MONITORING AND MODELING URBAN LAND-USE CHANGE WITH MULTITEMPORAL SATELLITE DATA

Roland Goetzke, Matthias Braun, Hans-Peter Thamm and Gunter Menz

Center for Remote Sensing of Land Surfaces (ZFL)
University of Bonn
53113, Germany

ABSTRACT

For Central Europe urbanization and urban sprawl are the major processes concerning land-use and land-cover change that alter the characteristic and functioning of the landscape permanently. Against the background of the current sustainable development debate there is an increasing demand of reliable information about the future trends of these land-use developments. Remote Sensing provides the required data and methods for an operational monitoring of land-use/-cover changes. However for the prediction of future trends additional information and a deep understanding of the driving forces of land-use/-cover change is needed. Here we show an exemplary pathway for obtaining the required future land-use information using accurate and stable techniques. We acquired historical land-use information using Landsat imagery. The measured changes (e.g. an increase of urban areas of about 37% between 1984 and 2005) are driven by a variety of socio-economic causes that have been evaluated statistically. We used this information to calibrate a spatial predictive model that has been implemented in the generic modeling framework XULU (eXtendable Unified LandUse Modeling Platform).

Index Terms— land-use change, urban sprawl, modeling, Landsat

1. INTRODUCTION

Monitoring and modeling of land-use (LU) and land-use change (LUC) are central subjects in environmental research [1, 2]. In Central Europe urban sprawl is one of the major problems in this context. Daily in Germany 114 ha of mostly arable land are converted to urban and infrastructure area, associated with an increasing amount of impervious surface with its impacts on the local and regional ecosystem [3]. The federal ministers of the environment have identified this as a problem by means of the Agenda 21 aims and determined to reduce the land consumption down to 30 ha per diem in 2020 [4]. For our research area, the state of North Rhine-Westphalia (NRW), we still have an opposite trend. Despite a stagnating population development every year about 5.500 ha are additionally converted to urban or infrastructure area. This indicates an increasing need of area per inhabitant. The settlement pattern in NRW is very heterogeneous, so the increase in urban area is also inhomogeneous which leads to the question which driving forces are relevant for the changes and how the future development will look like under certain conditions, which is central in LU modeling [5]. Depending on those questions three aims could be formulated. First, an operational monitoring of the LUC must be implemented, especially in order to map the changes in impervious area. Second, the relation between driving forces and LUC must be revealed and third a method to model future land-use trends must be developed.

2. DATA AND METHODS

2.1. Remote Sensing data

For the LUC monitoring we used Landsat data from the years 1975, 1984, 2001, and 2005. The Landsat data have been classified using a combination of knowledge based and supervised methods. For a detailed description of the preprocessing and classification refer to [6, 7]. Altogether 12 LU-classes have been created. It should be noted that three urban classes have been distinguished, depending on their degree of imperviousness. This distinction could be created by a linear regression model using imperviousness information from Quickbird imagery at 1m resolution resampled to a 30m raster and the NDVI from the Landsat data.

2.2. Geophysical and socioeconomic drivers

For the analysis of the driving forces of LUC a large set of geophysical and socioeconomic data were available. The geophysical data have been primarily derived from a Digital Elevation Model (DEM) and an official soil map. Furthermore the fragmentation was calculated based upon the amount of LU classes occurring within a defined matrix around a pixel. Socioeconomic data were available aggregated on community level. Those data had to be disaggregated in order to make them spatially explicit.

Therefore we transferred the e.g. population density information to district level by first mapping the population information to the single district centers and second by weighting the data by the single district sizes. To disaggregate the now point-shaped data, we interpolated them spatially using a circular kernel with 10km radius (approx. mobility radius). As a result we got maps of not 'true' population density, but maps for human activity. For example large industrial areas that contain no noteworthy population have been weighted stronger by this method. For other socioeconomic data that are not linked to district sizes we interpolated using inverse distant weighting.

To display other spatially explicit processes we measured distances, e.g. to rivers etc. Distance measures that display accessibility were weighted by the road network. For a detailed description of the driving forces used in this study refer to [8].

We had to reduce our set of LU classes to 6 (urban, agricultural areas, grassland, forest, water, other) and resample the original 30m data to 100m, because of (1) computational reasons and (2) because the intention of the study was to map trends and too much of detail would have concealed the general relations and processes in LU. Detailed studies about aggregating LU categories have been carried out by [9] and about changing resolution in predictive models by [10].

2.3. Correlation analysis

To extract from the plenty of information the important drivers for LU and LUC we tested which data correlate significantly with LU at the specified points in time and with the changes in between. Therefore we conducted a simple correlation analysis with all variables aggregated at community level. For latter analysis we considered only variables that correlated significantly with LU on the 95% confidence interval. To avoid multicollinearity all variables were additionally correlated in a regression analysis among each other and all variables with an R^2 of > 0.8 were discarded [11]. The remaining set of variables was to be considered to be in relation with LU or LUC.

2.4. Logistic regression

In the logistic regression analysis all driving forces were disaggregated and so spatially explicit. The emergence of a certain LU class at a specific point in time is the result of the effects of a set of driving forces. To describe the emergence of a certain LU we produced probability maps. Those maps were created by using a logistic regression analysis whereby the probability of occurrence (P) for a LU class or dependent variable (Y) is calculated in dependence of the occurrence of a set of driving forces or independent variables (X_n)

$$P(Y_i = 1 \mid X_n) = \pi_i = \frac{e^{\beta_0 + \beta_n X_n}}{1 + e^{\beta_0 + \beta_n X_n}}$$

with β_0 as the regression constant and β_n as the regression parameters to be calculated. A normal linear regression is here not adequate because the dependent variable can only have two values (1 = LU is present, 0 = LU is absent). To estimate the regression parameters the equation has to be transferred:

$$\frac{\pi_i}{1 - \pi_i} = e^{\beta_0 + \beta_n X_n}$$

On the left hand side of the equation are now the odds or the ratio of probability and counter-probability. By taking the logarithm of the odds we get the logit of the dependent variable (Y) and the function can be treated as a linear regression:

With the calculated betas the function can be retransformed.

$$\log_e\left(\frac{\pi_i}{1 - \pi_i}\right) = Logit(Y) = \beta_0 + \beta_n X_n$$

To estimate the effects of the single variables on the probability of occurrence for a LU class, we calculated the odds ratios (OR)

$$OR = \frac{\pi_A}{1 - \pi_A} / \frac{\pi_B}{1 - \pi_B}$$

where A and B are different parameter values, here the 25 and 75 quartiles.

The goodness of fit of a logistic regression model cannot be measured by R^2. Therefore we used the Receiver Operating Characteristic (ROC) to estimate the goodness of fit, which is the ratio between the True Positives and the False Positives of the prediction that has been made by the regression model. With this method the goodness of fit can be measured by the area under the curve (AUC), which is with a perfect prediction 1 and 0 when a complete random distribution is predicted. All values above 0.7 are considered acceptable in LU applications [12, 13].

2.5. Land-use modeling

The results from the logistic regression analysis are a major input for the Clue-S model that we used to model LU in NRW. Besides the so revealed relation between LU and driving forces other parameters are needed for the model, e.g. neighborhood probabilities, conversion elasticity, LU demand, etc. The Clue-S model has been developed by Verburg et al. [14]. We used the Clue-S model as a plugin in XULU (eXtendable Unified LandUse modeling platform), a JAVA based modeling framework that provides data management, visualization, and parallel computing functions. The XULU modeling framework has been developed by Schmitz et al. [15] to overcome difficulties in the handling of the Clue-S model and to provide an easy to use basis for model interoperability.

3. RESULTS

3.1. Land-use classification

The LU changes that could be detected by the combined knowledge based and supervised classification of the Landsat data for the four time steps are described in detail in [7]. A loss of agricultural areas and grassland of –27% (=-8.7% of the total land area) between 1975 and 2005 face a gain in impervious surface areas (high to low) of about 63% (=5.4% of the total land area). To achieve a satisfying overall accuracy of ~86% the results had to be visually inspected and errors in class assignment corrected that mostly occurred in classes that were spectrally similar like high density impervious surface and barren soil in gravel pits or agricultural areas. The use of the NDVI as a measure for vegetation fraction that has been correlated with imperviousness maps based on Quickbird and aerial images of test areas showed significantly better results in the class assignment of the imperviousness classes than tests with the maximum likelihood classifier. For further analysis the classes have been reduced to six classes (urban, agriculture, grassland, forest, water, other) and resampled to 100m.

3.2. Driving forces analysis

The correlation and logistic regression analysis have been carried out for calculating probabilities for all six LU classes for the year 1984 and for the prediction of LUC between 1984 and 2001. We decided to omit the 1975 data because of uncertainties that occurred because of the coarser resolution of the Landsat MSS data leading to an underestimation of highly clustered LU classes (e.g. urban).
The correlation analysis showed that the probability for urban areas in 1984 can be best described by the parameters shown in Tab. 1.

TABLE I. RESULTS OF THE CORRELATION ANALYSIS FOR THE LU CLASS 'URBAN' AND THE LUC BETWEEN 1984 AND 2001 (HIGHEST R'S ONLY)

Parameters	r (LU 1984)	r (LUC 1984-2001)
Fragmentation	0.617	-
Proximity to highways (500m buffer)	0.637	0.603
Distance to historic city 1840	-0.488	-
Distance to historic city 1900	-0.536	-
Distance to train station	-0.685	-0.623
Distance to highway entrance	-0.537	-0.582
Distance to city >100.000	-0.604	-0.566
Population density	0.969	0.731
Population density change	-	0.532
Construction permits	-	0.567
Unemployment rate	-	0.434

A comparison showed that the 8 driving forces that correlated best (highest Pearson's correlation coefficient) with LU added up to no inferior results than all driving forces. The same has been calculated for the other LU classes and for the LUC between 1984 and 2001 (for urban areas refer to Tab. 1).

We used the selected parameters as independent variables in the logistic regression. The AUC values taken from the ROC analysis as a measure for the goodness of fit showed that urban areas (0.901), agricultural areas (0.770), forest (0.791), and water (0.804) in 1984 could be predicted well by the variables used. The poorer results of grassland (0.668) and "other" (0.655) indicate the problem of improper class assignment. Grassland consists of meadow and pasture which are indistinguishable spectrally by multispectral satellite data, but occur at different locations. In "other" we subsumed the classes gravel pits & quarries, lignite mining and military training areas, which also cannot be described by the same variables. The regression weights and the odds ratios for the exemplary LU class "urban" can be taken from Tab. 2.

TABLE II. REGRESSION WEIGHTS AND ODD RATIOS AS A RESULT OF THE LOGISTIC REGRESSION ANALYSIS PREDICTING LU IN 1984

Predictor	β	OR
Intercept	0.89994	-
Fragmentation	32.2680	1.88
Proximity to highways (500m buffer)	-0.71762	0.47
Distance to historic city 1900	0,00001	1.30
Distance to train station	-0.00018	0.07
Distance to highway entrance	-0.00001	0.69
Distance to city >100.000	-0.000004	0.78
Population density	0.00063	1.82

all variables are significant at $p < 0.001$

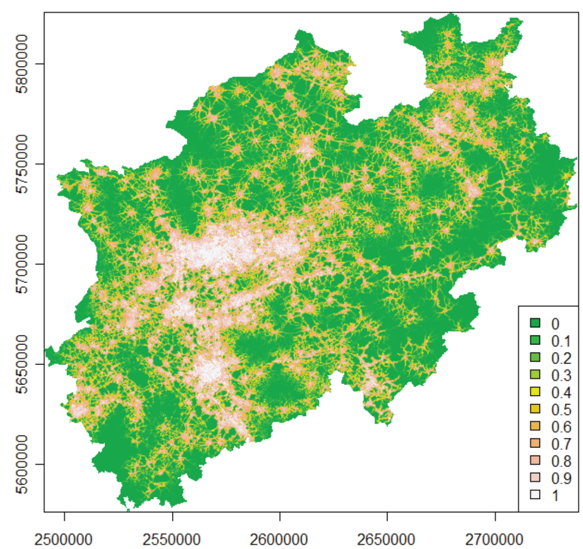

Figure 1. Probabiliy for urban LU in 1984 as a result of the logistic regression analysis

The resulting probability map for the occurrence of urban areas in 1984 is depicted in Fig. 1. The logistic regression has also been applied to calculate the probability for LUC between 1984 and 2001 resulting in an AUC value of 0.822. As independent variables we used the parameters shown in the right column of Tab. 1.

3.3. Land-use modeling with XULU

The results of the logistic regression analyses have been used as a major input for the Clue-S model that has been implemented as a plugin in the XULU modeling framework. The model could be calibrated by modeling the LUC between 1984 and 2001, which now offers the possibility to run the model up to 2020 with two different scenarios, which are “business as usual” and “reduction of land consumption”. So far visual inspection showed good modeling results although a quantitative comparison of the modeled LU of 2001 with the reference data has yet to be carried out. Problems occurred in the modeling of the classes “other” and “grassland”.

4. DISCUSSION AND OUTLOOK

The presented study shows an exemplary pathway to enhance remotely sensed data with additional socioeconomic or geophysical information to gain a better understanding of processes in LUC, which is in the literature often described as “from pixels to processes”. Crucial within that work is the logistic regression analysis that identifies relations between LU and driving forces.

XULU has proven to be a very useful tool for spatial modeling that makes the Clue-S model, which is implemented as a plugin, even more powerful. The researcher has direct access to all parameters, it offers the possibility to alter parameters at runtime, visualize intermediate results, and use parallel computing, which is because of the very large amount of data very important.

The study also showed that in complex LU compositions like NRW LU modeling is possible due to the good data basis concerning socioeconomic an geophysical data. But also difficulties could be detected, e.g. reasonable class assignment, handling of the large amount of data and for this reason the required reduction of complexity by means of resolution and number of classes.

5. REFERENCES

[1] E.F. Lambin and H. Geist (Ed), *Land-use and land-cover change. Local processes and global impacts*, Springer, Berlin / Heidelberg / New York, 2006.

[2] P.M. Vitousek, H.A. Mooney, J. Lubchenko, and J.M. Melillo, “Human domination of earth’s ecosystems,” *Science*, vol. 277, pp. 494-500, 1997.

[3] C.L. Arnold and C.J. Gibbons, “Impervious surface coverage: The emergence of a key environmental indicator,” Journal of the American Planning Association, vol. 62/2, pp. 243-258, 1996.

[4] Umweltbundesamt (Ed.), *Reduzierung der Flächeninanspruchnahme durch Siedlung und Verkehr*, Umweltbundesamt, Berlin, 2003.

[5] P.H. Verburg, J.R. Ritsema van Eck, T.C.M. de Nijs, M.J. Dijst, and P. Schot, “Determinants of land-use change patterns in the Netherlands,“ *Environment and Planning B: Planning and Design*, vol. 31, pp. 125-150, 2004.

[6] B. Schoettker, M. Over, M. Braun, A. Siegmund, and G. Menz, “Statewide land use change detection using remote sensing,” Proc. SPIE 2003, Barcelona, Spain, 2003.

[7] R. Goetzke, M. Over, and M. Braun, “A method to map land-use change and urban growth in North Rhine-Westphalia (Germany),” Proc. 2nd Workshop of the EARSeL SIG on Land Use and Land Cover, September 2006, Bonn, Germany, pp. 102-111, 2007.

[8] R. Goetzke, M. Judex, M. Braun & G. Menz, “Evaluation of Driving Forces of Land-Use Change and Urban Growth in North Rhine-Westphalia (Germany),” Proc. IGARSS, Barcelona, Spain, 2007.

[9] R.G. Pontius Jr. and N.R. Malizia, “Effect of Category Aggregation on Map Comparison,” in M.J. Egenhofer, C. Freska, and H.J. Miller (Eds.), *GIScience 2004*, Springer, Berlin / Heidelberg, pp. 251-268, 2004.

[10] R. Costanza and T. Maxwell, “Resolution and predictability: An approach to the scaling problem,” *Landscape Ecology*, vol.9, no.1, pp. 47-57, 1994.

[11] S. Menard, *Applied logistic regression analysis. Quantitative Applications in the Social Sciences,* Sage Publications, Thousand Oaks, CA, 2002.

[12] R.G. Pontius Jr. and L.C. Schneider, “Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA”, *Agriculture, Ecosystems and Environment*, vol. 85, pp. 239-248, 2001

[13] J.P. Lesschen, P.H. Verburg, and S.J. Staal, “Statistical methods for analyzing the spatial dimension of changes in land use and farming systems,” *LUCC Report Series No. 7*, 2005.

[14] P.H. Verburg, W. Soepboer, A. Veldkamp, R. Limpiada, V. Espaldon, “Modeling the Spatial Dynamics of Regional Land Use: The Clue-S Model”, *Environmental Management*, Vol. 30, No. 3, pp. 391-406, 2002.

[15] M. Schmitz, T. Bode, H.P. Thamm, and A.B. Cremers, “XULU – A generic JAVA-based platform to simulate land use and land cover change (LUCC),” Proc. MODSIM 2007, New Zealand, pp. 2645-2651, 2007.

[16] P.H. Verburg, T.C.M de Nijs, J. Ritsema van Eck, H. Visser, and K. de Jong, “A method to analyse neighbourhood characteristics of land use patterns,” *Computers, Environment and Urban Systems*, vol. 28, pp. 667-690, 2004.