

# LAND-COVER MODELLING USING CORINE LAND COVER DATA AND MULTI-LAYER PERCEPTRON

PIOTR DZIESZKO

Adam Mickiewicz University in Poznań, Poznań, Poland

Manuscript received: January 07, 2014

Revised version: February 21, 2014

DZIESZKO P., 2014. Land-cover modelling using CORINE Land Cover data and multi-layer perceptron. *Quaestiones Geographicae* 33(1), Bogucki Wydawnictwo Naukowe, Poznań, pp. 5–22, 9 figs, 7 tables. DOI 10.2478/quageo-2014-0004, ISSN 0137-477X

**ABSTRACT:** Last decades of research have revealed the environmental impacts of Land-Use/Cover Change (LUCC) throughout the globe. Human activities' impact is becoming more and more pronounced on the natural environment. The key activity in the LUCC projects has been to simulate the syntheses of knowledge of LUCC processes, and in particular to advance understanding of the causes of land-cover change. Still, there is a need of developing case studies regional models to understand LUCC change patterns. The aim of this work is to reveal and describe the main changes in LUCC patterns occurring in Poznań Lakeland Mesoregion according to CORINE Land Cover database. Change analysis was the basis for the identification of the main drivers in land cover changes in the study area. The dominant transitions that can be grouped and modelled separately were identified. Each submodel was combined with all submodels in the final change prediction process. Driver variables were used to model the historical change process. Transitions were modelled using multi-layer perceptron (MLP) method. Using the historical rates of change and the transition potential model scenario for year 2006 was predicted. Corine Land Cover 2006 database was used for model validation.

**KEY WORDS:** Land-Use/Land-Cover (LULC), Land-cover change, GIS, Spatial Model, Landscape, CORINE Land Cover

*Address of the corresponding author: Piotr Dzieszko, Adam Mickiewicz University, Dzięgielowa 27, 61-680 Poznań, Poland, e-mail: dzieszko@amu.edu.pl*

## Introduction

Last decades of research have revealed the environmental impacts of Land-Use/Cover Change (LUCC) throughout the globe, ranging from changes in atmospheric composition to the extensive modification of Earth's ecosystems (Foley *et al.* 2005). Analysing Land-Cover changes would not be possible without significant quality increase of remote sensing data which are one of the most common data source in GIS (Zwoliński 2012). Land-cover change is the result of complex interactions between human and environmental driving factors (Schaldach, Priess 2008, Szpikowski 2002, Macias, Dryjer 2010). The

impact of human activities is becoming more and more visible in the natural environment. One of the most important and obvious areas of concern of these activities is LUCC (Wang *et al.* 2012). But also responses of LUCC to natural processes are purpose of scientific research (Marfai 2011, Kolendowicz, Bednorz 2010). Sometimes LUCC patterns are considered as variables in spatial structure analysis algorithms (Mackiewicz *et al.* 1979). The key activity in the LUCC projects is to simulate the syntheses of knowledge of LUCC processes, and in particular to advance understanding of the causes of land-cover change. Such efforts normally follow one of two approaches: broad scale analyses and detailed case studies at

the local scale. First approach gives us an idea of feedback loops on a global scale but it cannot be used in modelling of regional case studies. Because regional scale modelling was one of the main tasks of this research, second approach was selected to perform the analysis.

There is no doubt that the modelling of Land-Use/Land-Cover (LULC) and environmental changes are the subject of increasing importance (Schweitzer *et al.* 2011). A broad range of models has been developed for this discipline (Haase, Schwarz 2009). There are two fundamental steps in every study of land change, i.e. detecting change in the landscape and describing that change to some set of causal factors. The last step is crucial for the research quality. In the past decade it appeared that the biggest challenge in LUCC modelling is to handle interactions at several spatial and temporal scales. Some of the recent publications have proved that it is possible to successfully face this challenge (Briassoulis 2000, Agarwal *et al.* 2001, Veldkamp, Lambin 2001, Parker *et al.* 2003, Verburg, Veldkamp 2005). But what is also worth noticing, understanding the causes of LUCC is dominated by simplifications. Global forces become the main determinants of land-cover change, as they amplify or attenuate local factors (Lambin *et al.* 2001). Understanding of how land-cover changes occur is critical since these anthropogenic processes can have broad impact on the environment (Pijanowski *et al.* 2002). Still, there is a need of developing regional models for case studies to understand LUCC changes patterns. And this is the reason for this research to be undertaken.

What is often unsaid, the lack of land-cover datasets is a huge problem, as they are an essential basis for any LUCC analysis. The causes and processes that create landscape structures are often poorly understood because of the format, availability and costs of data that represent these landscapes (Schmit *et al.* 2006). In Geographical Information Science it is tempting to consider maps as being absolute truth (Evans 1997). However, uncertainties and errors are intrinsic to spatial data (Burrough, McDonnel 1998).

Nowadays there are plenty of models and approaches in LUCC modelling. In the past decade researches revealed that many factors are re-

sponsible for changes in LUCC patterns. Among these factors there are biophysical, economic, social, cultural, political or institutional ones. It is impossible to use all of these approaches and to take into account every factor, but often more complex and interdisciplinary models can have more predicting power.

Modelling involves the use of artificial representations of the interactions within the land-use system to explore its dynamics and possible future development. Modelling approach is also beneficial because it reveals all the gaps in knowledge at the stage of model-building. Real-life experiments in land-cover systems are difficult to carry out, if not impossible, but computer models can be considered as laboratories in which land-cover patterns can be observed and tested.

Artificial Neural Networks (ANNs) are powerful tools that use a machine learning approach to quantify and model complex behaviour and patterns. ANNs are used for pattern recognition in a variety of disciplines, such as image analysis (Fukushima *et al.* 1983), landscape classification (Brown *et al.* 1998), pattern classification (Ritter *et al.* 1988), climate forecasting (Drummond *et al.* 1998), and remote sensing (Atkinson, Tatnall, 1997). The use of neural networks has increased substantially over the last several years because of the advances in computing performance (Skapura 1996) and the increased availability of powerful and flexible ANN software.

The aim of this work is to reveal and describe the main changes in LUCC patterns according to CORINE (COoRdination of INformation on the Environment) Land Cover database (European Commission 1993) occurring in Poznań Lakeland Mesoregion. Changes analysis was the basis for the identification of the main drivers in land-cover changes in the study area. The dominant transitions that can be grouped and modelled separately were identified. Each submodel was combined with all submodels in the final change prediction process. Transitions were modelled using multi-layer perceptron (MLP). Using the historical rates of change and the transition potential model the future scenario for the year 2006 was predicted. CORINE Land Cover 2006 database was used for model validation. This work presents changes in land-cover between 1990 and 2000 as well as modelling outputs for the study area.

## Material and methods

### The case study area

The work presented here is based on the Poznań Lake District Mesoregion case study area located in the central-west Poland. It is mainly located within Wielkopolskie Province and partly within Lubuskie Province. Case study area is located in the west of Poznań and it is closed by  $15^{\circ}24.7' - 16^{\circ}38.8' E$  and  $52^{\circ}0.5' - 52^{\circ}43.0' N$  coordinates. The study region is characterised by heterogeneous landscape with a predominance of mixed agriculture. It covers an area of 4,158 square kilometres. More than half of this area was covered by arable lands and more than a quarter of this area was covered by forests in 2006. This region was chosen for two main reasons, i.e. covering by the pan-European land-cover raster data set (Corine) and the availability of additional spatial information for this region.

Spatial differentiation of Poland after the transition in the economical aspect is well identified and featured (Szlachta 1993, Czyż 2000). There are some works presenting changes in land use or land cover of such an area of interest (Łowicki 2008, Ciołkosz, Bielecka 2005, Łowicki, Mizgajski

2005) but there are no such a work focusing on modelling these changes in this area.

Figure 1 presents the location of the case study area in the background of Poland. Figure 2 shows the shape of the study area. Satellite images were used to present the study area as well as digital elevation model with hill-shade effect.

### Data

Land-cover data is stored in different types and formats with varying thematic and special content and accuracy. Linking the scale of analysis with proper data accuracy is still very challenging task. The scale in modelling is important because of two reasons: the area being covered by the model and the level of detail, inherent in the model (Parker *et al.* 2002). In this study the CORINE Land Cover (European Commission 1993) database has been used. It provides a pan-European inventory of biophysical land-cover. It uses 44 class nomenclatures on the third detail level and is a key database for integrated environmental assessment. Database was transposed to level 2 which contains 15 class nomenclatures. The first level (five classes) corresponds to the main categories of the land cover/land use (artificial areas,

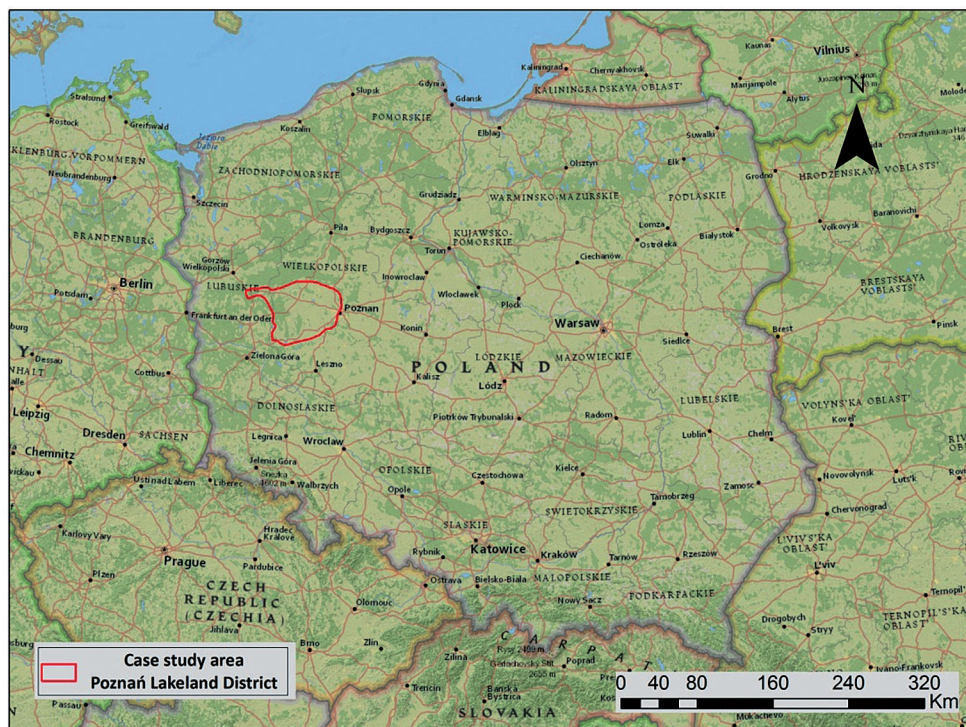


Fig. 1. Location of Poznań Lakeland District

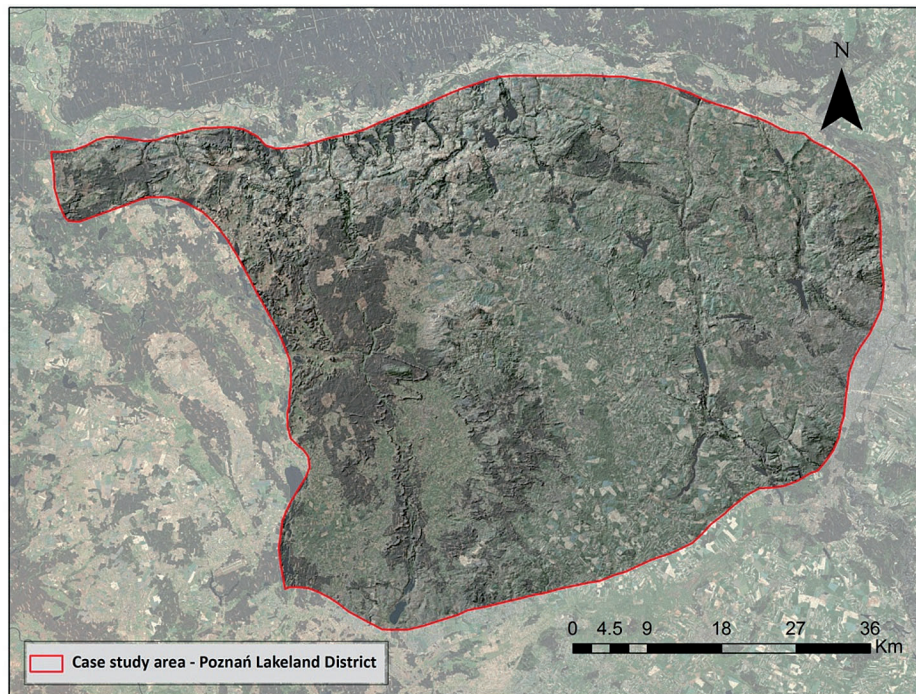


Fig. 2. Poznań Lakeland District. Satellite orthophotomap drapped on the hillshaded DEM

agricultural land, forests and semi-natural areas, wetlands, water surfaces). The second level (15 classes) covers physical and physiognomic entities at a higher level of detail (urban zones, forests, lakes, etc.); finally level 3 is composed of 44 classes (Willems *et al.* 2000).

Database is available with 100 m cellsize and the original data was in the form of a vector format at a scale of 1:100 000. The dataset is available from European Environment Agency (EEA 2013) website. CORINE was created from remotely sensed data. The datasets used in this work represent the land-cover in 1990, 2000 and 2006. A new version of CORINE for year 2012 will be available soon.

CORINE Land Cover provides information about land-cover changes for a substantial part of Europe. Availability of this data can contribute to new approaches of the assessment of the European landscape especially in modelling of its properties (Feranec *et al.* 2010).

CORINE Land Cover database has been validated. The official classification accuracy of CORINE was said to be 87%. It is a satisfying result but Schmit *et al.* (2006) say that the comparison of CORINE with the Integrated Administration and Control System (IACS) data revealed lower accuracy than the official validation. How-

ever, there is no other accurate data available for this study area and the data used for the analysis is considered appropriate for the spatial and temporal scale of the model.

Further data used in this analysis includes digital elevation model (DEM) at 30 m × 30 m spatial resolution and road and river network.

## Methods

CORINE Land Cover data for years 1990, 2000 and 2006 were downloaded from EEA website. Then, they were reclassified from the third detail level to the second detail level. 44 land cover classes were aggregated to 15 classes according to CORINE Land Cover legend (European Commission 1993). In the study area the following land-cover classes were identified: Urban fabric; Industrial, commercial and transport units; Mine, dump and construction sites; Artificial, non-agricultural vegetated areas; Arable land; Permanent crops; Pastures; Heterogeneous agricultural areas; Forests; Scrub and/or herbaceous vegetation; Open spaces with little or no vegetation; Inland wetlands; Inland waters.

Datasets were clipped to the extent of the case study area. Additional datasets were also collected i.e. DEM, road and river network, lakes and

topographic maps scans. Data were stored in ESRI Geodatabase format.

Next, datasets were imported to Idrisi Selva software and used for the analysis of the past land-cover changes. The main land-cover transitions in years 1990–2000 were identified. Driver variables for these transitions were developed and combined into final change prediction process.

Transition potential of land was identified and modelled with Land Change Modeller. Then historical rates of change and transition potential model were used to predict future scenario. 1990 CORINE Land Cover database was used as the earlier land-cover image and 2000 CORINE Land Cover database was used as the later land-cover image. MLP neural networks method was used to predict land cover in 2006. 2006 CORINE Land Cover database was used for model calibration and validation.

Land change prediction in Idrisi Selva Land Change Modeller is an empirically driven process that moves in a stepwise fashion from 1) Change Analysis through 2) Transition Potential Modelling to 3) Change Prediction. It is based on the historical change from time 1 to time 2 land cover maps to project future scenarios (Eastman 2012).

The multi-layer perceptron (MLP) neural net described by Rumelhart *et al.* (1986) is one of the most widely used ANNs. The MLP consists of three layers: input, hidden, and output (Fig. 1B) and thus can identify relationships that are non-linear in nature. ANN algorithms calculate weights for input values, input layer nodes, hidden layer nodes and output layer nodes by introducing the input in a feed forward manner, which propagates through the hidden layer and the output layer.

### Change Analysis

In the Change Analysis step, change is assessed between time 1 and time 2, between the two land-cover maps. The changes that are identified are transitions from one land cover state to another. It is likely that with many land-cover classes the potential combination of transitions can be complex. The main task is to identify the dominant transitions that can be grouped and modelled within submodels. Each group, or submodel of transitions, can be modelled separately, but ultimately each submodel is com-

bined with all submodels in the final change prediction process.

### Transition Potential Modelling

The second step in the Change Prediction process is Transition Potential Modelling, where potential of land to transition is identified. At this stage transition potential maps are created.

### Change Prediction

The third and the final step in Change Analysis is Change Prediction. Using the historical rates of change and the transition potential model, future scenario for a specified date is predicted.

The flow chart of this analysis is presented in Figure 3 (Eastman 2012). Two land cover maps, the earlier and the later one, and driver variables are the inputs. In Change Analysis step Change Map is determined. On the basis of driver variables transitions to the model are selected. After running the transition model, calculated transition potentials as well as the Demand for Land are used for land-cover projection and for calculation of the finally projected land-cover map.

### Multi-Layer Perceptron

The model uses MLP technique. Originally MLP was implemented to undertake the classification of remotely sensed imagery using the back propagation algorithm. MLP starts training on the samples it has been provided of grids that have and have not experienced the transition being modelled. Because this method relies on neural networks, firstly MLP is operating in automatic mode whereby it makes its own decisions about what parameters to use and how they should be changed to model the data in a better way. The module is iterating until the maximum iterations number or maximum accuracy rate (Eastman 2012).

## Results

### The change analysis: 1990–2000

The case study area covers 4158 km<sup>2</sup>. Changes between 1990 and 2000 in level 2 classes according to CORINE Land Cover equalled only 20.78 km<sup>2</sup>. It means that changes in this time interval

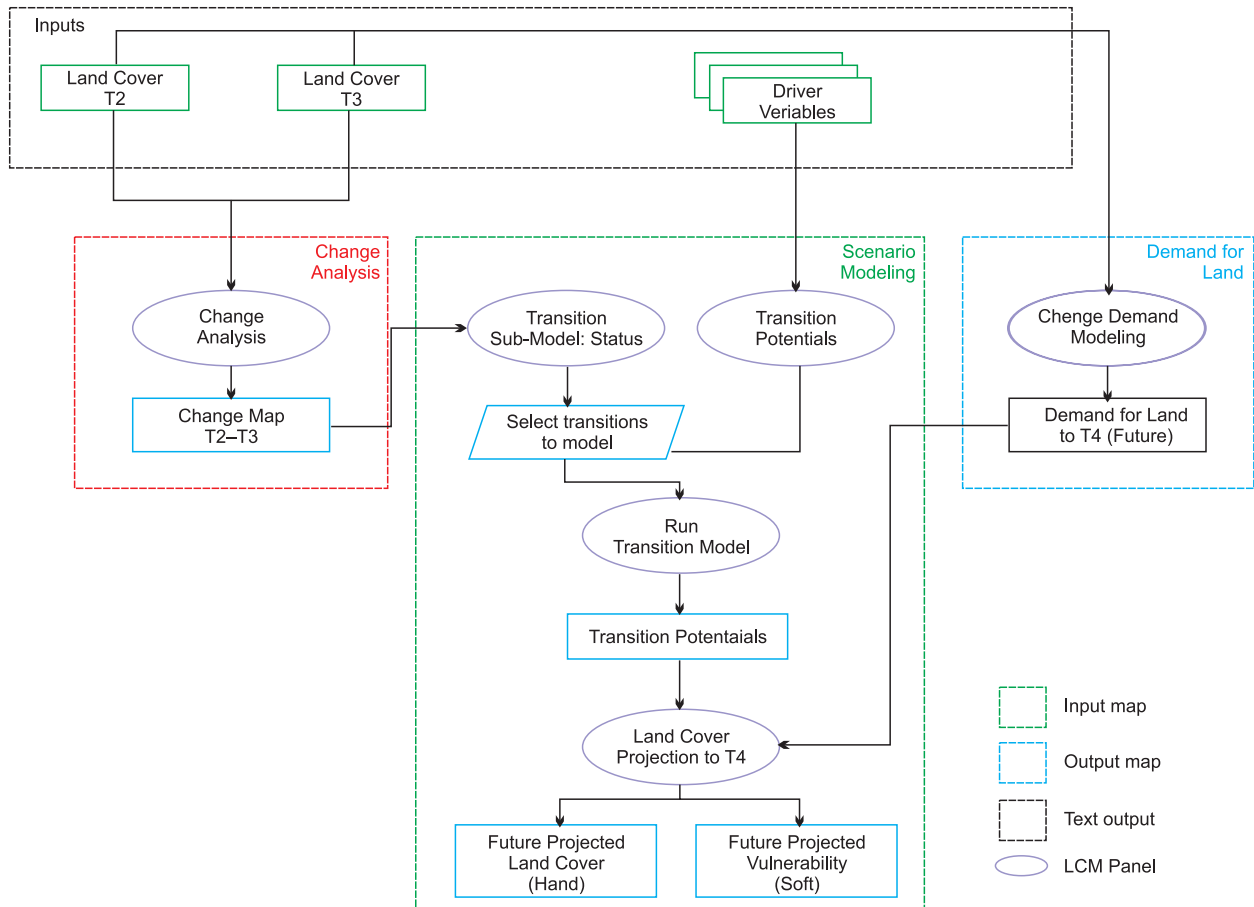


Fig. 3. Land Cover Prediction Flow Chart (Eastman 2012)

amounted 0.5% of the case study area. It shows that in the last decade of twentieth century these changes were not severe in the case study area. However, such changes appeared in definite classes so that some conclusions can be made on the basis of these data.

The amount of change in every class is presented in Table 1. It is easy to observe that arable land is dominating class in land cover structure. It covers almost 60% of the study area. What is also worth noticing is that forests cover more than 25% of the study area. So what is important for this case study area is that almost 85% of the land is arable land and forests. The other 11 land cover classes of the study area are only 15% of the area. This is the first sign that conversions in these two classes will be the main drivers of changes in Poznań Lakeland District.

According to Figure 4 and Table 2 we can see that the biggest change in the last decade of twentieth century took place in the class of permanent crops. This class lost more than 7 km<sup>2</sup>. What is

also visible in Figure 4 is that open spaces with little or no vegetation lost large amount of area (2.4 km<sup>2</sup>). The last class that lost the area between 1990 and 2000 is pastures. It lost less than quarter of square kilometre so it is not a significant change.

The three land cover classes increased their areas significantly. These are urban fabric (3.44 km<sup>2</sup>), forests (3 km<sup>2</sup>) and scrub and/or herbaceous vegetation (2.76 km<sup>2</sup>). Other classes also increased their area but the change was not significant. Industrial, commercial and transport units, mine, dump and construction sites, arable land and heterogeneous agricultural areas increased their area but these values equalled less than 0.5 km<sup>2</sup>.

Changes between CORINE Land Cover classes were obtained for study area in years 1990–2000 giving gains and losses of these classes. Changes provide information about an increase and decrease of the classes' areas but do not provide information about their dynamics. Gains and losses, however, do. Gains and losses of the classes in

Table 1. Types of land use and land cover for Poznań Lakeland District in years 1990, 2000 and 2006 (acc. CORINE Land Cover database, EEA 2013)

Level 1 classes	Code	Level 2 classes	Area (sq km)			Percent of area		
			1990	2000	2006	1990	2000	2006
1. Artificial surfaces	1.1	Urban fabric	110.85	114.29	125.82	2.67	2.75	3.03
	1.2	Industrial, commercial and transport units	15.92	16.33	25.72	0.38	0.39	0.62
	1.3	Mine, dump and construction sites	3.10	3.50	4.07	0.07	0.08	0.10
	1.4	Artificial, non-agricultural vegetated areas	15.41	15.41	21.79	0.37	0.37	0.52
2. Agricultural areas	2.1	Arable land	2451.64	2451.87	2426.22	58.96	58.96	58.35
	2.2	Permanent crops	13.79	6.44	5.24	0.33	0.15	0.13
	2.3	Pastures	169.59	169.09	165.1	4.08	4.07	3.97
	2.4	Heterogeneous agricultural areas	184.53	184.68	187.53	4.44	4.44	4.51
3. Forest and semi natural areas	3.1	Forests	1076.14	1079.14	1084.69	25.88	25.95	26.09
	3.2	Scrub and/or herbaceous vegetation associations	39.95	42.71	36.26	0.96	1.03	0.87
	3.3	Open spaces with little or no vegetation	3.78	1.38	0.99	0.09	0.03	0.02
4. Wetlands	4.1	Inland wetlands	8.27	8.27	8.21	0.20	0.20	0.20
	4.2	Maritime wetlands	-	-	-			
5. Water bodies	5.1	Inland waters	65.24	65.1	66.57	1.57	1.57	1.60
	5.2	Marine waters	--	-	-			
		sum	4158.21	4158.21	4158.21	100	100	100

CORINE Land Cover database are presented in Figure 5 and Table 2.

Comparing Figure 4 and Figure 5 we can see that the most dynamic class in case study area is arable land. Adding up gains and losses for this class, the difference was smaller than a quarter of 1 km<sup>2</sup>. But in fact arable land gained 8.26 km<sup>2</sup> and lost 8.03 km<sup>2</sup>. It means that the area change in this class is small but this class was really dynamic in the last decade of the twentieth century.

The analysis of arable land changes was performed and contributions to net change were calculated for this class. The results of these calculations are presented in Figure 6 and Table 3. What is worth noticing is that almost all area of arable land was gained at the expense of permanent crops class. The losses of arable land were spread into 6 classes: forests (1.9 km<sup>2</sup>); mine, dump and construction sites (1.98 km<sup>2</sup>); heterogeneous agricultural areas (1.61 km<sup>2</sup>); urban fabric (0.86 km<sup>2</sup>); scrub and/or herbaceous vegetation (0.69 km<sup>2</sup>) and industrial, commercial and transport units (0.24 km<sup>2</sup>). It is important information during

cultural areas (1.61 km<sup>2</sup>); urban fabric (0.86 km<sup>2</sup>); scrub and/or herbaceous vegetation (0.69 km<sup>2</sup>) and industrial, commercial and transport units (0.24 km<sup>2</sup>). It is important information during

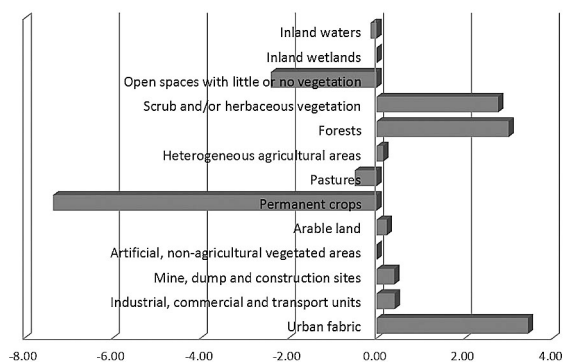


Fig. 4. Changes of LULC level 2 [km<sup>2</sup>] for Poznań Lakeland District between 1990 and 2000 (acc. CORINE Land Cover database, EEA 2013)

Table 2. Dynamics of land use/cover change for Poznań Lakeland District between 1990 and 2000 (acc. CORINE Land Cover database, EEA 2013)

Class	Changes	Losses	Gains
	[km <sup>2</sup> ]		
Urban fabric	3.44	0.00	3.44
Industrial, commercial and transport units	0.41	0.00	0.41
Mine, dump and construction sites	0.40	-1.96	2.36
Artificial, non-agricultural vegetated areas	0.00	0.00	0.00
Arable land	0.23	-8.03	8.26
Permanent crops	-7.35	-7.61	0.26
Pastures	-0.50	-1.16	0.66
Heterogeneous agricultural areas	0.15	-1.46	1.61
Forests	3.00	-3.34	6.34
Scrub and/or herbaceous vegetation	2.76	-3.65	6.41
Open spaces with little or no vegetation	-2.40	-2.40	0.00
Inland wetlands	0.00	0.00	0.00
Inland waters	-0.14	-0.26	0.12

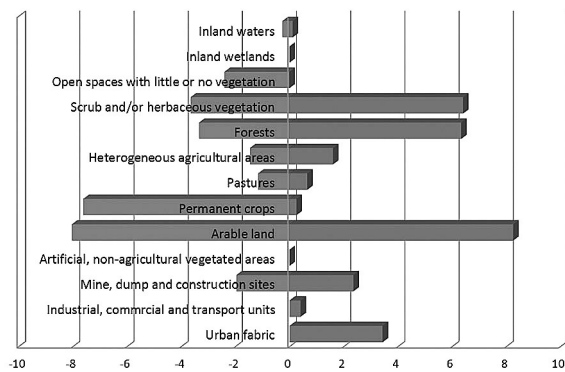


Fig. 5. Gains and losses of LULC level 2 [km<sup>2</sup>] for Poznań Lakeland District between 1990 and 2000 (acc. CORINE Land Cover database, EEA 2013)

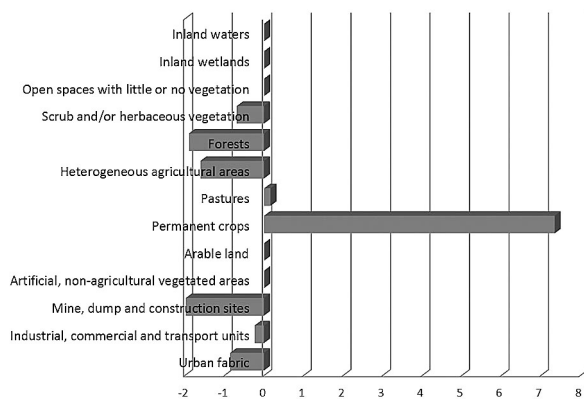


Fig. 6. Contributions to net change in arable land level 2 [km<sup>2</sup>] for Poznań Lakeland District between 1990 and 2000 (acc. CORINE Land Cover database, EEA 2013)

variables identification for modelling purpose. The conclusion is that only one class was converted to arable land (permanent crops) while 6 classes took part in the interception of arable land's area. Permanent crops lost 7.61 km<sup>2</sup> of the area. They gained only 0.26 km<sup>2</sup>. We can only observe that they lost their area and only at the expense of arable land.

Because arable land and forests are the key classes for the case study area, contributions to net change in forests were also calculated and they are presented in Figure 7 and Table 3. Figure 7 clearly shows that forests mainly gained area in the last decade of the twentieth century. Classes transformed to forests in this time are Arable land, Het-

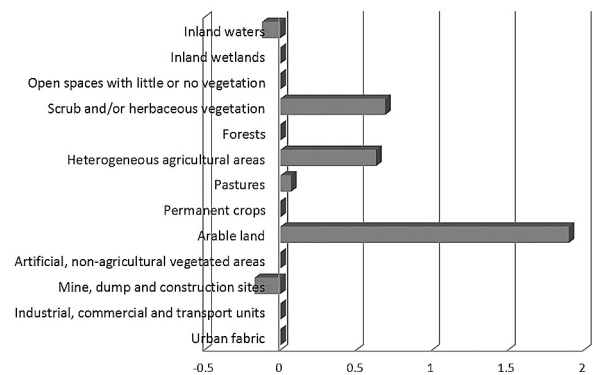


Fig. 7. Contributions to net change in forests level 2 [km<sup>2</sup>] for Poznań Lakeland District between 1990 and 2000 (acc. CORINE Land Cover database, EEA 2013)



Table 3. Contributions to net change in arable land and forests level 2 [km<sup>2</sup>] for Poznań Lakeland District between 1990 and 2000 (acc. CORINE Land Cover database, EEA 2013)

Class	Contributions to net change in Arable land	Contributions to net change in Forests
Urban fabric	-0.86	0
Industrial, commercial and transport units	-0.24	0
Mine, dump and construction sites	-1.98	-0.17
Artificial, non-agricultural vegetated areas	0.00	0
Arable land	0.00	1.9
Permanent crops	7.35	0
Pastures	0.16	0.07
Heterogeneous agricultural areas	-1.61	0.63
Forests	-1.90	0
Scrub and/or herbaceous vegetation	-0.69	0.69
Open spaces with little or no vegetation	0.00	0
Inland wetlands	0.00	0
Inland waters	0.00	-0.12

erogeneous agricultural areas and Scrub and/or herbaceous vegetation. Forests lost their areas at the expense of Inland waters and Mine, dump and construction sites but it was not on a large scale.

### Land cover change prediction

The key task of this research was to determine the key driving factors responsible for land-cover changes in the last decade of the twentieth century for the case study area. Land change modeller in Idrisi Selva software allows to group drivers into submodels but it makes prediction more difficult to perform. The main 8 transitions were mapped (Fig. 9) and identified with their areas: Permanent crops to Arable land (761 ha); Scrub and/or herbaceous vegetation to Forests (365 ha); Forests to Scrub and/or herbaceous vegetation (296 ha); Open spaces with little or no vegetation to Scrub (240 ha); Arable land to Forests (199 ha); Arable land to Mine, dump and construction sites (198 ha); Mine, dump and construction sites to Urban fabric (196 ha) and Arable land to Heterogeneous agricultural areas (161 ha). Transitions within the area less than 100 ha were ignored in this analysis. The main reason for ignoring these transitions is that they do not have a lot of influence for the modelling. What is more, the method used in this research works more efficiently when the number of modelled transition is lower.

Finally, 8 transitions were identified in model and they were grouped into 5 submodels with exact variables (Table 4). Grouping transitions is allowed only when we can assume that drivers for all transitions in a submodel are the same. In Table 4 variables determined for every submodel are also presented. There is also a piece of information about the data transformation in Table 4.

First submodel has only one transition. It is a transition from Permanent crops class to Arable land class. This is the biggest transition in this research (761 ha). For this transition 3 variables were determined. They were distance to arable land, distance to permanent crops class losses and distance to transition from all classes to arable land. For every variable in the first submodel natural log transformation was performed. The natural log transformation is effective in linearizing distance decay variables. It makes modelling tasks easier to solve (Eastman 2012). For every variable in every submodel the test of the potential explanatory power of a variable was performed. To do this a test contingency Table analysis was used. For qualitative variables, it uses the native categories of the variable to test association with the distribution of land-covers in the later land-cover map. The quantitative measure of association that was used was Cramer's V. A high Cramer's V indicates that the potential explanatory value of the variable is good, but does not guarantee a strong performance since it can-

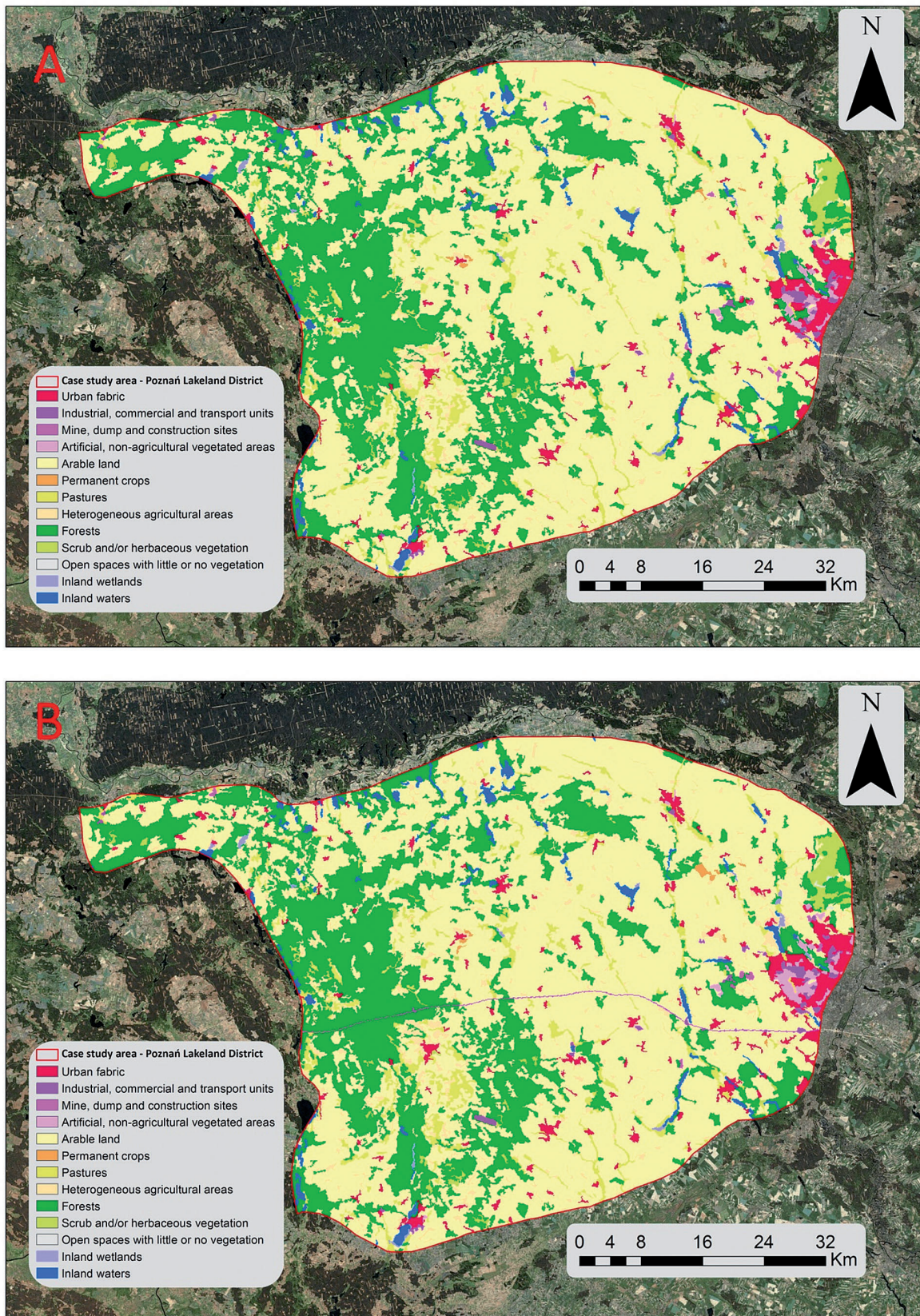


Fig. 8. The output map calculated by the model (A) and reference map (CORINE Land Cover 2006) (B) for the case study area (Poznań Lakeland District)

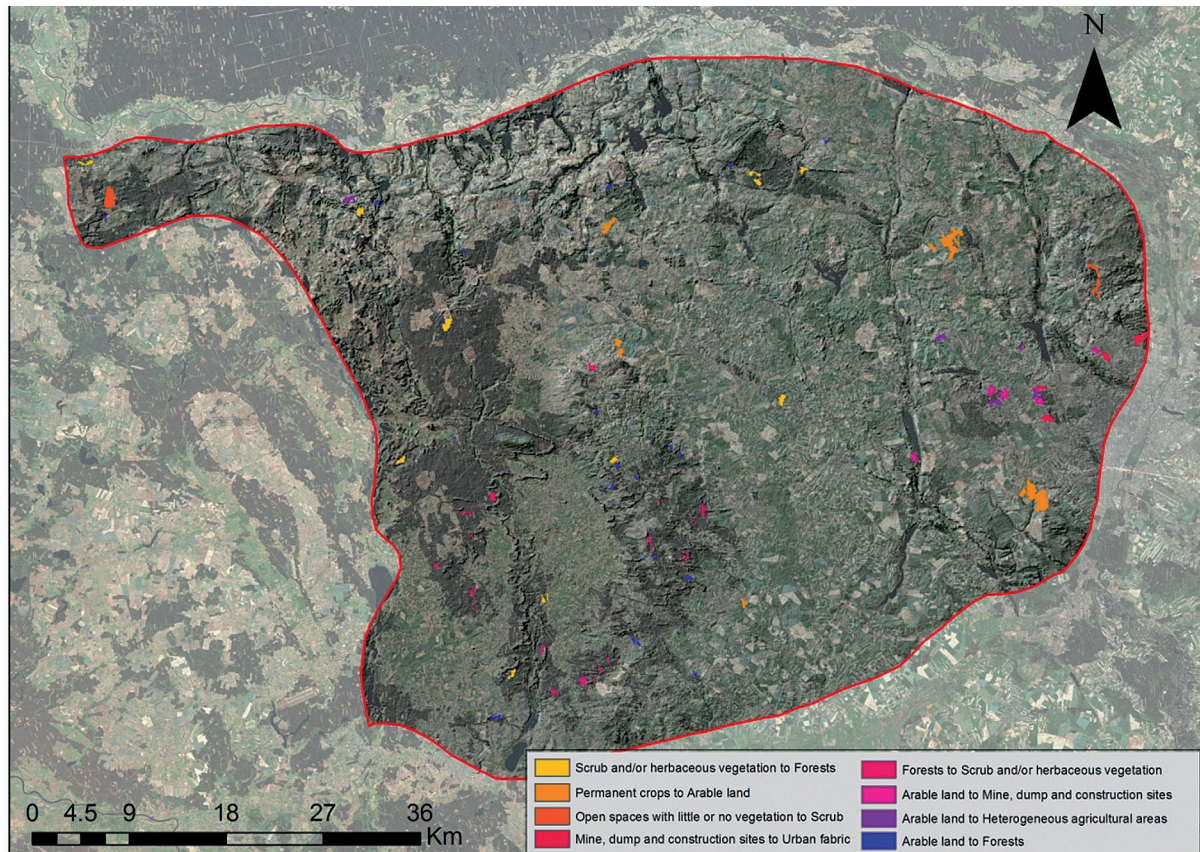


Fig. 9. Main transitions in land cover level 2 for Poznań Lakeland District between 1990 and 2000 (acc. CORINE Land Cover, EEA 2013)

Table 4. Submodels used in the land cover change model with their transitions, variables and information about transformation

Sub-model	Transition(s)	Variables	Transformation
1	Permanent crops to Arable land	Distance to arable land	natural log
		Distance to Permanent crops class losses	natural log
		Distance to transition from all classes to arable land	natural log
2	Forests to Scrub and/or herbaceous vegetation	Distance to transition from all classes to Scrub and/or herbaceous vegetation	natural log
		Open space with little or no vegetation to Scrub and/or herbaceous vegetation	natural log
		Distance to transition from forests to all classes	none
3	Arable land to forests	Distance to transition from all classes to forests	none
	Scrub and/or herbaceous vegetation to forests	Distance to forests	natural log
4	Arable land to Mine, dump and construction sites	Distance to transition from all classes to arable land	natural log
		Distance to arable land	natural log
	Arable land to heterogenous agricultural areas	Distance to heterogenous agricultural areas	natural log
		Distance to transition from all classes heterogenous agricultural areas	natural log
5	Mine, dump and construction sites to Urban fabric	Distance to transition from all classes tu urban fabric	natural log
		Distance to Urban fabric	natural log
		Distance to transition from mine, dump and construction sites to all classes	natural log

not account for the mathematical requirements of the modelling approach used and the complexity of the relationship. However, it is a good indication that a variable can be discarded if the Cramer's  $V$  is low (Eastman 2012). Cramer's  $V$  was calculated for every variable. Only variables with Cramer's  $V$  equalling about 0.3–0.4 were selected to be taken into account in the final model.

The second submodel has two transitions. Both transitions are transitions to scrub and/or herbaceous vegetation class. The first is from forest class, the second is from open space with little or no vegetation class. It was concluded that these two transitions depend on the same driver variables. 3 variables were selected for this submodel. They were distance to scrub and/or herbaceous vegetation; distance to scrub and/or herbaceous vegetation and distance to transition from forests to all classes. First two variables were transformed with natural log transformation. The last one was not transformed. The use of transformation depended on Cramer's  $V$  value calculation result. If this value was higher after transformation, then transformed variable was chosen to the model. If natural log transformation decreased Cramer's  $V$  value, variable without transformation was used.

The third submodel has 2 transitions. These are arable land to forests and scrub and/or herbaceous vegetation to forests. These two transitions were grouped because they show the occurrence of forestation process. Two variables were chosen to be used in the model. These are distance to transition from all classes to forests and distance to forests. Distance to forests was transformed with natural log transformation because after transformation it had higher explanatory power.

The fourth submodel has two transitions. These are arable land to mine, dump and construction sites and arable land to heterogeneous agricultural areas. Both transitions show the occurrence of arable land losses. Four variables were chosen for this submodel. These are distance to transition from all classes to arable land; distance to arable land; distance to heterogeneous agricultural areas and distance to transition from all classes heterogeneous agricultural areas. Every variable was transformed using natural log transformation.

The last, fifth submodel contains only one transition. It is transition from Mine, dump and construction sites to urban fabric. This transition was not grouped with any other transitions because its driver variables have only anthropogenic origin. Three variables were chosen because of its explanatory power in this submodel. The variables are: Distance to transition from all classes to urban fabric; distance to urban fabric; distance to transition from mine, dump and construction sites to all classes. All of these variables were transformed with the natural log transformation.

When all five submodels were constructed, transition potentials could be calculated. Only 8 transition potentials were calculated for the transitions bigger than 100 hectares that were taken into account. Procedure started with initial learning rates (start and end) and reduced these rates by half if significant oscillations in the Root Mean Square (RMS) Error were detected within the first 100 iterations. With each reduction, the process was restarted. Since specific transitions are being modelled, Land Change Modeller masks out the transition potentials of all cases that do not match the form case of any specific transition. For example, if the transition being modelled is "from forest" to agriculture, output values in the transition potential maps will only exist in initial forest areas (Eastman 2012).

### Change Demand Modelling

When all transitions potentials were calculated it became possible to predict the amount of change that will occur in 2006. Year 2006 was selected for prediction because CORINE Land Cover database is available for this year so this data can be used for model validation. For prediction Markov Chain method was used. This method determines the amount of change that will occur at the specific date according to calculated transition potentials. The procedure determines exactly how much land would be expected to transition from the later date to the prediction date based on a projection of the transition potentials into the future and procedure creates a transition probabilities matrix (Eastman 2012). The final transition probabilities matrix calculated for this model is presented in Table 5. The

model was run several times. Output map was compared with CORINE Land Cover 2006 map using Cross Tabulation. Cross tabulation allows evaluating changes between two maps. Specific Table that summarizes the number of grids of each land cover class was created. The Kappa Index of Agreement (KIA) (Cohen 1968, Rosenfield, Fitzpatrick-Lins 1986, Carstensen 1987) was also calculated. The Kappa index of Agreement indicates the degree of agreement between two maps. In addition, the Cramer's V value was calculated (Ott *et al.* 1983). Cramer's V is a way of calculating correlation in Tables, which have more than 2x2 rows and columns. It is used to determine the strength of association between variables and it varies between 0 and 1. Values of the statistics close to 0 denote little association between variables. Values close to 1 indicate a strong association.

After calculation of cross tabulation transition probabilities matrix was edited and model was restarted. It was iterated to obtain the best results of the model so that Cramer's V and Kappa index would be the highest. Final transition probabilities matrix after all iterations is presented in Table 5.

## Validation

Final transition probabilities matrix (Table 5) was used for predicting land cover map for year 2006. The output map and reference map (CORINE Land Cover 2006) are presented in Figure 8. For model validation cross tabulation method and Validate module were used. Validation is the process of making sure that an implemented model matches the real-world. (North, Macal 2007). We cannot expect that a model will perfectly match the reality but this is the point where a modeller should aim. Cross tabulation Table with Cramer's V value and overall Kappa are contained in Table 6. Both values are high. Cramer's V equals 0.8303 and overall Kappa equals 0.9620. It means that there is a really strong association between output and reference map. Table 7 presents the Kappa Index of Agreement for every class in output map generated by the model using CORINE Land Cover 2006 as a reference map.

What we can clearly identify in Table 7 is that the lowest KIA occurs for a class: mine, dump

and construction sites. When we take a look at a cross-tabulation Table (Table 6) we can see that about 50 pixels in an output map have the same classification as in a reference map in this class but more than 200 pixels are classified as arable land (column 3). It means that the model overestimates transitions from Mine, dump and construction sites into Arable land. The reason of that could be that normally class Mine, dump and construction sites often after a few years should be transformed into urban and anthropogenic classes. In this case some enhancements to probabilities matrix should be taken into account.

The next, lowest KIA is identified for class open spaces with little or no vegetation. This low value can be caused by low significance of this class in case study area. Open spaces with little or no vegetation amount to less than 1% of the area in 1990 and 2000 according to CORINE Land Cover database. In the output map generated by model this class has only 69 cells while in the reference map this class has only 99 cells. It can be concluded that for this class it is a relatively large error but for the whole model it is not. This is why KIA for this class is below 0.5 while for the whole model overall KIA is higher than 0.95.

Permanent crops class also has a KIA lower than 0.5. In cross-tabulation table (Table 6) we can see that most of cells are classified as correct (column 6) but also many of them are misclassified as arable land. It is caused by the fact that arable land gained a lot of their area during the last decade of the twentieth century at the expense of permanent crops. All the amount of arable land gain was transitioned from permanent crops. In this kind of modelling we are using information from the past to predict the state in the future. But we have to be aware of the fact that sometimes the past does not have the best explanatory power for the future. This is why the model overestimates transition from permanent crops to arable land even if a modeller has lowered transition probability for these classes in the transition probabilities matrix in modelling iterations. What is also worth noticing in both maps (output and reference ones) is that this class has about 500 cells which means that this class, even if it has a relatively large classification error, it does not have a lot of influence on the whole model.



Table 6. Cross-tabulation of output land cover map for 2006 generated by model (columns) against reference land cover map for 2006 acc. CORINE Land Cover database (EEA 2013) (rows). Digits 1–13 – see Table 5

	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
1	11244	0	5	4	434	10	8	838	1	0	0	0	5	12549
2	69	1513	62	8	692	0	7	18	203	0	0	0	0	2572
3	47	57	45	0	171	0	14	0	44	0	0	0	0	405
4	65	23	0	1528	41	27	2	0	487	0	0	0	0	2173
5	37	27	209	0	241170	162	342	288	163	18	27	20	2	242465
6	0	0	0	0	283	241	0	0	0	0	0	0	0	524
7	0	0	0	0	467	0	15984	12	35	0	0	0	0	16498
8	82	0	0	0	1148	0	265	17228	6	0	0	0	0	18729
9	0	13	0	0	268	0	71	114	106747	1124	0	0	6	108343
10	0	0	0	0	182	0	85	0	183	3172	0	0	0	3622
11	0	0	0	0	6	0	0	0	0	51	42	0	0	99
12	0	0	0	0	0	0	14	0	0	0	0	807	0	821
13	0	0	0	0	2	0	105	42	11	0	0	0	6456	6625
Total	11571	1633	321	1540	244864	440	16897	18540	107880	4365	69	827	6478	415754
Cramer's V			0.8303											
Overall Kappa			0.9620											

What should also be noticed is that artificial, non-agricultural vegetated areas class also has a relatively low (about 0.7) KIA. When we take a closer look at the cross-tabulation table (Table 6) we can see that in column number 4 KIA value is really high. In the output map almost every cell was classified correctly. But row 4 reveals some disadvantages of the model, i.e. that the transition from that Artificial, non-agricultural vegetated areas class to Forests is underestimated. This can be so because the transitions to forests were modelled in submodel 2 and in this submodel transition from that Artificial, non-agricultural vegetated areas was omitted. It could have more influence that the model predicted and thus this underestimation problem occurred.

As it was already said more than half of case study area was covered by arable land and more than a quarter of case study area was covered by forests in 2006. It means that these two classes covered about 85% of this area. It can be further concluded that case study area landscape can be defined as agricultural and forest landscape. These two crucial land cover classes has very high KIA (for both of them higher than 0.95) which can be a sign of the successfulness of the model. To confirm this conclusion VALIDATE module was used.

VALIDATE method according to Eastman (2012), Pontius (2000), Pontius (2002) and Pontius, Suedmeyer (2004) is a method dedicated to provide information about an agreement between two categorical images. VALIDATE can answer two questions which are crucial for modelling, i.e. how well a pair of maps agrees in terms of the quantity of cells in each category and how well

Table 7. Kappa Index of Agreement for every class in output map generated by model using CORINE Land Cover 2006 as a reference map

Land Cover class	Kappa Index of Agreement
Urban fabric	0.8930
Industrial, commercial and transport units	0.5866
Mine, dump and construction sites	0.1104
Artificial, non-agricultural vegetated areas	0.7021
Arable land	0.9870
Permanent crops	0.4594
Pastures	0.9675
Heterogeneous agricultural areas	0.9161
Forests	0.9801
Scrub and/or herbaceous vegetation	0.8744
Open spaces with little or no vegetation	0.4241
Inland wetlands	0.9829
Inland waters	0.9755

a pair of maps agrees in terms of the location of cells in each category (Eastman 2012). To validate the model various Kappa Indices of Agreement and related statistics were calculated to answer these questions. The statistics indicated how well the output map generated by the model agrees with the reference map (CORINE Land Cover 2006). The analysis separates agreement and disagreement between the two images into the following components:

- agreement due to chance,
- agreement due to quantity,
- agreement due to location at the stratified level,
- agreement due to location at the grid cell level,
- disagreement due to location at the grid cell level,
- disagreement due to location at the stratified level,
- disagreement due to quantity.

Because in this research every cell was taken into account in analysis stratified level statistics could be omitted. Two more statistics were also calculated: Kappa for no information (Kno) and kappa for grid-cell level location (Klocation). All statistics are presented in Table 8. For the study area Klocation equals 0.9762 – this indicates how well the grid cells are located on the landscape. Kno equals 0.9761. Division of agreement and disagreement for the study area is: 7% agreement due to chance, 34% agreement due to quantity, 56% agreement at the grid cell level, 1% disagreement at the grid cell level and less than 1% disagreement due to quantity (Table 8). The values of kappa and its variations are close to 1, which indicates that the model is much closer to the perfect agreement than to the level of agreement expected by chance.

Spatial modelling and simulation are not about creating models that can perfectly predict future states. It is and will always be impossible. But our efforts should bring us as close to this state as possible. In this context a created model can be considered a successful modelling tool.

## Conclusion and discussion

The model presented in this research is efficient in predicting land-cover patterns in the fu-

ture. As it was already mentioned there is a need to develop case studies regional models to understand LUCC patterns. In this context models have high explanatory power and can be used in spatial simulations in this region. But of course it has some disadvantages.

The basis for this model is the past. Past transitions in this case are crucial to predict future transitions. But the past is not always the best driver to predict future. In the context of case study area we have to be aware of the fact that in 2004 Poland accessed European Union (EU). This was the time when infrastructural investment started. In many cases transforming arable land to forests became profitable due to becoming the part of UE. Thus, predicting future using information from the past can be helpful but it might be insufficient. It is easy to observe that in Table 1. Changes between the areas of exact classes between 1990 and 2000 are dramatically different from those between 2000 and 2006. When we are operating in study area where the land-cover patterns are complex because of political reasons, longer time period and different modelling approaches should be used. This is the direction where further analysis should go. But even if the changes between areas of exact classes between 1990 and 2000 are dramatically different, the model has strong predicting power and can be used for this area and similar ones that are of interest.

Actually, in LUCC modelling interdisciplinary research and models are used very often. Apart from being a learning tool in unravelling the driving factors and system dynamics, land-cover change models are important in exploring possible future scenarios in the land-cover system. The functioning of the system can be explored with a model through “what-if” scenarios and the visualization of alternative land-cover configurations. These exploratory and projective capacities allow models to be used as communication and learning environment for stakeholders involved in land-cover decision making.

The large diversity of modelling approaches that have evolved over the past years has challenged different authors to review and classify different approaches. Such classification systems are mostly based on the dominant land-use change processes addressed by the model, the simulation technique used in the model or the



underlying theory. One of the very prospective and fast-developing techniques is agent-based modelling. At the most local level, with the unit of analysis being a plot, field or a farm, the match with agents of land-cover change, e.g. a farmer, is very good. Another group of models uses individual agents as units of simulation. There are several characteristics that define agents: they are autonomous, they share an environment through agent communication and interaction and they make decisions that tie behaviour to the environment. Such multi-agent systems give emphasis to the decision-making process of the agents and to the social organization and landscape in which these individuals are embedded.

In the context of this research the attempt of connecting presented method with agent-based modelling method should be made. Also biophysical and system approaches for simulating land-use change (Lee *et al.* 2008) could be applied and merged with presented modelling method. The model that is presented has a significant explanatory and exploratory power and can be a good start for finding solutions incorporating more simulation techniques.

Three crucial tasks are waiting to be solved in future researches for this area. Firstly, the most actual CORINE Land Cover database 2012 should be incorporated into the model. This database should be available for the research in the year 2013. Historical data from the year 1982 should also be used for modelling to enhance its predicting power. CORINE Land Cover 1982 does not exist but it is possible to obtain land-cover maps from earlier years using Landsat images classification techniques. Finally, an attempt of connecting presented method with other techniques should be made.

## References

- Agarwal C., Green G.M., Grove J.M., Evans T.P., Schweik C.M., 2001. *A review and assessment of land use change models: Dynamics of space, time, and human choice*. Center for the Study of Institutions, Population, and Environmental Change, Indiana University and USDA Forest Service, Bloomington and South Burlington.
- Atkinson P., Tatnall A., 1997. Neural networks in remote sensing. *International Journal of Remote Sensing* 18(4): 699–709.
- Barredo J.I., Kasanko M., McCormick N., Lavallo C., 2003. Modelling dynamic spatial processes: simulation of urban future scenarios through cellular automata. *Land-use Urban Plan.* 64: 145–160.
- Briassoulis H., 2000. Analysis of land use change: Theoretical and modeling approaches. In: S. Loveridge (ed.), *The web book of regional science* <http://www.rri.wvu.edu/regscweb.htm>. West Virginia University, Morgantown
- Brown D.G., Lusch D.P., Duda K.A. 1998. Supervised classification of glaciated landscape types using digital elevation data. *Geomorphology* 21(3–4): 233–250.
- Burrough P.A., McDonnell R.A., 1998. *Principles of Geographical Information Systems*. Oxford University Press, Oxford, 333 p.
- Carstensen L.W., 1987. A Measure of Similarity for Cellular Maps, *The American Cartographer* 14(4): 345–358.
- Ciołkosz A., Bielecka E., 2005. *Pokrycie terenu w Polsce; Bazy danych CORINE Land Cover*, Biblioteka Monitoringu Środowiska, Warszawa.
- Cohen J., 1968. Weighed kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* 70 (4): 213–220.
- Czyż T., 2000. Zróżnicowanie regionalne i nowa organizacja terytorialna Polski. In: Ilnicki, D., Ciok, S. (Eds.), *Przekształcenia regionalnych struktur funkcjonalno-przestrzennych*, Wrocław, pp. 55–69.
- Drummond S., Joshi A., Sudduth K. 1998. Application of neural networks: precision farming. *IEEE Transactions on Neural Networks*, 211–215.
- Eastman J.R., 2012. *Idrisi Selva Manual*. Clark University, Worcester, 324 p.
- EEA [European Environmental Agency], 2013. CORINE Land Cover database. Online: <http://www.eea.europa.eu/data-and-maps> (accessed 1 March 2013).
- European Commission, 1993. *CORINE Land cover map and technical guide. European Union Directorate-General Environment, Nuclear Safety and Civil Protection*. Luxembourg.
- Evans B.J., 1997. Dynamic display of spatial data-reliability: does it benefit the map user? *Computers, Geosciences*. 23 (4): 409–422.
- Feranec J., Jaffrain G., Soukup T., Hazeu G., 2010. Determining changes and flows in European landscapes 1990–2000 using CORINE land cover data. *Applied Geography* 30(1): 19–35.
- Foley J., 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., Kucharik C.J., Monfreda, C., Patz J., Prentice C., Ramankutty N., Snyder P.K., 2005. Global consequences of land use. *Science* (New York, N.Y.) 309(5734): 570–574.
- Fukushima K., Miyake S., Takayuki 1983. Neocognitron: a neural network model for a mechanism of visual pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC 13(5): 826–834.
- Haase D., Schwarz N., 2009. Simulation models on human nature interactions in urban landscapes: a review including spatial economics, system dynamics, cellular automata and agent-based approaches. *Living Reviews in Landscape Research* 3.
- Kolendowicz L., Bednorz E., 2010. Topoclimatic differentiation of the area of the Słowiński National Park, northern Poland. *Quaestiones Geographicae* 29(1): 49–56.
- Lambin E.F., Turner B.L., Geist H.J., Agbola S.B., Angelsen A., Bruce J.W., Coomes O.T., Dirzo R., Fischer G., Folke C., George P.S., Homewood K., Imbernon J., Leemans R., Li, X., Moran E.F., Mortimore M., Ramakrishnan P.S., Richards J.F., Skånes H., Will S., Stone G.D., Svedin U., Veldkamp T.A., Vogel C., Xu J., 2001. The causes of land-

- use and land-cover change: moving beyond the myths. *Global Environmental Change* 11(4): 261–269.
- Lee C.L., Huang S.L., Chan S.L., 2008. Biophysical and system approaches for simulating land-use change. *Landscape and Urban Planning* 86(2): 187–203. doi:10.1016/j.landurbplan.2008.02.006
- Łowicki D., 2008. Land use changes in Poland during transformation: Case study of Wielkopolska region. *Landscape and Urban Planning* 87(4): 279–288.
- Łowicki D., Mizgajski A., 2005. Zmiany krajobrazu kulturowego Wielkopolski w okresie transformacji i opisujące je kategorie użytkowania terenu (Changes in the cultural landscape of Wielkopolska in the transformation period (1989–2000) and land use categories determining them). *Przegląd Geograficzny* 77(4): 551–568.
- Macias A., Dryjer M., 2010. Forest cover dynamics in the city of Poznań from 1830 to 2004. *Quaestiones Geographicae* 29(3): 47–57.
- Mackiewicz A., Parysek J.J., Ratajczak W., 1979. A multivariate study of Poland's socio-economic spatial structure in 1975: a principal components analysis with eigenvalues obtained using modified Q R algorithm. *Quaestiones Geographicae* 79(5).
- Marfai M.A., 2011. Impact of coastal inundation on ecology and agricultural land use case study in central Java, Indonesia. *Quaestiones Geographicae* 30(3): 19–32.
- North M.J., Macal C.M., 2007. *Managing business complexity: Discovering strategic solutions with agent-based modelling and simulation*. New York: Oxford University Press.
- Ott L., Larson R.F., Mendenhall W., 1983. *Statistics: A Tool for the Social Sciences*. Boston: Duxbury Press.
- Parker D.C., Berger T., Manson S.M., 2002. Agent-based models of land-use and land-cover change. Report and Review of an International Workshop, October 4–7, 2001. *LUCC Report Series No. 6*, 124 p.
- Parker D.C., Manson S.M., Janssen M.A., Hoffman M., Deadman P., 2003. Multi-agent systems for the simulation of land-use and land cover change: A review. *Annals of the Association of American Geographers* 93: 314–337.
- Pijanowski B.C., Brown D.G., Shellito B.A., Manik G.A., 2002. Using neural networks and GIS to forecast land use changes: a Land Transformation Model. *Computers, Environment and Urban Systems* 26(6): 553–575.
- Pontius R.G., 2000. Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering, Remote Sensing* 66(8): 1011–1016.
- Pontius R.G., 2002. Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogrammetric Engineering, Remote Sensing* 68(10): 1041–1049.
- Pontius R.G., Suedmeyer B., 2004. Components of agreement in categorical maps at multiple resolutions. In: R.S. Lunetta, J.G. Lyon (eds), *Remote Sensing and GIS Accuracy Assessment* p. 233–251. Chapter 17 CRC Press, Boca Raton FL.
- Ritter N., Logan T., Bryant N., 1988. Integration of neural network technologies with geographic information systems. *Proceedings of the GIS symposium: integrating technology and geoscience applications* (pp. 102–103). Denver, Colorado. United States Geological Survey, Washington, DC.
- Rosenfield G.H., Fitzpatrick K.L., 1986. A Coefficient of Agreement as a Measure of Thematic Classification Accuracy. *Photogrammetric Engineering and Remote Sensing* 52(2): 223–227.
- Rumelhart D., Hinton G., Williams R., 1986. Learning internal representations by error propagation. In: D.E. Rumelhart, J.L. McClelland (eds), *Parallel distributed processing: explorations in the microstructures of cognition* (Vol. 1; pp. 318–362). Cambridge: MIT Press.
- Sangermano F., Eastman J.R., Zhu H., 2010. Similarity weighted instance based learning for the generation of transition potentials in land change modeling. *Transactions in GIS* 14(5): 569–580.
- Schaldach R., Priess A., 2008. Integrated Models of the Land System: A Review of Modelling Approaches on the Regional to Global Scale. *Living Reviews in Landscape Research* (2) 1–34.
- Schmit C., Rounsevell M.D.A., La Jeunesse I., 2006. The limitations of spatial land use data in environmental analysis. *Environmental Science, Policy* 9(2): 174–188.
- Schweitzer C., Priess J.A., Das S., 2011. A generic framework for land-use modelling. *Environmental Modelling, Software* 26(8): 1052–1055.
- Skapura D., 1996. *Building neural networks*. New York: ACM Press.
- Szlachta J., 1993. *Rozwój regionalny Polski w warunkach transformacji gospodarczej*, Warszawa.
- Szpikowski J., 2002. Contemporary processes of soil erosion and the transformation of the morphology of slopes in agricultural use in the postglacial catchment of the Chwałimski Potok (upper Parseta, Drawskie Lakeland). *Quaestiones Geographicae* 22: 79–90.
- Veldkamp A., Lambin E.F., 2001. Predicting land-use change: Editorial. *Agriculture Ecosystem Environment* 85(1–3): 1–6.
- Verburg P.H., Veldkamp A., 2005. Editorial: Spatial modeling to explore land use dynamics. *International Journal of Geographical Information Science* 19: 99–102.
- Wang S.H., Huang S.L., William W., Budd W.W., 2012. Integrated ecosystem model for simulating land use allocation. *Ecological Modelling* 227: 46–55.
- Willems E., Vandervoort C., Willekens A., Buffaria B., 2000. Landscape and land cover diversity index. Online: <http://ec.europa.eu/agriculture/publi/landscape> (accessed 1 August 2013).
- Zwołański Zb. (ed.), 2012. GIS – teledetekcja środowiska (GIS – remote sensing of the environment), Bogucki Wyd. Naukowe, Poznań.