Guest editorial: Integrated spatio-temporal analysis and data mining

Tao Cheng

With the arrival of the Big Data wave, the open availability of data with higher spatial, temporal, and thematic resolution enables us to better address complex scientific and social questions. Meanwhile, understanding, sharing, and reusing these data becomes more challenging given their high dimensionality and inter-connectedness. Spatio-temporal data mining (STDM) is the extraction of unknown and implicit knowledge, structures, relationships, or patterns from large amount of space-time data [1]. It emerged out of a need to create effective and efficient techniques in order to turn massive data into meaningful information and knowledge and is a key tool in analysing Big Data with spatiotemporal extent. This special issue includes three articles on understanding patterns and relationships of natural phenomena in three different applications.

In "*Multi-scale Decomposition of Point Process Data*", the patterns of earthquakes in a reservoir area are analysed as a mixture of a finite number of homogeneous point processes. The decomposition is proposed to automatically identify arbitrarily-shaped clusters in point data in three steps. Firstly, an objective function of the kth nearest distance is constructed, where a point data set is modelled as a mixture of probability density functions (pdf) of different homogeneous processes. Secondly, the mixture pdfs are separated into distinct pdfs, as a binary tree in which each leaf represents a homogeneous process. Thirdly, distinct clusters are generated from each homogeneous point process according to the density connectivity of the points. The case study clearly shows the spatial point patterns of earthquakes in a reservoir area. The spatiotemporal relationship between the main earthquake and the clustered earthquake (namely, foreshocks and aftershocks) was also revealed.

In "*Cluster recognition in spatial-temporal sequences: the case of forest fires*", forest fire sequences are modelled as a stochastic point process. The space-time scan statistics permutation model (STSSP) is applied to identify the hotspots in forest fire sequences. STSSP uses a scanning window, moving across space and time, to detect abnormal events in specific areas over a certain period of time. The method is tested by the case study of forest fires registered in Canton Ticino (Switzerland) from 1969 to 2008. Results revealed that forest fire events in Ticino are mainly clustered in the southern region where most of the population is settled. It also uncovered local hot spots arising from extemporaneous arson activities.

"*Assessing the changing flowering date of the common Lilac in North America: a random coefficient model approach*" investigates changes in the onset of the North American spring

T. Cheng (✉)
SpaceTimeLab, Department of Civil, Environmental & Geomatic Engineering, University College London, Gower Street, London WC1E 6BT, UK
e-mail: tao.cheng@ucl.ac.uk

by using the first bloom dates of lilacs from 1956 to 2003. The data were collected through Volunteered Geographical Information (VGI) and by expert researchers. It is argued that care must be taken when analysing data of this kind, with particular focus on the issues of lack of experimental design, and Simpson's paradox. It makes use of random coefficient modelling and bootstrapping approaches to overcome this issue, and a gradual advance in the onset of spring is suggested by the results of the analysis. A key lesson learned is that the appropriateness of the model calibration technique used given the process of data collection needs careful consideration.

Reference

1. Cheng T, Haworth J, Anbaroglu B, Tanaksaranond G, Wang J (2012) Spatio-temporal data mining. In: Fischer MM, Nijkamp P (eds) Handbook of regional science, Springer, Heidelberg (in press)

Tao Cheng is a Professor in GeoInformatics, and Director of SpaceTimeLab at the University College London. Her research interests span Geocomputation, network complexity, integrated spatio-temporal data mining (prediction, clustering, visualisation and simulation) and uncertainty of geographic information, with applications in transportation, crime, health, natural hazards, resource management, and location-based services.