Understanding information landscapes through space-time analysis

Li An, lan@mail.sdsu.edu; Sarah Wandersee, wanderse@rohan.sdsu.edu

Department of Geography, San Diego State University

Space-time analysis is a rapidly growing research frontier in many disciplines including geography, regional science, and GIScience. Recent advances in information science and GIS technologies, along with the increasing availability of large space-time datasets, have engaged scientists of varying background in space-time analysis and modeling. In this context, we have collected and analyzed weekly search data for the two keywords "climate change" and "global warming". Our goal in this space-time analysis of climate change data is two-fold. First, we are interested in knowing the information landscape related to people's understanding and perception of climate change and global warming, a hot topic of global interest; second, we aim to establish a space-time analysis methodology that may be applied to other domains or fields.

In regards to the information landscape for "climate change", our analysis results reveal that 1) climatic patterns (amount of precipitation, summer days, etc.) significantly predict the Internet traffic for "climate change" and "global warming", indicating that where you live may strongly affect how you perceive these terms (Figure 1); 2) socioeconomic and demographic features, such as residential characteristics (e.g., commute time, population density), age, gender, ethnicity, education, and income all have significant influences on such reflections or perceptions; 3) attitudes towards the terms "climate change" and "global warming" may be partially explained by the dominant political party membership of a certain region; and 4) "global warming" as a term returns more negative and neutral websites (Figure 2a) and expired websites (Figure 2b).

Climate predicts interest in climate change

Where you live

	Interest in CC	Interest in GW	Anti-GW	Pro-GW	Net support for GW
More summer time	More	More	More	More	
More drought	More	Less/More	More	More	More
More frost		More	More	More	
More precipitation		More	More		Less

Figure 1: The table indicating coefficient directions for survival analysis of "climate change" and "global warming" website activity in relation to climatic variables.

Related to our second goal, our major activity has been devoted to developing tools (modules, procedure, macros, and metrics) that advance space-time analysis methodology.

1) Data collection and preparation: We assembled a large amount of data for later regression analysis. Sources include the American Community Survey (2005-09), North America Climate Extremes Monitoring Program (2006), US Census (2010), US Agricultural Census (2007), National Agricultural Statistics Service (2010), and National Weather Service Storm Prediction Center (2011). In order to automate the processing of keyword search results, we developed R scripts (for data cleaning and formatting) and several models for ArcGIS using Python. These models facilitate batch processing to handle the large time series of weekly data we are collecting and can be adjusted to process any of our keyword search results in preparation for further space-time analysis.

2) Metrics development: Several metrics were employed to quantify spatial patterns at discrete times, including those used in landscape ecology (e.g., dominance, proximity), map comparison (e.g., Pontius et al. 2011), and spatial statistics (e.g., univariate and bivariate LISA, Geary's G; Anselin et al. 2006). Specifically, we have pursued exploratory spatial analysis in GeoDa, including the use of spatial lag and error models, univariate and bivariate local indicators of spatial association. In particular, we

used the "hazard" (a term quite often used in sociology, demography, and epidemiology) concept to depict and quantify the risk of an area being dominated or substantially influenced by certain events (e.g., ideas). The concept hazard is calculated as a function of survival time, or the time within which a place is free from a certain event (Allison 1995, p.63-66; An and Brown 2008). Two things on our to-do list are a) to explore the impact of choosing different spatial units on analysis results (Pontius et al. 2011), and b) to calculate different matrices (e.g., similarity matrix in Chen, Shaw, Yu et al. 2011) to quantify landscape changes over the chosen spatial units.

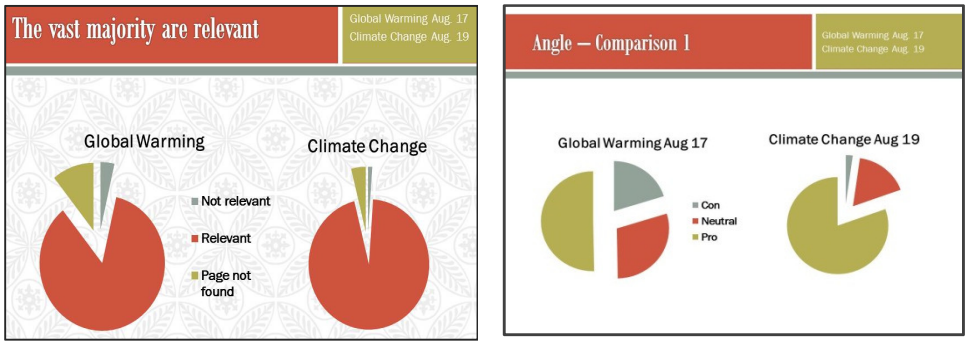

Figure 2. a (left): The classified websites are mostly relevant; b (right): Attitudes (pro, con, or neutral) towards "global warming" and "climate change" are different (from data of mid-August 2011 only).

3) Multivariate regression: We applied multivariate regression (e.g., OLS, binomial or multinomial logistic regression) in SAS in order to explore what biophysical, socioeconomic, and demographic factors may explain the space-time patterns related to reflections, attitudes, or perceptions towards "climate change" and "global warming". Various model correctness or fit measures (e.g., R-square or pseudo R-square, AIC) were used. The R-squares varied from 0.35 to 0.68. In particular, we conducted multilevel regression analysis in MLwiN. Preliminary multilevel regression results indicate that a multilevel model is very useful for the data. Further modeling will investigate multilevel (the levels of occasion, county, state, and region), cross-level, and temporal relationships between website activity (as resulting from our processing models) and our county and state database variables.

4) Model development: We developed tools, including modules, procedure, and macros in SAS, to further facilitate space-time analysis. Survival analysis (a special type of regression) has been heavily used in this regard, in which hazards (of a certain place being "dominated" by interest in climate change or global warming) were linked with a set of independent variables. Survival analysis is especially appropriate to deal with independent variables taking changing values over time or data related to imprecise time stamps for events.

References

1. Allison, P. D. 1995. Survival analysis using SAS: A practical guide. Cary, NC: SAS Institute.
2. An, L., and D. G. Brown. 2008. Survival analysis in land-change science: integrating with GIScience to address temporal complexities. Annals of Association of American Geographers, 98(2): 323-344.
3. An, L., D. G. Brown, J. Nassauer, and B. Low. 2011. Variations in Development of Exurban Residential Landscapes: Timing, Location, and Driving Forces. Journal of Land Use Science, 6 (1): 13–32.
4. Anselin, L., Syabri, I., and Kho, Y. 2006. GeoDa: An introduction to spatial data analysis. Geographical Analysis, 38(1):5–22.
5. Chen, J., S-L Shaw, H Yu, F. Lu, Y, Chai, Q. Jia. 2011. Exploratory data analysis of activity diary data: a space-time GIS approach. Journal of Transport Geography 19: 393-404.
6. Pontius Jr, RG, S Peethambaram, and J-C Castella. 2011. Comparison of three maps at multiple resolutions: a case study of land change simulation in Cho Don District, Vietnam. Annals of the Association of American Geographers 101(1): 45-62.