Spatio-temporal Clustering for Geographic Events

Prepared by Evgeny Noi

Spatial Moran's

$$I = \frac{n \sum \sum w_{ij} (a_i - \bar{a})(a_j - \bar{a})}{\sum \sum w_{ij} (a_i - \bar{a})(a_j - \bar{a})}$$

where w_{ij} is a spatial weight matrix of 0/1.

$$I_i = \frac{z_i \sum_j w_{ij} z_j}{\delta^2}$$

where δ^2 is variance of a and $z_i = (a_i - \bar{a})$.

Extensions to *time*

- Assume stationarity
- Empirical estimations based on data
- In Betazzon (2003) solved a set of 8 equations, where coefficients were set up as matrices in matrices
- Joining spatial and temporal matrices into one (multiplicative)
- Subject to scaling effects

Difficulties in implementing *space-time* autocorrelation

- Integrating different units in space and time
- how the different levels of non-stationarity in spatial data and in temporal data may exist
- spatial and temporal trends of geographical events may not be symmetrical or always be separable

Typical *space-time* representation

- Binary matrices based on threshhold (0/1)
- Inverse of distance ($1/d_{ij}, 1/t_{ij}$)

Alternatives: clustering

- Non-parametric spatio-temporal clustering (aka *Scan statistic*). Iterative in nature - repetitive computations.

Global and Local Temporal Moran's I

$$I_t = \frac{n \sum \sum w_{ij}(t_i - \bar{t})(t_j - \bar{t})}{W \sum (t_i - \bar{t})(t_j - \bar{t})}$$

$$I_{ti} = u_i \sum w_{ij} u_i$$

where $u_i = \frac{(t_i - \bar{t})}{\delta_t^2}$

Spatio-Temporal Moran's I

$$I_{st} = \frac{n \sum \sum w_{ij} t_{ij} (a_i - \bar{a})(a_j - \bar{a})}{\sum \sum w_{ij} t_{ij} (a_i - \bar{a})(a_j - \bar{a})}$$

where $t_{ij} = \left| \frac{1}{t_i - t_j} \right|$, and is binary depending on the threshold

$$I_{sti} = \frac{z_i \sum v_{ij} z_i}{\delta^2}$$

Assumptions

- Normality - the attribute value of each geographic event is considered to be drawn from a normal distribution (as well as timing)
- Randomization - equally likely random permutation of n events (Getis & Ord, 1981)

Microblogs autocorrelation (normality)

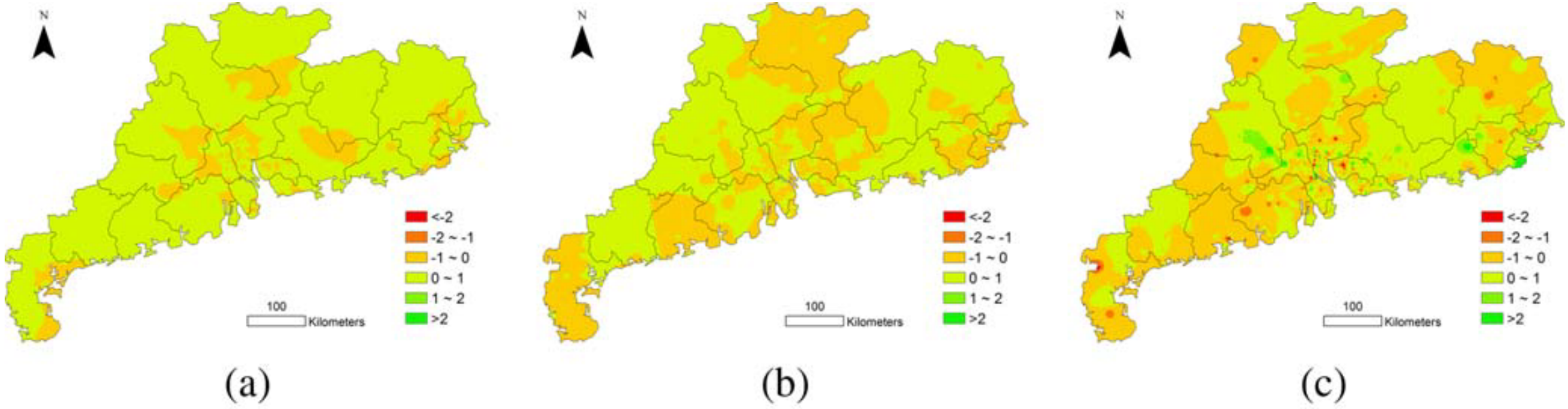

Figure 2. Distribution of Z scores under normality assumption based inversed distance. (a) Spatial autocorrelation, I. (b) Temporal autocorrelation, I_t. (c) Spatiotemporal autocorrelation, I_{sti}.

7 days temporal window

Table 1. Comparison Among Different Forms of Moran's Index Using Microblogs on Dengue Fever in Guangdong in 2014 Threshold of Date with 7 Days

	classic			t in place of a			$\sum w_{ij}t_{ij}$			$\sum w_{ij} \sum t_{ij}$		
Case#	I	Z_N	Z_R	I_t	Z_N	Z_R	I_{UTI}	Z_N	Z_R	I_{ITI}	Z_N	Z_R
$w_{ij}=1/d_{ij}$	0.063	**4.634**	**4.689**	0.080	**5.877**	**5.877**	0.099	**4.066**	**4.115**	0.120	**4.062**	**4.111**
$w_{ij}=1$ *if* $d_{ij} < 1$ km	−0.001	**−3.056**	**−3.092**	−0.001	**−4.083**	**−4.083**	0.000	0.844	0.853	0.000	0.844	0.852
$w_{ij}=1$ *if* $d_{ij} < 2$ km	−0.001	**−2.525**	**−2.554**	−0.001	**−2.277**	**−2.277**	0.000	0.890	0.899	0.000	0.890	0.899
$w_{ij}=1$ *if* $d_{ij} < 3$ km	−0.001	**−2.761**	**−2.791**	−0.001	**−2.980**	**−2.980**	0.000	0.675	0.682	0.000	0.675	0.682
$w_{ij}=1$ *if* $d_{ij} < 4$ km	−0.001	**−2.723**	**−2.752**	−0.001	**−4.162**	**−4.162**	0.000	0.693	0.701	0.000	0.694	0.701
$w_{ij}=1$ *if* $d_{ij} < 5$ km	−0.001	**−2.559**	**−2.584**	−0.001	**−4.490**	**−4.490**	0.000	0.673	0.680	0.000	0.673	0.680

Note: Bold text indicate statistical significance at the level of: $z > 1.96$ or $z < -1.96$.

14 days temporal window

Table 2. Comparison Among Different Forms of Moran's Index Using Microblogs on Dengue Fever in Guangdong in 2014 Threshold of Date with 14 Days

	classic			t in place of a			$\sum w_{ij}t_{ij}$			$\sum w_{ij} \sum t_{ij}$		
Case#	I	Z_N	Z_R	I_t	Z_N	Z_R	I_{UTI}	Z_N	Z_R	I_{ITI}	Z_N	Z_R
$w_{ij}=1/d_{ij}$	0.063	**4.634**	**4.689**	0.080	**5.877**	**5.877**	0.086	**4.854**	**4.912**	0.094	**4.851**	**4.909**
$w_{ij}=1$ *if* $d_{ij}<1$ km	−0.001	**−3.056**	**−3.092**	−0.001	**−4.083**	**−4.083**	0.000	1.046	1.054	0.000	1.045	1.053
$w_{ij}=1$ *if* $d_{ij}<2$ km	−0.001	**−2.525**	**−2.554**	−0.001	**−2.277**	**−2.277**	0.000	1.097	1.106	0.000	1.097	1.105
$w_{ij}=1$ *if* $d_{ij}<3$ km	−0.001	**−2.761**	**−2.791**	−0.001	**−2.980**	**−2.980**	0.000	0.877	0.884	0.000	0.878	0.884
$w_{ij}=1$ *if* $d_{ij}<4$ km	−0.001	**−2.723**	**−2.752**	−0.001	**−4.162**	**−4.162**	0.000	0.863	0.869	0.000	0.863	0.870
$w_{ij}=1$ *if* $d_{ij}<5$ km	−0.001	**−2.559**	**−2.584**	−0.001	**−4.490**	**−4.490**	0.000	0.666	0.672	0.000	0.667	0.672

Note: Bold text indicate statistical significance at the level of: $z>1.96$ or $z<-1.96$.

Microblogs autocorrelation (randomization)

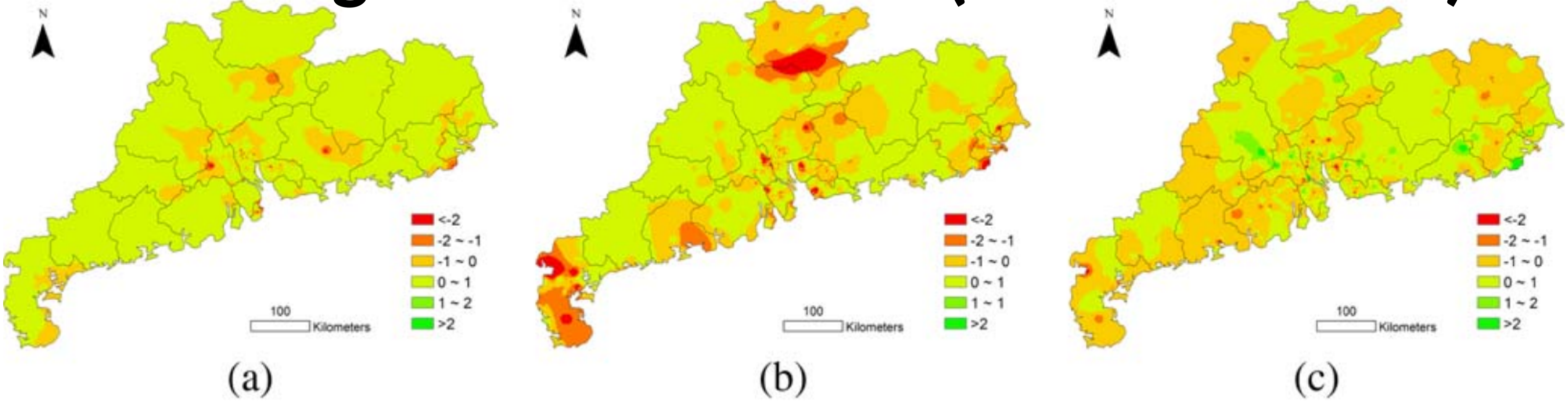

Figure 3. Distribution of Z Scores under Randomization Assumption Based Inversed Distance. (a) Spatial autocorrelation, I. (b) Temporal autocorrelation, I_t. (c) Spatiotemporal autocorrelation, I_{sti}.. [Color figure can be viewed at wileyonlinelibrary.com]

Conclusions

- Different results under normality and randomization
- Multiplicative nature of geographic events
- Time dimension alone cannot explain spatial autocorrelation (temporal effects in geo phenomena are not constant over space)

Critiques

- The authors assessed st autocorrelation on point pattern data only, but not on areal data (which might yeild different results, especially for different contiguity metrics)
- *spatio-temporal* autocorrelation is only significant for $1/d_{ij}$ and not for binary threshholded variables. Otherwise there could be **too many** zeroes.

Questions? 

Evgeny Noi

> “Make Geography Great Again”