Spatio-temporal analysis and visualisation of the anthrax epidemic situation in livestock in Kazakhstan over the period 1933-2016

Sarsenbay K. Abdrakhmanov,[1] Yersyn Y. Mukhanbetkaliyev,[1] Fedor I. Korennoy,[2] Bolat Sh. Karatayev,[3] Aizada A. Mukhanbetkaliyeva,[1] Aruzhan S. Abdrakhmanova[1]

[1]S. Seifullin Kazakh Agrotechnical University, Astana, Kazakhstan; [2]Federal Center for Animal Health (FGBI ARRIAH), Vladimir, Russia; [3]Kazakh Veterinary Research Institution, Almaty, Kazakhstan

Abstract

An analysis of the anthrax epidemic situation among livestock animals in the Republic of Kazakhstan over the period 1933-2016 is presented. During this time, 4,064 anthrax outbreaks (mainly in cattle, small ruminants, pigs and horses) were recorded. They fall into five historical periods of increase and decrease in the annual anthrax incidence (1933-1953; 1954-1968; 1969-1983; 1984-2001; and 2002-2016), which has been associated with changes in economic activity and veterinary surveillance. To evaluate the temporal trends of incidence variation for each of these time periods, the following methods were applied: i) spatio-temporal analysis using a *space-time cube* to assess the presence of hotspots (*i.e.*, areas of outbreak clustering) and the trends of their emergence over time; and ii) a linear regression model that was used to evaluate the annual numbers of outbreaks as a function of time. The results show increasing trends during the first two periods followed by a decreasing trend up to now. The peak years of anthrax outbreaks occurred in 1965-1968 but outbreaks still continue with an average annual number of outbreaks of 1.2 (95% confidence interval: 0.6-1.8). The space-time analysis approach enabled visualisation of areas with statistically significant increasing or decreasing trends of outbreak clustering providing a practical opportunity to inform decision-makers and allowing the veterinary services to concentrate their efforts on monitoring the possible risk factors in the identified locations.

Introduction

The zoonotic disease anthrax, caused by the spore-forming bacterium *Bacillus anthracis*, affects most mammals, especially herbivores but also humans. The spores are highly resistant and can survive in the environment for decades. Although vaccination and several effective antibiotics exist, treatment is generally sporadic in the less developed parts of the world where anthrax is common. The course of the disease is usually rapid and almost always fatal.

The first reliable information of anthrax in the Republic of Kazakhstan (RK), formerly part of the Union of Soviet Socialist Republics (USSR), refers to the end of the 19th century, while official recording of cases did not start until the early 1930s. The sapronotic existence of the infectious agent determines its natural focality (Lukhnova *et al.*, 2004). Anthrax soil foci have been accumulating over a long period of time in the RK. Throughout the 20th century, burial of animals that had died from anthrax was carried out at the site or in its immediate vicinity until the Ministry of Agriculture, USSR issued a decree in 1951 requiring the mandatory burning of all animal corpses without prior action in an effort to end further soil contamination (Cherkassky, 2002). However, a significant number of anthrax cases in animals and humans were still recorded during the period from the mid-1950s to the 1980s. It should be mentioned that in the 1930s until the 1950s a lot of anthrax outbreaks remained unaccounted for due to a lack of staff, lack of proper accounting and illiteracy of the population leading to thousands of anthrax-infected corpses of farm animals being buried in the ground. Due to the long-term survival of *B. anthracis* spores in the soil, the sites of these spontaneous burials contributed to an epidemic instability of the area.

From the early 1950s, vaccination of agricultural animals against anthrax was introduced in the RK, but mass vaccination was only launched in 1961. Nevertheless, even during this period the vaccination coverage was far from complete due to intensive growth of livestock populations and inadequate local provision of vaccines, lack of staff and poor accountability. In parallel, a far-reaching campaign to develop virgin and fallow lands was started

OPEN ACCESS

in the USSR that largely affected the territory of the RK, accounting for 61% of developed land (McCauley, 1976; Rowe, 2011). The development of virgin and fallow land was accompanied by a massive influx of people into the Republic from other regions of the USSR (more than 2 million people), as well as a concomitant increase in the number of livestock (the number of small ruminants and pigs increased four times during the period 1941 to 1961). This resulted in a significant growth of incidence in both humans and animals in 1950s. Although there was a downward trend in the annual incidence of anthrax after 1969, the situation remained tense and more 50 anthrax outbreaks per year were registered in the Republic until the early 1980s. Following the introduction of follow-up vaccination of susceptible animals in 1981 and strict monitoring of compliance with veterinary legislation, the situation was stabilised at an average of 20 outbreaks per year.

After the collapse of the USSR, the number of livestock decreased by more than half between 1991 and 1998 leading to the disintegration of the existing management system and the large collective and state farms and significant migration of the rural population to cities producing a decrease in annual anthrax mortality. The gradual improvement of the economic situation in the country since the late 1990s followed by large investments in livestock and upgraded veterinary services also contributed to an improvement of the epizootic situation. The current epizootic situation can be characterised as stable with an average annual number of anthrax outbreaks of 1.2 (95% confidence interval: 0.6-1.8). Despite the success achieved, animals and humans still succumb to this disease today in various parts of the RK. As a result, the risk of death from anthrax infection in humans, and the huge financial costs associated with eliminating outbreaks in animals require a study of the epidemic situation in the country, where old foci continue to be identified and new ones registered. A specific feature of the causative agent is that the reactivation of conserved soil foci usually occurs because of human economic activity, such as wasteland reclamation or construction work that brings *B. anthracis* back to the surface from deeper soil layers (Knop, 1981). The formation of new foci due to natural factors, *e.g.*, the relocation of soil with spores by wind, rain or flooding, is also likely (Hugh-Jones and Blackburn, 2009).

Access to records of anthrax outbreak in the country since the early 20th century enables in-depth studies of patterns of emergence and re-emergence of the disease. Available datasets can be analysed using special research methods, in particular those including space-time analysis that can visualise patterns based on geographical information systems (GIS). However, only few studies employing GIS-based methods to search for patterns of spatial and spatio-temporal clustering of anthrax exist, *e.g.*, Kracalik *et al*. (2011, 2012) who studied outbreaks using a dataset covering the period 1960 to 2006.

The purpose of the present research was to identify and visualise spatio-temporal anthrax hotspots and their temporal trends taking into account data for time period as broad as possible (1933-2016). We applied the user-friendly *space-time cube* technique supplemented by linear regression as this would contribute to the development of preventive and contra-epizootic measures as well as rational use of material, labour and financial resources. The technical approaches used made it possible to comprehensively analyse epidemiologically significant information related to the geographical spread of anthrax. They also allowed the compilation of an accurate review and, potentially, a forecast of the epidemic situation by revealing regularities in the spatio-temporal distribu-

tion and the activity of stationary unfavourable anthrax, significantly contributing to the assessment of the current epidemic situation.

Materials and Methods

The anthrax epidemic situation in the RK was investigated for the longest period for which official records exist, *i.e.*, 1933-2016.

Study area

RK is a central Asian country with an area of more than 2.7 million km^2 and a current population of 17.8 million. The average population density is about 6.5 people/km^2. The country has extensive natural resources with agriculture, in particular livestock breeding, being one of the main branches of the national economy. Administratively, the country is divided into 14 regions or *oblasts*, which in turn are divided into 179 districts or *rayons*.

Data

The data were obtained from the RK register of areas with a high risk of anthrax (Cadastral, 2002), supplemented with material from statistical veterinary reports posted on the official websites of regional and republican veterinary authorities, as well as with official data from the World Organization for Animal Health (OIE) and material from our own expeditionary research.

In total, there were 4,064 anthrax outbreaks in the RK during the study period (Figure 1). In this research, we defined an outbreak as a recorded occurrence of anthrax cases in a geographically localised population of animals (a herd on pasture, on a farm, in a village, *etc*.), followed by the burial of dead animals in the immediate vicinity (with or without cremating the remains). Since some outbreaks were repeatedly recorded in the same locations (villages, pastures) at different years, the total number of unique locations was 1,798.

Methods of analysis

In order to facilitate the space-time analysis, we divided the entire historical period from 1933 to 2016 into five time intervals corresponding to the main phases of changes in economic development and veterinary surveillance of the RK: 1) 1933-1953; 2) 1954-1968; 3) 1969-1983; 4) 1984-2001; 5) 2001-2016. For each of these periods we 1) estimated the general temporal trend of
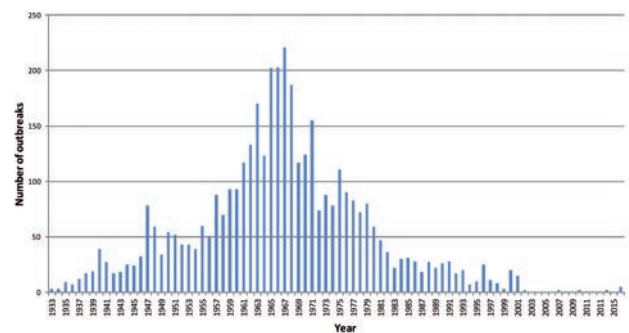

Figure 1. Annual dynamics of the registration of anthrax outbreaks in the Republic of Kazakhstan, 1933-2016.

annual incidence; 2) identified statistically significant hotspots and 3) indicated territories with increasing/decreasing tendencies in hotspot emergence. We carried out the analysis through the construction of a set of space-time cubes for hotspot analysis as used by individual researchers (Andrienko *et al.*, 2003; Harris *et al.*, 2017), which are available from the Space Time Pattern Mining toolbox within ArcGIS geographical information software package (ESRI, 2017). The space-time cube can be defined as a regular three-dimensional structure based on the geographic space along the x an y coordinates, with the vertical dimension representing time. The cube consists of space-time bins aggregating studied data points. For each bin, space and time dimensions were set. The further analysis relied on counting the number of data points within individual bins and application of various statistical approaches to reveal patterns of data distribution through both space and time looking for statistically significant cluster of data points or associated values. Epidemiologically, a hotspot is the clustering of outbreaks in a particular location over particular time periods. For example, the presence of a large number of outbreaks within a particular bin indicates the presence of a statistically significant hotspot in a given space-time cell provided that it is surrounded by bins (both in space and time dimensions) that also have a large number of outbreaks. We selected 50 km for the spatial bin dimension and one year as time period, which would reveal the patterns of annual changes in the epidemic situation in a 50 x 50 km space (2,500 km^2). The size of 50 km allows accounting for possible reporting bias when known outbreaks could be linked to the nearest villages.

When doing this type of analysis, it is useful to bear in mind that hotspots come in various guises, *e.g.*, new, consistent, increasing, sporadic, decreasing, fluctuating and historical (ESRI, 2017) depending on the patterns of their emergence over time. The identification of a specific type of hotspot at a given location demonstrates an increased clustering of outbreaks of the disease at the location during different time periods.

Software

The analysis using the space-time cube was performed using a Space Time Patterns Mining Tools software tool package, embed-

ded in the geographical information system (GIS) ArcGIS 10.4.1 by ESRI (Redlands, CA, USA). The systematisation of the anthrax database and fitting the distributions were carried out using a standard Microsoft Excel package with @Risk add-in (Palisade Inc., Ithaca, NY, USA).

Statistical approach

We applied Getis-Ord Gi* spatial statistics in its temporal interpretation to identify statistically significant hotspots of anthrax outbreaks (Ord and Gettis, 1995), where a conclusion on the presence/absence of a hotspot in a space-time bin is reached on the basis of z-values. We also used Mann-Kendall statistics (Mann, 1945; Kendall and Gibbons, 1990) to identify statistically significant trends of hotspot emergence. Mann-Kendall statistics, based on a comparison of the values in consequent space-time bins, are used to estimate trends in the evolution of the z-statistics over time. The value delivered by the Mann-Kendall statistics allows a conclusion to be reached about increasing or decreasing tendencies in the emergence of hotspots at a given location, which tendencies can be treated as indicators of improvement or deterioration of an epidemic situation in this particular location. In addition to the spatiotemporal analysis, we used a linear regression model to estimate the intensity of increases/decreases in the annual incidence during each time period. In this case, the annual number of anthrax outbreaks (N_i) was the dependent variable, with the year (Y_i) as the independent variable. Thus, the regression equation was $N_i \sim \beta + \alpha Y$, where β is the intercept and α the slope. Goodness of a model fit were assessed by the coefficient of determination R^2, which indicates how much of the data variability is explained by the model (range: 0 ... 1). The statistical software environment R was applied to calculate the temporal correlation of the annual incidence (R Core Team, 2014).

Results

Table 1 summarises the results of the analysis of the annual anthrax incidence for the five chosen periods using both space-time analysis and linear regression.

Table 1. Summary of annual anthrax incidence in the Republic of Kazakhstan for the five historical periods 1993-2016.

Period	Number of outbreaks							Spatio-temporal analysis		Linear regression					
	Total	Cattle	Horse	Pig	Small ruminants	Ann min°	Ann max#	Trend and Z-value	P	Intercept	Intercept SD	Slope	Slope SD	P	R^2
1933-1953	615	415	34	22	144	3	78	Increasing Z=4.3543	<0.001	0.8286	5.5952	2.5870	0.4456	<0.001	0.6395
1954-1968	1,850	1,051	90	95	614	39	221	Increasing Z=4.4098	<0.001	21.305	10.851	12.754	1.193	<0.001	0.8978
1969-1983	1,236	747	73	30	386	22	155	Decreasing Z=-3.5631	<0.001	135.457	10.216	-6.632	1.124	<0.001	0.7283
1954-2001	346	218	34	5	89	3	31	Decreasing Z=-3.1863	0.001	30.1830	3.0813	-1.1538	0.2847	<0.001	0.5066
2002-1916	17	15	1	1	0	0	5	NA§	NA	0.14286	0.66566	0.13214	0.07321	0.0943	0.2004

SD, standard deviation; NA, not available. °Minimum annual number of outbreaks for the corresponding time period; #maximum annual number of outbreaks for the corresponding time period; §corresponding analysis method could not be applied due to insufficient number of outbreaks in this time period.

1933 to 1953

The annual number of outbreaks during this period ranged from 3 to 78 with an average value of 29±4. The maximum number of outbreaks was registered in 1947. The annual incidence increased with an average rate of 2.587±0.446 outbreaks per year. New hotspots in this period emerged in East Kazakhstan region, and consecutive and sporadic hotspots in the Kostanay and South Kazakhstan regions. Statistically significant increasing trends of hotspots with wide distributions were registered in the Akmola and Karaganda regions (Figure 2).

1954 to 1968

There was a sharp increase in the annual incidence from 40 to 220 outbreaks with an average rate of 12.754±1.193 outbreaks per year in this period. The consecutive hotspots were registered in the West Kazakhstan, South Kazakhstan as well as in the Zhambyl and
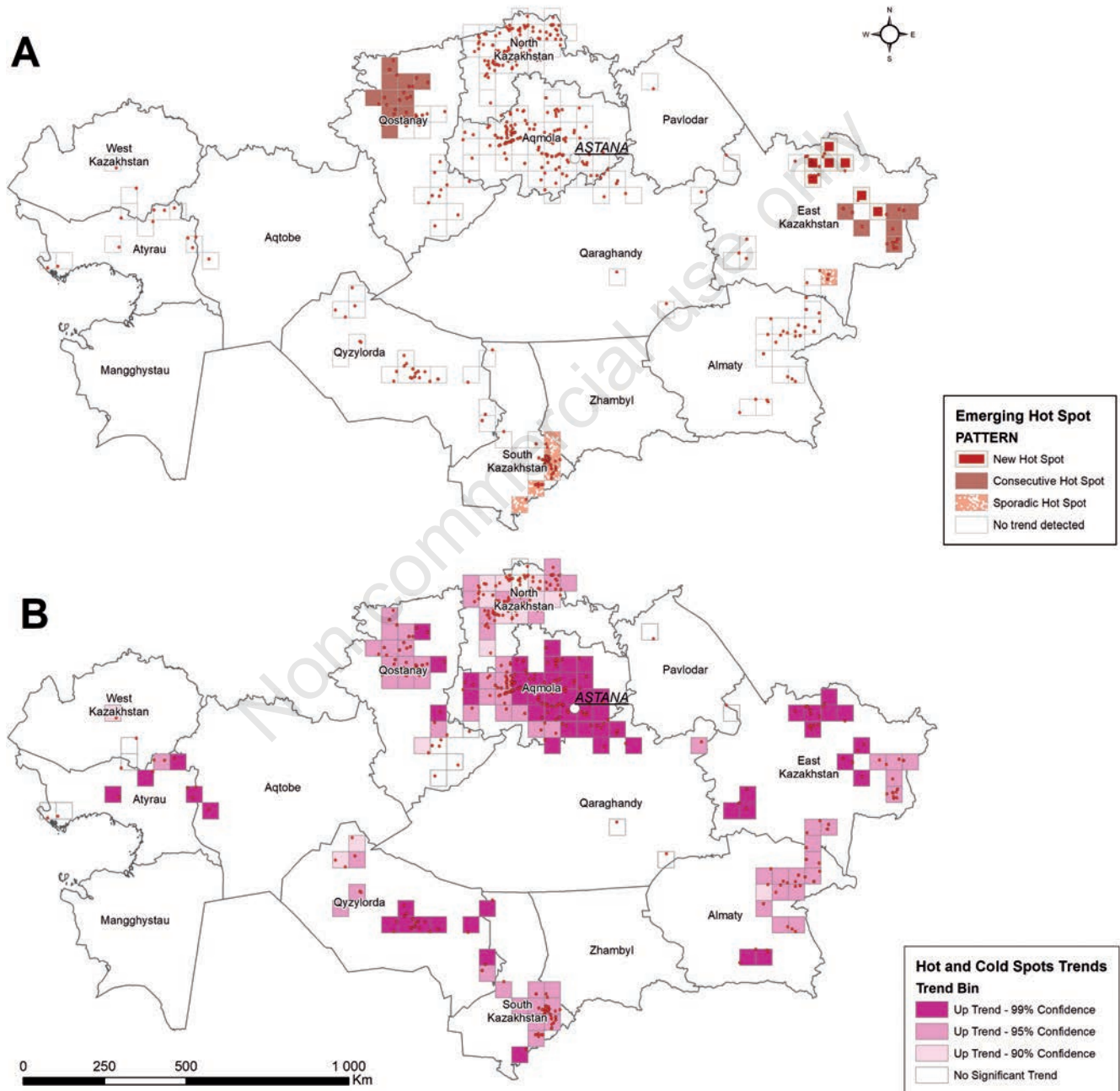

Figure 2. Spatio-temporal analysis of anthrax outbreaks in the Republic of Kazakhstan 1933-1953. Statistically significant hot/cold spots (A); statistically significant trends of hot/cold spots emergence (B).

East Kazakhstan regions. In this case, statistically significant upward trends of hotspot emergence were recorded on most of the inhabited and cultivated territory of the RK (Figure 3).

cally significant decreasing trends of hotspots were present in the vast territories of the northern, eastern and southern regions of the RK (Figure 4).

1969 to 1983

There was a decline in the annual incidence from 220 to 20-30 annual outbreaks with an average rate of 6.632±1.124 outbreaks per year. No hotspots were detected during this period and statisti-

1984 to 2001

In this period, the anthrax incidence showed a tendency to slowly decline at a rate of 1.154±0.285 outbreaks per year. The general trend was negative. During this period, sporadic hotspots
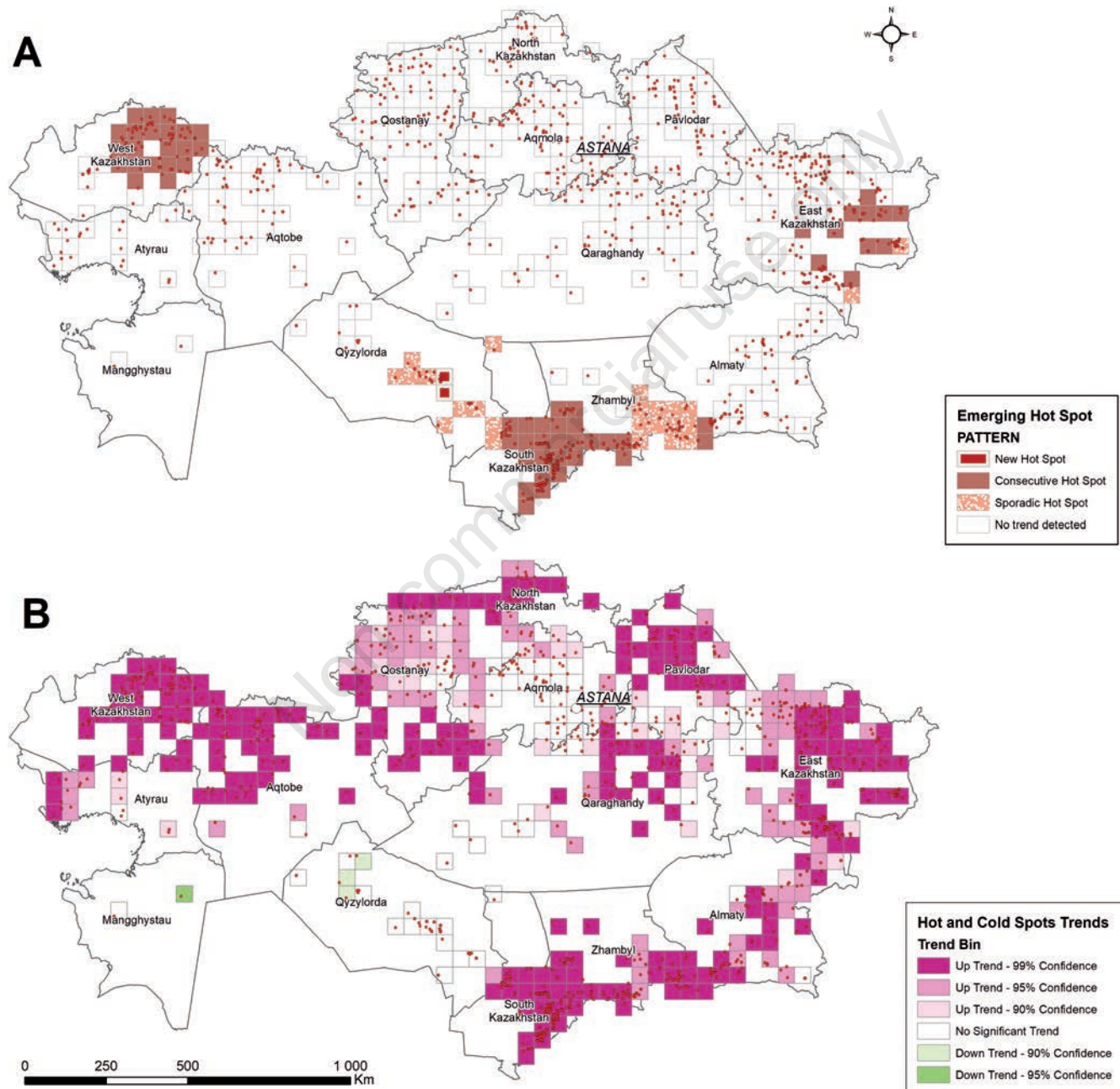

Figure 3. Spatio-temporal analysis of anthrax outbreaks in the Republic of Kazakhstan, 1954-1968. Statistically significant hot/cold spots (A); statistically significant trends of hot/cold spots emergence (B).

were recorded in the South Kazakhstan and Zhambyl regions, but the overall trend of hotspot emergence decreased (Figure 5).

2002 to 2016

From 0 to 5 outbreaks per year were registered in the latest period providing an average number of 1.2 (0.6-1.8). Given the relative constancy of the annual number of outbreaks, this value can be accepted as the average predicted incidence rate per year for the next time period. As the space-time analysis method requires at least 60 events, it could not be applied due to the low total number of outbreaks for the given time. The outbreak map for this period is shown in Figure 6.

Discussion

Our analysis of the epizootic situation for anthrax in the RK over the whole study period shows that the main temporary stages
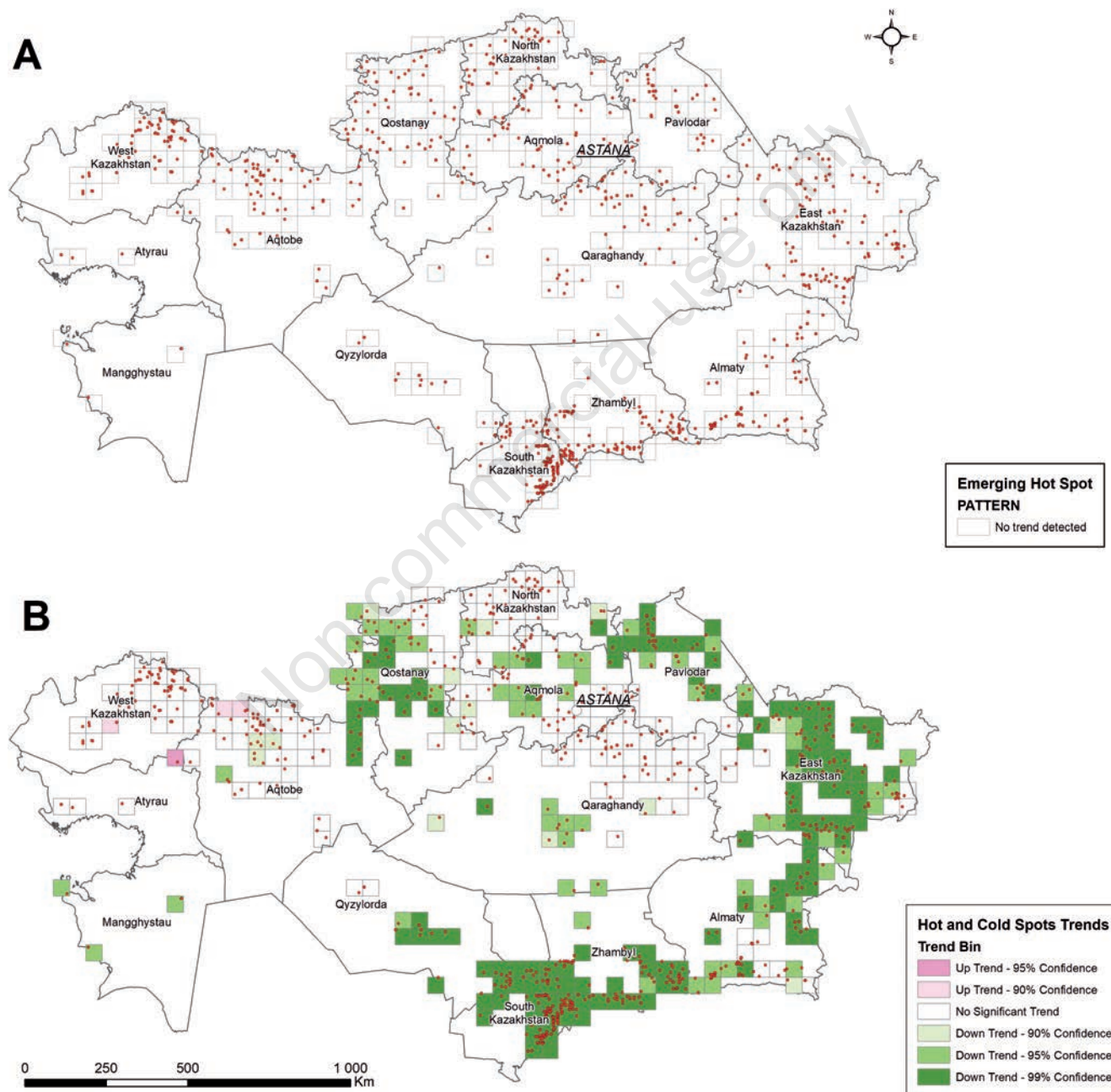

Figure 4. Spatio-temporal analysis of anthrax outbreaks in the Republic of Kazakhstan 1969-1983. Statistically significant hot/cold spots (A); statistically significant trends of hot/cold spots emergence (B).

of economic and agricultural growth in the Republic satisfactory explain the variant periods of anthrax incidence. Thus, there was a significant increase in annual incidence throughout the period 1933 to 1953, likely to be associated with the development of the national economy leading to an increase in the number of livestock without any measures for disinfection and disposure of dead animals. According to the national statistics agency, the cattle stock in the Republic consisted of 3.3 million heads in 1941 and reached 4.5 million heads by 1951. It should also be noted that the absence

of widespread registration of anthrax cases along with the population's low literacy rate with respect to pathogenesis and disease control measures probably led to a significant underestimation of the incidence rate during this time period.

The campaign to develop virgin and fallow lands that started in 1954 led to a dramatic growth of anthrax incidence. The increase of the susceptible populations (both of humans and animals) together with expansion of activities into virgin lands led to the removal of the causative agent from the soil surface at old (unac-
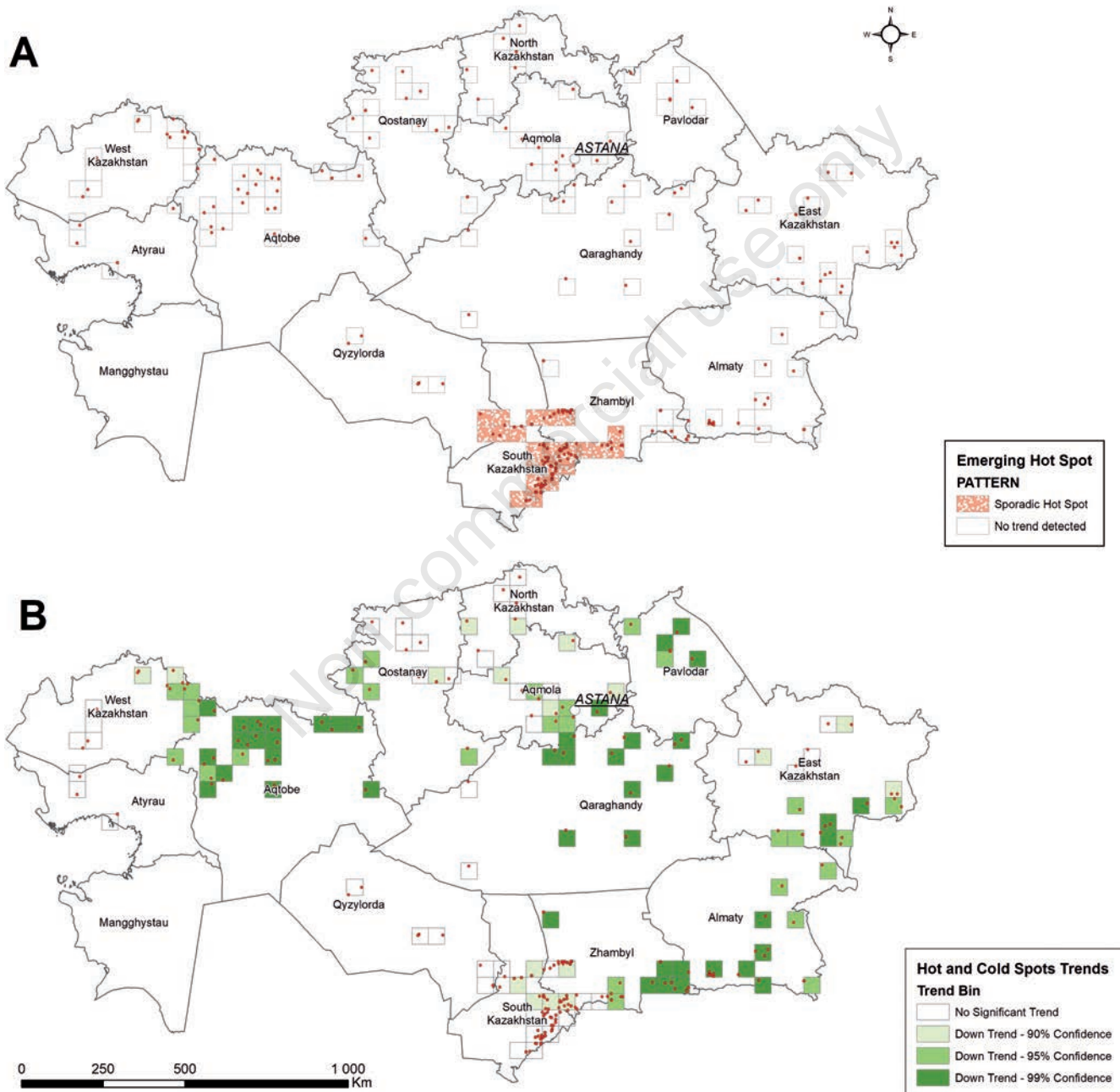

Figure 5. Spatio-temporal analysis of anthrax outbreaks in the Republic of Kazakhstan, 1984-2001. Statistically significant hot/cold spots (A); statistically significant trends of hot/cold spots emergence (B).

counted) anthrax burial sites and to intensive wind erosion of the soil, which contributed to the wind-blown distribution of anthrax spores over considerable distances. All this significantly increased exposure to anthrax spores, as confirmed by studies by Soviet scientists (Adamovich and Nikonov, 1970). In contrast, improvements in the veterinary surveillance along with introduction of mass vaccination led to a considerable reduction in the number of foci and brought anthrax away from its previous status of large-scale epizooty in the period 1961-2010. The spatio-temporal analysis of the epizootic situation demonstrates the absence of hotspots and a confident downward trend in their formation in nearly all the affected territories in the Republic starting from 1969 (Figure 4). Still, there was a number of sporadic hotspots in the southern part of the Republic after the 1980s, though combined with a downward trend in their emergence in previously affected territories (Figure 5).

The low number of outbreaks in the most recent period (2010-2016) did not allow the application of the chosen method of space-time analysis, but a significant spatial fragmentation of outbreaks can be observed during this period as well as the absence of repeated outbreaks in particular locations. Currently, as a result of the systematic implementation of comprehensive preventive and antiepizootic measures, the area of anthrax registration has gradually diminished. Compared to the 1950-1970s, large territories have become disease-free, including the North Kazakhstan, Aqtobe, Atyrau, Kostanay and Mangystau regions, where neither animal deaths due to anthrax, nor human cases have been recorded

in the last 20 to 30 years (in Mangystau for 50 years). Nevertheless, disease outbreaks are being registered in regions that previously experienced intensive epidemics, which indicates a continuing threat of the removal of the anthrax disease agent from old (unrecorded) burials as a result of agriculture, construction and other human activities, as well as the impact of various natural factors related to old soil foci (spring floods, inundations, earthquakes, etc.).

Our analysis shows a clearly expressed non-uniformity in the territorial distribution of anthrax foci and differences in their epidemiological activity throughout the country. Currently, new data on the epidemiology and ecology of anthrax are being accumulated, which are associated with the introduction of new forms of economic relations in the country, changes in people's social conditions and the formation of private property in agriculture. Today, anthrax in the RK is recorded in areas where developed livestock breeding exists and where the soil is neutral or slightly alkaline with a high humus content together with high temperatures and sufficient humidity (Aikembayev et al., 2010; Abdrakhmanov et al., 2017). In this connection, there is a need for an in-depth causal analysis of the incidence of anthrax in relation to natural (soil types, precipitation, temperature etc.) and anthropogenic (livestock, construction, land improvement, etc.) factors.

We found analysis using the space-time cube useful for studying retrospective data on disease incidence as it allowed us to draw conclusions regarding both trend variations within the target territory over time and localisation of sites with the most intensive incidence.
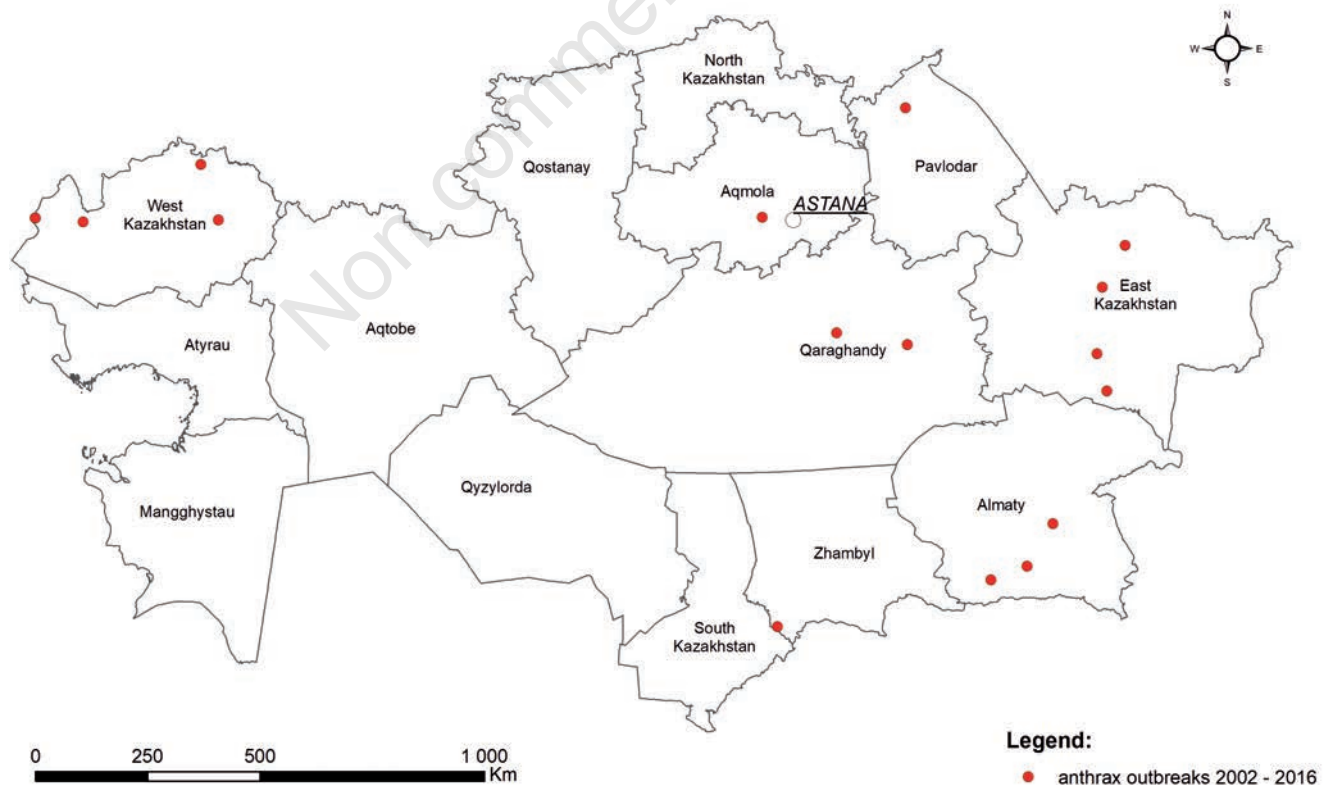

Figure 6. Anthrax outbreaks in the Republic of Kazakhstan over the period 2002-2016.

From the epidemiological point of view, it is especially useful to assess the extent of the manifestation of the epidemiological process in the target territory of the country by visualising spatial cells where a steady increase in the incidence exists (*i.e.*, repeated outbreaks). In the case of anthrax as a natural focal disease, this tendency may indicate the presence of foci in the given territory, such as spontaneous, unregistered and insufficiently disinfected anthrax burial sites. As for the shortcomings of this method, the need to set temporal and spatial dimensions of the cube cells was noted, a fact which is not always obvious and requires additional study. The choice of spatial and temporal resolutions of the space-time cube has a significant effect on the returned result. In our case, since the dates of the outbreaks could not be known more accurately than years, the minimum possible time step had to be one year. If more detailed time data were available, and if necessary to analyse the situation for shorter time periods, time steps such as months or weeks could be chosen. For our analysis, we accepted that the spatial size of the bin automatically calculated by ArcGIS was based on the spatial distribution of the data analysed. When studying a disease situation with a strongly pronounced local transmission, the spatial step can be estimated as the maximum clustering distance, for example, using the K-function (Dixon, 2012).

In general, the use of studies with information and communication technologies allows critical information to be obtained in advance on the various scenarios for the emergence, development and counteraction of natural focal epidemics. The most urgent problem is the development of an epidemic or outbreak as a result of the accidental removal of the causative agent of anthrax from conserved soil foci. Therefore, more complete and detailed information permitting a more detailed analysis of the spatio-temporal patterns of the disease's spread could serve a basis for the development of effective preventive and anti-epizootic measures that would ensure epidemiological well-being throughout the Republic.

Conclusions

We show here an example of the application of the space-time analysis technique using a space-time cube to identify trends in increasing and decreasing anthrax incidence on the territory of the RK over a long historical period. This type of analysis is well suited for working with historical data on disease incidence and allows a visual assessment of the development of the epidemic situation during certain time periods that are associated with the main stages in the development of the country's economy. The visualisation of hotspots and their trends enabled us to reveal the territories most at risk for an epidemic during the period under consideration. The obtained data should facilitate veterinary surveillance leading to the development of plans for preventive and anti-epizootic measures against anthrax.

References

Abdrakhmanov SK, Mukhanbetkaliyev YY, Korennoy FI, Sultanov AA, Kadyrov AS, Kushubaev DB, Bakishev TG, 2017. Maximum entropy modelling risk of anthrax in the Republic of Kazakhstan. Prev Vet Med 144:149-57.

Adamovich VL, Nikonov NN, 1970. The importance of landscape-ecological factors in the epizootology of anthrax. Message 2: Comparative method of assessment of epizootic tension of a territory. J Microbiol 8:113-7.

Aikembayev AM, Lukhnova L, Temiraliyeva G, Meka-Mechenko T, Pazylov Y, Zakaryan S, Denissov G, Ryan Easterday W, Van Ert MN, Keim P, Francesconi SC, Blackburn JK, Hugh-Jones M, Hadfield T, 2010. Historical distribution and molecular diversity of *Bacillus anthracis,* Kazakhstan. Emerg Infect Dis 16:789-96.

Andrienko N, Andrienko G, Gatalsky P, 2003. Visual data exploration using space-time cube. Proceedings of the 21st International Cartographic Conference (ICC), Durban, South Africa.

Cherkassky BL, 2002. Epidemiology and prevention of Anthrax. INTERSEN, Moscow, Russia.

Dixon PM, 2012. Ripley 's K function. In: El-Shaarawi AH and Piegorsch WW (eds.) Encyclopedia of Environmetrics. John Wiley & Sons, Chichester, UK, pp. 1796-803.

ESRI, 2017. How emerging hot spot analysis works. Available from: http://desktop.arcgis.com/en/arcmap/latest/tools/space-time-pattern-mining-toolbox/learnmoreemerging.htm#GUID-09587AFC-F5EC-4AEB-BE8F-0E0A26AB9230

Harris NL, Goldman E, Gabris C, Nordling J, Minnemeyer S, Ansari S, Lippmann M, Bennett L, Raad M, Hansen M, Potapov P, 2017. Using spatial statistics to identify emerging hot spots of forest loss. Environ Res Lett 12:024012.

Hugh-Jones M, Blackburn J, 2009. The ecology of *Bacillus anthracis*. Mol Aspects Med 30:356-67.

Kendall MG, Gibbons JD, 1990. Rank correlation methods. Griffin, London, UK.

Knop AG, 1981. The influence of anthropogenic transformation of environment on Anthrax soil foci. Actual problems of zoonotic infections. pp. 25-27.

Kracalik IT, Blackburn JK, Lukhnova L, Pazilov Y, Hugh Jones ME, Aikimbayev A, 2012. Analysing the spatial patterns of livestock anthrax in Kazakhstan in relation to environmental factors: A comparison of local (Gi*) and morphology cluster statistics. Geospat Health 7:111-26.

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 Spatiotemp Epidemiol 2:11-21.

Lukhnova LY, Ajkimbaev AM, Pazylov EK, 2004. Epidemic process of Anthrax in Kazakhstan. Bull Agric Sci Kazakhstan Bastau 7:44.

Mann HB, 1945. Nonparametric tests against trend. Econometrica 13:245-59.

McCauley M, 1976. Khrushchev and the development of Soviet agriculture. The Virgin Land Programme 1953-1964. Palgrave Macmillan, UK.

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

R Core Team, 2014. R foundation for Statistical Computing, Vienna. Available from: http://R-project.org/

Rowe WC, 2011. Turning the Soviet Union into Iowa: The Virgin Lands Program in the Soviet Union. In: Brunn S. (ed.) Engineering earth. Springer, Dordrecht, Germany. pp. 237-56.

Mapping displaced populations with reference to social vulnerabilities for post-disaster public health management

Junaid Ahmad,[1] Anees Ahmad,[2] Mokbul Morshed Ahmad,[3] Nafees Ahmad[4]

[1]Department of Disaster Preparedness, Mitigation and Management, Asian Institute of Technology, Pathumthani, Thailand; [2]Department of Space Science, Institute of Space Technology, Islamabad, Pakistan; [3]School of Environment, Resources and Development, Asian Institute of Technology, Pathumthani, Thailand; [4]Department of Development Studies, Pakistan Institute of Development Economics, Islamabad, Pakistan

Abstract

Millions of people are currently displaced from their homes because of local and international conflicts. In the last two decades, a substantial increase in the number of displaced people has been recorded. We measured the social vulnerabilities of displaced populations using a mathematical approach in combination with application of geographical information systems (GIS) tools and techniques to visualise movement and draw attention to the location of significant concentration of vulnerabilities. A retrospective study approach based on datasets collected from governmental and non-governmental organisations working with refugees and internally displaced persons in Pakistan was used. We applied simple mathematical formulas to calculate and map various types of vulnerability, such as refugee population, absorption capacity, unmet needs and overall vulnerability. This approach displays risks and vulnerabilities of displaced populations in an easily understood and straightforward manner that can be replicated in other parts of the world.

Introduction

Hundreds of active conflicts engross a large part of the world at present (Schultz, 2017). The most prominent of these involve Arabic countries, including Syria, Iraq, Yemen and Libya (Byman and Speakman, 2016) together with Afghanistan and Pakistan (Taylor, 2016). As a result, millions of people have lost their lives, and many have been forced to migrate from their homes and communities (Taylor, 2016). According to recent estimates by the United Nations High Commissioner for Refugees (UNHCR), about 6.5 million people are currently displaced, either as refugees (people forced to leave their country) or as internally displaced persons (IDPs). The number of refugees in the world is at the highest level since the Second World War (UNHCR, 1992, 2015) with the majority coming from Syria and Afghanistan (Donini et al., 2016). The two countries currently hosting the greatest numbers of refugees are Turkey with 2.5 million and Pakistan with 4.5 million (UNHCR, 2015).

The United Nations Office for Disaster Risk Reduction (UNISDR) emphasises that displacement in itself is a driver of future disaster risks, and it places people at a higher risk of impoverishment and human rights violations while exacerbating any pre-existing vulnerability (UNISDR, 2013). Wherever homes and livelihoods are destroyed, and where it is recurrent or unresolved, serious political differences for prolonged periods, displacement follows. Forced from their homes, people face heightened protection risks with regard to family separation, as well as sexual and gender-based violence, which particularly affect women and children (Cernea, 1999). People displaced by disasters are thus among the most vulnerable populations (Cutter et al., 2003) whose only option to avoid even worse situations is to leave their homes to seek a new life (Vecere et al., 2017). This most often leads to

dependence on external assistance; thus in the long-term displaced persons are the ones most affected by disasters.

Public health analyses of refugee and IDP populations entail the study of the number, distribution, movement, age structure, birth and death rates of a particular group of people, *i.e.* those forced to live in camps. Overcrowding affects health both physically and mentally, undermining resistance and increasing the risk of contracting infections. Malnutrition is another constant and subtle threat leading to individuals being more prone to diseases; they often fall sick and, in the case of pregnant women, child mortality increases. People living in refugee camps also receive inadequate medical service, either due to inaccessibility, high price or lack of needed supplies making them vulnerable to disease (Haysom and el Sarraj, 2013). Certain groups of the study population (*i.e.* displaced people in camps) are the most vulnerable. These groups include children aged less than five years, childbearing mothers and people with increased risk of contracting disease and other adverse health conditions.

After the beginning of the Soviet war in Afghanistan in 1979, an estimated 5 million Afghan refugees fled to neighbouring Pakistan, a developing country with limited resources to accommodate this influx (Iram and Butt, 2004; Kronenfeld, 2008). There is now a disparity between different parts of the country, with most of the refugees and IDPs hosted in the least developed provinces, *e.g.*, Khyber-Pakhtunkhwa (Shahbaz, 2013; Ahmad *et al.*, 2017). In 2014 alone, Pakistan hosted approximately 1.6 million Afghan refugees (UNHCR, 2015), most of them residing in urban areas and refugee villages.

After more than thirty years of conflict, many Afghans have been born in Pakistan and assimilated into Pakistani communities (Kronenfeld, 2008). The unorganised exodus and return of refugees from and to Afghanistan makes it difficult to accurately characterise the population of refugees. Out of 1.6 million registered Afghan refugees, 1.01 million (63%) live in urban and rural areas, while 0.59 million (37%) are in 76 refugee villages in Khyber Pakhtunkhwa, Baluchistan and Punjab (UNHCR, 2015). Khyber-Pakhtunkhwa Province alone has received millions of refugees and IDPs fleeing areas of conflict stressing resources and capacity of this province beyond its limits. At present (August 2017), in addition to 1.6 million Afghan refugees, Khyber-Pakhtunkhwa accommodates more than 1 million IDPs from North Waziristan Agency, a mountainous region of north-western

Pakistan bordering Afghanistan (UNHCR, 2015). The refugee population has given rise to enormous pressure on service delivery related to health, education, livelihood and governance in general.

Disasters have historically been quantified with regard to direct loss of life and the capital stock depleted following the occurrence of natural events (Patankar and Patwardhan, 2016) or war. However, there is now a greater focus on the secondary effects of disasters that comes closer to capturing the critical component of livelihood in the disaster risk equation (Lim, 2016). The magnitude of displacement is related to the extent and frequency of extreme and non-extreme natural events (de Almeida, *et al.*, 2016). The total number of people displaced by such events, both in relative and absolute terms, provides an important quantitative measure of their underlying vulnerability (Adeagbo *et al.*, 2016). The distance of displacement, whether to another part of the same community or to a different country, is also an important factor contributing to the level of vulnerability and lack of resilience affecting affected communities (Bonanno *et al.*, 2010).

Due to the Afghan war, which went on for over three decades, Pakistan remained at the top of the list of countries hosting the highest number of refugees until 2014, when it was surpassed by Turkey due to the Syria crisis (UNHCR, 2015). According to the Internal Displacement Monitoring Center (IDMC) (http://www.internal-displacement.org/internal-displacement/), Pakistan has not only been host to the largest number of refugees, but also of the highest number of IDPs for the last few years (IDMC, 2014). These population displacements have resulted in long-term social, political and economic consequences (Nomura *et al.*, 2016). With a vulnerability score of 22 and a resilience score of 3.66, Pakistan ranks 2nd in South Asia's Disaster Displacement Risk Index (DDRI) (Ginnetti *et al.*, 2015). The DDRI, developed by the IDMC, predicts annual displacement for each country and each prevailing hazard type every year for the next ten years. The index assumes a scenario of *business as usual* and takes into account the actual intensity and frequency of the hazards of the past, as well as population growth and change in exposure and vulnerability with current rate (Ginnetti *et al.*, 2015). An overview for Pakistan in this respect is presented in Tables 1 and 2.

As is common practice around the world, the Pakistani Government allocates resources based on population size. We discovered that budgets for the various fundamental needs are generally allocated without taking the actual refugee population into

Table 1. Pakistan's future annual displacement risk (next 10 years) (Ginnetti *et al.*, 2015).

Population (n)	Average risk of annual displacement	Regional rank in risk of annual displacement	Relative annual displacement (per million)	Regional rank	Change
187,250,400	1,805,600	2	9,600	3	2.2%

Table 2. Pakistan's displacement risk profile (Ginnetti *et al.*, 2015).

| Relative physical exposure | Risk configuration index | | | Historic displacement | |
	Vulnerability	Resilience	Risk configuration	Average annual displacement 2013 - trend level*	Relative annual displacement (per million) 2013 - trend level*
95,112	22	3.6	0.58	1,566,000	8,400

*Trend level here means calculating the displacement estimates based on estimated past displacement using the best available disaster-loss data and calibrating it using the five best years of high-confidence estimates recorded in International Displacement Monitoring Centre's Disaster-induced Displacement Database (DiDD) (Ginnetti *et al.*, 2015).

account since refugees and IDPs are not properly captured by census statistics and documentation is often out of date. The last census was conducted in 1998 and with respect to Khyber-Pakhtunkhwa, displaced people were dispersed all over the province in the subsequent years. This means that refugees in districts with high refugee/host-community population ratios are in a particularly vulnerable situation. We propose a simple, innovative approach based on geographical information systems (GIS) visualising the magnitude and severity of these concerns and communicating the needs to public health professionals so that appropriate resources can be allocated according to the level of vulnerability in the target populations.

Materials and Methods

Although there are refugees in other Pakistani provinces, Khyber Pakhtunkhwa has the highest number of registered refugees, which made us choose this province for study. More specifically, we chose camps and refugee villages with a population of 5,000 or more refugees and IDPs, which are recognised by UNHCR and the Pakistani Government. We focused on camps with the highest populations and on those that have experienced a large influx of displaced populations in the past. Khyber Pakhtunkhwa Province currently hosts a total of 97 refugee/IDP camps, all of which are situated in 14 out of its 26 districts. The highest number of displaced people is in the district of Peshawar with 140,136 registered, displaced people (UNHCR, 2015) and the second highest is in the district of Haripur with 116,514.

Study site

Khyber Pakhtunkhwa is a province in north-western Pakistan comprising 26 districts. It is situated between latitudes 31°15' and 36°57' North and longitudes 69°5' and 74°7' East and located at the intersection of three mountain ranges: Himalaya, Hindukush, and Karakorum. Kashmir is to its Northeast and Punjab to its Southeast with Afghanistan and Federally Administered Tribal Areas (FATA) to its West; Afghanistan also borders the province in the North. The district of Peshawar has the highest population and harbours the provincial capital.

Data sources

The data, covering the period January 2012-December 2014 and related to refugee populations in Khyber Pakhtunkhwa, used in this study were provided by the organisations working with refugee/IDP camps in 14 districts: Peshawar, Nowshehra, Mardan, Haripur, Buner, Kohat, Dir, Mansehra, Swabi, Hangu, DI Khan, Chitral, Charsaddha and Lakki-Marwat. Many local and international organisations work with IDPs and refugees, but for reliability and authenticity, only data emanating from Commissionerate Afghan Refugees Khyber Pakhtunkhwa (KP) (http://kpkcar.org/carnewsite/CAR/) and Frontier Primary Health Care (FPHC) (http://www.unhcr.org/partners/ngodirectory/48fdec0928/frontier-primary-health-care.html). The study is thus based on secondary, retrospective datasets. The use of secondary data (according to Sorensen *et al.* (1996) defined as information not collected with a specific research purpose) in public health and epidemiological research is currently increasing (Boslaugh, 2007).

According to these data, the total registered refugee population in Khyber Pakhtunkhwa is about 0.75 million. However, due to population movement and government policies that often shift populations from one place to another, the numbers vary over time. To reach an actual picture of our study population, data from the three study years (2012-2014) were included and an average obtained. Accordingly, our calculations are based on an average of 741,013 people residing within the camps in Khyber-Pakhtunkhwa Province at the time of the study.

The population in the camps had a ratio of 50.2% males and 49.8% females, similar to that of the host communities. To attain a structure of the study population based on age, we divided the population into three classes: infants up to 1 year of age (3.9% of the total population, with roughly as many males as females); children aged 1-4 (13.39% of the total population, with 6.7% males and 6.6% females) and those older than 5, which included 82.6% of the study population (41.5% males and 41.1% females) as shown in Table 3.

Table 3. The study population: demographic profile 2012-2014.

		2012	2013	2014	Average
Total		752,899	728,618	741,523	741,013
Males		375,384 (49.9%)	366,568 (50.3%)	373,947 (50.4%)	50.2%
Females		372,083 (50.2%)	362,050 (49.7%)	367,576 (49.7%)	49.8%
0-1 years	Male	15,688 (2.08%)	14,478 (1.98%)	13,784 (1.85%)	2.0%
0-1 years	Female	15,242 (2.0%)	14,063 (1.9%)	13,531 (1.8%)	1.9%
>1-4 years	Male	50,045 (6.6%)	49,045 (6.7%)	50,753 (6.8%)	6.7%
>1-4 years	Female	49,057 (6.5%)	47,683 (6.5%)	49,215 (6.6%)	6.6%
≥5 years	Male	309,914 (41.2%)	303,046 (41.6%)	309,419 (41.7%)	41.5%
≥5 years	Female	308,486 (41.0%)	300,304 (41.2%)	304,827 (41.1%)	41.1%

Mapping

To generate a map of the study area, the GIS data set for this research was downloaded from two websites (http://pakresponse.info/, and http://www.diva-gis.org/) and processed in QGIS 2.16 open source software (http://qgis.org/api/). GIS makes it possible to prepare separate layers for each attribute dataset and thus allows display of layers representing different information separately or together (Sadia *et al.*, 2016). The descriptive data included the refugee camps including the name of the districts, numbers of host population, numbers of refugee population, numbers of refugee families, *etc.* The spatial data for Pakhtunkhwa Province included the longitude and latitude of the districts, topographic layers and the districts referred to by national administrative divisions. After compiling the datasets for each study period, we entered this information into the QGIS and built the layers needed adding North direction and legends.

Classification

We classified the study population according to gender and age because these variables are associated with different needs (Wijk-Sijbesma, 2016). For example, children require vaccinations to prevent certain diseases, schools for learning and family support for psycho-social health; pregnant women need proper antenatal and postnatal care to prevent mortality and morbidity; married couples need family planning, counselling and information regarding contraception (Sedgh *et al.*, 2016); while elderly people must have regular medication for diseases such as diabetes and hypertension.

Definitions

Displaced population (DP): The total number of displaced people settled in a district when from there from their home district, province or country.

Local population (LPV): The total number of people living in their home district.

Highest district displaced population (HDDP): the district that accommodates the highest number of displaced population among all 14 districts.

Displaced population families (DPF): The total number of displaced families settled in a district when moved there into a district from their home district, province or country.

While a family may consist of a large variation with regard to the number of members, it constitutes a single unit for some aid purposes. For example, in any emergency or humanitarian situation, assistance is typically provided based on family units, *e.g.*, one tent for one family, while the number of individuals in the family becomes relevant with regard to the distribution of other aid items such as food and certain non-food items (NFIs), *i.e.*, blankets, sleeping mats and plastic sheeting to safeguard them from harsh weather and other environmental conditions (IASC, 2007). To account for these different needs, we propose two different definitions of vulnerability: Unmet needs vulnerabilities (UNV) for intervention on a family-unit basis and absorption capacity vulnerability (ACV) for interventions on an individual basis. Taking these issues into consideration along with relative refugee vulnerability (RRV), we calculated the coefficient for four factors that influence the vulnerability of displaced population and host communities using the following formulas:

$$RRV=DP/HDDP \qquad \text{Formula 1}$$

$$ACV=DP/LPV \qquad \text{Formula 2}$$

$$UNV=DPF/HDDP \qquad \text{Formula 3}$$

In addition to these three formulas, a fourth formula defining the total vulnerability (TV) corresponding to the unweighted average, can be expressed as follows:

$$TV=\text{Unweighted average of all three vulnerabilities} \qquad \text{Formula 4}$$

According to IASC (2007) NFIs are divided into three broad categories to assist population displaced by any natural or man-made disasters. As there is always sacristy of resources during the time of emergency, conflict or disaster, development practitioners and planners need to prioritise their actions for supplying food items and NFIs to the most needed areas first. The Formulas 1 and 2 are proposed to help make such decisions of prioritisation.

A posting by Price (2017) entitled *Opinion: see the most vulnerable - see the human landscape* states *Understanding the extent and specifics of humanitarian need is often a significant unknown. The current food security and IDP crisis in South Sudan underscores the importance of using remote sensing approaches to know exactly where vulnerable populations are migrating, how many displaced persons are on the move, what sort of resources they have access to and what threats and/or risks are nearby. A comprehensive foundational data set is incredibly valuable for intervention planning, but also critical to feasible exit strategies.* In this regard, in order to allow policy-makers to better understand the amount of support needed in each area settled by refugees or IDPs, we propose using Formula 3.

Results

The calculated refugee population vulnerability at the district level is shown in Figure 1. This type of vulnerability mapping is proposed for strategies, *e.g.*, based on the *water, sanitation and hygiene* (WASH) approach as well as for educational and health services. It is common practice that schools and hospital in a community are established based on the number of people residing in that community. Due to the unforeseen influx of displaced people from other districts only a limited number of these *outsiders* can be accommodated. This type of vulnerability will communicate the level of assistance needed in areas hosting displaced populations. Figure 1 clearly conveys the importance of calculating the impact of this vulnerability. In addition, general developmental practitioners and policymaker may not consider the district of Peshawar as particularly vulnerable because it is the capital city with a main hospital as well as many other healthcare institutions, both private and public. However, as we calculated the real vulnerability impact, we came to know that it is the most vulnerable district of the whole province due to an unforeseen influx of refugees. This may be because many refugees prefer to stay in urban area as there are more economic opportunities there compared to the rural situation. Figure 1 makes this clear in the shortest possible time and if available during the planning stage, maps of this kind would not only get the attention of decision-makers, but should also facilitate effective planning.

The calculated ACV (Figure 2) is proposed to express the extent of human resource shortage in the refugee affected and hosting areas (RAHA). For example, the number of doctors or the
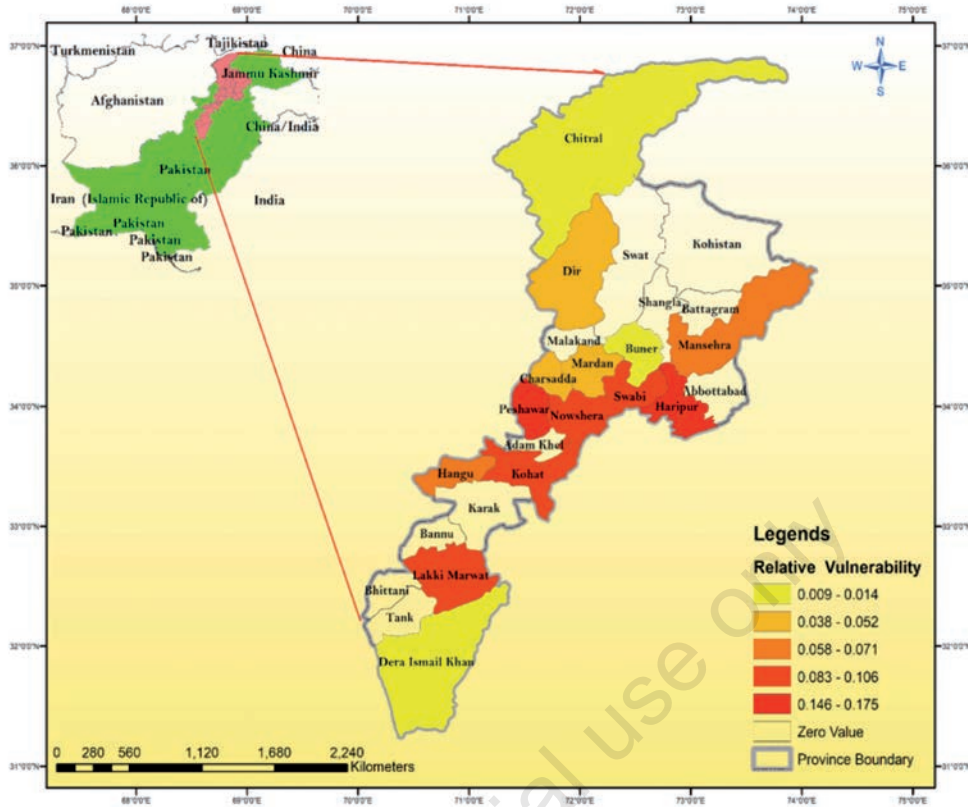
Figure 1. Calculated refugee population vulnerability at the district level.
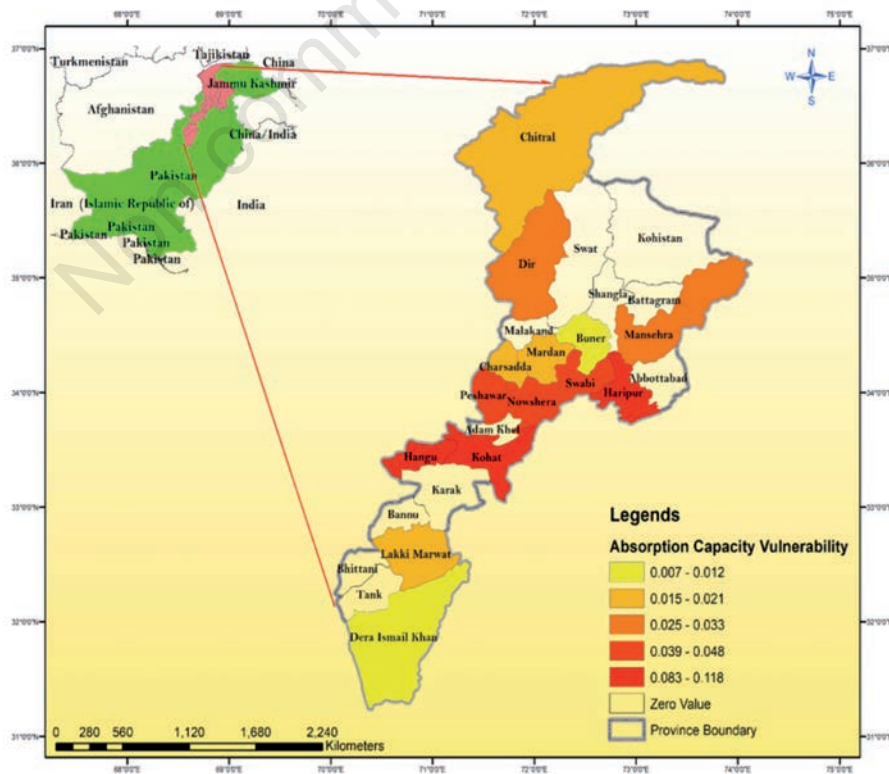

Figure 2. Calculated absorption capacity vulnerability at the district level.