Geospatial Health 7(1), 2012, pp. 111-126

Analysing the spatial patterns of livestock anthrax in Kazakhstan in relation to environmental factors: a comparison of local (G_i*) and morphology cluster statistics

Ian T. Kracalik[1,2], Jason K. Blackburn[1,2], Larisa Lukhnova[3], Yerlan Pazilov[3], Martin E. Hugh-Jones[4], Alim Aikimbayev[5]

[1]*Spatial Epidemiology and Ecology Research Laboratory, Department of Geography, University of Florida, Gainesville, FL 32611, USA;* [2]*Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611, USA;* [3]*Kazakh Science Center for Quarantine and Zoonotic Diseases, Ministry of Health of the Republic of Kazakhstan, Almaty, Kazakhstan;* [4]*Department of Environmental Science, School of the Coast and Environment, Louisiana State University, Baton Rouge, LA 70803, USA;* [5]*Scientific and Practical Center of Sanitary and Epidemiological Expertise and Monitoring, Ministry of Health of the Republic of Kazakhstan, Almaty, Kazakstan*

Abstract. We compared a local clustering and a cluster morphology statistic using anthrax outbreaks in large (cattle) and small (sheep and goats) domestic ruminants across Kazakhstan. The Getis-Ord (G_i*) statistic and a multidirectional optimal ecotope algorithm (AMOEBA) were compared using 1st, 2nd and 3rd order Rook contiguity matrices. Multivariate statistical tests were used to evaluate the environmental signatures between clusters and non-clusters from the AMOEBA and G_i* tests. A logistic regression was used to define a risk surface for anthrax outbreaks and to compare agreement between clustering methodologies. Tests revealed differences in the spatial distribution of clusters as well as the total number of clusters in large ruminants for AMOEBA (n = 149) and for small ruminants (n = 9). In contrast, G_i* revealed fewer large ruminant clusters (n = 122) and more small ruminant clusters (n = 61). Significant environmental differences were found between groups using the Kruskall-Wallis and Mann-Whitney U tests. Logistic regression was used to model the presence/absence of anthrax outbreaks and define a risk surface for large ruminants to compare with cluster analyses. The model predicted 32.2% of the landscape as high risk. Approximately 75% of AMOEBA clusters corresponded to predicted high risk, compared with ~64% of G_i* clusters. In general, AMOEBA predicted more irregularly shaped clusters of outbreaks in both livestock groups, while G_i* tended to predict larger, circular clusters. Here we provide an evaluation of both tests and a discussion of the use of each to detect environmental conditions associated with anthrax outbreak clusters in domestic livestock. These findings illustrate important differences in spatial statistical methods for defining local clusters and highlight the importance of selecting appropriate levels of data aggregation.

Keywords: spatial cluster analysis, cluster morphology, anthrax, logistic regression, Kazakhstan.

Introduction

In general, tests for clustering in health research are performed to locate hotspots or areas with an unusual increase in the presence of a disease. These tests frequently take the form of the following types of analysis: global, focused, local and, more recently, cluster morphology. Some of the more commonly utilised algorithms in health data analysis include: Moran's *I* (1950), local Moran's *I* (Anselin, 1995), local G_i and G_i* statistic (Getis and Ord, 1992; Ord and Getis, 1995), and the spatial scan statistic (Kulldorff, 1997). In particular, the local G_i* statistic has been well established as a useful tool for defining the extent of local spatial autocorrelation and identifying clusters of disease including malaria (de Castro et al., 2007), schistosomiasis (Clennon et al., 2006), dengue (Jeffery et al., 2009), and typhoid (Hinman et al., 2006). Additionally, the G_i* statistic has also been employed to uncover geographic and anthropogenic features associated with the distribution of spatial data. Kelly-Hope et al. (2009) used G_i* to compare environmental factors associated with variations in the abundance of malaria vectors. This study found there were correlations between environmental factors and clusters of *Anopheles* spp.

More recently, a class of techniques concerned with the shape or morphology of spatial clusters was introduced including: simulated Annealing (Duczmal and Assuncao, 2004), flexible scan (Tango and Takahashi, 2005), a multidirectional optimal ecotope algorithm (AMOEBA) (Aldstadt and Getis,

Corresponding author:
Jason K. Blackburn
Spatial Epidemiology and Ecology Research Laboratory
3141 Turlington Hall, University of Florida, Gainesville, Florida, 32611, USA
Tel. +1 352 273 9374; Fax +1 352 392 8855
E-mail: jkblackburn@ufl.edu

2006), Greedy Growth Scan (Yiannakoulias et al., 2007), and Cluster Morphology Analysis (CMA) (Jacquez, 2009). These methods are designed to identify contiguous, irregularly shaped clusters of events on the landscape that may often be overestimated in size, and/or shape by other methodologies such as circular scan statistics (Duczmal and Assuncao, 2004; Tango and Takahashi, 2005). Locating irregularly shaped clusters may be advantageous since they can possibly reveal patterns that are associated with geographical features such as elevation or water bodies (Yiannakoulias et al., 2010). AMOEBA derived from local G_i^* was shown to perform well in tests identifying clusters of high spatial values when compared to the circular Kulldorff spatial scan statistic (Aldstadt and Getis, 2006). Additionally, Doi et al. (2008) applied FleXScan to investigate Creutzfeldt-Jacob mortality in Japan and found that irregularly shaped hotspots appeared to follow the Fuji River basin; identifying a potential link between spatial variations in health outcomes and the environment. While local and morphology cluster statistics (i.e. G_i^* and AMOEBA) have been shown to be useful, there is little guidance regarding potential advantages and/or disadvantages in selecting a particular method.

Here we explore the differences in the ability of local and morphology methods to identify hotspots of disease, while also examining relationships between environmental characteristics and clusters. A number of statistical methods have been used to model environmental relationships in lieu of spatial clustering algorithms including discriminant function analysis (Rogers et al., 2002), and logistic regression (Eisen et al., 2007). However, researchers have shown that incorporating cluster analyses with statistical models such as logistic regression may better assess the level of risk (Ruiz et al., 2004). We used a long historical record of anthrax outbreaks in livestock from Kazakhstan (Aikembayev et al., 2010; Kracalik et al., 2011) to uncover possible spatial clustering while also trying to elucidate environmental factors associated with areas of high anthrax outbreaks.

Anthrax is a soil-borne zoonosis caused by the Gram-positive, spore-forming bacterium *Bacillus anthracis*, which primarily infects livestock and wildlife and secondarily afflicts humans (Hugh-Jones and Blackburn, 2009). It remains a problem in many developing countries including former states of the Soviet Union, which have undergone dramatic changes in their public and veterinary health infrastructure (Hugh-Jones, 1999). In particular, Kazakhstan has been burdened by outbreaks of the disease, both in humans (Woods et al., 2004) and in livestock (Aikembayev et al., 2010). Previous studies examining anthrax ecology have used ecological niche models to elucidate environmental characteristics that may either constrain or promote the persistence of *B. anthracis* spores across the landscape in the United States of America (Blackburn et al., 2007) and Kazakhstan (Joyner et al., 2010). While these studies have provided crucial information on the potential geographical range of *B. anthracis*, they have not taken into account environmental factors that may contribute to unusually high occurrences or clusters of the disease.

The available literature suggests that the anthrax bacterium has an ecological affinity for a range of specific soil conditions preferring alkaline, calcium rich soils with a high percentage organic material (Van Ness, 1971). Outbreaks of the disease have been linked to cyclical climatic patterns, specifically wet springs followed by hot, dry summers followed by a precipitation event (Gates et al., 1995; Turner et al., 1999; Parkinson et al., 2003). Subsequent studies have corroborated this affinity for specific environments (Dragon and Rennie, 1995; Gates et al., 1995; Hugh-Jones and Blackburn, 2009) though few have provided statistical correlations to link spatial patterns of disease occurrence to environmental factors. Smith et al. (1999, 2000) incorporated cluster analysis to show that spatial variations of anthrax infections in wildlife were possibly associated with genetic diversity and soil conditions.

This current study had three specific objectives: (i) identify local clusters of anthrax in large and small ruminants in Kazakhstan, while elucidating any potential differences in the local (G_i^*) and morphology (AMOEBA) cluster methods; (ii) explore the use of the AMOEBA methodology and the effects of using multiple settings in running the algorithm; and (iii) investigate differences in environmental variables in relation to statistically significant high cluster locations and non-cluster areas.

Methods

Geographical information systems (GIS) database development

A spatial database consisting of 1,206 anthrax outbreaks in large ruminants (cattle) and 1,318 outbreaks in small ruminants (sheep and goats) from the time period 1960 to 2006 was digitised from historical cadastres to describe and analyse the spatial distribu-

tion of livestock anthrax in Kazakhstan. These data were obtained from a larger database development effort in collaboration with the Kazakh Science Center for Quarantine and Zoonotic Disease in Almaty, Kazakhstan, described in detail elsewhere (Aikembayev et al., 2010; Joyner et al., 2010; Kracalik et al., 2011). The data were divided into livestock groups based on results of Aikimbayev et al. (2010), which suggested that cases within each group where distributed differently. Broadly, this subset of data reflects the time period after mass vaccination was implemented and corresponds to the averaged climate data from the WorldClim data set use in the environmental analysis (see below).

For the purposes of this study an outbreak was defined as a location that reported one or more positive confirmations of an anthrax infection in either large or small large ruminants. Confirmation was either through laboratory diagnostics (e.g. blood film evaluation, bacterial culture isolation and, more recently, molecular confirmation by polymerase chain reaction [PCR]) or classical clinical signs in animal carcasses at the outbreak site (Aikembayev et al., 2010). The locations and dates of livestock outbreaks were provided as latitude/longitude coordinate pairs georeferenced to the nearest community (Fig. 1). All GIS data were managed in ArcGIS 9.3.1 (ESRI, Redlands, CA, USA).

Both the G_i* statistic and AMOEBA are group-level statistics that require data to be aggregated. A 25 km hexagonal grid surface (measured from the centroid to the center of a straight line) was generated for the entire country of Kazakhstan using the Jennes repeating shapes tool (http://www.jennessent.com/arcgis/repeat_shapes.htm). Outbreak locations were then aggregated to the grid surface using the Hawth's Tool point-in-polygon count (http://www.spatialecology.com/htools/pntpolycnt.php). This resulted in 1,754 hexagons representing Kazakhstan (Fig. 1). A spatial grid surface was chosen in order to analyse the distribution of anthrax at a smaller scale than the rayon level (district equivalent political boundary; administrative level 2). A number of different sized grid cells were tested, but due to the computational time required to run the AMOEBA statistic (Duque et al., 2010) the 25 km grid cell was used in order to complete analyses in a manageable period of time. Specifically, runs on small grid cells could take days to weeks to run on a high-end desktop computer. Because Kazakhstan sits at relatively high latitude and has a large landmass, we used the a conic map projection, often used for aeronautical charts, i.e. the Asia Lambert Conformal Conic projection in ArcGIS (http://spatialreference.org/ref/esri/102027/) to preserve both size and shape of the grid cells across the entire country.

*Getis-Ord G_i**

This statistic (Getis and Ord, 1992; Ord and Getis, 1995) was used to identify local clusters in both livestock groups. The G_i* statistic is a useful tool for determining the spatial dependence between an observation and neighbouring observations within a user specified distance threshold (Getis et al., 2003). The G_i* statistic is written as:

$$G_i(d) = \frac{\Sigma_j \; w_{ij}(d) \times x_j - w_i^* \times \bar{x}}{S \times \{[(nS_{1i}^*) - w_i^{*2}]\,(n-1)\}^{/2}} \qquad \text{(Eq. 1)}$$

where $\bar{x}$ is the mean of all anthrax outbreaks located within the 25 km hexagonal grid cells and S the standard deviation. This statistic is distance-based and

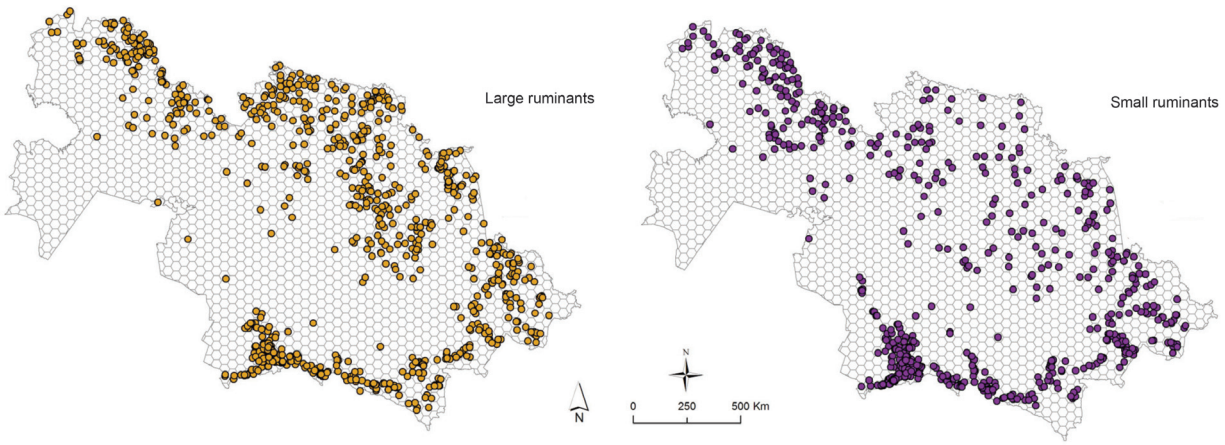

Fig. 1. The spatial distribution of anthrax outbreaks in Kazakhstan during the time 1960 to 2006 for large ruminants (in orange) n = 1,206 and small ruminants (in purple) n= 1,318.

three separate spatial weights matrices w_{ij} were utilised based on 1st, 2nd and 3rd order Rook contiguity relationships (O'Sullivan and Unwin, 2003) between all grid cells *i* that share a border each contiguity order are 1, including *i*, and all others are 0. G_i^* values are given as a standard normal variant with an associated probability from the z-score distribution (Getis et al., 2003). All contiguity matrices were constructed in GeoDa 0.9.1 (http://geodacenter.asu.edu/).

As G_i^* values approximate normal variants of the z-distribution as the number of neighbours reaches eight (Ord and Getis 1995), only those values greater than 3.18 representing a $P \leq 0.001$ were used to define high clusters. Following Getis et al. (2003) values of G_i^* must increase with increasing distance in order for a grid cell or rayon to be deemed a significant cluster at a particular distance. For example, a grid cell may have a G_i^* value exceeding the significance threshold of 3.18 at a 2nd order weights designation, but if the G_i^* value at 1st order is greater, then the grid cell is considered a significant cluster at the 1st order. This is defined as the critical distance, d_c, in Getis and Aldstadt (2004).

Multidirectional optimal ecotope-based algorithm (AMOEBA)

In order to identify the morphology or the geometric nature of anthrax clusters on the landscape an AMOEBA algortitm (Aldstadt and Getis, 2006) was used. The AMOEBA clustering algorithm is derived from the calculation of G_i^* values (Ord and Getis, 1995) aforementioned and is intended to search for irregularly shaped continuous clusters of high and low values (Aldstadt and Getis, 2006; Weeks et al., 2010). For each hexagonal grid cell *i* denoted G_i^* (0) and all contiguous neighbours *j*, the statistic is calculated following (Aldstadt and Getis, 2006) where *N* is the number of hexagonal spatial units, x_j the number of anthrax outbreaks at location *j*, and x the mean of all outbreaks. AMOEBA searches in a multidirectional manner from G_i^* (0), if one of the *j* neighbouring hexagonal units in the total set of *n* with in Kazakhstan maximises the G_i^* value of the statistic either negatively (low cluster) or positively (high cluster) then unit *j* becomes included in the continuous cluster (Aldstadt and Getis, 2006). If additional hexagonal units fail to increase the value of the statistic, they are removed from the continuous cluster; otherwise, they remain. In this study a statistical significance alpha $\alpha = 0.001$ was used derive hotspots of anthrax.

Additional settings within AMOEBA include the selection of a Bonferroni, false discovery rate (FDR), or a core cut-off threshold that limits the maximum cluster size. The selection criteria for a core cut-off in AMOEBA have not been described in detail so in order to evaluate the effects of a threshold we chose core cut-offs of: 2, 3, 6 and 12.

Environmental comparison

To identify potential correlations between the presence of clusters and specific landscape characteristics an environmental dataset was constructed. A total of 22 raster variables, at a spatial resolution of approximately 1 x 1 km, were used in the analysis, including six WorldClim variables (www.worlclim.org), 15 variables from the harmonised soils world database (HWSD) version 1.1 (http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/), and one variable from the Trypanosomiasis and Land Use in Africa (TALA) research group (Hay et al., 2006) (Table 1). Variables from HWSD were comprised of various measures of soil composition and physical properties available at the 1 x 1 km resolution, while measurements of soil characteristics in HWSD were tabulated by soil sampling units, which are comprised of a larger group of grid cells. These larger soil sampling units may include a range of measurements for a single soil variable, thereby resulting in a range of values for a given pixels rather than assigning a single unique value. In order to calculate specific soil raster variables the "summarise" feature in ArcGIS 9.3.1 was used to generate statistics on specific soil measurements. For example, in order to calculate measurements of pH, the total range of values for a set of pixels were used to create an average minimum and average maximum pH that would then be assigned to a single pixel in that sampling unit. This process was repeated for all soil variables used in Table 1. The zonal statistics routine in ArcGIS 9.3.1 was then applied to extract environmental information to each grid surface across Kazakhstan.

Statistical analyses

Environmental signatures of clusters

All statistical tests were performed in SAS v 9.2. The Kruskal-Wallis test (Sokal and Rohlf, 1995)was performed using the Proc Npar1way procedure to compare potential differences in environmental variables between multiple groups: AMOEBA, non-

Table 1. Summary statistics for the database of environmental variables.

Variables	Mean	Standard deviation	Percentiles			Correlated variables
			25th	50th (median)	75th	ρ s < 0.8
Average altitude*	376.18	480.00	116.27	242.40	450.53	
Average Bio1* (annual temperature range)	57.96	37.37	27.02	57.52	84.76	
Sum Bio12* (total annual precipitation)	4.87 x 10^5	2.84 x 10^5	3.13 x 10^5	4.72 x 10^5	6.32 x 10^5	
Average NDVI^	0.15	0.14	0.11	0.16	0.22	
Standard deviation pH¥	0.30	0.70	0.02	0.10	0.24	
Minimum pH¥	6.73	1.03	6.40	7.02	7.31	
Maximum pH¥	7.72	0.90	7.44	7.96	8.13	
Standard deviation $CaCO_3$¥ (calcium carbobnate)	0.88	1.14	0.07	0.46	1.18	X
Minimum $CaCO_3$¥	1.42	1.91	0.26	1.00	1.40	X
Maximum $CaCO_3$¥	9.09	6.41	4.02	5.87	15.86	X
Standard deviation OC¥ (organic carbon)	0.12	0.15	0.01	0.06	0.19	X
Minimum OC¥	0.49	0.31	0.29	0.40	0.62	
Maximum OC¥	1.51	0.90	0.70	1.61	1.97	
Standard deviation $CaSO_4$¥ (gypsum)	0.20	0.33	0.00	0.02	0.31	X
Minimum $CaSO_4$¥	0.17	0.58	0.03	0.10	0.10	X
Maximum $CaSO_4$¥	1.51	1.88	0.10	0.27	2.40	
Standard deviation ECE¥ (salinity)	0.54	0.80	0.02	0.15	0.83	
Minimum ECE¥	0.41	1.28	0.08	0.20	0.30	X
Maximum ECE	4.04	4.30	1.20	1.88	5.91	X
Average Bio7* (annual temperature range)	266.50	216.63	31.38	413.24	474.55	
Average Bio13* (precipitation of wettest month)	35.62	17.41	21.01	31.12	47.13	
Average Bio14* (precipitation of driest month)	10.44	5.01	6.71	11.37	13.16	

* WorldClim variables; ¥ HWSD variables; ^ TALA variable. Variables exhibiting significant correlations with other covariates were removed from logistic and discriminant function models.

AMOEBA, G_i* and non-G_i*. An $\alpha = 0.05$ was used with a correction for multiple comparisons (k-1)/2, where k equals the number of groups being tested (in this case $k = 4$) resulting in an $\alpha = 0.0167$. For the purposes of this study we were only interested in comparing statistically significant hotspots to every other spatial unit not identified as high cluster values, which were deemed non-hotspots. Since the Kruskal-Wallis test does not specify which groups were statistically different a Mann-Whitney U test (Sokal and Rohlf, 1995) was employed to perform pair-wise comparisons on the following individual groups: AMOEBA and non-AMOEBA, AMOEBA and G_i* and G_i* and non-G_i*, using an $\alpha = 0.0167$.

Due to the small number of hotspots clusters obtained from the AMOEBA and G_i* tests when examining anthrax outbreaks in small ruminants, all statistical analyses comparing groups and their corresponding environmental signatures were limited to large ruminants outbreaks. In this manner we avoided any issues that we may have had with extremely low sample sizes ($n \leq 9$).

Logistic regression analysis

Multivariate logistic regression models were built to evaluate the association between the probability that an area was able to support anthrax in large ruminants and its environmental characteristics. Logistic models took the form of the following equation:

$$Logit(P) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \ldots + \beta_k X_k \quad \text{(Eq. 2)}$$

where P is the probability of a hexagonal cell being classified as suitable for anthrax in large ruminants, β_0 the intercept and β_1 the coefficient assigned to independent variable X_1. For this modelling approach, a grid cell was defined as present if one or more large ruminant anthrax outbreaks occurred in that cell during the period 1960 to 2006 (n = 427); all other cells were defined as absent (n = 1,327). Subsets of training and testing data were created by randomly splitting each dataset. A subset of ~20% was withheld from the model-building process for post-hoc model evaluation (testing data) and the remaining subset (~80%) was used for model-building (training data) (Table 3). A total of 23 environmental variables were initially used in the model, including the latitude and longitude centroid of each hexagonal grid cell (Table 1). Correlations between variables were tested with a Spearman's rank correlation test and significantly correlated variables were removed using a cutoff of ≥0.85.

A stepwise modelling approach was adopted to reduce the number of potential environmental factors and select a single best model. Predictive accuracy metrics were performed using the testing data (~20% withheld from training). Sensitivity and specificity of the model were calculated (Fielding and Bell, 1997), as well as the area under the curve (AUC) score produced using a receiver operating characteristic (ROC) (Zweig and Campbell, 1993). We used a probability cutoff of P ≥0.46 from a classification table that maximised sensitivity and specificity.

Predicted probabilities from the logistic regression for each hexagonal cell were then mapped using the "Spatial Analyst Extension" in ArcGIS v9.3.1 using the following equation:

$$P = \frac{\exp(\beta_0 + \beta_1 X_1 \ldots + \beta_1 X_1)}{1 + [\exp(\beta_0 + \beta_1 X_1 \ldots + \beta_1 X_1)]} \quad \text{(Eq. 3)}$$

A risk surface was created using the same cut-off threshold with *high risk* defined as any value ≥0.46 with any lower values defined as *low risk*. This probability threshold was derived from a classification table by selecting a value that maximised the sensitivity and specificity. The probability surface was reclassified to high risk and low risk and mapped against AMOEBA and G_i* clusters for large ruminants. The total percentage of Kazakhstan defined as high risk was calculated in Spatial Analyst Extension.

Results

*AMOEBA and G_i**

The total number of statistically significant high clusters of anthrax across Kazakhstan varied among small and large ruminants as well as between the AMOEBA and G_i* statistics (Table 2). The number of clusters for large ruminants using a 1st, 2nd and 3rd, order Rook contiguity matrix was greater for AMOEBA (n = 119, 108 and 109, respectively) compared to G_i* (n = 53, 81 and 102, respectively). The opposite was true for small ruminants, with the number of clusters at the 1st, 2nd and 3rd contiguity relationships being greater for G_i* (n = 24, 77 and 55, respectively) compared to AMOEBA (n = 6, 5 and 6, respectively). In addition to differences in the number of clusters detected by each method there were pronounced spa-

Table 2. Results from the spatial statistical analyses of anthrax outbreaks in large and small ruminants using the local G_i* statistic and the AMOEBA cluster morphology.

AMOEBA	1st order[1]		2nd order[2]		3rd order[3]		Core 2		Core 3		Core 6		Core 12	
	H	O	H	O	H	O	H	O	H	O	H	O	H	O
Large ruminants outbreaks	119	150	108	38	109	37	137	132	119	150	25	244	6	263
Small ruminants Outbreaks	6	12	5	6	6	5	6	12	6	12	4	14	2	16

Local *Gi**	1st order[1]	2nd order[2]	3rd order[3]
	High clusteers[A]	High clusteers	High clusteers
Large ruminants outbreaks	53	81	102
Small ruminants outbreaks	24	77	55

[1] Contiguity analyses were performed using a core cutoff equal to 3; [2] Outbreak numbers were adjusted using the 2007 large ruminants density estimates obtained from GLIPHA; [3] The number of high clusters determined by the *Amoeba* analysis; [4] The number of spatial units deemed outside of a cluster by the *Amoeba* analysis; [A] The total number of significant *Gi** clusters for adjusted and raw case numbers were derived using a cluster assignment procedure whereby inclusion in a cluster group, e.g. 1st order or 2nd order was based on the concept of increasing *Gi** with increasing distance; using an $\alpha < 0.001$.

tial differences. The spatial distributions of clusters in large ruminants using AMOEBA (n = 149) appeared more dispersed and greater in total numbers in the north compared to G_i* clusters (n = 122), which were predominantly concentrated in southern Kazakhstan (Figs. 2 and 4). Spatial clusters of small ruminants were only found in southern Kazakhstan with the G_i* statistic identifying more clusters (n = 61) compared to AMOEBA (n = 9) (Figs. 3 and 5).

The selection of critical distances via the increasing order of the contiguity matrices revealed that, for the most part, greater magnitudes of distance or order show a trend towards an increase in the number of significant clusters. That is, as the distance increases the number of spatial clusters assigned to the higher order also increased as shown in Figs. 2, 3 and 4. The exception in this case was that of small ruminant clusters for AMOEBA, which had a relatively low number of clusters (n = 9) and did not display this trend.

Core cut-off values for AMOEBA showed an inverse relationship between the number significant high clusters for both large and small ruminants (Figs. 6 and 7). The results indicated that selection of a core cut-off value correspond to the number of spatial units designated as outside of a cluster. The higher the core cutoff value selected in AMOEBA the greater the number of spatial units defined as outside clusters, indicating that particular grid cells were neither a high or low cluster (Table 2). There were no core cut-off criteria for the G_i* statistic so no direct comparisons were made. Spatial agreement between the methods revealed that for large ruminants 29.5% of AMOEBA clusters and 36.1% of clusters defined by G_i* overlapped spatially, while agreement between methods for small ruminants revealed 77.8% of AMOEBA clusters and 11.5% of G_i* clusters overlapped spatially (Fig. 8).

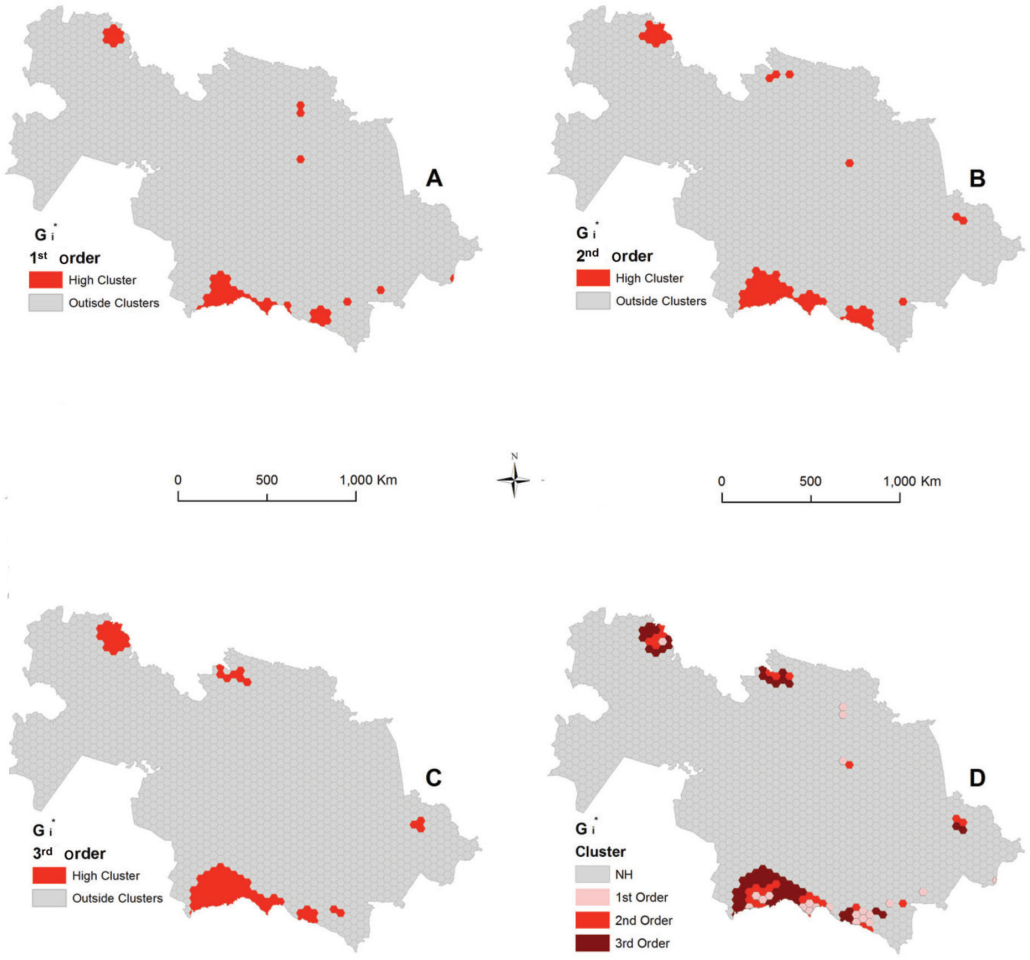

Fig. 2. Local clustering of anthrax outbreaks in large ruminants across Kazakhstan during the time period 1960-2006 using the G_i* statistic with three different Rook contiguity thresholds used. Map portrays the differences in the spatial extent of clusters using a 1st, 2nd and 3rd order contiguity matrix. Critical distance thresholds are displayed in G_i* cluster map with 1st order d_c shown in pink, 2nd order d_c shown in red and 3rd order d_c shown in burgundy.

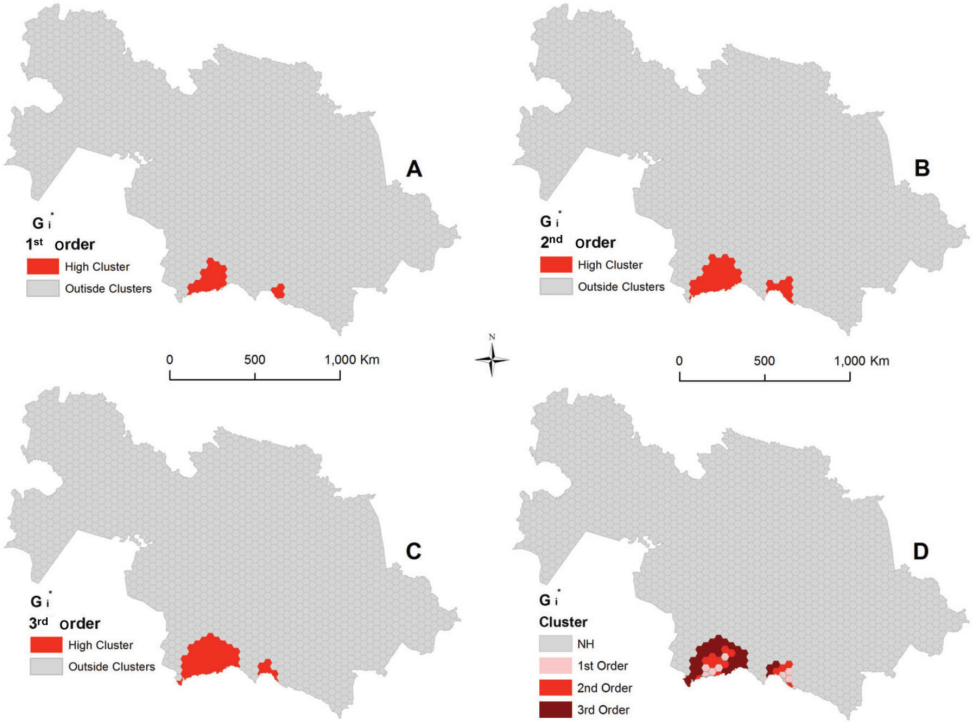

Fig. 3. Local clustering of anthrax outbreaks in small ruminants across Kazakhstan during the time period 1960-2006 using the G_i* statistic with three different Rook contiguity thresholds used. Map portrays the differences in the spatial extent of clusters using a 1st , 2nd and 3rd order contiguity matrix. Critical distance thresholds are displayed in G_i* cluster map with 1st order d_c shown in pink, 2nd order d_c shown in red and 3rd order d_c shown in burgundy.

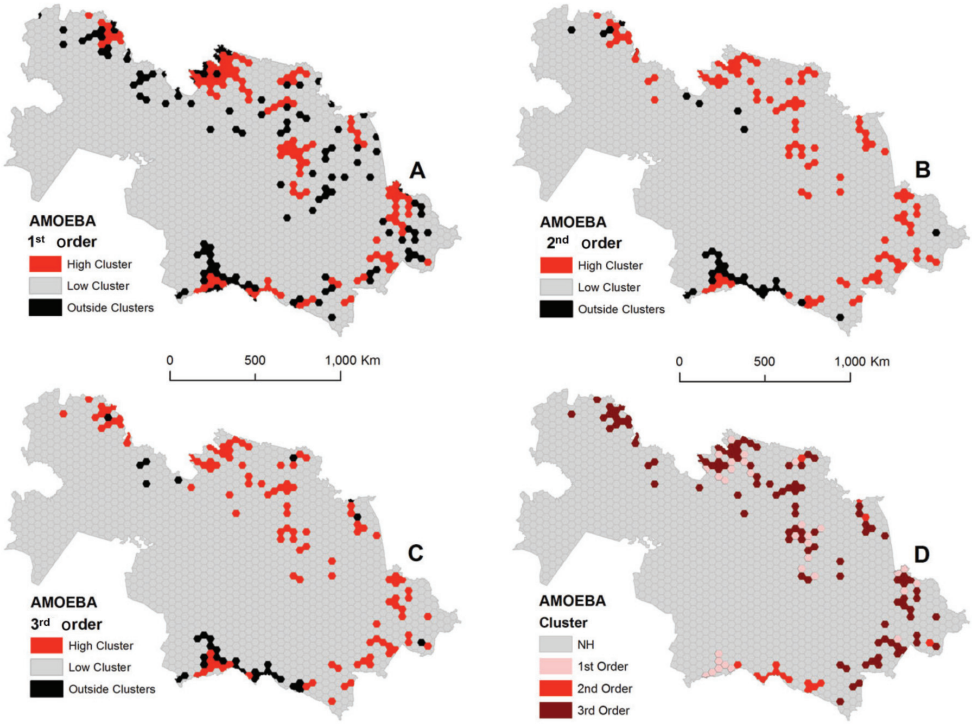

Fig. 4. AMOEBA cluster morphology of anthrax outbreaks for large ruminants in Kazakhstan during the period 1960-2006. Map portrays 25 km grid cells that were clusters of high values shown in red and outside of a cluster shown in black. Rook contiguity matrices were used at 1st order, 2nd order and 3rd order. The AMOEBA cluster map illustrates the use of critical distance d_c incorporating the Rook contiguity matrix to determine significance with 1st order d_c shown in pink, 2nd order d_c shown in red and 3rd order d_c shown in burgundy.

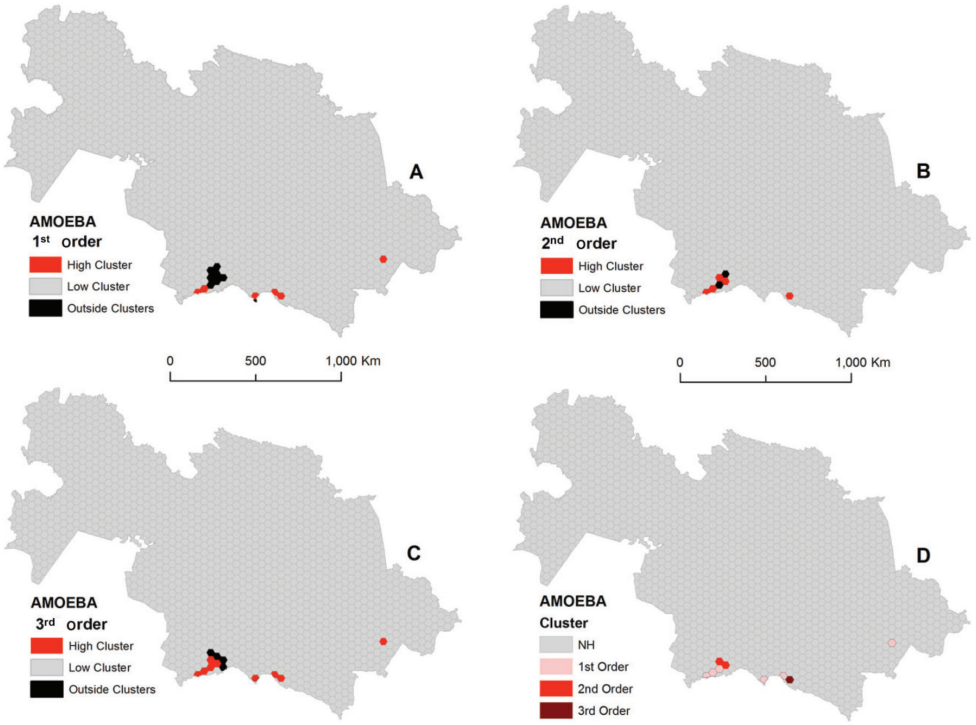

Fig. 5. AMOEBA cluster morphology of anthrax outbreaks for small ruminants in Kazakhstan during the period 1960-2006. Map portrays 25 km grid cells that were clusters of high values shown in red and outside of a cluster shown in black. Rook contiguity matrices were used at 1st order 2nd order and 3rd order. The AMOEBA cluster map illustrates the use of critical distance d_c incorporating the Rook contiguity matrix to determine significance with 1st order d_c shown in pink, 2nd order d_c shown in red and 3rd order d_c shown in burgundy.

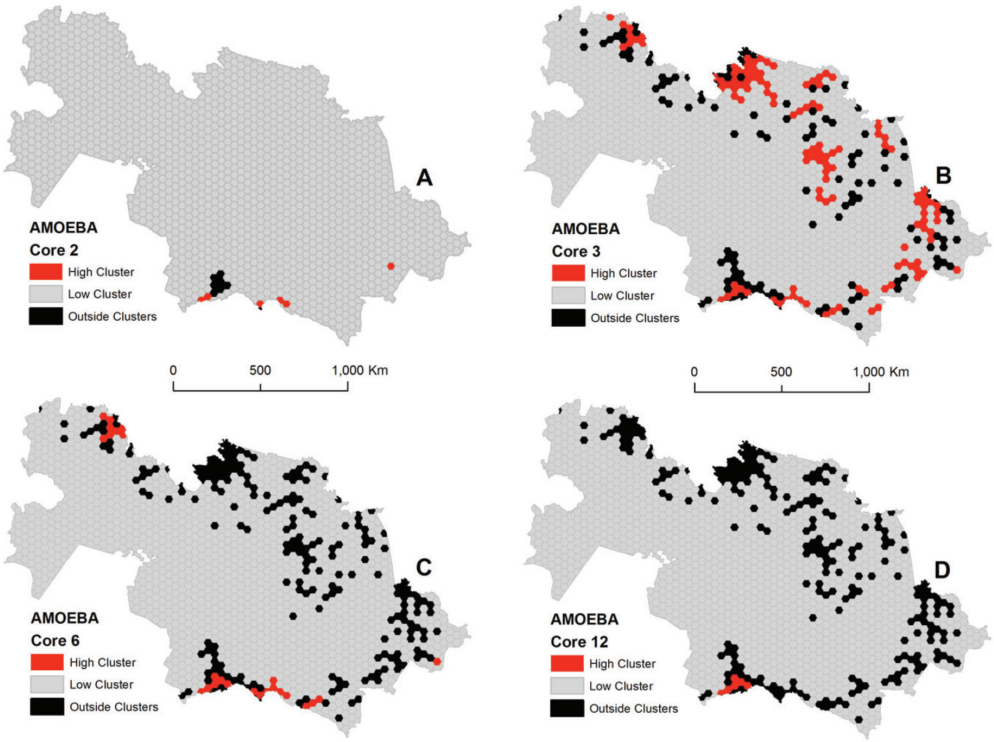

Fig. 6. AMOEBA core cut-off comparison for large ruminants showing core thresholds of 2, 3, 6 and 12. Hotspot cluster values are shown in red while areas deemed outside of a cluster are shown in black.

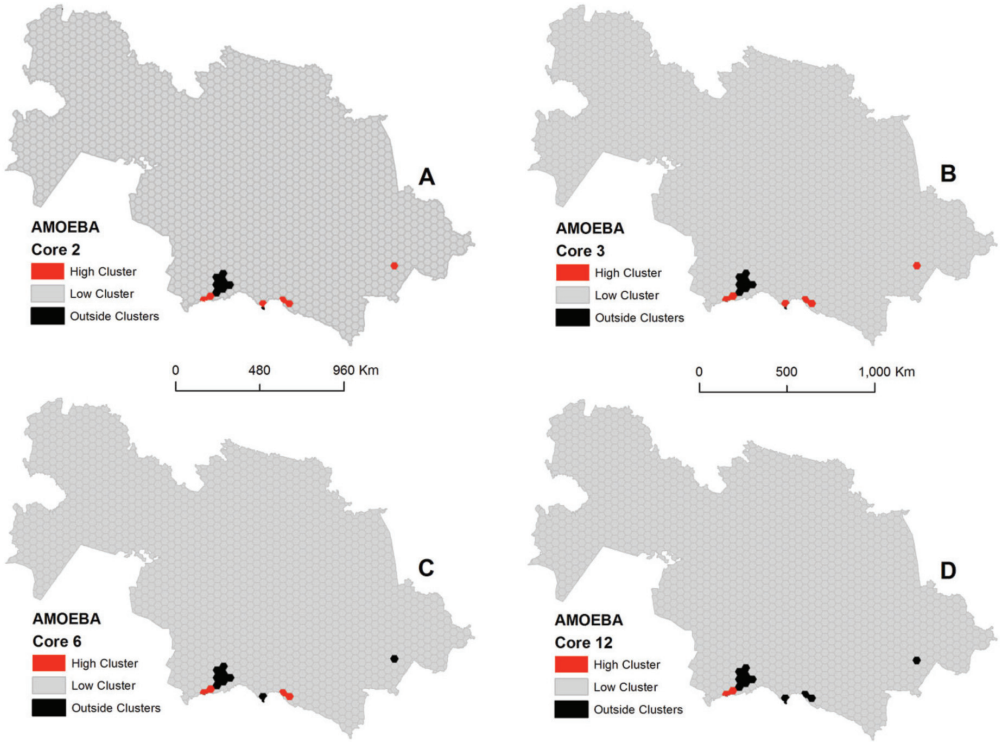

Fig. 7. AMOEBA core cut-off comparison for small ruminants showing core thresholds of 2, 3, 6, and 12. Hotspot cluster values are shown in red while areas deemed outside of a cluster are shown in black.

Environmental comparison

The Kruskal-Wallis test analysing clusters and non-clusters from each methodology indicated a significant difference between groups P <0.001. Fig. 9 shows the outcomes for the Mann-Whitney U test analysing pair wise comparisons of environmental variables and the following three groups: AMOEBA and non-AMOEBA, G_i* and non-G_i* as well as AMOEBA and G_i*, also showed a statistical difference P <0.0167 or -2.13 > z > 2.13 between the rank sums of environmental variables in Table 1. In the AMOEBA/G_i* comparison group, statistical differences were found between the following variables: average altitude, average Bio1 (annual mean temperature), sum Bio12 (annual precipitation), maximum pH, min OC, maximum $CaSO_4$, average Bio7 (annual temperature range), average Bio14 (precipitation of the driest

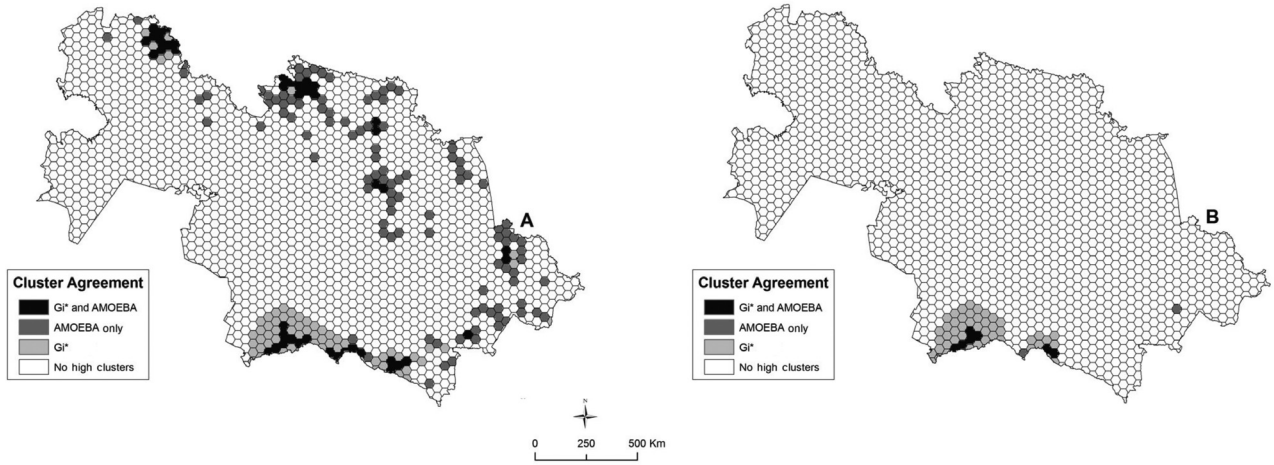

Fig. 8. Spatial agreement among hotspot clusters in large ruminants (A) and small ruminants (B) using the local G_i* statistic and the AMOEBA cluster morphology statistic. Spatial clusters that overlap between methodologies are displayed in black, clusters that were spatially unique for are displayed in light gray and spatially unique clusters for AMOEBA are shown in dark gray.

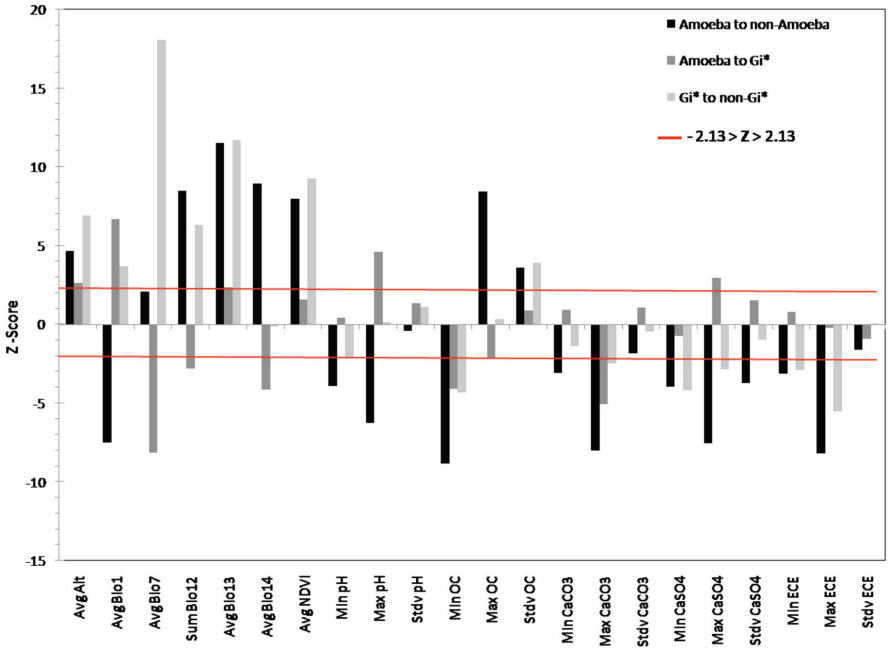

Fig. 9. Pair wise group comparisons of environmental variables using the Mann-Whitney U Test. Y-axis represents the z-score associated with the individual test statistic and the x-axis represents the environmental variables used in the analysis. Red lines show the critical significance level using a P = 0.0167 on a two-tail test representing a z-score threshold greater than 2.13 or less than -2.13.

month), and maximum $CaCO_3$. Differences in environmental factors between all paired groups analysed using the Mann-Whitney U tests were also found in the following variables: average altitude, average Bio1, average Bio13, minimum OC, maximum $CaCO_3$ and maximum $CaSO_4$. Notable differences in the ecological characteristics between AMOEBA and G_i* comparison group were maximum pH, the sum Bio12 and minimum organic carbon.

Logistic regression

The probability of suitable habitats for anthrax in Kazakhstan was defined by our best model (Table 4, Fig. 10). The model prediction showed that the data were a good fit and had an appropriate number of variables based on the Hosmer and Lemeshow goodness-of-fit-test (P = 0.346). The results indicated that bioclimatic variables as well as the normalized difference vegetation index (NDVI), altitude and latitude are all positively associated with the presence of anthrax (Table 4). The accuracy of the best model, based on the AUC obtained from the ROC plot, was 0.83 indicating good discrimination between cells with anthrax and cells without. The sensitivity of the model was 77.7% and the specificity was 73.8%.

The classified risk maps show that ~75% of AMOEBA clusters were within the high-risk zones *versus* ~64% of G_i* clusters (Fig. 10). The total percentage of the Kazakh landscape defined as having a high risk of large ruminant anthrax outbreaks was 32.2%. Thus, overall, large areas within the northern latitudes of Kazakhstan can be predicted to be highly suitability for anthrax and should therefore be classified as high risk for large ruminants.

Table 3. Metrics used to build the stepwise multivariate logistic regression model using an 80% training and 20% testing data split.

Metric	Logistic
n presence to build model	320
n absence to build model	1,327
n presence to test model	107
n absence to test model	332
Sensitivity	77.7%
Specificity	73.8%
AUC	0.83

Discussion

Cluster detection methods have been well established in epidemiological research as important tools for identifying potential variation in the distribution of a disease (Anderson and Titterington, 1997). We compared two different methodologies: a local clustering test and a cluster morphology test to delineate and compare the spatial distribution of anthrax outbreaks in Kazakhstan during the period 1960 to 2006.

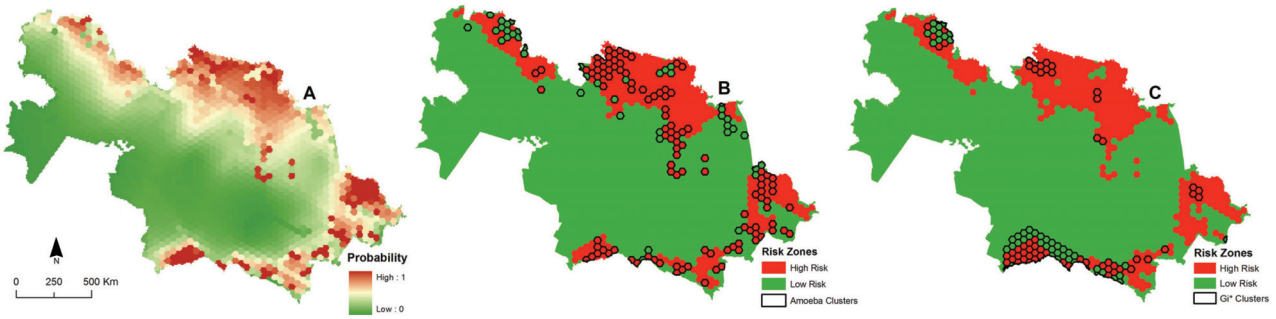

Fig. 10. Map (A) displaying the spatial prediction from the multivariate logistic regression with a higher probability of an anthrax outbreak shown in red and lower probability shown in green. Spatial risk map (B) displaying areas of high risk in red and areas of low risk in green compared to AMOEBA clusters, using a probability cut-off threshold of 0.46 and above to represent presence or high risk and everything below that threshold representing low risk.

In addition we examined the potential association with clusters and environmental factors. The local G_i* and the AMOEBA morphology statistic used in this study indicated that there were distinct differences in the spatial pattern of clusters identified between methodologies. The findings presented here illustrate the importance in choosing a spatial statistical methodology while also selecting an appropriate level of data aggregation.

Despite the fact that the AMOEBA statistic is derived from G_i* there were marked differences in their cluster patterns and each methodology delineated spatially unique hotspots (Fig. 8). The AMOEBA statistic, as expected, located irregularly shaped patterns while G_i* was restricted to more undefined, circular clusters. Additionally, the G_i* statistic appeared to overestimate the boundary of clusters in small and large ruminants in the south compared to that of AMOEBA, while simultaneously also not detecting outbreak clusters from large ruminants in the north. Observed variations in the cluster patterns of G_i* and AMOEBA could in part be a product of the low statistical power of the G_i* statistic (Jacquez, 2009). The low level of power would potentially result in a higher number of false negatives. However, there are no similar tests on AMOEBA that would allow for direct comparison of each methodology's statistical power. It is noteworthy though that another cluster morphology technique, Flex Scan, appears to have a relatively high power and accuracy when compared to G_i* (Jacquez, 2009).

Spatial differences in the location of hotspots between large and small ruminants highlight the distributional and aggregation effects the data had on both AMOEBA and G_i*. Despite the fact that there were more anthrax outbreaks in small ruminants, there were fewer clusters detected compared to that of large ruminants. This outcome might have been strongly influenced by separate geographical ranges for each livestock group. AMOEBA detected a greater number of large ruminant clusters while G_i* detected a greater number of small ruminant clusters, suggesting that the distribution of the outbreak data played a role in the ability of each method to identify hotspots. Patterns of clusters in this study were also most likely dependent on the level of aggregation used. In this case a 25 km hexagonal grid cell was selected to increase the minimum number of neighbors included in each analysis and also due to the intensive computational time

Table 4. Model summary from stepwise multivariate logistic regression used to predict the presence/absence of anthrax outbreaks in large ruminants across Kazakhstan. For variable definitions see Table 1.

	Parameter estimates		Wald x^2		
Variables	Coefficient	SE	x^2	*df*	*P*
Intercept	-52.899	7.912	44.70	1	< 0.001
Bio13	0.047	0.008	36.66	1	< 0.001
Bio 14	0.092	0.018	25.87	1	< 0.001
Bio12	2.82E-06	3.10E-07	82.51	1	< 0.001
Bio 7	0.027	0.006	20.42	1	< 0.001
Altitude	0.003	<0.001	19.47	1	< 0.001
NDVI	1.658	0.809	4.20	1	0.040
Bio 1	0.074	0.012	38.85	1	< 0.001
Latitude	0.602	0.106	32.04	1	< 0.001

required for AMOEBA (Duque et al., 2010). Selecting smaller grid cells for these analyses may have resulted in different patterns of clusters. However, tests on grid cells ≤10 km resulted in computational times greater than 10 days for each Rook contiguity order analysed.

The application of multiple orders of the Rook contiguity matrix to both AMOEBA and G_i* showed that each order (1st, 2nd or 3rd) produced unique varying spatial patterns of clusters. That is, the 1st order contiguity matrix elicited hotspots that 2nd order matrix did not and *vice versa*. The findings correspond with the suggestions of Ord and Getis (1995), which proposed analysing data at multiple distances. In this case the contiguity matrix served as our distance threshold and illustrates that the spatial structure of the data may change with distance. Due to the relatively large grid cell size of 25 km the use of a 2nd and 3rd order contiguity matrices may have exceeded the expectation of searching for cluster locally. Searching too far in this situation may bias the significance of the critical distance d_c towards the larger distance or contiguity order.

Similarly, core cut-off levels for AMOEBA elicited varying patterns of clustering at each threshold (2, 3, 6 and 12). However, there has been no previous discussion in the literature to guide the choice of an appropriate threshold. While a lower core threshold produced a greater number of hotspots this may not always be conducive to uncovering patterns of disease. Overestimating the burden of disease with high cluster numbers may strain public health resources, therefore finding a balance between the core threshold and the critical distance is crucial for appropriately analysing the spatial distribution of outbreaks.

Differences in the environmental signatures for clusters and non-clusters were consistent with research suggesting precipitation (Turner et al., 1999), soil pH (Van Ness, 1967; Smith et al., 2000), organic carbon content (Hugh-Jones and Blackburn, 2009), calcium (Himsworth, 2008), and temperature (Blackburn et al., 2007) may play a role in spore persistence and subsequent outbreaks. In terms of associating clusters with the environment, the shape or morphology of clusters may be an important aspect since ecological characteristics related to the presence of a disease may be distributed irregularly or unevenly across the landscape (Doi et al., 2008). In both the AMOEBA/non-AMEOBA and the G_i*/non-G_i* comparisons, significant differences between variables were noted. For example, Bio 13 and NDVI values were drastically different for both cluster techniques from non-clusters (Fig. 9). While there were clear differences between cluster and non-cluster comparisons, there were not necessarily discrete differences between AMOEBA and G_i*. Although the comparison of variables between AMOEBA and non-AMOEBA elicited the greatest differences, the Mann-Whitney U test did not establish any conclusive link between the ability of the cluster morphology statistic to better detect environmental signatures associated with anthrax outbreaks when compared to the local G_i*.

The logistic model focused exclusively on the risk of outbreaks in large ruminants in order to compare agreement between high risk areas defined by the model and hotspots from AMOEBA and G_i*. It is important to note that this model does not identify all areas in the country where anthrax is expected to occur (Joyner et al., 2010). Epizootic or enzootic anthrax activity has been detected throughout the region in several livestock species and this model does not capture the entire spatial heterogeneity of the disease across Kazakhstan (Aikembayev et al., 2010). Our model did, however, yield high (~83%) accuracy in correctly classifying grid cells as outbreak present or absent. Additionally it is likely that our model did not capture all of the potential factors related to the occurrence of the anthrax since there are several unrecognised factors that may contribute to the disease. Overall our classified logistic model predicted 32.2% of the landscape across Kazakhstan as being of high risk for anthrax outbreaks. In contrast, Joyner (2010) estimated that 54.1% of the landscape having the potential for spore habitat using an 8-variable, ecological niche model (Joyner, 2010).

While the scope of this current study was focused on local cluster techniques, it is worthwhile to note some of the differences in outbreak risk mapping from our logistic regression model and ecological niche modelling. The latter models estimate (by definition) the potential distribution of the target organism, *B. anthracis* spores, and may identify areas suitable for spores, where the disease may not occur or having been reported (Blackburn, 2010). In contrast, the risk model presented here targets outbreaks, or the likelihood of increased cases. It is also important to note that Joyner (2010) was modelling the pathogen from a combination of large and small ruminant occurrences, where our risk model is limited to large ruminants. We suggest that there is potential complimentary power in using both approaches to inform public health and veterinary management. For example, the ecological niche models can be used to identify the larger area of Kazakhstan where *B. anthracis* may persist, which can be used to identify where passive surveillance should

include diagnostics for anthrax. In contrast, active control areas can be better defined by the cluster analyses and the logistic modelling approach.

In comparing the distribution of clusters from both techniques with the logistic model, AMOEBA identified a greater number of clusters within areas designated as of high risk compared to G_i*. Clusters, however, were present across methodologies in the south of Kazakhstan indicating the possibility of an increased burden of disease there. Agreement between clustering methodologies across areas defined as being of high risk may illustrate spatial differences between endemic and sporadic areas. In this case endemic areas would be signified by greater agreement. Although cluster analyses were performed independently of the logistic model, the AMOEBA clusters showed a higher level of agreement with statistically significant areas classified as high risk for anthrax in large ruminants. This may be due to the ability of the AMOEBA algorithm to detect arbitrarily shaped hotspots; however, additional experiments are needed to confirm this.

AMOEBA was able to detect more irregular cluster patterns compared to G_i*, yet there is no definitive answer as to which methodology is the more robust one. The interpretation of G_i* as a tool for delimiting the extent of local spatial autocorrelation is clearer in the literature (Ord and Getis, 1995; Getis et al., 2003; Getis and Aldstadt, 2004), while the use of multiple distances in the application of AMOEBA is not well defined. Additionally, while AMOEBA has the ability to more accurately describe the shape of clusters its computational time limits its utility. Recently the intensive computational time of AMOEBA has been recognised and potential solutions may make AMOEBA a more viable option in the future (Duque et al., 2010). Given the potential strengths between methodologies, employing the use of more than one spatial modelling approach, may provide a better indication of the status of a disease (Jacquez and Greiling, 2003). For example, researchers can use G_i* to identify the maximum extent of local spatial autocorrelation and then incorporate that information to specify the weights matrix in a subsequent AMEOBA analysis of cluster morphology.

Implementing spatial clustering in the investigation of health events may provide information on irregular spatial patterns of disease as well as aid in the identification of environmental factors associated with the presence of a disease. Clustering can also be used to target specific areas with an elevated presence or potential presence for a disease to allow for a more efficient distribution of control measures such as vaccines (Eisen and Eisen, 2008). Areas in this study identified as being of high risk within a livestock group could be considered for active control measures, or priority areas for vaccine delivery, particularly if recent cases have been reported. However, it is noted that the investigation of disease occurrence is a complex, multifaceted problem. Population data on livestock were not available for this study, which is a major limitation in interpreting the proportion of disease in relation to animal densities and changes in population spatially over time. Despite the lack of population data, the raw outbreak numbers in this study may be useful for illustrating areas of increased disease presence. Future analyses should examine the level of statistical power for AMOEBA in addition to investigating clustering of anthrax at higher spatial resolution while investigating the effects of distance thresholds.

Acknowledgements

This work was funded by the US Defense Threat Reduction Agency through the Cooperative Biological Engagement Program in Kazakhstan. ITK and JKB are supported through the DTRA Joint University Partnership administered through the University of New Mexico. G. Temiraliyeva, Y. Sansyzbayev, T.A. Joyner, A.J. Curtis assisted with the development and maintenance of the anthrax GIS database.

References

Aikembayev A M, Lukhnova L, Temiraliyeva G, Meka-Mechenko T, Pazylov Y, Zakaryan S, Denissov G, Easterday WR, Van Ert MN, Keim P, Francesconi SC, Blackburn JK, Hugh-Jones ME, Hadfield TL, 2010. Historical distribution and molecular diversity of *Bacillus anthracis*, Kazakhstan. Emerg Infect Dis 16, 789-796.

Aldstadt J, Getis A, 2006. Using AMOEBA to create a spatial weights matrix and identify spatial clusters. Geogr Anal 38, 327-343.

Anderson N, Titterington D, 1997. Some methods for investigating spatial clustering, with epidemiological applications. J R Stat Soc Ser A 160, 87-105.

Anselin L, 1995. Local indicators of spatial association—LISA. Geogr Anal 27, 93-115.

Blackburn J, 2010. Integrating geographic information systems and ecological niche modeling into disease ecology. A case study of *Bacillus anthracis* in the United States and Mexico. In: Emerging and Endemic Pathogens. Advances in Surveillance, Detection, and Identification. EW SKP O'Connell, A Sulakvelidze, L Bakanidze (eds), pp. 59-88.

Blackburn J, McNyset K, Curtis A, Hugh-Jones M, 2007. Modeling the geographic distribution of *Bacillus anthracis*, the

causative agent of anthrax disease, for the contiguous United States using predictive ecologic niche modeling. Am J Trop Med Hyg 77, 1103-1110.

Clennon JA, Mungai PL, Muchiri EM, King CH, Kitron U, 2006. Spatial and temporal variations in local transmission of *Schistosoma haematobium* in Msambweni, Kenya. Am J Trop Med Hyg 75, 1034-1041.

de Castro MC, Sawyer DO, Singer BH, 2007. Spatial patterns of malaria in the Amazon: implications for surveillance and targeted interventions. Health Place 13, 368-380.

Doi Y, Yokoyama T, Sakai M, Nakamura Y, Tango T, Takahashi K, 2008. Spatial clusters of Creutzfeldt-Jakob disease mortality in Japan between 1995 and 2004. Neuroepidemiology 30, 222-228.

Dragon D, Rennie R, 1995. The ecology of anthrax spores: tough but not invincible. Canadian Vet J 36, 295-301.

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

Duque JC, Aldstadt J, Velasquez E, Franco JL, Betancourt A, 2010. A computationally efficient method for delineating irregularly shaped spatial clusters. J Geogr Syst, 1-18.

Eisen RJ, Eisen L, 2008. Spatial modeling of human risk of exposure to vector-borne pathogens based on epidemiological versus arthropod vector data. J Med Entomol 45, 181-192.

Eisen RJ, Reynolds PJ, Ettestad P, Brown T, Enscore RE, Biggerstaff BJ, Cheek J, Bueno R, Targhetta J, Montenieri JA, 2007. Residence-linked human plague in New Mexico: a habitat-suitability model. Am J Trop Med Hyg 77, 121-125.

Epp T, Argue C, Waldner C, Berke O, 2010. Spatial analysis of an anthrax outbreak in Saskatchewan, 2006. Canadian Vet J 51, 743-748.

Fielding A, Bell J, 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 24, 38-49.

Gates CC, Elkin BT, Dragon DC, 1995. Investigation, control and epizootiology of anthrax in a geographically isolated, free-roaming bison population in northern Canada. Can J Vet Res 59, 256-264.

Getis A, Aldstadt J, 2004. Constructing the spatial weights matrix using a local statistic. Geogr Anal 36, 90-104.

Getis A, Morrison AC, Gray K, Scott TW, 2003. Characteristics of the spatial pattern of the dengue vector, *Aedes aegypti*, in Iquitos, Peru. Am J Trop Med Hyg 69, 494-505.

Getis A, Ord JK, 1992. The analysis of spatial association by use of distance statistics. Geogr Anal 24, 189-206.

Hay S, Tatem A, Graham A, Goetz S, Rogers D, 2006. Global environmental data for mapping infectious disease distribution. Adv Parasitol 62, 37-77.

Himsworth CG, 2008. The danger of lime use in agricultural anthrax disinfection procedures. The potential role of calcium in the preservation of anthrax spores. Canadian Vet J 49, 1208-1210.

Hinman S, Blackburn J, Curtis A, 2006. Spatial and temporal structure of typhoid outbreaks in Washington DC, 1906-1909: evaluating local clustering with the Gi* statistic. Int J Health Geogr 5, 13.

Hugh-Jones M, 1999. 97 global anthrax report. J Appl Microbiol 87, 189-191.

Hugh-Jones M, Blackburn J, 2009. The ecology of *Bacillus anthracis*. Mol Aspects Med 30, 356-367.

Jacquez G, Greiling D, 2003. Local clustering in breast, lung and colorectal cancer in Long Island, New York. Int J Health Geogr 2, 3.

Jacquez GM, 2009. Cluster morphology analysis. Spat Spatiotemporal Epidemiol 1, 19-29.

Jeffery JAL, Yen NT, Nam VS, Hoffmann AA, Kay BH, Ryan PA, 2009. Characterizing the *Aedes aegypti* population in a Vietnamese village in preparation for a *Wolbachia*-based mosquito control strategy to eliminate dengue. PLoS Negl Trop Dis 3, e552.

Joyner T, Lukhnova L, Pazilov Y, Temiralyeva G, Hugh-Jones M, Aikimbayev A, Blackburn J, 2010. Modeling the potential distribution of bacillus anthracis under multiple climate change scenarios for Kazakhstan. PLoS One 5, e9596.

Joyner TA, 2010. Ecological niche modeling of a zoonosis: a case study using anthrax and climate change in Kazakhstan, Master's Thesis. University of Florida, Gainesville, 150 pp.

Kelly-Hope LA, Hemingway J, McKenzie FE, 2009. Environmental factors associated with the malaria vectors *Anopheles gambiae* and *Anopheles funestus* in Kenya. Malar J 8, 268.

Kracalik I, Lukhnova L, Aikimbayev A, Pazilov Y, Temiralyeva G, Blackburn JK, 2011. Incorporating retrospective clustering into a prospective cusum methodology for anthrax. Evaluating the effects of disease expectation. Spat Spatiotemporal Epidemiol 2, 11-21.

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

Moran PAP, 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17-23.

O'Sullivan D, Unwin DJ, 2003. Geographic information analysis. John Wiley & Sons Inc.

Ord JK, Getis A, 1995. Local spatial autocorrelation statistics: distributional issues and an application. Geogr Anal 27, 286-306.

Parkinson R, Rajic A, Jenson C, 2003. Investigation of an anthrax outbreak in Alberta in 1999 using a geographic information system. Canadian Vet J 44, 315-318.

Rogers DJ, Randolph SE, Snow RW, Hay SI, 2002. Satellite imagery in the study and forecast of malaria. Nature 415, 710-715.

Ruiz MO, Tedesco C, McTighe TJ, Austin C, Kitron U, 2004. Environmental and social determinants of human risk during

a West Nile virus outbreak in the greater Chicago area, 2002. Int J Health Geogr 3, 8.

Smith K, De Vos V, Bryden H, Hugh Jones M, Klevytska A, Price L, Keim P, Scholl D, 1999. Meso scale ecology of anthrax in southern Africa: a pilot study of diversity and clustering. J Appl Microbiol 87, 204-207.

Smith K, DeVos V, Bryden H, Price L, Hugh-Jones M, Keim P, 2000. *Bacillus anthracis* diversity in Kruger National Park. J Clin Microbiol 38, 3780-3784.

Sokal R, Rohlf F, 1995. Biometry (3rd edn). WH Freman and Company, New York.

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

Turner A, Galvin J, Rubira R, Condron R, Bradley T, 1999. Experiences with vaccination and epidemiological investigations on an anthrax outbreak in Australia in 1997. J Appl Microbiol 87, 294-297.

Van Ness G, 1967. Geologic Implications of Anthrax. Geol S AM S 90, 61-64.

Van Ness GB, 1971. Ecology of anthrax. Science 172, 1303-1307.

Weeks JR, Getis A, Hill AG, Agyei-Mensah S, Rain D, 2010. Neighborhoods and fertility in Accra, Ghana. An AMOEBA-based approach. Ann Assoc Am Geogr 100, 558-578.

Woods C, Ospanov K, Myrzabekov A, Favorov M, Plikaytis B, Ashford D, 2004. Risk factors for human anthrax among contacts of anthrax-infected livestock in Kazakhstan. Am J Trop Med Hyg 71, 48-52.

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

Yiannakoulias N, Wilson S, Kariuki HC, Mwatha JK, Ouma JH, Muchiri E, Kimani G,. Vennervald BJ, Dunne DW, 2010. Locating irregularly shaped clusters of infection intensity. Geospat Health 4, 191-200.

Zweig MH, Campbell G, 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 39, 561-577.