Cédric Grueau
Jorge Gustavo Rocha (Eds.)

Geographical Information Systems Theory, Applications and Management

First International Conference, GISTAM 2015
Barcelona, Spain, April 28–30, 2015
Revised Selected Papers

Springer

Communications in Computer and Information Science   582

More information about this series at http://www.springer.com/series/7899

Cédric Grueau · Jorge Gustavo Rocha (Eds.)

Geographical Information Systems Theory, Applications and Management

First International Conference, GISTAM 2015
Barcelona, Spain, April 28–30, 2015
Revised Selected Papers

Springer

Identifying Local Deforestation Patterns Using Geographically Weighted Regression Models

Jean-François Mas[(✉)] and Gabriela Cuevas

Centro de Investigaciones en Geografía Ambiental,
Universidad Nacional Autónoma de México,
Antigua carretera a Pátzcuaro 8701,
Col. Ex Hacienda de San José de la Huerta,
58190 Morelia, Michoacán, Mexico
{jfmas,gcuevas}@ciga.unam.mx
http://www.ciga.unam.mx

Abstract. This study aimed at identifying drivers and patterns of deforestation in Mexico by applying Geographically Weighted Regression (GWR) models to cartographic and statistical data. We constucted a nation-wide multidate GIS database incorporating digital data about deforestation from the Global Forest Change database (2000–2013); along with ancillary data (topography, road network, settlements and population disribution, socio-economical indices and government policies). We computed the rate of deforestation during the period 2008–2011 at the municipal level. Local linear models were fitted using the rate of deforestation as dependent variable. In comparison with the global model, the use of GWR increased the goodness-of-fit (adjusted R2) from 0.20 (global model) to 0.63. The mapping of GWR models' parameters and its significance, anables us to highlight the spatial variation of the relationship between the rate of deforestation and its drivers. Factors identified as having a major impact on deforestation were related to topography, accessibility, cattle ranching and marginalization. Results indicate that the effect of these drivers varies over space, and that the same driver can even exhibit opposite effects depending on the region.

Keywords: Deforestation · Drivers · Geographically weighted regression · Mexico

1 Introduction

Mexico, with a total area of about two million square kilometres, is a megadiverse country, but it presents high rates of deforestation [9]. Various studies have attempted to assess land use/cover change (LUCC) over the last decades [21] but there have been few attempts to assess the main causes of deforestation at national level [5,10,24]. Given the complexity of the Mexican territory, the processes of change and its factors are expected to be different depending on

the region. Geographically Weighted Regression (GWR) has been applied in exploring spatial data in the social, health and environmental sciences. The goal of this study is to evaluate the spatial patterns of deforestation with respect to drivers reported to influence LUCC using GWR models.

2   Materials and Methods

2.1   Material

In order to elaborate the geospatial database, the following data were used:

– Maps of tree cover, forest loss and forest gain from the Global Forest Change 2000–2013 database at 30 m resolution [14]. The map of tree cover in 2000 estimates canopy closure for all vegetation taller than 5 m in height and is encoded as a percentage per cell. The map of forest cover loss (2000–2013) is a binary map (loss/no loss) which indicates deforestation, defined as a change from a forest to non-forest state. An additional map (loss year) indicates the year forest loss occurred and is encoded as either 0 (no loss) or else a value in the range 1–13. The map of forest cover gain 2000–2012 indicates a non-forest to forest change within the study period and is encoded as either 1 (gain) or 0 (no gain). There is no map of forest gain year because forest gain was not allocated annually.
– Maps of ancillary data (digital elevation model, slope, roads maps, human settlements, climate, soils, municipal boundaries) [15].
– Socio-economic data from the National Institute of Geography, Statistics and Informatics (INEGI for its Spanish acronym) organized by municipality (Population census for 2005 and 2010) [16,17].
– Index of marginalization 2010 by municipality from the National Council of Population [8].
– Government policies (rural and cattle-rearing subsidies, and protected areas) [7,28].

GIS operations were carried out with Q-GIS [25]. Statistical analysis and graphs were created using R [26]. Geographically weighted regressions were carried out using the packages spgwr Bivand and Yu, [4] and GWmodel [13,19] in R.

2.2   Deforestation Rate Computing and Database Elaboration

In this study, forest area and forest change were estimated at municipality level (2456 municipalities) using the Global Forest Change 2000–2013 database [14]. A map of 2000 forest area was obtained thresholding the tree cover map at 10 %. For each municipality, the 2007 forest area was estimated updating 2000 forest area taking into account forest gain and loss during 2000–2006. As forest gain was allocated over the entire period 2000–2012, we computed a gain area proportional to the 2008–2011 period duration with respect to the 2000–2012 period. Deforestation rate was computed as the average annual proportion of 2007 forest

deforested during 2008–2011 for each municipality. In order to determine which ancillary variables are most likely to be indirect drivers of deforestation, we calculated, for each municipality, various indices describing resources accessibility, population, economic activities, and governmental policies:

– Road density (km of road per km^2 taking into account dirt and paved roads),
– Population density in 2010 (people per km^2),
– Settlements density (number of settlements per km^2),
– Index of marginalization, which takes into account incomes, level of schooling and housing conditions [8],
– Cattle density (heads per km^2),
– Goat density (heads per km^2),
– Average slope (degrees),
– Average elevation,
– Amount of governmental subsidies for agriculture (PROCAMPO, thousand of Mexican pesos per km^2),
– Amount of governmental subsidies for cattle ranching (PROGAN, thousand of Mexican pesos per km^2),
– Proportion of municipality within protected areas.

As the dependant variable is the proportion of 2007 forest cleared during 2008–2011, the variance of this proportion is likely to decrease as it approaches 0 or 1, which is a problem because the regression analysis assumes the error terms to have constant variance. To remove this problem, the proportion was transformed using the Eq. 1 to produce a variance-stabilised rate of deforestation [11].

$$TP = \arcsin[\sqrt{p}] \tag{1}$$

where TP is the transformed rate of deforestation and p is the original proportion.

Explanatory variables were also transformed using logarithm, square, square-root and exponential transformation in order to improve linearity.

2.3   Statistical Analysis

GWR is a local spatial statistical technique for exploring spatial non-stationarity [11]. It supports locally modelling of spatial relationships by fitting regression models. Regression parameters are estimated using a weighting function based on distance in order to assign larger weights to closer locations. Different from the usual global regression, which produces a single regression equation by summarizing the overall relationships among the explanatory and dependent variables (for the whole Mexican territory in that case), GWR produces spatial data that express the spatial variation in the relationships among variables. Maps that present the spatial distribution of the regression coefficient estimates along with the level of significance (e.g. t-values) have an essential role in exploring and interpreting spatial non-stationarity. Fotheringham et al., [11] provide with a

full description of GWR, and Mennis [22] gives useful suggestions to map GWR results.

Collinearity amongst the explanatory variables of a regression model is a well known problem which can lead to a loss of precision in the coefficient estimates [3]. This issue can be more pronounced in local regression models than in global ones due to smaller samples used to calibrate local regression and because some localities may exhibit high collinearity levels when others do not due to spatial heterogeneity [19]. The first stage of the study was a global correlation analysis between explanatory variables using the Spearman coefficient in order to discard highly correlated variables. Local collinearity between explanatory variables was assessed computing diagnostics as local correlations between pairs of explanatory variables, local variance inflation factors (VIFs) and local variance decomposition proportions (VDPs) for each explanatory variable and local condition numbers (CN). According to Lu et al. [19], local collinearity problems likely occur in the local regression model when local correlation values are greater than 0.8 for a given explanatory variable pair, when VIFs and VDPs are respectively greater than 10 and 0.5 for a given explanatory variable and when CN values are greater than 30 for the entire set of explanatory variables. Local correlation between the dependant variable (deforestation rate) and each explanatory variable was also calculated and mapped in order to detect explanatory variables with a high explanatory power. Explanatory variables were selected or discarded in order to reduce collinearity, keeping the ones with higher correlation with the deforestation rate and, as possible, trying to conserve variables describing different aspects of deforestation causes (accessibility, topography, human activities, public policies...).

## 2.4   Geographically Weighted Regression (GWR)

Due to the uneven distribution and size of the municipalities, the weighting function used an adaptive kernel which selects a proportion of the observations (k-nearest neighbours) assigned to each municipality and calculates the weights using a Gaussian model. The optimal size of the bandwidth (in this case the proportion of observations) was found by minimising the Akaike Information Criterion (AIC) [1], which is a model fit diagnostic that takes into account the model parsimony (trade-off between model complexity and prediction accuracy). A map was elaborated for each explanatory variable showing the value of the regressions coefficients (color scaling of the symbol) and statistical significance (hatched mask layer).

3   Results

3.1   Deforestation and Drivers Assessment

As shown in Fig. 1, the rate of deforestation varies greatly over space. The coastal floodplains of the Gulf of Mexico and the southern part of the country exhibits
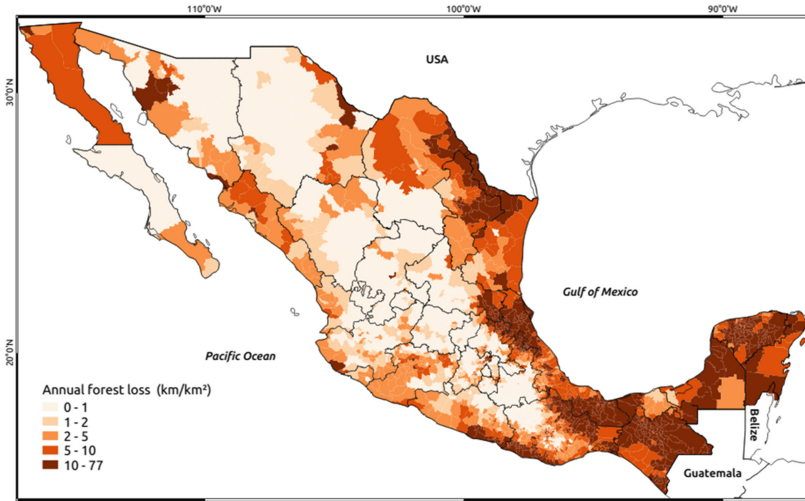
**Fig. 1.** Rate of deforestation during the period 2008–2011.

high rates of deforestation. It is worth mentioning that this rate of deforestation is a gross rate which does not include forest gain. It is also a relative rate, expressed as a proportion of previously existing forest that has been deforested during the period of study. While absolute rates of deforestation are expected among the municipalities with the largest areas of forests, high relative rates can occur when small area of forest are cleared in scarcely forested municipalities. Figures 2, 3 and 4 show the spatial distribution of some of the explanatory variables (road density, cattle density and marginalization).

3.2   Explanatory Variables Selection

The pairs of variables population and human settlements densities and, amount of cattle-ranching subsidies and cattle density present a global Spearman correlation of 0.71 and 0.99 respectively. For this reason, population density and cattle-ranching subsidies were removed from the set of explanatory variables. However, with the nine remaining variables there is still a problem of collinearity (Fig. 5). Therefore, more variables were discarded from the analysis. Global diagnostic give a poor information to carry out the process of selection. For example, slope, which has a low global correlation with deforestation (Table 1) exhibits high local values of correlation (Fig. 6). In most of the cases the relation is negative but in some cases, which correspond mainly to municipalities with large flooded areas or with a high proportion of anthropic cover, the relation is negative (more deforestation in steeper sloping areas). However, slope was dropped out the set of explanatory variables because it presented local collinearity. The analysis of local correlation between the rate of deforestation and the explanatory variables along with collinearity indices enable us to identify the
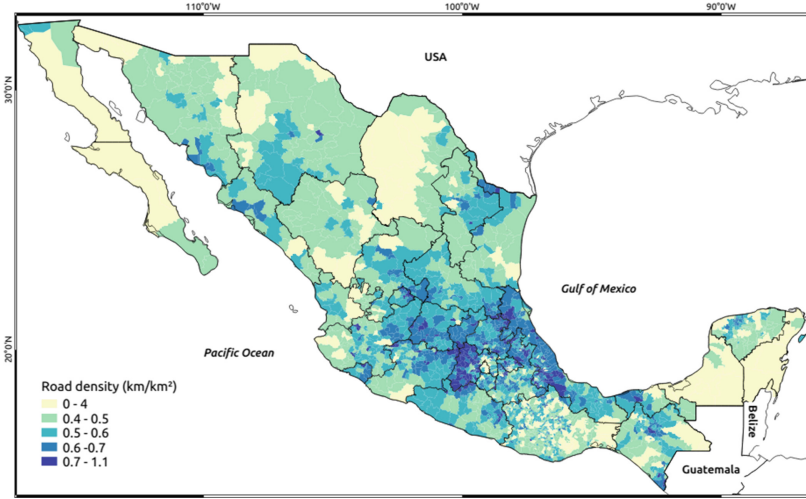
Fig. 2. Road density (km per km^2).
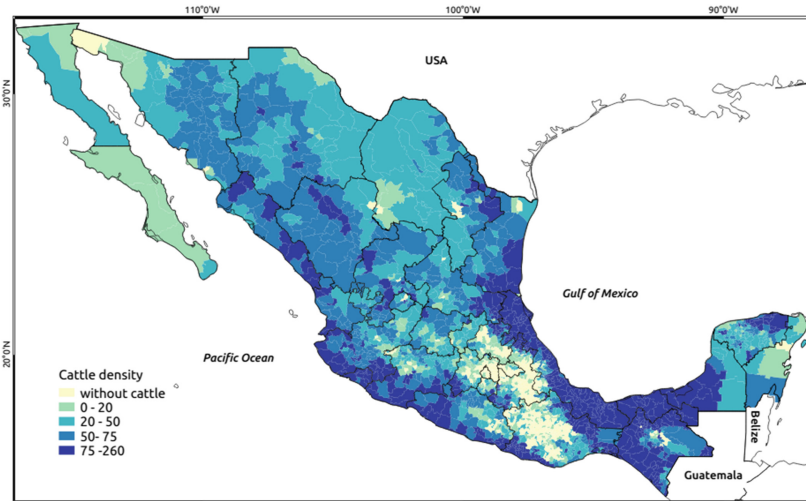
Fig. 3. Cattle density (heads per km^2).

variables with low explanatory power and high collinearity with other explanatory variables.

In order to decrease collinearity at tolerable levels, the number of explanatory variables was finally reduced to five: Index of marginalization, cattle density, elevation, road density, subsidies for agriculture and, protected areas. Figure 7 shows that the Condition Number is inferior to 30 in almost the entire territory.
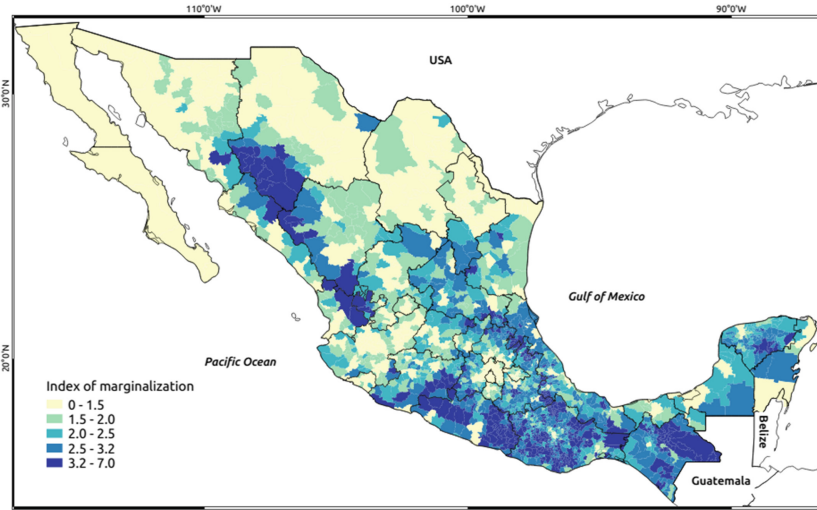
Fig. 4. Index of marginalization 2010 (CONAPO).
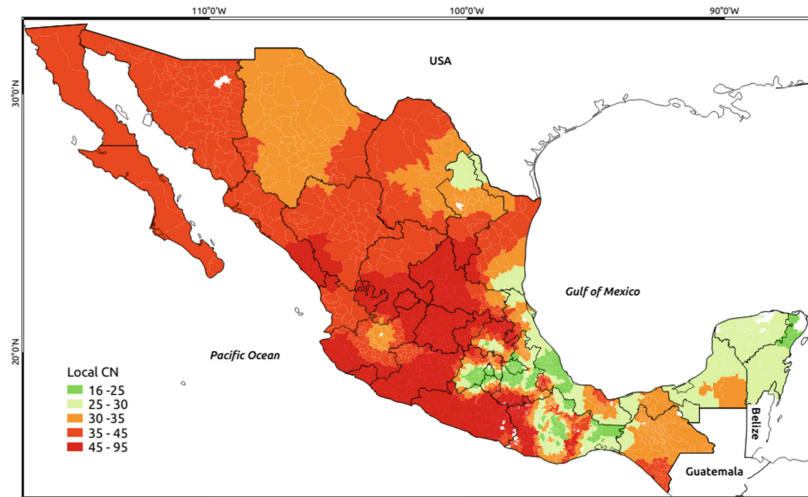
Fig. 5. Local condition number (CN) using the nine input explanatory variables. A value above 30 (red tones) indicates collinearity (Color figure online).

3.3   Geographically Weighted Regression (GWR)

For comparison purpose, a global model was fitted and obtained an adjusted-R^2 of 0.20 (Table 2). A GWR was fitted using the five selected explanatory variables and a weighting function based on a 7 % of the observations (167 neighbours). The use of GWR increased the strength in the relationship in terms of the goodness-of-fit (adjusted R^2) to 0.63.
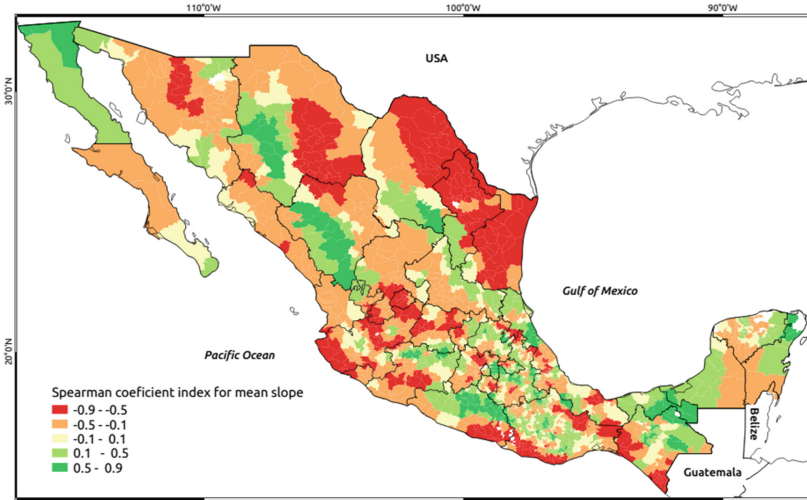
Fig. 6. Local Spearman correlation between the rate of deforestation and the slope. Red and green tones indicates negative and positive relationship respectively (Color figure online).

Table 1. Global coefficient of correlation (Spearman) between the rate of deforestation and explanatory variables.

Explanatory variable	Correlation
Road density	0.05
Population density	0.06
Settlements density	0.11
Marginalization index	0.25
Cattle ranching subsidies	0.36
Goat density	-0.16
Slope	-0.05
Elevation	-0.49
Subsidies for agriculture	0.11
Subsidies for cattle ranching	0.36
Protected areas	-0.06

Road density presents a positive relationship with deforestation in the north of the country and in the southern part, which is an expected behaviour as roads are often reported as a deforestation driver (Fig. 8). In the center of the country, which presents the higher road density values (Fig. 2), this relation is no significant or even negative, likely due to the fact that municipalities with the highest road densities are likely already almost totally deforested and therefore

Table 2. Global model summary.

	Estimate	Std. error	t value	Pr(>t)
(Intercept)	−0.0078	0.0074	−1.05	0.2924
Road density	0.0098	0.0114	0.86	0.3890
Subsidies for agriculture	0.0001	0.0000	3.40	0.0007
Protected areas	−0.0244	0.0067	−3.65	0.0003
Marginaization	0.0139	0.0016	8.78	0.0000
Cattle density	0.0007	0.0000	20.48	0.0000
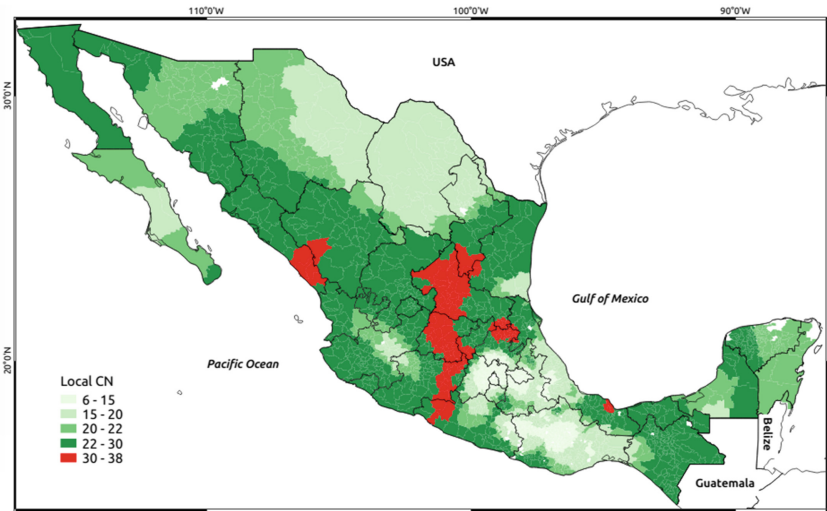
Fig. 7. Local condition number using the five inputs variables. Values above 30 (red), which indicates collinearity, appear only in a small part of the center of the country (Color figure online).

present low rates of deforestation during the period 2008–2011. Cattle density presents a positive relationship with deforestation in all the country (Fig. 9). The marginalization index presents a significant positive relationship with the rate of deforestation in the eastern part of Mexico (Fig. 10). Many studies have associated poverty and deforestation [27]. There are many regions with high level of marginalization where the relationship between marginalization and deforestation is not significant (and in some cases negative). Previous researches have reported that the most conserved natural areas in Mexico are often located in poor rural areas and/or community lands [2,6,10,12,18].
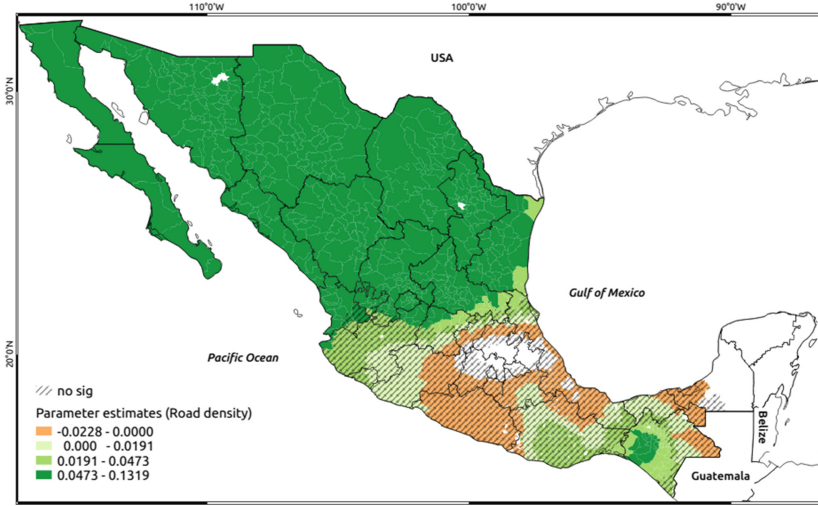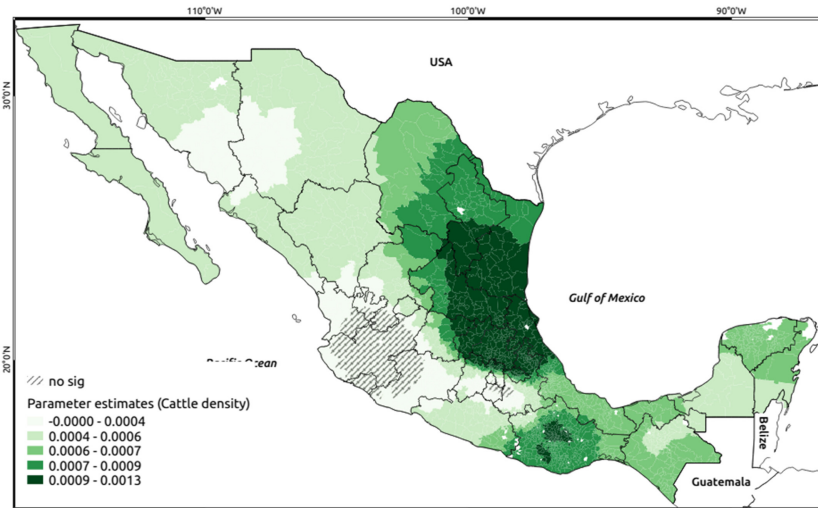
Fig. 8. Distribution of the local parameter estimates associated with road density.

4   Discussion

Some limitations of this study related with input data and with the way information is summarized at municipality level have been identified:

– Change data are based only on a drastic change of land cover (forest loss), they do not consider cover degradation. This factor has to be considered during the

Fig. 9. Distribution of the local parameter estimates associated with cattle density.
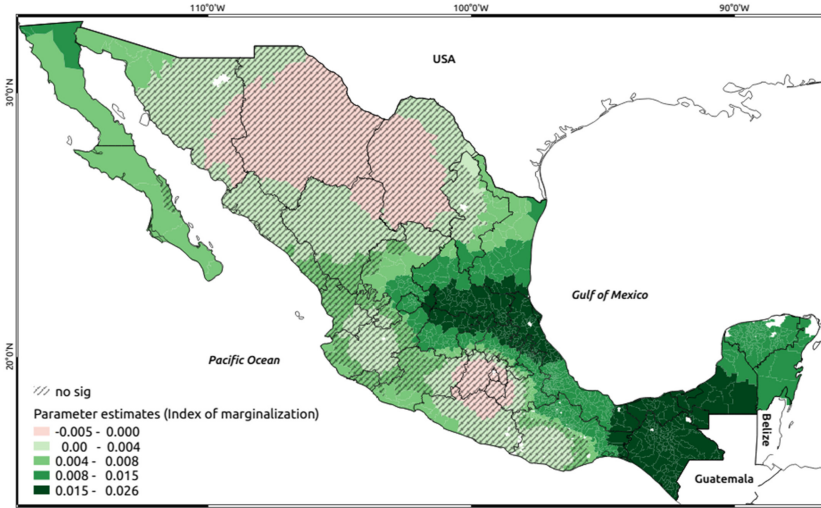
Fig. 10. Distribution of the local parameter estimates associated with the marginalization index.

results interpretation. In Mexico, the rate of deforestation has been decreasing during the last decades and most of the change processes are related with vegetation cover degradation rather than deforestation. Moreover, as the study used a rate of deforestation as dependant variable it does not take into account large extensions of scrublands in the north of the country.

– The rate of deforestation shows change from 2008 to 2011, but the drivers variables (population, marginalization, government subsidies) are from a particular date at different times of the period depending on data availability. The temporal issue cannot be totally addressed due to the lack of information. Additionally, in some cases, it could be interesting to calculate rates of change of these indices. For instance deforestation may be more related to the increase of population density than to population itself.

– Another limitation is the averaging of indices at municipality level which may end up with a figure that does not reflect the actual situation over much of the area. For instance, a municipality with both flat and steep slope areas will present the average value corresponding to moderate slope. Moreover, deforestation can occur in small regions which present very different features from the average figure. This effect, known as the modifiable areal unit problem (MAUP) [23] was evaluated in the case of municipal data for Mexico [20]. The evaluation concluded that, in most of the cases, MAUP did not make large difference to the results. However, counterintuitive relationship between deforestation and slope in some areas (Fig. 6) is maybe due to MAUP. A way to decrease this problem could be to calculate the indices taking into account only the forested area. For instance, average slope of municipality forest area

is used to explain deforestation instead of the slope average over the entire municipality.

– Finally, as depicted by the moderate R^2 of the model, the set of explanatory variables we used did not allow to explain the dependent variable in a satisfactory manner for the entire territory. More drivers have to be taken into account for future analysis.

Other limitations are related with the method used and the deforestation process itself: Deforestation is a complex process that depends on interacting environmental, social, economic and, cultural drivers. Some of them cannot be used into the model because they are unable to be mapped. Moreover, the GWR uses municipality information to explain deforestation but is unable to take into account shifting effects (deforestation in a given municipality is due to the actions from inhabitants from other municipalities) and effects at different scale (as the GWR use the same bandwidth for all the explanatory variables). It is worth noting that some drivers cannot act with very fuzzy spatial pattern or no pattern at all (e.g. global economy effect such as import/export of agriculture goods).

It is likely that the effect of a driver on a given region is related to the time such driver has been shaping the landscape and that different drivers have affect at different temporal and spatial scales, which makes the interpretation of the results difficult. Considering the rate of deforestation during different past periods of time will enable us to analyze the dynamic of deforestation in its temporal and spatial dimensions.

In this study, a special attention has been paid to evaluate and avoid collinearity removing the offending explanatory variables. However, the strategy of removing an explanatory variable is not ideal, particularly when only a local collinearity effect is present, because it limits the number of useful explanatory variables in the regression model due to the high correlation existing between spatial variables. An alternative strategy is to use models with a locally-compensated ridge term [13]. Other alternative GWR models are robust models to identify and reduce the effect of outliers and, mixed models which are able to manage some explanatory variables as constant over space (or stationary) and other as local (non-stationary). In future researches, alternative deforestation rates will be also computed, new explanatory variables such as land tenure will be integrated into the model and, a workshop will be organized to carry out deep interpretation of the results.

5   Conclusions

The results we obtained clearly show the advantages of a local approach such theGWR models over a global one, to evaluate drivers' effect on LUCC processes as deforestation over such a diverse and complex territory as Mexico. The GWR model enabled us to describe spatial relationships between drivers and deforestation. Local models gave a much better explanation of deforestation patterns than the average changes identified by global models such as global regression models.

GWR models can be useful in different types of LUCC study. The exploration of space can help account for differences between regions not captured by standard global measures and thus explain causes of LUCC in different areas. However, the interpretation of the maps of regression estimates is not an easy task and need to be supported by a profound knowledge of the area and of its history. In this regard, GWR offers an attractive visual tool to motivate discussion among specialists from different fields of knowledge (e.g. geographers, anthropologists, economists) and to communicate scientific results with policy-makers and general public. GWR can also be useful in policy design and assessment. Different governmental policies and interventions for deforestation reduction may be appropriate in different areas; depending on local conditions. Therefore the design of locally sensitive policies is likely to be more efficient than nation-wide policies. Alternatively, GWR can be used to evaluate the success of a given policy already in place by determining areas where the intervention was more successful and eventually why.

References

1. Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: Petrov, B., Csaki, F. (eds.) 2nd Symposium on Information Theory, pp. 267–281. Akademiai Kiado, Budapest (1973)
2. Alix-García, J., de Janvry, A., Sadoulet, E.: A tale of two communities: explaining deforestation in Mexico. World Dev. **33**(2), 219–235 (2005)
3. Belsley, D., Kuh, E., Welsch, R.: Regression Diagnostics: Identifying Inuential Data and Sources of Collinearity. Wiley, New York (1980)
4. Bivand, R., Yu, D.: Package spgwr, Geographically weighted regression. http://cran.open-source-solution.org/web/packages/spgwr/spgwr.pdf
5. Bonilla-Moheno, M., Redo, D.J., Mitchell Aide, T., Clark, M.L., Grau, H.R.: Vegetation change and land tenure in Mexico: a country-wide analysis. Land Use Policy **30**(1), 355–364 (2013)
6. Bray, D.B., Duran, E., Ramos, V.H., Mas, J.F., Velázquez, A., McNab, R.B., Barry, D., Radachowsky, J.: Tropical deforestation, community forests, and protected areas in the Maya Forest. Ecol. Soc. **13**(2), 56 (2008)
7. Bezaury Creel, J.E., Torres, J.F., Ochoa-Ochoa, L., Castro Campos, M., Moreno Díaz, N.G.: Bases de datos georeferenciadas de áreas naturales protegidas y otros espacios dedicados y destinados a la conservación y uso sustentable de la biodiversisad en México. The Nature Conservancy (2011). Database on CD. Mexico
8. CONAPO: Índices de marginaci por localidad. http://www.conapo.gob.mx/es/CONAPO/Indice_de_Marginacion_por_Localidad_2010

9. FAO: Global resources assessment. Forestry paper, 140 (2001)
10. Figueroa, F., Sánchez-Cordero, V., Meave, J.A., Trejo, I.: Socioeconomic context of land use and land cover change in Mexican biosphere reserves. Environ. Conserv. **36**(3), 180–191 (2009)
11. Fotheringham, S.A., Brunsdon, C., Charlton, M.: Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley, Chichester (2002)
12. García-Barrios, L., Galván-Miyoshi, Y.M., Valdivieso Pérez, I.A., Masera, O.R., Bocco, G., Vandermeer, J.: Neotropical forest conservation, agricultural intensification and rural outmigration: the Mexican experience. BioScience **59**(10), 863–873 (2009)
13. Gollini, I., Lu, B., Charlton, M., Brunsdon, C., Harris, P.: GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. J. Stat. Softw. **63**(17), 1–50 (2015). http://www.jstatsoft.org/v63/i17/
14. Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G.: High-resolution global maps of 21st-century forest cover change. Science **342**, 850–853 (2013)
15. INEGI: Carta topográfica escala 1: 250000. INEGI, México (2004)
16. INEGI: Conteo de población y vivienda 2005. Indicadores del censo de Población y vivienda. INEGI, México (2005)
17. INEGI: Censo de población y vivienda 2010. INEGI, México (2010)
18. Klooster, D.: Beyond deforestation: the social context of forest change in two indigenous communities in highland Mexico. J. Lat. Am. Geogr. **26**, 47–59 (2000)
19. Lu, B., Harris, P., Charlton, M., Brunsdon, C.: The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. Geospatial Inf. Sci. **17**(2), 85–101 (2014). http://www.tandfonline.com//abs/10.1080/10095020.2014.917453
20. Mas, J.F., Pérez, V.A., Andablo, R.A., Castillo Santiago, M.A., Flamenco, S.A.: Assessing modifiable areal unit problem in the analysis of deforestation drivers using remote sensing and census data. Int. Arch. Photogrammetry Remote Sens. Spat. Inf. Sci. (ISPRS Archives) **XL-3/W3**, 77–80 (2015)
21. Mas, J.F., Velázquez, A., Díaz-Gallegos, J.R., Mayorga-Saucedo, R., Alcántara, C., Bocco, G., Castro, R., Fernández, T., Pérez-Vega, A.: Assessing land/use cover changes: a nationwide multidate spatial database for Mexico. Int. J. Appl. Earth Obs. Geoinformatics **5**, 249–261 (2004)
22. Mennis, J.L.: Mapping the results of geographically weighted regression. Cartographic J. **43**(2), 171–179 (2006)
23. Openshaw, S.: Ecological fallacies and the analysis of areal census data. Environ. plann. A **16**, 17–31 (1984)
24. Pineda-Jaimes, N.B., Bosque Sendra, J., Gómez Delgado, M., Franco Plata, R.: Exploring the driving forces behind deforestation in the state of Mexico (Mexico) using geographically weighted regression. Appl. Geogr. **30**, 576–591 (2010)
25. QGIS Development Team: QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org
26. Core, R., Team, R.: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna (2014). http://www.R-project.org/
27. Rudel, T.A., Horowitz, B.: Tropical Deforestation: Small Farmers and Local Clearing in the Ecuadorian Amazon. Columbia University Press, New York (2013)
28. SAGARPA: Listas de beneficiarios de PROCAMPO y PROGAN (2008–2011)