MEASURING PERFORMANCE OF LAND USE MODELS

An Evaluation Framework for the Calibration and Validation of Integrated Land Use Models Featuring Cellular Automata

Alex HAGEN-ZANKER[1], Joyce van LOON[1,2], Arjan MAAS[1], Bas STRAATMAN[1,3], Ton de NIJS[4] and Guy ENGELEN[1,5]

[1] Research Institute for Knowledge Systems

[2] Maastricht University, Quantitative Economics

[3] Memorial University of Newfoundland, Department of Geography

[4] Netherlands Environmental Assessment Agency

[5] Flemish Institute for Technological Research, Centre for Integrated Studies

The automated calibration and validation of integrated land use models requires objective quantification of model performance at different levels of abstraction. This paper puts forward a calibration and validation routine and specifically focuses on the procedures for evaluating the model output. It is found that although a fully objective procedure is not yet available, a major part of the analytical tasks can be automated.

KEYWORDS

land use, cellular automata, integrated modelling, calibration, validation

INTRODUCTION

We are considering a land use model as proposed by Engelen et al. (2003). This model combines a spatial interaction (gravity) model at the regional level with a constrained cellular automata to disaggregate regional land use claims to cells. A feedback of cellular aggregates, such as mean suitability of available land, is input to the spatial interaction model. Thus, a tightly integrated and dynamic system is established. This model has been applied amongst other for ex post analysis (Geurs et al. 2003) and ex ante analysis (de Nijs et al 2004) of Dutch spatial policy. A major challenge in the application of the model is the calibration and validation (White and Engelen 2003, Straatman et al. 2004).

The reality of using the model learns that as time progresses it is necessary to recalibrate the model, because new data become available or new expectations of the model require adjustments. As always limited time and resources are available, calling for a manageable procedure for setting parameter values. This paper describes some steps towards a fully automated calibration that were taken as part of a calibration exercise for the Dutch Environment Explorer model. The results of the exercise were not only a calibrated model but also guidelines for future calibrations. These guidelines where put to the test, when calibrated models had to be produced for two other regions; Estonia and North-Italy. This paper will focus on the evaluation framework applied in the calibration and validation. Nevertheless, in accordance with the handbook Good Modelling Practice (van Waveren et al. 1999) other analytical tasks such as verification, global behaviour analysis, sensitivity analysis, and robustness tests have been performed as well. Also for the precise procedures to adjust parameter values, we refer to the calibration report (RIKS 2004).

METHODOLOGY – CALIBRATION PROCEDURE

The main intended use of the model is to explore possible future land use for approximately 30 years ahead. The historical land use data that is available however, does not stretch such a long period. Therefore the calibration is not only aimed at a best fit with historical data, but also at the general landscape structure or morphology that unrolls from the model dynamics when it is applied for periods that long surpass the available data.

It is important to realize that the model under consideration is an integrated model consisting of several model components that are dynamically linked. The difficulty of attributing discrepancies in the integrated model results to one or the other model component makes it attractive, possibly essential, to calibrate the model components individually. Thus the dynamic link between the model components is temporarily cut and the models are applied in a chain. After their individual calibration the models are reconnected. It is then necessary to assess how the different goodness of fit measures are affected by the restored dynamics.

The model components are the *national model*, the *regional model* and the *local model*. The regional model can further be split into the *gravity model* and the *density model*. At the national level the model is constituted simply of timelines that are the driving force of the regional model. For the national model historical data can be used and it requires no calibration. The regional model poses constraints to the local model in the form of land use demands per region. Therefore in the decoupled state these land use demands need to be supplied. A feedback to the regional model is formed by the regional aggregates of cellular characteristic as they transpire from the local model. These are once logged by the local model, after which they are used as input for the regional model. The local model is the constrained cellular automata model, this model is the most calculation intensive and requires the most parameters.

The flowchart of figure 1 illustrates the general calibration procedure. Six iteration loops are recognized (labelled in Roman numbers I to VI). Each iteration loop is started when the previous one is satisfied. The model starts by tuning the parameters of the cellular automata model to best match the historical data (iteration I). The optimization criterion is the agreement between the actual and simulated land use at the end of the calibration period as measured by the Fuzzy Kappa metric (Hagen 2003). The premise of this comparison method is that it not only credits exact cell to cell agreement but also near cell-to-cell agreement. Additional analysis of the maps helps to determine which land use categories require the most adjustment (Fuzzy Kappa per category, following Hagen-Zanker et al 2005) and also what interactions between land uses are likely to fix the discrepancies. The followed procedure is a variation on the one described in Straatman. This part of the calibration (iteration I) can be performed automatically, although it is found that the transition rules found by the automatic procedure need to be scrutinized to exclude aberrations.

Even if the model performs well at reproducing the historical data, it is possible that following these trends into the future leads to unrealistic landscape morphology. The evaluation of morphology and the transition rule adjustment follows the expert judgment based procedure laid out by White and Engelen(2003). An experienced user of the model has good understanding of the relation between transition rules and resulting patterns. A regularity in the distribution of cluster sizes, known as Zipf's law may be used to automatise this iteration. In this instance it has been used only for a single parameter (alpha) which has a strong influence on the morphology of the resulting landscape. The alpha parameter sets the stochastic volatility of the model and is thus responsible for the degree to which new clusters are formed.

Iteration III alternately invokes Iteration I and iteration II and thus balances the morphological coherence and historical accuracy. The general idea is that iteration II (morphology coherence) is meant for some coarse corrective parameter adjustments, whereas iteration I is meant for refined calibration. Once both coherence and accuracy are satisfactory, regional aggregates are calculated (such as the mean suitability of yet undeveloped land) to be used as input for the calibration of the regional model. 'Satisfactory' is still a non-quantified property and is mainly based on the modeller's perception whether prolonged a calibration effort of the cellular automata model will yield substantial improvements.

Finally one parameter that is essential to the morphology of the model is calibrated by an analysis of the cluster size distribution. It was found that the maps over time demonstrated to obey Zipf's law which assumes a linear relation between the log of the cluster size and the proportion of clusters exceeding that size. It was then found that this linear relation changes over time and the parameter stimulating stochastic perturbation was adjusted to best fit a projection of the cluster size distribution

Iteration IV is the calibration of the regional model. Within the regional model the gravity model and the density model can be distinguished. These have a reciprocal dynamical relation, as the density of activities in a region are a contributing factor to the attractivity of that region and vice versa the activity level of a region determines to a large extent the density of that activity (jobs or population per ha). The parameters of the two model components are calibrated separately although the links between the components remain intact. The same search algorithm is used for the two components, Golden Section Search to find local optimums and Random Search to escape the local optimum in the search for the global optimum. The procedure has been described in depth by Van Loon (2004) and it is fully automated. The goodness-of-fit measure is different for both components, the gravity model is optimised with respect to levels of activity and the density model with respect to density. These criteria are chosen (instead of for instance the land use area demands that are the product of activity and density) because they do not allow wrong parameter values in both modules to cancel each other out.

Once both the local and the regional model are calibrated, the dynamic link between the models is re-established and the integrated model is run once more. Now, an evaluation takes place to make sure that the re-coupling does not negatively distort the goodness-of-fit obtained in the decoupled modules. This is done for both the regional model and the cellular model, but in first instance the regional model (iteration V) because recalibration of the regional model is a minor task compared to recalibration of the local model. If it is found that goodness-of-fit deteriorated, the run with the integrated model is used to generate a new time series of regional aggregates and the regional model is recalibrated.

Once the regional model is stable, the same procedure is applied with regards to the cellular automata model. If it is found that over the calibration period the Fuzzy Kappa has diminished or the landscape morphology in the long run is not consistent, than the whole procedure is started again (Iteration VI) with the difference that the time series for regional land use demands are not based on historical data, but the run with the integrated model instead.

METHODOLOGY – VALIDATION PROCEDURE

To asses the predictive value of the model it is applied over a validation period. The goodness-of-fit measures for the validation are essentially the same as for the calibration, except that for the regional model measures are chosen that are more intuitive to interpret. Instead of the sum of squared errors of activity and density growth, the mean absolute error of activity growth and land use demands are calculated. For the cellular model the Fuzzy Kappa statistic was applied, but additional also the spatial distribution of similarity is considered.

In order to get a feeling for the meaning of the quantitative results, they were compared against the results obtained by naive predictors. Naive predictors are alternative models that satisfy the constraints put upon the actual model by minimally changing the initial situation. The general idea behind these models is that 'the best prediction for the weather of tomorrow is the weather of today'. For the regional model the naive predictor is the constant share model. This model distributes the national growth of an activity by keeping the relative distribution over the different regions constant. This means that the same growth factor is applied for all regions. A second aspect of the constant share model is that the density of different activities remains constant. In effect this means that the growth factor for a given activity is also applied for the land use claim associated to that activity. For example, the area of a region taken in by the land use "Industrial" follows the national trend of "Employment in the industrial sector".

The naive predictor for cellular land use change is that of minimum change to satisfy the constraints and random selection of the location for those changes. Thus, the naive predictor is pushed to satisfy the same regional constraints as the Constrained Cellular Automata. The model starts with the initial map and for a region randomly selects cells of land uses that are overrepresented (compared to the regional constraints) and then randomly assigns those cells to land uses that are underrepresented, until all constraints are satisfied. An alternative would be to apply a 'no change at all' naive predictor (as in Hagen 2003), but the current approach has the advantage that the overall composition of the maps is identical, making it better possible to focus on the quality of the configuration. Otherwise it would be difficult to separate composition (quantity) and configuration (location). Also the naive predictor and the cellular automata model would not be subject to the same constraints, leaving them less comparable.

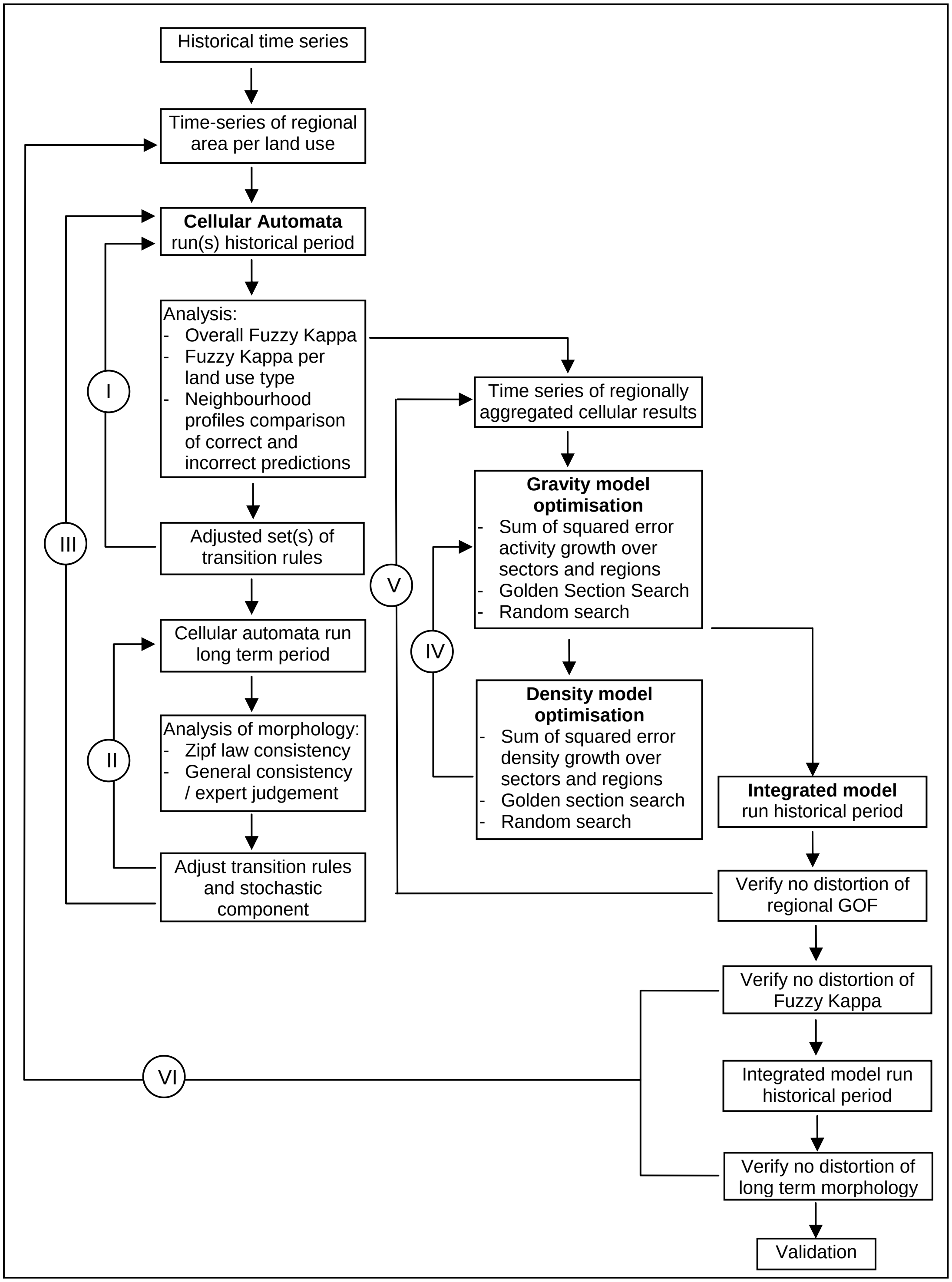


Figure 1 Flowchart overview of the full calibration procedure.

RESULTS - THE PILOT CASE: ENVIRONMENT EXPLORER FOR THE NETHERLANDS.

Four land use maps were available, for the years 1989, 1993, 1996 and 2000. They are all raster maps with a cell size of 500 m. The map of the year 2000 has consistency issues because it is known that definitions of some land use categories have changed. The calibration period is 1989-1996 to ascertain that the model is calibrated over the longest available period of time without being affected by the year 2000 inconsistency.

The validation period is 1996-2000. Despite the problems, the 2000 map is used anyway, under the rationale that the naive predictor is affected by this inconsistency as much as the Cellular Automata model.

The results (table 1) indicate that the calibration of the cellular model is successful in the sense that the model outperforms the naive predictors over the calibration period (1989-1996). Over the validation period 1996-2000 the model does not outperform the naive predictors. Different reasons can be called to explain this. The main believe is that the calibration period was too short to pick up on large scale spatial processes. In such a period relatively little land use changes do occur, and to aggravate the relative proportion of mapping errors over true land use change is large as well. An analysis on the basis of the (thematically aggregated) contingency table learns that 25% of the changes are unlikely, in the sense that they are transitions from urban area to natural or agricultural area or because they are changes of land use types that are not expected to change (mainly fresh water , but also salt water, airports and foreign country). This is a large percentage, but given the situation that only 3% of all cells change, a mapping accuracy of just less than 99.6% in both maps can be sufficient to cause such a disturbance. Note that this does not imply that the accuracy is larger than 99.6 %, since an unknown number of cells may harbour identical inadequacies for both years.

	1989-1996	1996-2000
Model	0.936	0.913
Naive	0.926	0.922

a. Fuzzy Kappa, map similarity

	1989-1996	1996-2000
Model	3.9%	5.2%
Naive	5.2%	3.9%

b. % error in growth of activities

	1989-1996	1996-2000
Model	3.3	7.7
Naive	5.7	6.4

c. errors in regional land use claims (25ha cells per land use type)

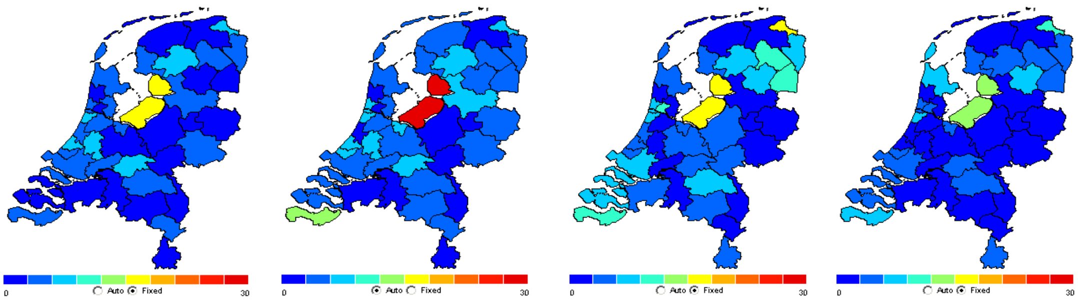
Table 1: Goodness of fit measures for the calibration and validation period

	Nature	Agriculture	Urban	Work	Features
Nature	11458	333	202	196	**64**
Agriculture	1371	97097	674	1127	**245**
Urban	**231**	**252**	3524	113	**104**
Work	**178**	**145**	81	17282	**44**
Features	**91**	**125**	**78**	**64**	4602

Total cells	139681
Unchanged	133963
Changed	5718
Suspect	1621

Table 2: Contingency table summarizing changes from 1989(rows) to 1996 (columns)

A spatial distribution of the error of the regional model (figure 2) clarifies that much of the error can be attributed to a single region which is the ‘Flevopolder’. This is a very young region; the land was only claimed from the sea in between 1939 and 1968. It is not surprising that this region is still developing at another pace than the rest of the Netherlands. This outlier may have contributed to an over-calibration of the regional model, meaning that the parameters have been adjusted too much to reproduce the behaviour of the ‘Flevopolder’ at the cost of the other regions.



a. Model 1989-1996 b. Naive 1989-1996 c. Model 1996-2000 d. Naive 1996-2000

Figure 2. Distribution of activity growth error over the regions

Figure 3. Spatial distribution of disagreement in 2000

A spatial distribution of errors at the cellular level (figure 3) indicates that errors are distributed more or less equally over the map, although errors are in general found in connection to urban areas. This is not surprising considering that these are the more dynamic areas and it is more difficult to correctly predict change than to predict non-change.

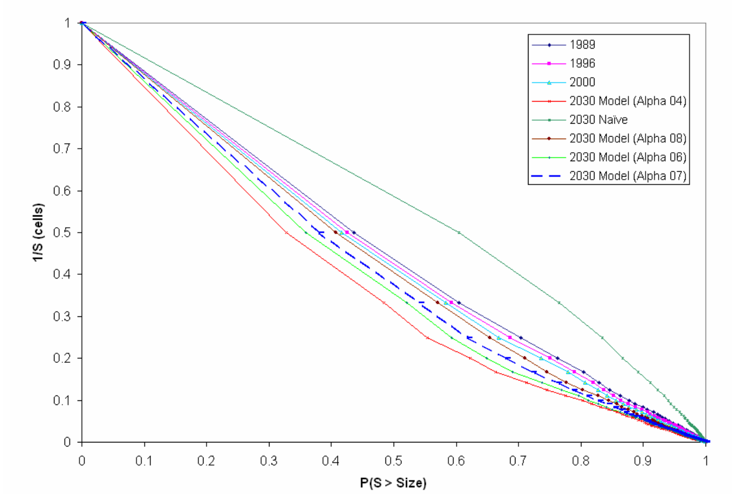
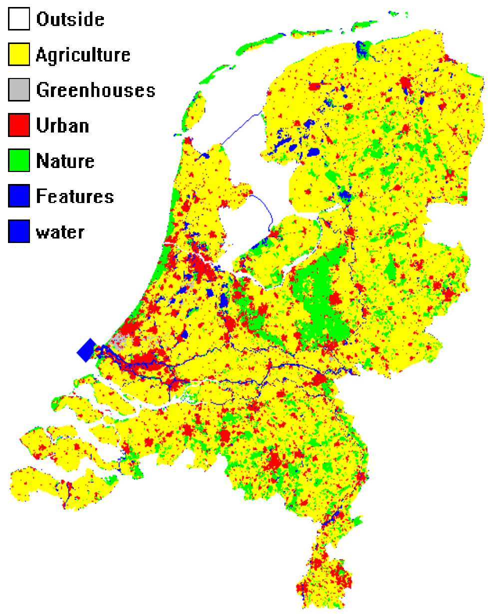
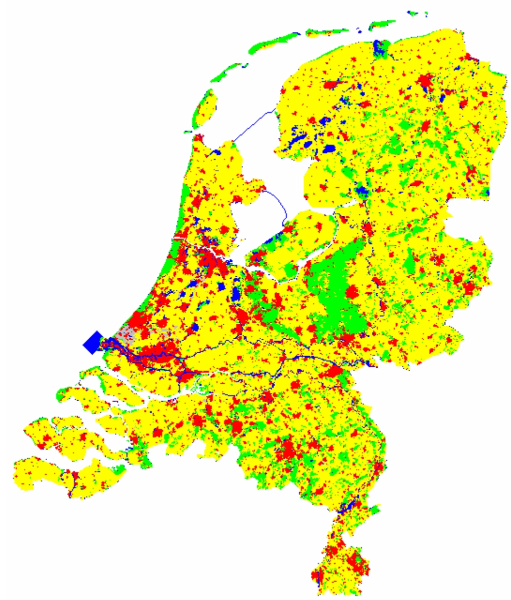


Figure 4. Cluster size distribution, for clusters of urban cells, the y-axis display the inverse of the cluster size and the x-axis the fraction of clusters equal or larger in size

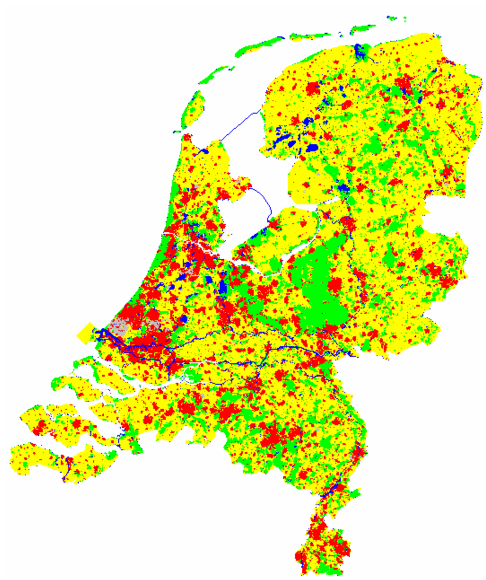
The cluster size distribution displayed in figure 4 illustrates how the cluster size changes over time. It also shows how cluster size distributions are impacted by the value of the parameter alpha. On the basis of these results alpha has been set at the value 0.7. Figure 5 shows input and output land use maps where categories have been collapsed into main land use types. The outcome of the cluster analysis is that the clusters in the model map of 2030 fits better than the naive predictor to the trend in cluster size distribution that can be recognized in the land use maps of 1989 and 2000.



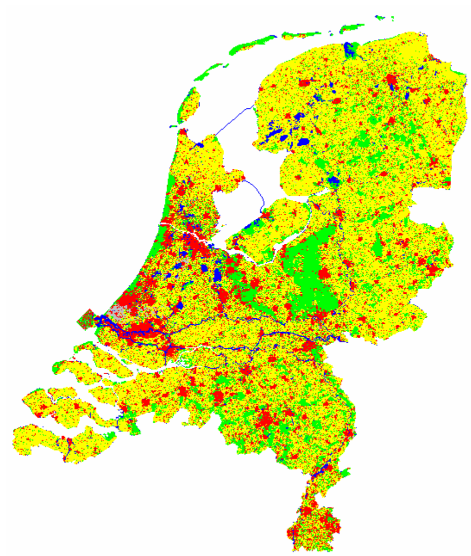
a. Land use 1989



b. Land use 2000



c. Land use 2030 Model (Alpha 0.7)



d. Land use 2030 Naive predictor

Figure 5. Land use maps used as input to the cluster analysis

RESULTS – FURTHER EXPERIENCE WITH ESTONIA AND NORTHERN ITALY

The pilot case for the Netherlands is followed by two more cases for Estonia and Northern Italy. The data on which the model is based is CORINE90 and CORINE2000 and regional economic indicators of the New Cronos REGIO database. The calibration of the regional spatial interaction model went fine but uneventful and is left out of consideration here. The true challenge has been the calibration of the Cellular Automata model. Although maps of two years were available the difference between the 1990 and 2000 map seem small and to an extent erratic. For instance it is striking that between 1990 and 2000 the Milano area in Northern Italy does not display any growth in urban area whereas the Torino area does. The only major change that occurred in the maps of Estonia is a large scale transition from pastures to intensive agriculture. If the land use categories are collapsed and only Agriculture, Nature, Industry, Urban and Features (water and roads) remain, then less than 0.5% of the cells are subject to change, of which 9% are 'suspect' changes. In the Italy case, less than 2 % of the cells changed, of which 15 % are suspect changes.

An additional challenge of the Estonia and Italy cases is that the model is intended to give spatially explicit explorations of scenarios that include land uses that in the past were not even present. Such land uses are biofuels and urban areas for specific population groups, gated communities for the affluent and thematic cities for alternative lifestyles. The purpose of the calibration here is not to make sure that the cellular automata transition rules optimally mimic historic time series, but instead that the rules exhibit the behaviour in accordance with the scenarios as they were delivered. Description of the daily activities of inhabitants of the new land use types and also their modes of transport were interpreted to rules of spatial configuration. These rules were then expressed in terms of cluster sizes and their dependence of other land use types.

	Nature	Agriculture	Urban	Industrial	Features
Nature	111460	254	10	41	8
Agriculture	334	59254	65	5	2
Urban	2	9	2161	2	0
Industrial	24	5	3	1288	0
Features	11	1	0	2	9438

Total cells	184379
Unchanged	183601
Changed	778
Suspect	64

Table 3: Contingency table summarizing changes in Estonia from CORINE90(rows) to CORINE2000(columns)

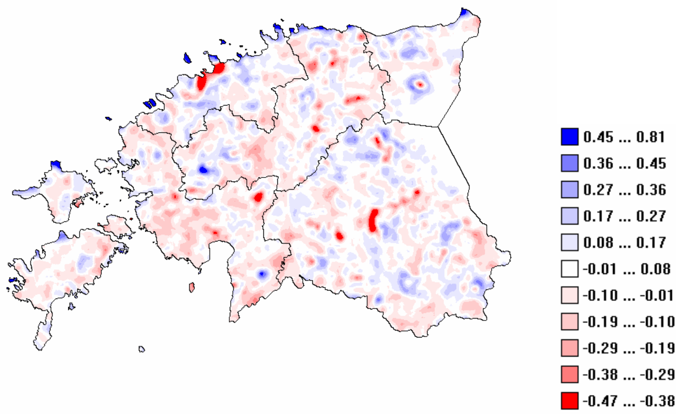
	Nature	Agriculture	Urban	Industrial	Features
Nature	274364	290	78	6	59
Agriculture	1086	257760	1786	93	142
Urban	16	71	23023	703	3
Industrial	28	49	4888	615	751
Features	31	24	14	331	7727

Total cells	573938
Unchanged	563489
Changed	10449
Suspect	1519

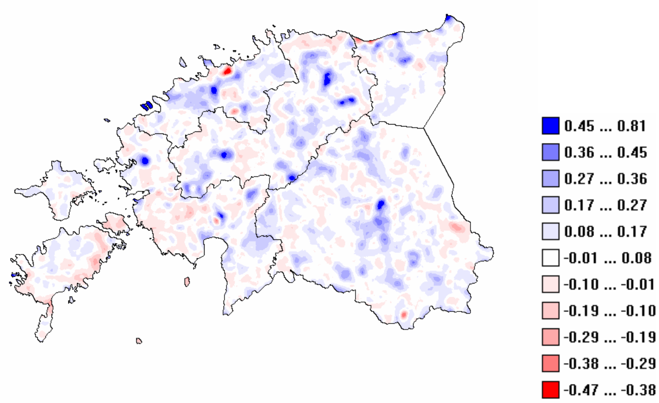
Table 4: Contingency table summarizing changes in Northern Italy from CORINE90 (rows) to CORINE2000(columns)

In the end, for lack of data, the transition rules as they followed from the Netherlands case were used. To test the robustness of the rules for both Italy and Estonia four test-scenarios were developed, not on the basis of story lines or trend extrapolations, but merely to force the model to cater considerable changes in land use, while preserving land use structure. On the basis of these test scenarios the rules were tuned to display realistic behaviour.

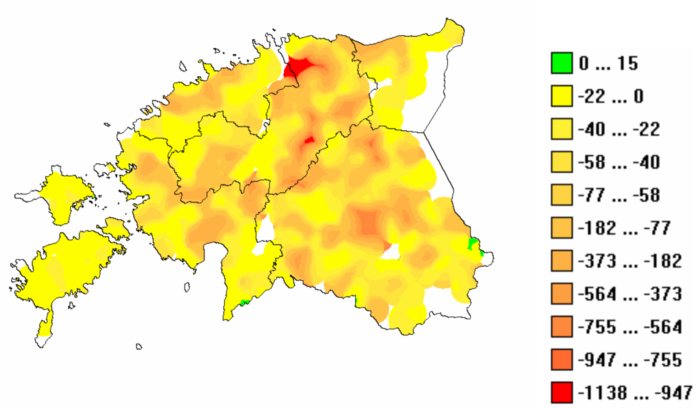
Finally the models were applied on the actual story-line based scenarios. The changes over time that these runs produced were evaluated on the basis of comparison methods that simultaneously account for structure and overlap (Hagen-Zanker, 2005). These methods apply a distance weighted moving window to obtain a spatial account of changes in structure (eg. mean patch size, Shannon diversity and prevalence). The outcome of this validation exercise was that structural land use changes that appear from the change analysis are in line with expectations. Figure 6 displays several of the maps that where used as quantifications of structural change.



a. Change in landscape diversity in Estonia over the period 2000-2030 for scenario 1



b. Change in landscape diversity in Estonia over the period 2000-2030 for scenario 2



c. Difference in patch size of agriculture, generally smaller clusters are found in Scenario 2 compared to Scenario 1.

Figure 5. Moving window based structure comparisons to validate the global behaviour of the Estonia model. The results confirm the expectations that in scenario 1 agricultural specialization will be stronger and the landscape will be less fragmented.

CONCLUSION

A typical approach to calibrating simulations models is to run them for a period in the past and to adjust parameters in order to obtain the best historical fit. Although intuitively sound, there are some drawbacks to this approach, the main one being the dependency on data availability and data quality. If a historical calibration is not possible, then a calibration aimed at the structuring quality of the model can be advised. A calibration on the basis of historical data is already a major challenge because of non linear relations between input and output and the difficulty of quantifying the agreement between model output and actual data. When the evasive concept of structuring quality needs to be analysed it is even more tempting to consider automatic calibration simply impossible and fully rely on expert judgement.

By making our calibration procedure explicit and recognizing the different iterations in it, we could split into a number of sub tasks. The methodologies put forward in this paper are a mixture of automatic (objective) and human judgment (subjective) procedures to deal with sub tasks of the calibration. It is found that although it is not yet possible to automatically calibrate the models we can step-by-step seek further quantitative approaches to replace human judgement. This will not only relieve us of some labour intensive and often boring tasks but also makes the models better transferable and transparent.

REFERENCES

1. **Engelen, G., White, R. and Nijs (de), T.** (2003). Environment Explorer: Spatial support system for the integrated assessment of socio-economic and environmental policies in the Netherlands. Integrated assessment, 4(2), 97-105.

2. **Geurs, K., Hagen, A., Hoen, A. and Wee (van), B** (2003, 6-10 October 2003). Ex post evaluation of Dutch spatial planning and infrastructure policies,. Paper presented at the European Transport Conference, Strasbourg, France.

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

4. **Hagen-Zanker, A.** (2005, accepted), Map comparison methods that simultaneously address overlap and structure, Journal of Geographical Systems

5. **Hagen-Zanker, A., Straatman, B. and Uljee, I.** (2005, accepted), Further developments of a fuzzy set map comparison approach, International Journal of Geographical Information Science

6. **Loon (van), J,** (2004). Towards model-specific automated calibration procedures for the LeefOmgevingsVerkenner model. Unpublished Master's, University Maastricht, Maastricht.

7. **Nijs (de), T., Niet (de), R. and Crommentuijn, L.** (2004). Constructing land-use maps of the Netherlands in 2030. Journal of Environmental Management, 72(1-2), 35-42.

8. **RIKS** (2004). LeefOmgevingsVerkenner: Kalibratie van de LOV. Research report commissioned by RIVM, Maastricht, 1-162

9. **Straatman, B., White, R., and Engelen, G.** (2004). Towards an automatic calibration procedure for constrained cellular automata. Computers, Environment and Urban Systems, 28(1-2), 149-170.

10. **Waveren (van), R. H., Groot, S., Scholten, H., F.C., G. v., Wösten, J. H. M., Koeze, R. D., and Noort, J. J.** (1999). Vloeiend modelleren in het waterbeheer, Handboek Good Modelling Practice (STOWA-rapport 99-05). Lelystad, the Netherlands: STOWA/RIZA.

11. **White, R. and Engelen, G**. (2003, 5-9 September). A Calibration Procedure for Constrained Large Neighbourhood Cellular Automata based Land Use Models. Paper presented at the 13th European Colloquium on Theoretical and Quantitative Geography, Lucca, Italy.

AUTHORS INFORMATION

Alex HAGEN-ZANKER
ahagen@riks.nl
Research Institute for Knowledge Systems
P.O.Box 463
6200 AL Maastricht
The Netherlands

Joyce van LOON
j.vanloon@ke.unimaas.nl
Maastricht University, Quantitative Economics
P.O.Box 616,
6200 MD Maastricht,
The Netherlands

Arjan MAAS
amaas@riks.nl
Research Institute for Knowledge Systems
P.O.Box 463
6200 AL Maastricht
The Netherlands sity

Ton de NIJS
ton.de.nijs@mnp.nl
Netherlands Environmental Assessment Agency
P.O.Box 303
3720 AH Bilthoven
The Netherlands

Bas STRAATMAN
g15bs@morgan.ucs.mun.ca
Memorial University of Newfoundland, Department of Geography
St. John's, Newfoundland
Canada, A1B 3X9

Guy ENGELEN
guy.engelen@vito.be
VITO - Flemish Institute for Technological Research,Centre for Integrated Studies
Boeretang 200
2400 Mol Belgium