Investigation of organized hydrological heterogeneity from a spatial analysis perspective

HONGYI LI[1] & MURUGESU SIVAPALAN[1,2]

1 *Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Illinois, USA*
hli23@uiuc.edu

2 *Department of Geography, University of Illinois at Urbana-Champaign, Illinois, USA*

Abstract Spatial analysis as a framework for exploring spatial patterns is well known in social-geography, econometrics and other fields, but has seldom been applied in catchment hydrology. This paper presents a new method for the investigation of the spatial patterns of hydrological variables from a *spatial analysis* perspective. Using the topographic index (TI) as an example, the preliminary yet promising results obtained from this analysis suggest that the Global Moran's I index constructed from an asymmetric inverse flow distance weight matrix is an effective bulk signature of hydrological spatial structure, and the Local Indicator of Spatial Association (LISA) captures well the transition from hydrological heterogeneity to hydrological homogeneity with increasing spatial scale. These results suggest that complementary to existing statistical methods, spatial analysis could be a promising framework for quantifying and modelling spatial patterns exhibited by hydrological variables.

Key words spatial analysis; TI; Global Moran's I; asymmetric inverse flow distance weight matrix; LISA

INTRODUCTION

The significance of spatial structure within hydrological variables has been extensively appreciated and studied (Western *et al.*, 2001; Milzow *et al.*, 2004; Marani *et al.*, 2006; Valdes *et al.*, 2006), and it has also been realized that standard statistical methods and even geostatistical methods are not sufficient to detect and quantify such spatial structures (Western *et al.*, 2001; Cai & Wang, 2006). The existence of spatial structure violates the assumption of standard statistical hypothesis tests (Cai & Wang, 2006), which assume that spatial variables are independent, thus automatically neglecting the close correlation that exists in space. Standard geostatistics is able to represent spatial continuity, but not spatial structure/organization. Western *et al.* (2001) convincingly demonstrated that the hydraulic connectivity among spatial hydrological variables, e.g. soil moisture, could not be captured and quantified by standard geostatistical methods.

Spatial analysis, however, is an alternative and effective framework for studying spatial patterns within hydrological variables, besides standard statistics and geostatistics. Cai & Wang (2006) used Moran's I to quantify the global autocorrelation of the Topographic Index (TI), and examined the impacts of spatial scale and resolution on the Moran's I of TI. However, the Moran's I in their work is constructed from an Inverse-Euclidean-Distance weight matrix without considering the hydrological connection among spatial locations, and therefore can actually only capture the spatial *continuity* of TI. We extend their work by constructing Moran's I from an asymmetric-Inverse-Flow-Distance weight matrix, which has more significant hydrological meaning than the Inverse-Euclidean-Distance weight matrix, and enables us to investigate spatial *connectivity*. Moran's I as a bulk indicator integrates the whole spatial structure into one single value. This is convenient for comparison among various basins (Cai & Wang, 2006) after standardization, but is not suitable for exploring the detailed spatial structures which might shed light on the underlying physical mechanisms. We therefore also apply another indicator, the Local Indicators of Spatial Association (LISA), to the TI of the same watershed. Interestingly, we find that LISA is able to effectively capture the apparent transition from hydrological heterogeneity to homogeneity that occurs with increasing spatial scale.

In this paper we first introduce Global Moran's I and LISA. We then apply these two indicators on TI from a natural watershed, and present the results along with some discussion of the results. Finally we draw a few preliminary conclusions.

METHODOLOGY

Spatial autocorrelation refers to the correlation of the values of one single variable through a spatial region, i.e. the value of a variable at one location is related to the values of the same variable at other neighbouring locations. Global Moran's I and LISA are the most widely used spatial statistical tools for quantifying spatial autocorrelation (Anselin, 1995, 2004). The mathematical description of Global Moran's I is as follows:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(X_i - \overline{X})(X_j - \overline{X})}{S_0 \sum_{i=1}^{n}(X_i - \overline{X})^2} \tag{1}$$

where X is the variable of interest, i and j indicate different locations, $\overline{X}$ is the average of X through the whole study area, W_{ij} is the so-called spatial weight between X_i and X_j, which quantifies the level of relation between these two values. The larger the value of W_{ij}, the more closely correlated X_i and X_j are. $S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}$ is used here as a normalization value such that the range of I is from –1 to 1. Positive values indicate spatial similarity, i.e. clustering of similar values, and negative values indicate spatial dissimilarity, i.e. clustering of dissimilar values. Zero value indicates no spatial autocorrelation, i.e. spatial randomness. Considering "similarity" might have a different meaning in hydrology. We hereafter refer to spatial similarity as spatial homogeneity, and spatial dissimilarity as spatial heterogeneity.

W_{ij} could be flexibly defined according to the types of spatial relation, for example, W_{ij} can be defined as inverse Euclidean distance D_{ij}, i.e. $W_{ij} = 1/D_{ij}$. This definition is consistent with the First Law of Geography, which states that "every thing is related to something else, but nearer things are more related than distant things" (Tobler, 1970). However, the global Moran's I based on this definition of spatial weight matrix is just a measure of spatial continuity. In this paper, W_{ij} is defined as inverse flow distance, and flow distance is the distance between any two locations along a natural flow path. The difference between Euclidean distance and flow distance is illustrated in Fig. 1. Note that $[W_{ij}]$ is a spatial weight matrix, and hence $[W_{ij}]$ constructed from inverse asymmetric flow distance is an asymmetric matrix, which is the one utilized in this work.

It is often assumed in spatial analysis that the values at two locations will be independent if the distance between them exceeds a certain value, i.e. threshold distance (hereby noted as $D_{threshold}$). That is, if the two locations are far enough, they are not related, thus the corresponding weight is automatically zero. Therefore, for each value of threshold distance $D_{threshold}$ there will be a corresponding weight matrix, and thus a corresponding value of Global Moran's I.

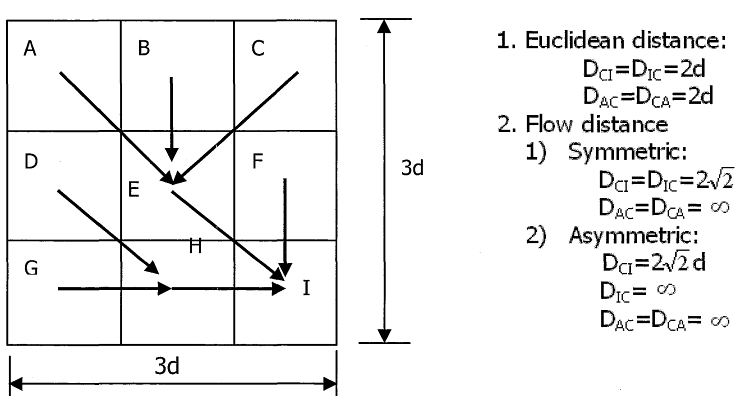

Fig. 1 Euclidean distance and flow distance (arrows indicate flow direction).

Notice that although Global Moran's I is a dimensionless number, there is no theoretical foundation to directly compare the values of Global Moran's I (for different variables within the same region or the same variable within different regions). A more rigorous way is to use standardized Global Moran's I (Luc Anselin, personal communication), calculated as follows:

$$z = E(I) / SE(I) \qquad (2)$$

where $E(I)$ and $SE(I)$ are the average and standard deviation of Global Moran's I, respectively. A positive standardized Global Moran's I means spatial similarity, whereas a negative value means spatial dissimilarity.

According to its definition, Global Moran's I is a bulk indicator of spatial correlation integrated over the whole region, and cannot help detect detailed information on spatial processes within a region. Local Indicators of Spatial Association (LISA) are able to represent local spatial autocorrelation for each location, and thus are a powerful tool to explore spatial processes within the region of interest (Anselin, 1995). Notice that LISA is a generic name of a series of local spatial statistics. Local Moran's I, as a special case of LISA, can be defined as:

$$I_i = (X_i / m_2) \sum_{j=1}^{n} W_{ij} X_j \qquad (3)$$

where $m_2 = \sum_i X_i^2$ does not vary with location i. Global Moran's I is the spatial mean of Local Moran's I.

RESULTS AND DISCUSSION

Global Moran's I of TI

A small watershed with a drainage area of about 3 km^2 in Brown County, Indiana, USA, was selected for this illustrative study. Topographic Index (TI) (Beven & Kirby, 1979) has significant hydrological meaning and exhibits spatial structure, and is therefore selected as the variable under study. But there is no doubt that these spatial analysis tools can be applied to any other hydrological variable that exhibits spatial structure. TI of this watershed is calculated based on a 30 × 30 DEM, as shown in Fig. 2.

Table 1 shows Global Moran's I calculated from different distance types and different threshold distances. Generally speaking, the values of standardized Global Moran's I from inverse

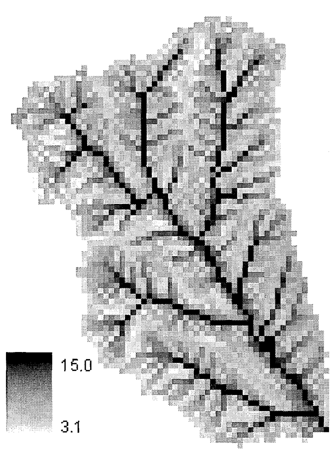

Fig. 2 Distribution of TI in the watershed under study.

Table 1 Moran's I and Z values under different weight matrices.

$D_{threshold}$	Distance type	Global Moran's I	Standardized Global Moran's I
30 m	Euclidean distance	0.4406	32.34
30 m	Flow distance	0.7284	44.06
150 m	Euclidean distance	0.1380	41.94
150 m	Flow distance	0.6045	80.37

flow distance weight matrices are usually greater than those from inverse Euclidean distance weight matrices. Although not fully presented here, the results under other threshold values also confirm this conclusion. The reason for this phenomenon is that the probability of two locations on the same flow path having similar values (inverse flow distance weight) is larger than that of two locations close to each other, but not on the same flow path (inverse Euclidean distance weight), while the probability of two locations on the same flow path having dissimilar values is smaller than that of two locations close to each other while not on the same flow path. This would be more easily understood by considering two locations, one in the river, and another on a hillslope close to the previous one, but with the two locations not being on the same flow path. The values of TI will be significantly different between them. The weight from inverse Euclidean distance is a fairly large value due to short Euclidean distance, which leads to smaller Moran's I due to the dissimilarity between the values of TI. The weight from inverse flow distance is zero since they are not hydraulically connected, and thus will not affect the value of Moran's I.

Impact of threshold distance

Threshold distance is another important factor affecting Moran's I. There are, in fact, two methods to define neighbouring area of a location with the help of threshold distance (Luc Anselin, personal communication). One is called the cumulative neighbourhood method, which defines all locations having an effective distance smaller than the given threshold distance as neighbours. Such a neighbourhood is like a dish. The previous results are based on this method. The other one is called the band neighbourhood method, which defines all locations having an effective distance larger than a minimum distance $D_{threshold,\min}$ while smaller than a maximum distance $D_{threshold,\max}$ as neighbours. Such a neighbourhood is like a circular band, or a donut. The difference between $D_{threshold,\max}$ and $D_{threshold,\min}$ is called the width of the band, and is usually a constant. The sensitivity of standardized Global Moran's I on threshold flow distance is examined in this work with the cumulative neighbourhood method and the band neighbourhood method, respectively.

Figure 3(a) shows that, under the cumulative neighbourhood method, standardized Global Moran's I of TI first increases quickly, reaches the maximum at a threshold flow distance of about 500 m, and then decreases relatively slowly with increasing threshold flow distance. This is consistent with Fig. 3(b), where the band width in Fig. 3(b) is 30 m in this work, and the threshold distance is estimated as the mean of $D_{threshold,\max}$ and $D_{threshold,\min}$. In Fig. 3(b), under the band neighbourhood method, the standardized Moran's I of TI starts as a positive value (similarity among nearby neighbours), decreases to zero at a threshold distance of 500 m, and then flips to negative. The abrupt increase of the standardized Moran's I of TI from about 2600 m is probably caused by error since the longest flow distance in the study area is about 2600 m, and can thus be neglected. A possible hypothesis which could be inferred from Fig. 3 is that the values of TI at two locations having a certain distance (here about 500 m) might not be related, i.e. they are totally independent.

Local Moran's I

As we have discussed previously, the values of local Moran's I are the local components of Global Moran's I, and a map of local Moran's I is a detailed description of the spatial structure of the hydrological variable of interest.

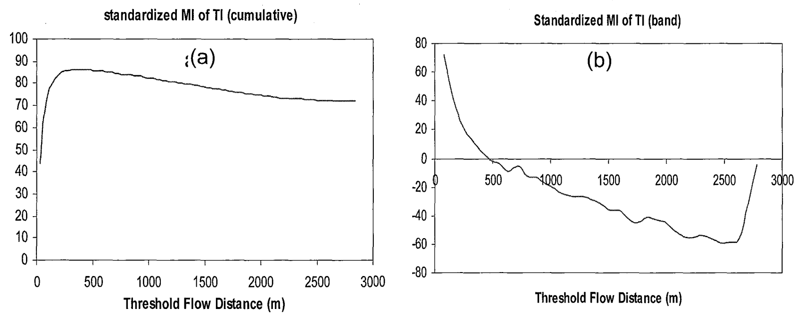

Fig. 3 Standardized Global Moran's I (under asymmetric inverse flow distance weight matrix) changing with threshold distance.

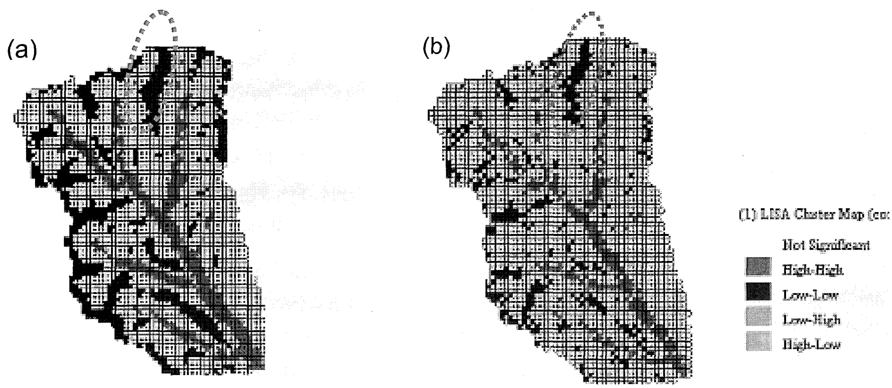

Fig. 4 LISA cluster maps under traditional weight matrices: (a) queen contiguity weight, (b) inverse Euclidean distance weight.

Figure 4 shows the maps of local Moran's I calculated from traditional weight matrices: queen contiguity weight matrix and inverse Euclidean distance weight matrix. Queen contiguity is a type of contiguity (the other two are rook contiguity and bishop contiguity). If two spatial units share a common boundary or a common vertex, then we define their relationship as queen contiguity. Notice that our work is based on 30 × 30 DEM, i.e. grid areas. These maps are calculated with software *GeoDa* (Anselin, 2004a) and *R* (Anselin, 2004b). The High-High clusters are the locations with high TI values surrounded by high TI values (similarity), Low-Low clusters are the locations with low TI values surrounded by low TI values (similarity), Low-High clusters are the locations with low TI values surrounded by high TI values (dissimilarity), and High-Low clusters are the locations with high TI values surrounded by low TI values (dissimilarity). The common feature of Fig. 4(a) and (b) is that the grids near ridges are defined as Low-Low clusters as highlighted by dashed circles, i.e. they are defined as similar neighbours.

However, some grids near ridges are not real neighbours from a hydrological perspective. Although some grids at the ridges are adjacent, they may not be hydrologically connected, since water will disperse in different directions. The probability of them being on a common flow path is considerably low. So according to the definition of inverse flow distance weight, the weight between such two locations at the ridges is most probably zero. This fact is well reflected by the maps of standardized local Moran's I shown in Fig. 5. It can thus be concluded that LISA from inverse flow distance weight matrix is more reasonable than the LISA from other weight matrices, in the sense that it effectively removes the hydrologic-pseudo-neighbours.

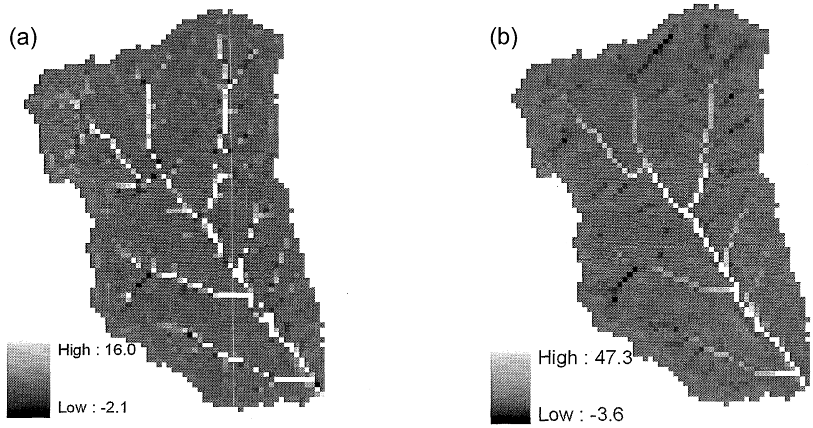

Fig. 5 Standardized Local Moran's I under inverse flow distance weight matrix: (a) threshold flow distance = 30 m, (b) threshold flow distance = 3000 m.

Another interesting pattern from the inverse flow distance weight matrix method (Fig. 5), which is not imposed by the results of the other two methods (Fig. 4), is that the LISA values increase along the streams in the downstream direction, i.e. from negative (significant heterogeneity) at the headwater cells to positive (significant homogeneity) at the downstream cells. Note that the grid cells within the river reaches are **river cells**. According to the definition of inverse flow distance weight matrix, and thus the corresponding definition of hydrological neighbours, for those river cells at the headwaters, their neighbours are mainly the cells on the hillslope (hereafter noted as **hillslope cells**). Noticing that the river cells have high TI values while the hillslope cells have low TI values, it is not difficult to expect dissimilarity among the most upstream river cells and their neighbours. However, for any cell on the main stream, river cells consist of a large part of its neighbours, and because the river neighbour cells have apparently larger weights (due to shorter flow distances) than hillslope neighbour cells, i.e. the river neighbour cells have more impacts than the hillslope cells, so it is reasonable to expect similarity among the cells on the main stream. The increasing of LISA values along the channel in the downstream direction is especially clearly shown in Fig. 5(b). Together with the increasing of LISA values, the drainage area, i.e. spatial size, is also increasing in the downstream direction. That is, a river cell with a larger drainage area is also having a larger LISA value, which implies more spatial homogeneity.

CONCLUSIONS

From the above discussions, a few preliminary conclusions can be drawn:

- Spatial analysis is a promising framework to detect and quantify spatial pattern/structure besides standard statistics and geostatistics, as illustrated here with standardized Global Moran's I and Local Moran's I.
- From a hydrological perspective, inverse flow distance weight matrix has more hydrological meaning than inverse Euclidean distance proposed by Cai & Wang (2006), and thus has more potential to be applied for hydrological modelling on a catchment scale.
- Comparing with Global Moran's I, LISA, here referring to local Moran's I based on inverse flow distance weight matrix, is more able to detect local spatial pattern and help us to detect spatial trends in a region.
- LISA of a hydrological variable nicely represents the transition from spatial heterogeneity (dissimilarity) to homogeneity (similarity) from the headwater to the watershed outlet, with

increasing drainage area. It is the authors' feeling that, although further theoretical clarification is necessary, this result could possibly be used as one of the ways to estimate the reasonable size of the Representative Elementary Watershed (REW) (Reggiani *et al.*, 1998) for distributed hydrological modelling.

Acknowledgements The first author wants to thank Luc Anselin, Dingbao Wang and Ximing Cai for the valuable discussions and suggestions.

REFERENCES

Anselin, L. (1995) Local indicators of spatial association–LISA. *Geograph. Anal.* **27**(2), 93–115.

Anselin, L. (2004a) Exploring spatial data with GeoDa. *A Workbook, Spatial Analysis Laboratory*, 129–147. Spatial Analysis Laboratory, Dept. of Agricultural and Consumer Economic and CSISS, University of Illinois at Urbana-Champaign, Illinois, USA.

Anselin, L. (2004b) Spatial regression analysis in R: *A Workbook, Spatial Analysis Laboratory*, 13–34. Spatial Analysis Laboratory, Dept. of Agricultural and Consumer Economics and CSISS, University of Illinois at Urbana-Champaign, Illinois, USA.

Cai X. & Wang, D. (2006) Spatial autocorrelation of topographic index in catchments. *J. Hydrol.* **328**, 581–591.

Marani, M., Silvestri, S., Belluco, E., Ursino, N., Comerlati, A., Tosatto, O. & Putti, M. (2006) Spatial organization and ecohydrological interactions in oxygen-limited vegetation ecosystems. *Water Resour. Res.* **42**, W06D06, doi:10.1029/-2005WR004582.

Milzow, C., Molnar, P., McArdell, B. W. & Burlando, P. (2006) Spatial organization in the step-pool structure of a steep mountain stream (Vogelbach, Switzerland). *Water Resour. Res.* **42**, W04418, doi:10.1029/2004WR003870.

Reggiani, P., Sivapalan, M. & Hassanizadeh, S. M. (1998) A unifying framework for watershed thermodynamics: Balance equations for mass, momentum, energy, entropy and the 2nd law of thermodynamics. *Adv. Water Resour.* **22**(4), 367–398.

Tobler, W. R. (1970) A computer movie simulating urban growth in the Detroit region. *Econom. Geography Suppl.* **46**, 234–240.

Valdes, D., Dupont, J.-P., Laignel, B., Ogier, S., Leboulanger, T. & Mahler, B. J. (2007) A spatial analysis of structural controls on Karst groundwater geochemistry at a regional scale. *J. Hydrol.* **340**, 244–255.

Western, A. W., Blöschl, G. & Grayson, R. B. (2001) Towards capturing hydrologically significant connectivity in spatial patterns. *Water Resour. Res.* **37**(1), 83–97.