SPATIO-TEMPORAL REGRESSION MODELS FOR DEFORESTATION IN THE BRAZILIAN AMAZON

Giovana M. de Espindola[a], Edzer Pebesma[b,c1], Gilberto Câmara[a]
[a] National institute for space research (INPE), Brazil
[b] Institute for Geoinformatics, University of Münster, Germany
[c] 52°North GmbH, Muenster

KEY WORDS: land use change, spatial simultaneous autoregression

ABSTRACT:
Deforestation in the Brazilian Amazon has sharply decreased over the past five years. In this study we try to explain the spatio-temporal pattern of deforestation in a selected area by relating data from 2002-2008 to a number of explanatory variables, part of which are related to control actions conducted by the government. We do so by considering the yearly fraction of deforestation for 25 km x 25 km cells, and spatial multiple regression models that incorporate autoregressive components in space and in time, as well as spatial, temporal and spatio-temporal physical and human-induced predictors. The ultimate goal is to evaluate the effect of control actions, and to obtain process knowledge needed for land change models needed to evaluate future actions.

1. INTRODUCTION

As one of the largest tropical forests in the world, the Brazilian Amazon is an area where deforestation affects environmental themes such as biodiversity and greenhouse gas emission with global proportions. After a long period of increase, deforestation in the Brazilian Amazon has sharply decreased over the past five years. Following Aguiar et al. (2007), in this study we try to explain the spatio-temporal changes of deforestation in the Brazilian Amazon by relating yearly data from 2002-2008 to a number of explanatory variables. We do so by considering the yearly fraction of deforestation for 25 km × 25 km cells, and by using spatial multiple regression models that incorporate autoregressive components in space and in time, and predictors that vary over space, over time and over space-time. The goal is to understand the changes in deforestation, and ultimately to understand the effect of control actions and to obtain process knowledge needed for land change models that are developed to evaluate future actions.

2. DATA

The dependent variable is yearly deforestation for 25 km × 25 km grid cells, shown in figure 1. The variability of deforestation is explained by (i) an autocorrelation effect (in space, time, or space-time) and (ii) by external variables. A large number of external predictor variables. Of these predictor variables, some varied only over time (e.g. world market prices), some varied only over space (e.g. distance to river), and some varied over space and time. For the variables that changed over space *and* time, both the initial (2002) value was offered as a prediction, and the temporal change of the time before (e.g., for time step 2003, the change 2003-2002) was offered as predictor. Several of the socio-economic spatio-temporally varying variables were only available at the spatial level of administrative units. In these cases, they were assigned to the grid cell for which the unit was dominant.

Temporal predictors included: price of soy bean and price of meat.

Spatial precitors included: distance to nearest municipality, distance to nearest capital in Legal Amazon Euclidian distance to Sao Paulo, Euclidian distance to nearest port, Euclidian distance to nearest river, Euclidian distance to nearest mineral deposit, Euclidian distance to nearest road, Euclidian distance to nearest timber industry, Percentage of high fertility soils, Percentage of low fertility soils, Percentage of very low fertility soils, Strength of connection to ports through road networks, Strength of connection to Sao Paulo through road networks, Strength of connection

[1] corresponding author, edzer.pebesma@uni-muenster.de

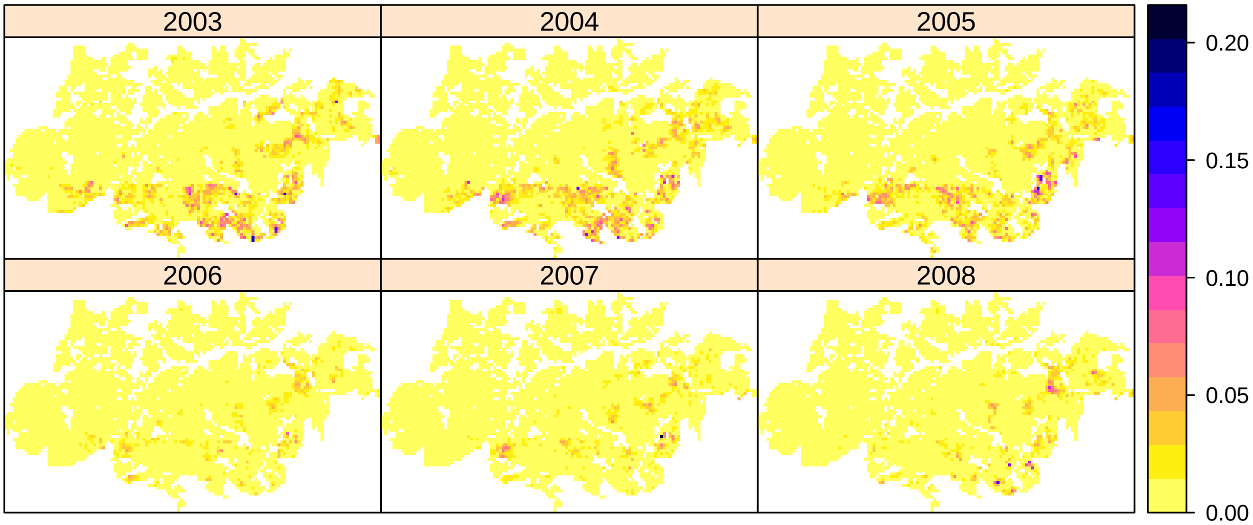

Figure 1: Yearly deforestation rates in the Brazilian Amazon, per year, as fractions of 25 km × 25 km cells, over the period 2003-2008

to Sao Paulo and Recive through road networks, Average temperature for the three driest months, Average precipitation for the three driest months, Seasonal index, Humidity index, Percentage of conservation units in 2002, Total area of soybean in 2002, Total area of sugarcane in 2002, control actions in 2002, population in 2002, and total of exports in 2003.

Spatio-temporal predictors included: change in percentage deforestation (as autoregressive predictor for model 1 only), change in cell percentage of conservation units, change in area of soybeans, change in area of sugarcane, change in control actions, change in population, and change in total exports.

3. METHODS

Regression modelling approximates a dependent variable with n observations $y = (y_1, ..., y_n)'$ to a set of p independent variables $x_j = (x_{1j}, ..., x_{nj})'$ by a linear function,

$$y = \sum_{j=1}^{p} \beta_j x_j + e = X\beta + e$$

where X is the design matrix that has x_{ij} on row i and column j. The regression coefficient vector β is typically estimated by minimizing the residual sum of squares, $e'e$.

Simultaneous autoregression (SAR) models (Cressie and Wikle, 2011) define the residual process $y - X\beta$ to follow an autoregressive process, i.e.

$$Y - X\beta = B(Y - X\beta) + v$$

which can be rewritten as

$$Y = X\beta + (I - B)^{-1} v \tag{1}$$

where v follows a zero-mean normal distribution with covariance matrix $\sigma^2 I$ (i.e., is independent), and B defines which residuals are correlated, and to which degree. Typically, B is sparse, and $B_{ii} = 0$. Non-zero values B_{ij} occur only when Y_i and Y_j are *neighbours*. Additionally, we assume that the non-zero values of B have a single value, which is the parameter that describes the degree of autocorrelation. This value will be called λ: for any non-zero B_{ij}, cells i and j are neighbours and $B_{ij} = \lambda$.

To define spatial neighbours, in this study we used the queen neighbours, meaning the 8 cells adjacent to each grid cell, or less in case of boundary cells or missing valued (or masked) pixels in the neighbourhood.

For a spatio-temporal regression model, where we will denote $y_{[t]} = (y_{1,t}, ..., y_{n,t})$ as the observation in grid cell i and time step $t \in \{1, ..., m\}$. As a first step from purely spatial SAR models towards spatio-temporal SAR models, in addition to the spatial autoregressive effect of the residuals we can take a temporally lagged observation $y_{[t-1]}$ into the regression, as in

$$y_{[t]} = X\beta + \gamma y_{[t-1]} + (I - B)^{-1} v, \;\; t = 2, ..., m \quad (2)$$

where B addresses spatial neighbours only. We will call this **model 1**.

In a second approach, the SAR model (1) is specified for all time steps, but the B matrix not only addresses spatial neighbours $y_{i,t}$ and $y_{j,t}$ with $i \neq j$, but also the two temporal neighbours of $y_{i,t}$, $y_{i,t-1}$ and $y_{i,t+1}$. A simplifying assumption here is that a single autocorrelation coeficient describes the correlation both in space and time. We will call this model **model 2**.

The third model, **model 3** extends model 2 with spatio-temporal neighbours, i.e. residuals $y_{i,t}$ and $y_{j,t+1}$ are correlated when grid cells i and j are neighbours. Again, a single correlation coefficient is fitted to describe correlations between all (spatial, temporal, and spatio-temporal) neighbours. Figure 2 shows the different neighbours defined in models 1, 2 and 3.

Regressions were carried out with the R function `spautolm` in R package `spdep` (Bivand et al., 2008). This function provides maximum likelihood estimation of β and λ, but does not simultaneously estimate λ, γ *and* β using maximum likelihood. One solution to this would be to define neighbours in space *and* time, combined with a weighting factor that defines how neighbouring in space compares to neighbouring in time, in terms of weights, would be a minimum requirement for this to make sense. The solution chosen here was to add the temporal factor to the fixed effects $X\beta$, effectively leading to an more least squares oriented solution.

4. RESULTS

Maps of yearly deforestation for the period 2002-2007 are shown in figure 1. The explanatory variables addressed are, for each grid cell and time step t defined in section 2. Each of the regression models 1, 2 and 3 (section 3) were computed for the full set of prectors. Table 1 lists the regression coefficients for those variables that were found significant for at least one of the three models at the $\alpha = 0.1$ level.

From these results it can be seen that a fair number of predictors is significant, and has similar standardized regression coefficient values, for each of the three models. Interestingly, the two purely temporal (time series) variables are highly significant. This is relatively easy, meat prices for instance gradually increase over the time period considered and can account for the gradual decreas in deforestation rate.

It is also clear from the λ values and the autoregression coefficient for *change in deforestation* $t - 1$ that autocorrelation in space and time is different. This was ignored for models 2 and 3, where a single λ value was fitted.

5. DISCUSSION AND CONCLUSIONS

Building on the work of Aguiar et al. (2007) who looked at spatial regressions, and Espindola et al. (in press) who compared regression models for two moments in time, This paper gives a first step into the direction of directly modelling and explaining *temporal changes* in deforestation for 25 km × 25 km grid cells covering the Brazilian Amazon. We did this by including predictors related to changes in protected areas, changes in amount of cattle, changes in soy bean and sugar cane plantation coverage. The regression model evaluated here considers yearly deforestation as it depends on the very limited set for which spatially distributed time series were available. In addition, it was only evaluated to which extent the change in deforestation depended on the *changes* in each of these independent variables, i.e. the variables as such were not included directly as predictor. As a consequence, a number of effects found significant may result from confounding effects. No pure time series (e.g. market prices) or purely spatial factors (e.g. climate) were included. Improved understanding of the governing processes may be obtained by evaluating a wider range of regression models.

The regression model entertained here (1) was held deliberately simple, and these first results should be interpreted with caution and some reservations. Improvement of these first results might be obtained when (i) transformation of the dependent and/or independent variables improve the linearity of the relationships, (ii) other grid cell

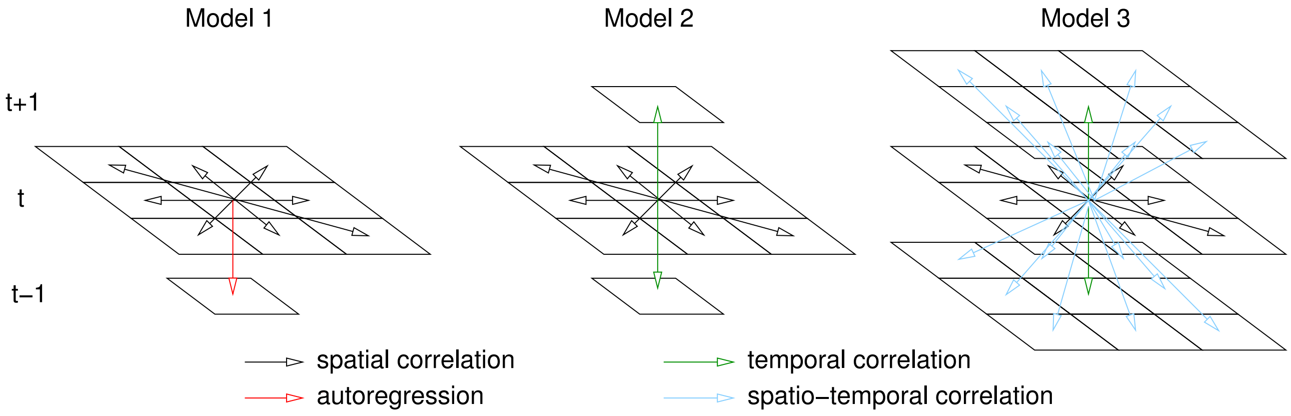

Figure 2: Neighbours addressed for models 1, 2 and 3.

sizes are used than the current 25 km × 25 km cells used here, (iii) more than one time lagged autoregressive terms are used (iv) an estimation procedure is used that can model autocorrelation in space and time separately. Not only variables may be omitted, leading to bias and confounding of those present, it is also possible that some variables need be omitted as they may explain variability for the wrong reason (e.g. the increase in meat price may *not* be the reason for decrease in deforestation rates).

Previous results have shown that protected areas are significant in preventing deforestation in high-pressure areas, and the creation of those areas have been increased as a control policy applied by the Brazilian government.

On the other hand, a debate is growing about the extent of the deforestation as a result of the expansion of cattle (pasture) and soy industry. Most recent analyses suggest that deforestation is driven by the expansion of cattle ranching, rather than soy bean. Soy bean and sugar cane seem to be replacing deforested areas previously under pasture.

References

Aguiar, A. P. D., Camara, G., and Escada, M. I. S. (2007). Spatial statistical analysis of land-use determinants in the Brazilian Amazonia: exploring intra-regional heterogeneity. Ecological Modelling, Volume 209 (Issues 2-4), Pages 169-188

Bivand, R. S., Pebesma, E. J., and Gomez-Rubio, V. (2008). Applied Spatial Data Analysis with R. Springer, NY.

Cressie, N. and C.K. Wikle, 2011. Statistics for spatio-temporal data. Wiley, NY.

De Espindola, G., A.P.D. de Aguiar, E. Pebesma, G. Câmara, L. Fonseca, accepted. Agricultural land use dynamics in the Brazilian Amazon based on remote sensing and census data, Applied Geography, http://dx.doi.org/10.1016/j.apgeog.2011.04.003.

predictor		**model 1**		**model 2**		**model 3**	
change in conservation units	ST	-0.0133	*				
change in percentage soybean	ST	0.0112	.			0.0145	*
change in control actions	ST					0.0078	*
change in deforestation $t-1$	ST	0.4450	***				
price soy bean	T	0.0593	***	0.0845	***	0.0700	***
price meat	T	-0.0736	***	-0.0496	**	-0.0217	**
distance to nearest municipality	S	-0.0177	*				
distance to Sao Paulo	S	-0.0474	.	-0.1487	*	-0.1859	**
distance to roads	S	-0.0437	*	-0.0721	**	-0.0543	.
Strength of connection to Sao Paulo through road networks	S	0.0824	***	0.1081	*	0.1100	*
Seasonal index	S	-0.2126	.	-0.3278	.		
Humidity index	S	0.2102	.	0.3207	.		
Percentage of conservation units in 2002	S	-0.0135	*	-0.0323	***	-0.0316	***
Total area of soybean in 2002	S	-0.0137	*	-0.0330	***	-0.0405	***
control actions in 2002	S	0.0174	***				
population in 2002	S	-0.0145	*	-0.0178	.	-0.0153	.
Percentage of very low fertility soils	S			0.0800	*	0.0753	*
Percentage of low fertility soils	S			0.0411	*	0.0390	.
Percentage of high fertility soils	S			0.0761	*	0.0726	*
Average precipitation for the three driest months	S					-0.0808	.
λ		0.667		0.874		0.895	
σ^2		0.371		0.382		0.460	
Nagelkerke R^2		0.595		0.565		0.508	

Table 1: Standardized regresion model coefficients, and their significance (codes: $0 <$ *** $\leq 0.001 <$ ** $\leq 0.01 <$ * $\leq 0.05 <$. ≤ 0.1). S indicates purely spatial predictors, T purely temporal predictors, and ST spatio-temporally varying predictors. The predictor *change in deforestation* $t-1$ was only offered to model 1, as the other models dealt with autocorrelation in a different way (see figure 2). λ is the estimated autocorrelation coefficient, σ^2 the residual variance.