (Table 4.8) with the values of *kappa simulation* (Table 4.10) shows significant differences. One can conclude that urban development has been implemented to a relatively small extent (i.e. planned land use changes are implemented only to 17% of the total area for which they were planned). As *kappa histo* value suggested, Surčin is a municipality that has the slowest rate of planned land use changes (only 5% of changes were implemented). The *kappa transloc* values indicate that the changes are not conducted entirely as planned. New Belgrade is the municipality with lowest discrepancies regarding locations of newly built objects and their use plan.

Looking at the *kappa simulation* values per classes (Table 4.11) with respect to total area, it can be seen that there are large dissimilarities in all classes. These dissimilarities are mainly caused by the unrealized urban development plan.

Table 4.11 Values of *kappa simulation, kappa transloc and kappa transition* per class above the total area.

	$K_{simulation}$	$K_{transloc}$	$K_{transition}$
1. Agriculture	0.057	0.488	0.116
2. Wetland	0.098	1.00	0.098
3. Traffic areas	0.055	0.859	0.065
4. Infrastructure	0.000	n.a.	0.000
5. Residential	0.300	0.468	0.641
6. Commercial	0.212	0.521	0.408
7. Industry	0.066	0.594	0.112
8. Special use	0.038	0.309	0.122
9. Green areas	0.076	0.693	0.109

The construction of residential and commercial structures dominated in all development projects during the period from 2001 to 2010. Kappa values for the *Infrastructure* class indicate that none of the planned infrastructure objects have been built above the total area. Based on kappa simulation values for agricultural and green areas, one can conclude that illegal construction is still underway in these areas as values in Tables 4.8 and 4.9 point out. Planned construction of traffic areas is realized by only 6.5% with minor location changes. Indices values for the *Water area* class indicate that planned construction for three bridges is realized by 9.8% with no deviation from the initially planned locations.

The detailed overview of the obtained results is published in paper "The application of different kappa statistics indices in assessment of similarity between planned an actual land use maps" (Samardzic-Petrović et. al, 2013b). Some of those differences between Master Plan of Belgrade and actual land use map for 2010 are presented in Figure 4.14: a) Planed for *Infrastructure* - used as *Commercial*, *Residential* and *Agriculture*, b) Planed for *Industry* - used as *Commercial*, c) Planed for *Green areas* - used as *Agriculture* and *Residential*, d) Planed for *Commercial* –

used as *Agriculture*, *Commercial* and *Residential*, e) Planed for *Green areas* - used as *Agriculture* (not built), f) Planed for *Commercial* and *Special - used* as *Green areas* (not built).

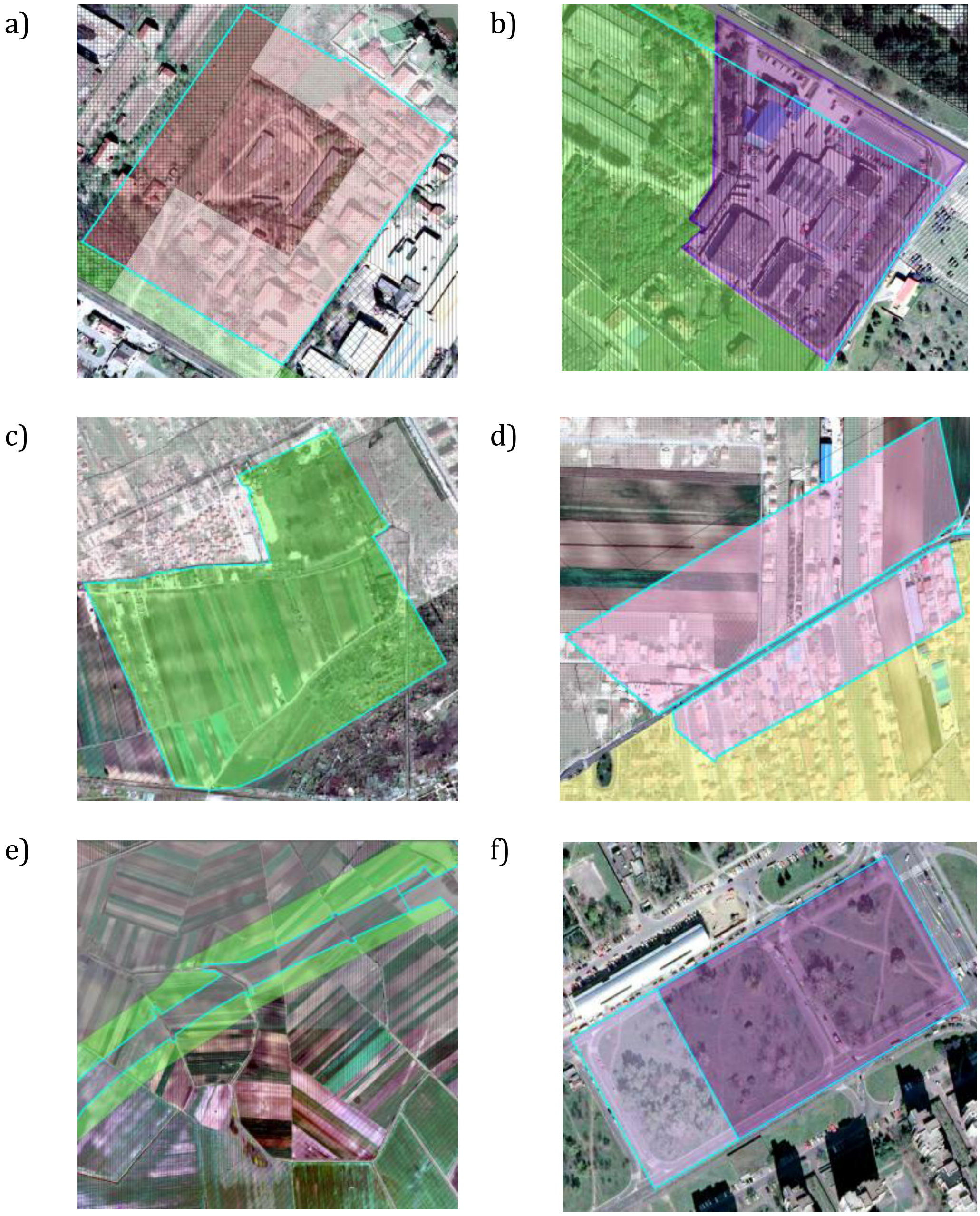


Figure 4.14 Some of the differences between Master Plan of Belgrade and actual land use map for 2010.

There is some degree of similarity in land use considering overlapping of the classes. Classes of *Green* and *Agriculture* areas are similar since they both belong to the unbuilt class whereas *Residential* and *Commercial* classes are often combined facilities. The similarity between adjacent classes was realized with the following Category Similarity Matrix (Table 4.12):

Table 4.12 Category similarity matrix.

Class	1	2	3	4	5	6.	7	8	9
1	1	0	0	0	0	0	0	0	0.6
2	0	1	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0
4.	0	0	0	1	0	0	0	0	0
5	0	0	0	0	1	0.4	0	0	0
6	0	0	0	0	0.4	1	0.2	0	0
7	0	0	0	0	0	0.2	1	0	0
8	0	0	0	0	0	0	0	1	0
9	0.6	0	0	0	0	0	0	0	1

Therefore, besides the previous kappa indexes calculation, the fuzzy set map comparison was also performed (Figure 4.15). In Figure 4.15 each cell has a value between 1 (for identical cells) and 0 (for total disagreement). The darker areas indicate more intensive disagreement. Unlike Figure 4.12 it is possible to obtain a gradual analysis of the similarity of two maps by distinguishing total agreement (white areas), medium similarity and low similarity (the shades of gray) and total disagreement (black areas).

Alongside the qualitative assessment of similarities between categorical (class) maps, previously calculated kappa indices provide valuable information regarding spatial assessment. Based on standard *kappa*, *kappa location* and *kappa histo* values, one can achieve similarity assessment regarding overall land use classes.

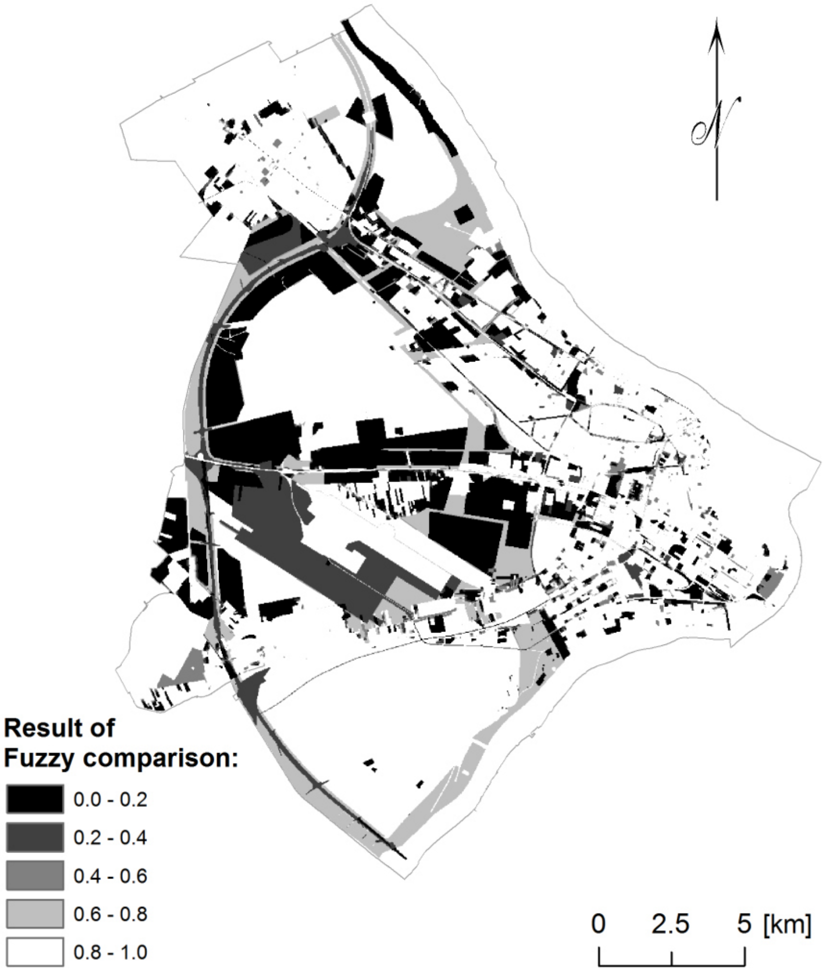


Figure 4.15 Spatial assessment of similarity in the fuzzy set approach.

However, it is necessary to conduct separate analysis for emerged land use changes since urban development is a space-temporal process. Obtained results indicate that *kappa simulation*, *kappa transition* and *kappa transloc* statistical analysis can provide information regarding similarity assessment of emerged land use change maps. Therefore, analysis of the realization of a Master plan can be carried out based on the assessment of similarities between the Master plan and actual land use along with implementation of kappa statistics indices.

Chapter 5:

Results and discussion

The results of conducted experiments are presented in two parts. In the first part, proposed methodology (presented in chapter 3) and performance evaluation of the data-driven methods with three different ML techniques, (Decision Trees, Neural Networks and Support Vector Machines) through four experiments, is presented.

In the second part, the sensitivity of a predictive model with regards to SVM parameters changes were examined on different data representations and on the same number of attributes selected by Info Gain, Gain Ratio and Correlation-based Feature Subset. Additionally, the capability to find appropriate optimal SVM parameters using only data from the past, in order to predict future land use changes was tested.

In the research reported herein, a total of nine data representations were used and around 1000 models were built. However, the comparison of the proposed DD methods has been presented by using 4 basic data representations accompanied by more than 150 models. In the second part of the research, more than 120 models were built using the additional three basic data representations. The time required to build each model was dependent on many different factors including: the used ML techniques, the selected parameters of a particular technique, the number of used attributes in each dataset, the number of selected cells in a study area and the processing power. Therefore, the time necessary to build each model ranged from a few seconds to one day.

The Weka software (Hall et al., 2009) was used to build the models using DT (Weka J48 implementation of the C4.5 algorithm), NN (Multilayer Perceptron implementation) and SVM (SMO - Sequential Minimal Optimization algorithm). The Map Comparison Kit (MCK) (Visser and de Nijs, 2006) software was applied for validation of generated predictions and ArcGIS for presentation of the results.

5.1 Performance evaluation of data-driven methods for modelling land use change

The proposed methodology and performance evaluation of DD methods with different ML techniques (DT, NN and SVM), different attribute ranking methods and different data representations was accomplished within four experiments:

- The first experiment focused on the appropriateness of the proposed data sampling procedure for land use change modelling presented in section 3.4.
- The second experiment compared land use model outcomes obtained from the four different dataset representations for each of the DT, NN and SVM techniques.
- The third experiment examined different attribute ranking methods χ^2, Info Gain (IG) and Gain Ratio (GR) presented in section 3.5
- The fourth experiment used information about attribute rankings to select those attributes that contribute to the best performance of the predictive land use model.

Before the comparison of DT, NN and SVM models, a set of parameters was found for each best performing model. Changing the values of parameters in the case of DT and NN models did not influence the results significantly, hence default parameters were used for these two techniques: for DT – a confidence factor used for pruning was set at 0.25 and a minimum number of instances per leaf was 2; and for NN – number of neurons in a hidden layer was selected to be (number of land use classes + number of attributes)/2. For SVM, the parameters were investigated separately for each data representation. Since the SVM parameters exhibited a significant influence on the outcome of the model, it was necessary to investigate them thoroughly, and hence the second part of the dissertation emerged. In the experiments from the first part, default parameters for $C(1)$ and $\gamma(0.01)$ (RBF kernel) were used.

5.1.1 Datasets creation for Training and Testing

In order to build and test a model, it is necessary to create two independent datasets (explained in section 3.2): the *training* dataset of the form ($\mathbf{x}^{t-1}$, y^t) is devised in order to learn the predictive function $f\,'\mathbf{x}^{t-1} \rightarrow y^t$ and the *test* dataset ($\mathbf{x}^t$, y^{t+1}) is devised in order to test the predictive function.

Four different data representations were generated for training and testing purposes using datasets in which each grid cell of the study area was represented as a vector of attributes $\mathbf{x}^t$ (presented in section 4.3):

- The first dataset is called the *simple dataset representation* S since the data contains information about accessibility and population.
- The second dataset contains attributes from the *simple dataset representation* and includes *explicit* information about land use classes in the cell's neighbourhood and is referred to as *dataset representation with neighbourhood* S^{n}.
- The third dataset contains attributes from the *dataset representation with neighbourhood* and includes information about the cell's previous land use class (history t-2-in training set and t-1 in test set) and is referred to as *dataset representation with neighbourhood & history* S^{nh}.
- The forth dataset is the *dataset representation with neighbourhood, history & changes* S^{nhc} and contains attributes from the *simple dataset representation* and includes information regarding the changes of spatial attributes that occurred in the past, information about land use classes in the cell's neighbourhood and cell's previous land use class.

Based on the various data representations presented, eight basic training and testing datasets were created: $S_{3\text{-}7}$ ($\mathbf{x}^{2003}$, y^{2007}) and $S_{7\text{-}10}$ ($\mathbf{x}^{2007}$, y^{2010}); $S^{n}_{3\text{-}7}$ ($\mathbf{x}^{2003}$, y^{2007}) and $S^{n}_{7\text{-}10}$ ($\mathbf{x}^{2007}$, y^{2010}); $S^{nh}_{3\text{-}7}$ ($\mathbf{x}^{2001,\ 2003}$, y^{2007}) and $S^{nh}_{7\text{-}10}$ ($\mathbf{x}^{2003,\ 2007}$, y^{2010}); $S^{nhc}_{3\text{-}7}$ ($\mathbf{x}^{2001,\ 2003}$, y^{2007}) and $S^{nhc}_{7\text{-}10}$ ($\mathbf{x}^{2003,\ 2007}$, y^{2010}). The attributes used to represent cells in datasets are given in Table 5.1.

Table 5.1 Attributes used for different data representation S, S^{n}, S^{nh} and S^{nhc}.

<table>
<tr><th rowspan="2">Attributes</th><th colspan="4">Data representation</th></tr>
<tr><th>S</th><th>S^{n}</th><th>S^{nh}</th><th>S^{nhc}</th></tr>
<tr><td>x_1</td><td colspan="4">Municipalities (Zemun, New Belgrade and Surčin)</td></tr>
<tr><td>x_2</td><td colspan="4">Euclidean distance of grid cell to city centre</td></tr>
<tr><td>x_3</td><td colspan="4">Euclidean distance of grid cell to municipality centre</td></tr>
<tr><td>x_4</td><td colspan="4">Euclidean distance of grid cell to the closest the rivers</td></tr>
<tr><td>x_5</td><td colspan="4">Euclidean distance of grid cell to the closest big green areas</td></tr>
<tr><td>x_6</td><td colspan="4">Euclidean distance of grid cell to the closest railway lines at time <i>t</i></td></tr>
<tr><td>x_7</td><td colspan="4">Euclidean distance of grid cell to the closest highway at time <i>t</i></td></tr>
<tr><td>x_8</td><td colspan="4">Euclidean distance of grid cell to the closest main road at time <i>t</i></td></tr>
<tr><td>x_9</td><td colspan="3">Euclidean distance of grid cell to the closest street of category I at time <i>t</i></td><td>Delta street of category I</td></tr>
<tr><td>x_{10}</td><td colspan="3">Euclidean distance of grid cell to the closest street of category II at time <i>t</i></td><td>Delta street of category II</td></tr>
<tr><td>x_{11}</td><td colspan="3">Number of inhabitants at time <i>t</i></td><td>New inhabitants</td></tr>
<tr><td>x_{12}</td><td colspan="4">Land use class at time <i>t</i></td></tr>
<tr><td>x_{13}</td><td></td><td colspan="3">Most frequent land use class in Moore neighbourhood 7x7 at time <i>t</i></td></tr>
<tr><td>x_{14}</td><td></td><td colspan="3">Second frequent land use class in Moore neighbourhood 7x7 at time <i>t</i></td></tr>
<tr><td>x_{15}</td><td></td><td></td><td colspan="2">Land use class at time <i>t</i>-1</td></tr>
<tr><td>y</td><td colspan="4">Land use class at time <i>t</i>+1</td></tr>
</table>

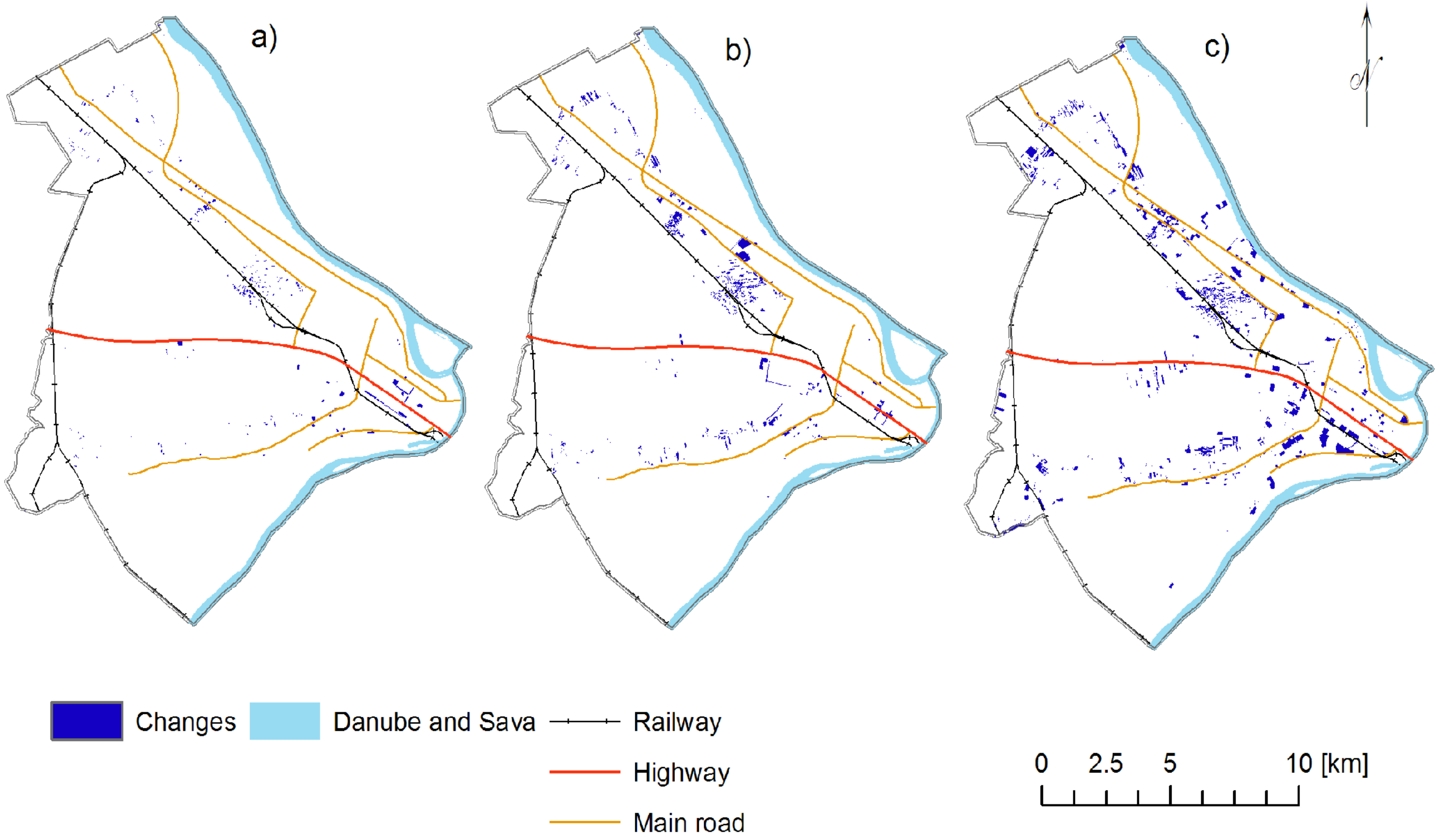


Figure 5.1 Changes in land use a) from 2001 to 2003, b) from 2003 to 2007, c) from 2007 to 2010.

5.1.2 Experiment 1: Proposed sampling data

After the analysis was completed, it was found that only 4% (93 073 cells out of 2 263 577cells) of the study area has been changed during the period of nine years (2001 – 2010) (Figure 5.1). In the period 2001 – 2003, changes have occurred in 0.40% of cells (9 134 cells), while greater changes in 1.13% (25 672 cells) and 2.65% (59 984 cells) of the total amount of cells located within the boundaries of the study area were noted in the periods encompassing 2003 – 2007 and 2007-2010, respectively (Figure 5.1).

Since a small amount of land use change can have negative consequences on model building and validation, it is important to create appropriate training and test datasets. Therefore, the method for data sampling proposed and presented in section 3.4 is tested in this experiment. For that purpose four datasets were created and used: U_{3-7}, U_{7-10}, B_{3-7}, and B_{7-10}. These datasets were sampled from the smaller and representative part of the study area. The cells in all four datasets were represented according to the *simple dataset representation* S, i.e. datasets used in this experiment contain same attributes as datasets S_{3-7}, S_{7-10} (Table 5.1). The symbol U denotes an unbalanced dataset (included all cells from the selected smaller area) while B denotes a balanced dataset (included the same amount of changed and unchanged cells). Sets B_{3-7} and B_{7-11} were created by including all cells that changed their land use state from a previous time epoch and an equal number of cells that did not change their state. Unchanged cells were pseudo randomly selected assuming a uniform distribution over the entire sampling area and preserving the original class proportions.

Models were built using all three ML techniques and the obtained results are presented in Table 5.2.

Table 5.2 *Kappa* values for balanced B and unbalanced U training and test datasets.

Datasets used for modeling	*Kappa*		
	DT	NN	SVM
Trained on U_{3-7} and tested on U_{7-10}	0.73	0.82	0.85
Trained on U_{3-7} and tested on B_{7-10}	0.23	0.31	0.35
Trained on B_{3-7} and tested on B_{7-10}	0.52	0.57	0.58

The obtained results for traditional *kappa* were very high when the model was built with dataset U_{3-7} containing all cells and then tested on dataset U_{7-10}. This result reflected the nature of both datasets in which vast amount of cells remained unchanged and therefore induced high *kappa* values. When the test dataset was changed to B_{7-11} with balanced amounts of changed/unchanged cells, the results for *kappa* decreased significantly. The obtained results were expected since the model was trained mainly on unchanged cells and was biased towards predicting unchangeable cells. The *kappa* values increased when model was trained on B_{3-7} and tested on B_{7-11}.

This experiment indicates that the balanced sampling strategy for ML techniques provides better model outcomes. The results obtained are in accordance with Santé et al. (2010) who indicated, while reviewing literature using *kappa* statistics, that *kappa* values have been inflated and are dependent on the proportion of land use changes.

Since it has been shown that the proposed sampling method provides better model outcomes and more realistic validations, similar data sampling was performed herein and was carried out for all datasets that were used for further experiments.

5.1.3 Experiment 2: Comparison of models built using DT, NN and SVM based on different data representations

This experiment aimed to compare LUC models that were built using all three ML techniques and four proposed data representations S, S^n, S^{nh} and S^{nhc}. *Simple* data representation was evaluated using $S_{3\text{-}7}$ as training and $S_{7\text{-}10}$ as a test dataset for all three ML techniques. Similarly, to evaluate the data representation with *neighbourhood* datasets $S^n{}_{3\text{-}7}$ and $S^n{}_{7\text{-}10}$ were used. Data representation with *neighbourhood & history* was evaluated using datasets $S^{nh}{}_{3\text{-}7}$ and $S^{nh}{}_{7\text{-}10}$. Data representation with *neighbourhood, history & changes* was evaluated using $S^{nhc}{}_{3\text{-}7}$ as a training and $S^{nhc}{}_{7\text{-}11}$ as a test set. The values obtained for *kappa* and *kappa simulation* for the second experiment are presented in Figure 5.2 for each learning technique and data representation.

The results of this experiment indicate that all three ML techniques are capable to predict land use changes. The bar chart (Figure. 5.2) indicates that the NN and SVM achieved better predictive models for the study area than DT. However, unlike DT, a drawback of the other two techniques (NN and SVM) is that they do not provide explanations of how the results were derived in order for a domain expert to interpret them.

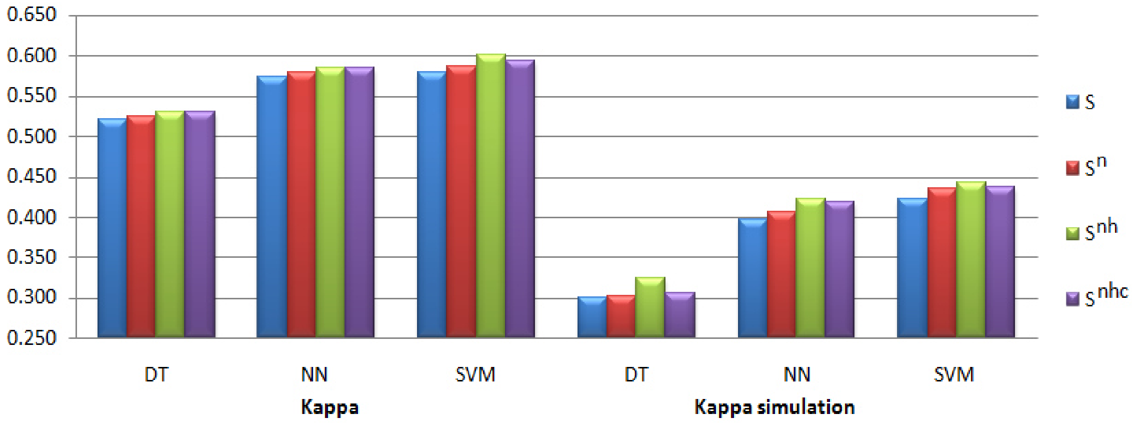


Figure 5.2 Comparison of the validation measures of models for four data representations S, S^n, S^{nh}, and S^{nhc} for all three machine learning techniques.

Furthermore, DT is more time efficient and creates models in seconds, while NN and SVM require lengthy processing times on the order of minutes to hours, respectively.

The representation that includes information from the neighbourhood and history showed slightly better performance for all three ML techniques. This result can lead to the assumption that information from the neighbourhood and history are important for prediction of land use, which will be tested in the next experiment.

5.1.4 Experiment 3: Ranking of attributes according to their significance for models

This experiment was conducted to estimate the importance of each considered attribute on the modelling process by using three different ranking methods (χ^2, IG and GR) and training datasets for the proposed data representations. The obtained attribute rankings using all three ranking method χ^2, IG and GR are presented for all four data representations in Figures 5.3 and 5.4.

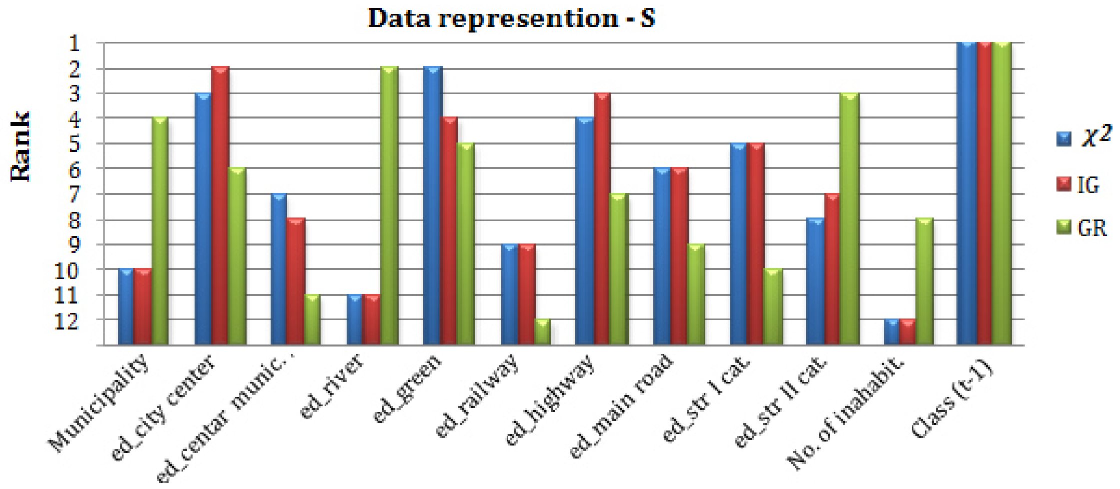


Figure 5.3 Attribute ranking values based on χ^2, IG and GR method for data representation S.

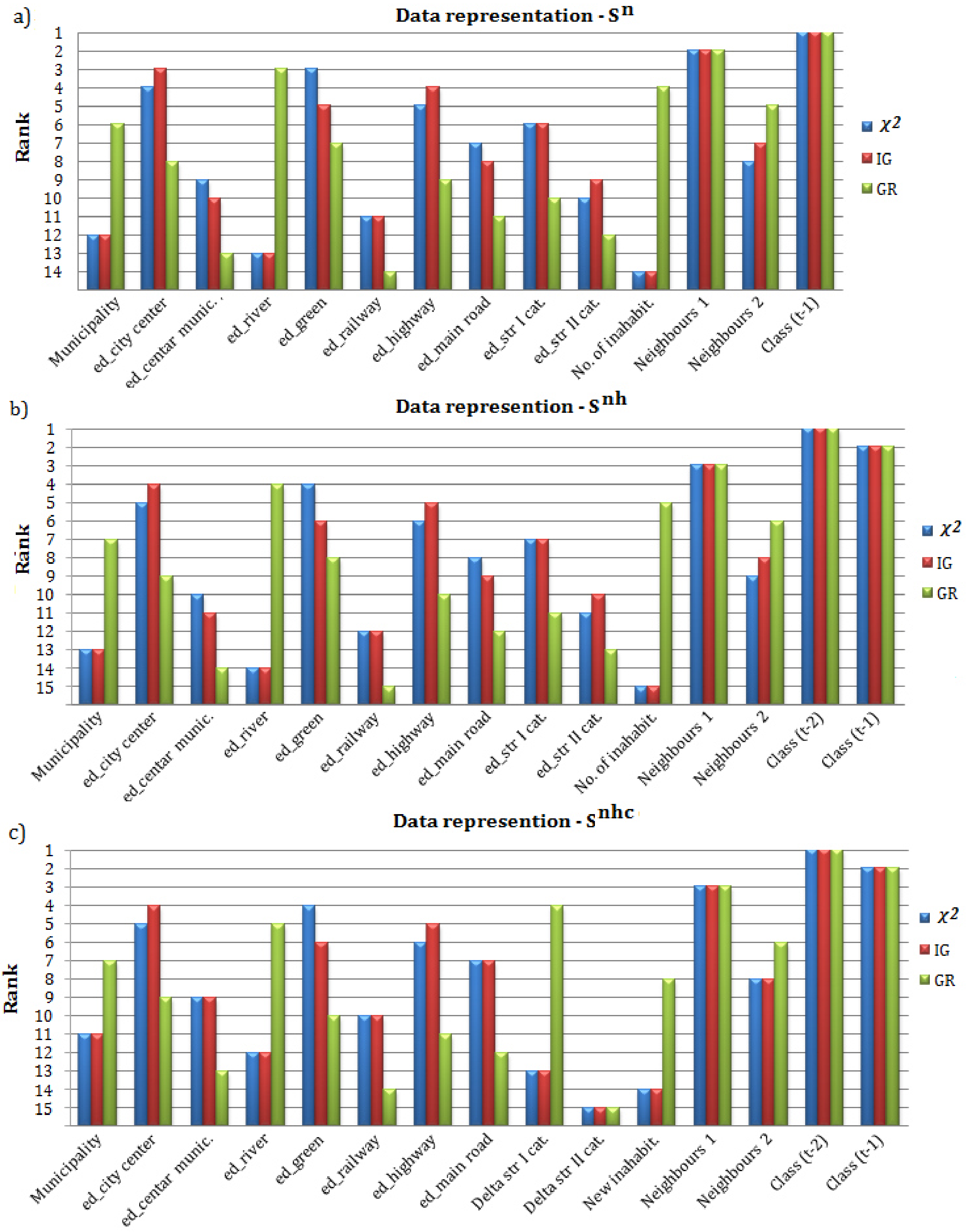


Figure 5.4 Attribute ranking values based on χ^2, IG and GR method for data representation a) S^n, b) S^{nh} and c) S^{nhc} .

Since the three aforementioned methods rank attributes independently of each other according to their measure of association with the land use class in time t and that data representations S^n and S^{nh} present expanded representation S, the results are different only for additional attributes (neighbourhood attributes for S^n and neighbourhood and history attributes for S^{nh}) (Figure 5.4 a, b). The results for the S^{nhc} differ slightly more due to changed attributes (Delta street of category I and II and New inhabitants attributes) (Figure 5.4 c).

All attribute ranking methods "come to the same point" that the information on previous class of land use is the most informative for the prediction of LUC. For both *simple* and *neighbourhood* dataset representations, the highest ranked attribute was the land use class in time *t*-1, while the previous land use class (land use class in time *t*-2) was the most informative attribute for the dataset representations with *neighbourhood* & *history* and *neighbourhood, history & changes*. Land use class in time *t*-1 was the second ranked attribute for those two datasets. Therefore, additional information about land use from the past is very important for the prediction.

The next ranked attribute for both datasets, S^{nh} and S^{nhc}, is the most frequent land use class in the neighbourhood. For that reason, additional information about land use classes in the cell's neighbourhood also play very important role.

The results suggest that both χ^2 and IG rank relevant attributes almost identically while GR indicates some differences. Hence, only outcomes of IG and GR will be discussed and used in the next experiment. It is well known that IG ranks higher attributes with wider ranges of values when compared to GR. Hence, the municipality attribute (only four distinct values) is ranked much lower than in the GR case and, from an expert's point of view, it would be a very important attribute (the municipalities from the study area are very different in the sense of urban growth).

Based on the IG, other important attributes for all data representations were related to the proximity of the city centre and the distance from highways. The lowest ranked attributes for the first three proposed dataset representations were related to distances from rivers and the number of inhabitants per cell. Based on the IG results, additional attributes in the fourth data representation that describe changes are very low ranked, while the GR ranks the change of distance to the first category streets very high.

5.1.5 Experiment 4: Finding a set of attributes that best describe the process of land use change

This experiment was conducted in order to find how attribute reduction affects the modelling outcomes for each ML technique. Ranking methods described in the previous experiment estimated the importance of each attribute independently of all other attributes. Regarding the correlation between attributes, it is important to estimate how they behave together in a real prediction process. Therefore, a recursive attribute elimination method is performed (removing the lowest-ranked one, and repeating the process until all attributes have been removed). Finding a set of attributes that best describe the process of land use change using the recursive attribute elimination method was conducted for attributes from data representation S^{nh} ranked based on the IG and attributes from data representation S^{nhc} ranked based on IG and GR rank methods. The obtained values for all *kappa* measures and all three ML techniques are presented in Figures 5.5 and 5.6.

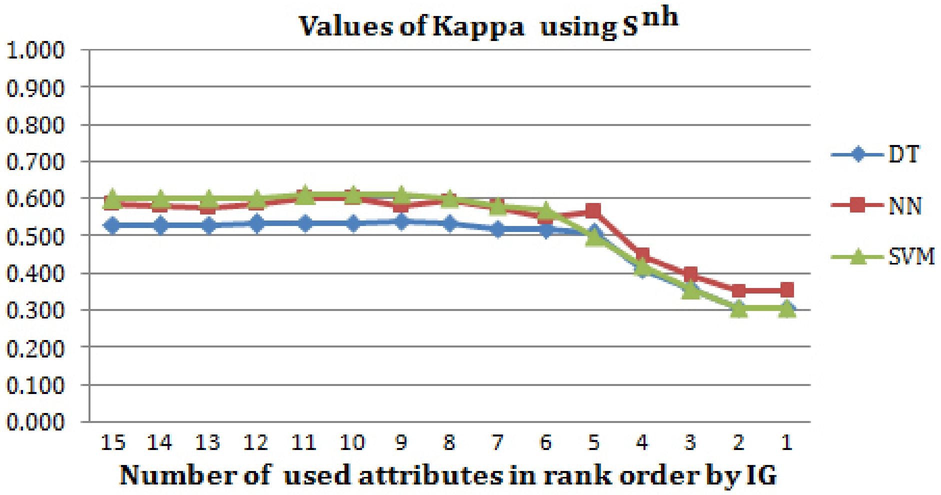


Figure 5.5 Obtained values of *kappa* in a recursive attribute elimination process using attributes from S^{nh} ranked based on IG.

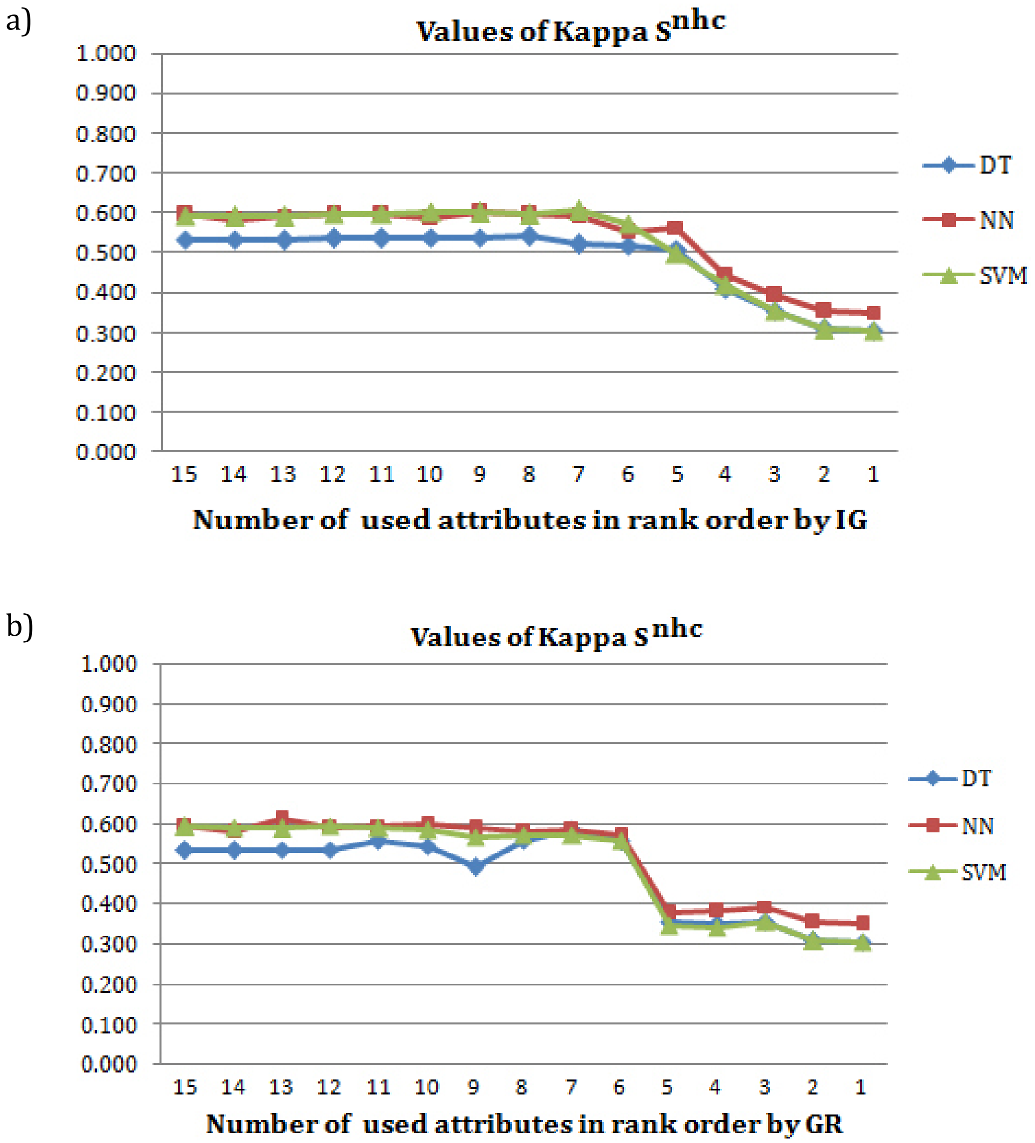


Figure 5.6 Obtained values of *kappa* in a recursive attribute elimination process using attributes from S^{nhc} ranked based on a) IG and b) GR.

The recursive attribute elimination method showed that the first 5+ ranked attributes by IG and first 6+ ranked attributes by GR produced satisfactory models for prediction of the land use change using all three ML techniques. In addition, *kappa* values indicate that the elimination of certain attributes improved the model outcomes.

The values of *kappa* indicate (Figure 5.5 and 5.6) that the DT method is 5% less efficient when compared to the two other used methods. By examining the results obtained for the GR, *kappa* values declined sharply for all three techniques when compared against the number of attributes used (for S^{nhc} with 6 highest ranked attributes). This is different to IG, whereby values also decrease significantly but more gradually when using 6 highest ranking attributes or less.

The first 6 IG ranked attributes in S^{nh} and S^{nhc} data representations are the same. Therefore, the results obtained for all *kappa* measurements for models built using 6 or less highest ranking attributes from S^{nh} and S^{nhc} are the same. On the other hand, models constructed with 7 or more ranked attributes provide slightly different values of used measures. For this reason, recursive attribute elimination method by GR was carried out only for one data representation (S^{nhc}).

The highest values of *kappa*, obtained after the recursive attribute elimination on S^{nh} ranked by IG and S^{nhc} ranked by IG and GR are presented in Table 5. 3.

Table 5.3 Highest *kappa* values obtained after the recursive attribute elimination on S^{nh} ranked by IG and on S^{nhc} ranked by IG and GR.

Used data representation	Used ML technique	*Kappa*	Used rank method	Number of used attributes
S^{nh}	DT	0.539	IG	9
	NN	0.607	IG	11
	SVM	0.612	IG	9
S^{nhc}	DT	0.541	IG	8
		0.559	GR	11
	NN	0.599	IG	9
		0.611	GR	13
	SVM	0.605	IG	7
		0.594	GR	12

The use of the first 9 attributes ranked by IG in data representation S^{nh} provides the highest obtained values of *kappa* for models built by using DT and SVM.

However, SVM is more capable for predicting future land use classes when compared to DT.

The results given in Table 5.3 indicate that the differences between the highest obtained values of *kappa* using the first ranked attributes by IG or GR are neglected for the same data representation (S^{nhc}). However, the number of used attributes ranked by IG is smaller. Having that in mind, it can be concluded that IG presents a more appropriate attribute ranking method than GR considering that the increase in the number of attributes implies the increase of model complexity.

The calculated *kappa* values obtained using different data representations (Table 5.3) S^{nh} and S^{nhc} and the same attribute ranking method IG are similar, while the number of used attributes is slightly different. The highest values of *kappa* are obtained by using one less attribute from S^{nhc} then from S^{nh} in modeling land use change by DT and by using two less attributes from S^{nhc} then from S^{nh} in modeling by NN and SVM. However, considering the complexity regarding the generation of some S^{nhc} attributes in this study area, and the fact that they do not contribute to a significant improvement of the model performance, the data representation S^{nh} was used for the further analysis.

For comprehensive result analysis, it is necessary to observe the measure in which the models predict the quantity of changes as well as to confirm if those changes are located where they should be. For that purpose *kappa location* and *kappa histo* were used. The obtained values of *kappa histo* and *kappa location* are presented in Figure 5.7.

The results presented in Figure 5.7 indicate that all three ML techniques are equally capable for modeling location and quantity of future land use classes. The obtained *kappa* and *kappa histo* values follow similar trends for DT, NN and SVM. Based on the analysed *kappa location*, the DT method performs slightly better than NN and SVM. The values of *kappa histo* decrease sharply when using three and less first-ranked attributes by IG.

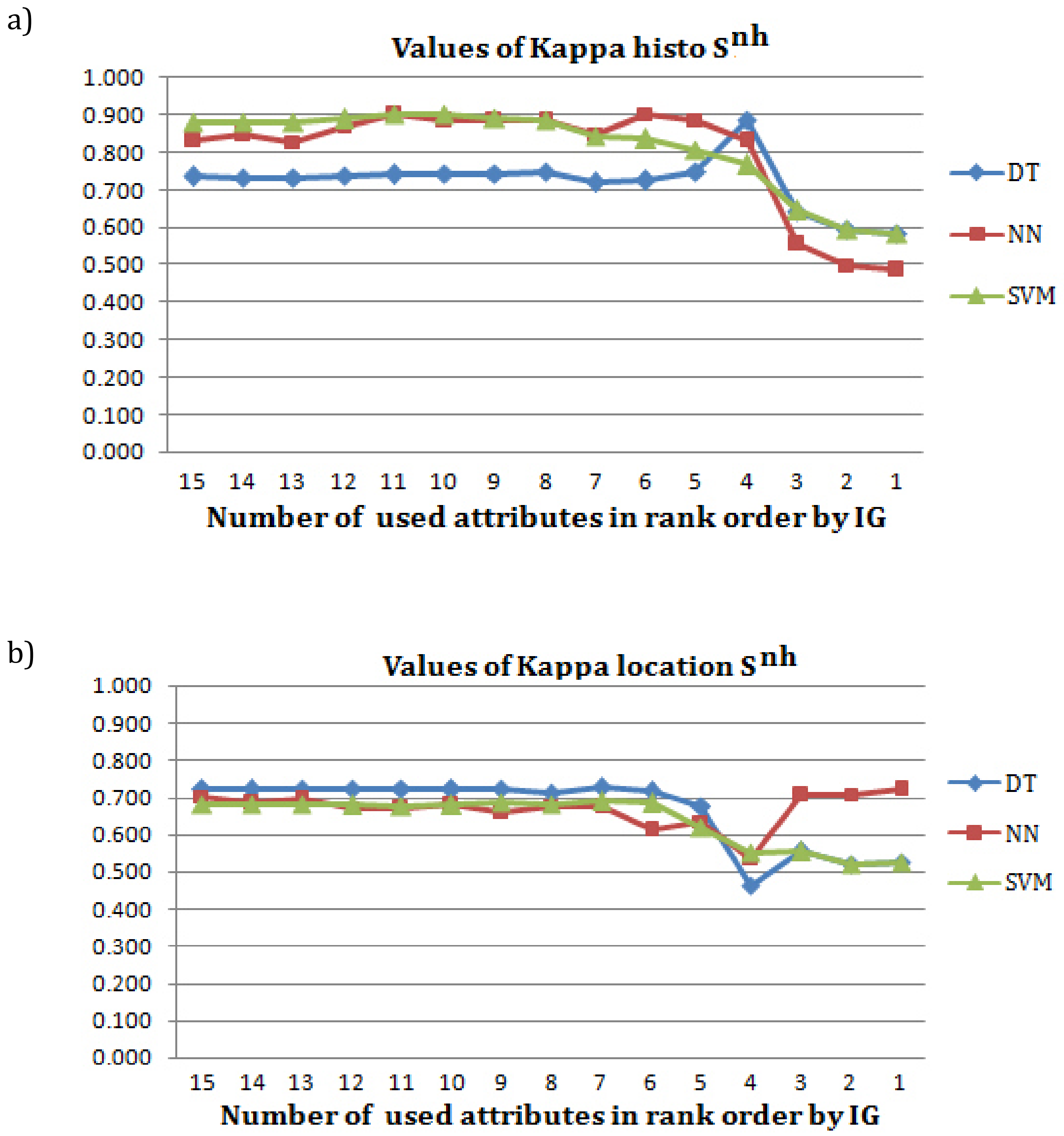


Figure 5.7 Obtained values of a) *kappa histo* and b) *kappa location* in a recursive attribute elimination process using attributes from S^{nh} ranked by IG.

In order to examine the extent to which the applied models predict changes, *kappa simulation* and its corresponding variations were used. The obtained results are presented in Figure 5.8.

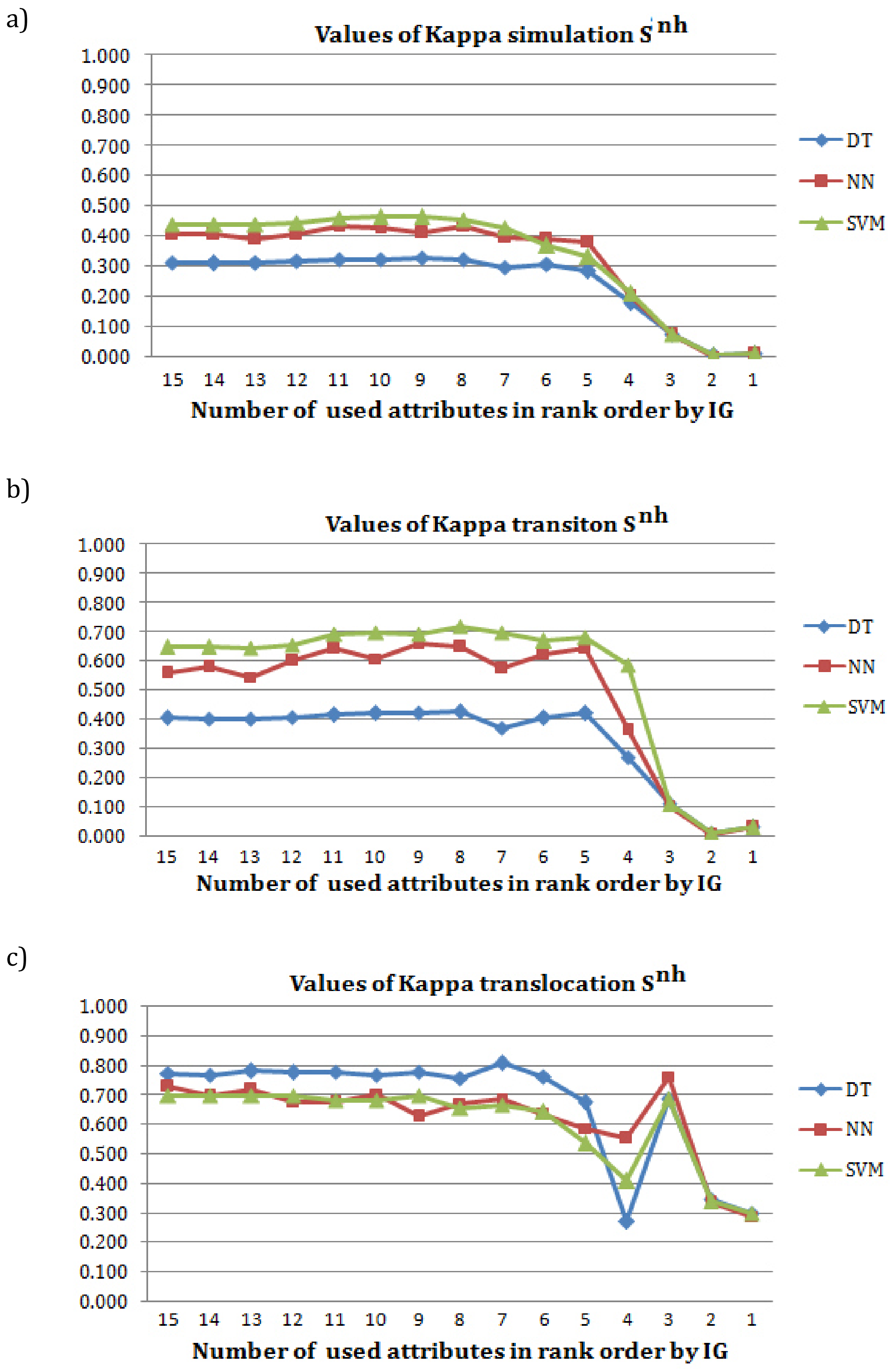


Figure 5.8 Obtained values of a) *kappa simulation* b) *kappa transition* and c) *kappa transloc* in a recursive attribute elimination process using attributes from S^{nh} ranked by IG.

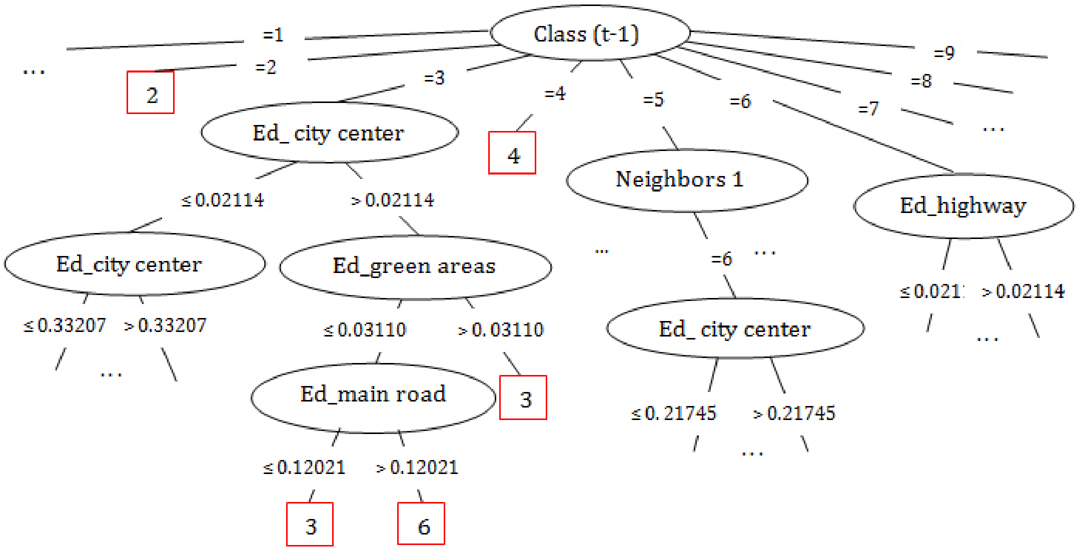


Figure 5.9 Part of the built decision tree for S^{nh} data representation.

The obtained values for *kappa simulation* indicated trends that are similar to the ones obtained for *kappa* and indicated that SVM and NN performed nearly 10% better than DT.

In general, all previous results indicate that while all three ML techniques can be used for modelling urban land use change, the accuracy can be improved by using appropriate balanced sampling schemes and relevant attribute selections.

Results indicate that SVM and NN have a slightly better capability to model changes than DT. Considering that the prediction of the correct location where the changes occurred is more important for the land use change modelling than overall quantity of changes, the SVM technique is more appropriate than NN. However, the DT technique helps reveal the explanation as to how the results were derived. This enables an expert to interpret the model because it is possible to visualize the tree from which the decisions have been made (Figure 5.9). This is not possible with SVM and NN and therefore these two techniques remain a black box and are closed for the user to understand.

Finally, based on the evaluation experiments, the modelling outcomes are presented for the highest *kappa/kappa simulation* values per each class in Table 5.4 and Figures 5.10, 5.11 and 5.12.

Table 5.4 Performance measures of selected models built by DT, NN and SVM for all classes based on *kappa, кappa location, kappa histo, kappa simulation*, *kappa transition* and *kappa transloc*.

Measure	ML techniq.	Land use class								
		Agricultural	Wetlands	Traffic areas	Infrastructure	Residential	Commercial	Industry	Special use	Green area
Kappa	DT	0.656	0.986	0.705	1.000	0.566	0.150	0.397	0.703	0.310
	NN	0.732	0.959	0.596	0.966	0.591	0.362	0.483	0.697	0.428
	SVM	0.736	0.985	0.551	1.000	0.612	0.311	0.563	0.694	0.395
$K_{location}$	DT	0.903	0.988	0.785	1.000	0.683	0.540	0.441	0.909	0.499
	NN	0.875	0.971	0.624	0.985	0.610	0.481	0.499	0.901	0.446
	SVM	0.820	0.987	0.629	1.000	0.638	0.593	0.599	0.880	0.412
K_{histo}	DT	0.726	0.998	0.898	1.000	0.829	0.278	0.898	0.773	0.620
	NN	0.836	0.987	0.955	0.981	0.969	0.752	0.969	0.774	0.961
	SVM	0.898	0.998	0.877	1.000	0.959	0.524	0.941	0.788	0.959
$K_{simulati}$	DT	0.471	0.000	0.077	1.000	0.456	0.034	0.242	0.012	0.059
	NN	0.605	0.007	0.004	0.000	0.495	0.173	0.359	0.002	0.308
	SVM	0.625	0.001	0.004	1.000	0.526	0.164	0.475	0.046	0.275
$K_{transloc}$	DT	0.872	n.a.	0.136	1.000	0.814	0.195	0.729	0.577	0.521
	NN	0.848	0.015	0.005	n.a.	0.667	0.307	0.807	0.071	0.571
	SVM	0.763	0.017	0.007	1.000	0.693	0.387	0.735	0.248	0.505
$K_{translition}$	DT	0.540	0.000	0.564	1.000	0.560	0.175	0.332	0.021	0.114
	NN	0.714	0.490	0.743	0.000	0.742	0.564	0.445	0.035	0.540
	SVM	0.819	0.046	0.623	1.000	0.759	0.423	0.646	0.186	0.545

The results presented in Table 5.4 enable a detailed analysis of the capability of all three used ML techniques to predict changes for each of land use classes.

DT, NN and SVM do not model each class with the same accuracy. For example, DT is less capable for predicting the *Commercial* land use class in comparison to the other two techniques (based on *kappa* and *kappa simulation* values), particularly with regard to the amount of that class in the future (based on *kappa histo* and *kappa transition* values). On the other hand, DT is more capable for predicting the *Traffic areas* land use class then the other two techniques, particularly with regards to the location of that class in the future (based on *kappa location* and *kappa transloc* values).

In accordance with the processed assessment of similarity between planned and actual land use maps (section 4.4), none of the infrastructure objects have been built above the total area from 2001 to 2010. Therefore, this is the class with no changes and DT and SVM successfully "learned" that class does not change whereas NN did not (*kappa* ≠ 1). The small amount of changes on the *Wetland* class during the observed time period was caused by the start of construction of the bridge over the Sava. All ML techniques registered those changes during the learning process. However, since the number of changed cells is very small (0.4%) when compared to all cells in the *Wetland* class, the resulting changes are not learned so well. During the reclassifications, the existing *Not built* class was predefined into the *Green area* or *Agriculture* class based on the actual state detected on ortophoto maps. Additionally, the *Green area* class contains areas used for several different purpose (*Cemetery*, *Parks*, *Recreation* and *Not built*). Since the *Green area* class contains areas with various purposes, ML techniques had difficulties to find (learn) rules of changes. In order to improve modeling of this class it is possible to predefine this class as two or more classes. The outcomes of these selected models built using DT, NN and SVM, along with the actual LUC for the period from 2007 to 2010 for all three municipalities are presented in Figure 5.10, 5.11 and 5.12.

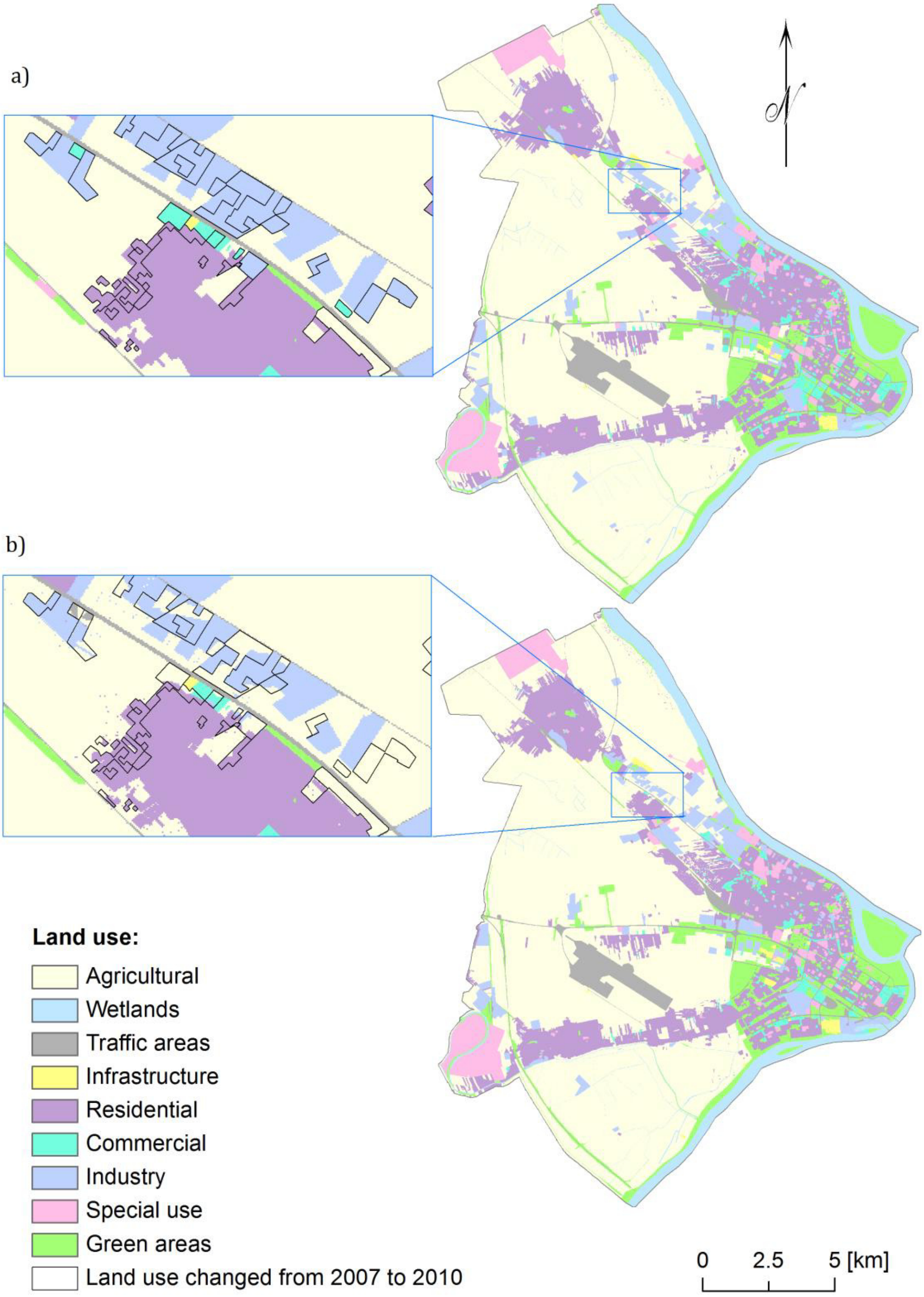


Figure 5.10 a) The actual land use and b) predicted land use for year 2010 obtained with Decision Trees**.**

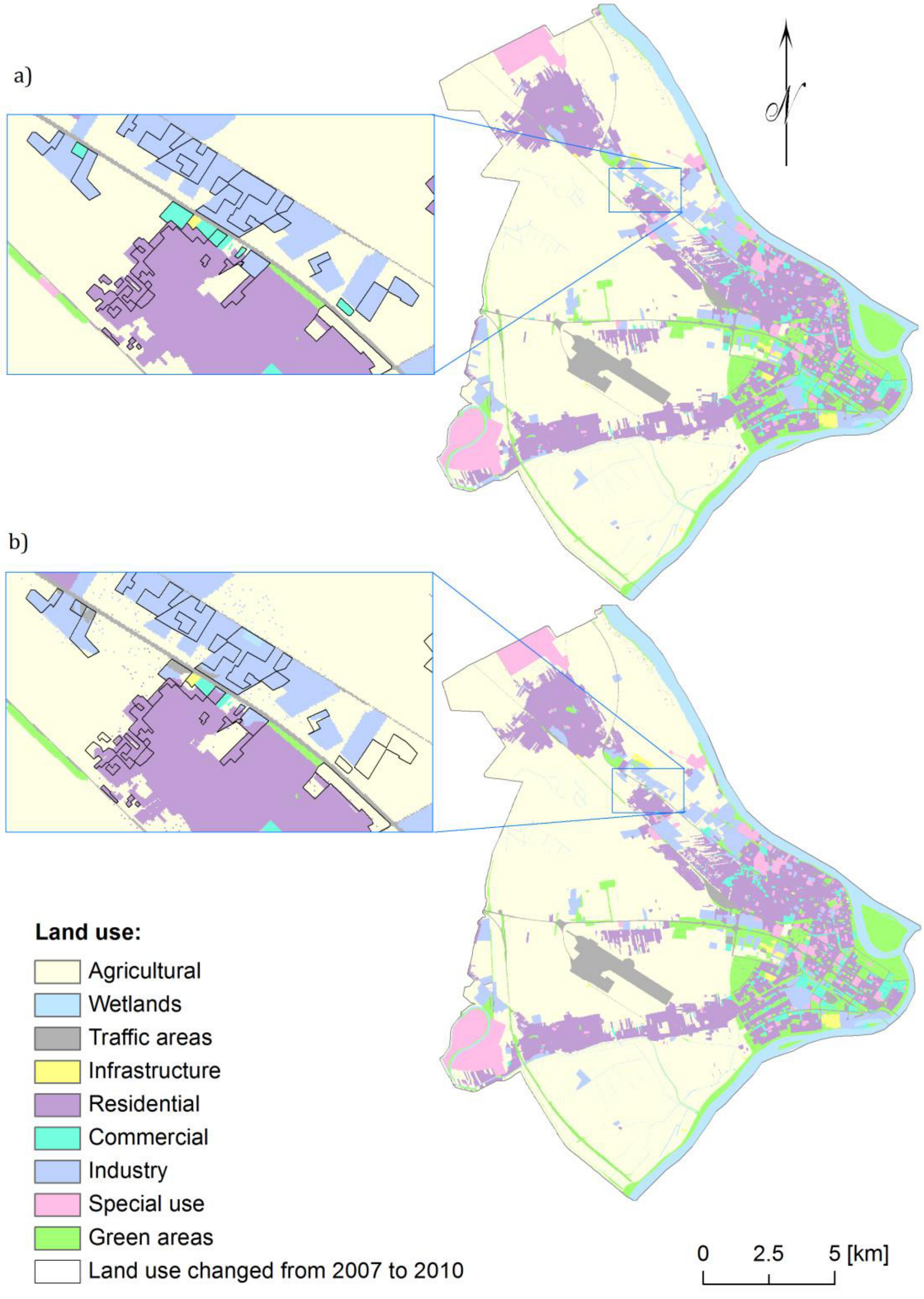


Figure 5.11 a) The actual land use and b) predicted land use for year 2010 obtained with Neural Networks**.**

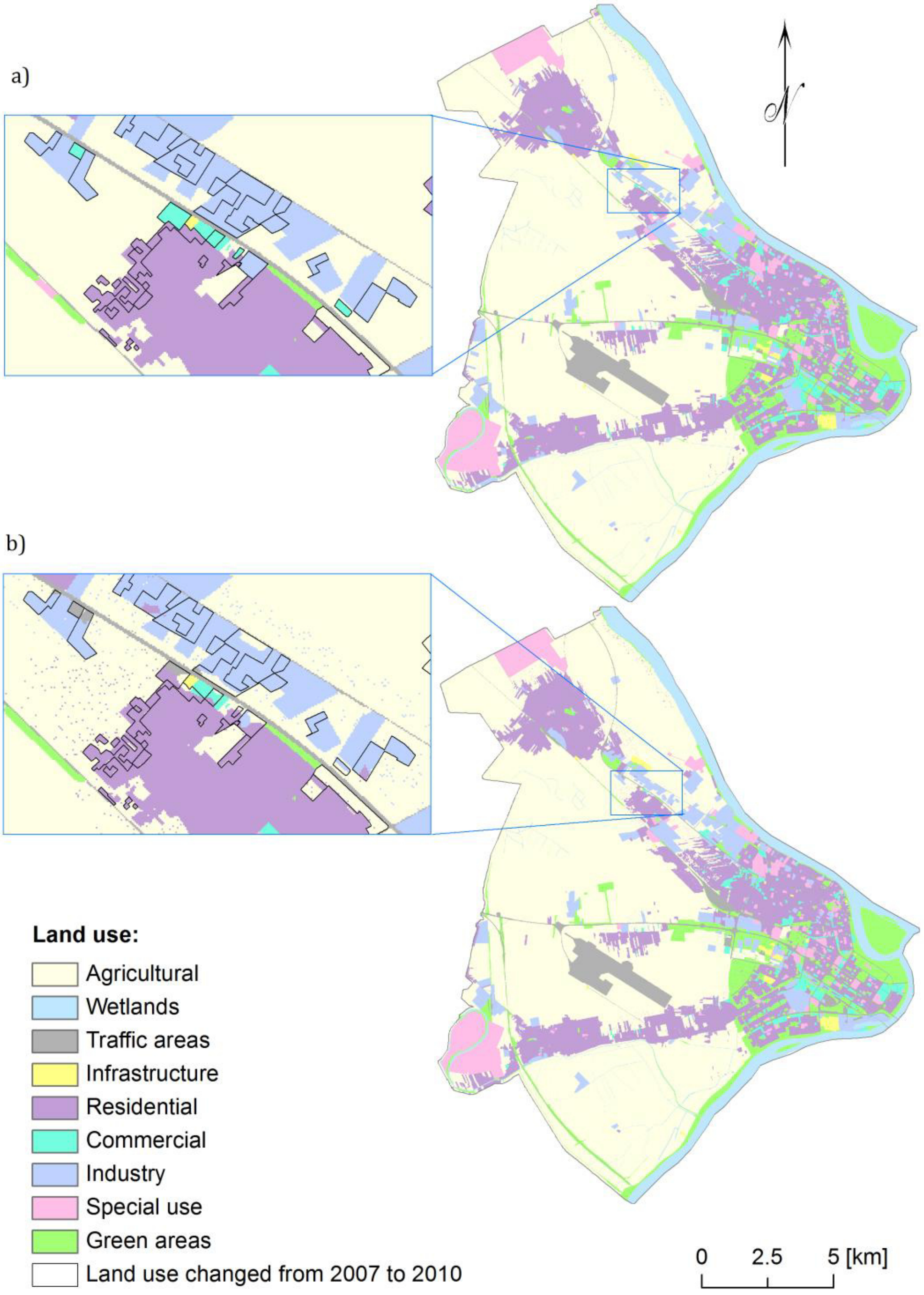


Figure 5.12 a) The actual land use and b) predicted land use for year 2010 obtained with Support Vector Machines.

5.1.6 Analysis of maps of modelling outcomes

The analysis of the maps of modelling outcomes (Figures 5.10, 5.11 and 5.12) was performed in consultation with urban planners. By comparing actual land use and predicted land use maps on marked polygons (areas where changes have taken place during the period from 2007 to 2010), it can be concluded that the differences are irrelevant from an urban point of view.

The detailed analysis of the distribution of classes in the map of actual land use differs from the map of predicted land use and can be performed based on the generic form of a contingency table (explained in section 3.6.1) (Table 5.5).

Table 5.5 Generic form of a contingency table for a selected model built by SVM.

Classes		Map B - predicted land use									
		1	2	3	4	5	6	7	8	9	Σmap A
Map A - actual land use	1	0.256	0.000	0.002	0.000	0.024	0.001	0.006	0.000	0.003	0.291
	2	0.000	0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.025
	3	0.000	0.000	0.021	0.000	0.005	0.002	0.001	0.000	0.002	0.032
	4	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.002
	5	0.057	0.000	0.004	0.000	0.264	0.004	0.004	0.001	0.009	0.344
	6	0.004	0.000	0.006	0.000	0.026	0.025	0.030	0.000	0.012	0.104
	7	0.016	0.000	0.002	0.000	0.012	0.001	0.063	0.001	0.004	0.098
	8	0.002	0.000	0.001	0.000	0.006	0.005	0.001	0.024	0.002	0.040
	9	0.000	0.000	0.004	0.000	0.025	0.002	0.004	0.001	0.026	0.064
	ΣmapB	0.335	0.025	0.041	0.002	0.363	0.039	0.109	0.027	0.059	1

During the visual examination, it was determined that the modelling outcome derived by SVM largely agreed with the actual land use map. However, SVM suffers (is corrupted) from “salt & pepper” noise (classification error) slightly more in comparison to the other two techniques. Salt & pepper noise can be defined as a kind of impulse noise in which only a few pixels are noisy and their values are often extreme (Boncelet 2005). In a gray scale image, these noisy pixels look like salt and pepper spread on the image. In the classified maps (as is the case in this dissertation), salt & pepper noise can be considered as a single pixel (or a small group of contiguous pixels) that is distinct from its (or their) spatial neighbourhood.

Reduction of that deficiency was accomplished using a median filter (Boncelet, 2005). The median filter works by moving a window (of 3x3 cell size) through the map pixel by pixel and replacing each pixel with the median value of neighboring pixels. In median filtering, the neighboring pixels are ranked according to intensity and the median value becomes the new value for the central pixel.

Using median filter on maps of modelling outcomes improves classification accuracy especially for prediction of the changed cells (Figure 5.13)

As a result, for a selected model built using SVM, the *kappa* value is increased by 0.042 (0.653) and the *kappa simulation* by 0.055 (0.518). The accuracy for predicting the future location of all cells increased by 0.056 (*kappa location*=0.733) and for future location of changed cells increased significantly by 0.142 (*kappa transloc*=0.812).

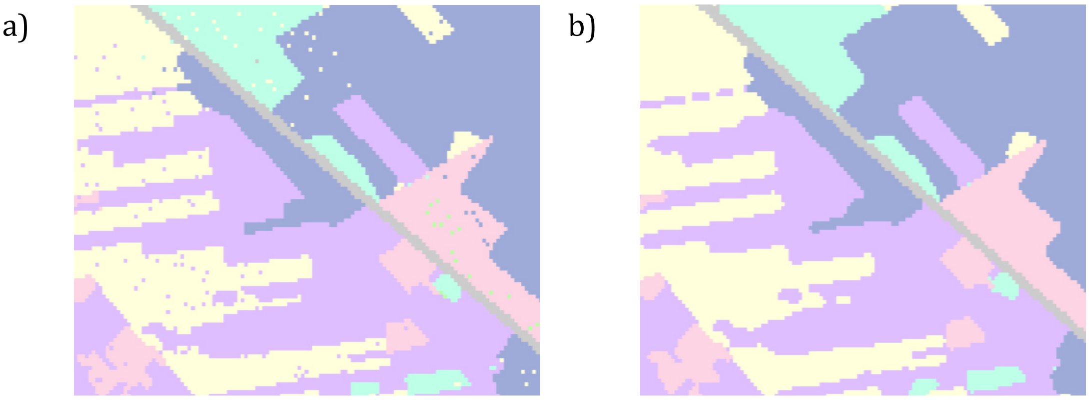


Figure 5.13 Part of map of modelling outcome a) before and b) after applying median filter.

However, the accuracy for predicting the future quantity of all cells stayed the same and the future quantity of changed cells decreased by 0.064 (*kappa translοc*=0.637).

The profound analysis can be performed using maps of modelling outcome which present the probability of occurrence of each class individually. Namely, the outcomes from ML techniques can be presented in two ways: as maps of probability of occurrence for each class individually and as a map in which each pixel is associated with the class that has the highest probability.

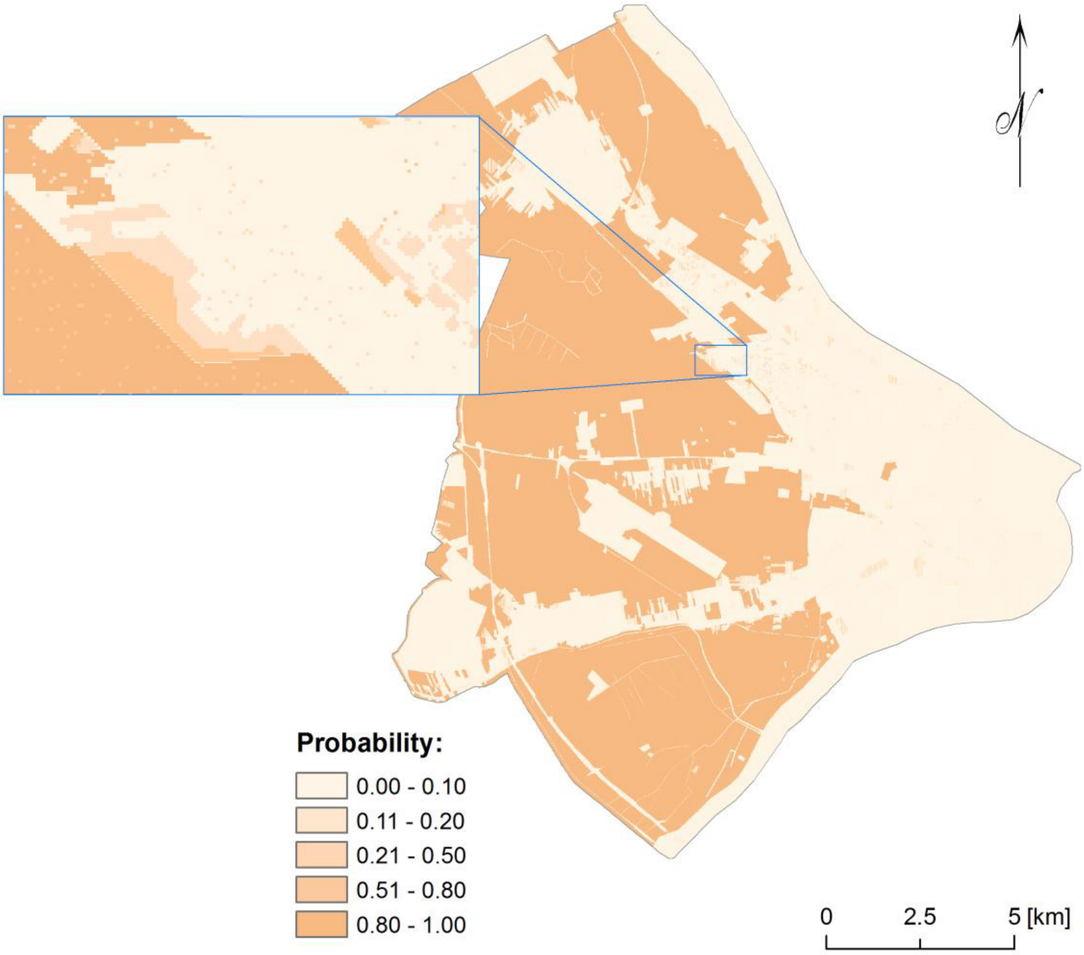


Figure 5.14 Map of probability of occurrence for *Agriculture* class.

Maps of probability can greatly contribute to the decision making processes of urban planners. The probability map for the *Agricultural* class is presented in Figure 5.14, while the probability maps for other classes are presented in appendices 1-8.

Moreover, there is some degree of similarity in land use considering the overlap of individual classes such as *Commercial* and *Residential* classes, which are often combined in one area.

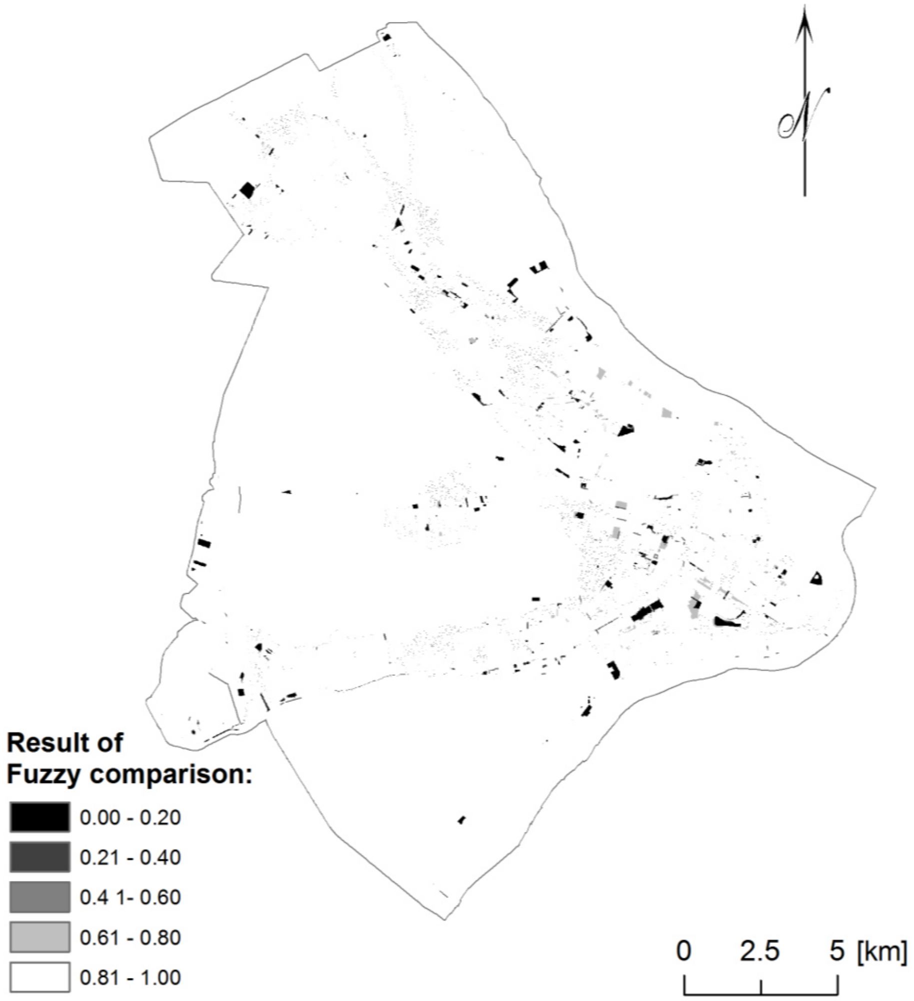


Figure 5.15 Comparison of actual and predicted land use maps for year 2010 in the fuzzy set approach.

For that reason, it is convenient to consider the probability of similar classes during the analysis of accuracy of modelling outcomes. From the planner's point of view, if one of the similar classes emerges instead of another one, it does not necessarily have to mean that the prediction was completely inaccurate.

The comparison of actual and predicted land use maps for year 2010 was carried out using fuzzy kappa in the same way as in the assessment of similarity between actual and planned land use maps (section 4.4). Since the fuzzy set approach provides gradual maps of similarity, it is possible to analyze differences between these two maps considering the level of similarity between particular classes (Figure 5.15).

5.2. Sensitivity of the predictive land use change model built by SVM

Obtaining the best performing model implies proper implementation of the following: adequate data representation, sampling data and selection of the appropriate subset of attributes as was shown in previous experiments. In addition, the efficient application of SVM requires the selection of optimal parameters. For this reason, special attention is given to this phase of LUC modeling using SVM.

The efficient application of SVM, which uses the Radial Basis Function kernel (explained in section 3.3.3), requires the selection of the optimal combination of penalty factor C and the Gaussian parameter γ. A standard parameter selection procedure assumes the existence of a separate validation set used to measure the performance of the model trained with selected parameter combinations. The best performing combination on the validation set is used to train the model on the whole training set. The validation set is usually obtained by dividing the training set into two (or more in the case of *k*-fold cross validation) independent parts.

Since the research problem has the form of a time series, the validation set should not be from the same time period as the test set. Therefore, the parameters are varied on the training set ($\mathbf{x}^{t-2}$, y^{t-1}) and the model performance is measured on the ($\mathbf{x}^{t-1}$, y^{t}) validation set. The best performing parameters were used to train the final SVM prediction model on ($\mathbf{x}^{t-1}$, y^{t}), which is then used to predict y^{t+1}.

In this part of the dissertation the following was examined:

- Sensitivity of the LUC model built by SVM using all considered attributes and the same number of attributes selected by Info Gain, Gain Ratio or Correlation based Feature Subset with regards to the SVM parameter changes,
- Sensitivity of the LUC model built by SVM using subsets of attribute selected from different data representations by Correlation based Feature Subset with regards to the SVM parameters. Additionally, a realistic performance of the SVM model was tested by finding the appropriate parameters using only the available data from the past (*t-2*, *t-1*; *t*) and then using those best performing parameters for modeling (predicting) "unknown" future land use, y^{t+1}.

5.2.1 Study area and datasets creation for Training and Testing

The Zemun municipality was used as a study area in this experiment. Since the area contained a small amount of cells with changed land use, the balanced sampling approach was used to create training and test datasets for building and evaluation of derived models according to the procedure described in section 3.4.

Three different data representations were used to build proposed SVM prediction models. The first representation included attributes x_1 to x_{10} and attribute x_{20} described in Table 5.6. These attributes were used to build the basic model based on distances to significant objects, population and previous land use (in further text referred as M). In the other two neighbourhood representations, information (x_{11} to x_{20}, Table 5.6.) was added to the basic model and presented with two variations: $M_{7 \times 7}$ with Moore neighbourhood of 7x7 and $M_{21 \times 21}$ with Moore neighbourhood of 21x21 (explained in section 4.3).

Table 5.6 Attributes used for different data representation M, M_{7x7} and M_{21x21}.

Attributes	Description
x_1	Euclidean distance of grid cell to municipality center
x_2	Euclidean distance of grid cell to city center
x_3	Euclidean distance of grid cell to the closest rivers
x_4	Euclidean distance of grid cell to the closest big green areas
x_5	Euclidean distance of grid cell to the closest railway lines at time *t*
x_6	Euclidean distance of grid cell to the closest highway at time *t*
x_7	Euclidean distance of grid cell to the closest main road at time *t*
x_8	Euclidean distance of grid cell to the closest street of category I at time *t*
x_9	Euclidean distance of grid cell to the closest street of category II at time *t*
x_{10}	Population change index
x_{11}	Number of agricultural cells in neighbourhood at time *t*
x_{12}	Number of wetlands cells in neighbourhood at time *t*
x_{13}	Number of traffic areas cells in neighbourhood at time *t*
x_{14}	Number of infrastructure cells in neighbourhood at time *t*
x_{15}	Number of residential cells in neighbourhood at time *t*
x_{16}	Number of commercial cells in neighbourhood at time *t*
x_{17}	Number of industry cells in neighbourhood in at time *t*
x_{18}	Number of special use cells in neighbourhood at time *t*
x_{19}	Number of green areas cells in neighbourhood at time *t*
x_{20}	Land use class at time *t*
y	Land use class at time *t+1*

5.2.2. Sensitivity of SVM predictive land use change model in regard to parameter changes and used attributes selection methods

Data representation used for this experiment was M and models were built using the training set (2003 – 2007) and tested over the test set (2007 – 2010).

In the first step of experimentation, an attribute selection was carried out using three methods: IG, GR and CFS. The obtained results are presented below, in Figure 5.16.

The CFS selected subset of five attributes including: land use class, PCI and Euclidian distance to the closest big green area, highway and main road. Whereas, the CFS automatically determines a subset of *k* relevant attributes, which are highly correlated with the land use class and are uncorrelated with each other; IG and GR rank all considered attributes independently of each other according to their measure of association with the future land use class. Therefore, in order to compare the sensitivity of models built with attributes selected with those three methods in regards to SVM parameters, the five highest ranked attributes by IG and GR were selected and three new data representations were created.

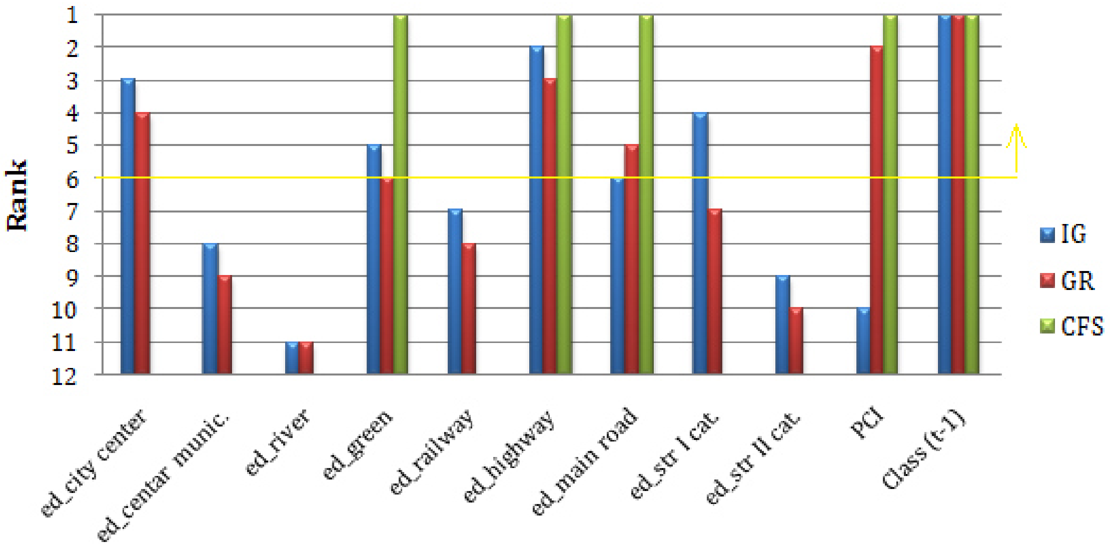


Figure 5.16 Selection of attributes by IG, GR and CFS.

Each data representation contains training and test datasets and five selected attributes by IG, GR and CFS.

In the second step, the model performance derived with all attributes (M) and with a subset of 5 selected attributes based on the CFS (M^{CFS}), IG (M^{IG}) and GR (M^{GR}) for different SVM parameters *C* and γ were compared. The range of parameters used to train SVM models was {1, 5, 10, 50, 100} for *C* and {0.5, 1, 5, 10} for γ, which makes the total of 20 combinations created for each individual data representation. Validation values measured for all models that were built using various SVM parameters and four different data representations are presented in Figure 5. 17.

Compared to the model based on all attributes, the use of a subset of attributes selected by CFS increases the kappa value by 6%. Whereas the use of a subset of attributes selected by GR increases the kappa by 4% and the use of IG decreases the kappa value by 10%. Results indicate that M^{CFS} and M^{GR} are more robust to different SVM parameter combinations and exhibit better kappa performance.

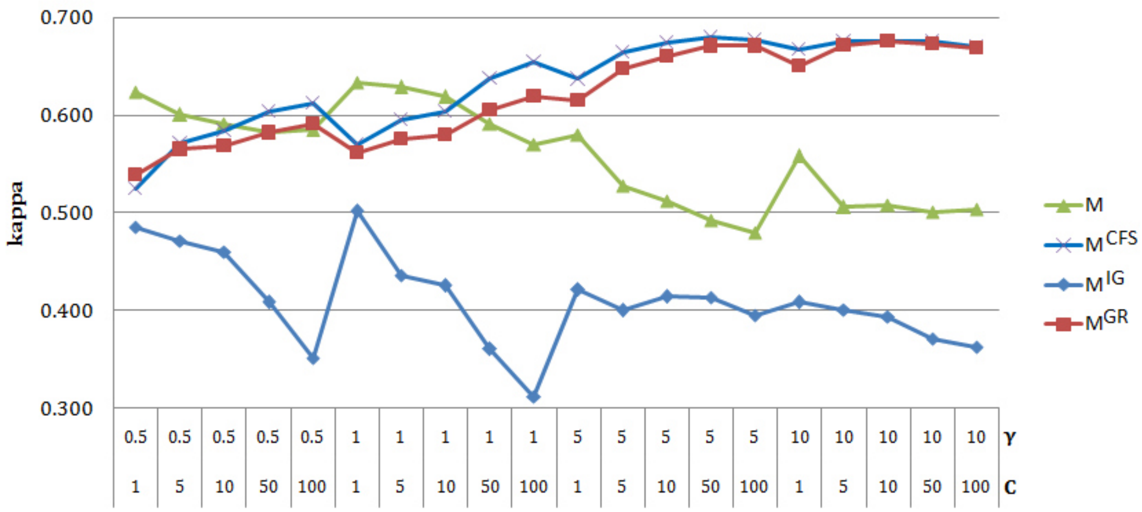


Figure 5.17 Comparison of the model performance derived with all attributes (M) and with subset of 5 selected attributes based on the CFS (M^{CFS}), IG (M^{IG}) and GR (M^{GR}) for different SVM parameters *C* and γ.

Models built based on M^{CFS} are slightly better than the ones based on M^{GR}. Using M and M^{IG} provides models that are less capable at predicting LUC and can be overfitted with higher values of parameters.

Generally, a subset of *k* attributes selected by CFS provides slightly better models when compared to *k* highest ranked attributes by GR and significantly better models compared to *k* highest ranked attributes by IG.

5.2.3 Sensitivity of SVM predictive land use change model in regard to parameter changes and subset of attributes selected by CFS from different data representation

The model built with attributes selected by CFS provides slightly better model performance when compared to others and therefore this attribute selection method was used for this experiment. In the first step of the experiment, a CFS method was performed for each data representation, M, $M_{7 \text{ x } 7}$ and $M_{21 \text{ x } 21}$, in order to find the most informative subset of attributes from Table 5.6. The results are shown in Table 5.7.

The number of selected attributes increased with the addition of neighbourhood information to the basic model M (M_{7x7}, M_{21x21}) and by expanding the size of neighbourhood (M_{21x21}). In all training datasets, the CFS selected land use class (in *t*), PCI and Euclidean distances to the closest big green area, highway and main road were used. Additionally, the most relevant classes in the 7×7 neighbourhood are *Wetlands*, *Traffic areas* and *Special use* (school, hospital, police station...). Beside those attributes, in the 21×21 neighbourhood, residential and industry class are joined to the subset of selected attributes. Models labelled as M^{CFS}, $M^{CFS}{}_{7x7}$ and $M^{CFS}{}_{21x21}$ were built on training sets that contain only selected attributes (Table 5.6).

Table 5.7 Subset of selected attributes based on the CFS method for three data representations.

M	$M_{7 \times 7}$	$M_{21 \times 21}$
Euclidean distance of grid cell to the closest big green areas	Euclidean distance of grid cell to city center	Euclidean distance of grid cell to city center
Euclidean distance of grid cell to the closest highway	Euclidean distance of grid cell to the closest big green areas	Euclidean distance of grid cell to the closest big green areas
Euclidean distance of grid cell to the closest main road	Euclidean distance of grid cell to the closest railway lines	Euclidean distance of grid cell to the closest railway lines
Population change index	Euclidean distance of grid cell to the closest highway	Euclidean distance of grid cell to the closest highway
Land use class	Euclidean distance of grid cell to the closest main road	Euclidean distance of grid cell to the closest main road
-	Population change index	Population change index
-	Number of wetlands cells in neighbourhood	Number of wetlands cells in neighbourhood
-	Number of traffic areas cells in neighbourhood	Number of traffic areas cells in neighbourhood
-	Number of special use cells in neighbourhood	Number of special use cells in neighbourhood
-	Land use class	Number of residential cells in neighbourhood
-	-	Number of industry cells in neighbourhood
-	-	Land use class

In order to examine the influence of the parameters on the performance of the models, individual models were built using the training set (2003 – 2007) and tested over the test set (2007 – 2010). The obtained *kappa* values are presented in Figure 5.18.

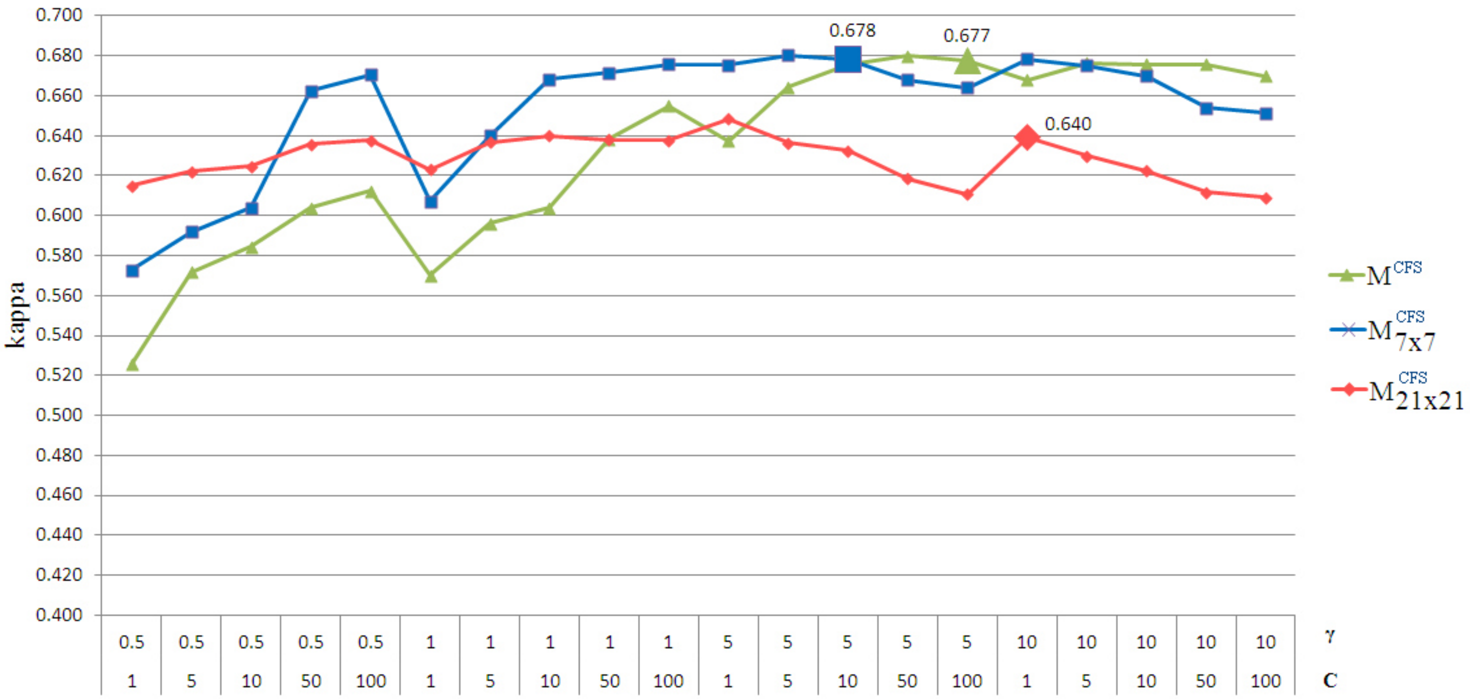


Figure 5.18 Comparison of the model performance derived with different data representation and different values of SVM parameters C and γ.

The results show that SVM provided an adequate prediction of LUC using all three data representations. The ranges of *kappa* for M^{CFS}, $M^{CFS}{}_{7x7}$, and $M^{CFS}{}_{21x21}$ are 0.52-0.68, 0.57-0.68 and 0.61-0.65, respectively. Although the use of $M^{CFS}{}_{21x21}$ obtains slightly lower results, this dataset is less sensitive to parameter changes than the other two. The highest *kappa* values for M^{CFS} and $M^{CFS}{}_{7x7}$ are similar. However, the $M^{CFS}{}_{7x7}$ performs better with smaller values of parameters which indicate that $M^{CFS}{}_{7x7}$ has a greater power of generalization.

In the previous step all models were evaluated on the corresponding test sets. The test sets belong to the future from the perspective of data used to train the models. In reality, land use experts would build an operative model by finding the appropriate parameters using only the available data from the past.

A more realistic case can be examined if 2010 is used as an unknown future land use class (y^{t+1}). A realistic performance of models was accomplished in following manner:

1. Models were built based on 2001 – 2003 training datasets and tested on 2003 – 2007 validation sets for each of 20 parameters combinations

2. Best performing parameters were selected for each data representation (M^{CFS} – [5,100]; $M^{CFS}{}_{7x7}$ – [5,10]; $M^{CFS}{}_{21x21}$ – [10,1])

3. Models were built based on 2003 – 2007 training datasets and tested on 2007 – 2010 using previously selected best performing parameters (Table 5.8).

Table 5.8 Values for different kappa measures for models that were built based on selected parameters, for all three data representations.

Dataset	*C*	γ	*Kappa*	$K_{location}$	K_{histo}
M^{CFS}	100	5	0.677	0.744	0.910
$M^{CFS}{}_{7x7}$	10	5	0.678	0.757	0.895
$M^{CFS}{}_{21x21}$	1	10	0.640	0.729	0.877

The model performance on the 2007 – 2010 test set could be regarded as a realistic assessment of its capability to predict future LUC (operative model performance emphasized in Figure 5.18 with bigger markers). Models M^{CFS} and M^{CFS}_{7x7} exhibited similar values of *kappa* and the value for M^{CFS}_{21x21} was slightly lower.

In order to provide a better overall evaluation, two additional kappa values had to be analyzed: kappa location and kappa histo. The M^{CFS}_{7x7} model was able to predict the location of changes slightly better than M^{CFS} and M^{CFS}_{21x21} models. However M^{CFS} was better at predicting the quantity of change than the other two models. Hence, it is difficult to judge what could be the best model independently for the final application. Therefore, the model performance for all classes from the study area is show in Table 5.9.

All classes, excluding the *Commercial* class, resulted with *kappa* values higher than 0.50 and indicated that SVM had difficulties to "learn" changes for that land use class. This could be explained by the kind of data, since this class is often combined with two different classes (*Industrial* and *Residential*). *Infrastructure* and *Wetland* were the only two classes without changes within the period between 2001 – 2010 years (*kappa* equals one). *Kappa location* values indicate that the inclusion of neighbourhood information improves the results significantly for certain classes. M^{CFS}_{7x7} predicts locations for future traffic areas significantly better when compared to the other two models. On the other hand, M^{CFS}_{21x21} significantly improves the prediction for *Commercia*l classes. The prediction for the quantity of changes is excellent for the majority of classes. It is the responsibility of the urban planner to decide which one of the presented models would be selected. The obtained predicted land use changes are presented in Figures 5.19, 5.20 and 5.21.

Table 5.9 Performance measures of selected models for all classes, based on *kappa*, *kappa location* and *kappa histo* values.

Land use class	*Kappa*			$K_{location}$			K_{histo}		
	M^{CFS}	$M^{CFS}{}_{7x7}$	$M^{CFS}{}_{21x21}$	M^{CFS}	$M^{CFS}{}_{7x7}$	$M^{CFS}{}_{21x21}$	M^{CFS}	$M^{CFS}{}_{7x7}$	$M^{CFS}{}_{21x21}$
Agriculture	0.722	0.723	0.679	0.754	0.764	0.763	0.958	0.946	0.889
Wetland	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Traffic areas	0.545	0.657	0.563	0.549	0.812	0.574	0.999	0.808	0.981
Infrastructure	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Residential	0.743	0.748	0.673	0.743	0.776	0.681	0.983	0.964	0.988
Commercial	0.168	0.165	0.168	0.312	0.296	0.397	0.538	0.558	0.423
Industry	0.680	0.654	0.661	0.686	0.670	0.692	0.992	0.976	0.955
Special use	0.594	0.594	0.600	0.932	0.932	0.934	0.645	0.638	0.642
Green area	0.506	0.518	0.477	0.953	0.954	0.962	0.531	0.542	0.496

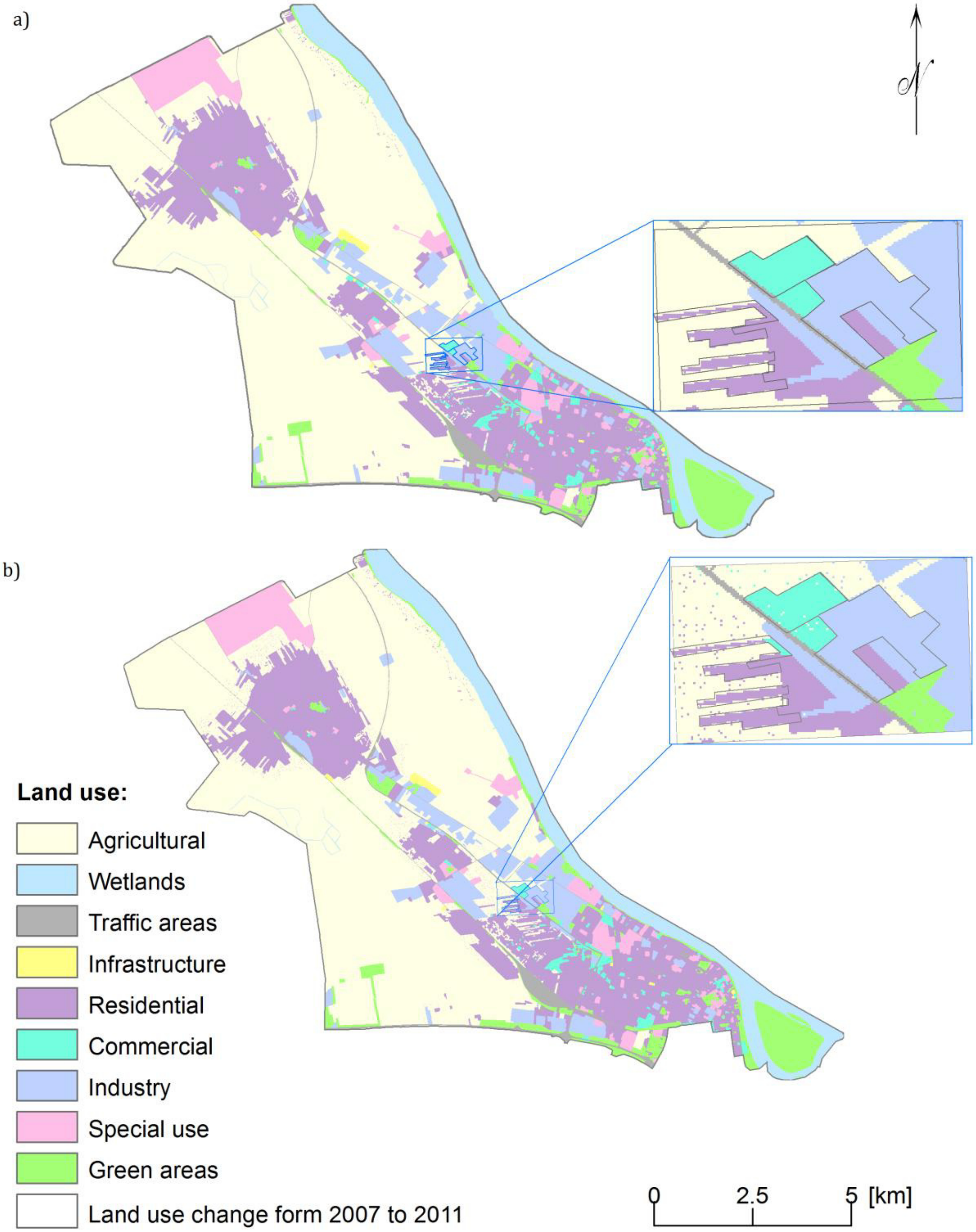


Figure 5.19 a) The actual land use and b) predicted land use for year 2010 obtained with selected parameters and M^{CFS} data representation.

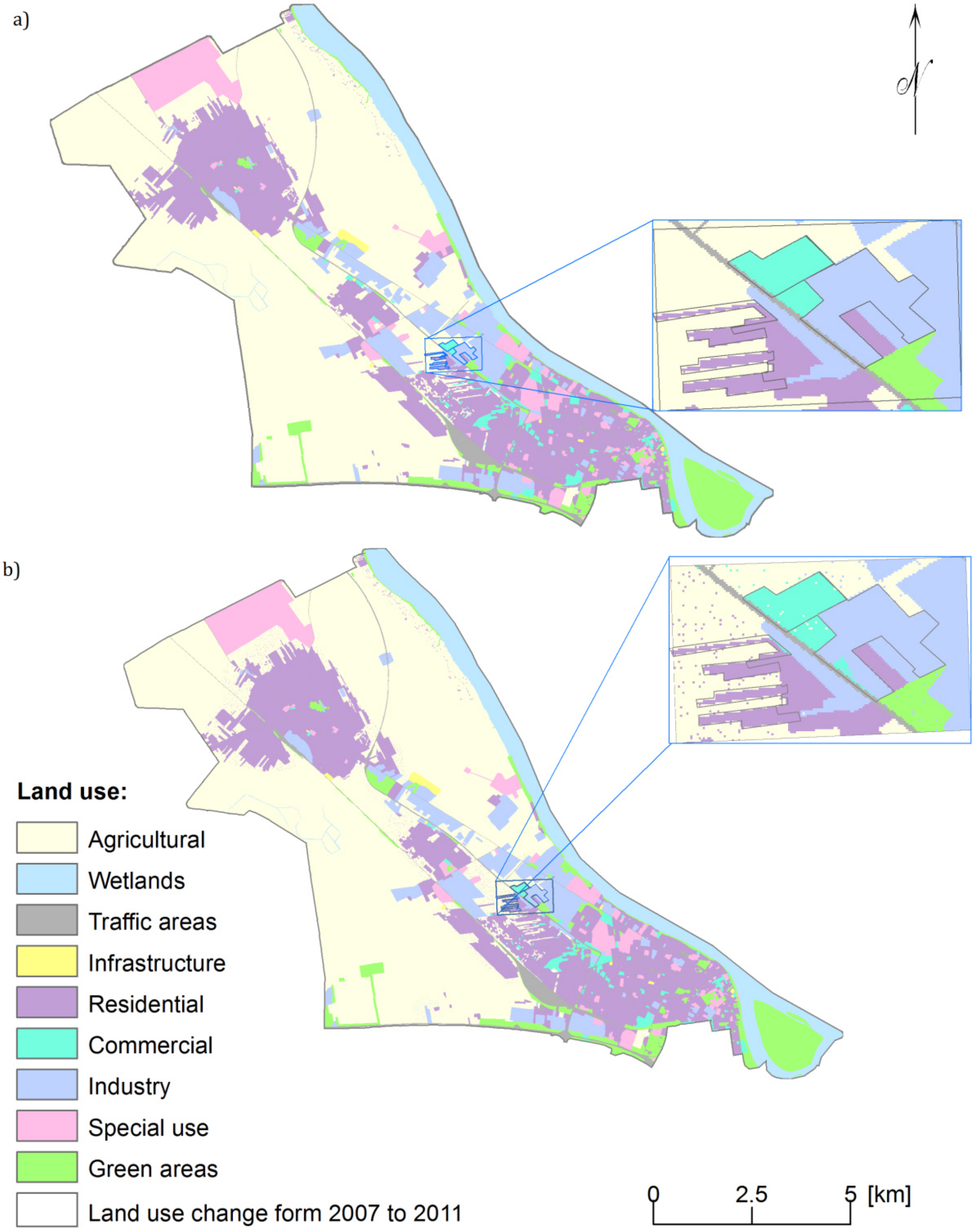


Figure 5.20 a) The actual land use and b) predicted land use for year 2010 obtained with selected parameters and $\mathrm{M^{CFS}}_{7x7}$ data representation.

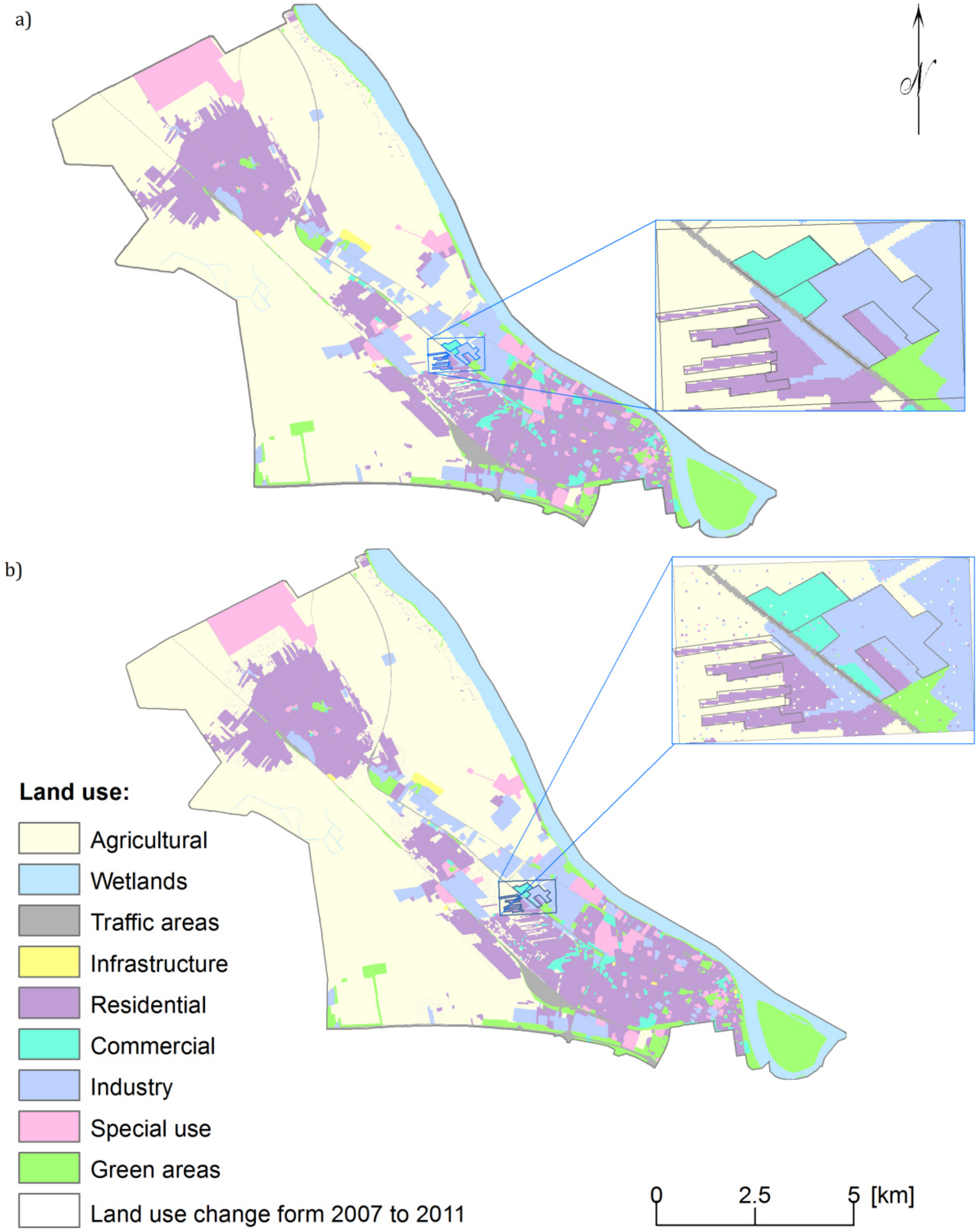


Figure 5.21 a) The actual land use and b) predicted land use for year 2010 obtained with selected parameters and $M^{CFS}{}_{21x21}$ data representation.

Chapter 6:

Conclusion

The development of models for the analysis and prediction of dynamic geographic phenomena has been spurred recently by the vast availability of geospatial data in digital form, development of GIS and related technologies. As a relatively new approach, which has the capability to develop modelling procedures for representation of the underlying complex phenomena from historical datasets, data-driven methods are insufficiently researched in the field of land use. This research examined the possibility of applying different data-driven methods to predict urban land use changes (LUC). The proposed methodology included problem definition, data collection and preparation, data sampling, analysis of urban LUC attributes importance, building, validating and analyzing models for LUC prediction.

Four experiments were conducted to examine the proposed methodology. Different representations of data, different attribute selection methods and different numbers of the attributes were considered. Furthermore, three machine learning techniques were used (Decision Trees, Neural Networks and Support Vector Machines) in order to build models and model outcomes were compared (using various forms of Kappa statistics) and analyzed. Moreover, nine land use classes were considered for building and validating the models which added to the complexity of the research. The study area that was used for these four experiments encompasses three Belgrade municipalities (Zemun, New Belgrade and Surčin) represented as 10×10 m grid cells in four different moments in time (2001, 2003, 2007 and 2010).

Land use changes present very complex spatial - temporal process which depends on many different factors (attributes). Creating database for study area in GIS environment enables integration, manipulation and analyzes of different types of considering attributes. In the proposed methodology the study area is represented as a grid of cells, where each cell is uniquely identified with its accompanied attributes and land use class. For each considering moment in time, the GIS data layer (point layer- where each point represents center of established grid cell) is created and presents distribution of considering attributes on a study area for each moment in time. By using those GIS data layers it is easy to create training and test datasets which are necessary to built and validate models.

In land use problem domain, the overall number of cells that change their land use over time is very small compared to the total number of cells in the study area. Therefore, the unbalanced nature of the data could be misleading for both the learning process and the evaluation of the model performance. A proposed procedure in this research that selects an equal number of land use changed and unchanged cells preserving original distribution over the classes yielded better predictive models.

The study experimented with four different data representations used as inputs to different machine learning techniques. Apart from the commonly used urban indicators in LUC modelling (such as present land use class, number of inhabitants, distances to city centre, highways, roads) representations used in this research included the neighbouring information about land use classes, history information about previous land use and information regarding the changes of some spatial attributes that occurred in the past at each cell in the grid. The experiments suggested that a model with neighbouring information performed better than a simple model with common urban indicators. However, further enhancement of the data representation with historical information related to previous land use improved the prediction over the neighbouring model. On the other hand, the representation with changed attributes information does not contribute to a significant improvement of the model performance, as might be expected, especially regarding to relatively small amount of changes in the considered attributes.

A detailed analysis of attribute importance was performed using three attribute ranking methods: χ^2, Info Gain and Gain Ratio in order to find the most relevant attributes for all data representations. The selection of the method for the attribute ranking depends on the machine learning techniques and on the types of attributes themselves because the methods favored differing types of attributes. It is found that both χ^2 and IG rank the attributes almost identically while GR indicates some differences. Both methods showed that the previous land use and neighborhood are very important but not sufficient for an accurate modeling result. By using recursive elimination method, a subset of the most informative attributes was found. The experiments suggested that relatively small number of attributes (6 or 5) was sufficient for realistic model predictions. Reducing the initial set of input attributes to an informative subset resulted in less complex models, in regard to number of used attributes, and models with better performance. Additionally, the proposed methodology allows the consideration of numerous attributes (categorical and continuous) followed by proposed selection of subset of the most informative ones for the learning process. This approach enables a wide range of user unbiased application of model outcomes.

Based on the obtained results it can be concluded that all three machine learning techniques are suitable for modeling land use change. When compared together, the NN, DT and SVM techniques all have some advantages and disadvantages. Generally, NN and SVM have a slightly better capability to model changes than DT. Considering that the prediction of the correct location where the changes occurred is more important than overall quantity of changes, the SVM technique is more appropriate than NN, from an urban point of view. The capability of the DT to reduce the derived model as a set of logical IF–THEN rules enables a domain expert to have an easier understanding of the problem and in many cases could be preferable to more complex functional methods such as SVM.

In addition, sensitivity analysis of the SVM-based model was performed in order to explore its sensitivity to attribute selection and parameter changes. Models were built and the outcomes were compared and analyzed by using the same number of attributes selected from the new three different data representations obtained through Info Gain, Gain Ratio and Correlation-based Feature Subset and various combinations of SVM parameters. Moreover, the capability to find the appropriate optimal SVM parameters using only data from the past is necessary to test the predictions of future land use changes. The municipality of Zemun was used as the main testing area, which is the largest municipality within the Belgrade city limits.

The obtained results indicate that a subset of *k* attributes selected by CFS provides slightly better models compared to *k* highest ranked attributes by GR and provides significantly better models compared to *k* highest ranked attributes by IG. Using selected attributes by CFS and GR resulted in a simple model (less attributes – less complicated model) with better performance and with less possibility to be overfitted with higher values of SVM parameters.

In conclusion, this research presents a novel means of enhancing data-driven methods and assesses their suitability for modelling land use change in cases of high thematic resolutions –i.e. large amount of classes. Based on the obtained results it can be concluded that the proposed data-driven methodology provides predictive LUC models which could be successfully used for creation of possible scenarios of urban LUC and presents helpful tools for modern urban planning decision making purposes. Considering the flexibility of the proposed methodology, in regard to used attributes, number of target classes, spatial and temporal resolution, it has great potential in application for modelling other spatial-temporal phenomena.

Bibliography

Abe S., 2010. Support Vector Machines for pattern classification. Springer, London: 471pp.

Abraham T.H., 2002. (Physio) logical circuits: The intellectual origins of the McCulloch–Pitts neural networks. Journal of the History of the Behavioral Sciences 38(1): 3-25.

Agarwal C., Green G.M., Grove J.M., Evans T.P. and Schweik C.M., 2002. A review and assessment of land-use change models: dynamics of space, time, and human choice. Gen. Tech. Rep. NE-297. Newton Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station: 61pp.

Almeida C.M., Glerianib J.M., Castejon E.F. and Soares Filhod B.S., 2008. Using neural networks and cellular automata for modelling intra urban land use dynamics, International Journal of Geographical Information Science, 22 (9): 943-963.

Anderson J.R., 1976. A land use and land cover classification system for use with remote sensor data. US Government Printing Office 964: 28pp.

Araghinejad S. 2014. Data-Driven Modeling: Using MATLAB in Water Resources and Environmental Engineering. Water Science and Technology Library 67, Springer Netherlands: 400pp.

Aronoff S., 2005. Remote Sensing for GIS Managers. ESRI Press, Readlands CA: 487pp.

Bajat B., Krunić N., Samardzić-Petrović M., Kilibarda M. 2013. Dasymetric modelling of population dynamics in urban areas. Geodetski Vestnik 57(4):777-792.

Belousov A.I., Verzakov S.A. and Von Frese J., 2002. Applicational aspects of support vector machines. Journal of Chemometrics, 16 (8-10): 482-489.

Benedict K.F. and Lauffenburger D.A., 2013. Insights into proteomic immune cell signaling and communication via data-driven modelling, Systems Biology, Current Topics in Microbiology and Immunology, 363: 201-233.

Bischof H., Schneider W. and Pinz A.J., 1992. Multispectral classification of Landsat-images using neural networks, Geoscience and Remote Sensing, IEEE Transactions, 30 (3): 482-490.

Böhner J., Blaschke T. and Montanarella L., 2008. SAGA – Seconds Out. Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie, 19: 113pp.

Boncelet C., 2005. Image noise models, in Handbook of Image and Video Processing, 2nd ed., A. C. Bovik, Ed. Academic Press, ch. 4.5. 397-409.

Breiman L., Friedman J.H., Olshen R.A. and Stone C.J., 1984. Classification and Regression Trees, Wadsworth Advanced Books and Software, Belmont, CA: 368pp.

Brimicombe A., 2010. GIS, environmental modeling and engineering. CRC Press: 378pp.

Brown D.G., Page S., Riolo R., Zellner M. and Rand W., 2005. Path dependence and the validation of agent-based spatial models of land use. International Journal of Geographic Information Systems, 19 (2): 153-174.

Burrough P.A. and McDonnell R.A., 1998. Principles of Geographic Information Systems, 2nd Ed. Oxford University Press, Oxford: 332pp.

Burrough P.A., 1996. Natural objects with indeterminate boundaries. In: Burrough, P. A. and Frank, A. U. (eds), Geographic Objects with Indeterminate Boundaries. Taylor and Francis, London: 3-28.

Castella, J.C., Trung T.N. and Boissau S., 2005. Participatory simulation of land use changes in the Northern Mountains of Vietnam: the combined use of an agent-based model, a role-playing game, and a geographic information system. Ecology and Society, 10(1): 27pp.

Charif O., Omrani H., Basse R.-M. and Trigano P., 2012. Cellular automata model based on machine learning methods for simulating land use change, Simulation Conference (WSC), 9-12 Dec. 2012, Berlin, GermanyProceedings of the 2012 Winter, IEEE: 1-12.

Cheng J., 2003. Modelling spatial and temporal urban growth. Doctoral Dissertation, Faculty of Geographical Sciences Utrecht University, Netherlands

Chorley R.J. and Haggett P., 1967. Models in geography. London: Methuen: 816pp.

Cihlara J. and Jansenb L. J. M. 2001. From Land Cover to Land Use: A Methodology for Efficient Land Use Mapping over Large Areas. The Professional Geographer 53 (2): 275-289.

Clarke K.C., Hoppen S. and Gaydos L.J., 1997. A self-modifying cellular automaton model of historical urbanization in the San Franciso Bay area. Environment and Planning B 24: 247–61.

Cohen J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1): 37-46.

Couclelis H., 1985. Cellular worlds: a framework for modelling micro-macro dynamics. Environment and Planning A 17: 585-96.

Cristianini N. and Shawe-Taylor J., 2000. An Introduction to Support Vector Machines and other kernel-based learning methods. Cambridge University Press, Cambridge: 202pp.

Das S., 2001. Filters, wrappers and a boosting-based hybrid for feature selection. In ICML, 1: 74-81.

Đorđević D., 1997. Basic approaches in twentieth century land use planning. Collection of the papers XL VII, Geographical Faculty University of Belgrade, 69-84.

Dragicevic S., 2013a Week/Lecture 1: Spatial Modeling: Introduction. GEOG-Advanced Spatial Analysis and Modeling. Simon Fraser University.

Dragicevic S., 2013b Week/Lecture 6: Introduction to Agent-based modeling. GEOG-Advanced Spatial Analysis and Modeling. Simon Fraser University

Egresits Cs, Monostori L., and Hornyák J., 1998. Multistrategy learning approaches to generate and tune fuzzy control structures and their application in manufacturing. Journal of Intelligent Manufacturing 9(4): 323-329.

Epstein J.M., 2008. Why Model?. Journal of Artificial Societies and Social Simulation 11(4)12 (http://jasss.soc.surrey.ac.uk/11/4/12.html).

ESRI 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

European Environmental Agency, 2010. Raster data set of built-up and non built-up areas including continuous degree of soil sealing ranging from 0 - 100% in aggregated spatial resolution (100 x 100 m) (http://www.eea.europa.eu/data-and-maps/data/eea-fast-track-service-precursor-on-land-monitoring-degree-of-soil-sealing-100m-1 Accessed 20 September 2012).

Fayyad U. Piatetsky-Shapiro G. and Smyth P., 1996. From Data Mining to Knowledge Discovery in Databases, AI Magazine, 17(3): 37-54.

Foley J.A., DeFries R., Asner G.P., Barford C., Bonan G., Carpenter S.R., Chapin F.S., Coe M.T., Daily G.C., Gibbs H.K., Helkowski J.H., Holloway T., Howard E.A., Kucharik C.J., Monfreda C., Patz J.A., Prentice I.C., Ramankutty N. and Snyder P.K., 2005. Global consequences of land use. Science, 309: 570–574.

Foody G.M., 2004. Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy. Photogrammetric Engineering and Remote Sensing, 70(5): 627-634.

Friedl M.A. and Brodley C.E., 1997. Decision tree classification of land cover from remotely sensed data, Remote Sensing of Environment, 61(3): 399–409.

Geisser S., 1993. Predictive Inference: An Introduction. New York: Chapman & Hall: 240pp.

Geurs K.T., and Van Wee B. 2004. Accessibility evaluation of land-use and transport strategies: review and research directions. Journal of Transport geography 12 (2):127-140.

Gilbert N., 2008. Agent-based models, University of Surrey, Guildford, UK, SAGE Publications: 112pp.

Grozdanić M., 2010. Prikaz metodologije planiranja u zaštićenim kulturno-istorijskim područjima na primeru starog jezgra Zemuna, Nasleđe, 11: 149-181.

Hagen A., 2002. Multi Method assessment of map similarity. Proceedings of the 5th AGILE Conference on Geographic Information Science, 25-27 AprilPalma, Spain: 171 –182.

Hagen A., 2003. Fuzzy set approach to assessing similarity of categorical maps. International Journal of Geographical Information Science, 17(3): 235-249.

Haines-Young R. and Petch J., 1986. Modelling. IN Haines-Young R. and Petch J., Physical Geography: Its Nature and Methods. London: Harper and Row: 144-157.

Hall M. and Smith L., 1999. Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper. Proceedings of the Twelfth FLAIRS conference, May 1-5, 1999, Orlando, Florida: 235-239.

Hall M., and Smith L., 1998. Practical feature subset selection for machine learning. In Proceedings of the 21st Australasian Computer Science Conference, Acsc'98. Singapore: Springer-Verlag Singapore Pte Ltd: 181-191.

Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P. and Witten I.H., 2009. The WEKA Data Mining Software: An Update; SIGKDD Explorations, 11(1):10-18.

Haykin S., 2009. Neural Networks And Learning Machines (3rd Edition). Prentice Hall, New York: 936pp.

Hu Z., and Lo C., 2007. Modeling urban growth in Atlanta using logistic regression. Computers, Environment and Urban Systems, 31(2): 667–688.

Huang B., Xie C. and Tay R., 2010. Support vector machines for urban growth modelling, Geoinformatica 14: 83–99.

Huang Z., Chen H., Hsu C.-J., Chen W.-H. and Wu S., 2004. Credit rating analysis with support vector machine and neural networks: A market comparative study. Decision Support Systems, 37: 543–558.

IT Glossary (https://www.gartner.com/it-glossary/predictive-modeling/)

Jain A.K., Jianchang M., and K. Moidin M., 1996. Artificial neural networks: A tutorial. IEEE computer 29 (3): 31-44.

Jones R., 2005. A Review of Land Use/Land Cover and Agricultural Change Models. Stratus Consulting Inc. for the California Energy Commission, PIER Energy-Related Environmental Research CEC-500-2005-056: 18pp.

Kantardzic M., 2011. Data mining: Concepts, models, methods, and algorithms, (Wiley - IEEE Press): 534 pp.

Khan J., Wei1 J.S., Ringnér M., Saal L.H., Ladanyi M., Westermann F., Berthold F., Schwab M., Antonescu C.R., Peterson C. and Meltzer P.S., 2001. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nature medicine, 7(6): 673-679.

Kim Y.S., Street W.N. and Menczer F., 2003. Feature selection in data mining. In: Wang, J. Data mining: opportunities and challenges. Idea Group Inc.: 80-105.

Knudby A., LeDrew E., and Brenning A., 2010. Predictive mapping of reef fish species richness, diversity and biomass in Zanzibar using IKONOS imagery and machine-learning techniques", Remote Sensing of Environment, 114(6): 1230-1241.

Kocabas V. and Dragicevi S., 2009. Agent-based model validation using Bayesian networks and vector spatial data. Environment and Planning B: Planning and Design, 36(5): 787-801.

Kocabas V. and Dragicevic S., 2006. Assessing cellular automata model behaviour using a sensitivity analysis approach. Computers, Environment and Urban Systems, 30(6): 921-953.

Kovačević M., Bajat B., Trivić B. and Pavlović R., 2009. Geological Units Classification of Multispectral Images by Using Support Vector Machines. Proceedings of International Conference on Intelligent Networking and Collaborative Systems INCoS 2009, 4-6 November, Barcelona, Spain: 267-272.

Lai T. and Dragićević S., 2011. Development of an urban landslide cellular automata model: a case study of North Vancouver, Canada. Earth Science Informatics, 4(2): 69-80.

Lakes T., John I., Müller D., Krüger C. and Rabe A., 2010. A support vector machine approach to model urban growth in the greater Tirana region, Albania. In The 13th AGILE International Conference on Geographic Information Science. Guimaraes, Portugal.

Lambin E.F. and Geist H., 2006. Land-Use and Land-Cover Change: Local processes and global impacts, Springer Berlin Heidelberg: 222pp.

Landis J.R. and Koch G. G., 1977. The measurement of observer agreement for categorical data. Biometrics, 33(1): 159-174.

Li X. and Yeh A.G.O., 2004. Data mining of cellular automata's transition rules, International Journal of Geographical Information Science, 18(8): 723-744.

Liu H. and Setiono R., 1995. Chi2: Feature selection and discretization of numeric attributes. In 2012 IEEE 24th International Conference on Tools with Artificial Intelligence: 388-391.

Liu W., Seto K., Sun Z. and Tian Y., 2007. Urban land use prediction model with spatiotemporal data mining and GIS. In Q. Weng and D. Quattrochi, Eds., Urban remote sensing. CRC Press, Taylor and Francis: 165-78.

López E., Bocco G., Mendoza M. and Duhau E., 2001. Predicting land-cover and land-use change in the urban fringe: A case in Morelia city, Mexico, Landscape and Urban Planning, 55 (4): 271–285.

Lwin K. and Murayama Y., 2009. A GIS Approach to Estimation of Building Population for Micro-spatial Analysis. Transactions in GIS 13(4): 401–414

Mahajan Y. and Venkatachalam P., 2009. Neural network based cellular automata model for dynamic spatial modeling in GIS, Springer, Berlin/Heidelberg: 341-352.

Makins M. (Ed.) 1995. Collins English Dictionary. 3rd ed. updated. Glasgow: Harper Collins (http://www.collinsdictionary.com/).

Marić I., Niković A. and Manić B., 2010. Transformation of the New Belgrade urban tissue: filling the space instead of interpolation, SPATIUM International Review, 22: 47-56.

Marjanovic M., Bajat B. and Kovacevic M., 2011. Landslide susceptibility assessment with machine learning algorithms, Engineering Geology, 123(3): 225–234.

Matthews R.B., Gilbert N.G., Roach A., Polhill J.G. and Gotts N.M., 2007. Agent-based land-use models: a review of applications. Landscape Ecology, 22(10): 1447-1459.

Meyer W. B. and Turner B.L., 1994. Changes in land use and land cover: a global perspective (Vol. 4). Cambridge University Press: 537pp.

Mitchell T., 1997. Machine Learning. McGraw-Hill, Now York: 432pp.

Muller M.R. and Middleton J., 1994. A Markov model of land-use change dynamics in the Niagara Region. Ontario, Canada, Landscape Ecology, 9(2): 151–157.

Nestorov I. and Protić D., 2009. CORINE kartiranje zemljisnog pokrivaca u Srbiji. Gradjevinska knjiga d.o.o.: 180pp.

Okwuashi O., McConchie J., Nwilo P., Isong M., Eyoh A., Nwanekezie O., Eyo E. and Danny Ekpo, A., 2012. Predicting future land use change using support vector machine based GIS cellular automata: a case of Lagos, Nigeria, Journal of Sustainable Development, 5(5): 132-139.

O'Sullivan D. and Haklay M., 2000. Agent-based models and individualism: is the world agent-based?. Environment and Planning A 32(8): 1409 – 1425.

Paliwal M. and Kumar U.A., 2009. Neural networks and statistical techniques: A review of applications, Expert Systems with Applications, 36(1): 2–17.

Petrić J., Maričić, T. and Basarić, J., 2012. The population conundrums and some implications for urban development in Serbia. Spatium 28: 7-14.

Pickett S.T.A., Cadenasso M.L., Grove J.M., Nikon C.H., Pouyat E.V., Zipperer W.C., and Constanza B., 2001. Urban ecological systems. Linking terrestrial ecological, physical, and socioeconomic components of metropolitan areas. Annual Review of Ecology and Systematics 32: 127-157.

Pijanowski B.C., Brown D. G, Manik G. and Shellito B., 2002. Using neural nets and GIS to forecast land use changes: a land transformation model, Computers, Environment and Urban Systems 26(6): 553-575.

Pontius Jr.R.G., 2000. Quantification error versus location error in comparison of categorical maps. Photogrammetric Engineering & Remote Sensing, 66: 1011-1016.

Quinlan J.R., 1986. Introduction to DecisionTrees. Machine Learning, 1: 81-106.

Quinlan J.R., 1993. C4.5: Programs for Machine Learning. Morgan Caufman, San Mateo, CA: 312pp.

Rosenblatt F., 1958. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Cornell Aeronautical Laboratory, Psychological Review, 65(6): 386–408.

Rumelhart D., Hinton G. and Williams R., 1986b. Learning representations by back-propagating errors. Nature 323 (6088): 533–536.

Rumelhart D., Hinton G. and Williams R., 1986a. Learning internal representations by error propagation. In D. E. Rumelhart, & J. L. McClelland (Eds.), Parallel distributed processing: explorations in the microstructures of cognition Cambridge: MIT Press, 1: 318–362.

Samardžić-Petrović M., Bajat B. and Kovačević M., 2013a. Assessing similarities between planned and observed land use maps: the Belgrade's municipalities case, Symposium GIS Ostrava 2013, CD Proceedings, 21th - 23th January Ostrava.

Samardžić-Petrović M., Bajat B. and Kovačević M., 2013b. The application of different kappa statistics indices in the assessment of similarity between planned and actual land use maps. 2nd International Scientific Conference RESPAG 2013, CD Proceedings, 22nd - 25th May, Belgrade, Serbia

Samuel A.L., 1959. Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development 3: 211–229.

Santé I., García A.M., Miranda D. and Crecente R., 2010. Cellular automata models for the simulation of real-world urban processes: A review and analysis, Landscape and Urban Planning, 96 (2): 108–122.

Schneider L.C. and Pontius Jr.R.G., 2001. Modeling land use change in the Ipswich watershed, Massachusetts, USA. Agriculture, Ecosystems and Environment, 85(1): 83–94.

Shannon C.E., 1948. A Mathematical Theory of Communication. Bell System Technical Journal 27 (3): 379–423.

Shen G., 2002. Fractal dimension and fractal growth of urbanized areas. International Journal of Geographical Information Science, 16(5): 419-437.

Simon P., 2013. Too Big to Ignore: The Business Case for Big Data. Wiley: 257 pp.

Solomatine D.P. and Ostfeld A., 2008. Data-driven modelling: some past experiences and new approaches, Journal of Hydroinformatics, 10(1): 3–22.

Solomatine D.P., 2002. Data-driven modelling: paradigm, methods, experiences, 5th International Conference on Hydroinformatics, Cardiff, UK: 1-5.

Sousa S., Caeiro S. and Painho M., 2002. Assessment of map similarity of categorical maps using Kappa Statistics: The Case of Sado Estuary. ESIG 2002, 13-15 November, Tagus Park, Oeiras, Portugal.

Stevens D. and Dragicevic S., 2007. A GIS-based irregular cellular automata model of land-use change. Environment and Planning B Planning and Design, 34(4): 708-724.

Sudhira H.S., 2004. Integration of Agent-based and Cellular Automata Models for Simulating Urban Sprawl. International Institute for Geo-information Science and Earth Observation: 78pp.

Talavera L., 2005. An evaluation of filter and wrapper methods for feature selection in categorical clustering. In Advances in Intelligent Data Analysis VI, Springer Berlin Heidelberg: 440-451.

Thekkudan T.F., 2008. Calibration of an Artificial Neural Network for Predicting Development in Montgomery County, Virginia: 1992-200, Published in 2008. Msc Thesis.

Theobald D.M. and Hobbs N.T., 1998. Forecasting rural land-use change: a comparison of regression and spatial-based models. Geographical and Environmental Modeling, 2: 65-82.

Tien Bui D., Pradhan B., Lofman O. and Revhaug I., 2012. Landslide Susceptibility Assessment in Vietnam Using Support Vector Machines, Decision Tree, and Naïve Bayes Models, Mathematical Problems in Engineering: 26 pp.

Tobler W.R., 1970. A Computer Movie Simulating Urban Growth in the Detroit Region, Geographical Analysis 46 (2): 234-240.

Tobler W.R., 1979a. Cellular Geography. Philosophy in Geography. Gale, S. and C. Olsson. Dordrecht,Kluwer: 379-386.

Tobler W.R., 1979b. Smooth pycnophylactic interpolation for geographical regions. Journal of the American Statistical Association, 74: 519–30.

Torrens P.M., 2006. 'Simulating Sprawl', Annals of the Association of American Geographers, 96(2): 248–275.

Triantakonstantis D.P., 2012. Urban Growth Prediction Modelling Using Fractals and Theory of Chaos, Open Journal of Civil Engineering, 2012, 2: 81-86.

Triantakonstantis D., Mountrakis G. and Wang J., 2011. A spatially heterogeneous expert based (SHEB) urban growth model using model regionalization, Journal of Geographic Information System, 3(3):195-210.

Turner B.L., Lambin E.F. and Reenberg A., 2007. The emergence of land change science for global environmental change and sustainability. Proceedings of the National Academy of Sciences. 104(52): 20666-20671.

Turner M.G., 1988. A spatial simulation model of LUCC in a piedmont county in Georgia. Applied Mathematics and Computation, 27: 39–51.

U.S. EPA, 2000. Projecting Land-Use Change: A Summary of Models for Assessing the Effects of Community Growth and Change on Land-Use Patterns. EPA/600/R00/098. U.S. Environmental Protection Agency, Office of Research and Development, Cincinnati, OH: 260 pp.

URBEL, 2003. Urban Planning Institute of Belgrade, The Master Plan of Belgrade 2021. Official Gazette of the City of Belgrade, 27.

URBEL Urban Planning Institute of Belgrade, http://www.urbel.com

van Vliet J., 2009. Assessing the Accuracy of Changes in Spatial Explicit Land Use Change Models, Proceedings of 12th AGILE International Conference on Geographic Information Science, 2-5 June Hannover, Germany

van Vliet J., Bregt A.K. and Hagen-Zanker A., 2011. Revisiting Kappa to account for change in the accuracy assessment of land-use change models, Ecological Modelling, 222: 1367–1375.

Vapnik V., 1995. The Nature of Statistical Learning Theory. Springer-Verlag, Now York: 314pp.

Veldkamp A. and Fresco L.O., 1996. CLUE: a conceptual model to study the conversion of land use and its effects. Ecological modelling, 85(2): 253-270.

Veldkamp A. and Lambin E.F., 2001. Predicting land-use change. Agriculture, ecosystems & environment, 85(1): 1-6.

Verburg P.H., Schot P.P., Dijst M.J. and Veldkamp A., 2004. Land use change modelling: current practice and research priorities, GeoJournal, 61(4): 309-324.

Visser H. and de Nijs T., 2006. The Map Comparison Kit. Environmental Modelling & Software, 21(3): 346–358.

Vojković G., Miletić R. and Miljanović D., 2010. Recent demographic-economic processes in the Belgrade agglomeration. Journal of the Geographical Institute "Jovan Cvijic" SASA 90(1): 215-235.

Vrzić Đ., 2010. Analysis of planning activities for Belgrade Fortress area – Consideration of the area in Master plans for Belgrade (in 1923, 1950, 1985 and 2003), Detailed urban plan and regulation plans, (http://www.betonhala.com/2011/BelgradeFortressPlans.pdf)

Wegener M., 1994. Operational urban models state of the art. Journal of the American Planning Association, 60(1): 17-29.

White R., and Engelen G., 1993. Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. Environment and Planning A, 25: 1175-1199.

White R., Engelen D. and Uljee I., 1997. The use of contrained cellular automata for high resolution modelling of urban land use dynamics, Environment and Planning B, 24 (3): 323–343.

Witten I. H., Eibe F. and Hall M. A., 2011. Data mining: practical machine learning tools and techniques, Elsevier: 662pp.

Wu F. and Yeh A. G., 1997. Changing spatial distribution and determinants of land development in Chinese cities in the transition from a centrally planned economy to a socialist market economy: a case study of Guangzhou. Urban Studies, 34: 1851-1879.

Xie Y.C., Batty M. and Zhao K., 2007. Simulating emergent urban form using agent-based modeling: Desakota in the Suzhou-Wuxian region in china. Annals of the Association of American Geographers, 97: 477–495.

Yang Q., Lia X. and Shid X., 2008. Cellular automata for simulating land use changes based on support vector machines, Computers & Geosciences, 34(6): 592–602.

Yeh A.G.O. and Li X., 2003. Simulation of development alternatives using neural networks, cellular automata, and GIS for urban planning, Photogrammetric Engineering and Remote Sensing, 2003, 69 (9): 1043-1052.

Zadeh L., 1965. Fuzzy sets. Information and Control, 8: 338–353.

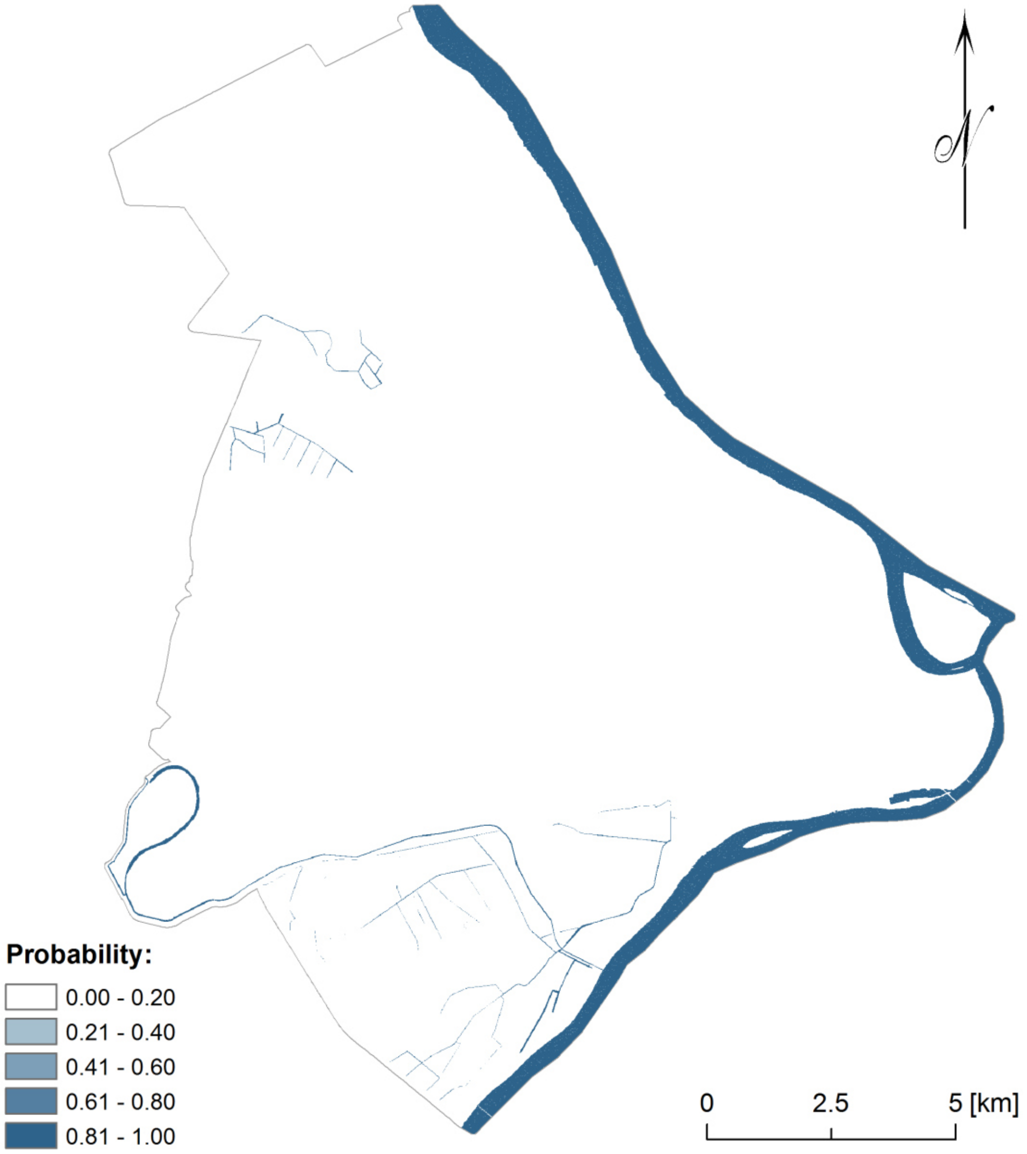
Zhao Q., Wu J., Yang G. and Chen J., 2011. The dynamic simulation and prediction of land use change based on GIS. In Remote Sensing, Environment and Transportation Engineering (RSETE), 24-06 June, 2011 International Conference on, Nanjing, China: 7942–7945.

Zhou Y., 2004. Data driven process monitoring based on neural network and classification trees, PhD Dissertation, Chemical Engineering Department, Texas A&M University, College Station, TX. PhD

Zhu Z., Yew-Soon O. and Manoranjan D., 2007. Wrapper–filter feature selection algorithm using a memetic framework. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions 37(1): 70-76.

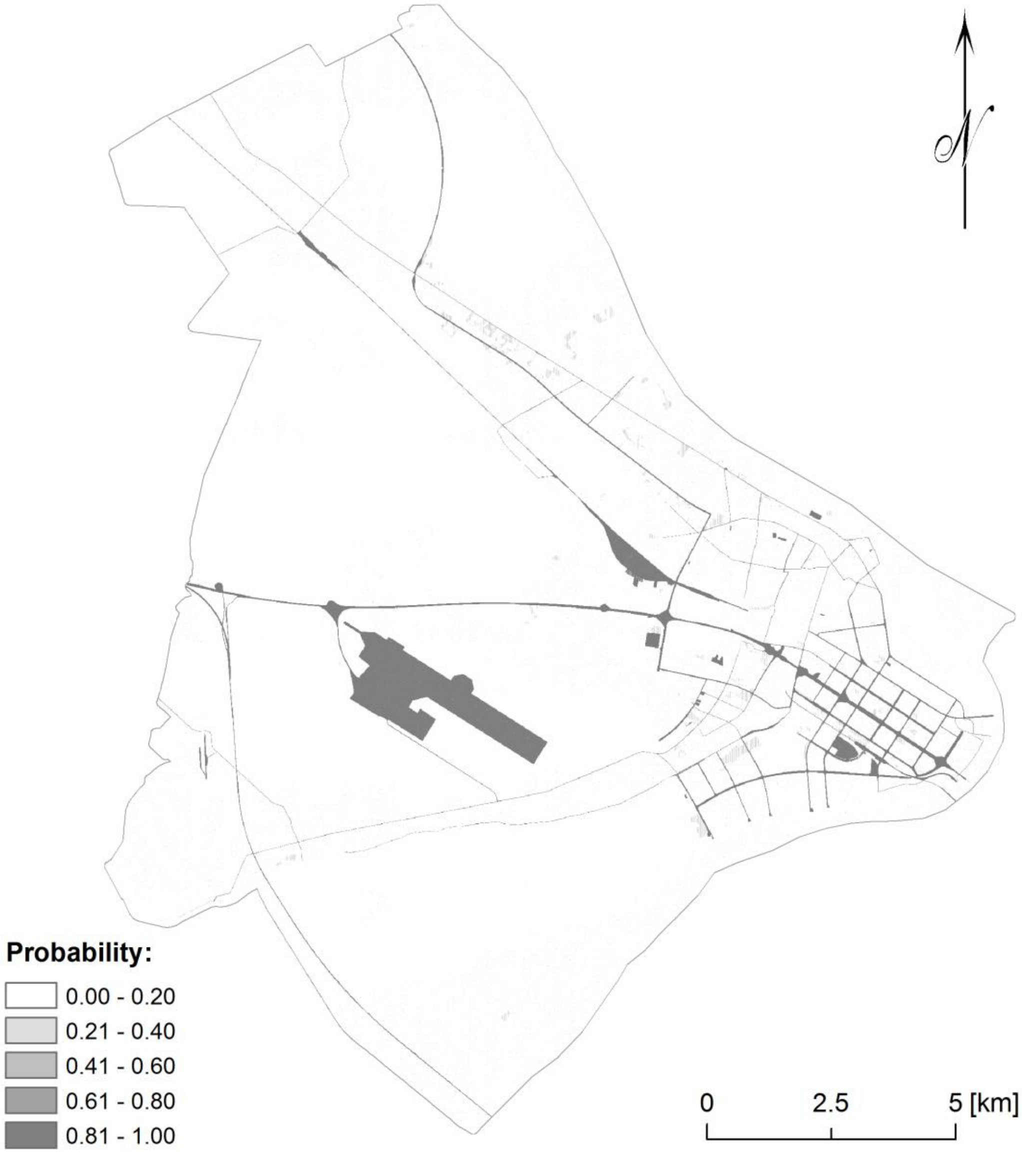
Appendix 1:

Map of probability of occurrence for *Wetland* class



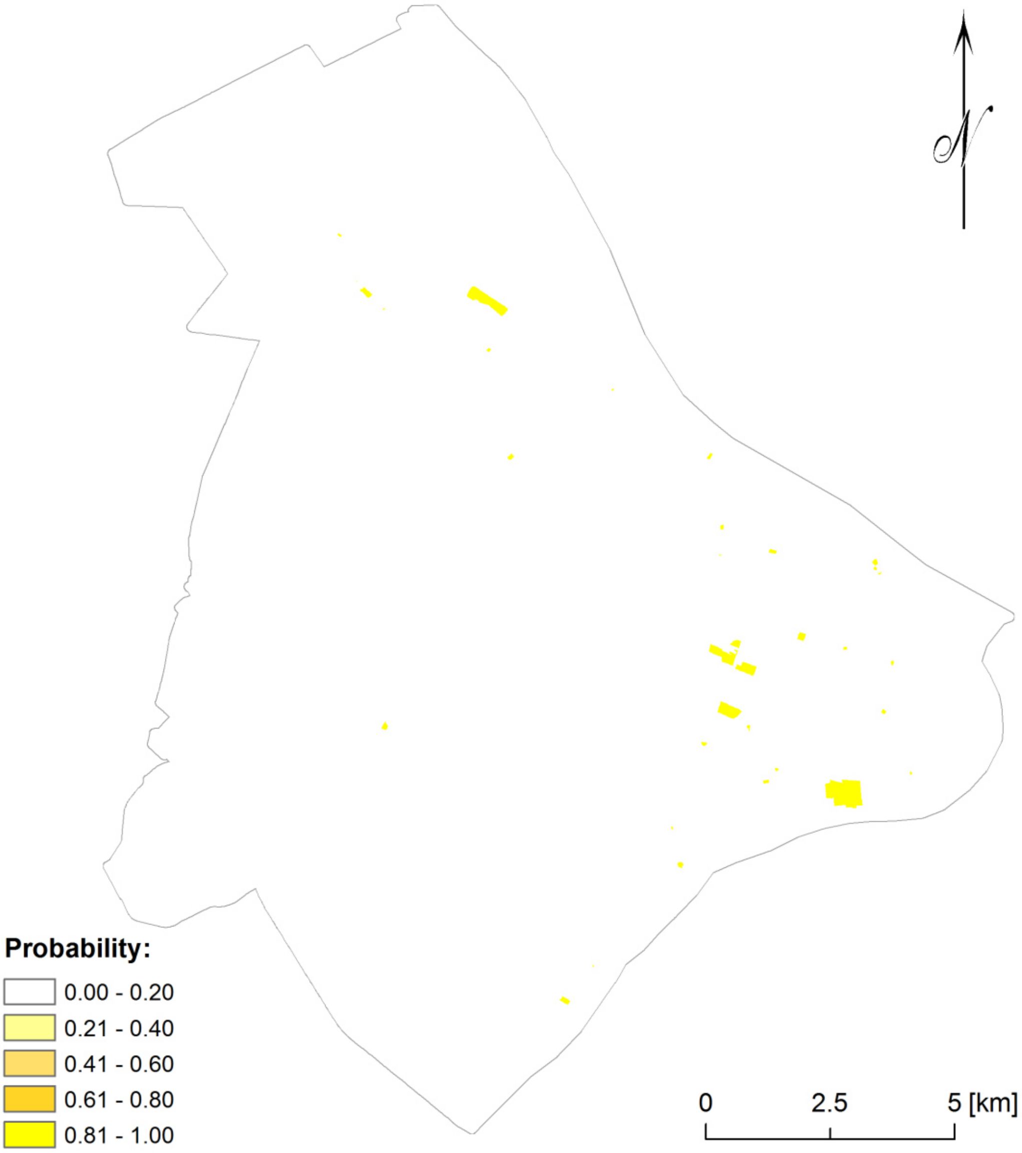
Appendix 2:

Map of probability of occurrence for *Transportation networks* class



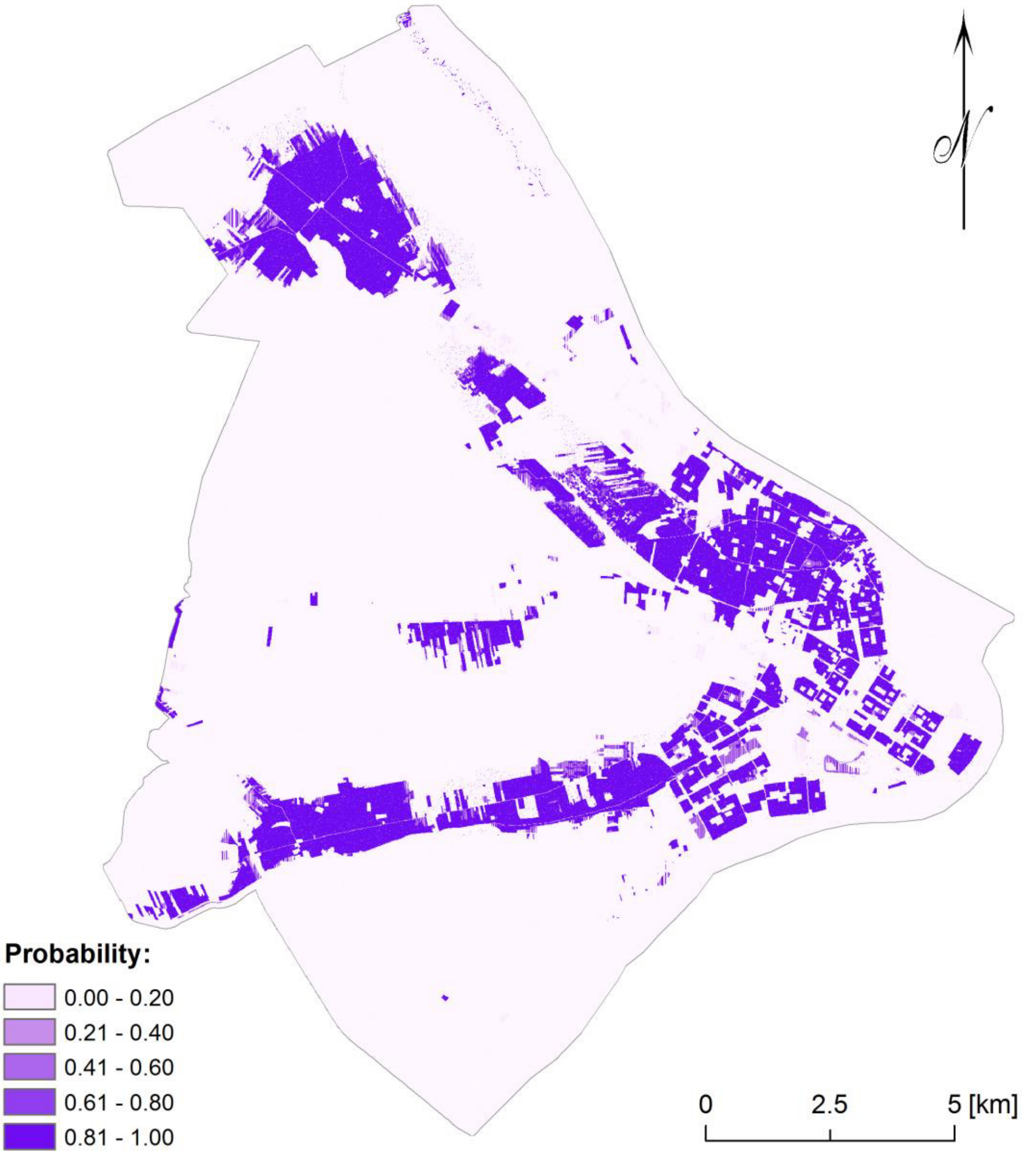
Appendix 3:

Map of probability of occurrence for *Infrastructure* class



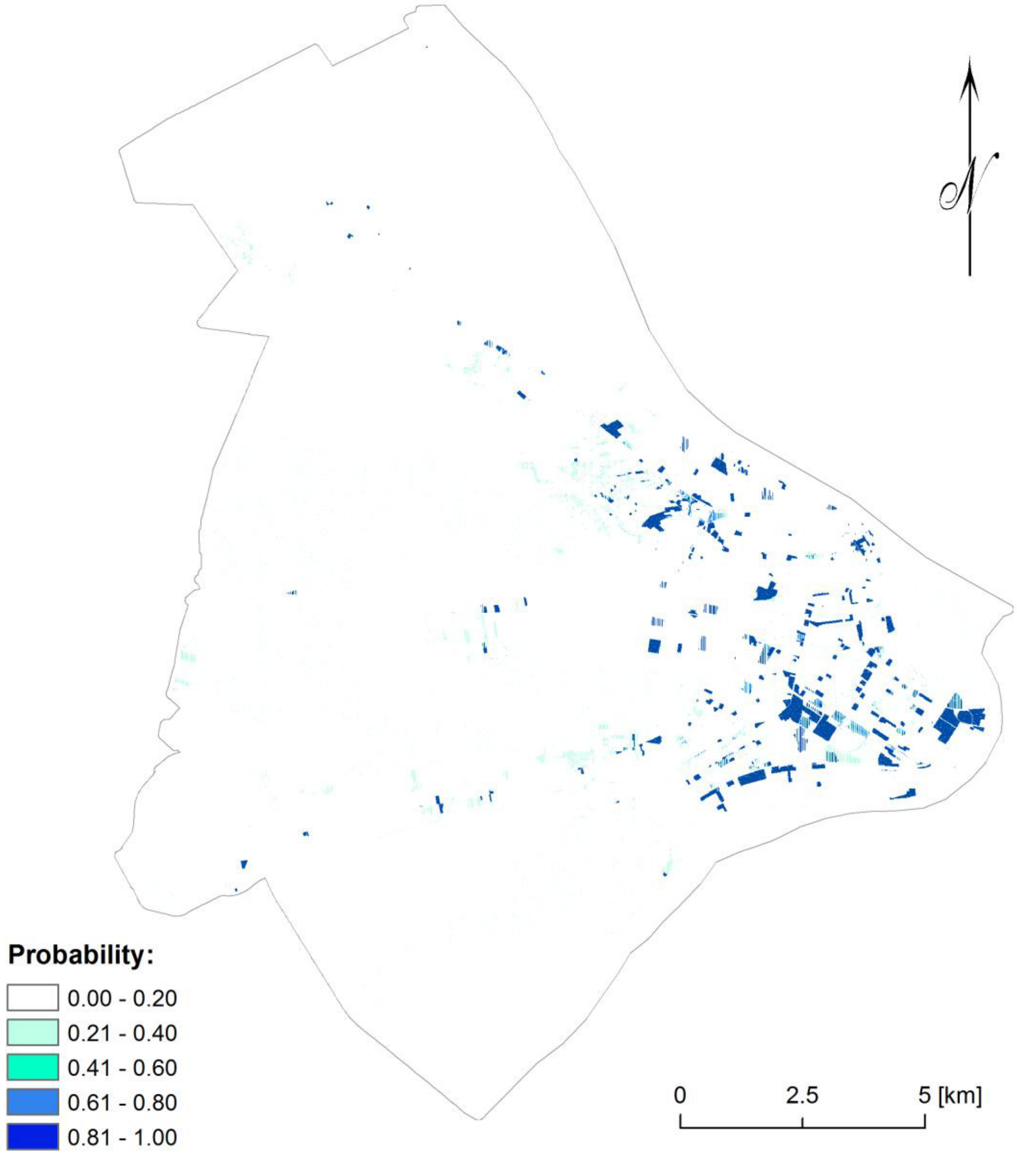
Appendix 4:

Map of probability of occurrence for *Residential* class



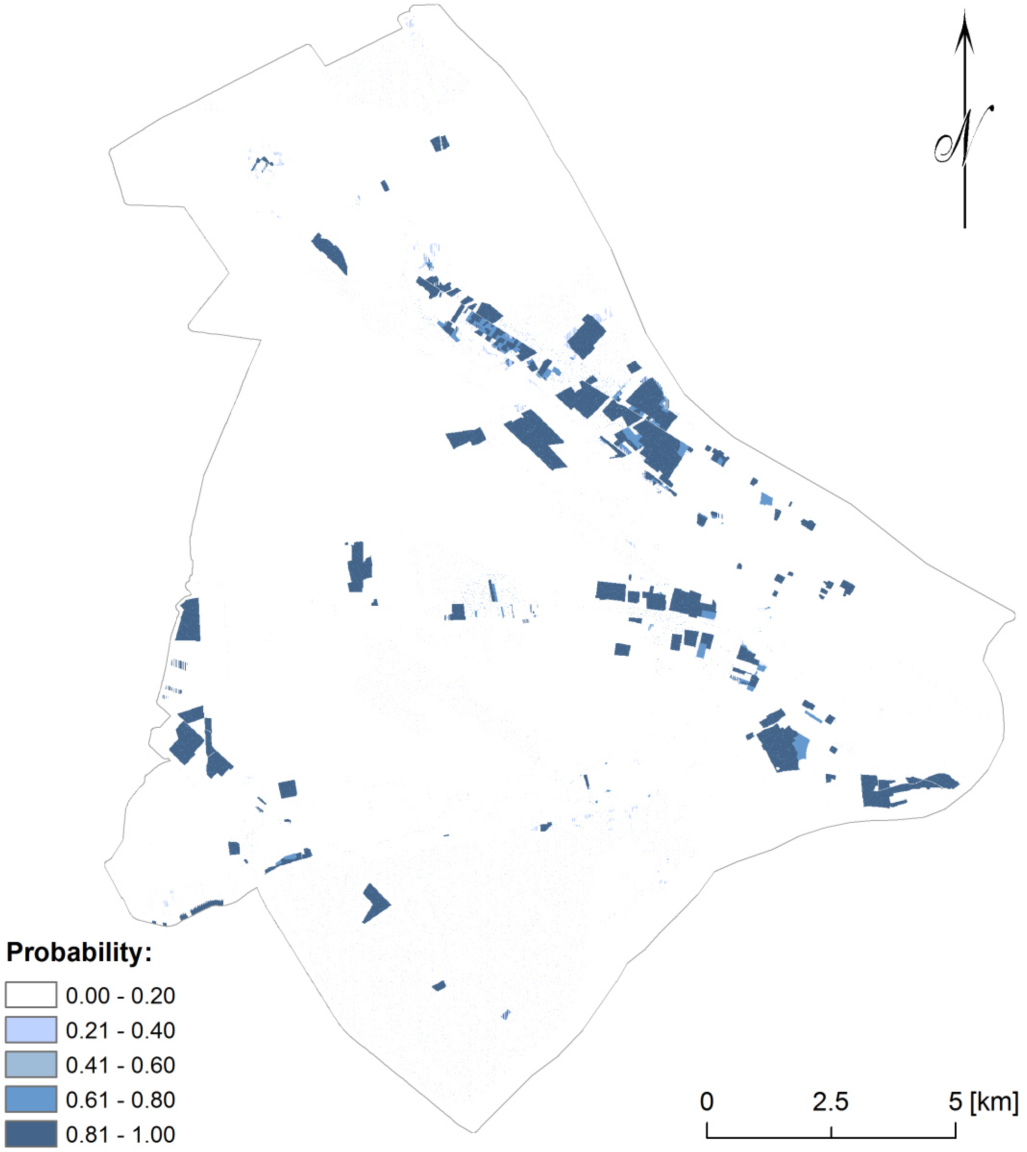
Appendix 5:

Map of probability of occurrence for *Commercial* class



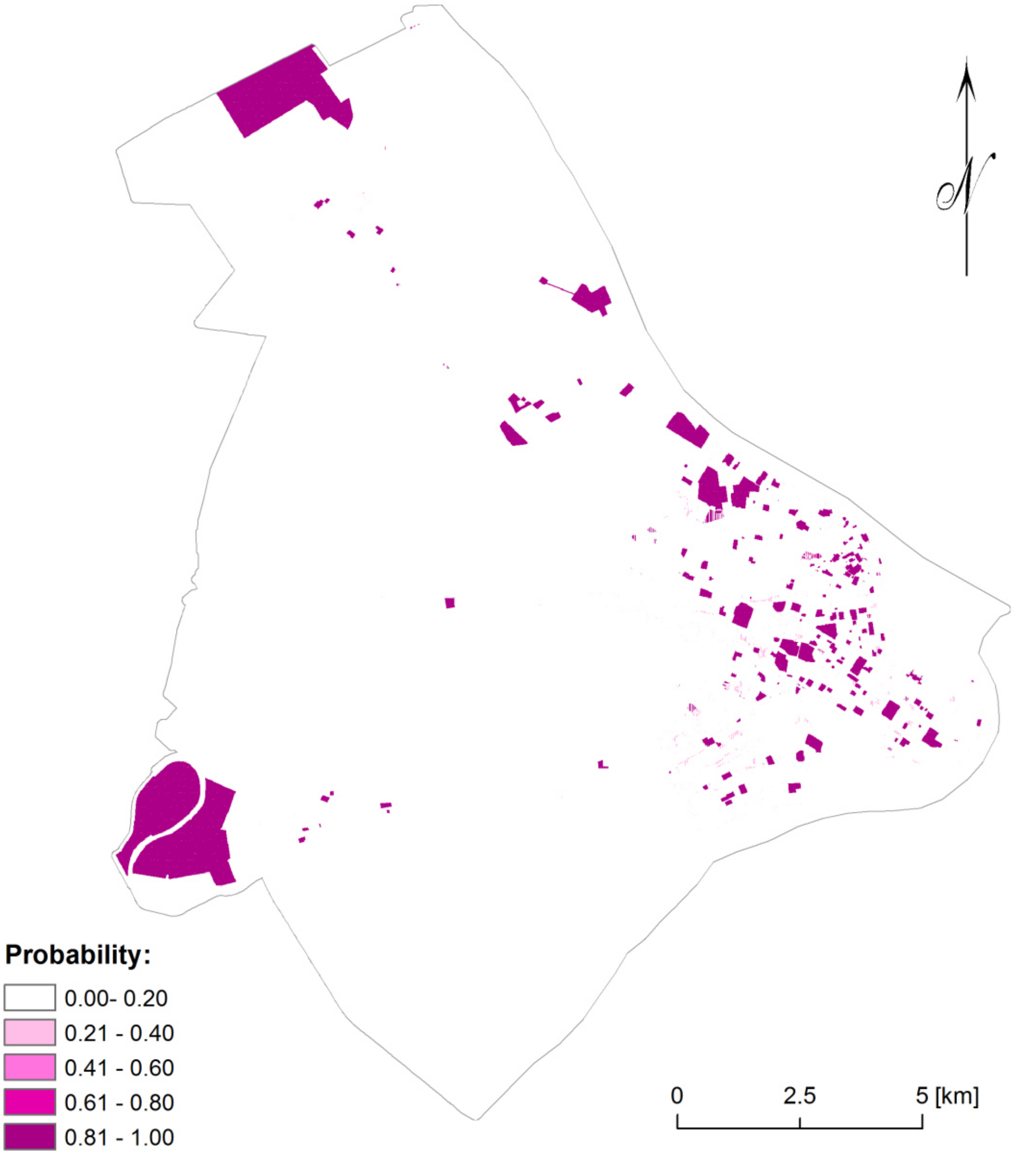
Appendix 6:

Map of probability of occurrence for *Industry* class



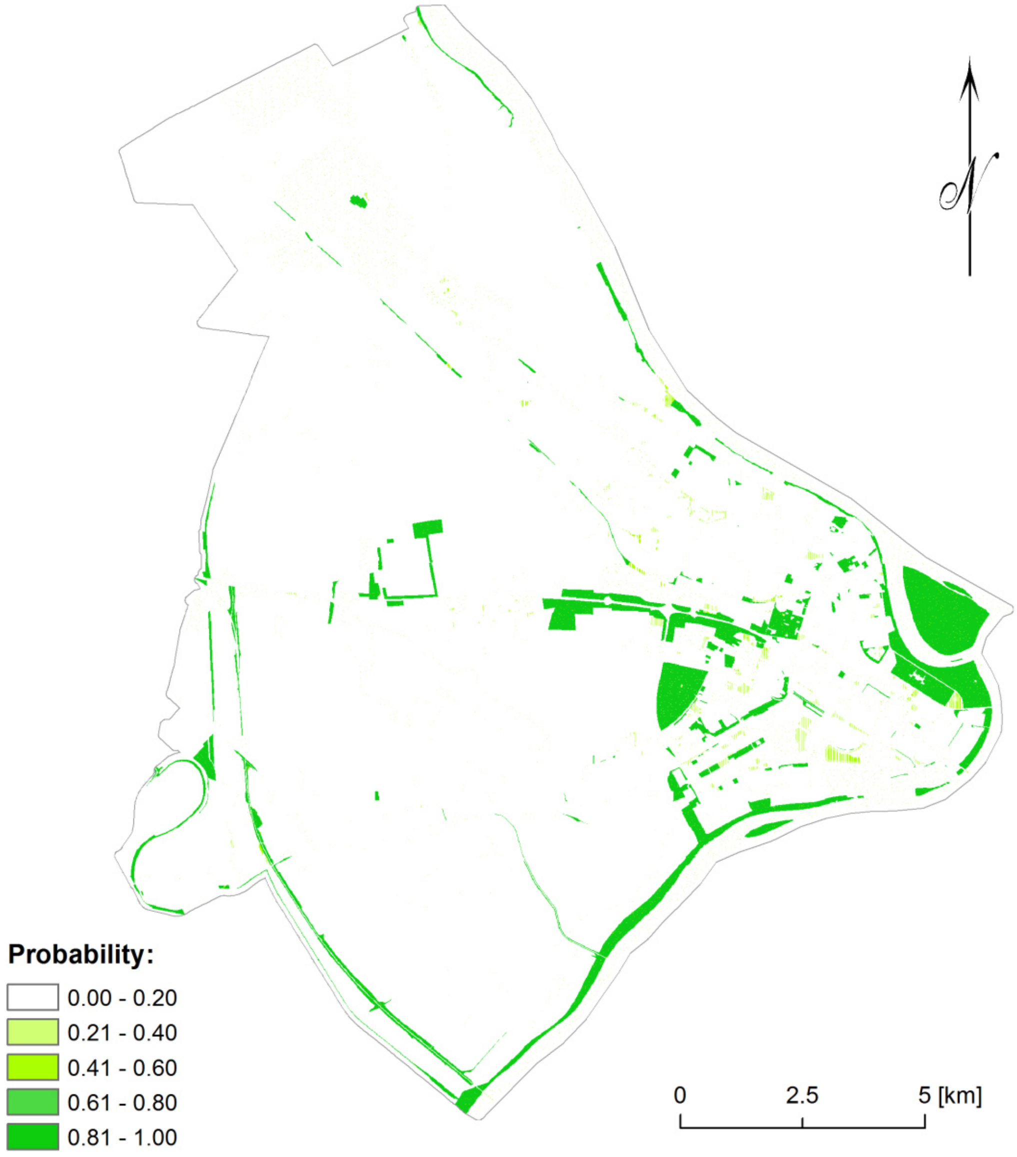
Appendix 7:

Map of probability of occurrence for *Special use* class



Appendix 8:

Map of probability of occurrence for *Green area* class



Biography

Mileva Samardžić-Petrović was born in Hamburg, Germany, on June the 3rd, 1980. She finished Elementary School "Milos Crnjanski" and Secondary Technical School "Jovan Vukanović", department of Geodesy in Novi Sad. She enrolled Faculty of Civil Engineering, Department of Geodesy, University of Belgrade in year 1999 and graduated in 2007 with the grade point average of 8,30 (eight and 30/100) on the topic “Geoid Determination for the territory of Republic of Serbia by GPS levelling method” (grade 10).

Mileva enrolled PhD studies at Faculty of Civil Engineering by the end of year 2007 and passed all required exams as of September 2010 with the grade point average of 9,88 (nine and 88/100). As a part of her PhD work and professional development, she attended Simon Fraser University, Canada, in 2013 as a visiting research fellow for a period of 5 months. During her PhD studies Mileva published, as author and co-author, papers related to spatial analysis and modelling: 1 journal paper (from SCI list), 5 papers in peer-reviewed journals, 3 chapters in monograph and 6 international conference papers.

For the Teaching Assistant position in the Field of Academic Expertise - Engineering Geodesy at Faculty of Civil Engineering, University of Belgrade, she was first elected on January 1 2008. Due to achieved results, by the decision of Electoral Council of Faculty of Civil Engineering she was re-elected to the same position by the end of November 2010. Mileva actively participates in preparation of lectures and lectures in laboratories for following courses: Geodesy, Geodesy in Transportation Engineering, Geoinformation Systems and Natural Resources. As a researcher, she participated in 3 scientific projects financed by Ministry of Science.

Прилог 1.

Изјава о ауторству

Потписани Милева Самарџић-Петровић

Број индекса 30/07

Изјављујем

да је докторска дисертација под насловом

„ПРЕДВИЂАЊЕ ПРОМЕНА У КОРИШЋЕЊУ ЗЕМЉИШТА ПРИМЕНОМ МОДЕЛА ВОЂЕНИХ ПОДАЦИМА (DATA-DRIVEN MODELS)“

- резултат сопственог истраживачког рада,
- да предложена дисертација у целини ни у деловима није била предложена за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио интелектуалну својину других лица.

Потпис докторанда

У Београду, 03.06.2014. године

Прилог 2.

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора Милева Самарџић-Петровић

Број индекса 30/07

Студијски програм Геодезија и геоинформатика

Наслов рада „ПРЕДВИЂАЊЕ ПРОМЕНА У КОРИШЋЕЊУ ЗЕМЉИШТА ПРИМЕНОМ МОДЕЛА ВОЂЕНИХ ПОДАЦИМА (DATA-DRIVEN MODELS)“

Ментор В.проф. др Бранислав Бајат, дипл. геод. инж

Потписани ______________________

Коментор В.проф. др Милош Ковачевић, дипл. електр. инж

Потписани ______________________

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла за објављивање на порталу **Дигиталног репозиторијума Универзитета у Београду.**

Дозвољавам да се објаве моји лични подаци везани за добијање академског звања доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис докторанда

У Београду, 03.06.2014. године

____

Прилог 3.

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

„ПРЕДВИЂАЊЕ ПРОМЕНА У КОРИШЋЕЊУ ЗЕМЉИШТА ПРИМЕНОМ МОДЕЛА ВОЂЕНИХ ПОДАЦИМА (DATA-DRIVEN MODELS)“

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигитални репозиторијум Универзитета у Београду могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство

2. Ауторство - некомерцијално

3. Ауторство – некомерцијално – без прераде

4. Ауторство – некомерцијално – делити под истим условима

5. Ауторство – без прераде

6. Ауторство – делити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, кратак опис лиценци дат је на полеђини листа).

Потпис докторанда

У Београду, 03.06.2014. године

1. Ауторство - Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.

2. Ауторство – некомерцијално. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.

3. Ауторство - некомерцијално – без прераде. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.

4. Ауторство - некомерцијално – делити под истим условима. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.

5. Ауторство – без прераде. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.

6. Ауторство - делити под истим условима. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.