Locating irregularly shaped clusters of infection intensity

Niko Yiannakoulias[1], Shona Wilson[2], H. Curtis Kariuki[3], Joseph K. Mwatha[4], John H. Ouma[4], Eric Muchiri[3], Gachuhi Kimani[4], Birgitte J. Vennervald[5], David W. Dunne[2]

[1]*School of Geography and Earth Sciences, McMaster University, Hamilton, Ontario L8S4K1, Canada;* [2]*Department of Pathology, University of Cambridge, UK;* [3]*Division of Vector Borne Diseases, Ministry of Health, Nairobi, Kenya;* [4]*Kenya Medical Research Institute, Nairobi, Kenya;* [5]*DBL-Centre for Health Research and Development, Faculty of Life Science, University of Copenhagen, Copenhagen, Denmark*

Abstract. Patterns of disease may take on irregular geographic shapes, especially when features of the physical environment influence risk. Identifying these patterns can be important for planning, and also identifying new environmental or social factors associated with high or low risk of illness. Until recently, cluster detection methods were limited in their ability to detect irregular spatial patterns, and limited to finding clusters that were roughly circular in shape. This approach has less power to detect irregularly-shaped, yet important spatial anomalies, particularly at high spatial resolutions. We employ a new method of finding irregularly-shaped spatial clusters at micro-geographical scales using both simulated and real data on *Schistosoma mansoni* and hookworm infection intensities. This method, which we refer to as the "greedy growth scan", is a modification of the spatial scan method for cluster detection. Real data are based on samples of hookworm and *S. mansoni* from Kitengei, Makueni district, Kenya. Our analysis of simulated data shows how methods able to find irregular shapes are more likely to identify clusters along rivers than methods constrained to fixed geometries. Our analysis of infection intensity identifies two small areas within the study region in which infection intensity is elevated, possibly due to local features of the physical or social environment. Collectively, our results show that the "greedy growth scan" is a suitable method for exploratory geographical analysis of infection intensity data when irregular shapes are suspected, especially at micro-geographical scales.

Keywords: disease clusters, schistosomiasis, hookworm, spatial scan, Kenya.

Introduction

Geographic cluster detection methods can be employed to identify localised geographic patterns of disease, and have been applied in a variety of decision support and research contexts. These methods are used to find geographic "hot-spots" where illness and/or infection is highly concentrated in the population. One of the most commonly used methods is the spatial scan (Kulldorff, 1997). In its general form, the spatial scan enumerates a large number of potential geographic clusters in a study area in order to determine which is the most likely to have caused the rejection of a null hypothesis of constant risk. This cluster is referred to as a "most-likely cluster", and is tested for significance using Monte Carlo methods. In most instances, there are a large number of potential clusters; in a study area of 40 locations, there are over one-thousand billion different possible cluster sets. However, geographic clusters usually require some sort of geographic structure to be of interest, so the original spatial scan method is restricted to circularly shaped clusters, which also reduces the total number of possible clusters searched. The drawback of this and similar fixed search geometries is that they are less able to detect hot-spots of irregular shape; for example, a

Corresponding author:
Niko Yiannakoulias
School of Geography and Earth Sciences
McMaster University, 1280 Main Street West
Hamilton Ontario L8S4K1, USA
Tel. +1 905 525 9140 ext. 20117; Fax +1 905 546 0463
E-mail: yiannan@mcmaster.ca

circular window may not detect a spatial anomaly that follows a river, highway or the shores of a lake.

For illnesses in which water is part of a parasite or vector life cycle or a route of transmission between hosts, living in close proximity to water is often an important predictor of infection prevalence and intensity in humans. Studies have shown a relationship between proximity to water and schistosomiasis (Clements et al., 2006; Brooker and Clements, 2009), malaria (Kleinschmidt et al., 2001; Gemperli et al., 2006), cholera (Ali et al., 2002) and other infectious diseases. In spite of the role of water as an exposure medium for many illnesses, few methods of spatial analysis are designed to explicitly identify spatial patterns related to water. While many methods of analysis can identify how proximity to water may be associated with risk of disease, few are specifically capable of finding spatial heterogeneities over and above the baseline association between risk and proximity to water. This is a particularly important shortcoming at micro-spatial scales, where subtle features of waterways—such as water velocity, water clarity, mineral content, soil type and local ecology—may result in important local heterogeneities in risk.

In recent years, new methods have been developed to search for disease clusters without using fixed search geometries (Duczmal and Assunçao 2004; Patil and Taille 2004; Tango and Takahashi 2005; Assunção et al., 2006; Yiannakoulias et al., 2007; Duczmal et al., 2008). Although many of these methods differ with respect to the details of how clusters are found, all are free of explicit constraints about shape and may be of particular value in spatial analysis of disease where roads, waterways and other irregularly-shaped features of the physical landscape may influence the geographic distribution of disease. These methods may be particularly well suited to the analysis of patterns of infectious disease in which local variations in a water regime may play an important role in characterising risk.

In this paper, we apply a recently developed method of cluster detection to identify clusters of infection intensity at a micro-spatial scale. We compare this method to the traditional circular window spatial scan by comparing the ability of these two methods to detect a synthesised cluster occurring near a river. Our hypothesis is that methods of cluster detection free from geometric constraints are more likely to find a cluster along a waterway than the traditional circular window spatial scan approach. We then apply both methods to detect clusters of *Schistosoma mansoni* and hookworm infection intensity in a small community in Kenya.

Materials and methods

There are two parts to this study. The first part compares the ability of a "greedy growth scan" (Yiannakoulias et al., 2007) and circular window spatial scan (Kulldorff, 1997) to detect clusters of infection intensity that may be of irregular (and particularly, non-circular) shapes. The second part applies these methods to real data on *S. mansoni* and hookworm infection intensity in humans, two helminth infections arising from faecal contamination of the environment but with differing modes of transmission (Roberts and Janovy, 1996). *S. mansoni* is dependent on permanent water bodies for its transmission, due to an indirect life cycle that involves aquatic snails of the *Biomphalaria* genus as an intermediate host (Gryseels et al., 2006). Hookworm has a direct life cycle, with larvae hatching from eggs in contaminated soil, which then develop into the infective L_3 stage (Bethony et al., 2006). Both parasites depend on environmental factors such as pH, temperature and moisture for the successful completion of their life-cycles. Coupled with the role of human behaviour in transmission, transmission and prevalence of both parasites is often at a focal or micro-geographical scale (Brooker et al., 2006; Wilson et al., 2007).

Data

Individual *S. mansoni* and hookworm infection intensity measurements were made for a random selection of individuals (n = 520) in the village of

Kitengei, Makueni district, Kenya, a small community of roughly 1,400 permanent residents. Infection intensity was measured as the arithmetic average from two Kato-Katz thick smears over three stool samples, to give egg per gram (epg) counts. Since there is likely to be household clustering of *S. mansoni* and hookworm infection intensity in rural communities (Brooker et al., 2006), detecting clusters in the village may be influenced by clustering within households. To avoid this, we restricted our analysis to the persons with the highest epg in each household. This reduction of data increases type II error, and makes our analysis more likely to fail to reject a null hypothesis of constant spatial pattern in infection intensity.

Locations of households were identified in the field using a global positioning system. We created a Voronoi tessellation based on these household points to create discrete zones for each household, and to construct a topology file defining neighbouring household zones. This file is required for the cluster detection process described below.

The "greedy growth scan"

In this application, the "greedy growth scan" searches for clusters by grouping geographic zones together in a way that creates subgroups or clusters of zones with high infection intensity (Yiannakoulias et al., 2007). Clusters must be geographically contiguous; two or more zones can be grouped together only if they neighbour each other. The search for these clusters is conducted using a greedy agglomeration algorithm which "grows" clusters by agglomerating areas in a way that increases a likelihood ratio; potential groupings that result in clusters with larger likelihood ratios are chosen in favour of potential groupings that result in clusters with lower likelihood ratios. To ensure that the algorithm searches an entire study area thoroughly, the search processes initiates from all zones within a study region. The cluster found with the largest likelihood ratio test statistic is treated as the most likely cluster to cause the rejection of a null

hypothesis of constant risk.

In disease endemic areas, morbidity associated with *S. mansoni* and hookworm infections is usually a function of infection intensity; infection with a greater number of parasites usually corresponds with more symptoms and sickness. The Poisson and Bernoulli likelihood ratio tests commonly used in spatial scan methods require that individual persons are characterized into dichotomous states (e.g. diseased or not, severely infected or not). Such dichotomization would discard available and potentially useful information about infection intensity. An alternative approach is to use an exponential model, which is suitable for searching for clusters in which observations are measured on a continuous and discrete scale (Huang et al., 2007). The exponential model was developed to search for clusters of survival under the null hypothesis that survival time is geographically homogenous. We replace survival time with infection intensity, and therefore, search for clusters under the null hypothesis of constant spatial infection intensity. For each potential cluster, the exponential likelihood ratio test statistic is calculated, and the cluster with the largest test statistic is treated as the cluster most likely to reject a null hypothesis of constant infection intensity. For most likelihood models associated with the spatial scan, tests of significance are based on Monte Carlo simulations. The nature of the exponential model makes this difficult, because the distribution of infection intensity is unknown. As an alternative, we conduct a permutation test in which a random re-arrangement of the intensity measurements over all locations is used to make inferences about statistical significance (Huang et al., 2007).

Like other irregular-shaped cluster-detection methods, the "greedy growth scan" tends to find clusters of highly irregular shape, particularly when there are considerable spatial heterogeneities in the data. To ameliorate this, a non-connectivity penalty is applied to the clusters to ensure that the cluster shape is not highly irregular (Yiannakoulias et al., 2007). The challenge of this penalty is that a user

has to set the non-connectivity parameter, ideally, before the data are actually analysed. In our analysis, we apply a weak non-connectivity penalty (of 0.25) to all analyses of simulated and real data since we suspect local features of the environment could result in clusters of truly irregular shape. We also set a population threshold such that no cluster should include more than 30% of the population for all runs (of both the circular window spatial scan and the "greedy growth scan" methods). This is because our emphasis is on finding localised clusters of unusual shape, rather than general heterogeneities in infection intensity.

Simulation analysis

To create synthesised clusters, we randomly permute observed infection intensity levels across locations. We demarcate a region near the river as a synthetic cluster area. To ensure that the synthetic cluster area has high infection intensity detectable by the "greedy growth scan" and circular window spatial scan methods, we weight the permutation process so that zones inside the cluster area have a greater tendency to include high infection intensity observations than locations outside the cluster. This weighting parameter, φ, reflects increasing concentration of infection intensity within the synthetic cluster area where a value of $\varphi = 1$ is a random permutation, and a value of $\varphi = 10$ has all the highest infection intensity areas inside the synthesised cluster area. For each synthesised dataset, the "greedy growth scan" and circular window spatial scan are used to search for a cluster most likely to cause the rejection of a null hypothesis of constant spatial infection intensity.

We compare the "greedy growth scan" and the circular window spatial scan methods in two ways. First, we calculate the true positive and false positive rates based on whether or not the synthesised hot-spot area was detected by the cluster detection algorithm. We consider the synthesised hot-spot area detected if at least half of this hot-spot area is found. This generous cut-off ensures some compa-

rability between the methods; it is very likely that if the detection threshold was restricted to the whole cluster area the circular window spatial scan would never detect it. By setting the threshold to half the area, a method is considered successful even when it encloses only part of the synthesised hot-spot area. Second, we map the proportion of simulations that each zone is included in a hot-spot found by both methods. Zones inside the synthesised cluster area should be found more often than zones outside of the synthesised cluster area. For each cluster, we report the average household infection intensity.

Analysis of clusters of S. mansoni *and hookworm infection intensity*

We use egg counts of *S. mansoni* and hookworm to characterise infection intensity in each zone. We search for clusters using the "greedy growth scan" and circular window spatial scan methods. The settings of the two methods are the same as in the simulation above, except a larger number of permutations are used (9,999). The most-likely clusters found by both methods are presented on maps. These maps are combined with maps of spatially-kriged average household infection intensity. These maps reveal the general spatial pattern of infection intensity by interpolating infection intensities between households, and helps provide a reference for comparing the two methods of cluster detection.

Results

Simulation

Figure 1 illustrates the relative abilities of the "greedy growth scan" and circular window spatial scan methods for detecting zones within the synthesised cluster. Each point estimate is based on the number of true positive clusters found over 1,000 simulations. When the distribution of infection intensity is random ($\varphi = 1$), neither method effectively detects clusters in the synthesised cluster area. As
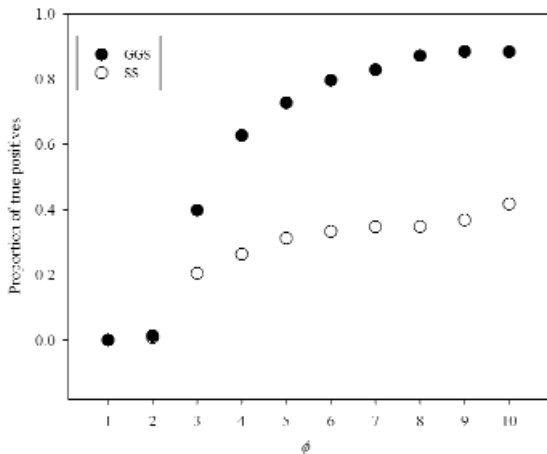
Fig. 1. Proportion of true positives for "greedy growth scan" (GGS) and circular window spatial scan (SS).

φ increases, and infection intensity is more likely to cluster within the synthesised cluster area, both search methods become more capable of identifying statistically significant clusters, though the "greedy growth scan" improves at a faster rate. Both methods show a flattening out of their proportion of true positives around $\varphi = 7$ or 8, with the "greedy growth scan" possessing a success rate almost twice the circular window spatial scan.

Figure 2 illustrates the locations included in found clusters for the simulation. If a method worked perfectly, all zones within the cluster would have a value of 1.00 (and would be shaded black) and all other zones would have a value of 0.00 (and would be shaded white). For $\varphi = 3$, both methods identify
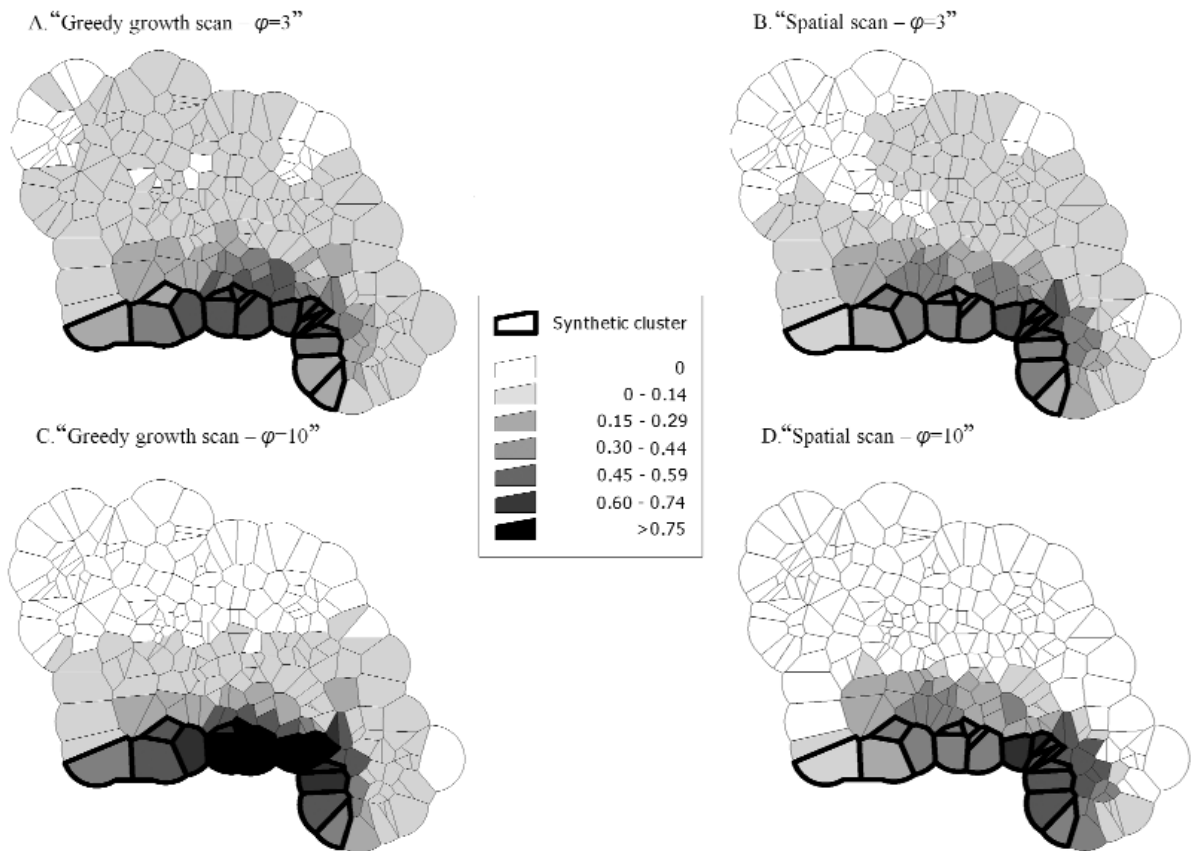

Fig. 2. Map illustrating the proportion of times a zone was included in a cluster for 1,000 simulations at a given value of φ.

many zones outside of the synthesised cluster as being part of a cluster. The "greedy growth scan" appears more likely to include false positive zones than the circular window spatial scan, as it more frequently includes non-cluster zones inside a cluster. For $\varphi = 10$, both methods focus more on the synthesised cluster area, though it is apparent that the "greedy growth search" includes more zones that are not part of the synthesised cluster. The circular window spatial scan tends to concentrate around the synthesised cluster area, however since the clustering is focused on the river area, and the circular search window cannot enclose it effectively, the method tends not to find large portions of the synthetic cluster area.

Clusters of S. mansoni *and hookworm*

Figure 3 shows the map of the found most likely clusters of *S. mansoni* infection intensity for both methods; the circle identifies the most-likely cluster found by the circular window spatial scan, and the irregular shape identifies the cluster found by the "greedy growth scan". The *S. mansoni* cluster found by the "greedy growth scan" is statistically significant (P = 0.007), with an average household infection intensity of 193 epg within the cluster. The *S. mansoni* cluster found by the circular window spatial scan is not statistically significant (P = 0.239) and has an average household infection intensity of 664 epg. The cluster found by the circular window spatial scan is included in the cluster found by the "greedy growth scan", but the latter also includes a number of other areas.

Both the "greedy growth scan" and the circular window spatial scan find statistically significant hookworm clusters, with significance values of 0.026 and 0.026, respectively (Fig. 4). The "greedy growth scan" finds a most-likely cluster with an average household infection intensity of 397 epg and the circular window spatial scan finds a most-likely cluster with an average household infection intensity of 320 epg. The two found clusters are located in similar areas, but are located farther from the river than the *S. mansoni* clusters.

Discussion

Spatial cluster detection is distinct from spatial modelling (e.g. Raso et al., 2005) in both method and purpose. The former supports surveillance activities by identifying areas where risk is anomalously high, and helps identify otherwise unknown geographically-patterned risk factors. The latter is a powerful tool for characterising geographic variations in risk, and can be used to test explicit hypotheses about putative risk factors. Recent developments that facilitate the efficient detection of irregular cluster shapes have helped to broaden the application of spatial cluster detection methods to micro-spatial scales (Wieland et al., 2007). At these scales, clusters may point to highly localised factors of the physical or social environment that could explain local variations in risk. These methods may also be useful at larger scales, for example, when features such as precipitation, soil type and socio-economic factors may explain geographic variations in risk (Knopp et al., 2008).

In our simulation, we found the "greedy growth scan" to be effective at identifying clusters of infection intensity when they occur in irregular shapes, such as along water sources. This is consistent with other research that has identified non-circular search methods as effective at identifying clusters of incidence and prevalence of disease (Duczmal and Assunçao, 2004; Patil and Taille, 2004; Tango and Takahashi, 2005; Assunção et al., 2006; Yiannakoulias et al., 2007; Duczmal et al., 2008). Our results also show that there is a high degree of over-fitting in the "greedy growth scan"; this method was more likely to include zones not part of a true cluster in the found cluster area. In theory, this can be controlled by imposing stricter penalties on irregularly shaped clusters (Assunção et al., 2006; Kulldorff et al., 2006; Yiannakoulias et al., 2007). However, this requires users to make *a priori* decisions about how much irregularity in shape is tolerable. Having contextual knowledge about the pathogen and environment could help inform the selection of these parameters, but it is important
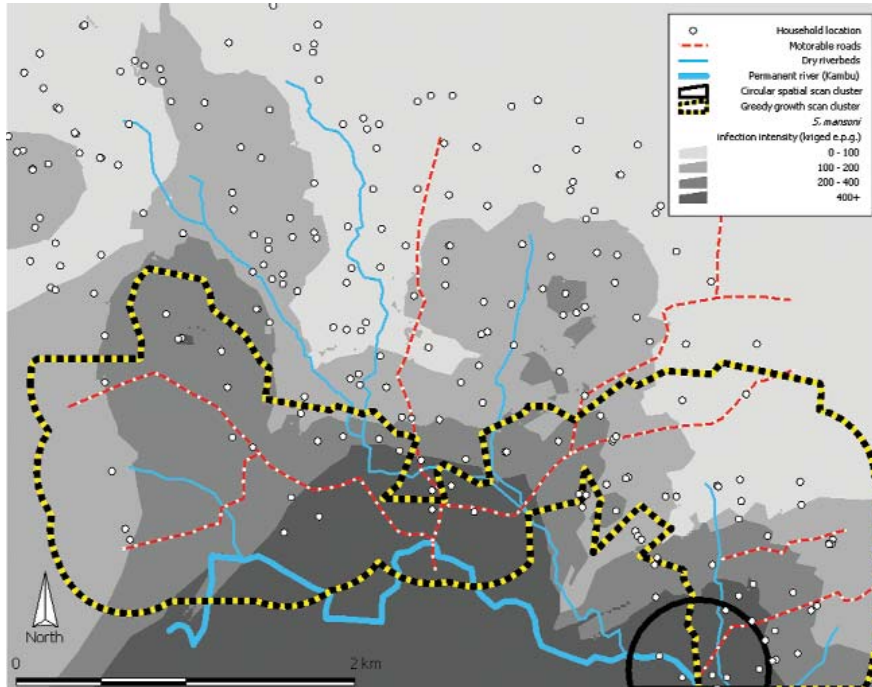
Fig. 3. Clusters of *S. mansoni* infection intensity, Kitengei, Makueni district, Kenya.
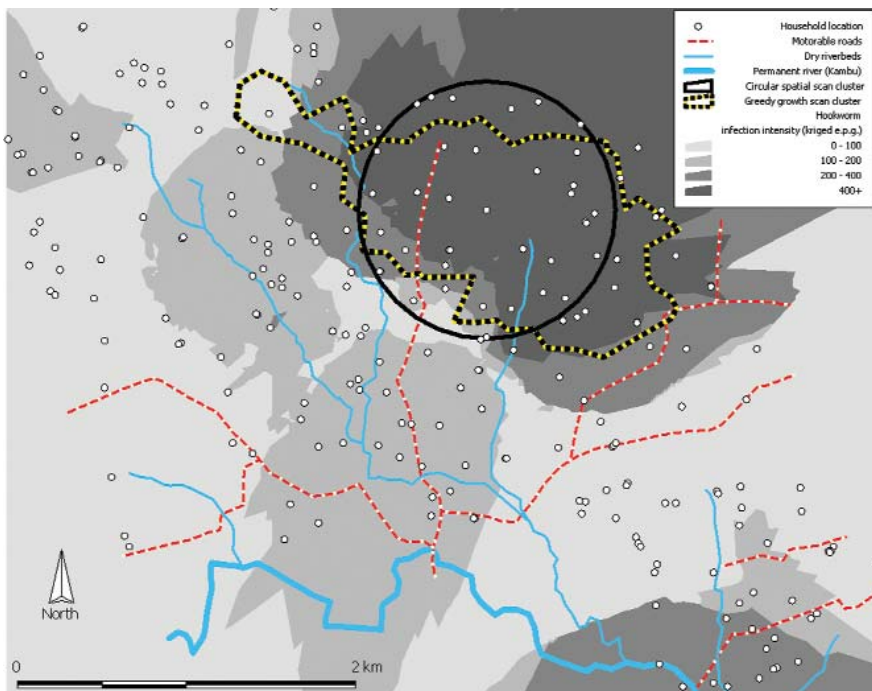

Fig. 4. Clusters of hookworm infection intensity, Kitengei, Makueni district, Kenya.

that this is done prior to looking at the data, since one is at risk of pre-selection bias (or "Texas sharpshooting"). In the simulation, we found our results to be fairly robust to the degree of irregular-shape penalization applied; the location of clusters found by the "greedy growth scan" was similar for all but the highest magnitude of penalisation.

In the analysis of real data, the "greedy growth scan" found a statistically significant cluster of *S. mansoni* infection intensity that tracked, roughly, the length of the main permanent river within the study area. Given the suspected role of water exposure as a risk factor for infection intensity, this is not surprising. We note that this cluster was found in high resolution spatial data, and illustrates the ability of this and similar methods to detect spatial clusters in infection intensity at micro-spatial scales. The real value of this method may be that it can identify clusters along the river that do not include the whole of the river, but sections of the river where exposure is particularly high, or where local factors related to sun exposure, soil type or snail habitat may influence infection intensity. This would be most effective if the analysis controlled for confounding by individual or household attributes that are not distributed in a geographically uniform manner.

At lower resolutions, or when more observations are available, the circular window spatial scan is often able to find clusters of irregular shape by identifying multiple circular clusters that, when seen together, form an irregular pattern. At micro-spatial scales detecting multiple spatial clusters may be more difficult since there is likely to be fewer observations, and in turn, less statistical power of detection. This may point to the methodological value of using the "greedy growth scan" or other irregular-shape spatial scan methods at micro-spatial scales. These methods can characterise irregular clusters as a single irregular form, and can in turn gather the numbers to reach statistical significance. It is important to note, however, that while the circular window spatial scanning method did not find statistically noteworthy real clusters along the river, it did

find a small possible cluster of high infection intensity near the southeast corner of the study area. This included two observations with particularly high infection intensity, and may suggest the possible presence of some localised clustering of higher transmission or vulnerability along this section of the river.

Both methods found statistically significant and similar clusters of hookworm infection intensity. We note that both are further from the river than the clusters of *S. mansoni* infection intensity, and neither follow the path of the river. Since infection by both of these parasites simultaneously is common in endemic regions, it seemed likely to find clusters of the infection intensity in the same location. Polyparasitism is an important concern since it increases the burden of infection on ill health, and may contribute to increased risk of future infection (Brooker and Utzinger, 2007; Pullan and Brooker, 2008). Apparent differences in the patterns here could be due to different spatial distribution of social or environmental risk factors associated with the two types of infections. While not conclusive, our findings are consistent with the hypothesis that close proximity to water plays more of a role in explaining infection intensity of *S. mansoni* than infection intensity of hookworm. Our analysis also illustrates the importance of scale in the analysis of infection intensity and disease in general; the patterns observed here are highly local, which suggests that larger scale patterns may be comprised of a mixture of local micro-geographical patterns easily obscured by course, more aggregate analyses.

We note that the patterns presented here do not account for variations in the distribution of population characteristics, such as the geographic variations in age, occupation or socio-economic status. The well known relationship between these socio-demographic variables and infection intensity could greatly influence underlying geographic pattern of infection intensity if these attributes themselves are geographically heterogeneous. For example, when combined with the general observation that hookworm infection is more common in adults and

S. mansoni infections are more common in children (Hotez et al., 2008), a geographically-patterned distribution of age where older people live further from the river, and younger people closer to the river could explain the clusters observed here. The purpose of this study was to test methodology (rather than provide explanations), so we did not explicitly adjust for age, occupation, sex or other factors that may have explained our observations in the cluster detection exercises.

This analysis shows the utility of irregular shaped cluster detection methods for finding clusters of infection intensity that may occur along water sources or other irregular features of the physical landscape. The data used in this study cover a small area, and present high-resolution information about the spatial distribution of infection intensity. Micro-scale cluster detection analysis may help identify heterogeneities in infection intensity due to features of the local environment, thereby aiding in field investigations of factors that could increase or decrease levels of infection. At high resolutions, spatial scan methods that can locate irregularly-shaped clusters may be particularly important, as it may be at these scales where clusters are least likely to occur in regular compact or circular patterns.

Acknowledgements

References

Ali M, Emch M, Donnay JP, Yunus M, Sack RB, 2002. Identifying environmental risk factors for endemic cholera: a raster GIS approach. Health Place 8, 201-210.

Assunção R, Costa M, Tavares A, Ferreira S, 2006. Fast detection of arbitrarily shaped disease clusters. Stat Med 25, 723-742.

Bethony J, Brooker S, Albonico M, Geiger SM, Loukas A, Diemert D, Hotez PJ, 2006. Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. Lancet 367, 1521-1532.

Brooker S, Alexander N, Geiger S, Moyeed RA, Stander J, Fleming F, Hotez PJ, Correa-Oliveira R, Bethony J, 2006. Contrasting patterns in the small-scale heterogeneity of human helminth infections in urban and rural environments in Brazil. Int J Parasitol 36, 1143-1151.

Brooker S, Clements ACA, 2009. Spatial heterogeneity of parasite co-infection: determinants and geostatistical prediction at regional scales. Int J Parasitol 39, 591-597.

Brooker S, Utzinger J, 2007. Integrated disease mapping in a polyparasitic world. Geospat Health 2, 141-146.

Clements ACA, Moyeed R, Brooker S, 2006. Bayesian geostatistical prediction of the intensity of infection with *Schistosoma mansoni* in East Africa. Parasitology 133, 711-719.

Duczmal L, Assunção R, 2004. A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. Comput Stat Data An 45, 269-286.

Duczmal L, Cançado A, Takahashi R, 2008. Delineation of irregularly shaped disease clusters through multiobjective optimization. J Comput Graph Stat 17, 243-262.

Gemperli A, Vounatsou P, Sogoba N, Smith T, 2006. Malaria mapping using transmission models: application to survey data in Mali. Am J Epidemiol 163, 289-297.

Gryseels B, Polman K, Clerinx J, Kestens L, 2006. Human schistosomiasis. Lancet 368, 1106-1118.

Hotez PJ, Brindly PJ, Bethony JM, King CH, Pearce EJ, Jacobson J, 2008. Helminth infections: the great neglected diseases. J Clin Ivest 118, 1311-1321.

Huang L, Kulldorff M, Gregorio D, 2007. A spatial scan statistic for survival data. Biometrics 63, 109-118.

Kleinschmidt I, Sharp BL, Clarke GPY, Curtis B, Fraser C, 2001. Use of generalized linear mixed models in the spatial analysis of small-area malaria incidence rates in KwaZulu Natal, South Africa. Am J Epidemiol 153, 1213-1221.

Knopp S, Mohammed KA, Khamis IS, Mgeni AF, Stothard JR, Rollinson D, Marti H, Utzinger J, 2008. Spatial distribution of soil-transmitted helmiths, including *Strongyloides stercoralis*, among children in Zanzibar. Geospat Health 3, 47-56.

Kulldorff M, 1997. A spatial scan statistic. Commun Stat A-Theor 26, 1481-1496.

Kulldorff M, Huang L, Pickle L, Duczmal L, 2006. An ellip-

tic spatial scan statistic. Stat Med 25, 3929-3943.

Patil G, Taillie C, 2004. Upper level set scan statistic for detecting arbitrarily shaped hotspots. Environ Ecol Stat 11, 183-197.

Pullan R, Brooker S, 2008. The health impact of polyparasitism in humans: are we under-estimating the burden of parasitic diseases? Parasitology 135, 783-794.

Raso G, Matthys B, N'Goran EK, Tanner M, Vounatsou P, Utzinger J, 2005. Spatial risk prediction and mapping of *Schistosoma mansoni* infections among schoolchildren living in western Côte d'Ivoire. Parasitology 131, 97-108.

Roberts L, Janovy J, 1996. Foundations of Parasitology 5th Edition. Wm. C. Brown Publishers. Dubuque, IA, USA.

Tango T, Takahashi K, 2005. A flexibly shaped spatial scan statistic for detecting clusters. Int J Health Geogr 4, 11.

Wieland S, Brownstein J, Berger B, Mandl K, 2007. Density-equalizing Euclidean minimum spanning trees for the detection of all disease cluster shapes. Proc Natl Acad Sci USA 104, 9404-9409.

Wilson S, Booth M, Jones F, Mwatha J, Kimani G, Kariuki H, Vennervald B, Ouma J, Muchiri E, Dunne D, 2007. Age-adjusted *Plasmodium falciparum* antibody levels in school-aged children are a stable marker of microgeographical variations in exposure to *Plasmodium* infection. BMC Infect Dis 7, 67.

Yiannakoulias N, Rosychuk R, Hodgson M, 2007. Adaptations for finding irregularly shaped clusters. Int J Health Geogr 6, 28.