Environmental Modelling & Software 33 (2012) 114–115

Contents lists available at SciVerse ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Software, data and modelling news

Environmental health risk detection with GeogDetector

Jin-Feng Wang [a,*], Yi Hu [a,b]

[a] State Key Laboratory of Resources and Environmental Information Systems, Institute of Geographic Sciences and Nature Resources Research, Chinese Academy of Sciences, A11 Datun Road, Beijing 100101, China
[b] School of Earth & Mineral Resources, China University of Geosciences, Beijing 100083, China

ARTICLE INFO

Article history:
Received 13 July 2011
Received in revised form
19 January 2012
Accepted 27 January 2012
Available online 29 February 2012

Keywords:
Spatial heterogeneity
Consistence of spatial patterns
Interaction effect

ABSTRACT

Human health is affected by many environmental factors. Geographical detector is software based on spatial variation analysis of the geographical strata of variables to assess the environmental risks to human health: the risk detector indicates where the risk areas are; the factor detector identifies which factors are responsible for the risk; the ecological detector discloses the relative importance of the factors; and the interaction detector reveals whether the risk factors interact or lead independently to disease.

Software available

Software name: GeogDetector
Developers: Yi Hu and Jin-Feng Wang
Contact: huy@lreis.ac.cn or wangjf@lreis.ac.cn
Available since: January 2010
Operating system: 32-bit Windows
Program language: C#
Availability: free from http://www.sssampling.org/GeogDetector

1. Introduction

Human health could be at risk if health related factors exceed the norm at any level. From the individual down to the organic system, texture, cell, and molecular levels, it is typically the duty of biologists and doctors using microscopes to detect pathogens, while epidemiologists and environmental scientists safeguard the individual, the family, the community, society, and ecological environments. For both of these groups and others (Christakos, 1992; Getis and Ord, 1992; Haining, 2003; Kulldorff, 1997; Rushton, 1992; Chen et al., 2011), we developed geographical detector to assist in exploring risks (Wang et al., 2010a; Hu et al., 2011).

The GeogDetector software explores environmental risks to health through two mechanisms. One method is the proxy approach. Pathogens of diseases are often difficult or too expensive to measure directly. However, these pathogens often exist in geographical strata, which is measurable. It is thus possible to explore direct pathogens of diseases by investigating their proxy in geographical space. The second mechanism used by the geographical detector involves testing the spatial consistence of spatial patterns between a disease and its environmental proxies; that is, comparing the spatial consistency of health risk spatial distribution (e.g., disease prevalence) versus the geographical strata (e.g., climate, soil, water, population, ethnicity, culture and lifestyle, poverty, nutrition, land use, factor, etc, individually or bilaterally) in which potential health risk determinants exist. We assume that a disease would exhibit a spatial distribution similar to that of an environmental factor if the environmental attribute leads to the disease (Wang et al., 2010a), such as people living in areas with heavier polluted soil may have higher prevalence of a disease.

The GeogDetector is grounded on the *PD*, i.e. Power of Determinant, which generates four detectors (Wang et al., 2010a).

$$PD = 1 - \frac{1}{N\sigma^2}\sum_{i=1}^{L} N_i\sigma_i^2$$

where N and σ^2 denote the area and the variance of disease prevalence of the study area respectively. The study area is stratified into L stratums, denoted by $i = 1, \ldots, L$ (Wang et al., 2010b), according to spatial heterogeneity (which is defined as an attribute whose statistical properties, e.g., mean and standard deviation, change in space) of a suspected determinant or its proxy of the disease. $PD \in [0, 1]$, 1 if the determinant completely controls the

* Corresponding author.
E-mail address: wangjf@lreis.ac.cn (J.-F. Wang).

1364-8152/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2012.01.015

disease, 0 if the determinant is completely unrelated to the disease. Thus, the *PD* reflects the degree to which a determinant explains the prevalence of the disease.

2. Software features

The main GUI of this software is based on the ARC-Engine 9.3 platform or a more advanced version implemented using C# that allows users to manipulate and visualize the spatial data more easily. The data format supported is ESRI Shapefiles, a popular format shared by most GIS software.

2.1. Risk detector

This is used to search for areas with potential health hazards. A *t*-test is used to compare the differences in average values between strata; the larger the difference, the greater is the danger to the population's health in the sub-region.

2.2. Factor detector

This quantifies the impact of an environmental factor on the observed spatial disease pattern using Power of Determinant (*PD*) value.

2.3. Ecological detector

This explores whether a geographical stratum C (associated with one suspected determinant) is more significant than another stratum D (associated with another suspected determinant) in controlling the spatial pattern of the disease by comparing the *PD* values of C and D.

2.4. Interaction detector

This probes whether two health determinants C and D when taken together weaken or enhance one another, or whether they are independent in developing a disease. It compares the sum of the disease contribution of two individual attributes to the contribution of the two attributes when combined:

Enhance : if $PD(C \cap D) > PD(C)$ or $PD(D)$

Enhance, bivariate : if $PD(C \cap D) > PD(C)$ and $PD(D)$

Enhance, nonlinear : if $PD(C \cap D) > PD(C) + PD(D)$

Weaken : if $PD(C \cap D) < PD(C) + PD(D)$

Weaken, univariate : if $PD(C \cap D) < PD(C)$ or $PD(D)$

Weaken, nonlinear : if $PD(C \cap D) < PD(C)$ and $PD(D)$

Independent : if $PD(C \cap D) = PD(C) + PD(D)$

where $\cap$ could easily be implemented in GIS by overlaying the two factor layers of C and D.

3. Concluding remarks

We believe that this method of geographical detectors is novel as it extracts the implicit interrelationships between risk factors and health event without any assumptions or restrictions with respect to explanatory and response variables, which are difficult to model using classic epidemiological methods. Additionally, the geographical detectors are usable for both quantitative data and nominal data. The later can cause trouble with classic regression when there are too many categories (Allen, 1997). The GeogDetector software is completely free of charge. All the steps needed by the user to interact with the software, from installation to how to use it, are well documented at the geographical detector web page. It provides a quick, easy and efficient way for researchers to grasp association between human health and environmental attributes from the perspective of spatial distribution.

Acknowledgments

This study was supported by CAS (XDA05090102), MOST (2012CB955503; 2012ZX10004-201; 2011AA120305), and NSFC (41023010) grants.

References

Allen, M.P., 1997. Regression Analysis with Dummy Variables. Understanding Regression Analysis. Plenum Press, New York.
Chen, H.Y., Wood, M.D., Malby, L.E., 2011. Uncertainty analysis in a GIS-based multi-criteria analysis tool for river catchment management. Environmental Modelling & Software 26 (4), 395–405.
Christakos, G., 1992. Random Field Models in Earth Sciences. Dover Publ., NY.
Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. Geographical Analysis 24, 189–206.
Haining, R., 2003. Spatial Data Analysis: Theory and Practice. Cambridge University Press, Cambridge.
Hu, Y., Wang, J.F., Li, X., Ren, D., Zhu, J., 2011. Geographical detector-based risk assessment of the under-five mortality in the 2008 Wenchuan Earthquake, China. PLoS ONE 6 (6), e21427.
Kulldorff, M., 1997. A spatial scan statistic. Communications in Statistics: Theory and Methods 26, 1481–1496.
Rushton, J.P., 1992. Cranial capacity related to sex, rank, and race in a stratified random sample of 6,325 U.S. military personnel. Intelligence 16, 401–413.
Wang, J.F., Li, X.H., Christakos, G., Liao, Y.L., Zhang, T., Gu, X., Zheng, X.Y., 2010a. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. International Journal of Geographical Information Science 24 (1), 107–127.
Wang, J.F., Haining, R., Cao, Z.D., 2010b. 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.