Looking Down on the Suburbs: Justification for transition matrices

What's the business with transition matrices?

- I want (Diego wants) a simple formal quantitative "proof" of how cities change locally (At the pixel level) yet retain a stable spatial distribution (Or, alternatively, to check why the spatial distribution is not stable)
- I want to construct counterfactuals of the form "What would have happened in terms of distributions of development intensity and sprawl index if the set of cities A had been constructing like the set of cities B, in the period of study?"

Simple approach: "Pseudo-markovian" transition matrices

Imagine we have an arbitrary set of pixels G, this might be a city or some other geography, or it might be a subset of this geography under other conditions. To characterize the process that this set underwent between 2001 and 2019, we might classify the set of pixels G into V bins of values of said pixels (Values of development intensity or sprawl index, for example). Then, we can compute a transition matrix $\underbrace{\begin{bmatrix} P_{1',1} & ... & P_{1',V} \\ \vdots & \ddots & \vdots \\ P_{V',1} & ... & P_{V',V} \end{bmatrix}}_{T_{2001-2019}}$, where each element $P_{v',v} = P(i_{2019} \in v' | i_{2001} \in v)$.

Then,

$$\underbrace{\begin{bmatrix} P_{1',1} & ... & P_{1',V} \\ \vdots & \ddots & \vdots \\ P_{V',1} & ... & P_{V',V} \end{bmatrix}}_{T_{2001-2019}} \times \underbrace{\begin{bmatrix} P_1 \\ \vdots \\ P_V \end{bmatrix}}_{Hist_{2001}} \approx \underbrace{\begin{bmatrix} P'_1 \\ \vdots \\ P'_V \end{bmatrix}}_{Hist_{2019}}$$

This is all under the assumption that $\{\forall \text{ pixels } i, j \in G : P_{v',v}(i) = P_{v',v}(j)\}$ (Homogeneity of transition process across pixels)

Building a simple counterfactual out of this

Imagine we have two sets of pixels, G_1 and G_2 (Say, two different geographies) each with its own process $T_{2001-2019}(G_k)$. Then, we could build a counterfactual $Hist_{2019}(G_k, T(G_{-k}))$ by applying the transition process of set G_k to the initial histogram of the set G_{-k}.

The goal is to obtain the counterfactual distribution of values of the variable of interest (Development intensity, sprawl) if set G_k "behaved" like the set G_{-k}.

It is obvious that this is severely flawed, because $P_{v',v}(i)$ depends on many observed characteristics of pixel i (That's why I need a proper model) but I gave it a try anyways. Some obstacles:

- In order for the transition probabilities to be comparable, the size of P_1 has to be equal $\forall$ sets G_k. The reason is that we can always take more empty space around a geography and consider it to be part of the set, so that the transition probability of unbuilt pixels would be "artificially lowered"
- In order for the transition probabilities of unbuilt pixels to be comparable, they have to be "buildable" pixels. For this reason, following Saiz (2010) I exclude pixels covered by water (Obviously) and pixels with a terrain ruggedness index higher than 50.

Pixel selection is done in the following way: For each city and year, I delimit "urbanized" areas as those with values of the sprawl index of 15% or less. Then, I draw a buffer around these areas adjusting the radius to create a city delimitation where, among pixels which are not covered by water and the ruggedness index is less than 50, the proportion of completely unbuilt pixels (Those in the lowest bin) in the given year is of 45%.

For the counterfactuals I have done, I consider the following sets: Pixels in the cities in the bottom quartile of the Wharton Regulation Index, and cities in the top quartile of the Wharton Regulation index.