第六届全国艾滋病学术大会 杭州 2019年10月15-18日

Modeling Spatio-temporal Variation of the HIV Epidemic

徐成东
xucd@Lreis.ac.cn
中国科学院地理科学与资源研究所

吕毅
中国疾病与预防控制中心

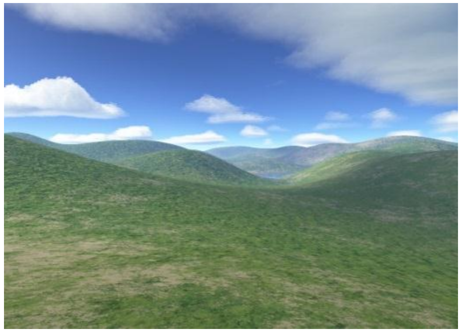
Outline

1. Nature of Spatial Data
2. Bayesian Spatio-temporal Hierarchical Model
3. Study Data
4. Results and Interpretation
5. Conclusions

Nature of Spatial Data

Spatial Autocorrelation

low



high

maximum

The coincidence of locational and attribute similarity
(Anselin, L. 1988. Spatial econometrics. Methods and models. Dordrecht: Kluwer)

Analogous neighborhood characteristics, similar socio-economic characteristics of their residents, and the quality of services (Dubin, R. 1992. Spatial autocorrelation and neighborhood quality. Regional Science and Urban Economics 22: 433-452).

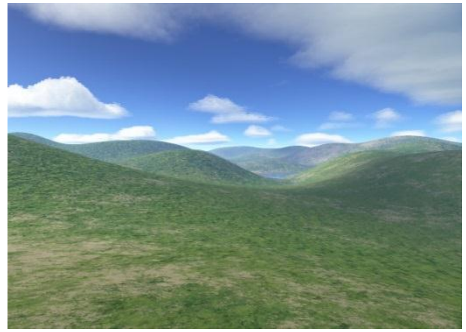
Spatial Autocorrelation: Consequences

iid population → autocorrelated population

- Population: iid → autocorrelated: Moran I, G
- Regression: OLS → SAR/MA/CAR
- Total: simple sum Σy → Bkriging Σwy
- Mapping: complete sample → Kriging
- Sampling: *n*(*iid*)↘*n*(ACR); *v*(*iid*)↘*v*(ACR)
- Cause other ACR: air pollution → lung cancer

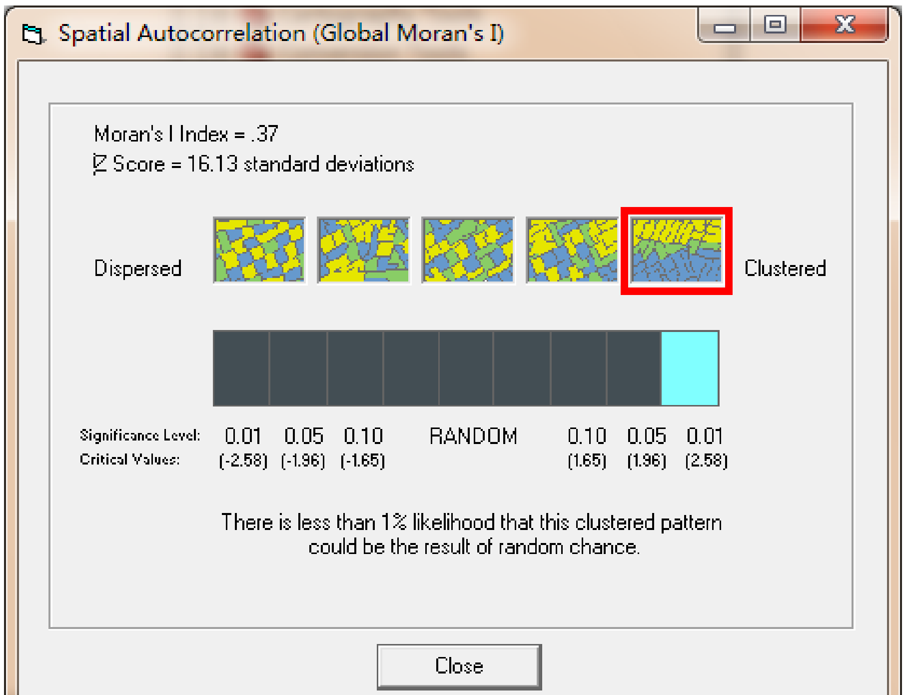
Spatial Autocorrelation: Measure

low $I = -0.3$



high $I = 0.6$

maximum $I = 1$



$$\text{Moran } I = \frac{N}{\sum_{ij} w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

$I \in [-1, 1]$

$i, j = 1, \ldots, N$

w_{ij}: join between i and j,

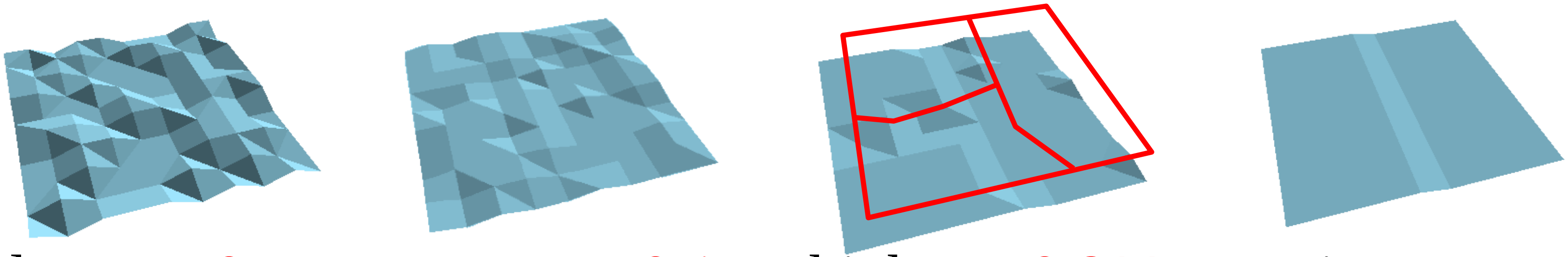
$E(I) = -\frac{1}{n-1}$, null hypothesis of no acr.

$$z = \frac{I - E(I)}{\sqrt{v(I)}} \sim N(0, 1)$$

Moran PAP. 1950. Notes on continuous stochastic phenomena. Biometrika 37: 17-23

Spatial Stratified Heterogeneity

Attribute similarity in strata



low $q = 0$ some $q = 0.1$ high $q = 0.8^{**}$ maximum $q = 1^{***}$

Nature: climate zone, landcover, watershed, plates

Human: states, classes, labor division, urban function zones

Biology: niche, footprint, fauna

Culture: religions, behavior, food

Miscellanies: east, middle and east

Spatial Heterogeneity: Consequences

Population → local, subpopulation

- Hotspot: Moran I → LISA, G → Gi
- Regression: SAR/MA/CAR →GWR,HB, MLM
- Total: Bkriging → MSN, B-shade, SPA
- Mapping: Kriging → Sandwich
- Sampling: Sample is biased when $n < L$
- Cause other heterogeneity

Wang JF, Haining R & Cao ZD. 2010. Sample surveying to estimate the mean of a heterogeneous surface: reducing the error variance through zoning. **International Journal of Geographical Information Science** 24(4): 523-543.

Spatial Heterogeneity: Local Measure

Gi, LISA, SatScan

$$G_i(d) = \frac{\sum_{j \neq i}^{n} w_{ij}(d) x_j}{\sum_{j \neq i}^{n} x_j}$$

Getis A, Ord JK. 1992. The analysis of spatial association by use of distance statistics. Geographical Analysis. 24(3): 189-206.

$$I_i = z_i \Sigma_j w_{ij} z_j, \quad z_i = \frac{x_i - \bar{x}}{SD_i},$$

Q2(X- , WX+): LH Q1(X+,WX+):HH
Q3(X- , WX-): LL Q4(X+, WX-): HL

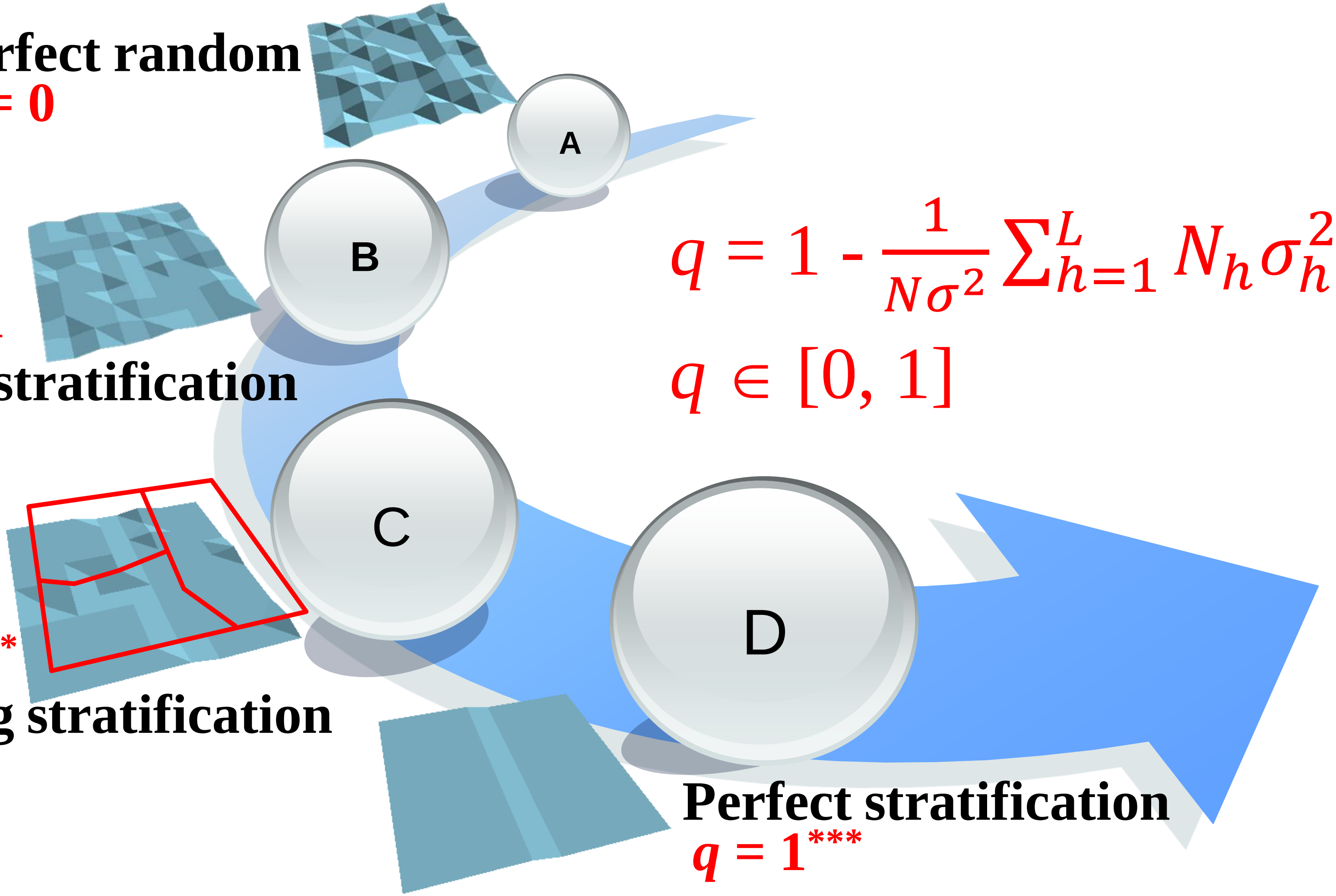
Anselin L. 1995. Local indicators of spatial association – LISA. Geographical Analysis. 27(2): 93-115

$$LR = [c/\mu]^{c} \mathrm{x} [(C\text{-}c)/(C\text{-}\mu)]^{C\text{-}c}$$

Kulldorff M. 1997. A spatial scan statistic. Communications in Statistics – Theory and Methodology 26(6): 1481-1496

Spatial Heterogeneity: Global Measure

Geographical Detector

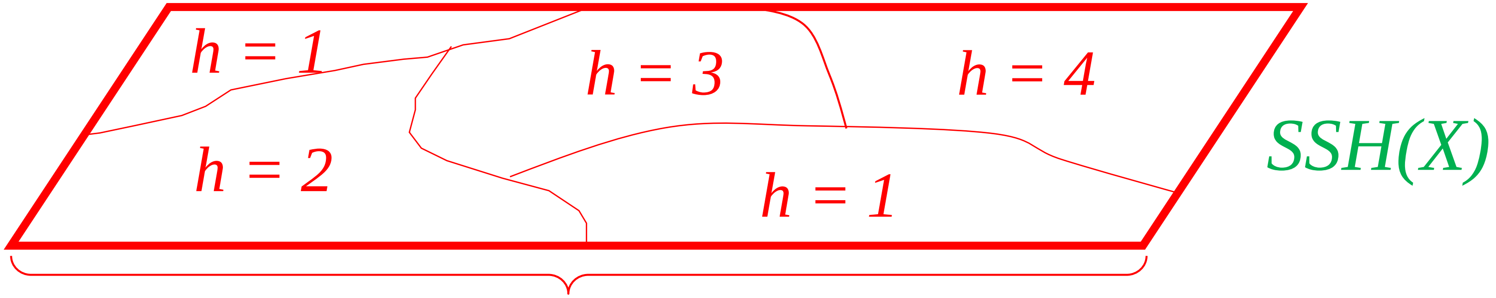


$$q = 1 - \frac{1}{N\sigma^2}\sum_{h=1}^{L} N_h \sigma_h^2$$

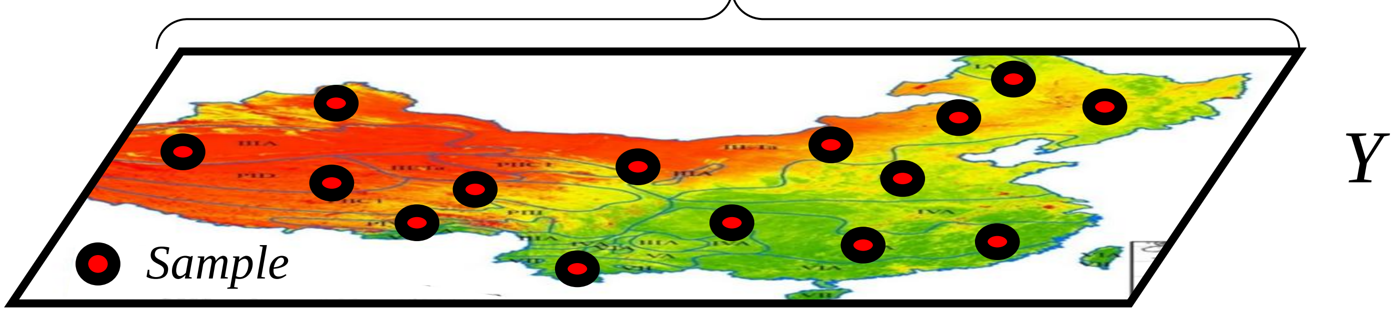
$$q \in [0, 1]$$

Spatial Stratified Heterogeneity (SSH): Attribute

(Axiom: If *X* causes *Y*, then their maps would be coupled)



$$q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2}$$



$q \in [0, 1]$, 100q% of Y is explained by X
0, if Y is independent to X
1, if Y is fully depended upon X

$$\frac{N-L}{L-1}\frac{q}{1-q} \sim F(L-1, N-L; \lambda)$$

Bayesian Spatio-temporal Hierarchical Model

Bayesian Spatio-temporal Hierarchical Model:

$$Y_{it} \sim MVN(\mu_{it}, \sigma_y^2)$$

$$\log(\theta_{it}) = \alpha + S[i] + b_0 * t + v(t) + B[i] * t + \epsilon_{it}$$

- $S[i]$ → Spatial process
- $b_0 * t + v(t)$ → Temporal process
- $B[i] * t$ → Spatio-temporal interaction

$$S[i] \sim \mathrm{CAR.Normal}(\mathrm{S.adj.sp}(i), \mathrm{S.num.sp}(i), \mathrm{S.weights.sp(i)}, \tau^2_{\mathrm{S,CAR}})$$

$$B[i] \sim \mathrm{CAR.Normal}(\mathrm{B.adj.sp}(i), \mathrm{B.num.sp}(i), \mathrm{B.weights.sp(i)}, \tau^2_{\mathrm{B,CAR}})$$

$$\mathrm{v(t)} \sim \mathrm{CAR.Normal}(\mathrm{v.adj.tm(t)}, \mathrm{v.num.tm(t)}, \mathrm{v.weights.tm(t)}, \tau^2_{\mathrm{T,CAR}})$$

Spatially structured random effect and a spatially unstructured can be commonly incorporate by the form of a priori information

Modelling with BSTHM using HIV/AIDS Data

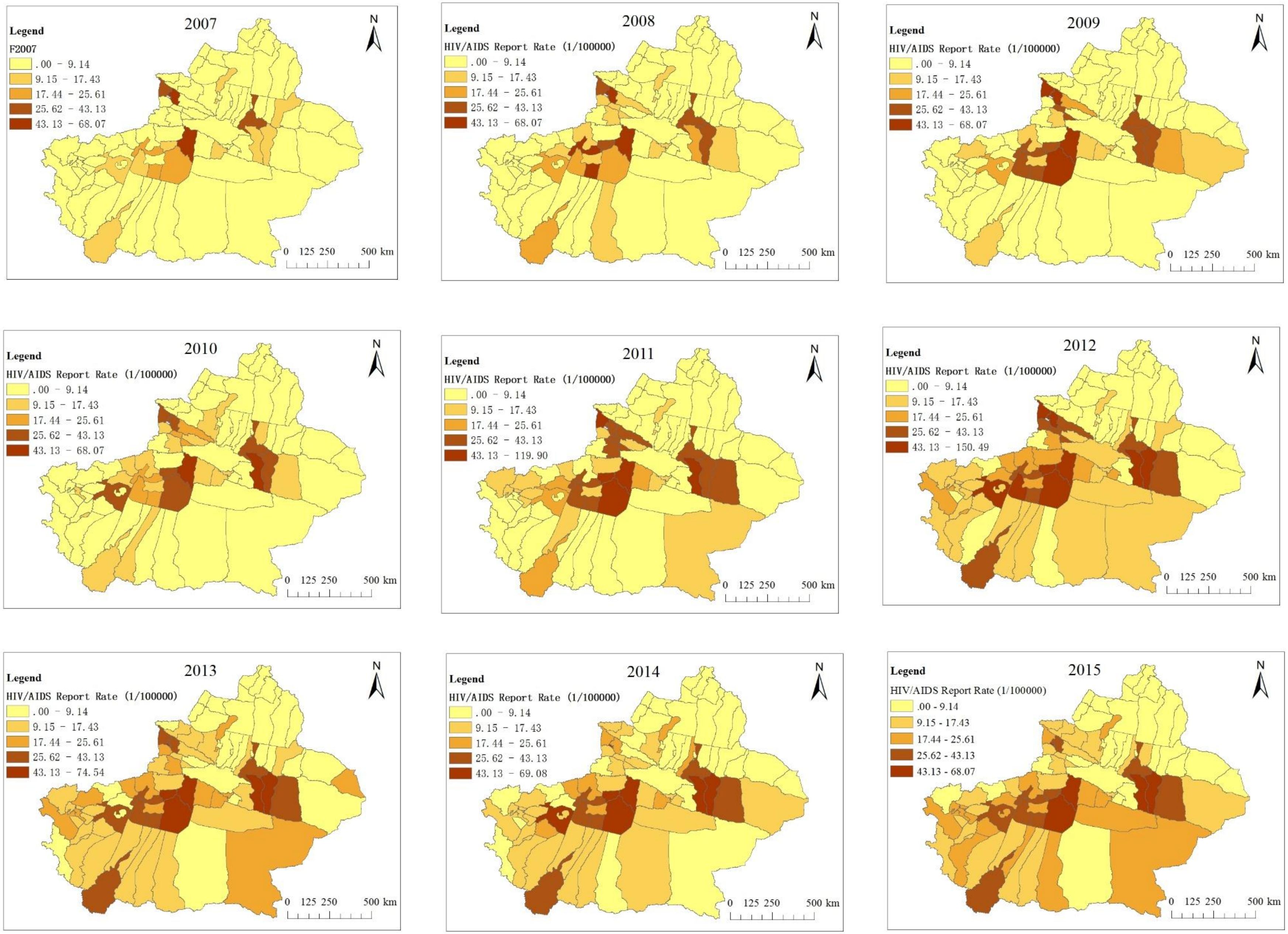
$$y_{it} \sim Possion(n_{it} r_{it})$$

$$\log(r_{it}) = \alpha + s_i + (b_0 t^* + v_t) + b_{1i} t^* + \varepsilon_{it}$$

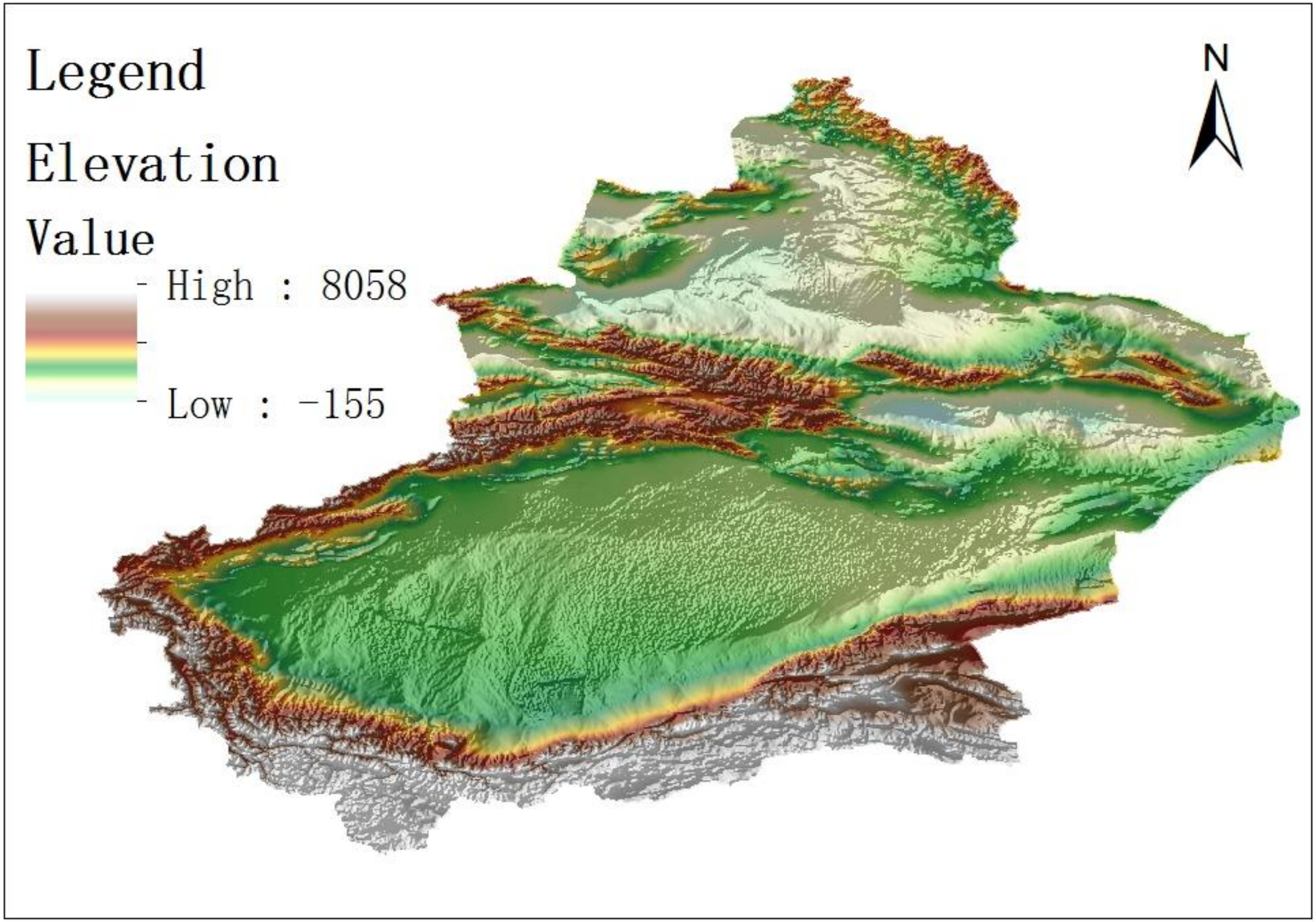
Poisson and log link regression functions were used to model the disease data. y_{it} and n_{it} were the number of cases and the total population, respectively. α was the overall log risk average for the entire study. The observed space-time heterogeneity in disease risk was divided into the the spatial term s_i, and the temporal term $(b_0 t^* + v_t)$.

Study Data

HIV/AIDS Report Rate from 2007 to 2015

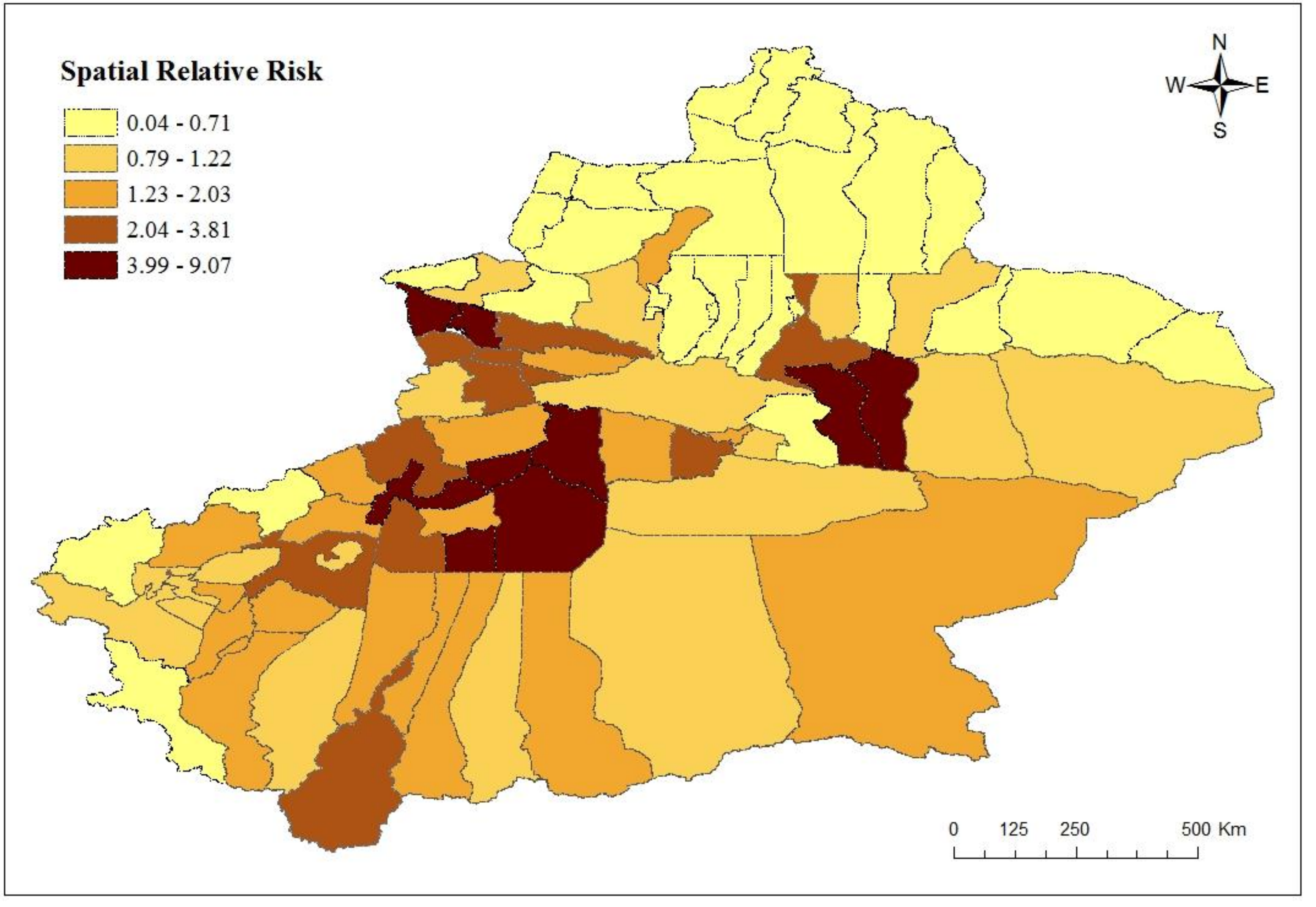


Terrain in the study region

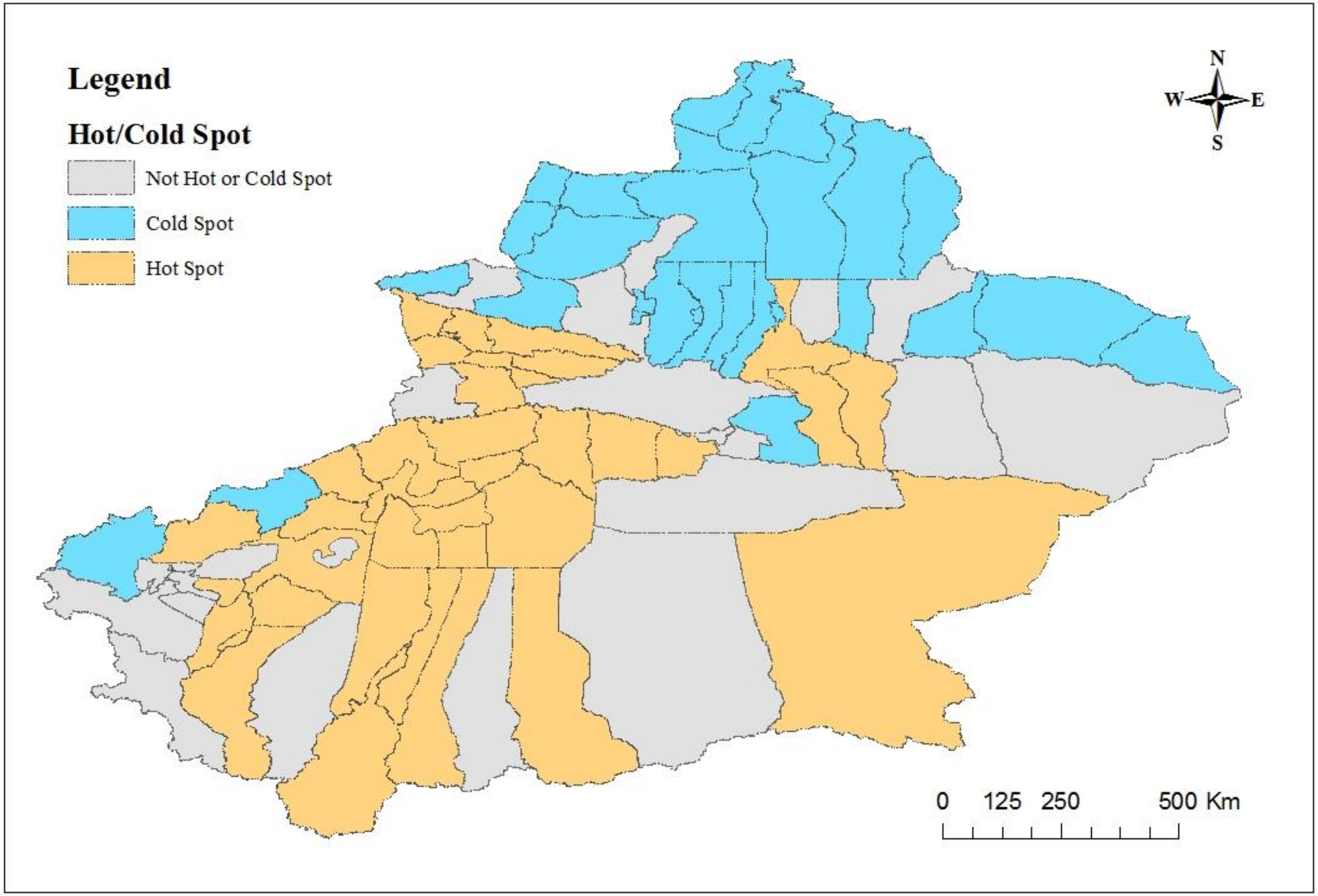


Results and Interpretation

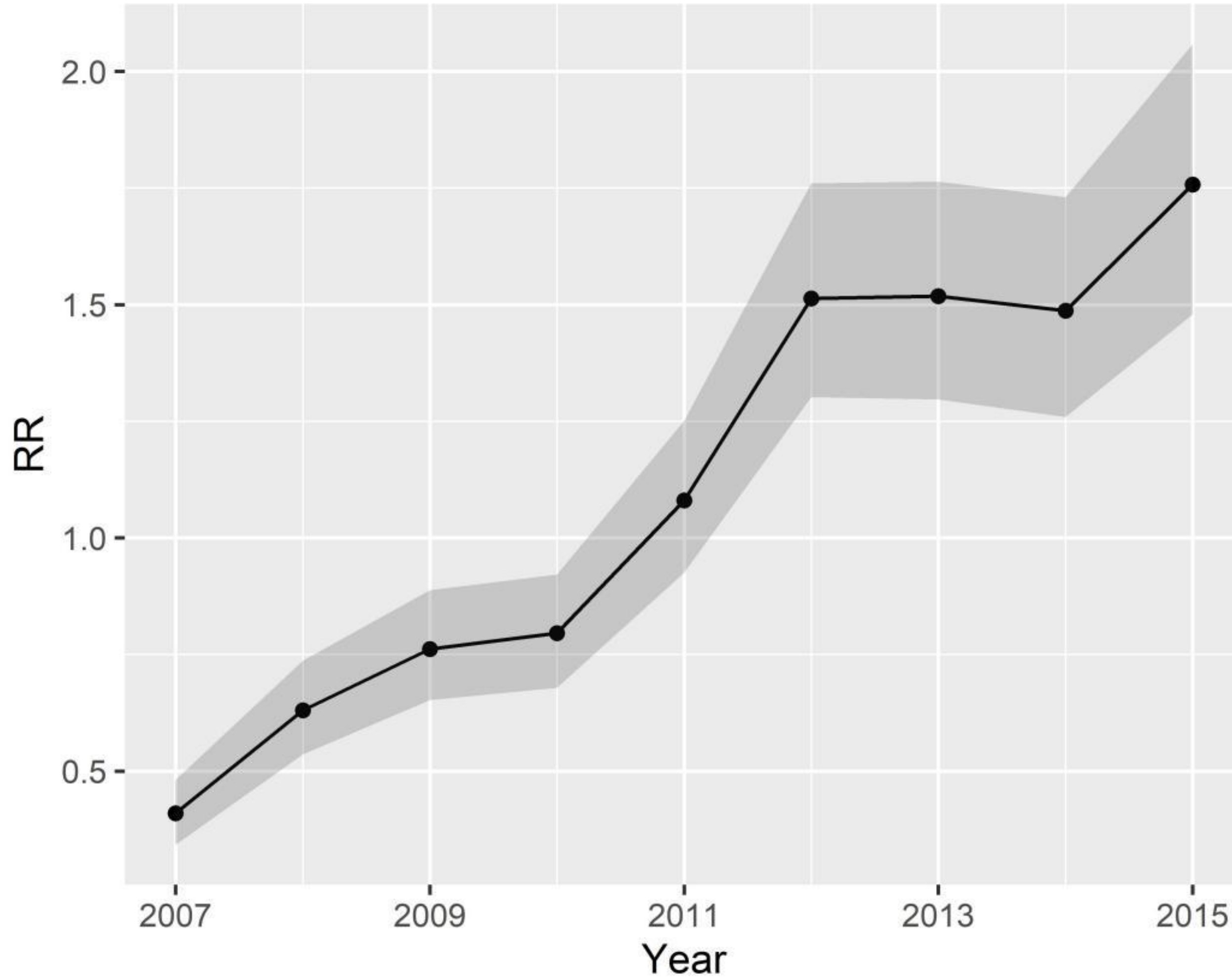
Common Spatial Pattern



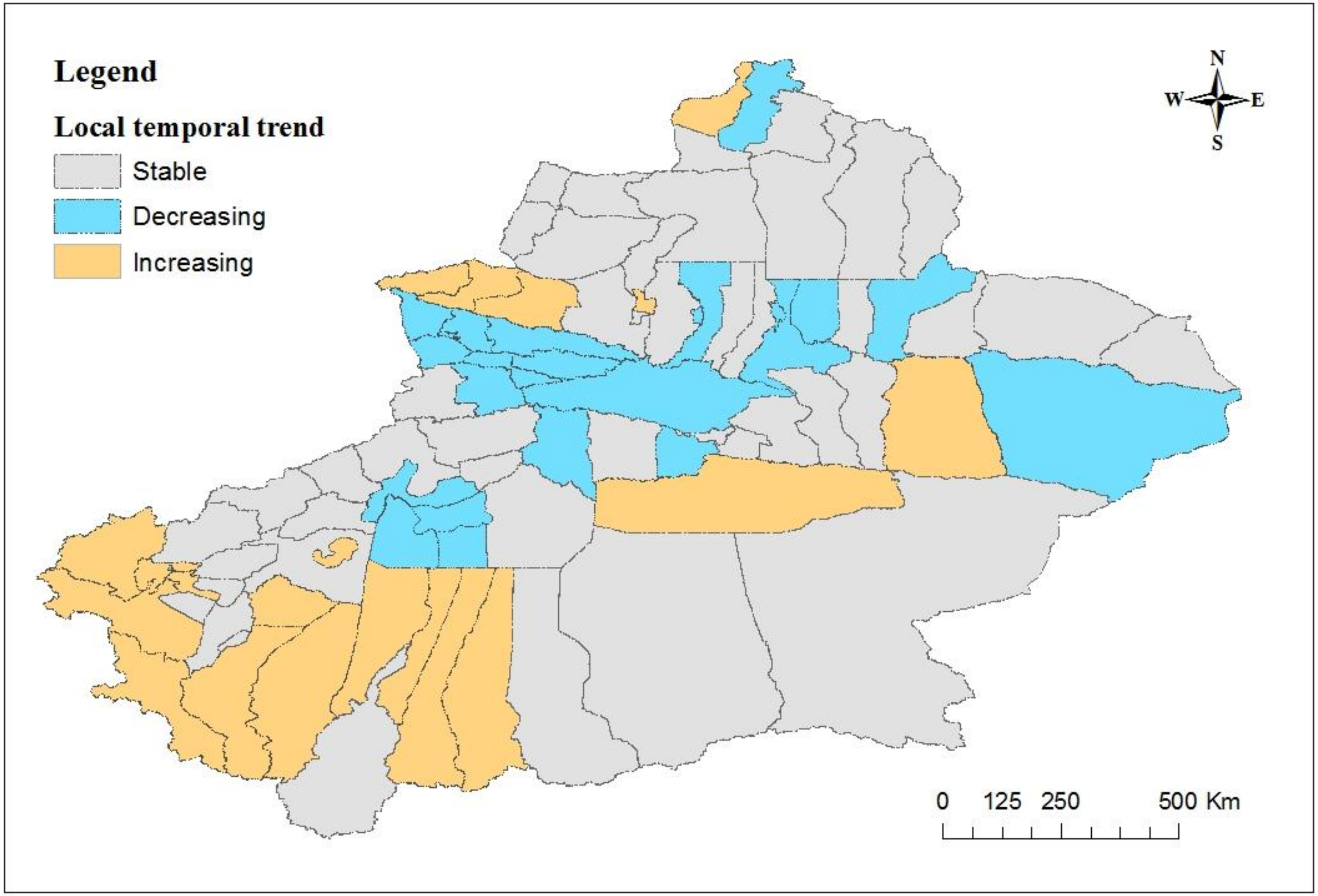
Hot and cold spots of spatial relative risks (RRs) based on the posterior probability of $\exp(S_i) > 1.0$



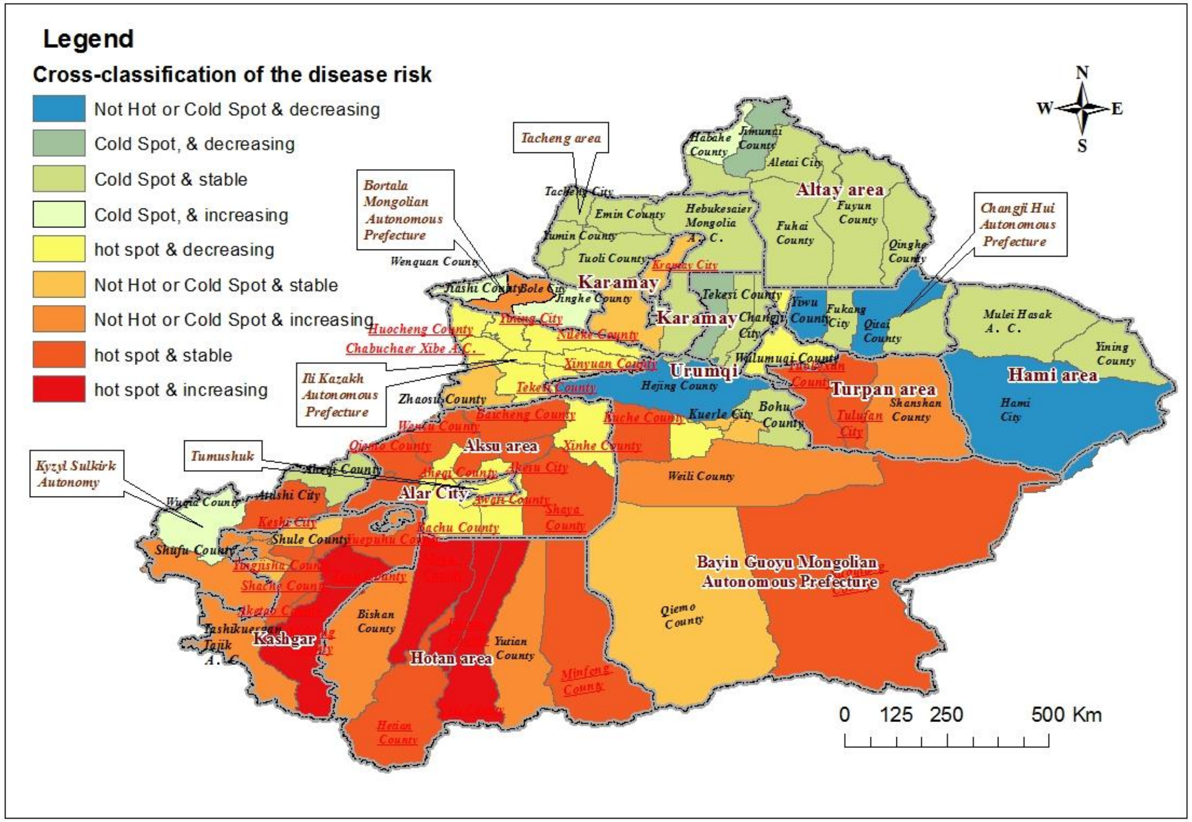
Overall temporal trend (exp [$b_0 t^* + v_t$])



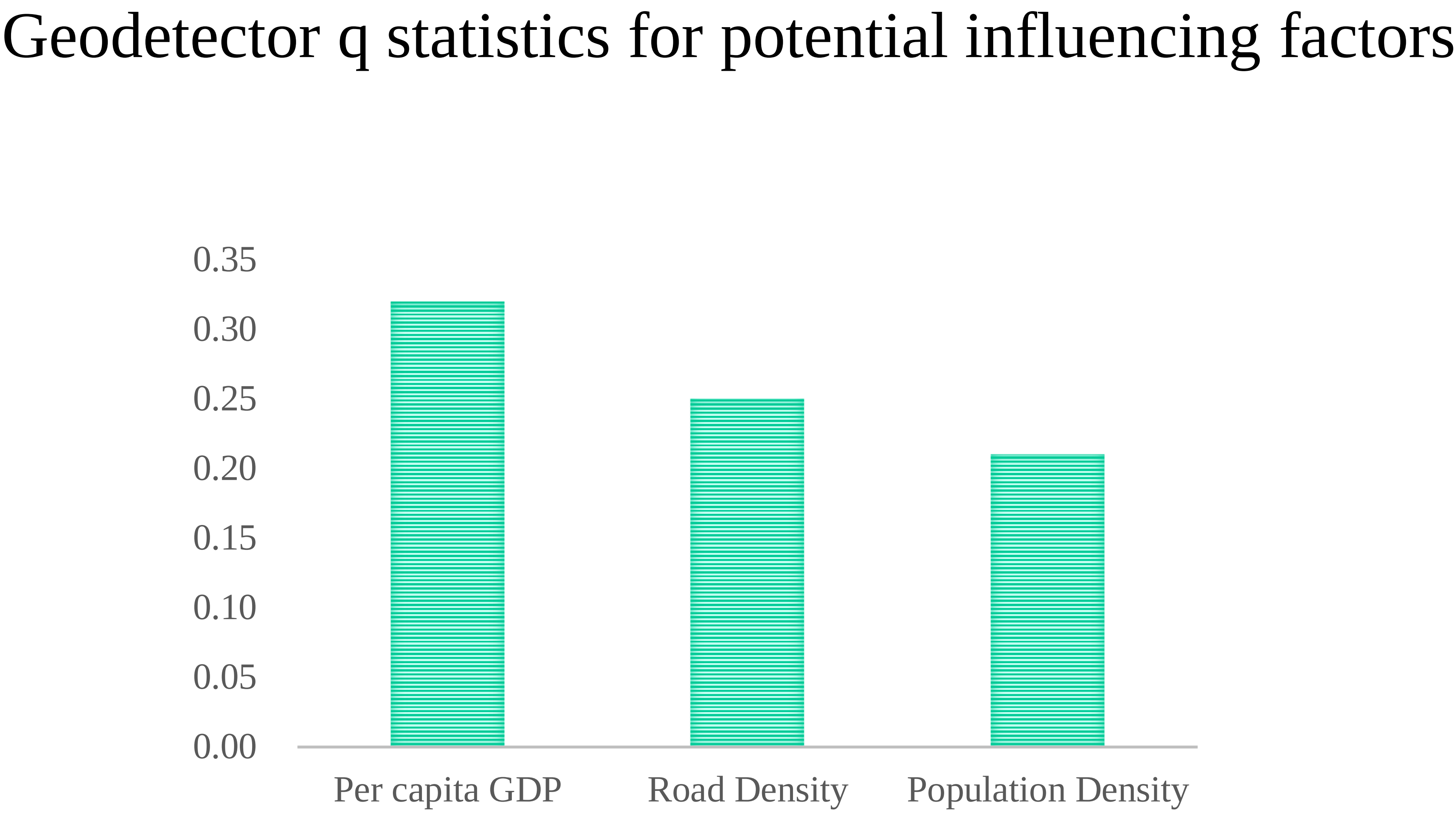
Deviations in local trends compared to the overall trend b_{1i} of the disease



Cross-classification of the disease risk in all counties



	Faster increase trend	Slower increasing trend	Not different from common trend	Total
Hot spots	17	15	5	37(42.05%)
Cold spots	20	4	5	29(32.95%)
Neither hot /cold spots	8	4	10	22(25%)


Geodetector q statistics for potential influencing factors
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Per capita GDP
Road Density
Population Density

Conclusions and Future Work

1. HIV/AIDS report rate in Xinjiang presented statistically significant spatio-temporal hegerogeneity
2. The HIV/AIDS risk presented a increased trend
3. Three was statistically significant local temporal variation
4. Population density, traffic and economic condition had apparent influence
5. More explanatory variables will be collected

地理探测器：www.geodetector.cn

Thanks