Analysing spatio-temporal co-location of categorical incidents: a case study of chronic respiratory diseases in Nanning City

Jianquan Cheng[*1], Ling Li[†2] , Xiongfa Mai[‡3], Lian Duan[4]

[1] Department of Natural Sciences, Manchester Metropolitan University
[2] Key Laboratory of Environment Change and Resources Use in Beibu, Nanning Normal University
[3] School of Mathematics and Statistics, Nanning Normal University
[4] Joint Centre for Urban Health and Security Intelligent Data Analytics, Nanning Normal University

January 17, 2022

Summary

Analysing spatial pattern of chronic respiratory diseases particularly in children is prerequisite for modelling its demographical, behavioural, and environmental contributions. This paper aims to explore the application of an innovative approach – geographically and temporally weighted co-location quotient (GTWCLQ) into such case study of Nanning City in 2016. The results exhibit the values of GTWCLQ in analysing the spatio-temporal associations, including symmetrical and asymmetrical dependence, between five categories of such diseases by considering their spatio-temporal dependence and heterogeneity. The paper has also discussed its scaling effect and unbalanced temporal scale problem in such urban big data analytics.

KEYWORDS: geographically and temporally weighted co-location quotient, spatio-temporal dependence and heterogeneity, unbalanced temporal scale problem, chronic respiratory diseases, Nanning city.

1. Introduction

Chronic respiratory diseases have become one of major concerns in the development of Healthy China 2030 due to heavy air pollution in many Chinese cities (Guan, et al., 2016). The impacts of air pollution including indoor pollution on the chronic respiratory diseases in children aged 3 and below are particularly challenging for local environmental planning and management. These diseases are usually recorded by local hospitals as incident data at residence location. It is interesting to explore the spatial association between the different categories of chronic respiratory diseases before modelling the demographical, environmental and behavioural factors. The co-location quotient, has been increasingly used to measure the directed spatial dependence between categorical variables. Many types of urban big data are categorical variables, such as POI data that have location and nominal attribute values but no numerical values (e.g., interval or ratio). The co-location quotient (CLQ) aims to quantify the spatial association between categories of a population that may exhibit spatial autocorrelation. It can also detect symmetry and asymmetry in spatial dependence. The geographically weighted co-location quotient (GWCLQ) that takes into account the spatial heterogeneity of the association between categorical data, like other methods of geographically weighed analytics (Li, et al. 2016; Zhang, et al., 2019), has been incorporated into ArcGIS Pro 2.8. but there is no tool of its spatio-temporal version. This paper aims to explore the application of geographically and temporally weighted co-location quotient (GWTCLQ) for the Chronic respiratory diseases analysis.

* j.cheng@mmu.ac.uk
† Liling81@nnnu.edu.cn
‡ maixf@nnnu.edu.cn

2. Data and Methods

The case study area is Nanning city, which is a provincial capital of Guangxi Zhuang Autonomous Region in Southern China (Shen, et al., 2019). The incident data are composed of five categories of chronic respiratory diseases reported for children aged 3 and below in 2016. Selection of this specific age group is their mobility only limited within residential neighbourhood, compared with school students. Then the spatial pattern of chronic respiratory diseases might be associated with the climate and environmental conditions at local level. These incident data are represented as points spatially with daily temporal resolution. The key approach for this case study is geographically and temporally weighted co-location quotient and its methodological and technical details can be referred to our recently published work (Li, et al., 2022). Here are only mathematical descriptions of this approach.

$\mathrm{GTWCLQ}^{\mathrm{g}}_{\mathrm{A-B}}$ (global geographically and temporally weighted co-location quotient), is formulated as follows (Equation 1):

$$GTWCLQ^{g}_{A-B} = \frac{\sum_{k\in E}\sum_{i\in C}\sum_{j\in D_k} W_{ijk}x_{kj}}{M\sum_{k\in E}\sum_{i\in C}\sum_{j\in D_k} W_{ijk}} \quad (1)$$

where $\mathrm{GTWCLQ}^{\mathrm{g}}_{\mathrm{A-B}}$ denotes the global geographically and temporally weighted co-location quotient of type A relative to type B at time *k*. W_{ijk} is a spatio-temporal weight denoting the relative importance of the *jth* point at time *k* to the *ith* A-point at actual time. x_{kj} is a binary value that equals 1 if the *jth* point at time *k* is a type B point, and is equal to 0 otherwise. E is the set of time periods in study. C is the set of type A points at actual time and D_k is the set of all points at time *k*. Values greater than one show that type A points tend to be spatially dependent on type B points at time *k*, while values less than one indicate that type A points are likely to be far from type B at time *k*. The larger the value, the stronger the dependence or attraction.

Its local geographically and temporally weighted colocation quotient is formulated as follows (Equation 2):

$$GTWCLQ^{l}(i) = \frac{\sum_{k\in E}\sum_{j\in D_k} W_{ijk}x_{kj}}{M\sum_{k\in E}\sum_{j\in D_k} W_{ijk}} \quad (2)$$

where all terms are the same as equation 1, and $\mathrm{GTWCLQ}^{\mathrm{l}}(i)$ is the local value of *ith* type A relative to type B at time *k*.

The spatio-temporal weight matrix is constructed by multiplying the spatial weight matrix with the temporal weight matrix, term by term, as shown in Equation 3.

$$f(d^{s}_{ij}) \times g(d^{t}_{ij}) = w_{ij}t_k = W_{ijk} \quad (3)$$

where f(.) and g(.) are the kernel functions for spatial and temporal weight computation respectively and a dot product scheme is used to calculate the final spatio-temporal weight values, which can save computation time for large-scale big data sets.

Particularly, the temporal weight value is calculated as follows (Equation 4):

$$T_{ij} = g(d_{ij}^t) = \begin{cases} (d^t_{ij})^{-\alpha} & if \quad and \quad d^t_{ij} \le t_0 \\ 0 & d^t_{ij} > t_0 \end{cases} \quad (4)$$

where T_{ij} is the temporal weight value between time point *i* and point *j*. t_0 is temporal bandwidth value and α is the time decay parameter, measuring the relative importance of the higher-order temporal neighborhoods. The value of α is often set to 0,1 or 2. A value of 0 indicates a "no autocorrelation" temporal effect. When α=1, the inverse distance function is used to measure temporal closeness and when α=2 the bi-square function is used, as in this study. To test the significance of spatio-temporal colocation quotient values, a Monte Carlo simulation method is used to calculate statistical test values.

GTWCLQ software package developed by the team is available with the identifier at the following link https://doi.org/10.6084/m9.figshare.16641763.

3. Results

There are 159,974 points across 12 months of 2016 for all the 5 categories: C1-Acute nasopharyngitis, C2-Breathing pneumonia, C3-Acute pharyngitis, C4 Bronchitis and C5- Polyangiitis-bronchitis. When running the GWTCLQ package for December 2016, a spatial bandwidth of 300 and temporal bandwidth of 12 (months) were selected and 1,000 times for Monte Carlo simulation. The global GWTCLQ values at significance level 1%, between all the five categories, are listed into Table 1.

Table 1. Global values of GWTCLQ at monthly level

Category	C1	C2	C3	C4	C5
C1	1.01	1.06	0.96	0.83	0.868
C2	0.865	1.198	0.955	0.904	
C3	0.857	1.045	1.863	0.875	0.985
C4	0.802	1.056	0.97	1.506	1.068
C5	0.793	1.073	1.015	1.056	4.358

Note: blank cells are not significant at 1%.

The comparisons in Table 1 indicate category C5 demonstrates the highest level of clustering, and their clusters with high significance level (<1%) is shown in Figure 1.

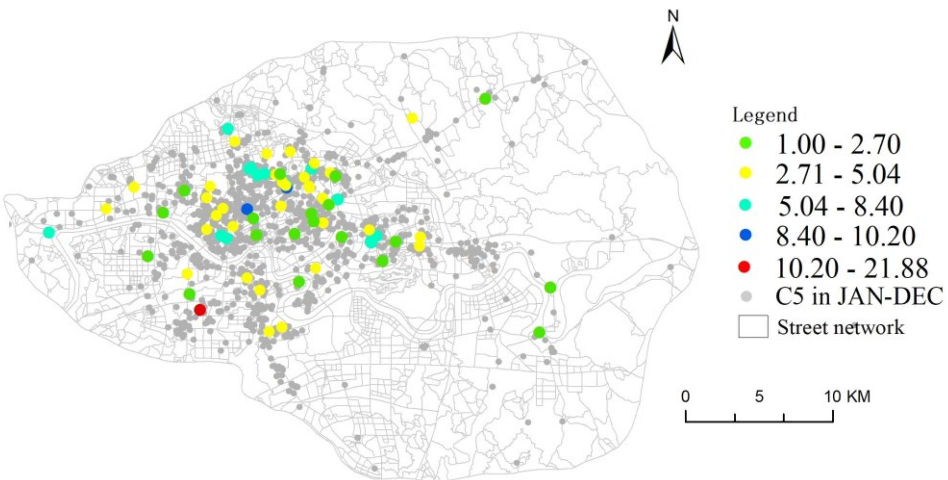

Figure 1. Local clusters of C5 (polyangiitis-bronchitis) from GWTCLQ at monthly level

It is also clear to see that there is only one pair of categories with symmetrical dependence or co-location with each other: category C4 and C5. Category C1 and C2 are unique as all other categories are dependent on C2 and by contrast C1 is independent of all other categories. Category C3 is dependent on C2 but not opposite so showing asymmetrical dependence. The clusters of category C5 in Figure 1 are concentrated around large-scale industrial parks in which plenty of factories might be the source of air pollutions.

To test the effects of temporal scale, the GWTCLQ was also run for 30 days of December 2016 at daily level but category C5 was merged into C4 due to small size sample and their co-location nature in Table 1. A spatial bandwidth of 60 (due to smaller size of sample at this level) and temporal bandwidth of 30 (days) were selected and 1000 times for Monte Carlo simulation. The global GWTCLQ values at significance level 1%, between all the four categories, are listed into Table 2.

Table 2. Global values of GWTCLQ at daily level

Category	C1	C2	C3	C4&C5
C1	1.313	0.790	0.381	1.085
C2	0.519	1.545	0.518	0.954
C3	0.128	1.159	6.025	0.314
C4&C5	0.581	0.851	0.245	2.989

Note: all significant at 1% level

On this temporal scale, there are some changes, Category C3 demonstrates the strongest level of clustering and its clusters shown in Figure 2. Category C1 is dependent on C4&C5 but not opposite so showing different asymmetrical dependence after being merged.

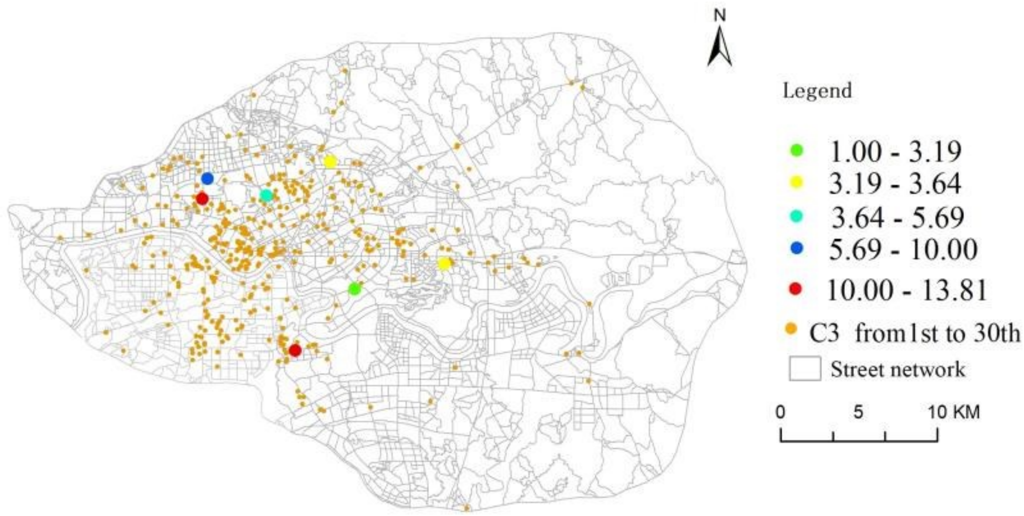

Figure 2. Local clusters of C3 (acute pharyngitis) from GWTCLQ at daily level

These results show GWTCLQ is (temporal) scale sensitive, determined by resolution (daily or monthly) and duration (one year or one month). In addition, the spatial heterogeneity shown in these maps is very related to the spatio-temporal dependence defined by different space and time decay functions and parameters in Equation 4. Using a different parameter α=1.5 (2.0 for above results), the GWTCLQ analysis produced similar results as those in Table 2, but different result for α=0. All these mean strong spatial and temporal effects in the pattern of chronic respiratory diseases in children across the city in 2016.

4. Conclusion and discussion

The above-mentioned results exhibit that GTWCLQ is able to analyse the spatio-temporal association (co-location, asymmetrical dependence and independence) between categorical variables from incident data through global and local geo-computations. However, this study has not compared the spatio-temporal patterns between different months or different days due to the unbalanced temporal scale problem (Li, et al. 2022), which will be explored in the future work.

Acknowledgement
This study has been supported by National Natural Science Foundation of China (No. 41961062), Key Research & Development Program of Guangxi Provence (No. 2019AB16010), and Program of Natural Science Foundation of Guangxi Province (No. 2018JJA150089).

References

Guan W, Zheng X, Chung K, and Nan-Shan Zhong, N (2016) Impact of air pollution on the burden of chronic respiratory diseases in China: time for urgent action, *The Lancet*, 388(10054): 1939-1951.

Li L, Cheng J, Bannister J, Mai X (2022). Geographically and temporally weighted co-location quotient: an analysis of spatiotemporal crime patterns in Greater Manchester. *International Journal of Geographical Information Science*. doi:10.1080/13658816.2022.2029454.

Li Z, Cheng J and Wu Q (2016) Analyzing regional economic development patterns in a fast developing province of China through geographically weighted principal component analysis. *Letters in Spatial & Resource Sciences*, 9(3), 1-13. doi:10.1007/s12076-015-0154-2.

Shen J, Cheng J, Huang W and Zeng F (2020) An Exploration of Spatial and Social Inequalities of Urban Sports Facilities in Nanning City, China. *Sustainability*. 12(11),4353-4353.

Zhang L, Cheng J and Jin C (2019). Spatial Interaction Modeling of OD Flow Data: Comparing Geographically Weighted Negative Binomial Regression (GWNBR) and OLS (GWOLSR). *ISPRS International Journal of Geo-Information*, 8(5), 220. doi:10.3390/ijgi8050220.

Biographies

Jianquan Cheng is a Reader in Urban Studies and deputy director of the Manchester Metropolitan Crime and Wellbeing Big Data Center. His recent interests focus on analysing and modelling how physical and built-environment impacts on public health and well-being at a variety of scales in Chinese and British cities, which aim to generate data driven evidence and frameworks for spatial planning and governance.

Ling Li is an associate professor in GIS. Her expertise includes geospatial modelling by fusing AI and mathematical modelling into big data analytics. Li's recent projects focus on analysing and modeling spatial-temporal data to support predictive policy making.

Xiongfa Mai is an associate professor in machine learning, whose expertise is deep learning and precipitation forecasting including intelligent optimisation algorithms. His research projects mainly focus on spatio-temporal series prediction using the deep learning approach.

Liang Duan is a professor in GIS and director of Joint Centre for Urban Health and Security Intelligent Data Analytics. His expertise is big data analytics for crime applications including machine learning.