

Manuscript Number: JTRG-D-15-00003R2

Title: Regional Transport and Its Association with Tuberculosis in the Shandong Province of China, 2009-2011

Article Type: Full Length Article

Keywords: Tuberculosis; transport networks; geospatial analysis; health geography; China

Corresponding Author: Prof. Poh-Chin Lai, PhD

Corresponding Author's Institution: The University of Hong Kong

First Author: Erjia Ge, PhD

Order of Authors: Erjia Ge, PhD; Poh-Chin Lai, PhD; Xiulei Zhang; Xiaohuan Yang; Xuezheng Li; Haiying Wang; Xiaolin Wei

Abstract: Human mobility has played a major role in the spread of infectious diseases such as tuberculosis (TB) through transportation; however, its pattern and mechanism have remained unclear. This study used transport networks as a proxy for human mobility to generate the spatial process of TB incidence. It examined the association between TB incidence and four types of transport networks at the provincial level: provincial roads, national roads, highways, and railways. Geographical information systems and geospatial analysis were used to examine the spatial distribution of 2,217 smear-positive TB cases reported between 2009 and 2011 in the Shandong province. The study involved factors such as population density and elevation difference in conjunction with the types of transport networks to predict the disease occurrence in space. It identified spatial clusters of TB incidence linked not only with transport networks of the regions but also differentiated by elevation. Our research findings provide evidence of targeting populous regions with well-connected transport networks for effective surveillance and control of TB transmission in Shandong.

## **Highlights**

- ◆ The emergence and transmission of tuberculosis (TB) is related to human movement.
- ◆ TB in Shandong showed significant spatial clustering at scales of 7 km and below.
- ◆ TB clusters were related to easy access to provincial/national roads but not rails.
- ◆ Transport effects on TB were different between low and high altitude regions.

## **Manuscript Title**

Regional transport and its association with tuberculosis in the Shandong province of China, 2009-2011.

## **Abstract**

Human mobility has played a major role in the spread of infectious diseases such as tuberculosis (TB) through transportation; however, its pattern and mechanism have remained unclear. This study used transport networks as a proxy for human mobility to generate the spatial process of TB incidence. It examined the association between TB incidence and four types of transport networks at the provincial level: provincial roads, national roads, highways, and railways. Geographical information systems and geospatial analysis were used to examine the spatial distribution of 2,217 smear-positive TB cases reported between 2009 and 2011 in the Shandong province. The study involved factors such as population density and elevation difference in conjunction with the types of transport networks to predict the disease occurrence in space. It identified spatial clusters of TB incidence linked not only with transport networks of the regions but also differentiated by elevation. Our research findings provide evidence of targeting populous regions with well-connected transport networks for effective surveillance and control of TB transmission in Shandong.

**Keywords:** transport effects, human mobility, tuberculosis, geospatial analysis, China.

## **1. Introduction**

There is compelling evidence that much of disease spread today is related to global movement of people, animals, and goods. It has also been observed that the most recent outbreak of Ebola and its intense transmission in West Africa are being monitored and tracked closely because there is real risk of new countries being affected (Gomes et al., 2014; Wesolowski et al., 2014). While migratory birds have contributed to the spread of avian influenza, it has been said that modern modes of transportation, especially air travel, are responsible for the unprecedented

volume and speed of cross-border and cross-continental transmission of diseases in the 21<sup>st</sup> century.

Hagerstrand's theory of diffusion (Ellegård & Svedin, 2012) is the basis for the formulation of many epidemic models. His time-geographic concepts link individuals in one or more places through movement in space and time. Diffusion of disease is examined by tracking where an infected individual has been and with whom the infected has been in contact. The outbreak of SARS in 2003 highlighted the importance of tracking a highly infectious index patient in Hong Kong that resulted in an acute outbreak almost went out of control. The 2009 Swine flu also painted a bleak picture of disease transmission and diffusion. But the tracking of individuals' space-time movement is a daunting task (Kwan, 2000; Chen et al., 2011). Even with today's technological advances in following people's small-scale movements and activity space, scientists are baffled by the immense detail and volume of data and how to make sense of the placing and patterning of human activities.

### ***1.1. The notion of distance decay***

Disease diffusion concerns the spread of a disease from its source to new locations and the pattern of diffusion is affected by barriers such as time, distance, physical, and cultural factors (Bossak and Welford, 2015). It is well documented that the farther away from the source of a disease, the more time it takes to feel the impact. This friction of distance is what geographers refer to as the distance decay effect. The likelihood of disease spread can be explained by the mechanisms of expansion and relocation (deBlij & Murphy, 2003). Expansion diffusion is said to occur when the number of infected individuals in an area grows continuously larger in space and time. Expansion can occur through an established structure, also known as hierarchical diffusion, or through a group of people or an area, also known as contagious diffusion (Cliff et al., 1981; Meade et al., 2000). Relocation is a sequential diffusion process whereby an infection is transmitted through movement of its carrier agents or a migrating population (Cliff and Haggett, 2004; Martens and Hall, 2000; Stoddard et al., 2009). The distance decay function in geographic profiling (the circle hypothesis and the distance decay theory) of disease investigation is highly influenced by human activities and environmental attributes.

As the impact of human activities spread more widely through global transportation networks, disease can be transported via many different processes and pathways. However, the processes generally result in some degree of distance decay.

The fact that disease emergence decays with distance from population centers has been reported time and time again (Fotheringham & Rogerson, 1993; Xia, Bjørnstad & Grenfell, 2004). With new evidence highlighting that disease spread tended to occur faster along established transportation routes (such as major roads, waterways, and coastlines) (Kausrud et al., 2010; Wen et al., 2012), it is more certain that both environmental factors and population movements play important roles in disease transmission. Although the modeling of complex spatial interaction of disease phenomena is far from being perfect, understanding the role of disease spread along these networks and the travel patterns would allow for better identification of distance jumps or rate of disease infection. For example, Balcan et al. (2009) found that the effect of short-range commuting flows was larger than that of long-range airline flows and that the epidemic behavior would be different due to multiscale mobility processes in the disease dynamics. The ability to estimate how fast or broad an infection would spread under different conditions could help public health officials refine disease control or intervention measures.

## **2. Tuberculosis**

*Mycobacterium tuberculosis* (TB) is a highly contagious bacterium that spreads from person to person via aerosols (Dye, 2006; Jones-López et al., 2013). The worldwide spread of TB has been a continuous threat to public health globally with almost one third of the world's population infected (Dye, 2006). The World Health Organization (WHO, 2012) estimated that the world had over 8.7 million new TB infection and 1 million TB-related deaths in 2012.

The transmission of TB is highly complex and dynamic, and varying spatially (Bryant et al., 2013; Tessema et al., 2013). It has been asserted that travelling by public transports has a role in the emergence and spread of infectious diseases such as TB (Wilson, 1995; Barnett and Walker, 2008). The importance of TB contact locations such as community drop-in centers, bars, and parks, and social networks

have been explored to great effect (Cook et al., 2007; Carter et al., 2009). Indeed, the movements of people in space and migrants, in particular, have contributed to the spread of TB in both developing (Long et al., 2008; Wei et al., 2009; Pace-Asciak et al., 2013) and developed countries (Haase et al., 2007; McPherson et al., 2008; Franzetti et al., 2010; Edelson and Phipers, 2011). An earlier study by Jia et al. (2008) reported a higher rate of TB prevalence in migrants than the local residents of Beijing. Long et al. (2008) also found factors such as long working hours, poor living and working environment, as well as insufficient healthcare support and medical insurance to be probable causes of higher TB prevalence among the migrants. However, findings from these analyses that are confined to sufficiently small population groups and geographic areas might not be appropriate for larger regions with more complex and dynamic patterns of human mobility. For example, China has a large number of labor workers traversing long distances regularly between urban cities and rural areas known to have a high prevalence of TB (CTCC, 2004). These migrant workers not only have increased the likelihood of TB spreading across wide geographic scales but also promulgated the disease to further propagate by local movements through public transports (Edelson and Phipers, 2011).

Wu et al. (2006) defined human mobility as a network of interacting communities where the connection and corresponding intensity represent the flow of people among them. But tracking massive population movement among infected localities is impractical, if not impossible. In a recent study, Balcan and colleagues (Balcan et al., 2009) applied long-range intercontinental airlines and short-range commuting flows to simulate the global process of human mobility and its impact on the spread of influenza-like diseases over multiple geographic scales. The authors acknowledged that the spread of infectious diseases is caused by a strong heterogeneity of transport networks.

Altitude (or elevation) has been applied as a significant predictor in a number of studies of infectious diseases, including TB (Saito et al., 2006) and avian influenza (Gilbert et al., 2008; Ge et al., 2012). In particular, altitude has a protective effect against TB in Peru in which lower disease prevalence and increased household clustering of TB was found in high altitude villages compared to those at sea level

(Saito et al., 2006). In this study, we included altitude to capture spatial variation in topography as well as to serve as a surrogate indicator of other unmeasured variables related to TB infection.

This paper makes use of transport networks to explore the association between human mobility and TB occurrences across multiple geographic scales. The primary objective is to study the association modeled on transport networks to address the following issues: (1) spatial patterns of TB incidents and disease clusters or hot spots, (2) degrees of spatial association between TB clusters and transport networks, and (3) types of spatial association differentiated by two elevation groups of high versus low altitude.

### **3. Data & Methods**

This was a retrospective study to test the null hypothesis of no association between TB clusters and transport networks differentiated into two separate regions of low and high elevation. We postulated that TB cases are spatially related to transport networks whereby well-connected roads facilitate peoples' movements. We also postulated that provincial roads in zones of low altitude had a higher association with people movement because of their greater accessibility to travelers.

**3.1. Study Area.** The study area was the Shandong province (eastern China) (Figure 1), the second most populous province of China in 2010. It has a population of 95 million with 45.6 million living in urban areas (SPSB, 2012). The province is a prosperous and relatively rich region with a strong agricultural and industrial foundation, although an imbalanced urban and rural economic development remains a problem. Around 7 million migrants lived in the province in 2010, and most of them were scattered in the developed coastal cities of northeast Shandong, e.g., Weihai, Dongying, Qindao, and Jinan (SPSB, 2007a). It was estimated that 60% of migrants were sub-regional travelers traversing between urban and rural areas within the province (SPSB, 2007b).

--- Insert Figure 1 ---

**3.2. Data Sources.** Three kinds of data were used in the study: health outcomes (TB cases), geographic features (transport networks and terrain elevation), and demographic data (population density). All data were converted into 1x1 square kilometer lattice. This spatial resolution was chosen because it is a broadly accepted spatial unit for ecological studies (Ge et al., 2012; Qiu et al., 2014). Moreover, research have shown that an individual's activity space varies between 800 and 1000 meters and that a finer spatial resolution may cause structural clutter that affects decision-making (Lai et al., 2015). However, social science and public health studies are faced with current limitations of neighborhood level measures. Neighborhood or spatial scales may involve shifting political/census boundaries or changing lattice size which have been found to affect clustering patterns. Wieczorek et al. (2012) suggested that an exploratory analysis should utilize more than one neighborhood scales to provide an improved understanding of spatial clustering across the entire region. For this reason, we also computed data at 2x2, 5x5, 10x10 square kilometers to examine the effects of lattice size on spatial clustering.

A total of 27,622 TB cases from 2009 to 2011 were obtained from the Shandong Laboratory Surveillance System and the Shandong Provincial TB Reference Laboratory. These case data were sampled from 19 counties and 3 districts among the 19 million population in the province (Table 1). Each record of TB cases contained the following inputs: patient identifier, age, gender, and residential address (complete with domestic names of villages in rural areas or the exact flat with street names and numbers in urban areas). 2,217 of the TB cases were confirmed as smear-positive<sup>1</sup> pulmonary TB according to the World Health Organization (WHO, 2013). We included all smear-positive TB cases in this study and geocoded their residential addresses by matching each nominal address (complete with five to six hierarchical administrative district names) against the gazetteer records and determining its longitude and latitude coordinates through the Google geocoding service (Figure 1). This process of geocoding was similar to that applied in previous studies (see Gilbert et al., 2014) and all cases were geocoded successfully based on the nominal addresses. The TB cases were then aggregated by various spatial units (1x1, 2x5, 5x5, 10x10 square kilometers). Because data aggregation may result in systematic bias or ecological fallacy, examining

clustering results by multiple spatial scales can shed light on the robustness of the analysis method.

Table 1. Positive TB cases by counties/districts of the Shandong province, 2009-2011\*

City	County/District	Population (1,000 persons)	2009		2010		2011	
			culture +	culture -	culture +	culture -	culture +	culture -
Linyi	Cangshan	1,162	417	297	343	337	277	408
	Tancheng	838	343	259	297	265	162	352
	Linshu	617	213	161	212	257	156	204
	Yishui	998	386	284	344	314	213	320
	Fei	757	281	234	134	294	124	209
Jinan	Chuqing	579	106	136	135	95	106	136
	Licheng	1,124	88	104	183	165	88	104
	Laiyang	879	134	467	167	389	162	318
Yantai	Yantai	830	142	255	179	235	120	231
Heze	Laizhou	884	149	117	120	99	89	139
	Dongming	711	146	84	136	115	78	135
	Yuncheng	1,041	136	199	386	173	202	172
	Sishui	536	252	49	255	39	176	80
Jining	Yanzhou	534	199	80	191	60	105	101
	Zhoucheng	1,117	329	198	330	203	306	214
Binzhou	Zhouping	779	261	102	225	78	157	128
Dezhou	Decheng	419	227	64	235	81	238	67
Weifang	Linju	834	215	39	198	79	210	88
Qingdao	Jiaonan	868	190	209	186	192	109	226
Tai'an	Xintai	1,316	554	151	552	180	310	189
Zaozhuang	Tanzhou	1,604	445	226	270	198	227	267
Liaocheng	Xin	959	474	188	475	263	476	189
A total of 27,622 positive TB cases			5,687	3,903	5,553	4,111	4,091	4,277

\* The data were collected and compiled from the Shandong Laboratory Surveillance System and Shandong Provincial TB Reference Laboratory.

Transport network data (in GIS line-based format) were obtained from the National Data Sharing Infrastructure of Earth System Science (Geodata Center, 2013). The data included 244,500 kilometers of roads comprising of highways, national, and provincial roads with a road density of 156 kilometer per hundred square kilometers in 2012 (STD, 2013). These well-connected transport networks, which covered all counties in the province, were used as a surrogate measure for population mobility because land transports accounted for the majority of travels within the province. The shortest distances from a lattice cell to each type of roads<sup>2</sup> and railroads<sup>3</sup>, as well as their respective densities over a range of regions (at 2x2, 5x5, 10x10 square kilometers) were computed<sup>4</sup>. The network density was considered indicative of the volume of traffic flow (Thamizh Arasan and Dhivya, 2009) especially when local traffic data were not available in rural areas of the province.

Altitude data showing topographic variation were derived from the 90-meter digital elevation model obtained from the Shuttle Radar Topography Mission provided by the CGIAR Consortium for Spatial Information. We applied the Gaussian mixture model and the expectation maximization algorithm (Dempster et al., 1977) to delineate regions of low (between -105 and 81 with 63 meters on average) and high altitude (over 81 with 181 meters on average). We used population density as an indicator of the risk of bacterial infection. The data on population density were estimated using geospatial technologies (remote sensing and GIS) based on the most updated Chinese census and economic data along with land use and land cover data (Yang et al., 2009).

**3.3. Data Analysis.** The data analysis for assessing possible associations between TB and transport networks were conducted in three stages: (1) examining the spatial patterns of TB cases, (2) linking TB clusters and transport networks, and (3) establishing association differentiated by high and low altitude regions.

*3.3.1. Stage 1.* We used point features to represent TB cases that can be regarded as a spatial point process (Ripley, 1977). The Ripley's K function, proposed for the identification of spatial point patterns, has been used in public health studies (Dixon, 2006; Wheeler, 2007; Delmelle et al., 2011). We examined the patterns of

TB cases using the Ripley's K-function statistics at a county level of 16 kilometers and below. The Monte Carlo simulation of 999 random replicates was performed to generate an envelope representing complete spatial randomness at the 99% confidence level.

We examined the local patterns of TB clusters (i.e. hot spots) using a modified local  $K$ -function statistics<sup>5</sup>. The local  $K$ -function statistics, proposed by Getis (1984), has been used to identify spatial patterns and epidemic spread of infectious diseases (Vazquez-Prokopec et al., 2010). In this study, we modified the local  $K$ -function statistics to estimate the risk of TB clusters after adjusting for the spatial auto-correlation of TB cases. We weighted the number of cases regarding distance and formulated the weighted  $K$ -function statistics as:

$$\hat{K}_i(h) = \frac{\sum_{j=1}^n w_{ij} I(d_{ij} < h)}{\lambda c_i} \quad (1)$$

where  $\lambda$  is TB density and  $d_{ij}$  was the distance between lattice points  $i$  and  $j$ . An indicator function  $I$  was defined as 1 for  $d_{ij} \leq h$  and 0 for  $d_{ij} > h$ , which indicated whether or not TB cases surrounding a lattice point  $i$  were within the range of distance  $h$ . Each case was weighted by an inverse Euclidean distance,  $w_{ij} = 1/d_{ij}$ . We applied the modified local  $K$ -function statistics to estimate the risks of clusters at the scales of 2, 5, and 10 kilometers, respectively, and defined the risk as the larger the estimates the intensive the cluster, and therefore the higher the risk to spreading. We standardized and mapped the modified  $K$ -function results over a continuous range (0, 1).

3.3.2. *Stage 2.* We used bivariate spatial association analysis or spatial correspondence analysis (Haining, 1991) to measure relations between the intensity of TB clusters and their distances to the four types of transport infrastructures (i.e. provincial roads, national roads, highways, and railways) at the same scales as stage 1. The method is a novel application of a modified T-test with its sample size  $n$  adjusted for correlations of spatial data like clusters of TB cases to test for statistical significance of the  $R$  correlation coefficient expressed as:

$$R = \left( \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) \right) / \left( \sum_{i=1}^n (y_i - \bar{y})^2 (x_i - \bar{x})^2 \right)^{1/2} \quad (2)$$

where  $\bar{y}$  and  $\bar{x}$  were the means of TB cluster intensity and transport infrastructures including densities of and distances to the four types of traffic networks. The  $n$  is sample size, i.e., the number of lattices within the study area. A reduction in the degrees of freedom was a function of the level of spatial autocorrelation between the two raster maps that were correlated. The purpose of the adjustment procedure was to identify the “equivalent” number of independent observations (Haining, 2003). As the mathematics went beyond our study, we did not go into details of the formula in this part.

3.3.3. *Stage 3.* We examined and compared the associations between the occurrence of TB cases and the four types of transport infrastructures between the low altitude and high altitude regions. The autologistic regression model (Besag, 1972) was employed to formulate the associations after adjustment of spatial dependence of variables and their covariates by adding an additional term (the autocovariate) as shown in the equations below.

$$\begin{aligned} \log \frac{p_i}{1 - p_i} = & \alpha + \beta_1 alt_i + \beta_2 pop_i + \beta_3 D2ProRd_i + \beta_4 D2NatRd_i \\ & + \beta_5 D2Highway_i + \beta_6 D2Railway_i + \beta_7 DenProRd_i \quad (3) \\ & + \beta_8 DenNatRd_i + \beta_9 DenHighway_i + \beta_{10} DenRailway_i \\ & + \beta_{11} autocov_i \end{aligned}$$

where

$$autocov_i = \frac{\sum_{j=1}^{k_i} w_{ij} p_j}{\sum_{j=1}^{k_i} w_{ij}} \quad (4)$$

was the autocovariate term that described the average probability of TB occurrence weighted by cases in neighbors  $k$  surrounding the lattice  $i$ . The weight  $w_{ij}$  was equal to the inverse of the Euclidean distance  $h_{ij}$  between lattices  $i$  and  $j$ . In Equation (3),  $p_i$  was probability of TB occurrences, and  $D2ProRd_i$ ,  $D2Nat_i$ ,  $D2Highway_i$ , and  $D2Railway_i$  were the respective shortest distances from lattice  $i$  to provincial roads, national roads, highways, and railways. The  $DenProRd_i$ ,  $DenNat_i$ ,  $DenHighway_i$ ,

and  $DenRailway_i$  were the respective network densities at lattice  $i$ . The variables of  $alt_i$  and  $pop_i$  were elevation and population density, respectively.

This was a nested case-control study. The Nagelkerke/Cragg and Uhler's pseudo- $R^2$  (Nagelkerke, 1991) were used as goodness-of-fit statistic for the auto-logistic regression models. In addition, the Cohen's Kappa index (Carletta, 1996) was used to predict the probability of TB occurrence pertaining to transport and population characteristics of the low and high altitude regions. The receiver-operating characteristics (ROC) curve provided a two-dimensional description of the predictive performance (Fawcett, 2006). The area under the ROC curve (AUC) is a measure of the probability of correct classification.

The study used Matlab 2010b to process matrices of all of spatial lattices (i.e. TB cases, altitude, population density, and the four types of transport infrastructures) and applied ArcGIS 9.3 to map the predictive results.

## 4. Results

**4.1. Exploring spatial patterns of TB cases.** The  $K$ -function analysis (Figure 2a) showed that the overall patterns of TB cases varied from clustered to dispersed with changing distance scales from 0 to 16 kilometers. This result indicated that TB cases were spatially heterogeneous with significant clustering (confidence intervals CI: 99.9%) at a scale of around 7 kilometers but became increasingly dispersed thereafter. The highest level of TB clustering occurred at about 4-5 kilometers.

--- Insert Figure 2 ---

Based upon results of the  $K$ -function analysis (Figure 2a), we modified the local  $K$  function and applied the method to map the intensity of TB clusters at a scale of 5x5 square kilometer (Figure 2b). We standardized the analysis values to range between 0 and 1 to represent weak to strong clustering effects for TB risks. A large number of TB clusters were identified near the county boundaries. The blotchy pattern suggested that risks of TB were unevenly distributed in geography and tended to form isolated clusters of 5 kilometers or smaller.

**4.2. Linking TB clusters and transport networks.** Table 2 shows that TB clusters were mostly negatively related to the shortest distance to transport infrastructures. The provincial roads exhibited the shortest and statistically significant association ( $p = 0.025, 0.049, 0.051$ ) over multiple lattice scales of 2, 5, and 10 kilometers, implying that TB cases tended to cluster in areas along provincial roads. TB's association with the remaining transport infrastructures, however, did not show very consistent or statistically significant observation probably because our results were clouded by the presence of intermittent green belts (Madders and Lawrence, 1985).

The positive association between TB clusters and network density for both provincial and national roads was found statistically significant ( $p < 0.05$ ). Indeed, the association with provincial roads became stronger ( $R = 0.127, 0.207, 0.237$ ) with increasing distances from 2 to 10 kilometers. TB clusters were also found positively associated with national roads albeit at the lesser geographic scales of 2 and 5 kilometers. According to results of the spatial correspondence analysis ( $p > 0.1$ , Table 2), both highway and railway densities did not exhibit any significant association with TB clusters.

Table 2. Spatial correspondence analysis between TB clusters and transport infrastructures at different geographic scales.

TB clusters <sup>a</sup>	Transport Infrastructure <sup>b</sup>	Assessment of Bivariate Spatial Association					
		R	n	Adjusted N	t-test (1-tailed)	df.	p-value
2km x 2km	Provincial Roads	-.069	47926	365	1.96	363	.025
	National Roads	-.040	47926	198	.553	196	>.10
	Highway	.007	47926	234	.106	232	>.10
	Railway	-.037	47926	223	.547	221	>.10
5km x 5km	Provincial Roads	-.109	47926	175	1.65	173	.049
	National Roads	-.071	47926	84	.641	82	>.10
	Highway	.015	47926	103	.148	101	>.10
	Railway	-.063	47926	97	.613	95	>.10
10km x 10km	Provincial Roads	-.137	47926	110	1.63	108	.051
	National Roads	.106	47926	46	.705	44	>.10
	Highway	.016	47926	58	.118	56	>.10
	Railway	.094	47926	54	.680	52	>.10
TB clusters	Network Density <sup>c</sup>						
2km x 2km	Provincial Roads	.127	47926	611	3.15	609	.000
	National Roads	.074	47926	630	1.86	628	.035
	Highway	-.011	47926	542	.260	540	>.10
	Railway	.025	47926	548	.584	546	>.10
5km x 5km	Provincial Roads	.207	47926	188	2.88	186	.001
	National Roads	.104	47926	194	1.64	192	.050
	Highway	-.014	47926	162	.175	160	>.10
	Railway	.064	47926	171	.838	169	>.10
10km x 10km	Provincial Roads	.237	47926	78	2.12	76	.015
	National Roads	.103	47926	81	.917	79	>.10
	Highway	.018	47926	68	.143	66	>.10
	Railway	.101	47926	73	.859	71	>.10

The spatial correspondence analysis is used for testing null hypotheses that there is no spatial association between TB clusters and each of the four types of transport infrastructures. The above results quantify spatial association between TB clusters and various transport infrastructures at the geographic scales of 2, 5, and 10 kilometers. R = Pearson's correlation coefficient; n = sample size; Adjusted N = adjusted sample size.

<sup>a</sup> TB clusters were estimated by the weighted local *K*-Function analysis at the geographic scale of 2, 5, and 10 kilometers;

<sup>b</sup> The shortest distance to various transport infrastructures; and

<sup>c</sup> Number of transport infrastructures per lattice.

#### ***4.3. Establishing associations differentiated by low and high altitude regions.***

The histograms and boxplots in Figure 3 showed a relatively normal distribution of elevation levels for the low altitude region but extremely varied elevations within the region of high altitude. Transport network was comparatively denser in low altitude than the more hilly areas on higher grounds. The *K*-function analyses indicated that TB cases in the low altitude region clustered at a distance of around 15 kilometers and below but those in the high altitude region displayed more compact spatial clustering of 10 kilometers and below.

--- Insert Figure 3 ---

Results of the auto-logistic regression indicated consistent associations between various factors and TB occurrences between two regions of different elevation, with the exception of the shortest distance to railway (D2RW) and its network densities (DRW) (Table 3). Population density (Pop) was slightly attenuated but showed statistically significant positive association with the occurrences of TB in both regions but more so for the high altitude region. This was probably due to the uneven distribution of population in the hilly terrain of high altitude areas. The clusters of TB cases were found to have negative association with altitude, but only statistically significant ( $\beta = -0.006, p < 0.000$ ) in high altitude region.

Table 3. Summary results of the auto-logistic regression analysis for TB incidents in Shandong province, 2009-2011. The results are differentiated by two regions of low and high altitude.

Model Summary												
	Con	Pop	Alt	D2Pd	D2Nd	D2HW	D2RW	DPd	DNd	DHW	DRW	AutoC
<b>Low Altitude</b>												
$\beta$	-2.22**	.000**	-.003	-.047*	-.018*	.022**	.020**	1.58*	.765	-1.91*	-.377	.075
<i>p</i> -value	.000	.000	.480	.003	.002	.000	.000	.006	.319	.050	.000	.659
<b>High Altitude</b>												
$\beta$	-2.14**	.001**	-.006**	-.043	-.020*	.014	-.001	1.74*	1.47	-2.39*	.907	1.11
<i>p</i> -value	.000	.000	.000	.071	.004	.012	.733	.022	.306	.023	.519	.219
Model Assessment												
	Low Altitude						High Altitude					
AUC±SD	.761±.002						.761±.005					
Kapp±SD	.133±.004						.137±.006					
Pseudo-R <sup>2</sup> ±SD	.157±.010						.172±.014					

These results are the averages of 1000 bootstrap replicates of the auto-logistic regression model. Con: constant of the regression models; Pop: population density; Alt: average altitude; D2Pd, D2Nd, D2HW, and D2RW: minimum distances to provincial road, national road, highway, and railway, highway, and railway; DPd, DNd, DHW, and DRW: traffic densities for provincial road, national road, highway, and railway based on 2km x 2km lattice; AutoC: spatial autocovariant; AUC: Area under curve measure on correct classification; Kapp: Cohen's index on misclassification; Pseudo-R<sup>2</sup>: Nagelkerke/Cragg & Uhler's Pseudo-R2 goodness-of-fit measure. \*\* *p*-value < 0.001; \* *p*-value < 0.05.

Results of the auto-logistic regression analysis indicated that the occurrence of TB in the low altitude region was negatively associated with the shortest distance to provincial ( $\beta = -0.047$ ,  $p < .003$ ) and national roads ( $\beta = -0.018$ ,  $p < .002$ ), but positively associated with highways and railways (Table 3). In addition, we also found that TB occurrences were positively related to provincial road density, but negatively associated with both highway density and railway density (Table 3).

Contrary to the low altitude region, transport networks in the high altitude region were less dense particularly in the more hilly areas. The occurrence of TB was found to be negatively associated with the shortest distance to national roads ( $\beta = -0.02$ ,  $p < 0.004$ ), but positively associated with toll highways (Table 3). The significance of the regression models and the model fit were assessed statistically using AUC, Kappa, and Pseudo  $R^2$  (Table 3). These statistics indicated fairly good performance of the two regression models. The ROC curve was applied to measure the probability of correct classification or the relative fit of the two models, as shown in Figure 4.

--- Insert Figure 4 ---

Figure 5 is a probability map derived from the auto-logistic regression models for the risk of TB occurrences. It shows that TB was more likely to occur in the proximity of provincial and national roads, and around their major junctions. It also indicates that TB trended towards areas of low elevation with a denser population, particularly in the high altitude region, e.g. hilly areas in the central province.

--- Insert Figure 5 ---

## 5. Discussion

Public transports have been observed to associate with TB transmission (CDC, 1995; Kenyon, 1996). Compared with air transports, passengers had nearly twice the risk of being infected when exposed to infectious passengers in mass public transports (Mohr et al., 2012). The current study is one of only few studies on the quantitative estimates of the risk of TB transmission by local and regional transport infrastructures. The results indicated that human mobility could have contributed to

the varying spatial patterns of TB distribution showing significant clustering of 7 kilometers and below against major transport infrastructures (Figure 2a). Short-distance accessibilities seemed most likely to associate with the between or within city transmissions of TB in the province. The isolated clusters (hot spots) identified along provincial roads and county boundaries (Figure 2b) implied that TB transmission could be a process of expansion or hierarchical diffusion through traffic networks. Similar transmission processes have been reported in other infectious diseases (Wilson, 1995; Balcan et al., 2009).

The occurrence of TB was found to be significantly associated with the inverse shortest distances to provincial and national roads in both the low and high altitude regions (Table 2). Moreover, the distance decay association was more pronounced and significant in the low altitude region. As provincial and national roads are often well connected and do not incur any user fee, they brought not only convenience to short-distance travelers but also increased TB diffusion locally. Having adjusted for potential confounders, i.e., population density and terrain variation within each region, we found that TB was more prevalent in areas of lower elevation within the high altitude region (Table 3). Previous studies have indicated that provincial roads played a major commuting role between cities and undeveloped villages in the lower plains of hilly regions, such as the central part of Shandong (Mao, 1996). Regional disparity in economic development has also encouraged population migration from rural to urban areas, where a previous study indicated over 60% of migrants worked in manufacturing workshops situated near provincial and national roads (Zhou, 2011). These short-distance transport links were likely pathways of TB diffusion, particularly among migrants from rural suburbs where the disease was usually prevalent and its incidence rates relatively high (Wei et al., 2009). We applied the shortest distance to transport infrastructures and their respective network densities as proxies to account for migrant association with TB occurrences. Although the present study was not powered to quantify the association due to absence of available data on migrants, e.g., travel duration and frequency per year, it did shed lights on the impacts of regional transports in the diffusion process. Our results of the spatial diffusion patterns of TB by short-distance transports are consistent with those of an epidemiological study conducted in South Africa (Andrews et al., 2013).

Highways and railways in China are usually built in the outskirts of a town and require users to pay at tollgates and stations. They are mainly used for long-haul or inter-regional travels. Most migrants are low-wage workers with little societal insurance who cannot afford the high cost of daily commute. The significant association between short-distance travels and TB occurrences in the both low and high altitude regions implied that TB was more persistent at the neighborhood scales. In other words, the disease was less likely transmitted from outside the province through long-haul transports along highways or railways. Such a pattern of TB diffusion and associations with regional transport networks in the province may not be consistent with observations in other parts of the world which are highly complex with confounding influences from society, culture, geography, human behaviors, and lifestyles (Horna-Campos et al., 2007).

Our findings about the associations between TB disease and provincial/national roads have implications on air pollution and its health consequences. As public health is affected by noise and air pollution caused by road traffics (Jephcote and Chen, 2012), more thoughts should be given to development density and mix of transportation in the planning process. Even though highways and railways may be associated with higher traffic flows and increased vehicular-related air pollution, provincial and national roads that do not separate themselves from communities by green belts (Madders and Lawrence, 1985) can exert more traffic-related health impacts on the residents (Kim et al., 2012).

There are a few limitations in this study. Our data lack personal information on patient residency status (whether migrants or permanent residents), occupation, smoking status, concurrent chronic respiratory diseases, and social economic status. The absence of residency status may lead to an underestimation of the contribution of short-distance accessibilities to TB diffusion, although the associations of provincial and national roads were attenuated but statistically significant. The absence of other data have prevented us to control for a variety of potentially confounding effects of socio-economic and health factors. However, we may include data on genotypes of TB cases in future studies to uncover bacterial evolution and its spatial spreads associated with human mobility, which will

augment our understanding of the TB diffusion process and the dynamics of microbial evolution.

## **6. Conclusion**

It is crucial for public health administrators to develop an understanding of people movement, especially in areas infected with TB but with limited access to healthcare facilities, so that they can plan for appropriate measures to carry out adaptive prevention measures. This study highlights important associations between TB and human mobility in two regions of different altitudinal variation. In the low altitude region, national roads and provincial roads might have facilitated regional and sub-regional transmissions of TB through short-distance travels and commuting among neighborhoods. In the high altitude region, TB tended to occur in areas of lower elevation with well-connected roads. Toll highways and railways seemed to have minimal effects on TB transmission irrespective of elevation differences. These results provide valuable evidence suggesting not only migrant workers but also the roles played by cheaper means of transportation in TB transmission.

The above understanding of the where and why TB occurs at a certain geographic location can assist the Chinese government to rethink public health and transport strategies for the control of TB transmission that considers human mobility. As the most effective way of TB control is to identify persons who have active TB and to screen persons who had contact with TB patients to determine whether they are infected, this study suggests that major transportation hubs (or disease hot spots such as bus and railway terminals) may be enlisted to target high-risk populations. For instance, promotion of TB control programs to help educate the public (e.g., through posters and information pamphlets) and to provide guidance and assistance for local screening and prevention services (e.g., establishing local screening and treatment facilities) can encourage awareness and prevent progression of detected cases to active TB. The government should also engage active consultation and oversight for the TB control programs and collaborate with local health and transport authorities to ensure successful implementation of these prevention measures.

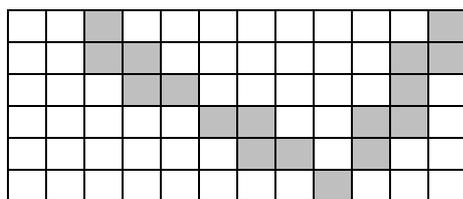
## 7. Footnotes

- <sup>1</sup> Smear-positive is based on the presence of at least one acid fast bacillus (AFB+) in at least one sputum sample. A patient with one or more initial sputum smear examinations positive for AFB is also considered to be a “definite” case, provided that there is a functional external quality assurance system with blind rechecking (WHO 2013).
- <sup>2</sup> China roads can be classed into five general types: highways, national roads (guodao 國道), provincial roads (shengdao 省道), rural/town roads (xiandao 縣道), and urban roads (xiangdao 鄉道). Highways are usually built in urban fringe areas connecting major cities or districts in different provinces or regions. These intercity freeways are largely built in the recent two decades and have dual carriageways with two or more lanes on each carriageway. The lanes are wider in width for better driving safety of high-speed vehicles. National roads comprise large transport networks that cover most populated areas, including urban and rural locations of the country. National roads are built with solid quality and large load-carrying capacity for heavy traffic flows. Compared to highways and national roads, provincial roads are roads with a narrower lane width, less traffic flows, and more localized coverage. Provincial roads have easier and more direct access to rural/town roads in counties or townships and urban roads in cities. Most provincial and national roads do not impose charges in use but highways are tolled trunk roads across the country. This study focused on three types of roads, namely highways, national, and provincial roads.
- <sup>3</sup> Railroads in China may be differentiated by passenger or freight depending on the time of construction. Most railroads for passengers built in recent decades are for high-speed trains. This form of transportation network is well connected for regional travel to provide linkage with larger cities like Shanghai, Beijing, Tianjin, Wuhan, Guangzhou, and Shenzhen. They are constructed to good quality standards to ensure travel safety at high speeds of over 200 kilometers per hour. Railroads for freight transport goods (such as

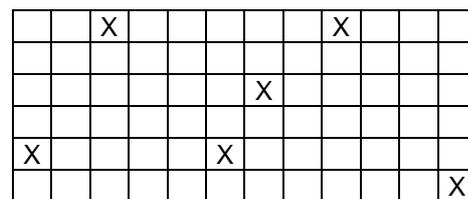
coal, metal, food, and wood) from source areas (mostly villages, hilly areas, and remote suburban locations where population is relatively sparse) to manufactories and markets. Many railroads built in the last 50 years are for shared use by passengers and freights; however, freight rail spurs are often routed into small villages and suburban areas. The present study included all railroads.

Railroads for transporting goods do not have the same potential to spread TB as passenger railroads because of locational and distributional differences in the networks and stations. Most railroads for passengers provide regional transportation linking major cities or counties of China where population densities are usually high. Railroads for transporting goods connect rural villages, townships, and source or market areas of raw materials which are situated away from urban areas and cities with large population. As a result, passenger railways in comparison to those for transporting goods are more likely to be hosts of TB transmission, although the two kinds of railroads overlap in some locations. This study included both types of railroads and considered them to have similar potential because of the following limitations: (1) there was no exact data to determine the incident rates of TB associated with each type of transportation; (2) there was no data to distinguish explicitly the overlapped portions of the two railways; and (3) the distinction is blurred in practice because railroads for transporting goods are sometimes taken over for passenger use in times of need.

- 4 Both road density for and the shortest distance between a TB case and a road were calculated using the lattice method. Firstly, all line-based road features were converted to raster cells showing presence of roads or otherwise (see diagram a below). Similarly, a raster layer was also available to show the presence of TB cases or not (see diagram b below).



(a) Sample raster for roads (1x1 km<sup>2</sup>)  
 indicates presence of roads



(b) Sample raster for TB cases (1x1 km<sup>2</sup>)  
 X indicates presence of TB cases

Each raster cell is identified by a coordinate pair ( $x$  column and  $y$  row). For raster of various sizes (e.g.,  $2 \times 2 \text{ km}^2$ ,  $5 \times 5 \text{ km}^2$ , or  $10 \times 10 \text{ km}^2$ ), the cell location of a TB case to the cell location of a road is computed using the distance formula  $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$ . The shortest distance was found by comparing the computed distances.

As for the density calculation, a number of lattice cells were selected around cells with TB cases (e.g.,  $2 \times 2 \text{ km}^2$ ,  $5 \times 5 \text{ km}^2$ , or  $10 \times 10 \text{ km}^2$ ). The proportion of lattice cells with roads within the selection was estimated to represent the average road density.

- <sup>5</sup> The modified local K-function extends the original version of the local K-function from a discrete to a continuous space. The reason for not using Ripley's L-function in this study is explained below.

In spatial statistics, the L-function is defined as  $\hat{L}(h) = \sqrt{\frac{\hat{K}(h)}{\pi}} - h$ , where  $\hat{L}(h) = 0$  is under complete spatial randomness (CSR). In other words, if the value of  $\hat{K}(h) > \pi h^2$  (or  $\hat{L}(h) > 0$ ) for  $h$ , the statistic indicates clustering at this scale  $h$ . If the value of  $\hat{K}(h) < \pi h^2$  (or  $\hat{L}(h) < 0$ ) for  $h$ , the statistic indicates regular distribution as a departure from CSR at this scale  $h$ . Consider the equation above, the Ripley's L-function is nothing more than a standardized version of the Ripley's K-function (see also <http://www.math.umt.edu/graham/stat544/ripleys.pdf>). Therefore, we chose not to involve the L-function because its inclusion would not add more information beyond that provided by the K-function in the present study.

## 8. Acknowledgement

This study was supported by the earmarked grant CUHK 14411614 of the Hong Kong Research Grants Council. We would like to thank participating laboratories in the county/district tuberculosis dispensaries, which collected and provided samples to the Shandong Chest Hospital for further analysis.

## 9. Ethical Approval

Ethical approval (No: 2011LINSHEN8) for GIS-based study was obtained from the Institutional Review Board (IRB) of the Shandong Chest Hospital (NIH IRB00006010). The authors declare they have no actual or potential competing financial interests.

## 10. References

- Andrews J.R., Morrow C., Wood R., 2013. Modeling the role of public transportation in sustaining tuberculosis transmission in South Africa. *American journal of epidemiology* 177(6): 556-561.
- Balcan D., Colizza V., Goncalves B., Hu H., Ramasco J.J., Vespignani A., 2009. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences* 106(51): 21484-21489.
- Barnett E.D., Walker P.F., 2008. Role of immigrants and migrants in emerging infectious diseases. *Medical Clinics of North America* 92(6):1447-1458.
- Besag J.E., 1972. Nearest-neighbor systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society, Series B*, 34:75-83.
- Bossak B.H., Mark R. Welford, M.R., 2015. Spatio-temporal characteristics of the medieval Black Death. In Kanaroglou P., Delmelle E., Páez A., Eds.: *Spatial analysis in Health Geography, Part II Infectious Disease*. Ashgate's Geographies of Health Series.
- Bryant J.M., Schürch A.C., van Deutekom H., Harris S.R., de Beer J.L., de Jager V., Kremer K., van Hijum S.A., Siezen R.J., Borgdorff M., Bentley S.D., Parkhill J., van Soolingen D., 2013. Inferring patient to patient transmission of *Mycobacterium tuberculosis* from whole genome sequencing data. *BMC infectious diseases* 13(1): 110.
- Carletta J., 1996. Assessing agreement on classification tasks: the Kappa statistic. *Computational linguistics* 22(2): 249-254.
- Carter A., Zwerling A., Olson S., Tannenbaum, T.N., Schwartzman K., 2009. Tuberculosis and the city. *Health & place* 15(3): 807-813.
- Centers for Disease Control and Prevention (CDC), 1995. Exposure of passengers and flight crew to *Mycobacterium tuberculosis* on commercial aircraft, 1992-1995. *Morbidity & Mortality Weekly Report* 44(8): 137-40 [Erratum appears in *Morbidity & Mortality Weekly Report* 1995; 44(9): 175].
- Chen, J., Shaw, S.L., Yu, H., Lu, F., Chai, Y., Jia, Q., 2011. Exploratory data analysis of activity diary data: a space-time GIS approach. *Journal of Transport Geography* 19(3): 394-404.

- China Tuberculosis Control Collaboration (CTCC), 2004. The effect of tuberculosis control in China. *The Lancet* 364(9432): 417-422.
- Cliff A.D., Ord J.K., Haggett P., Versey G.R. 1981. *Spatial Diffusion: An History Geography of Epidemics in an Island Community*. Cambridge, UK: Cambridge University Press.
- Cliff A. and Haggett P., 2004. Time, travel and infection. *British Medical Bulletin* 69(1): 87-99.
- Cook V.J., Sun S.J., Tapia J., Muth S.Q., Argüello D.F., Lewis BL, Rothenberg RB, McElroy PD. Network Analysis Project Team. 2007. Transmission network analysis in tuberculosis contact investigations. *Journal of Infectious Diseases* 196(10): 1517-1527.
- deBlij H.J., Murphy, A.B., 2003. *Human Geography: Culture, Society, and Space*. New York: John Wiley & Sons, 7th edition.
- Delmelle E., Delmelle E.C., Casas I., Barto T., 2011. HELP: a GIS-based health exploratory analysis tool for practitioners. *Applied Spatial Analysis and Policy* 4(2): 113-137.
- Dempster A., Laird N., Rubin D., 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society, Series B (methodological)* 39: 1-38.
- Dixon P.M., 2006. *Ripley's K Function*. Chichester, UK: John Wiley & Sons, Ltd.
- Dye C., 2006. Global epidemiology of tuberculosis. *The Lancet* 367(9514): 938-940.
- Edelson P.J., Phypers M., 2011. TB transmission on public transportation: a review of published studies and recommendations for contact tracing. *Travel medicine and infectious disease* 9(1): 27-31.
- Ellegård, K., Svedin, U., 2012. Torsten Hagerstrands time-geography as the cradle of the activity approach in transport geography. *Journal of Transport Geography* 23: 17-25.
- Fawcett T., 2006. An introduction to ROC analysis. *Pattern recognition letters* 27(8): 861-874.
- Fotheringham S.A., Rogerson P.A., 1993. GIS and spatial analytical problems. *International Journal of Geographical Information Science* 7(1): 3-19.
- Franzetti F., Codecasa L., Matteelli A., Degli E.A., Bandera A., Lacchini C., Lombardi A., Pinsi G., Zanini F., El-Hamad I., Gori A., 2010. Genotyping analyses of tuberculosis transmission among immigrant residents in Italy. *Clinical Microbiology and Infection* 16(8): 1149-1154.

- Ge E., Haining R., Li C.P., Yu Z.G., Waye M.M.Y., Chu K.H., Leung Y., 2012. Using knowledge fusion to analyze avian influenza H5N1 in East and Southeast Asia. *PLoS ONE* 7(5): e29617. doi:10.1371/journal.pone.0029617.
- Getis A., 1984. Interaction modeling using second-order analysis. *Environment and Planning A* 16(2): 173-183.
- Geodata Center. 2013; National Data Sharing Infrastructure of Earth System Science Homepage. Available: <http://www.geodata.cn/Portal/dataCatalog/dataList.jsp>. [accessed 10 Oct 2013]
- Gilbert M., Xiao X., Pfeiffer D.U., Epprecht M., Boles S., Czarnecki C., Chaitaweesub P., Kalpravidh W., Minh P.Q., Otte M.J., Martin V., Slingenbergh J., 2008. Mapping H5N1 highly pathogenic avian influenza risk in southeast Asia. *Proceedings of the National Academy of Sciences* 105(12): 4769-4774.
- Gomes M.F., y Piontti A.P., Rossi L., Chao D., Longini I., Halloran M.E., Vespignani A., 2014. Assessing the international spreading risk associated with the 2014 West African Ebola outbreak. Version 1, *PLOS Currents* 9(6). doi: 10.1371/currents.outbreaks.cd818f63d40e24aef769dda7df9e0da5.
- Haase I., Olson S., Behr M.A., Manyeki I., Thibert L., Scott A., Zwerling A., Ross N., Brassard P., Menzies D., Schwartzman K., 2007. Use of geographic and genotyping tools to characterize tuberculosis transmission in Montreal. *The International Journal of Tuberculosis and Lung Disease* 11(6): 632-638.
- Haining R., 1991. Bivariate correlation with spatial data. *Geographical Analysis* 23(3): 210-227.
- Haining R., 2003. *Spatial Data Analysis: Theory and Practice*. Cambridge, UK: Cambridge University Press.
- Horna-Campos O.J., Sánchez-Pérez H.J., Sánchez I., Bedoya A., Martín M., 2007. Public transportation and pulmonary tuberculosis, Lima, Peru. *Emerging infectious diseases* 13(10): 1491-1493.
- Jephcote C., Chen H., 2012. Environmental injustices of children's exposure to air pollution from road-transport with the model British multicultural city of Leicester: 2000-09. *Science of the Total Environment* 414: 140-151.
- Jia Z.W., Jia X.W., Liu Y.X., Dye C., Chen F., Chen C.S., Zhang W.Y., Li X.W., Cao W.C., Liu H.L., 2008. Spatial analysis of tuberculosis cases in migrants and permanent residents, Beijing, 2000-2006. *Emerging infectious diseases* 14(9): 1413.
- Jones-López E.C., Namugga O., Mumbowa F., Ssebidandi M., Mbabazi O., Moine S., Mboowa G., Fox M.P., Reilly N., Ayakaka I., Kim S., Okwera A., Joloba M., Fennelly K.P., 2013. Cough aerosols of *Mycobacterium tuberculosis* predict new infection. *American journal of respiratory and critical care medicine* 187(9): 1007-1015.

- Kausrud K.L., Begon M., Ari T.B., Viljugrein H., Esper J., Büntgen U., Leirs H., Junge C., Yang B., Yang M., Xu L., & Stenseth N.C. (2010). Modeling the epidemiological history of plague in Central Asia: palaeoclimatic forcing on a disease system over the past millennium. *BMC Biology* 8(1): 112.
- Kenyon T.A., Valway S.E., Ihle W.W., Onorato I.M., Castro K.G., 1996  
Transmission of multidrug-resistant Mycobacterium tuberculosis during a long airplane flight. *New England Journal of Medicine* 334(15): 933-938.
- Kim M., Chang S.I., Seong J.C., Holt J.B., Park T.H., Ko, J.H., Croft J.B., 2012.  
Road traffic noise: annoyance, sleep disturbance, and public health implications. *American journal of preventive medicine* 43(4): 353-360.
- Kwan, M.P., 2000. Human extensibility and individual hybrid-accessibility in space–time: a multi-scale representation using GIS. In: Janelle, D.J., Hodge, D.C. (Eds.), *Information, Place, and Cyberspace. Issues in Accessibility*. Springer-Verlag, Berlin, Germany, pp. 241–256.
- Lai P.C., Chow C.B., Wong H.T., Kwong K.H., Kwan Y.W., Liu S.H., Tong W.K., Cheung W.K., Wong W.L., 2015. An early warning system for detecting H1N1 disease outbreak – a spatio-temporal approach. *International Journal of Geographical Information Science* (ahead-of-print): 1-18.
- Long Q., Li Y., Wang Y., Yue Y., Tang C., Tang S., Squire S.B., Tolhurst R., 2008.  
Barriers to accessing TB diagnosis for rural-to-urban migrants with chronic cough in Chongqing, China: A mixed method study. *BMC Health Services Research* 8(1): 202.
- Madders M., Lawrence M., 1985. The contribution made by vegetation buffer zones to improved air quality in urban areas. In D.O. Hall, N. Myers & M.S. Margaris (Eds.), *Economics of Ecosystems Management*. Netherland: Springer, pp: 175-81.
- Mao H., 1996. The research about an indicator system of sustainable development in Shandong Province. *Geographical Research* 15(4): 16-23.
- Martens P., Hall L., 2000. Malaria on the move: human population movement and malaria transmission. *Emerging infectious diseases* 6(2):103-109.
- McPherson M.E., Kelly H., Patel M.S., 2008. Persistent risk of tuberculosis in migrants a decade after arrival in Australia. *Medical Journal of Australia* 188(9):528-531.
- Meade M., Florin J., Gesler M., 2000. *Medical Geography*, 2<sup>nd</sup> ed., Guilford, New York.
- Mohr O., Askar M., Schink S., Eckmanns T., Krause G., Poggensee G., 2012.  
Evidence for airborne infectious diseases transmission in public ground transports – A literature review. *Euro surveillance* 17(35): pii=20255.

- Nagelkere N.J.D., 1991. A note on a general definition of the coefficient of determination. *Biometrika* 78(3): 691-692.
- Pace-Asciak A., Mamo J., Calleja N., 2013. Tuberculosis among undocumented boat migrants to Malta: implications for a migrant tuberculosis policy. *The International Journal of Tuberculosis and Lung Disease* 17(8): 1065-1070.
- Qiu J., Lia R., Xu X., Hong X., Xia X., Yu C., 2014. Spatiotemporal pattern and risk factors of the reported novel avian-origin influenza A (H7N9) cases in China. *Preventive Veterinary Medicine* 115(3): 229-237.
- Ripley B.D., 1977. Modeling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological)* 39: 172-212.
- Saito M., Pan W.K., Gilman R.H., Bautista C.T., Bamrah S., Martín C.A., Tsiouris S.J., Argüello D.F., Martínez-Carrasco G., 2006. Comparison of altitude effect on *Mycobacterium tuberculosis* infection between rural and urban communities in Peru. *The American journal of tropical medicine and hygiene* 75(1): 49-54.
- Shandong Provincial Statistics Bureau (SPSB). Population. In: Shandong Statistical Yearbook 2010. China Statistics Press. P.R. China. Available: <http://chinadataonline.org/member/yearbook/default.asp?StartYear=1981&EndYear=2010&ybcode=SHANDONG>. [accessed 12 Nov 2013].
- Shandong Provincial Statistics Bureau (SPSB). Employment and Payment for Labor. In: Shandong Statistical Yearbook 2007. 2007a; China Statistics Press. P.R. China. Available: <http://chinadataonline.org/member/yearbook/default.asp?StartYear=1981&EndYear=2010&ybcode=SHANDONG>. [accessed 12 Nov 2013].
- Shandong Provincial Statistics Bureau (SPSB). Major Statistics of Counties (Cities and Districts at County Level). In: Shandong Statistical Yearbook 2007. 2007b; China Statistics Press. P.R. China. Available: <http://chinadataonline.org/member/yearbook/default.asp?StartYear=1981&EndYear=2010&ybcode=SHANDONG>. [accessed 12 Nov 2013].
- Shandong Transportation Department (STD) 2013; Introduction of Shandong Transportation Homepage. Available: <http://www.sdjt.gov.cn/publish/main/41/index.html>. [accessed 6 Dec 2013].
- Stoddard S.T., Morrison A.C., Vazquez-Prokopec G.M., Soldan V. P., Kochel T., Kitron U., Elder J.P., Scott T.W., 2009. The role of human movement in the transmission of vector-borne pathogens. *PLoS neglected tropical diseases* 3(7): e481. doi: 10.1371/journal.pntd.0000481
- Tessema B., Beer J., Merker M., Emmrich F., Sack U., Rodloff A.C., Niemann, S., 2013. Molecular epidemiology and transmission dynamics of *Mycobacterium tuberculosis* in Northwest Ethiopia: new phylogenetic lineages found in Northwest Ethiopia. *BMC infectious diseases* 13(1): 131.

Thamizh Arasan V. and Dhivya G., 2009. Concentration of Heterogeneous Road Traffic. In: Kankesu J., (Ed.), *Advanced Technologies*. InTech, Available: <http://www.intechopen.com/books/advanced-technologies/concentration-of-heterogeneousroad-traffic>. [accessed 14 Dec 2013].

The CGIAR-CSI Consortium for Spatial Information (CGIAR-CSI) Homepage. Available: <http://srtm.csi.cgiar.org>. [accessed 14 Dec 2013].

Vazquez-Prokopec G.M., Kitron U., Montgomery B., Horne P., Ritchie S.A., 2010. Quantifying the spatial dimension of dengue virus epidemic spread within a tropical urban environment. *PLoS neglected tropical diseases* 4(12): e920.

Wei X., Chen J., Chen P., Newell J.N., Li H., 2009. Barriers to TB care for rural-to-urban migrant TB patients in Shanghai: a qualitative study. *Tropical Medicine & International Health* 14(7): 754-760.

Wen T.H., Lin M.H., Fang C.T., 2012. Population movement and vector-borne disease transmission: Differentiating spatial-temporal diffusion patterns of commuting and non-commuting dengue cases. *Annals of the Association of American Geographers* 102(5): 1026-1037.

Wesolowski A., Buckee C.O., Bengtsson L., Wetter E., Lu X., Tatem A.J., 2014. Commentary: Containing the Ebola outbreak-the potential and challenge of mobile network data. Version 1, *PLoS Currents* 9(6). doi: 10.1371/currents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e.

Wheeler D.C., 2007. A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in Ohio, 1996-2003. *International Journal of Health Geographics* 6(1): 13.

Wieczorek, W. F., Delmerico, A. M., Rogerson, P. A., & Wong, D. W. (2012). Clusters in irregular areas and lattices. *Wiley Interdisciplinary Reviews: Computational Statistics* 4(1): 67-74.

Wilson M.E., 1995. Travel and emergence of infectious diseases. *Emerging infectious diseases* 1(2): 39.

World Health Organization. *Global Tuberculosis Report 2012*. 2012 Report No.: WF300. Geneva: the World Health Organization. Available: [http://apps.who.int/iris/bitstream/10665/75938/1/9789241564502\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/75938/1/9789241564502_eng.pdf). [accessed 5 August 2014].

World Health Organization (WHO), *Global Tuberculosis Report 2013*. Geneva: World Health Organization. Available: [http://apps.who.int/iris/bitstream/10665/91355/1/9789241564656\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/91355/1/9789241564656_eng.pdf). [accessed 30 May 2014]

Wu Z., Braunstein L.A., Havlin S., Stanley H.E., 2006. Transport in weighted networks: partition into superhighways and roads. *Physical review letters* 96(14): 148702.

Xia Y., Bjørnstad O.N., Grenfell B.T., 2004. Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *The American Naturalist* 164(2): 267-281.

Yang X., Huang Y., Dong P., Jiang D., Liu H., 2009. An updating system for the gridded population database of China based on remote sensing, GIS, and spatial database technologies. *Sensors* 9(2): 1128-1140.

Zhou C., 2011. Study on current status and countermeasures of tuberculosis patients health-seeking behavior and treatment management among migrants in Shandong. Ph.D. thesis. Shandong University, China.

### Figure Captions

**Figure 1.** Spatial distribution of TB cases in Shandong province, China (2009-2011). The base map portrays four types of transportation networks (provincial road, national road, highway, and railway) on a terrain characterized by elevation differences. TB cases appeared to aggregate in areas of low altitude and heavy traffic.

**Figure 2.** Results of the *K*-function and modified local *K*-function analyses. (a) A plot of global measures of clustering. The global pattern of TB showed clustering below 7 km but dispersion beyond that value. Clustering of TB cases is most significant at around 4-5 km. (b) A map showing the intensity of TB clusters (with values ranging between 0 and 1) in Shandong. At the optimum scale of 5 km, most TB clusters were found in areas of low altitude and nearby towns or counties with well-connected and accessible transportation networks.

**Figure 3.** The frequency, boxplots, and *K*-function plots of TB cases of two regions of altitudinal difference in Shandong. Left column: The frequency and boxplots show that low altitude regions vary between -105 and 81 meters in elevation. The *K*-function plot exhibits clustering (significant at 95%) to occur at around 15 km and below. Right column: The frequency and boxplots indicate that elevation for high altitude regions is over 81 meters with an average height of 181 meters. The *K*-function analysis reveal clustering (significant at 95%) of TB cases

at 5 km and below although at a lesser degree compared to that of low altitude region.

**Figure 4.** ROC curves for the auto-logistic regression models. The area under the ROC curve (AUC) indicates the probability of correct classifications. (a) and (b) show the 2009-2011 average AUC of the models for TB cases in low and high altitude regions respectively. The shaded areas represent the envelopes of the 1000 bootstrap replicates.

**Figure 5.** Predicted TB risks in Shandong. The probability map shows the risks of having the disease, with probability values approaching one indicating higher risks. Areas with higher risks of TB occurrences are located near provincial and national roads, as well as at major transport junctions across the low altitude regions.

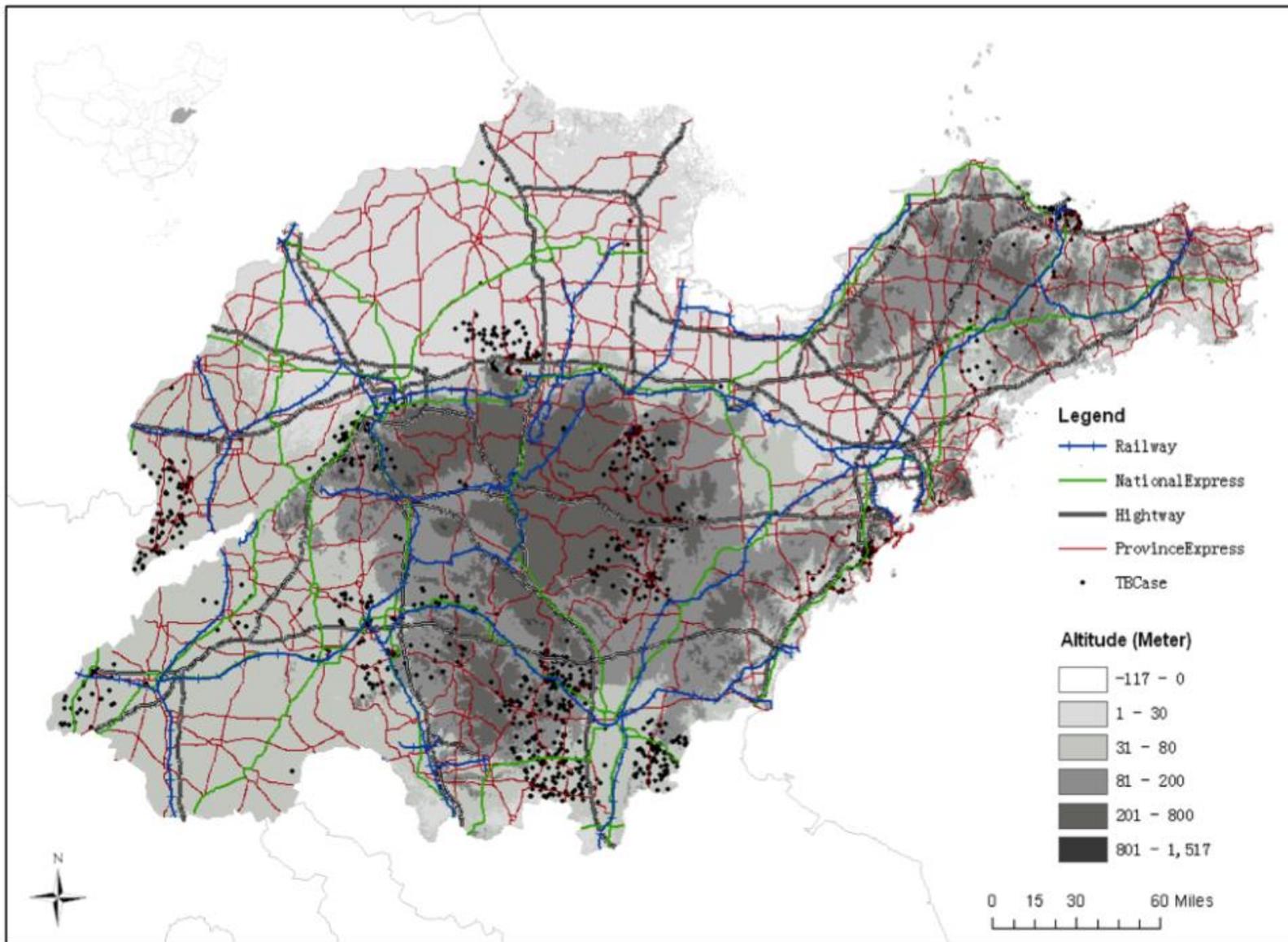


Figure 1

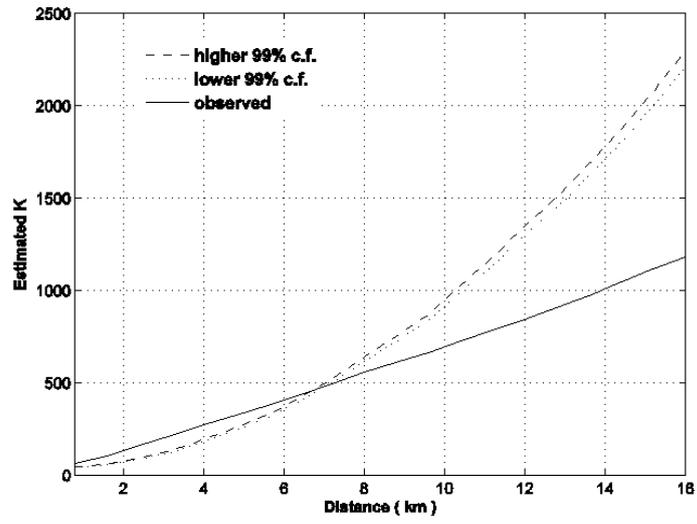


Figure 2a



Figure 2b

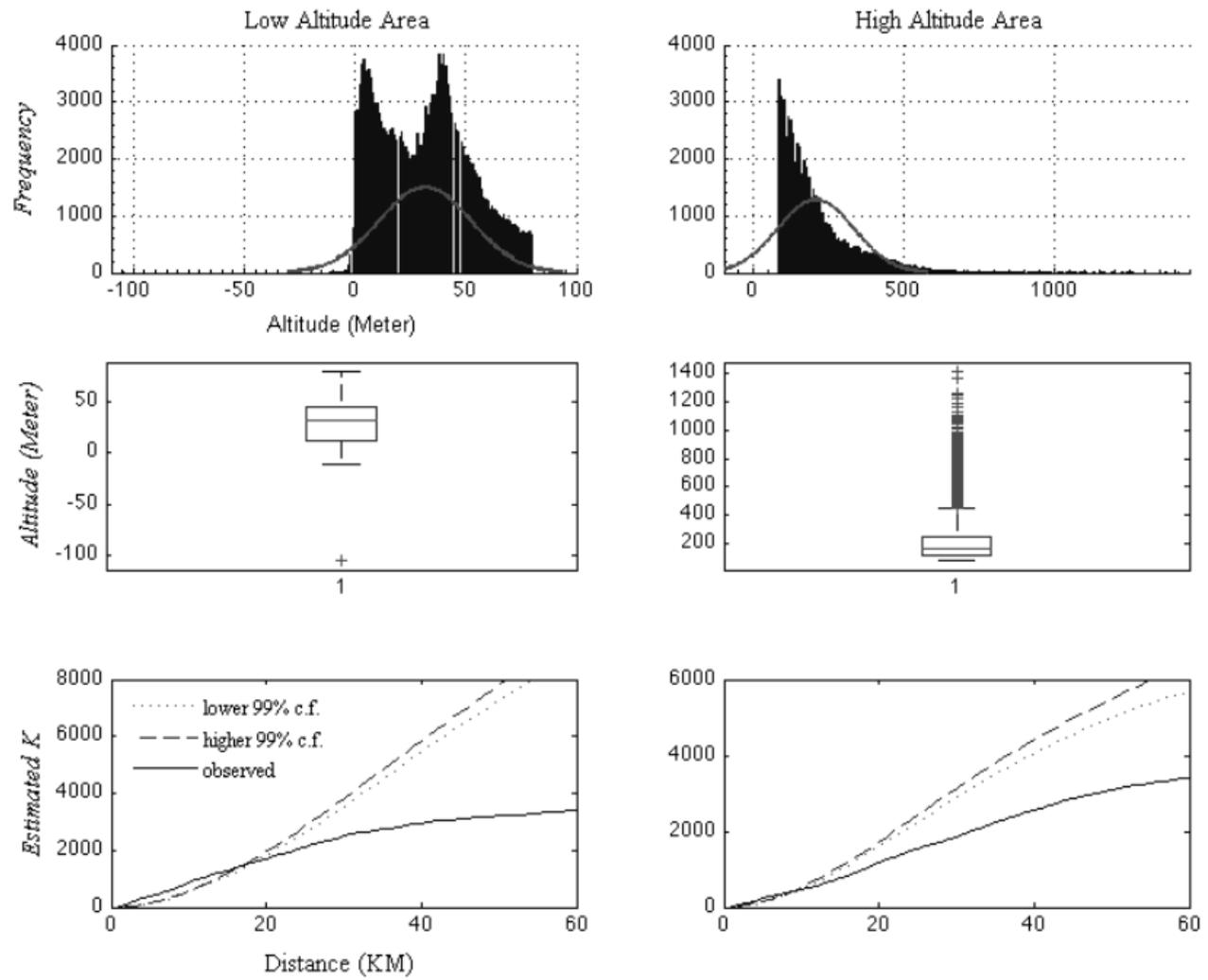
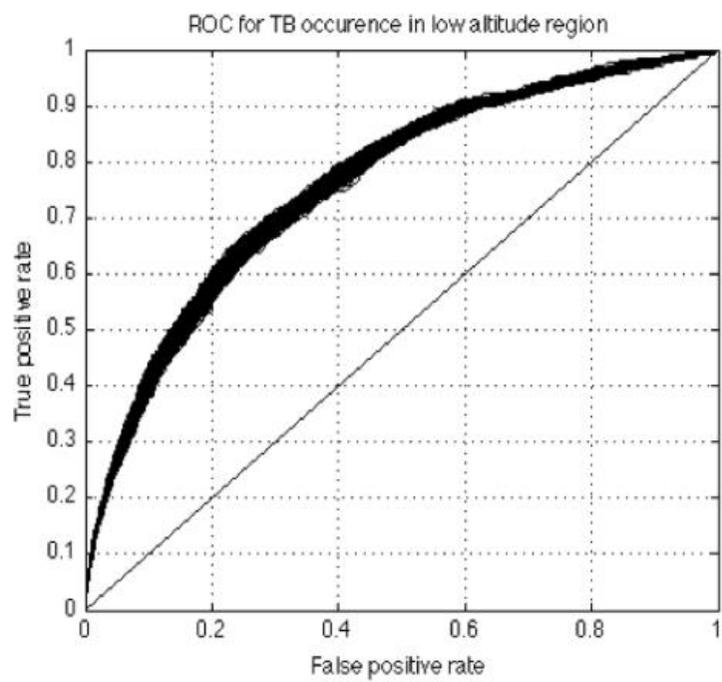
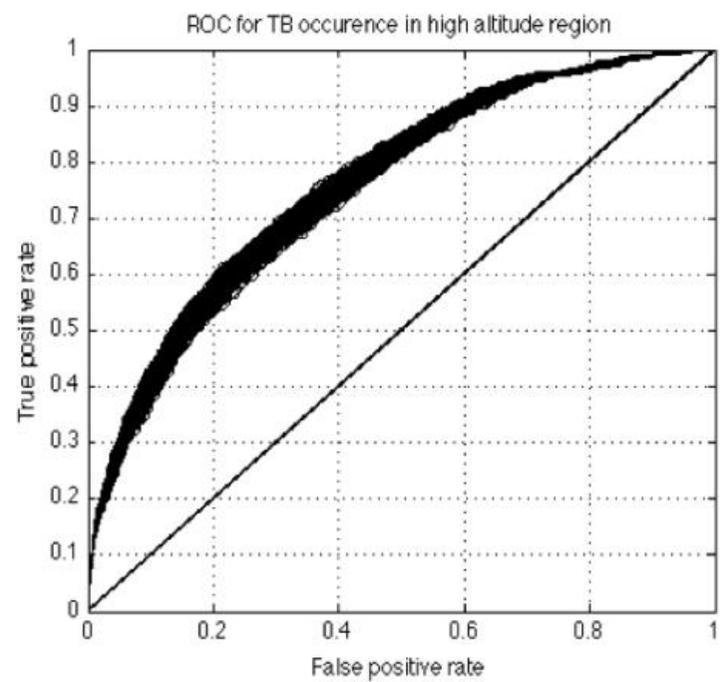


Figure 3



(a)



(b)

Figure 4



Figure 5