Cartographic Scale and Minimum Mapping Unit Influence on LULC Modelling

David García Álvarez
Departamento de Análisis Geográfico Regional y Geografía Física, Universidad de Granada, Granada, Spain

Keywords: Scale, Minimum Mapping Unit, CORINE, SIOSE, LULC Modelling, Dinamica Ego.

Abstract: Two models at two different scales (1:25,000 and 1:100,000) were calibrated using two different Land Use and Land Cover maps at such cartographic scales (SIOSE and CORINE) and with a different Minimum Mapping Unit (0.2-0.5ha and 25ha). Differences between models were assessed through cross-tabulation analysis (quantity and allocation disagreement) and spatial metrics (pattern disagreement). The models results have been very different depending on the scale considered, although most of the disagreement comes from the contrasting input maps. In any case, the scale at which the models were calibrated have proved to influence the pattern modelled and the quantity and allocation of changes.

1 INTRODUCTION

Depending on the considered scale, spatial data can offer different information about the studied features and the relationship between them. In consequence, scale influences any analysis of geographical data, including Land Use and Land Cover (LULC) modelling.

Usually scale is understood as cartographic scale (ratio), extent (map size or study are size) or grain, which is sometimes referred as spatial scale (O’Sullivan and Perry, 2013). The temporal and thematic resolution are also considered part of the concept of scale, together with the Minimum Mapping Unit (MMU) (Castilla et al., 2009), that is, the smallest size area unit to be mapped. A smaller MMU means a more detailed map, whereas a bigger MMU reduces such detail. In the last case, smaller features are not drawn and, consequently, the map representation only focus on the dominant features.

Several papers have addressed the scale influence on LULC modelling, focusing on the grain or spatial resolution (Blanchard, Pontius Jr. and Urban, 2015), extent (Verburg A. Veldkamp, 2004), temporal resolution (Rosa et al., 2015) and, in the case of CA models, neighbourhood size (Pan et al., 2010). However, there is a lack of research about how the cartographic scale and the Minimum Mapping Unit (MMU) of the data vary the model results.

Several studies proved the MMU influence on pattern analysis and landscape metrics calculation (Saura, 2004; Kelly, Tuxen and Stralberg, 2011). This shows how MMU affects GIS analysis and, therefore, the need to evaluate this component of scale.

Figure 1: Asturias Central Area location. Sources: National Topographic Map 1:200.000.

The objective of this paper is to study the effects of cartographic scale and MMU on LULC modelling through the comparison of two models calibrated at two different scales (1:100,000 and 1:25,000). We study the quantity and allocation disagreements as
well as the pattern disagreement consequence of the dissimilar data used in the scenarios generated by the models. From this point forward, when referring to scale we refer to the cartographic scale and MMU.

2 STUDY AREA AND DATA SETS

2.1 Study Area

The test area was the Asturias Central Area, the most dynamic space of Asturias (Spain) (Fig. 1). The main changes are from rural covers to urban and industrial spaces.

2.2 Data Sets

Two LULC maps at two cartographic scales and with different MMU were employed: CORINE (1:100.000; 25ha) and SIOSE (1:25.000; 0.5-2ha) (Fig. 2).

Whereas SIOSE was obtained by photo-interpretation of aerial imagery, CORINE was made from a generalization of SIOSE. Therefore, both maps refer to the same base dates (2005 and 2011) and the differences between them are a result of the generalization process, that is a result of the different scale rules (MMU).

There is not a final land cover map for SIOSE. Its data base gives information about the proportions of every cover that compose every polygon, but there is not a unique label that identifies all polygons. Therefore, we carried out a generalization of that statistical information in a way that the geometry (polygons) is defined by a unique cover (label). This was made through the implementation of translation rules according to the proposal of Delgado Hernández (2016).

To make comparable the two maps, they were reclassified according to the same legend. Although coarser scales usually tie in with simpler thematic resolutions, since our objective is to analyse the influence of cartographic scale and MMU on LULC modelling, we have kept the thematic resolution constant. Otherwise, the results would show the general influence of all components of scale on LULC models. Moreover, both map cartographic scales (1:25.000 and 1:100.000) are regional. Accordingly, both fit well with the proposed legend.

Finally, both data sets were rasterized at 12.5m (SIOSE) and 50m (CORINE) following the criteria proposed by Hengl (2006) in the search for the

Figure 2: Input maps comparison for an example area (Gijón). Sources: SIOSE (2011) and CORINE (2012)
minimum influence of the rasterization on our analysis. That is why we have not used the reference CORINE resolution (100m).

3 METHODS

3.1 Model Calibration and Simulation

Two models were implemented in Dinamica EGO, one for SIOSE (1:25.000 model) and another one for CORINE (1:100.000 model). Dinamica is a recognized stochastic cellular automata model, which is in addition very flexible (Mas et al., 2014). This allowed to set up both models according to the same criteria.

Two functions compose Dinamica EGO: expander and patcher. The expander function models new pixels as an expansion of previous patches, whereas the patcher function models new pixels as a new patch, isolated from previous patches of the same class. More information about these functions and the model architecture can be found in Soares, Cerqueira and Pennachin (2002).

Dinamica Ego models transitions. Therefore, different transitions were selected for each model according to the changes measured by each pair of input maps (Table 1). Only those transitions with a minimum quantity of changes (>10ha) were considered. Like the modelling objective is to study how artificial surfaces expand, there were selected only those transitions which transition to an artificial cover.

Drivers were chosen according to expert criteria (interviews) and literature review. When a correlation greater than 0.5 between two drivers was detected, one of them was removed from the model.

The driving forces included in the model are: roads, train stations, residential and industrial buildings, coastline, leisure facilities, population density, slopes, planning, substratum and industrial ports. When possible, drivers were obtained from sources with similar scales to the implemented models (1:25.000 and 1:100.000).

Driving forces relation with changes was calibrated through the Weights of Evidence method, which is part of Dinamica EGO. The two models were run with the same weights of evidence, according to expert criteria. This is possible because Dinamica allows the user to modify manually the obtained weights.

The model parameters (size and variance of new patches) were established according to real changes (2005-2011). Finally, when some strange or incorrect behaviour was detected, it was corrected manually. Thus, it was applied a manual and expert calibration.

Once the model was calibrated, a simulation was run to the year 2020, which fits well with the short calibration period (six years). Transition rates for the simulation year (2020) are a modification of the rates of change for the calibration period according to real trends of change for the modelled period, as pointed out by experts.

Table 1: In grey, transitions modelled by the two models. In white, transitions modelled by only one.

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction sites</td>
<td>Continuous urban fabric</td>
</tr>
<tr>
<td>Pastures</td>
<td>Discontinuous urban fabric</td>
</tr>
<tr>
<td>Construction sites</td>
<td>Industrial and commercial units</td>
</tr>
<tr>
<td>Pastures</td>
<td></td>
</tr>
<tr>
<td>Construction sites</td>
<td></td>
</tr>
<tr>
<td>Arable lands</td>
<td></td>
</tr>
<tr>
<td>Pastures</td>
<td></td>
</tr>
<tr>
<td>Complex cultivation patterns</td>
<td></td>
</tr>
<tr>
<td>Land principally occupied by agriculture</td>
<td></td>
</tr>
<tr>
<td>Forests</td>
<td></td>
</tr>
<tr>
<td>Natural grasslands</td>
<td></td>
</tr>
<tr>
<td>Scrubland</td>
<td></td>
</tr>
<tr>
<td>Construction sites</td>
<td>Infrastructures</td>
</tr>
<tr>
<td>Forests</td>
<td>Mineral extraction sites</td>
</tr>
<tr>
<td>Scrubland</td>
<td>Dump sites</td>
</tr>
<tr>
<td>Arable land</td>
<td></td>
</tr>
<tr>
<td>Pastures</td>
<td></td>
</tr>
<tr>
<td>Forests</td>
<td>Construction sites</td>
</tr>
<tr>
<td>Dump sites</td>
<td></td>
</tr>
<tr>
<td>Arable land</td>
<td></td>
</tr>
<tr>
<td>Pastures</td>
<td></td>
</tr>
<tr>
<td>Complex cultivation patterns</td>
<td></td>
</tr>
<tr>
<td>Land principally occupied by agriculture</td>
<td></td>
</tr>
<tr>
<td>Forests</td>
<td></td>
</tr>
<tr>
<td>Natural grasslands</td>
<td></td>
</tr>
<tr>
<td>Scrubland</td>
<td></td>
</tr>
</tbody>
</table>

3.2 Data Analysis and Assessment

Disagreements were calculated for the input maps and for the changes simulated, that is without considering the permanent areas. Disagreements for input maps give us information about how the difference of the initial data can explain the results generated by the models.

Quantity and allocation disagreements were analysed through the matrix proposed by Pontius Jr. and Millones (2011). For the pattern disagreement, a series of spatial metrics were calculated through FRAGSTATS 4.2. These are: Number of Patches...
(NP), Area-Weighted Mean Patch Area (AWMPA) and Patch Cohesion Index (PCI). Their selection was based in how much information they provided, that is how well they express the difference between the compared maps.

4 RESULTS

4.1 Quantity and Allocation Disagreement

There is an important difference in the quantity and allocation of classes between the two input maps (SIOSE and CORINE) because of their different scale. Only around the 44% of the area in one map corresponds to the same category in the other map (Fig. 3).

In consequence, each map measures different types and quantity of changes. This has resulted in the consideration of different transitions for the two models (Table 1). Also, like the areas where every class is located are different (25% allocation disagreement), the simulated changes will locate in a different position. Since there are two models which simulate different transitions and the location of the classes where the transition takes place are probably different, there is a low probability that the changes simulated by both models would be similar.

This is what Figure 3 tell us: the only agreement between simulated changes by the two models is due to chance. Depending on the input maps used, the model produces a very different result. The 95% of the changes simulated by the two models are different (Fig. 4).

Figure 4 allows to see the quantity disagreement between the simulated changes depending on the class considered. Each confusion bar is composed by various sections, which represent the proportion of pixels that are allocated to a different class on the other simulation. When the section for any particular class (e.g. continuous urban fabric) is larger in one bar than on the other, there is a quantity disagreement, which is proportional to the difference between the two sections in both bars.

The simulated changes are greater in the 1:25.000 model than in the 1:100.000 model: the size of the confusion bar for permanence is greater in the 1:100.000 model than in the 1:25.000 model (Fig. 4). Hence, regarding the total area simulated as change by both models, in the 1:100.000 model only the 38% of the area is change, whereas in the 1:25.000 model that is true for the 68% of the area. This is because, due to the smaller MMU, SIOSE allows to detect small changes. Whereas only changes over 5ha are drawn in CORINE, SIOSE represents every change bigger than 0.4ha.
The quantity disagreements at the class level are related to the quantity disagreements between input maps. E.g., there is more quantity disagreement for continuous urban fabric in the 1:25,000 model than in the 1:100,000 model because the area of the continuous urban fabric is bigger in the input maps of the first model (SIOSE) than in the input maps of the second model (CORINE).

![Figure 5: Agreement bars per category for simulated changes in the two models (1:100,000 and 1:25,000).](image)

Like the only agreement between simulated changes is due to chance, the allocation disagreement bars at the class level don’t give us extra information (Fig. 5). Most of the area in the bars are disagreements and, therefore, their information corresponds to the disagreements showed by Figure 4.

No allocation agreement is achieved because, whereas the drivers are the same in the two models, the candidate areas to transition are located in different places. The bigger the quantity and allocation disagreement between input maps, the bigger the probability that a same pixel is located in a different place in the two maps and, therefore, the bigger the probability that the candidate pixel to transition would be different in the two models.

### 4.2 Pattern Disagreement

The pattern simulated by the two models is related to the input maps pattern. However, when one compares real changes (2005-2011) to simulated changes, the results show how the model behaves similarly independent of the considered scale.

Despite of the bigger MMU for CORINE (25ha) than for SIOSE (0.5-2ha), there are not big differences in the fragmentation of changes for the 1:100,000 and 1:25,000 models (Table 2). In fact, some classes show a bigger area-weighted mean patch area (polygon mean area corrected by the polygon size) for the model at a finer scale than for the model at a coarser scale. Likewise, the number of changing patches increases with the simulation for the coarser model, whereas it falls for the finer model.

Whereas the effect of the MMU rule is evident for the real changes (input maps cross tabulation), we can’t perceive it in the simulated changes (Table 2).
2). That is because the models work at the pixel level, regardless of the MMU. Since the pixel size is much smaller than the MMU (156m² (1:25.000) vs 0.2-0.5ha (SIOSE) and 0.25ha (1:100.000) vs 25ha (CORINE)), there is not much difference in the model behaviour because of the MMU.

The bigger the contrast between the MMU and the pixel size, the more evident the effects of the model behaviour in the resultant pattern. That is the reason why the 1:100.000 model show a more contrasted behaviour regarding to real changes than the 1:25.000 model.

Like there are not MMU rules, the connection or aggregation of simulated changes (patch cohesion index) is smaller than the aggregation of real changes in both models, although the contrast is again more pronounced for the coarse scale model than for the fine scale one.

5 DISCUSSION

5.1 Input Maps

Input maps play an essential role on the model results. Therefore, knowing the uncertainty of the data sets which we are using is critical in modelling research (Verburg, Neumann and Nol, 2011), since most of the model conclusions will be a consequence of how these maps reflect reality.

The results have showed important differences between input maps (SIOSE and CORINE). This has been a great limitation for the models agreement: the dissimilar quantities and allocations of the same categories turn out on different possibilities to allocate the same transitions.

Working with maps at lower thematic resolutions can help to achieve a higher agreement between input maps. Thus, uncertainty is usually lower at coarser scales since local changes are omitted (Verstegen et al., 2012).

5.2 Model Calibration

The finer the scale considered, the bigger the information that input maps provide. Accordingly, maps at finer scales (SIOSE) show a bigger quantity and types of changes than maps at coarser scales (CORINE). In consequence, transition rates (estimated quantities of changes) and potential transitions (type of changes modelled) are different depending on the scale of the model: the quantity of changes and the number of transitions are bigger for the 1:25.000 model than for the 1:100.000 model.

The provision of more information about reality can be seen as an advantage because we can understand better the dynamics of our study area. However, it is also a limitation when we need to manage tons of complex information to calibrate the model. At finer scales, transitions rarely occur alone and different transitions happen together. The patterns of change are also more complex.

When using fine scale maps, we also need to pay attention to the possible noise in the data. The finer the scale considered, the greater the possibility to find noise (small changes that are not real changes). This noise will influence the results of our model.

Thus, the modeller has to find a balance between data detail and model complexity. More detail but much more complexity is worse than less detail and a very simple model (Wainwright and Mulligan, 2013). The perfect balance would be a manageable complexity level which is in accordance with the detailed added to the model.

Also, depending on the dynamics that the modeller can explain, a finer or coarser scale should be chosen. The 1:100.000 and 1:25.000 models were calibrated using the same driving forces, despite of the fact that their input maps show different dynamics.

The SIOSE maps show small changes, because of the small MMU, that are not correctly modelled since there are not additional drivers to explain them. For certain classes, like dump sites, the 1:25.000 model identifies more changes. However, most of these new changes come from processes that are different to the processes which cause the changes identified by the 1:100.000 model. Like we model in both cases the changes with the same drivers, the 1:25.000 model extrapolates changes from one process to changes from other processes.

If there is only information for the main dynamics of the study area, a coarse scale model, like the 1:100.000 model, is advisable. However, if we can explain also the small changes which are visible in SIOSE, the 1:25.000 model is maybe the best option.

Nevertheless, CORINE maps only reflects the bigger changes in the Asturias Central Area. Because of the scarce dynamics of this area when compared with metropolitan areas or other big cities, the changes showed by CORINE are few and with very specific locations. Therefore, it is difficult to extract an organic growing pattern from that data.

Consequently, depending on the area studied and its characteristics, most of the dynamics can only emerge at specific scales. If the urban sprawling comes from small urban patches, a fine scale map is
needed. However, in the opposite case, a coarse map can be sufficient.

Therefore, every scale has some advantages and limitations. The modeller mission is, as pointed out previously, to find a balance between all the requirements.

5.3 Simulations

The two scenarios generated by the two calibrated models are very different. The agreement between them is only by chance. Most of this difference comes from the contrasted information in the input maps (quantity and allocation disagreement). The stochastic component of the model must have influenced the results also. Nevertheless, the scale and resolution at which the model is set up also have played a role in the resultant scenario.

Grain is considered as spatial scale and it is related to the others concepts of scale: small MMU imply finer spatial resolution than larger units. Then, models at finer scales (1:25,000, 12.5m) simulate more pixels than models at coarser scales (1:100,000, 50m). As a result, the quantity of pixels to allocate is not the same for models at different scales and resolutions: the bigger the resolution, the bigger the quantity of pixels to allocate and the more likely the model to make a mistake. Therefore, the probability to make a mistake is greater for finer scale models than for coarser scales models.

Similar studies which have focused the analysis on the influence of the spatial resolution on LULC modelling have reached similar conclusions (Marceau et al., 2005; Pan et al., 2010).

In addition, there is an incoherence between the model resolution (12.5m and 50m) and the MMU (0.2-0.5ha and 25ha). This makes the pattern of the simulated scenarios more fragmented than the initial pattern, especially when we are working with maps that have big MMU, like CORINE. The model allocates changes as pixels whereas input maps only show changes that meet the MMU. In consequence, the changes allocated by the model will be smaller than the changes measured by the input maps.

Model validation through techniques that compare the generated scenario with the real map for the same date are not completely reliable. Whereas the scenario doesn’t meet with the MMU rules, the reference map does. Consequently, they are never going to show the same information. A real change that only affects a pair of pixels won’t be reflected in the reference maps because it doesn’t comply with the minimum required size. However, the model does can simulate correctly that change.

Dinamica EGO allows the user to achieve the wished simulation pattern through the functions expander and patcher. One can decide how much pixels will be allocated as expansion of previous patches and how much pixels will conform new patches for every simulated category. The user can decide also, for every transition, the mean and variance of the new patches generated.

Although that seems a solution for the proposed problem (Soares-Filho et al., 2003), that did not work for our study area. The mean and variance parameters are only considered when there is an enough variety of candidate areas of different sizes for a specific transition. A candidate area is possible when inside a polygon of the destination category of the transition there is a suitable area, that is an area that, according to the model driving forces, has a value above 0.

Suitable areas for transitions are going to be smaller in models at finer scales (1:25000) than in models at coarser scales (1:100000) because of the respective size of the polygons in each model (MMU). Therefore, models at finer scales (smaller MMU), as far as their input maps are composed by small polygons, find more difficult to vary the desired pattern than models at coarser scales (bigger MMU).

Patch-based models can be a solution for all these problems (Wang and Marceau, 2013).

4 CONCLUSIONS

There is an important source of uncertainty consequence of the chosen scale in LULC modelling, as it is in any GIS analysis.

Firstly, this uncertainty comes from the input maps dissimilarity. LULC data for the same area offer different information depending on the cartographic scale and minimum mapping unit (MMU). The input maps selected, as far as they show a specific representation of a given area, will provide different input parameters to the LULC model.

Making these maps simpler (e.g. decreasing thematic resolution) can reduce the dissimilarity between them at the expense of model complexity.

Secondly, uncertainty comes from the scale at which the model is set up. Modelled patterns are dependent on the spatial resolution, which is linked with the MMU: small MMU imply finer spatial resolution than larger units. The quantity and detail of changes also vary with the scale. Models at finer scales manage more information than models at
coarser scales, although they are more complex to calibrate since the greater the quantity of information, the higher the model complexity. How the modeller manages this complexity can introduce additional uncertainty in the model. Therefore, the user must strike a balance between model complexity and explanatory power.

ACKNOWLEDGEMENTS

This work has been supported in part by project SIGEMOD 2020. BIA2013-34462-P (Spanish Ministry of Economy and Competitiveness and the Feder European Regional Development Fund). The author is also grateful to the Spanish Ministry of Economy and Competitiveness and the European Social Fund for the funding of his research activity (Ayudas para contratos pre-doctorales para la formación de doctores 2014).

REFERENCES


