Spatial autocorrelation

Kevin Li

12/20/2021

Moran's I

Moran's I is defined as

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}(y_i - \overline{y})(y_j - \overline{y})}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

where y_i = observations, $w_{i,j}$ = distance weight, n = number of observations, $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}$

The null hypothesis of no spatial correlation is tested assuming normality of I under this null hypothesis. If the observed value of I is significantly greater than the expected value, then the values of y are positively autocorrelated, whereas if $I_{observed} < I_{expected}$, this will indicate negative autocorrelation.

Here I use weights based on the inverse of the distance between farms. Below I plot the p-values resulting from testing each land cover at each buffer distance. I was expecting that over longer buffer distances the spatial autocorrelation would be higher because some buffers would overlap and therefore values in land cover would be correlated. However, you can see that spatial autocorrelation is significant at most distances except at some low and high values, mostly for pasture.

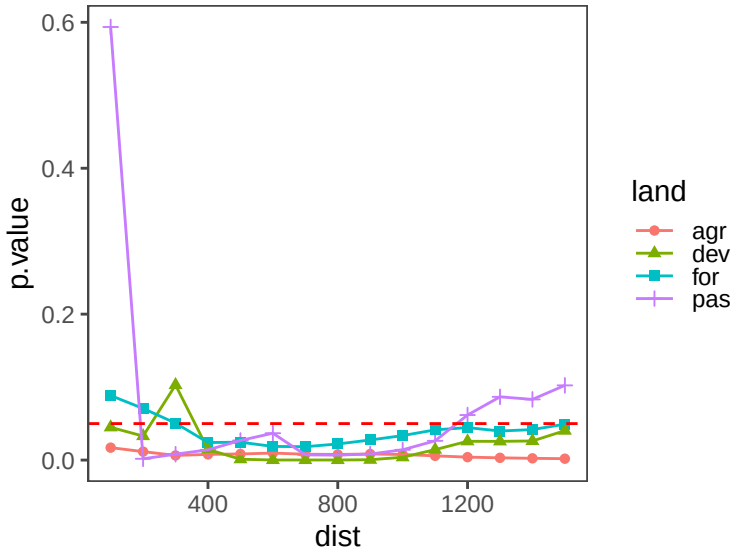

Figure 1 Moran's I p values of % land cover over different buffer distanaces (x axis). Moran's I calculated using the inverse of distance between farms as weight.

Spatial autocorrelation due to local and regional groupings

These were not the results I expected. I thought that land cover taken over larger radii would be more correlated (significant p value), possibly because the buffer radius would be within the distance to the next site/farm. However, most of these farms are well above the range of buffer radii we looked at (see Figure 2).

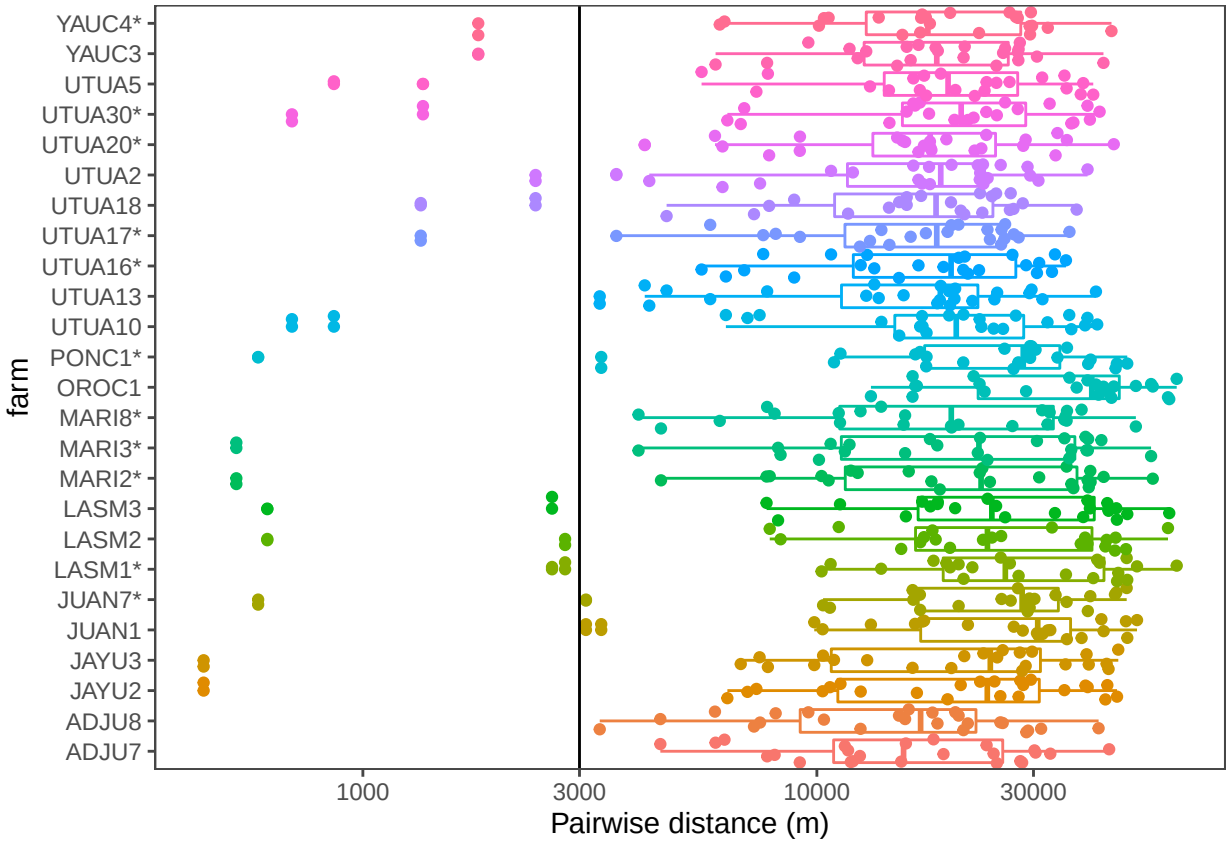

Figure 2 Distribution of pairwise farm distances. Vertical line is 3000m, 2 times the largest buffer distance. Points represent a pairiwse distance between two farms.

On the other hand, I wonder if this is because we are taking a biased sample of the landscape by selecting coffee farms only. We might expect coffee farms to have similar surroundings, so this might be influencing the results. Or, it could be that these correlations are reflecting regional similarities. For example, maybe the landscape in Utuado is made up of similar-sized patches of forest and farms when compared to farms in Adjuntas.

I tried two other weighting schemes instead of the inverse of pairwise distances.

Alternative weight 1: only consider farms within the max buffer radius of each other for correlation

Here, weights between 2 farms can only be 1 or 0: 1 if the farms are within 2x the max buffer distance of each other (3000m), or 0 if they are not.

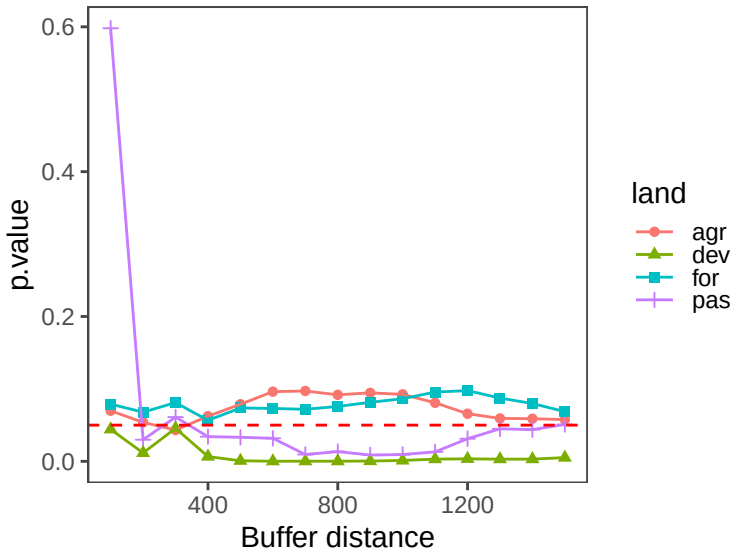

Figure 3 Moran's I significance when only considering farms within 3000m of each other as neighbors.

This suggests that, comparing between farms within 3000m of each other versus all other farms, the developed and pasture cover are more correlated, measured over most buffer sizes (x axis). Forest and agriculture are mostly not significantly more correlated.

Alternative weight 2: compare within the same state vs between municipalities

I wonder if these patterns are reflecting differences between the municipalities. I tried a different weighting where 1's are assigned to between farms of the same state (e.g. the comparision between 2 farms in Utuado gets 1).

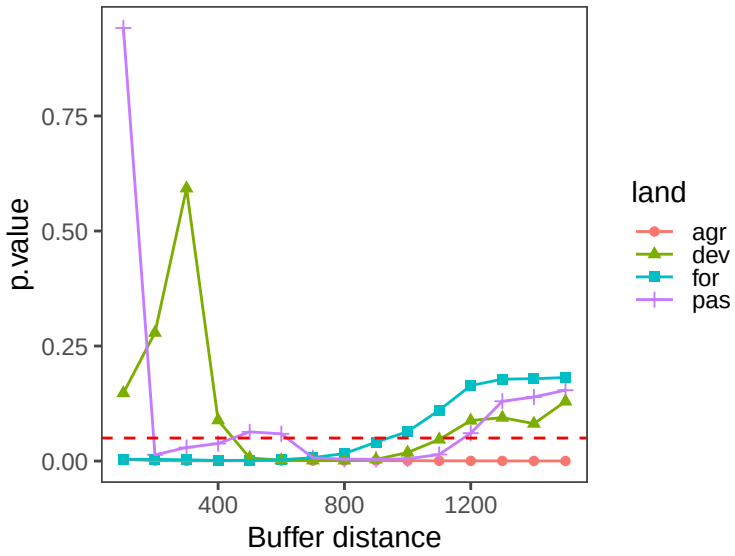

Figure 4 Moran's I significance when only considering farms within the same municipality as neighbors.

At middle buffer distances, farms within the same municipality are similar to each other in terms of % agriculture, development, forest, and pasture. At longer greater buffer distances, these correlations aren't significant comparing between municipalities, except for % agriculture.

Conclusions

This analysis only looked at spatial autocorrelation for the independent variables (land cover). There seems to be some kind of pattern over the farms. Farms nearer to each other have more similar land cover compositions. Shared local landscape within 3000m might be explaining similarities in pasture and development (Figure 3). Regional characteristics (defined as the same municipality) could be explaining similarities in all four land cover types within buffer distances of about 500-1000m (Figure 4).

I didn't look at spatial autocorrelation in the dependent variables, i.e. roya and the natural enemies. This would be the other side of this and maybe would be even more important. On the other hand, spatial autocorrelation in these things could be reflecting some biological or seasonal dynamics, especially if they don't match up with the landscape variables.