Areal Wombling

Sameer Singh Rawat, LMU

Overview

1. Introduction - What is wombling?
2. Crisp Vs Fuzzy Wombling
3. Crisp Areal Wombling
    a. Traditional approach
    b. Hierarchical Modelling approach
    c. Traditional Vs hierarchical approach
    d. Results

Introduction

- Wombling (also known as edge detection or barrier analysis) refers to the techniques of identifying edges or boundaries of rapid change.
- It is named after William H. Womble (statistician).
- Uses - Environmental Sciences (to detect boundaries between different ecological zones). Public health (to detect boundaries where disease rate changes significantly).
- Two type of wombling - Crisp areal wombling and Fuzzy areal wombling.

Crisp Vs Fuzzy Wombling

1. Boundaries are sharp and clear identifying the regions of significant change.
2. A threshold c is fixed and boundary is considered if the Boundary likelihood value (difference between the regions) exceeds this threshold c.
3. It can be used in fields for understanding significant change of attribute like detecting significant change of disease rate or significant change of environmental or ecological variables.

1. It allows for gradual change between the regions hence boundaries can be unclear.
2. It makes use of probability to assign likelihood of boundary existing.
3. It can be used to understand gradual change of attribute values.

Crisp Areal Wombling

Crisp Areal wombling

- In this paper we will be focussing on crisp areal wombling.
- For this we have taken the dataset for covid cases occurring in USA in April 2020.
- Data contains the deaths recorded and confirmed cases in April 2020 and we have taken Standardized Death Rate (SDR) as a measure to investigate which are the states where covid control measures differ significantly perhaps indicating the scope of improvement in such states.
- Also for performing bayesian hierarchical modelling we have included the covariates such as number of Hospital beds per 1000 people and percentage of people with age above 65. And this data is taken for year 2019 to avoid any biases that might have been caused due to covid measures or the changes in these figures that might have occurred after April 2020.

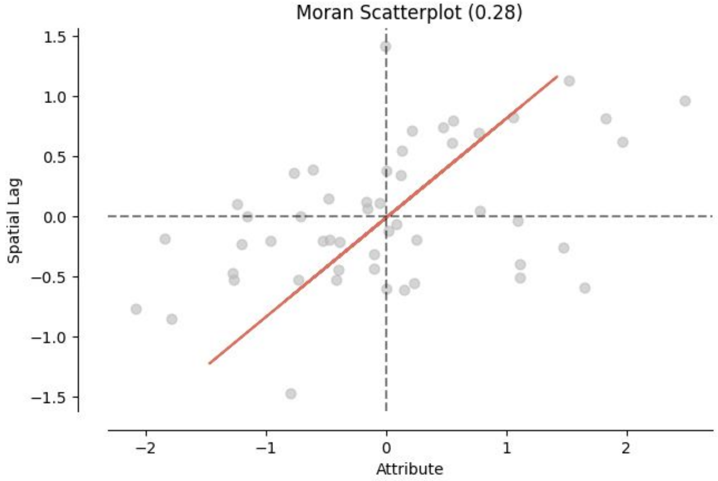
Spatial Autocorrelation by Moran’s I

- This shows that there is moderate positive spatial autocorrelation among the values of SDRs. This means that the states with similar value of SDRs are likely to be close to each other. Also the statistic is highly significant.
- Moran scatterplot shows the relationship between variable and its spatial lag. Spatial lag is the average value of variable in its neighbouring region.

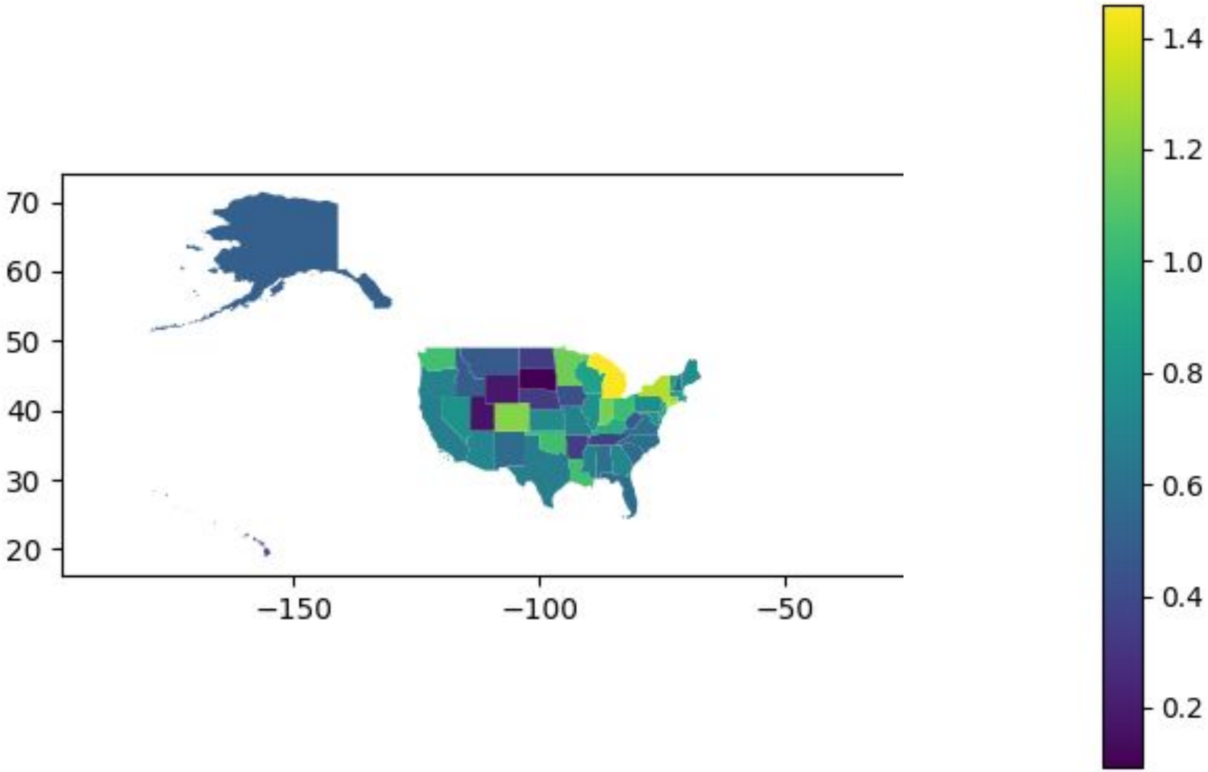
```python
# Conduct Moran's I test for spatial autocorrelation of SDRs
moran_cases = Moran(gdf['SDR'], w)
print("Moran's I for SDR:", moran_cases.I)
print("Moran's I p-value:", moran_cases.p_sim)
```

✓ 0.0s

```
Moran's I for SDR: 0.28029218285750596
Moran's I p-value: 0.005
```

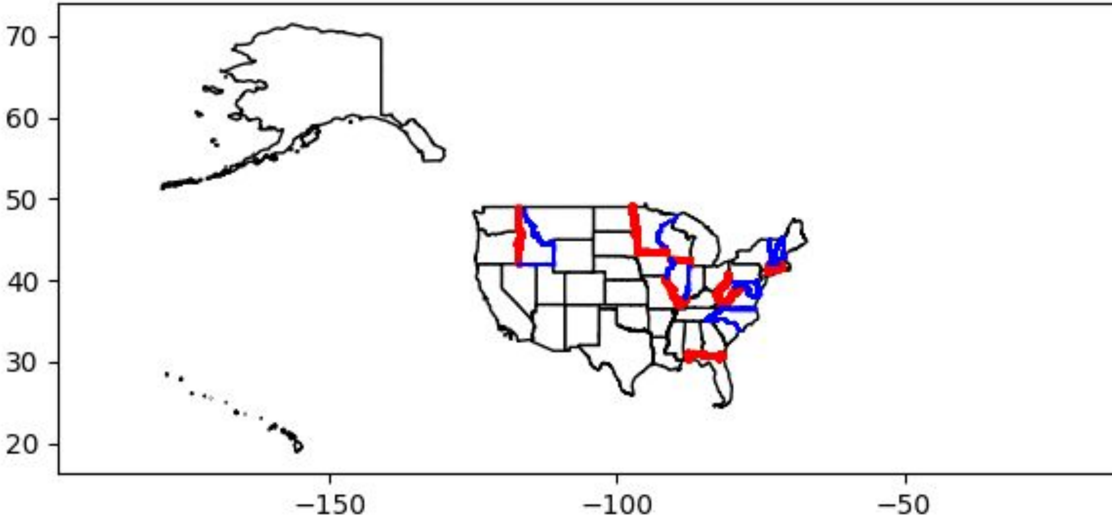


Layman’s approach - SDR plot



Traditional Crisp Wombling

1. Calculate difference between SDRs between each state and its neighbours.
2. Find the 50 percentile and 80 percentile of these boundary differences and mark the boundaries of the states whose boundary likelihood value exceed 80 percentile value with red and with blue where boundary likelihood value exceeds 50 percentile value.



Hierarchical Modelling approach

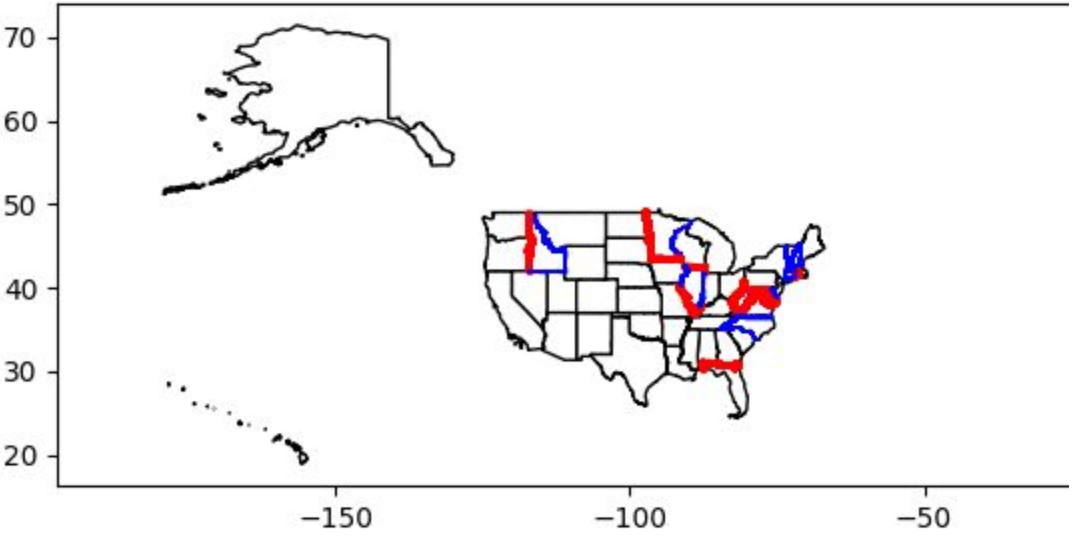
1. Model Specifications -

$$Y_i \sim \text{Poisson}(\mu_i)$$
$$\log(\mu_i) = \log(E_i) + X_i^T \beta + \phi_i$$
$$\phi_i \mid \phi_{-i} \sim \text{ICAR}$$

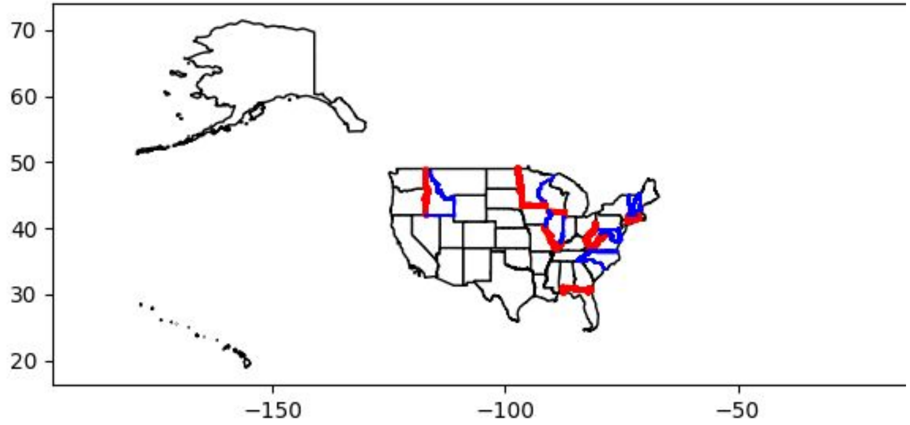
2. Markov chain Monte Carlo (MCMC) samples $\mu_i^{(g)}$ g=1,2,...,1000
3. SDR is calculated by $$Z_i = \frac{\mu_i}{E_i}, \quad i = 1, \ldots, 51$$
4. BLVs are obtained by $D_{ij} = |Z_i - Z_j| \quad \text{for all } i \text{ adjacent to } j$
5. $$\bar{E}(D_{ij}|y) = \frac{1}{1000} \sum_{g=1}^{1000} D_{ij}^{(g)} = \frac{1}{1000} \sum_{g=1}^{1000} |Z_i^{(g)} - Z_j^{(g)}|$$

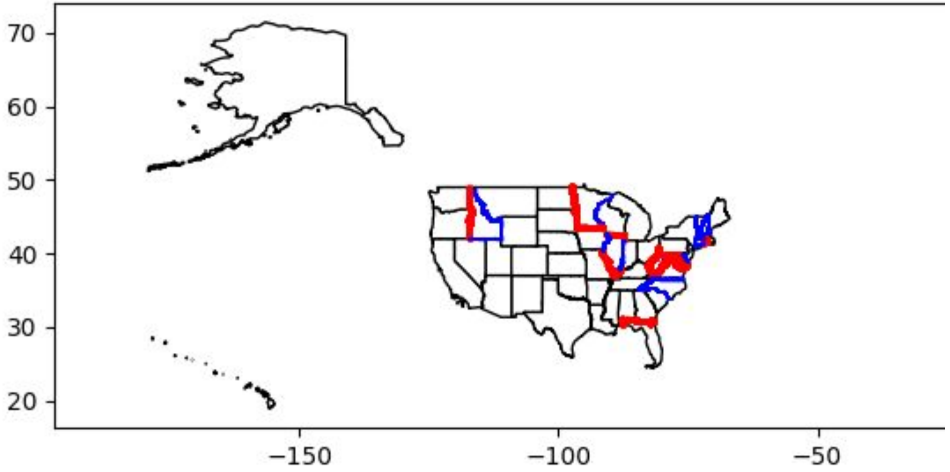
Hierarchical Modelling approach (contd.)

6. Find the 50 percentile and 80 percentile of these boundary differences and mark the boundaries of the states whose boundary likelihood value exceed 80 percentile value with red and with blue where boundary likelihood value exceed 50 percentile value.



Traditional Vs Hierarchical Approach





Results

The states of Minnesota, Colorado, Louisiana, Oklahoma, Michigan, Nevada, Kentucky, New York, Arizona, Washington, Connecticut, Ohio are the states where the deaths to confirmed ratio is significantly different than their neighbouring states, hence signifying probable scope for improvement in healthcare facilities.

```
sorted.head(100)
```

✓ 0.1s

	State1	State2	boundary_diff	significant_20	significant_50	greater_SDR
16	Minnesota	South Dakota	1.060156	True	True	Minnesota
74	Utah	Colorado	1.042117	True	True	Colorado
84	Colorado	Wyoming	1.018903	True	True	Colorado
58	Nebraska	Colorado	0.872592	True	True	Colorado
14	Minnesota	North Dakota	0.822343	True	True	Minnesota
17	Minnesota	Iowa	0.727532	True	True	Minnesota
67	Louisiana	Arkansas	0.701225	True	True	Louisiana
87	Oklahoma	Arkansas	0.695042	True	True	Oklahoma
9	Illinois	Michigan	0.676979	True	True	Michigan
46	New Mexico	Colorado	0.636852	True	True	Colorado
76	Utah	Nevada	0.619074	True	True	Nevada
90	Tennessee	Kentucky	0.592891	True	True	Kentucky
23	Rhode Island	New York	0.590675	True	True	New York
54	Wisconsin	Michigan	0.575721	True	True	Michigan
77	Utah	Arizona	0.554301	True	True	Arizona
26	Idaho	Washington	0.550048	True	True	Washington
65	Pennsylvania	New York	0.547764	True	True	New York
22	Rhode Island	Connecticut	0.547129	True	True	Connecticut
2	West Virginia	Ohio	0.544593	True	True	Ohio
86	Colorado	Arizona	0.487816	True	True	Colorado
47	New Mexico	Oklahoma	0.469465	True	True	Oklahoma
3	West Virginia	Kentucky	0.462867	True	True	Kentucky
85	Colorado	Kansas	0.457784	True	True	Colorado

Thank you!