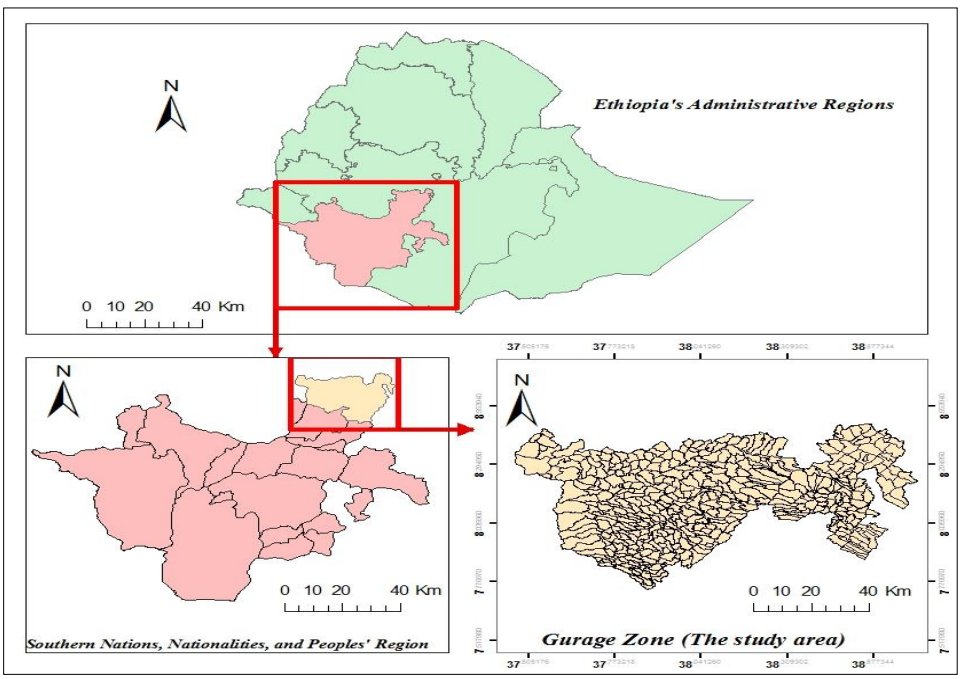
COLLEGE OF HEALTH SCIENCES

SCHOOL OF PUBLIC HEALTH

THE SPATIAL EPIDEMIOLOGY OF TUBERCULOSIS IN GURAGE ZONE, SOUTHERN ETHIOPIA

BY: SEBSIBE TADESSE ONKE



A DISSERTATION SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES OF ADDIS ABABA UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD) IN PUBLIC HEALTH

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

THE SPATIAL EPIDEMIOLOGY OF TUBERCULOSIS IN GURAGE ZONE, SOUTHERN ETHIOPIA

A DISSERTATION SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES OF ADDIS ABABA UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD) IN PUBLIC HEALTH

BY: SEBSIBE TADESSE ONKE

SUPERVISORS

PROF. FIKRE ENQUESELASSIE GASHE (PHD)

DR. SEIFU HAGOS GEBREYESUS (PHD)

SCHOOL OF PUBLIC HEALTH, COLLEGE OF HEALTH SCIENCES

ADDIS ABABA UNIVERISTY

DECEMBER, 2018

DISSERTATION APPROVAL

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

THE SPATIAL EPIDEMIOLOGY OF TUBERCULOSIS IN GURAGE ZONE, SOUTHERN ETHIOPIA

BY: SEBSIBE TADESSE ONKE

SCHOOL OF PUBLIC HEALTH, COLLEGE OF HEALTH SCIENCES

ADDIS ABABA UNIVERISTY

APPROVED BY THE EXAMINING BOARD

Chairman, Examining Board	**Signature**	**Date**
Supervisor (Primary)	**Signature**	**Date**
Supervisor (Secondary)	**Signature**	**Date**
External, Examiner	**Signature**	**Date**
Internal, Examiner	**Signature**	**Date**
Internal, Examiner	**Signature**	**Date**

LIST OF ORIGINAL PAPERS

Paper I: Tadesse S, Enqueselassie F, Hagos S. Spatial and space-time clustering of tuberculosis in Gurage zone, southern Ethiopia. PLoS One 2018; 13(6):e0198353.

Paper II: Tadesse S, Enqueselassie F, Gebreyesus SH. Estimating the spatial risk of tuberculosis distribution in Gurage zone, southern Ethiopia: a geostatistical kriging approach. BMC Public Health 2018; 18:783.

Paper III: Tadesse S, Enqueselassie F, Gebreyesus SH. Ecological factors affecting spatial distribution of tuberculosis in Gurage Zone, Southern Ethiopia. BMC Public Health 2018 (Under review).

ACRONYMS AND ABBREVIATIONS

AFB	Acid Fast Bacilli
AIC	Akaike's Information Criterion
ASE	Average-Standard Error
BCG	Bacille Calmette-Guerin
BIC	Bayesian Information Criterion
CSA	Central Statistical Agency
DOTS	Directly Observed Treatment-Short course
ETB	Extra-pulmonary Tuberculosis
GeoDa	Geographic Data analysis tool
GIS	Geographic Information Systems
GWR	Geographically Weighted Regression
LLR	Log-Likelihood Ratio
MTB	*Mycobacterium tuberculosis*
MDR-TB	Multi-Drug Resistant Tuberculosis
MODIS	Moderate Resolution Imaging Spectroradiometer
MSE	Mean-Standardized Error
NDVI	Normalized Difference Vegetation Index
PTB-	Smear-negative Pulmonary Tuberculosis
PTB+	Smear-positive Pulmonary Tuberculosis
RMSE	Root-Mean-Square Error
RMSSE	Root-Mean-Square-Standardized Error
RR	Relative Risk
SAC	Spatial Autocorrelation
SAR	Spatial Autoregressive
SDM	Spatial Durbin Model
SEBS	Spatial Empirical Bayes Smoothing
SEM	Spatial Error Model
TB	Tuberculosis
TB/HIV	Tuberculosis co-infection with Human Immunodeficiency Virus

WHO	World Health Organization
XSMLE	Estimating spatial panel data using Maximum Likelihood Estimator

ACKNOWLEDGMENTS

First and foremost, I give my sincerest thanks to the Almighty God for His unconditional and everlasting love!

Next, I heartily thank my primary supervisor Prof. Fikre Enqueselassie for his dedicated and highly qualified support, swift feedback, friendliness, patience and understanding. His advice gave me confidence to think critically and laid a foundation to work independently. Without his motivation and encouragement, I would not have completed this thesis work. I extend my best wishes for his future.

I would also like to thank my second supervisor Dr. Seifu Hagos for his valuable inputs and encouragement throughout this thesis process. His spatial epidemiological insight and critical appraisal were always helpful to me.

My deepest gratitude also goes to University of Gondar and Addis Ababa University for giving me this PhD training opportunity, and for funding the research project. Furthermore, I would like to express my appreciation for academic staff and PhD students at the School of Public Health, College of Health Sciences, Addis Ababa University from whom I have learnt during the various seminars, classes and other academic endeavors.

My mentors Prof. Tefera Abula and Mr. Yohannes Jorge in particular have made the process easier by providing me their offices with internet access, and offering good advice and coffee.

I am very thankful for the data collectors, supervisors and personnel at health facilities in Gurage Zone for their support and patience during the field work.

Finally, my special thanks go to my family, especially my wife Simenesh Mulugeta for her support, love, and understanding.

TABLE OF CONTENTS

Contents **Pages**

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Background: The global distribution of tuberculosis is skewed heavily toward low-and-middle income countries, which accounted for about 87% of all estimated incident cases. Ethiopia is a low-income country in east Africa that remains highly afflicted by tuberculosis, with varying degrees of magnitudes across settings. However, there is a dearth of studies clarifying about the spatial epidemiology of the disease in Ethiopia. Lack of such information may contribute to the partial effectiveness of tuberculosis control programs.

Objectives: The specific objectives of this study were: 1) to detect spatial and space-time clustering of tuberculosis, 2) to estimate spatial risk of tuberculosis distribution using limited spatial datasets, and 3) to identify ecological factors affecting spatial distribution of tuberculosis in Gurage Zone, Southern Ethiopia.

Methods: The study data were obtained from different sources. Specific objectives 1 and 3 included a total of 15,805 tuberculosis patients diagnosed at health facilities in Gurage Zone during 2007 to 2016, whereas specific objective 2 included 1,601 patients diagnosed in 2016. The geo-location and population data were obtained from the Central Statistical Agency of Ethiopia (specific objectives 1-3). The altitude data were extracted from global digital elevation model v2 (specific objective 2). The normalized difference vegetation index data were derived from the moderate resolution imaging spectroradiometer imagery, and the temperature and rainfall data were obtained from the Meteorological Agency of Ethiopia (specific objective 3). The global Moran's I, Kulldorff's scan and Getis-Ord G_i^* statistics were used to analyze purely spatial and space-time clustering of tuberculosis (specific objective 1). The geostatistical kriging approach was applied to estimate the spatial risk of tuberculosis distribution (specific objective 2). The spatial panel data analysis was used to estimate the effects of ecological factors on spatial distribution of tuberculosis prevalence rate (specific objective 3).

Results: The prevalence of tuberculosis varied from 70.4 to 155.3 cases per 100,000 population in the Gurage Zone during 2007 to 2016. Eleven purely spatial clusters (relative risk: 1.36–14.52, P-value < 0.001) and three space-time clusters (relative risk: 1.46–2.01, P-value < 0.001) for high occurrence of tuberculosis were detected. The clusters were mainly concentrated in border areas of the zone. The predictive accuracies of ordinary cokriging models have improved with the inclusion of anisotropy, altitude and latitude covariates, the change in detrending pattern from local to global, and the increase in size of spatial dataset (mean-standardized error = 0, root-

mean-square-standardized error = 1, and average-standard error ≈ root-mean-square error). The spatial risk of tuberculosis was estimated to be higher (i.e., tuberculosis prevalence rate > 100 cases per 100,000 population) at western, northwest, southwest and southeast parts of the study area, and crossed between high and low at west-central parts**.** The tuberculosis prevalence rate observed in a given kebele was determined by both tuberculosis prevalence rate (spatial autoregressive coefficient = 0.83) and unobserved factors (spatial autocorrelation coefficient = -0.70) in the neighboring kebeles. By controlling the spatial effects, a 1°C rise in temperature was associated with an increase in the number of tuberculosis prevalence rate by 0.72, and a 1 person per square kilometer increase in population density was related to an increase in the number of tuberculosis prevalence rate by 1.19.

Conclusions: The spatial and space-time clusters for high occurrence of tuberculosis were mainly concentrated at border areas of the Gurage Zone. The prevalence rate of tuberculosis in a given kebele was determined by both the prevalence rate of tuberculosis and other unobserved factors in its neighboring kebeles in the zone, indicating sustained transmission of the disease within the communities. The spatial risk of tuberculosis distribution between kebeles in the zone was partly explained by spatial variations in temperature, population density, altitude, and latitude. The geostatistical kriging approach can be applied to estimate the spatial risk of tuberculosis distribution in data limited settings.

Recommendations: Tuberculosis control programs should consider the cooperation of neighboring kebeles in the design and implementation of tuberculosis prevention and control strategies to interrupt the chain of disease transmission between the communities. Moreover, the designing of locally effective tuberculosis prevention and control strategies should consider spatial locations with higher temperature and population density. Further research is required to evaluate the effectiveness of geographically targeting tuberculosis prevention and control interventions using the inputs from spatial epidemiological methods.

Keywords: Ecological factors, Geostatistical kriging approach, Purely spatial clusters, Space-time clusters, Spatial autocorrelation, Spatial epidemiology, Spatial heterogeneity**,** Spatial panel data analysis, Tuberculosis distribution

1. INTRODUCTION

1.1 Background

Tuberculosis (TB) is an infectious disease affecting millions worldwide and is caused by the *Mycobacterium tuberculosis* (MTB*)* complex, with MTB being the most common infecting species in humans. Despite intensified efforts to control TB under Directly Observed Treatment-Short course (DOTS) strategy, the disease continues to place an extraordinary public health, financial, and social burden on those afflicted by the disease and their families, and on government (1, 2). The global distribution of the disease is skewed heavily toward resource-limited settings due to inequitable distribution of health services, poor socio-economic conditions, co-infection with Human Immunodeficiency Virus (TB/HIV) and emergence of Multi-Drug Resistant TB (MDR-TB) (1, 3, 4).

The Government of Ethiopia initiated a pilot TB control project based on the DOTS strategy in the mid of 1992 (5). Since then the program has been subsequently scaled up in the country and has reached 100% district and 90% health facility coverage (6). However, studies from northern, central and southern Ethiopia have shown that the prevalence of TB ranges from 76 to 189 per 100,000 population suggesting the spatial distribution of TB is nonrandom and cases are clustered at specific-geographical locations (7-10). This calls for an urgent need to identify high-risk geographical areas that require targeted interventions.

The recent advancements in spatial statistics and Geographic Information Systems (GIS), and increasing availability of public health data from the health facilities, demographic and health surveillance sites and national surveys have provided good opportunities for epidemiologists to explore the spatial nature of TB, even at smaller administrative levels (9, 11, 12). These techniques can be used to assess the local inequalities in TB distribution in Ethiopia, where uniform interventions are being implemented across the country (9, 10). Therefore, this study attempted to assess the spatial epidemiology of TB in Gurage Zone, Southern Ethiopia.

1.2 Statement of the problem

TB is a major global public health problem. In 2016, there were an estimated 10.4 million incident cases and 1.7 million deaths worldwide (1). The rate varied widely among countries from under 10 per 100,000 population in most high-income countries to 150-300 in most of the 30 high TB burden countries, and above 500 in a few countries (1). Africa, home to 13% of the world's population, carries 25% of the global burden of TB cases (1).

Ethiopia is a low-income country in east Africa that remains highly afflicted by TB and is ranked among the list of 14 countries with high burden of TB, TB/HIV and MDR-TB (1). In 2016, there were an estimated 182,000 incident cases and 30,000 death (1). About 7.6% of the incident cases were estimated among people living with HIV in 2016. The MDR-TB cases were estimated to be 2.7% (95% CI: 1.5-4%) among new TB cases and 14% (95% CI: 5.6-23%) among previously treated TB cases (1). Moreover, TB is a leading cause of death among infectious diseases in the country (13).

Several studies have been carried out at district, regional and national levels to investigate the epidemiology of TB in Ethiopia (5, 14, 15). Although these studies disclose important information for TB control programs, spatial nature of the disease have rarely been taken into account. The studies are based on administrative units that may include cases outside of the catchment areas or miss cases from their catchment that are enrolled in neighboring health facilities. Moreover, most of the studies were conducted in a short period of time, which make them deficient in detecting the pattern of the disease distribution over time. Spatial analyses of TB may help public health officials discover high-risk geographical locations for targeted interventions, identify neighborhood factors affecting TB distribution, explore the transmission pattern, and evaluate the impacts of interventions (9, 16). Lack of such information may contribute for the partial effectiveness of TB control programs in Ethiopia (9, 10).

The spatial distribution of TB at a given geographic area can be affected by the spatial heterogeneity of underlying ecological factors, like climatic, socioeconomic and environmental factors (11, 17). Geographically-diverse literatures have revealed that the ecological factors affecting spatial distribution of TB in one region may not be the same as those in another region (11, 17, 18). This spatial heterogeneity proves the place-specific nature of the risk factors.

However, most previous studies did not incorporate the spatial effects of the factors into statistical modeling of TB distribution in Ethiopia (9, 10, 12). This might bias the estimation results, and finally could lead to wrong public health policy formulation.

This study attempted to detect spatial and space-time clustering of TB, estimate spatial risk of TB distribution using limited spatial datasets, and identify ecological factors affecting spatial distribution of TB in Gurage Zone, southern Ethiopia. The findings may help local public health authorities as a guide for planning, budgeting and resource mobilization.

1.3 Rationale and significance

Within the context of Sustainable Development Goal 3, the World Health Organization (WHO) defined the End TB Strategy 2016-2035—an ambitious framework for ending the global TB epidemic by 2035. It calls for spatial analysis of TB data to assess within-country inequalities, with findings used to identify geographical areas where progress is lagging and greater attention is needed (1). Moreover, epidemiological evidences have revealed that applying similar interventions to control TB across settings is not effective and efficient (9, 10). Thus, TB control programs should consider strategies to reach the vulnerable populations at high-risk geographical areas where on-going transmission sustains the epidemic. However, there is a dearth of studies clarifying about the spatial epidemiology of TB in Ethiopia. Lack of such information may contribute to the partial effectiveness of TB control programs.

Therefore, this study was intended to detect spatial and space-time clustering of TB, estimate spatial risk of TB distribution, and identify ecological factors affecting spatial distribution of TB in Gurage Zone, Southern Ethiopia. The study findings may help to identify geographical locations that require targeted interventions, measure the burden of TB within-and-between smaller administrative units, active case detection, evaluate the impacts of intervention programs, optimize public health resource utilization, increase communities' awareness, improve patient treatment outcome, and reduce drug resistance. Moreover, the findings may also add for the growing body of spatial epidemiological research on TB in high-burden countries.

2. LITERATURE REVIEW

2.1 The biology of TB

2.1.1 Historical perspective of TB

TB is an ancient disease. Pathological signs of tubercular decay discovered in Egyptian mummies proved the existence of human TB since prehistoric era. The disease has been described by various names, such as *schachepheth* in Hebrew, *phthisis* in Greek, and c*onsumption*, *wasting disease* and the *white plague* in English (19).

In the 18th century, the first TB epidemic occurred in Europe as a result of industrialization (20). Cities were being rapidly developed and urban settlements were improperly arranged. People were thrown together in squatters where housing and hygiene condition was poor. Crowded living environments, poor hygiene and malnutrition constituted ideal circumstances for the accelerated spread of TB. Following the global diffusion of industrialization and population migration, TB epidemic swept through the world in the next 150 years (20).

Sanatorium care was introduced as the first specific treatment for TB in mid 1800s. Infected people were sent to sanatorium where they were provided with plenty of food, large living area and fresh air. The improved living environment allowed patients to strengthen host immunity to fight against the TB bacteria in their body, and gradually led to recovery. Medical care in sanatorium had been used for almost 100 years as a primary treatment to TB. The discovery of MTB by Robert Koch in 1882 and the discovery of X-rays by Wilhelm Konrad Röntgen in 1895 led to a wave of research that initiated the development of the Bacille Calmette-Guerin (BCG) vaccine in the 1912, and effective medical treatment. Some of the major breakthroughs were the discovery of the anti-TB drugs, like streptomycin in 1944, isoniazid in 1952, and rifampicin in 1965 (19).

2.1.2 Aetiological agents of TB

TB is caused by the members of the MTB complex, with MTB being the most common infecting species in humans. The MTB complex also comprises the human pathogens *M. africanum* and *M. canettii* as well as the primarily animal infecting species *M. bovis*, *M. microti*, *M. pinnipedii* and *M. caprae*, all of which have been identified as causative agents of TB in humans (21, 22).

2.1.3 Transmission of TB

The MTB is spread by small airborne droplets, called droplet nuclei (each of them contain 5,000 to 10,000 bacteria), generated by the coughing, sneezing, talking, or singing of a person with pulmonary or laryngeal TB. About 3,000 droplet nuclei can be generated when one coughs, talks for 5 minutes, and sings for one minute. Sneezing generates about 40,000 droplet nuclei, which can spread to individuals up to 10 feet away. These tiny droplets can remain airborne for minutes to hours after expectoration (23). Each of them may transmit the disease since inhaling less than 10 bacteria may cause an infection (24). If not treated, an infectious person could infect 10 to 15 people on average every year (25). The probability of transmission depends upon infectiousness of the person with TB, environment of exposure, duration of exposure, and virulence of the organism (26). The chain of transmission can be stopped by isolating patients with active disease and starting effective anti-TB therapy (27).

2.1.4 Clinical manifestations of TB

TB may develop differently in each patient according to the patient's status of immune system. Stages include latency, primary disease, primary progressive disease, and extrapulmonary disease. Each stage has different clinical manifestations.

Latent TB

MTB organisms can be enclosed, but are difficult to completely eliminate (28). Persons with latent TB have no signs or symptoms of the disease, do not feel sick, and are not infectious (29). However, viable bacilli can persist in the necrotic material for years or even a lifetime (30), and if the immune system later becomes compromised, the disease can be reactivated. Factors that can trigger reactivation of an infection include co-infection with HIV, uncontrolled diabetes mellitus, renal failure, malnutrition, smoking, chemotherapy, organ transplantation, and long-term corticosteroid usage (29).

Primary disease

Primary pulmonary TB is often asymptomatic that the results of diagnostic tests are the only evidence of the disease. Although primary disease essentially exists subclinically, some self-limiting findings might be noticed in an assessment. Associated paratracheal lymphadenopathy may occur because the bacilli spread from the lungs through the lymphatic system. If the primary

lesion enlarges, pleural effusion is a distinguishing finding. This effusion develops because the bacilli infiltrate the pleural space from an adjacent area. The effusion may remain small and resolve spontaneously, or it may become large enough to induce symptoms, such as fever, preuritic chest pain, and dyspnea. Dyspnea is due to poor gas exchange in the areas of affected lung tissue. Dullness to percussion and a lack of breath sounds are physical findings indicative of a pleural effusion because excess fluid has entered the pleural space (31).

Primary progressive TB

Active TB disease develops in only 5 to 10% of persons infected with MTB. When a patient progress to active TB, early signs and symptoms are often nonspecific. Manifestations often include progressive fatigue, malaise, weight loss, and a low-grade fever accompanied by chills and night sweats (32). Wasting is due to lack of appetite and altered metabolism associated with the inflammatory and immune responses.

Wasting involves the loss of both fat and lean tissue; the decreased muscle mass contributes to the fatigue (33). Finger clubbing, a late sign of poor oxygenation, may occur; however, it does not indicate the extent of the disease (34). A cough eventually develops in most patients. Although the cough may initially be nonproductive, it advances to a productive cough of purulent sputum. The sputum may also be streaked with blood. Hemoptysis can be due to destruction of a patent vessel located in the wall of the cavity, the rupture of the dilated vessel in a cavity, or the formation of an aspergilioma in an old cavity. The inflamed parenchyma may cause pleuritic chest pain. Extensive disease may lead to dyspnea or orthopnea because the increased interstitial volume leads to a decrease in lung diffusion capacity. Although many patients with active disease have few physical findings, rales may be detected over involved areas during inspiration, particularly after a cough. Hematologic studies may reveal anemia, which is the cause of the weakness and fatigue. Leukocytosis may also occur due to the large increase in the number of leukocytes in response to the infection (31).

Extrapulmonary TB

Although the pulmonary system is the most common location for TB, extrapulmonary disease occurs in more than 20% of immunocompetent patients, and the risk for extrapulmonary disease increases with immunosuppression (35). The most serious location is the central nervous system,

where infection may result in meningitis or tuberculomas. If not treated, tubercular meningitis is fatal in most cases; making rapid detection of the myobacteria essential (36). Headaches and change in mental status after possible exposure to TB or in high risk groups should prompt consideration of this disease as a differential diagnosis. Another fatal form of extrapulmonary TB is infection of the bloodstream by mycobacteria; this form of the disease is called disseminated or miliary TB. The bacilli can then progresses rapidly throughout the body, leading to multi-organ involvement (37). Miliary TB progresses rapidly and can be difficult to diagnose because of its systemic and nonspecific signs and symptoms, such as fever, weight loss, and weakness (31). Lymphatic TB is the most common extrapulmonary TB, and cervical adenopathy occurs most often. Other locations include bones, joints, pleura, and genitourinary system (35).

2.1.5 Diagnosis and treatment of TB

In low-income countries diagnosis of TB relies on detection of Acid Fast Bacilli (AFB) on microscopic examination and on cultivation of MTB. Diagnosing MDR-TB is more demanding since it takes six to eight weeks to grow the mycobacteria on a culture media containing anti-TB drugs and requires special laboratory facilities (38). Rapid tests like GeneXpert and line probe assays are also being used to diagnose drug resistance. The sputum smear result is often negative in patients with extra-pulmonary TB and TB/HIV co-morbidity. Because of insufficient laboratories for histo-pathological or culture examinations, the most common way of diagnosing such cases is clinical diagnosis supported by chest radiography.

Chemotherapy regimens that are used for the treatment of all types of TB are classified as first- and second-line anti-TB drugs. First-line anti-TB drugs include Isoniazid, Rifampicin, Pyrazinamide, Ethambutol and Streptomycin. Isoniazid and Rifampicin are the two most commonly used drugs for treatment of TB. First-line anti-TB drugs are safe and effective if used correctly. Second-line drugs are used for the treatment of MDR-TB. These are listed as aminoglycosides, polypeptides, fluoroquinolones and thioamides. Second-line anti-TB drugs are less potent, need to be administered for a much longer time, are more toxic and are high-cost compared to first-line anti-TB drugs. Agents with unclear roles in drug-resistant TB treatment are called third-line anti-TB drugs, such as Clofazimine, Linezolid, Amoxicillin/Clavulanate, Thioacetazone, Imipenem/Cilastatin and high-dose Isoniazid (39).

2.2 Spatial epidemiology of TB

2.2.1 Spatial epidemiology

Spatial epidemiology is the description and analysis of geographically indexed health data with respect to demographic, environmental, and socioeconomic risk factors. The core idea of spatial epidemiology is based on the Tobler's First Law of Geography. In 1970, Tobler stated that everything is related to everything else, but near things are more related than distant things (40). Therefore, in the analytical framework of spatial epidemiology, disease is not evenly distributed across space. The resultant disease pattern could be influenced by the varying distribution of risk factors over space (40).

The discipline of spatial epidemiology has developed greatly since the 19th century with the development of spatial statistics and advances in specialist GIS software and handheld Global Positioning System receivers in the past two decades. The utilization of these methods and software packages can allow the spatial linkage of various datasets of relevance to disease distributions and the description and understanding of complex spatial patterns in the risk of disease. There is increasing acceptance of evidence that people's area of residence may influence their health either in addition to or in interaction with their individual characteristics. Most of the time environmental determinants could not be conceptualized in individual level. In this case, it is necessary to aggregate disease pattern and suspected determinants by different places, treating spatial units, rather than individuals, as the basic analytical unit of the study (40).

Spatial epidemiology is of particular significance in studying infectious disease. As infectious disease must be transmitted via direct contact, the presence of disease in certain area may imply possible episodes of transmission. There is growing number of studies applying spatial statistics in studying infectious disease, such as schistosomiasis, filariasis, soil transmitted helminths, dengue, malaria, H1N1 influenza pandemic, HIV and TB (11, 41-47).

2.2.2 Global distribution of TB

The spatial distribution of TB is skewed heavily toward low-income and emerging economies. Most of the estimated number of cases in 2016 occurred in South-East Asian Region (45%), the African Region (25%) and Western Pacific Region (17%); smaller proportions of cases occurred in the Eastern Mediterranean Region (7%), the European Region (3%) and the Region of the Americas (3%). The 30 high-burden countries that have been given highest priority at the global level since 2000 accounted for 87% of all estimated incident cases worldwide (**Figure 1**) (1).

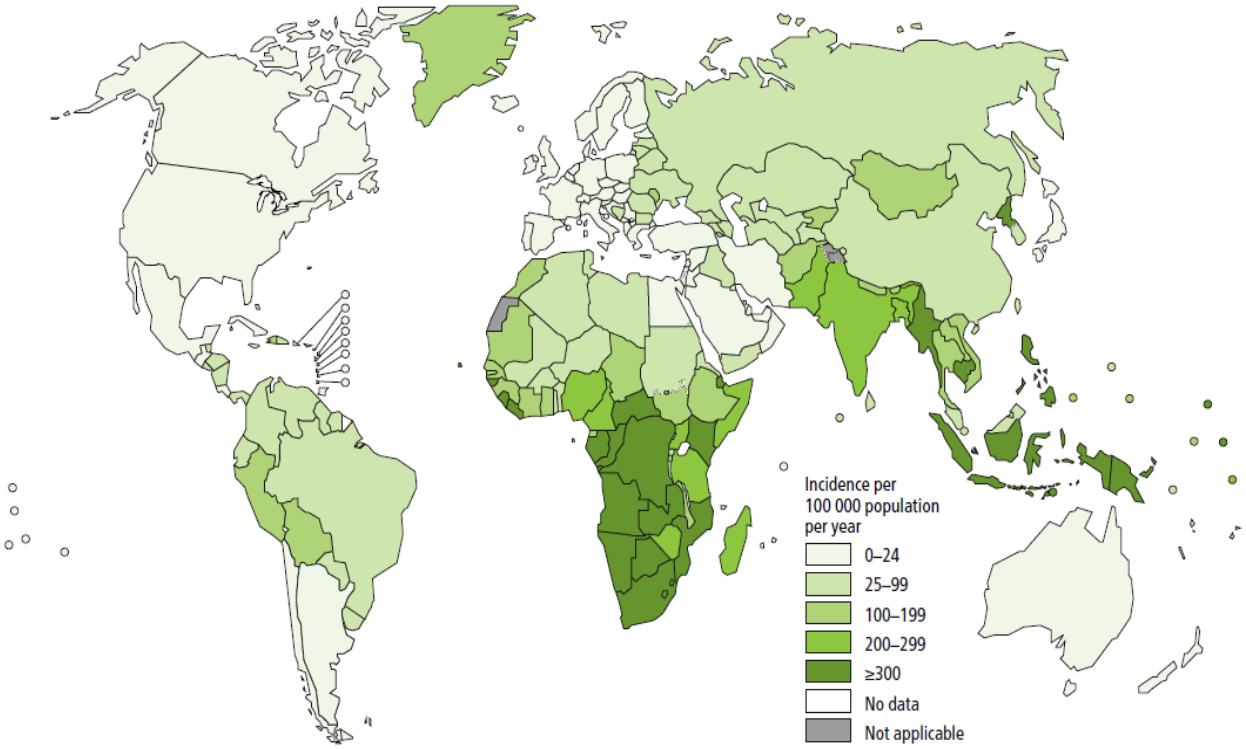


Figure 1 Estimated TB incidence rates, 2016. **Source:** The WHO global TB report of 2017 (1).

2.2.3 Spatial and space-time clustering of TB

The findings from spatial and space-time cluster analysis of TB may help local public health authorities identify geographical locations that require more attention, evaluate the impacts of intervention programs, and allocate public health resources. Few studies have been conducted to detect spatial and space-time clustering of TB in high-burden countries.

Ge E, et al. (2016) employed Moran's I, Getis-Ordi Gi*, and Kulldorff's space-time scan statistics to explore the spatial patterns of TB in Zhejiang Province, China. The disease was spatially clustering with statistically significant Moran's I values (P-value < 0.001). The most likely cluster and ten secondary clusters were identified in the province (48).

Dangisso MH, et al. (2015) conducted a study to identify spatial and space-time clusters of smear-positive pulmonary TB in Sidama Zone, southern Ethiopia. Scan statistics, Global Moran's I and Getis-Ordi Gi* statistics were used to analyze the spatial distribution and clusters of the disease. In a purely spatial analysis, they identified the most likely cluster with Relative Risk (RR) of 2. The Getis-Ordi Gi* statistic also identified the clusters in the same areas, and the spatial clusters showed stability in most areas in each year during the study period. The space-time analysis also detected the most likely cluster with a RR of 1.92 (9).

Tadesse T, et al. (2013) conducted a population-based study to detect the spatial clustering of smear-positive TB cases in Dabat District, northwest Ethiopia. A spatial scan statistic was used to identify purely spatial clusters of TB among permanent residents. Two significant spatial clusters were identified in the study district (10).

Tiwari N, et al. (2006) made a population-based study to test for the presence of statistically significant spatial clusters of TB in India. The Global Moran's I was used to test spatial clustering of TB for the period 2003 to 2005. The study showed that there were statistically significant high rates of spatial clusters of TB mainly in three areas of the district (49).

2.2.4 Estimating spatial risk of TB distribution

In low-income countries obtaining complete data that show spatial heterogeneity in the risk of TB within-and-between smaller administrative units is difficult. The existing epidemiological data are from surveys that are conducted through sampling a limited number of locations due to logistical and financial limitations. This presents a considerable obstacle to measure the disease burden at all locations (1, 50). In such settings geostatistical kriging method can be used to produce prediction estimates and associated prediction errors at all unsampled locations using data from the neighboring sampled locations (11). The method uses semivariogram models to found out weights from the sampled data. The weights dictate how each sampled value contributes to the estimated value at unsampled location (51).

2.2.5 Ecological determinants of TB distribution

Spatial unit in conventional statistics are treated as independent entity assuming that the spatial pattern of these spatial units along with its associated values do not pose any influence in the observed pattern. This assumption violates the basic assumption of spatial analysis that value in one place is influenced by values in nearby places (52). Detecting spatial dependency would help researchers justify their regression models in an ecological analysis and their smoothing techniques when mapping in small boundaries (40). Few studies adopted spatial modeling techniques to examine the relationship between TB distribution and its determinants.

Sun W, et al. (2015) explored geographic, climatic and socioeconomic factors, and employed methods of spatial statistical analyses to evaluate the role of spatial heterogeneity in the complex ecological causes of TB prevalence. Social and environmental variables were tabulated to investigate the latent factor structure of the data using exploratory factor analysis. Partial least square path modeling was used to analyze the complex causal relationship between the factors and TB prevalence. A GWR model was used to explore the local association between factors and TB prevalence. Exploratory factor analysis and partial least square path modeling indicated significant associations between TB prevalence and its latent factors. Altitude, longitude, climate, and education burden played an important role. Additionally, the GWR model showed that each latent factor had different effects on TB prevalence in different areas (53).

Li XX, et al. (2014) conducted a study to determine factors contributing to the regional inequity of TB burden in China by using an ecological approach. Latent ecological variables were identified by using exploratory factor analysis. Partial least squares path modeling was chosen to construct the structural equation model to evaluate the relationship between TB prevalence and ecological variables. Furthermore, a Geographically Weighted Regression (GWR) model was used to explore the local spatial heterogeneity in the relationships. The latent ecological variables in terms of TB prevalence, TB investment, TB service, health investment, health level, economic level, air quality, climatic factor and geographic factor were identified. With the exception of TB service and health level, other ecological factors had significant impacts on TB prevalence to varying degrees. Moreover, each ecological factor had different impacts on TB prevalence in different regions (18).

2.3 Public health, economic and social impacts of TB

TB is a public health, economic and social disaster of immense magnitude (1, 54). The WHO data show that approximately 10.4 million people become infected annually. It is currently the ninth leading cause of death globally and the leading cause of death from a single infectious agent (1).

The TB/HIV co-epidemic remains to be a major public health challenge, particularly in low-income countries. There were an estimated 1.2 million TB/HIV co-infected patients globally; about 74% of these cases were in the African Region (55). There has been strong link between TB and HIV, as they are capable of disarming the host's immune responses. TB is the most common opportunistic disease which kills those infected with HIV. Similarly, HIV co-infection increases the risk of latent TB reactivation by 20-folds (56). The proportion of smear-negative pulmonary TB and extra-pulmonary TB is high among HIV co-infected TB patients (57, 58).

Resistance of TB strains to numerous anti-TB agents is recognized as an increasing problem in many countries. In 2016, estimated 4.1% of new and 19% of previously treated TB cases had MDR-TB globally with noticeable geographic variations (1). Patients with MDR-TB must take a daily cocktail of drugs, as many as 20 pills a day, and a daily painful injection in the early stages of treatment for more than two years, resulting in social isolation, loss of employment and long-term socio-economic effects (59). Furthermore, the side effects of such treatment range from

persistent nausea to psychosis and total deafness. Some patients find the side effects too arduous to bear and interrupt or stop treatment altogether, which again can lead to extensively drug-resistant TB, resistant to Isoniazid and Rifampicin plus any fluoroquinolone and at least one of three injectable second-line drugs. On average, an estimated 6.5% of patients with MDR-TB had extensively drug-resistant TB in 2016 (1).

TB treatment requires significant amounts of scarce resources from families, government and insurance schemes (60). In 2017, an estimated US$ 9.2 billion per year is required to ensure a full response to the global TB epidemic: about 75% for detection and treatment of drug susceptible TB; 21.7% for MDR-TB and 2.1% for collaborative TB/HIV activities (1). Although TB drugs are provided free of charge in most countries, becoming ill with TB can have a substantial financial impact on patients and households because TB causes physical weakness to continue working or find work, requires regular visits to a health facility requiring time off work and incurring travel expenses, and needs nutritious food or supplementary vitamins to counter the side effects of TB drugs (61). Evidences show that TB causes as much as 40% of patient's average annual income loss, an average 3 months of work time loss, about 30% of annual household income loss, and an average 15 years of income loss if the individual dies of disease (62, 63). Catastrophic costs can lead to profound social and public health impacts. For example, children of parents ill with TB may drop out of school to seek paying work or care for parents, and some patients stop treatment before cure and may suffer worsening health, transmit disease or die. Patients often have to resort to coping mechanisms that may be irreversible: up to 75% of TB patients must take out a loan; up to 50% sell household items; and up to 66% rely on financial support from relatives (64).

Apart from economic costs TB has social costs too (59). Evidences from high-burden countries have shown that TB patients face various levels of isolation and rejection from families and communities, including loss of employment, reduced education opportunities, vulnerability to disability, divorce or spoiled marriage prospects, and isolation at home that forbids sharing food, utensils or sleeping space (54). This may influence the patients' desire to seek medical help. The fear of stigmatization makes individuals with very obvious signs and symptoms to attribute it to non-stigmatized diseases or hide the diagnosis from others. Those put on treatment may end up

defaulting from treatment because of lack of support. These may be responsible for sustained community transmission of TB and the emergence of drug resistance strains (2).

2.4 TB prevention and control strategies in Ethiopia

TB control efforts in Ethiopia were dated back to the early 1960s when a few TB centers and sanatoriums were set up in some urban settings. These efforts were not well coordinated until 1976 when the national TB control program office was established. However, the program had staff and budgetary constraints. In the early 1990s the national TB and leprosy control program was reorganized to strengthen the regional TB program units and integrate the program into the general health services. In the meantime, the TB and leprosy programs were merged to form a single program (65, 66).

Currently, Ethiopia is working towards interrupting transmission dynamics of TB, reducing morbidity and mortality, and preventing emergence and spread of drug resistance in the general population by expanding TB diagnostic and treatment services in line with the increasing number of public and private health facilities. A total of 266 hospitals, 3,622 health centers, and 16,660 health posts are providing DOTS services (13).

TB/HIV collaborative activities are being scaled-up in health centers and hospitals to provide TB patients' access to HIV testing and HIV care including antiretroviral treatment, and reducing TB burden in people living with HIV through TB screening and scale-up of Isoniazid preventive therapy (66).

TB prevention and control is one of the 16 package programs of the health service extension program at the community level. The health extension workers are engaged in awareness creation, promotion of TB prevention, better TB diagnosis, and treatment through early referral of presumptive TB cases and treatment support (67). Through the use of advocacy, communication and social mobilization, the TB program aims to promote awareness of key prevention and control strategies throughout the general population (14).

A systematic review shows that symptom screening, physical examination, chest X-ray, histology, histopathology, cytology, smear microscopy, tuberculin skin test, QuantiFERON TB gold in tube and culture are being used to diagnose TB in Ethiopia (68). The sputum smear

microscopy remains as the most accessible test for TB in the country. To strengthen the TB control program, the Ministry of Health established one National Reference Laboratory at the Ethiopian Public Health Institute, 10 laboratories with TB culture and drug susceptibility tests, and 315 laboratories equipped with Gene Xpert MTB/RIF (to detect MTB and Rifampicin resistance, which is a proxy marker for MDR-TB) (65). At the region level, the provision of laboratory reagents, supplies, and external quality assurance is managed by the regional reference laboratories (14).

According to the Ministry of Health, the treatment regimens for the new TB cases consist of Isoniazid-Rifampicin-Pyrazinamide-Ethambutol for the first two months followed by Isoniazid-Rifampicin for four months. The first two months of intensive phase treatment are under direct supervision of the health workers. During this phase, with the exception of those who are critically ill, TB patients receive treatment on ambulatory basis. Whereas, for previously treated TB cases an eight month regimen containing Streptomycin-Rifampicin-Isoniazid-Rifampicin-Ethambutol for two months followed by Isoniazid-Rifampicin-Pyrazinamide-Ethambutol for another one month during the intensive phase, followed by five months Isoniazid-Rifampicin-Ethambutol, is recommended. The standard treatment regimens for MDR-TB cases consist of Ethambuto-Pyrazinamide-Kanamycin or Amikacin-Levofloxacin-Ethionamide-Cycloserine for six months followed by Ethambutol-Pyrazinamide-Levofloxacin-Ethionamide-Cycloserine for 12 months (69).

2.5 Conceptual framework

This study is designed of three specific objectives (papers) addressing the major components of spatial epidemiology of TB (i.e., cluster analysis, spatial risk mapping and ecological analysis of risk factors) in Gurage Zone, Southern Ethiopia. Specific objective 1 (Paper I) has detected the location, size and severity of purely spatial and space-time clusters for high occurrence of TB. Specific objective 2 (Paper II) has estimated spatial risk of TB distribution co-impacted by geographic factors using data from sample locations, with findings important to measure the disease distribution at all locations. Finally, specific objective 3 (Paper III) has identified environmental and climatic factors affecting spatial distribution of TB across the study area (**Figure 2**).

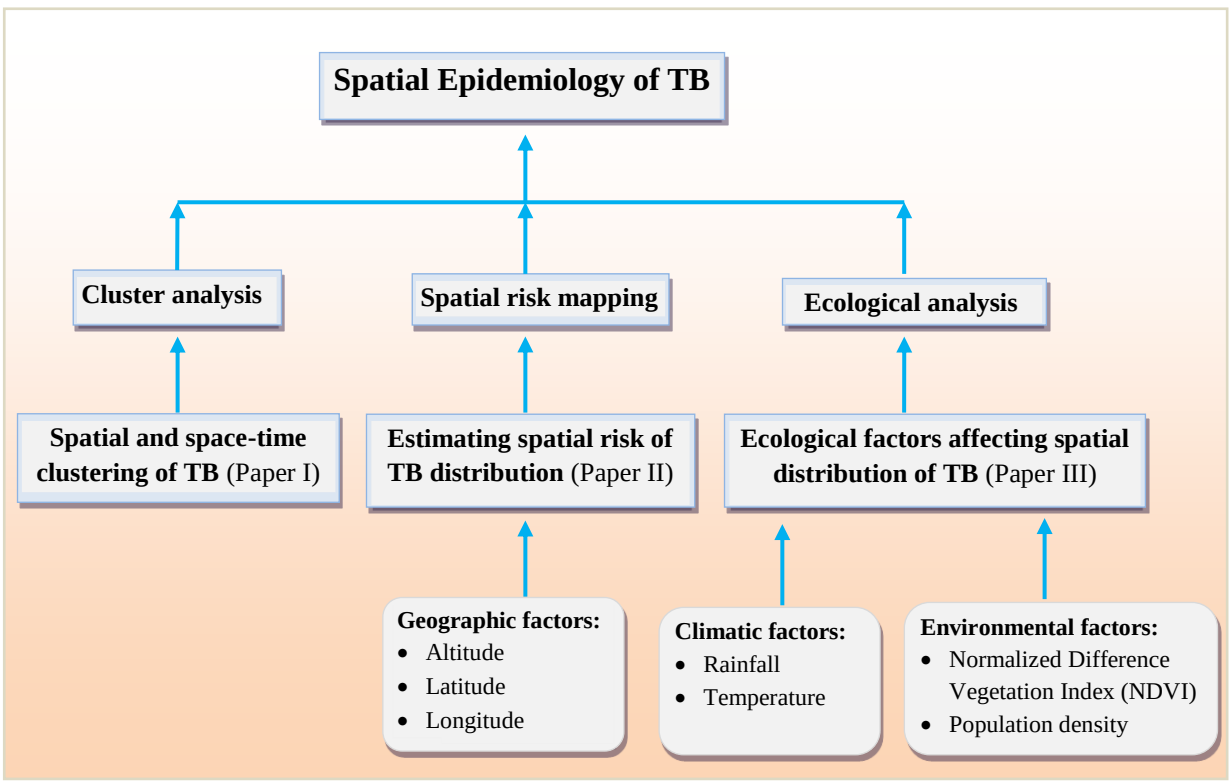


Figure 2 The conceptual framework of the study

3. OBJECTIVES

3.1 General objective

- The overall objective of this study is to assess the spatial epidemiology of TB in Gurage Zone, Southern Ethiopia.

3.2 Specific objectives

1. To detect spatial and space-time clustering of TB in Gurage Zone, Southern Ethiopia **(Paper I)**
2. To estimate spatial risk of TB distribution in Gurage Zone, Southern Ethiopia **(Paper II)**
3. To identify ecological factors affecting spatial distribution of TB in Gurage Zone, Southern Ethiopia (**Paper III**)

4. MATERIALS AND METHODS

4.1 Study area and setting

The study was conducted in the Gurage Zone, Southern Ethiopia, which is located between 7°76' and 8°45' N latitude and 37°46' and 38°71' E longitude (**Figure 3**). The zone has 13 districts, two town administrations (at Butajira and Wolkite), and 403 rural and 20 urban kebeles (the smallest administrative units with a population of 5,000 on average). It covers an area of about 5,932 km^2. According to the 2007 census, the zone has a total population of 1,279,646 (48.6% males and 51.4% females) in 2007. About 84% of the population lived in the rural areas (70).

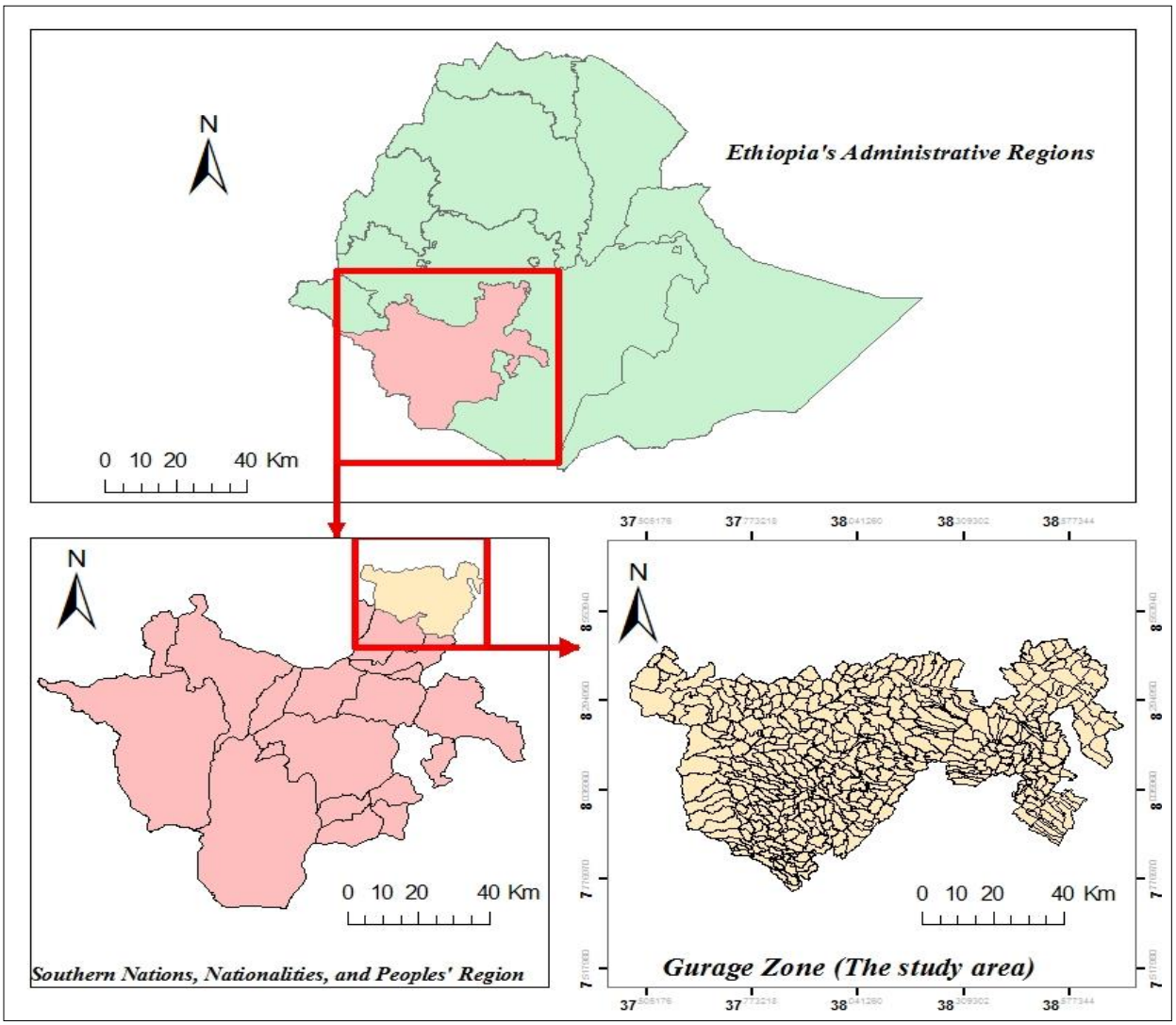


Figure 3 Map of the study area (Gurage Zone)

The zone is selected for conducting this study due to reasons described below. First, the zone has a geographic diversity characterized by mountain ranges and valleys, which may impact a geographic access to TB diagnosis and treatment centers (71). Studies show that problems of geographic access to DOTS-providing health facilities are the most important factors influencing early TB treatment initiation, compliance, and treatment outcome (72). As a result of this, it is assumed that TB cases are clustered in specific-geographic locations. Hence, implementing a uniform TB control measures may not be effective in mitigating the burden of TB in the zone. Second, according to the Southern Nations, Nationalities, and Peoples' Regional State Health Department Report of 2016 the zone was among the highest hit areas by TB in the region (73). Third, the zone covers a wider geographical area containing urban and rural areas. Fourth, the zone houses the Butajira Health and Demographic Surveillance site, which has about two decades of field experience (74).

The zone has a total of 6 hospitals, 70 health centers, 414 health posts and 92 clinics that are involved in the prevention and control of TB (73). The health posts and the clinics provide community education, identify and refer presumptive TB cases to health facilities for further investigation, give BCG vaccination, contact locating and screening, trace and link lost to follow up cases, and support treatment adherence through female health extension workers. The health centers carryout all activities as health posts and clinics, and additionally provide intensified case finding, sputum microscopy services, provide isoniazid preventive therapy for eligible persons, diagnose and manage adverse drug reactions and other complications, carry out TB/HIV collaborative activities, refer smear negative presumptive TB, extra-pulmonary TB and MDR-TB patients to higher level facilities, provide support to health post staff, keep patient records and manage medicines stocks, plan and implement TB infection control. Health centers additionally provide DOTS services for patients with MDR-TB referred by treatment initiation centers. The hospitals carry out activities as health centers, and additionally provide referral services and admission care for seriously ill TB patients. Selected hospitals provide diagnosis and treatment for MDR-TB patients, including inpatient care. The GeneXpert machines are installed at hospital laboratories. Private health facilities are also engaged in TB diagnosis, treatment and/or referral of presumptive TB and MDR-TB cases depending on their capacity (75).

4.2 Study design

An ecological study is an observational study that analyzes aggregated data to assess the burden of a disease at the population level, rather than the individual level (76). In this study, the data were aggregated at the kebele level before final analyses. The kebele was chosen as a unit of spatial analysis because it was the finest resolution used for healthcare planning in the study area, and could provide information closer to the individual level.

4.3 Study population

The study population was all TB patients who were permanent residents of the Gurage Zone and attended DOTS-providing health facilities during 2007 to 2016.

4.4 Sample size determination

All TB cases (15,805) diagnosed between 2007 and 2016 were included in specific objectives 1 and 3 (Papers I and III), and those 1,601 cases diagnosed in 2016 were used in specific objective 2 (Paper II).

4.5 Study variables

4.5.1 Dependent variable

The spatial distribution of TB prevalence rate at each kebele (the number of TB cases of a given kebele divided by the population of a given year and multiplied by 100,000) was used as a dependent variable in specific objectives 2 and 3 (Papers II and III).

4.5.2 Independent variables

The altitude, latitude and longitude were used as independent variables in specific objective 2 (Paper II), whereas normalized difference vegetation index, population density, rainfall and temperature in specific objective 3 (Paper III).

4.6 Data sources

The list of DOTS-providing health facilities were obtained from the Health Department database of Gurage Zone. TB patients' data retrieved from unit TB registers at the facilities during 2007 to 2016 were used in specific objectives 1 and 3 (Papers I and III), whereas the 2016 TB data in specific objective 2 (Paper II). A data collection format was used to retrieve the patient data (**Annex II**). The latitude, longitude and population data obtained from the Central Statistical

Agency of Ethiopia (CSA) were used for the three papers. Altitude data extracted from ASTER Global Digital Elevation Model V2 were used in specific objective 2 (Paper II). The NDVI data derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) imagery, and the temperature and rainfall data obtained from the Meteorological Agency of Ethiopia were used in specific objective 3 (Paper III).

4.7 TB diagnosis and case definition

The diagnostic criteria of the national TB diagnosis guideline of Ethiopia were used to diagnose the TB cases (66).

Smear-positive pulmonary TB (PTB+): is diagnosed when at least two initial sputum smear examinations are positive for AFB or one smear-positive result for AFB and culture-positive result for *M. tuberculosis* or one smear-positive result for AFB and radiographic abnormalities indicative of active TB, in addition to a clinician's judgment. The regional laboratory carries out external quality assurance on all slides, and provides a feedback to the health facility providing DOTS services.

Smear-negative pulmonary TB (PTB-): is diagnosed when there are symptoms evocative of TB, at least three smear-negative initial results for AFB, lack of response to antibiotics, smear-negative and radiological abnormalities indicative of pulmonary TB, and judgment of a clinician.

Extra-pulmonary TB (ETB): is diagnosed when a specimen from an extra-pulmonary site is culture positive or histo-pathological abnormality from a biopsy, and strong clinical evidence indicative of active ETB. However, because of insufficient laboratories for histo-pathological or culture examinations, most of the health facilities diagnose ETB based on a clinician's judgment.

Newly diagnosed case of TB: is a patient who had never taken anti-TB drugs or had taken for less than a month.

Retreatment TB case: is a patient who has previous treatment failure, or relapse or default.

4.8 Data management and processing

In specific objectives 1-3 (Papers I-III), data were entered, validated, cleaned, and coded using MS Excel (MicroSoft, Redmond, WA, USA). The patients' data were linked to their actual address using CSA codes to prevent duplication, and were aggregated at kebele level for spatial analyses.

In specific objectives 1 and 2 (Papers I and II), kebele centroids were used to represent a geographically weighted central location as coordinates.

In specific objective 2 (Paper II), the aggregated dataset from all the 423 kebeles in the Gurage Zone were used to examine the actual spatial risk of TB distribution. Then, the aggregated TB datasets co-impacted by geographic factors from sample of 169(40%), 254(60%) and 338(80%) kebeles were used to estimate the spatial risk of TB distribution and associated standard error by using a geostatistical kriging approach.

In specific objective 3 (Paper III), the spatial weighting matrix, which was needed to perform the spatial panel regression analyses, was constructed in Geographic Data analysis tool (GeoDa) by using the kebele-level polygon shape file. A first order queen contiguity (adjacency) weights matrix, which defines the neighbors as those with either a shared border or vertex, was used for spatial weights (77). The zero TB prevalence rates in the spatial panel dataset were replaced by one in order to make the data structure strongly balanced (i.e., all kebeles have data for all years). All the variables were log transformed since it improved estimation.

4.9 Data analyses

4.9.1 For specific objective 1 (Paper I)

Spatial autocorrelation analyses

The global Moran's I statistic was run in ArcGIS 10.2 to examine the presence of spatial clustering of TB in the whole study area. The value of Moran's I is calculated based on the deviation from the mean of two neighboring values. The following equation is used to calculate the Moran's I statistic (78, 79):

$$\mathrm{I} = \frac{n}{S_0} \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{\mathrm{j}=1}^{\mathrm{n}} \omega_{\mathrm{i,j}} \mathrm{z_i z_j}}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{z_i^2}}$$

where $\mathrm{z_i}$ is the deviation of a prevalence of TB for kebele *i* from its mean $(X_i - \bar{X})$, $\mathrm{z_j}$ is the deviation of a prevalence of TB for kebele *j* from its mean $\left(X_j - \bar{X}\right)$, $\omega_{\mathrm{i,j}}$ is the spatial weight between kebele *i* and *j*, *n* is the total number of kebeles, and S_0 is the aggregate of all the spatial weights: $S_0 = \sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{\mathrm{j}=1}^{\mathrm{n}} \omega_{\mathrm{i,j}}$. The $\mathrm{z_I}$-score for the statistic is computed as: $\mathrm{z_I} = \frac{I - E[I]}{\sqrt{V[I]}}$, where $E[I] = -1/(n-1)$ and $V[I] = E[I^2] - E[I]^2$. The spatial relationships among kebeles were conceptualized by calculating the spatial weights from the input file containing the prevalence rate of TB for each kebele and the geo-coordinates data. A first order queen polygon contiguity (adjacency) weights matrix, which defines the neighbors as those with either a shared border or vertex, was used for spatial weights (9). The spatial weighing matrix was constructed in GeoDa by using the kebele-level polygon shape file (77). Statistical significance for high occurrence of TB was decided when the Z-score ≥1.96 and a P-value ≤ 0.05.

Purely spatial and space-time cluster analyses

The Kulldroff's scan statistic was performed in SaTScan 9.2 to identify the location, size and severity of purely spatial and space-time clusters using the number of TB cases, the population for each kebele, the year of TB diagnosis and the geo-coordinates data as input files (80). The discrete Poisson model was used with the assumption that the number of TB cases at each location was Poisson distributed with a known population at risk. Scan circles of various sizes were used to identify the purely spatial clusters for high occurrence of TB. The upper limit for the maximum cluster size was set to 50% of the population at risk, which allowed small and large clusters to be detected.

To identify space-time clusters for high occurrence of TB, a cylindrical window with the circular geographic base representing to the space and height to time was used. The size of the window was limited to 50% of the expected number of TB cases, and the time was set to the time period from 2007 to 2016.

In both cases, the likelihood ratio was computed to measure a RR of TB occurrence within the cluster when compared to the risk outside using Monte Carlo simulations. The maximum number of replications for Monte Carlo simulation was set to 99,999. The null hypothesis stated that the risk of TB remains the same inside and outside the cluster. The cluster with the maximum Log-Likelihood Ratio (LLR) and containing more cases than expected was defined as the most likely cluster. The likelihood function for a specific scanning window is proportional to:

$$\left(\frac{c}{E[c]}\right)^{c}\left(\frac{C-c}{C-E[c]}\right)^{C-c} I(c > E[c])$$

where C is the total number of TB cases, c is the observed number of TB cases in the scanning window, and $E[c]$ is the expected number of TB cases in the window under the null-hypothesis, I is an indicator function that is equal to 1 where the window has more cases than those expected under the null hypothesis, and 0 otherwise.

The P-value was created using the combination of approximation. A standard of 'no geographical overlap' was selected to report secondary clusters (80). Statistical significance was reported when a P-value was ≤ 0.05.

The Getis-Ord G_i^* statistic was also implemented in ArcGIS 10.2 to identify the locations of clusters for high occurrence of TB. The G_i^* statistic performs the spatial analysis by looking at each kebele within the context of a neighboring kebele. The local sum for a kebele and its neighbors is proportionally compared to the sum of all kebeles. When the local sum is much different than the expected local sum and that difference is too large to be the result of random chance, a statistically significant Z-score result. The following equation is used to compute the G_i^* statistic (78, 81):

$$G_i^* = \frac{\sum_{j=1}^{n} \omega_{i,j} x_j - \bar{X} \sum_{j=1}^{n} \omega_{i,j}}{S \sqrt{\frac{\left[n \sum_{j=1}^{n} \omega_{i,j}^2 - \left(\sum_{j=1}^{n} \omega_{i,j}\right)^2\right]}{n-1}}}$$

where x_j is the prevalence of TB for kebele j, $\omega_{i,j}$ is the spatial weight between kebeles *i* and *j*, *n* is the total number of kebeles, $\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n}$, and $S = \sqrt{\frac{\sum_{j=1}^{n} {x_j}^2}{n} - (\bar{X})^2}$. Therefore, the G_i^* statistic is a Z-score. The spatial relationships among kebeles were conceptualized by calculating the spatial weights from the input file containing the prevalence rate of TB for each kebele and the geo-coordinates data. A first order queen polygon contiguity weights matrix, which defines the neighbors as those with either a shared border or vertex, was used for spatial weights (9). Statistical significance for high occurrence of TB was decided when the $G_i^* \geq 1.96$ and a P-value ≤ 0.05.

4.9.2 For specific objective 2 (Paper II)

Spatial smoothing

Spatial Empirical Bayes Smoothing (SEBS) method was employed in GeoDa in order to overcome small areas variance instability, which is due to variations in population size as well as few cases of TB in some areas. The population for each kebele was used as a base variable and number of TB cases was used as an event (9, 77). A queen contiguity weights matrix that defines the neighboring kebeles as those with either a shared border or vertex was used for spatial weights (9). The SEBS method was not applied for the datasets that were used for spatial prediction since the geostatistical kriging would result smoothed estimates by using a weighted linear combination of the known measured values.

Ordinary kriging

Varieties of kriging have been developed, such as ordinary, universal, simple and indicator. Ordinary kriging was preferred to other types of kriging because it predicts an estimate for unsampled kebele by assuming a constant mean in the local neighborhood of each estimation kebele, which is a characteristic of focal diseases like TB. Besides, it is a good geostatistical method to model data that exhibit spatial trend (82). It uses a semivariogram model to measure spatial autocorrelation between pairs of prevalence rates as follows (83):

$$\gamma(\mathrm{h}) = \frac{1}{2\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{Z}(\mathrm{x}) - \mathrm{Z}(\mathrm{x}+\mathrm{h})\right)^2 \quad \ldots\ldots (1)$$

where *n* is the total number of pairs of sample kebeles, *Z(x)* and *Z(x + h)* are the prevalence rates at any two kebeles *x* and *x + h* separated by distance *h*. Calculations of $\gamma(h)$ are repeated for *2h, 3h, 4h, ..., kh*. The models of spatial autocorrelation commonly exhibit similar characteristics, which are called the sill, range, and nugget. The sill is the maximum variability between pairs of prevalence rates. The separation distance at which the sill is reached is termed the range and represents the maximum distance beyond which prevalence rates are spatially independent. The nugget effect refers to the situation in which the difference between prevalence rates taken at sampling kebeles that are close together is not zero. It represents spatial sources of variation at distances smaller than the sampling interval (i.e. spatial variations of prevalence rates at village level, which is a spatial subset of kebele) or measurement error (e.g. passive case detection).

As described in detail previously (84), an unknown prevalence rate $\hat{Z}_u$ at kebele *u* is estimated as a weighted-linear combination of *n* known samples as follows:

$$\hat{Z}_u = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{W_i}\,\mathrm{Z_i} \quad \ldots\ldots (2)$$

where $\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{W_i} = 1$

The optimal weights which produce the minimum estimation error in equation (2) can be determined by using the following simultaneous equations:

$$\begin{aligned}
&\mathrm{W_1}\gamma(\mathrm{h_{1,1}}) + \cdots + \mathrm{W_n}\gamma(\mathrm{h_{1,n}}) + \lambda = \gamma(\mathrm{h_{1,u}}) \quad \ldots\ldots (3)\\
&\vdots \qquad \ddots \qquad \vdots \qquad \ddots \quad \vdots \quad \vdots\\
&\mathrm{W_1}\gamma(\mathrm{h_{n,1}}) + \cdots + \mathrm{W_n}\gamma(\mathrm{h_{n,n}}) + \lambda = \gamma(\mathrm{h_{n,u}})\\
&\mathrm{W_1} + \cdots + \mathrm{W_n} = \gamma(\mathrm{h_{n,u}})
\end{aligned}$$

where $\gamma(h_{i,j})$ is a semivariogram model which is a function of distance $h_{i,j}$ between prevalence rates *i* and *j*, and λ is the Lagrange Multiplier to minimize the kriging error.

The correlation between prevalence rates *i* and *j* is expected to decrease as their separation distance $h_{i,j}$ increases. The optimal weights in equation (2) are calculated as follows:

$$\begin{bmatrix} W_1 \\ \vdots \\ W_n \\ \lambda \end{bmatrix} = \begin{bmatrix} \gamma(h_{1,1}) & \cdots & \gamma(h_{1,n}) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(h_{n,1}) & \cdots & \gamma(h_{n,n}) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma(h_{1,u}) \\ \vdots \\ \gamma(h_{n,u}) \\ 1 \end{bmatrix} \quad \text{..... (4)}$$

Therefore, ordinary kriging produces an unbiased estimate with minimum variance.

Ordinary cokriging

Ordinary cokriging is an extension of ordinary kriging method that uses both the spatial autocorrelation for prevalence rate (i.e. the main variable of interest) and the spatial cross-correlations between prevalence rate and geographic variables (i.e. altitude, latitude and longitude) to make estimations of the prevalence rates at unsampled kebeles. The development of the ordinary cokriging system is identical to the development of ordinary kriging system. The mathematical formulation of ordinary cokriging has been described in detail by Yalcin (85).

In this objective both ordinary kriging and ordinary cokriging models were tested for the three categories of datasets, and ordinary cokriging models were selected as the best-fitted ones.

Model selection

In this study the effects of the different types of semivariogram models (i.e., stable, spherical, circular, tetraspherical, pentaspherical, Gaussian, exponential, rational quadratic, K-Bessel, hole effect and J-Bessel), detrending (i.e., neighborhood, global and local), anisotropy (i.e., false and true) and geographic covariates (i.e., longitude, latitude and altitude) on the predictive performance of kriging were checked by using a cross-validation technique. The technique leaves and adds each sample points in the dataset turn by turn to provide pairs of predicted and measured values that can be compared to evaluate the model's performance. A total of 528 geostatistical kriging models were generated for each category of spatial dataset (i.e., 40%, 60% and 80%).

The final models for each category of the spatial dataset were decided based on the lowest total error, obtained by sorting values of Root-Mean-Square Error (RMSE), absolute value of Mean-Standardized Error (MSE), Root-Mean-Square-Standardized Error (RMSSE) and absolute value of the difference of Average-Standard Error (ASE) from RMSE in ascending order, and then ranking and summing up the ranks. All these errors are expressed by equations (5)-(8) below (86):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[Z^*(x_i) - Z(x_i)]^2} \quad \text{..... (5)}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{Z^*(x_i)-Z(x_i)}{\sigma^2(x_i)}\right] \quad \text{..... (6)}$$

$$RMSSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{Z^*(x_i)-Z(x_i)}{\sigma^2(x_i)}\right]^2} \quad \text{..... (7)}$$

$$ASE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\sigma^2(x_i)} \quad \text{..... (8)}$$

where $\sigma^2(x_i)$ is the kriging variance for location x_i, and $Z^*(x_i)$ and $Z(x_i)$ are the predicted and the sampled values at the location x_i, respectively.

Sensitivity analyses

The Semivariogram Sensitivity tool, which is found under the Geostatistical Analyst toolbox of ArcGIS 10.2, was used to perform sensitivity analyses on the predicted values and associated standard errors by varying the nugget and range within a percentage of the original values. The outputs of the analyses were a table indicating which parameter values were used and what the resulting predicted and standard error values were. Small fluctuations in the output with small changes in the input parameter values indicate more confident predictions which can be used to make decisions.

4.9.3 For specific objective 3 (Paper III)

Spatial panel data models

The previous study has shown that there were statistically significant similarities in TB prevalence rates between neighboring kebeles in the Gurage Zone for each year during 2007 to 2016 (87, 88), suggesting that the traditional ordinary least squares regression should incorporate the spatial effects in the analysis to improve estimation. Therefore, this study used spatial panel data modeling techniques to estimate the effects of ecological factors (i.e., rainfall, temperature, NDVI, and population density) on spatial distribution of TB prevalence rate.

Four spatial panel data models have been compared to select the best fit one for the data analyses. These are:

1. **Spatial lag or Spatial Autoregressive model (SAR).** This model explains the interaction among prevalence rate of TB as:

   $y_t = \rho W y_t + X_t\beta + \partial + \varepsilon_t$ ..... (1)

   where y_t denotes the prevalence rate of TB, ρ denotes the spatial autoregressive (or lag) coefficient reflecting the severity of spatial interdependence in the distribution of TB prevalence between neighboring kebeles (the normal range is 0 to 1, and a high value indicated strong neighborhood effect), W denotes the spatial weight matrix describing the spatial proximity between kebeles, X_t denotes the matrix of ecological variables, β denotes the regression coefficients, ∂ denotes the kebele-specific fixed effects whose omission could bias the estimates, and ε_t denotes the error terms.

2. **Spatial Error Model (SEM).** This model specifies the interaction among error terms as:

   $y_t = X_t\beta + \partial + \varphi_t$, with $\varphi_t = \lambda W \varphi_t + \varepsilon_t$ ..... (2)

   where φ_t denotes the spatially autocorrelated error term and λ denotes the coefficient of spatial autocorrelation in error terms.

3. **Spatial Durbin Model (SDM).** This model includes spatially lagged values of TB prevalence rates and spatially weighted ecological variables from neighboring kebeles as independent variables. It reads as:

$y_t = \rho W y_t + X_t\beta + WZ_t\theta + \partial + \varepsilon_t$……...……..……...…….. (3)

where WZ_t represents the weighted average effect of the neighboring kebeles on the ecological variables and θ represents the coefficient of spatial dependence between the ecological variables.

4. **Spatial Autocorrelation model (SAC).** This model combines the SAR model with autoregressive errors, and is also known as SARAR model. It is specified as:

$y_t = \rho W y_t + X_t\beta + \partial + \varphi_t$, with $\varphi_t = \lambda W\varphi_t + \varepsilon_t$……..……...…..….. (4)

Specification tests

The Hausman test has revealed that the SDM with spatial fixed-effects specification is appropriate when compared to the SDM with random-effects specification (chi^2 = 107.95, P-value < 0.001). Next, the SAR and SEM models are compared against SDM since they are nested in SDM [24]. The SDM is preferred to the SAR (chi^2 = 93.66, P-value < 0.001) and the SEM (chi^2 = 143.84, P-value < 0.001) models. Finally, since the SDM and the SAC model are non-nesting models, the Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are used to compare them. The SAC model is proved to be the best fit model for explaining the variations in TB prevalence in the different kebeleswith AIC = 8585.79 and BIC = 8630.27 against AIC = 8732.05 and BIC = 8795.59 for SDM. The spatial panel data analyses and the specification tests were executed in Stata 14 (Stata Corp, College Station, Texas) using the XSMLE command (estimating spatial panel data using Maximum Likelihood Estimator) (89).

4.10 Data quality assurance

Supervisors and data collectors were trained on the field methods, data extraction and record keeping. The data completeness and consistency was checked page-by-page by year, health facilities, kebele, and district against unit TB registers. Errors related to the geo-coding of cases were avoided by linking each case to the correct home address using geo-codes from the CSA. Moreover, the representativeness of the study was ensured by including all DOTS-providing health facilities and cases from all kebeles in the study.

4.11 Ethical considerations

The study protocol was reviewed and approved by the Research and Ethical Committee of the School of Public Health, and the Institutional Review Board of the College of Health Sciences, Addis Ababa University. Because of the retrospective nature of the study, informed consent from patients was not required. A letter of support was obtained from the Gurage Zone Department of Health to obtain information from all districts and health facilities. The anonymity of cases was kept by using pseudo identification. Medical records were stored in a secure place to help maintain the confidentiality of the clinical information of cases. Moreover, spatial confidentiality was ensured by aggregating the patient data at kebele level **(Annex III)**.

Table 1 Summary of specific objectives and methods used for the study in Gurage Zone, Southern Ethiopia

Objectives	**Study design**	**Study population**	**Data sources**	**Data analyses**
Specific objective 1: Spatial and space-time clustering of TB	Ecological study design	All TB patients of the Gurage Zone who attended DOTS-providing facilities from 2007 to 2016	• Unit TB Register - *TB patients data* • CSA - *Population data* - *Geo-location data*	• Global Moran's I statistic • Kulldroff's scan statistic • Getis-Ord G_i^* statistic
Specific objective 2: Estimating spatial risk of TB distribution	Ecological study design	All TB patients of the Gurage Zone who attended DOTS-providing facilities in 2016	• Unit TB Register - *TB patients data* • CSA - *Population data* - *Geo-location data* • GDE Model V2 - *Altitude data*	• SEBS • Geostatistical kriging • Cross-validation technique • Sensitivity analyses
Specific objective 3: Ecological factors affecting spatial distribution of TB	Ecological study design	All TB patients of the Gurage Zone who attended DOTS-providing facilities from 2007 to 2016	• Unit TB Register - *TB patients data* • CSA - *Population data* - *Geo-location data* • MODIS imagery - *NDVI data* • Meteorol. Agency - *Rainfall data* - *Temperature data*	• Spatial panel data analyses

5. RESULTS

5.1 Specific objective 1 (Paper I): Spatial and space-time clustering of TB

5.1.1 Demographic characteristics of TB cases

A total of 16,618 TB cases were diagnosed between 2007 and 2016. Of these, 4.9% were excluded from the final analyses due to incomplete addresses or being outside of the study area. Out of the 15,805 cases included in this study 55.3% were males. The mean age with a standard deviation of the cases was 34.0 ± 16.4. About 93.2% of the cases were newly diagnosed, while 6.8% were retreatment cases. Fifteen percent of the cases were from urban areas. The prevalence of TB varied from 70.4 to 155.3 cases per 100,000 population during 2007 to 2016 **(Figure 4)**.

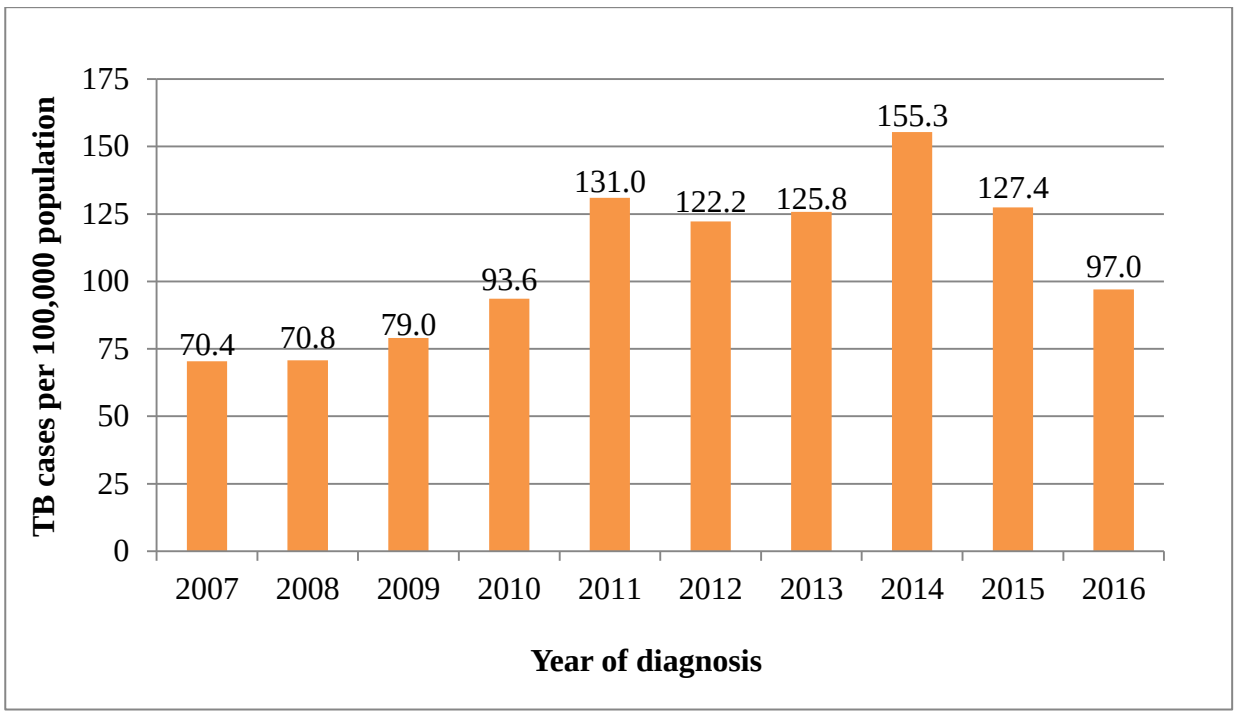


Figure 4 Trend of TB prevalence in Gurage Zone, Sothern Ethiopia

5.1.2 Spatial autocorrelation

The global Moran's I statistic was significant for each year (P-value < 0.001), implying that there were spatial and space-time clustering in the distribution of TB in the Gurage Zone **(Table 2)**.

Table 2 Global spatial autocorrelation of TB distribution in Gurage Zone, Southern Ethiopia, 2007-2016

Year	Moran's I	Z-score	P-value	Pattern
2007	0.266568	12.3	<0.001	Clustered
2008	0.143781	6.8	<0.001	Clustered
2009	0.194124	9.6	<0.001	Clustered
2010	0.288589	13.4	<0.001	Clustered
2011	0.191062	9.1	<0.001	Clustered
2012	0.161519	7.7	<0.001	Clustered
2013	0.116817	5.4	<0.001	Clustered
2014	0.130803	6.1	<0.001	Clustered
2015	0.174448	8.2	<0.001	Clustered
2016	0.098785	4.6	<0.001	Clustered

5.1.3 Purely spatial clusters

The purely spatial cluster analyses identified the clusters for high occurrence of TB at peripheral areas of the geographic zone. The most likely cluster with 291 TB cases (71.19 expected cases) was detected at southwest of Abeshege district. The size of the cluster was within a radius of 4.45 km. People within this cluster had about four times higher risk of TB infection than those outside the cluster (LLR = 192.87, P-value < 0.001). Eleven significant secondary clusters for high occurrence of TB were also detected (LLR = 2.56-135.75, P-value < 0.001). Most of these locations belong to the districts of Gumer, Endegagn, Enemor Ener, eastern Abeshege, Kebena, Welkite Town, northern Ezha, northwestern and eastern Cheha, Meskan, Butajira Town, northern Merako, southwest Sodo (**Table 3, Figure 5**).

Table 3 Purely spatial clusters for high occurrence of TB in Gurage Zone, Southern Ethiopia, 2007-2016

Cluster type	Cluster Year	Cluster center/radius	Observed cases	Expected cases	LLR	RR	P-value
Most likely cluster	2007-2016	(8.27 N, 37.58 E) / 4.45 km	291	71.19	192.87	4.16	<0.001
Secondary cluster	2007-2016	(8.12 N, 38.53 E) / 16.50 km	3325	535.65	135.75	1.39	<0.001
2nd secondary	2007-2016	(7.85 N, 37.82 E) / 9.97 km	1365	998.14	64.99	1.40	<0.001
3rd secondary	2007-2016	(8.33 N, 37.88 E) / 12.32 km	1253	942.62	49.53	1.36	<0.001
4th secondary	2007-2016	(8.05 N, 37.64 E) / 14.69 km	685	460.81	49.21	1.51	<0.001
5th secondary	2007-2016	(7.97 N, 38.06 E) / 3.98 km	340	213.02	32.51	1.61	<0.001
6th secondary	2007-2016	(8.24 N, 38.06 E) / 0 km	36	8.93	23.14	4.04	<0.001
7th secondary	2007-2016	(8.14 N, 38.31 E) / 1.56 km	218	136.27	20.91	1.61	<0.001
8th secondary	2007-2016	(8.18 N, 37.85 E) / 0 km	58	23.55	17.86	2.47	<0.001
9th secondary	2007-2016	(8.32 N, 38.55 E) / 0 km	140	86.04	14.29	1.63	<0.001
10th secondary	2007-2016	(8.41N, 38.24 E) / 0 km	44	17.22	2.56	14.52	<0.001

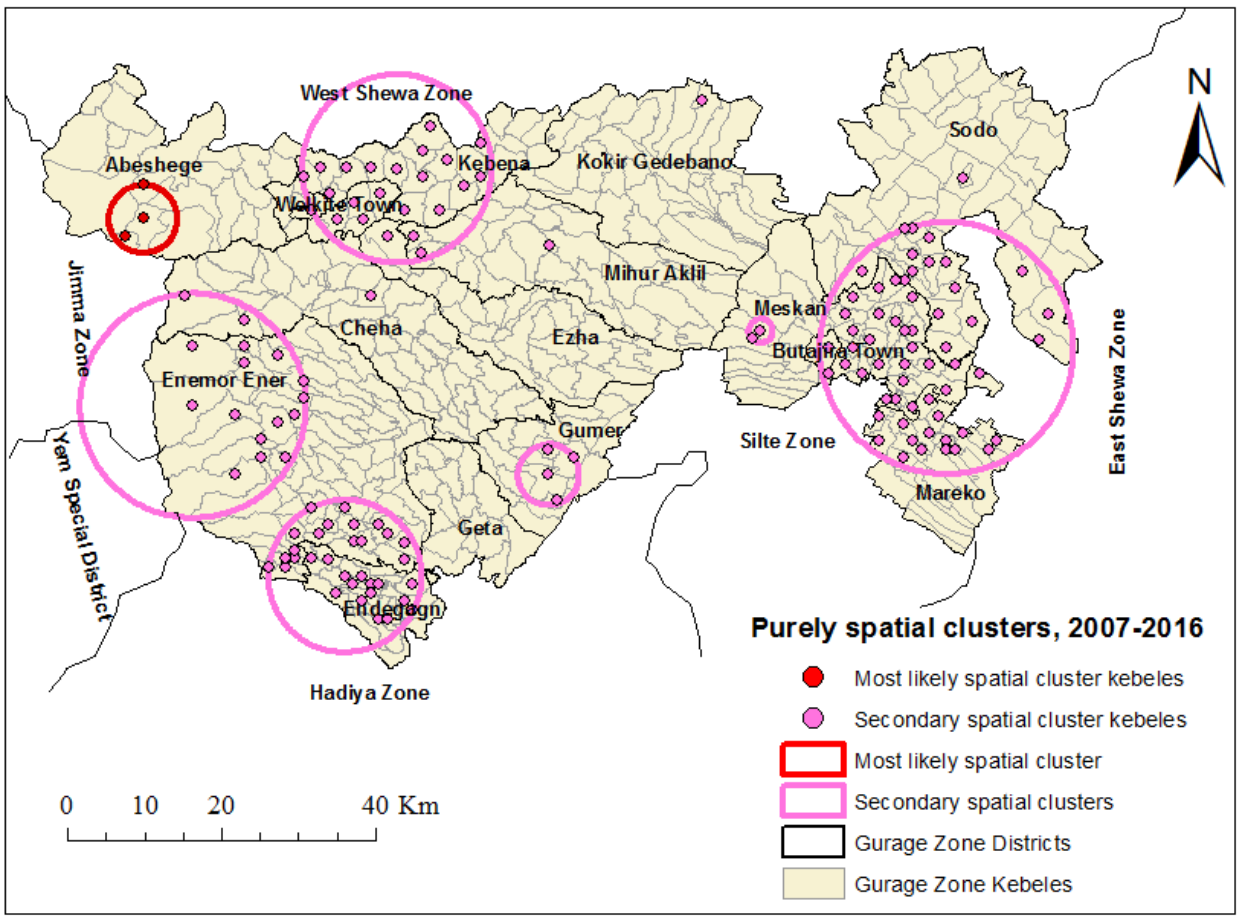


Figure 5 Purely spatial clusters for high occurrence of TB in Gurage Zone, Southern Ethiopia

The nature of the clusters for high occurrence of TB in the study area was evaluated for each year from 2007 to 2016. There were considerable spatial variations in the risk of TB by year **(Table 4, Figure 6).** The G_i^* statistic showed spatial variations in the risk of TB at nearly similar geographic locations **(Figure 7)**.

Table 4 Annual purely spatial clusters for high occurrence of TB in Gurage Zone, Southern Ethiopia, 2007-2016

Cluster type	Cluster Year	Cluster center/radius	Observed cases	Expected cases	LLR	RR	P-value
Most likely cluster	2007	(8.15 N, 38.17 E) / 32.07 km	635	406.44	118.53	2.91	<0.001
Secondary cluster	2007	(8.29 N, 37.51 E) / 14.61 km	86	24.92	47.64	3.71	<0.001
Most likely cluster	2008	(8.36 N, 37.49 E) / 14.45 km	101	22.78	75.66	4.85	<0.001
Secondary cluster	2008	(8.28 N, 37.99 E) / 23.45 km	325	215.73	32.70	1.78	<0.001
2nd secondary	2008	(8.23 N, 38.49 E) / 4.54 km	44	12.91	23.40	3.53	<0.001
3rd secondary	2008	(7.94 N, 38.07 E) / 0 km	14	2.13	14.56	6.65	<0.001
4th secondary	2008	(7.84 N, 37.83 E) / 7.78 km	68	34.01	13.78	2.08	<0.001
5th secondary	2008	(8.14 N, 38.31 E) / 1.56 km	25	8.04	11.55	3.17	0.004
Most likely cluster	2009	(8.43 N, 37.56 E) / 47.88 km	382	194.53	92.08	2.50	<0.001
Secondary cluster	2009	(7.89 N, 37.84 E) / 6.97 km	85	39.00	21.26	2.28	<0.001
2nd secondary	2009	(8.22 N, 38.51 E) / 7.02 km	61	26.13	17.44	2.42	<0.001
3rd secondary	2009	(8.14 N, 38.31 E) / 1.56 km	28	9.20	12.53	3.10	<0.001
4th secondary	2009	(8.33 N, 38.20 E) / 11.49 km	100	60.88	11.28	1.71	0.005
5th secondary	2009	(7.97 N, 38.11 E) / 9.16 km	71	39.17	10.90	1.87	0.006
Most likely cluster	2010	(7.86 N, 37.73 E) / 12.99 km	137	54.72	46.24	2.68	<0.001
Secondary cluster	2010	(7.97 N, 38.11 E) / 5.98 km	89	31.16	36.90	2.99	<0.001
2nd secondary	2010	(8.25 N, 38.51 E) / 10.06 km	111	52.84	25.61	2.20	<0.001
3rd secondary	2010	(8.06 N, 38.46 E) / 3.13 km	60	21.73	23.24	2.85	<0.001
4th secondary	2010	(8.13 N, 38.30 E) / 0 km	23	6.54	12.56	3.56	<0.001
5th secondary	2010	(8.31 N, 37.96 E) / 4.93 km	35	14.97	9.85	2.38	0.017
Most likely cluster	2011	(8.13 N, 38.64 E) / 23.49 km	507	267.67	103.35	2.23	<0.001
Secondary cluster	2011	(8.36 N, 37.49 E) / 14.45 km	109	45.68	32.59	2.47	<0.001
2nd secondary	2011	(8.03 N, 37.83 E) / 0 km	26	7.19	14.71	3.65	<0.001
3rd secondary	2011	(7.84 N, 37.83 E) / 1.56 km	27	7.84	14.32	3.48	<0.001
4th secondary	2011	(8.17 N, 38.25 E) / 7.07 km	72	39.60	10.94	1.85	0.006
Most likely cluster	2012	(8.00 N, 38.58 E) / 27.86 km	525	319.81	69.98	1.91	<0.001
Secondary cluster	2012	(8.12 N, 37.76 E) / 4.58 km	60	21.56	23.39	2.84	<0.001
2nd secondary	2012	(8.26 N, 37.73 E) / 5.93 km	93	57.16	9.80	1.66	0.018
Most likely cluster	2013	(8.06 N, 38.47 E) / 11.50 km	380	231.48	46.66	1.80	<0.001
Secondary cluster	2013	(7.86 N, 37.75 E) / 2.20 km	42	12.21	22.33	3.49	<0.001
2nd secondary	2013	(8.28 N, 38.53 E) / 4.96 km	49	21.85	12.62	2.28	<0.001
3rd secondary	2013	(8.27 N, 37.58 E) / 4.45 km	25	8.55	10.44	2.95	<0.011
Most likely cluster	2014	(8.00 N, 38.50 E) / 4.96 km	151	52.54	63.04	3.00	<0.001
Secondary cluster	2014	(8.33 N, 37.85 E) / 3.30 km	48	11.27	33.11	4.33	<0.001
2nd secondary	2014	(8.10 N, 37.70 E) / 11.50 km	144	87.60	15.87	1.68	<0.001
3rd secondary	2014	(8.19 N, 37.94 E) / 3.98 km	46	22.48	9.53	2.07	0.024
Most likely cluster	2015	(8.06 N, 38.51 E) / 17.92 km	481	334.77	34.58	1.57	<0.001
Secondary cluster	2015	(8.18 N, 37.63 E) / 40.94 km	843	688.91	25.27	1.38	<0.001
2nd secondary	2015	(7.82 N, 37.89 E) / 3.13 km	43	18.53	11.88	2.35	0.003
Most likely cluster	2016	(8.29 N, 37.51 E) / 9.92 km	66	23.11	26.96	2.94	<0.001
Secondary cluster	2016	(7.91 N, 37.86 E) / 3.98 km	57	21.01	21.32	2.78	<0.001
2nd secondary	2016	(8.00 N, 38.58 E) / 25.06 km	343	251.21	18.24	1.47	<0.001
3rd secondary	2016	(8.33 N, 37.88 E) / 12.32 km	148	95.13	13.49	1.61	<0.001
4th secondary	2016	(8.05 N, 37.64 E) / 7.78 km	26	9.82	9.22	2.68	0.030

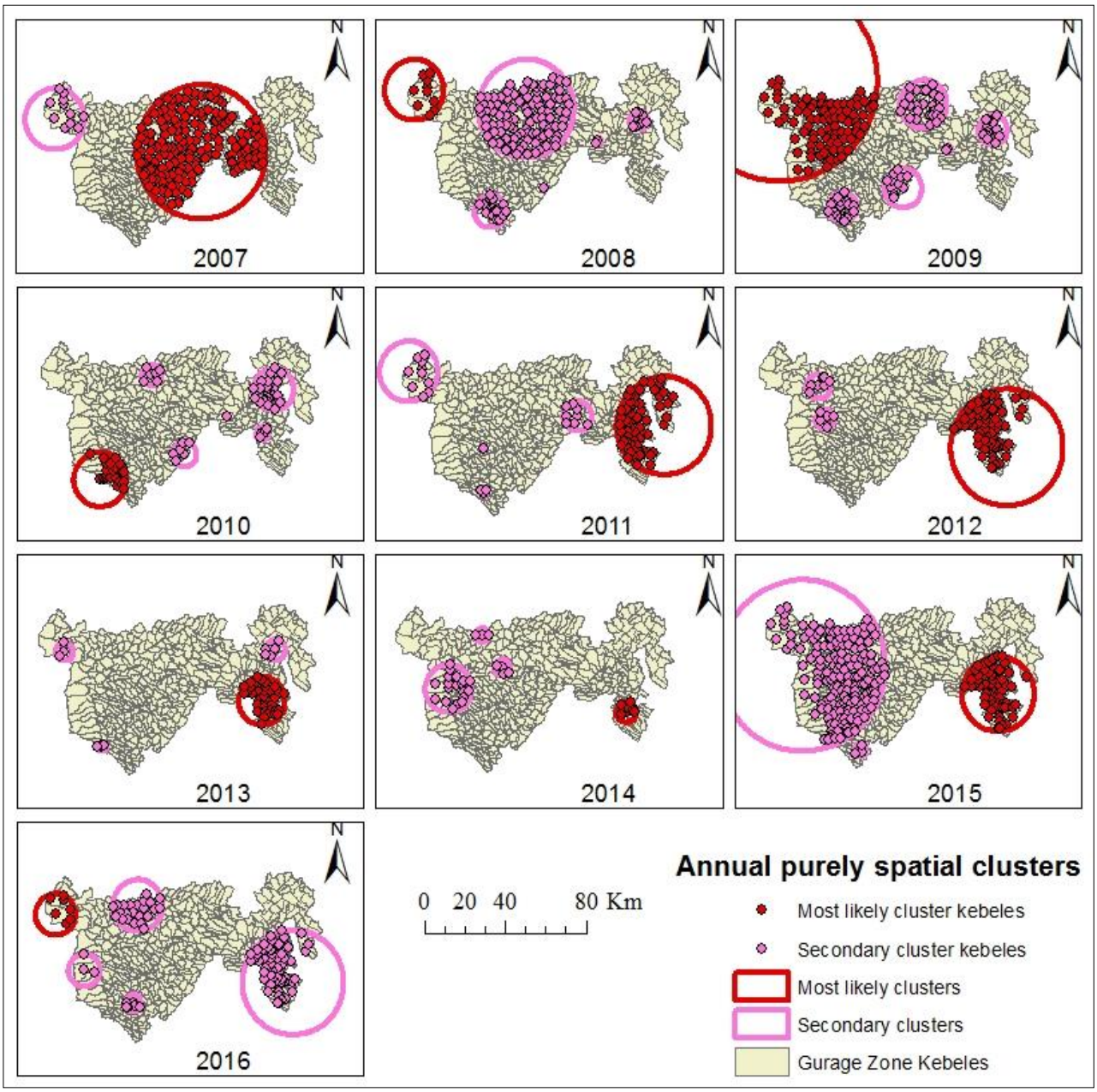


Figure 6 Annual purely spatial clusters for high occurrence of TB identified by using SaTScan statistic in Gurage Zone, Southern Ethiopia

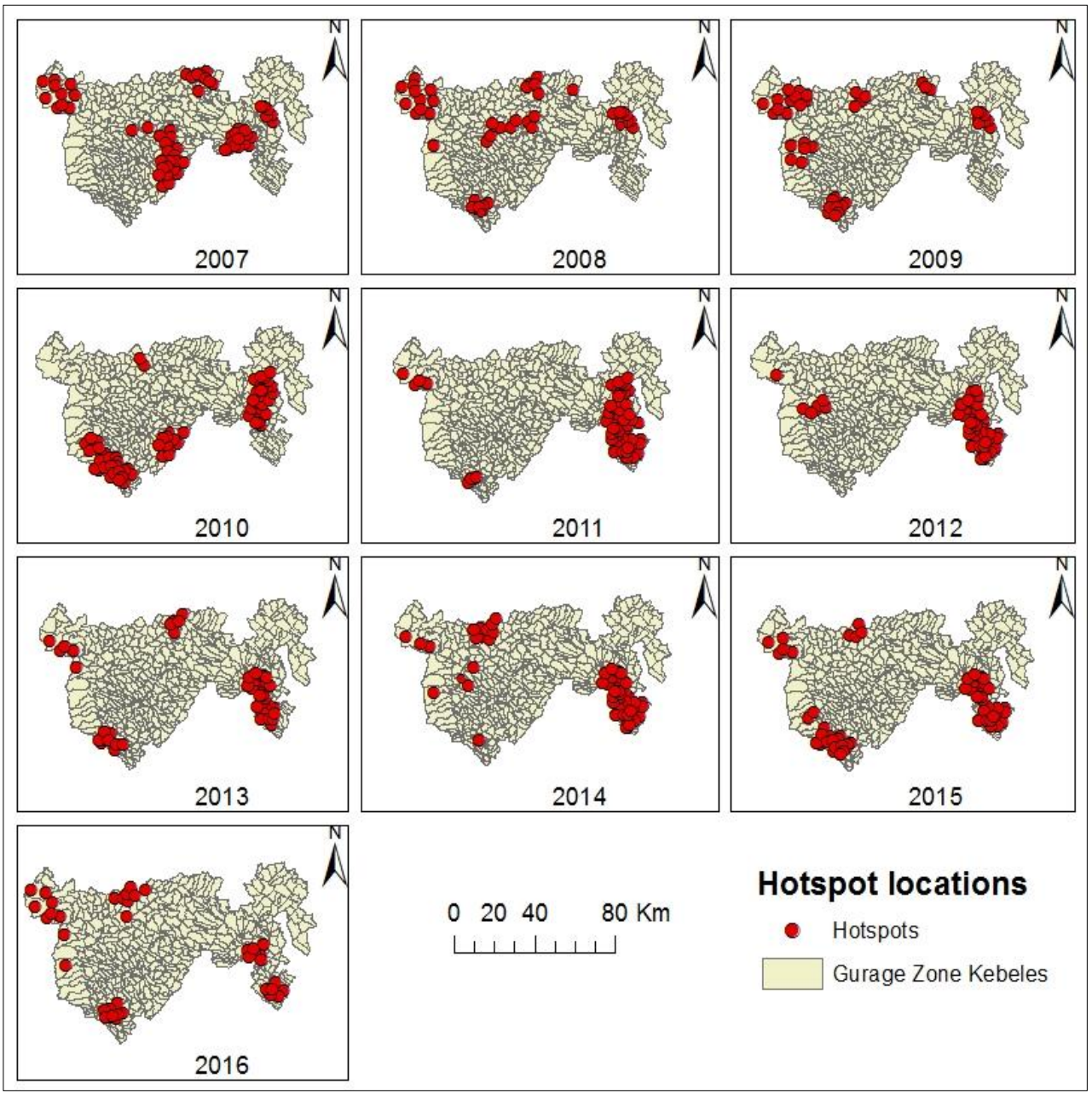


Figure 7 Spatial locations of significant hotspots of TB identified by using Getis-Ord G_i^* statistic in Gurage Zone, Southern Ethiopia

5.1.4 Space-time clusters

The space-time cluster analysis identified the clusters for high occurrence of TB at peripheral areas of the study area. The most likely cluster with 2,502 observed cases (1,353.51 expected cases) was detected for the period 2011 to 2015. The cluster covered Meskan, Butajira Town, Merako and southwest of Sodo areas within a radius of 17.92 km. People living within this cluster had about two times higher risk of TB infection than those outside the cluster (LLR = 435.61, P-value < 0.001). A secondary cluster covering Endegagn, Enemor Ener, Cheha, Abeshege, Welkite Town, weastern Kebena, northern Ezha and western Mihur Aklil areas was detected for the period 2011 to 2015 (LLR = 202.81, P-value < 0.001). The second secondary cluster was detected at central Gumer district for the period 2007 to 2011 (LLR = 37.64, P-value < 0.001) **(Table 5, Figure 8)**.

Table 5 Space-time clusters for high occurrence of TB in Gurage Zone, Southern Ethiopia, 2007-2016

Cluster type	Year	Cluster center/radius	Observed cases	Expected cases	LLR	RR	P-value
Most likely cluster	2011-2015	(8.06 N, 38.51 E) / 17.92 km	2502	1353.51	435.61	2.01	<0.001
Secondary cluster	2011-2015	(8.12 N, 37.64 E) / 39.95 km	4117	3076.12	202.81	1.46	<0.001
2nd secondary	2007-2011	(7.97 N, 38.06 E) / 3.98 km	197	99.21	37.64	2.00	<0.001

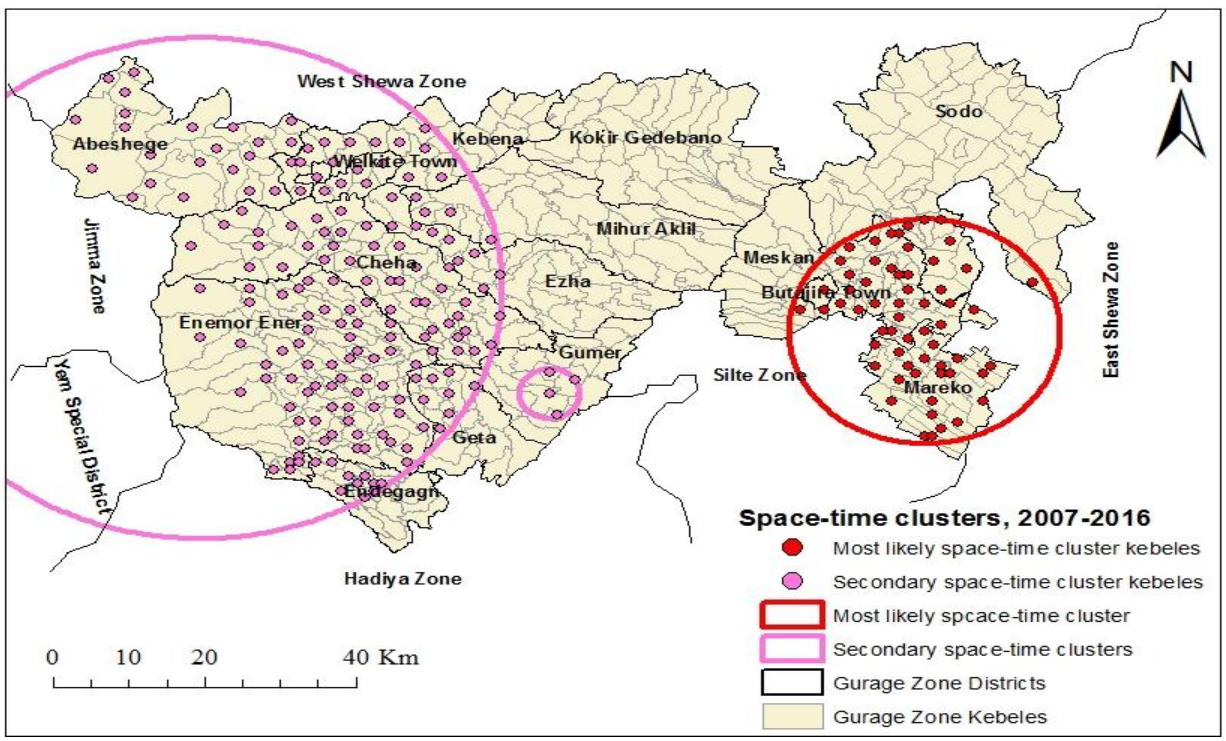


Figure 8 Space-time clusters for high occurrence of TB in Gurage Zone, Southern Ethiopia

5.2 Specific objective 2 (Paper II): Estimating spatial risk of TB distribution

5.2.1 Patient characteristics

A total of 1,626 TB cases were diagnosed during January to December, 2016. Only 1.5% of them were excluded from the final analyses because of incomplete addresses or being outside of the study area. Out of 1,601 cases included in this study 57.5% were males and 42.5% were females, yielding a male to female ratio of 1.3:1. The mean age with a standard deviation was 36 ± 17 years for all cases (34 ± 16 years for males and 38 ± 17 years for females). About 89.6% of the cases were newly diagnosed, while 10.4% were retreatment cases. Of the cases 41.2% were PTB+, 31.9% PTB- and 26.9% ETB. Residentially, 86.6% were from rural areas (**Table 6)**.

Table 6 Characteristics of TB patients in Gurage Zone, Southern Ethiopia, 2017 (n = 1,601)

Variables	Number	Percent
Sex		
Male	920	57.5
Female	681	42.5
Residence		
Rural	1387	86.6
Urban	214	13.4
Category of TB		
Newly diagnosed	1435	89.6
Retreatment	166	10.4
Type of TB		
PTB+	661	41.2
PTB-	510	31.9
ETB	430	26.9

5.2.2 The actual spatial risk of TB distribution

The risk distribution of TB varied from 59 to 173 cases per 100,000 population across districts of the Gurage Zone (**Figure 9**). Moreover, the smoothed rates of TB varied from zero to 634 cases per 100,000 population across kebeles of the zone. High risk of TB was observed at northwest, western, southwest and southeast parts. The risk distribution crossed between high and low at west-central parts (**Figure 10**).

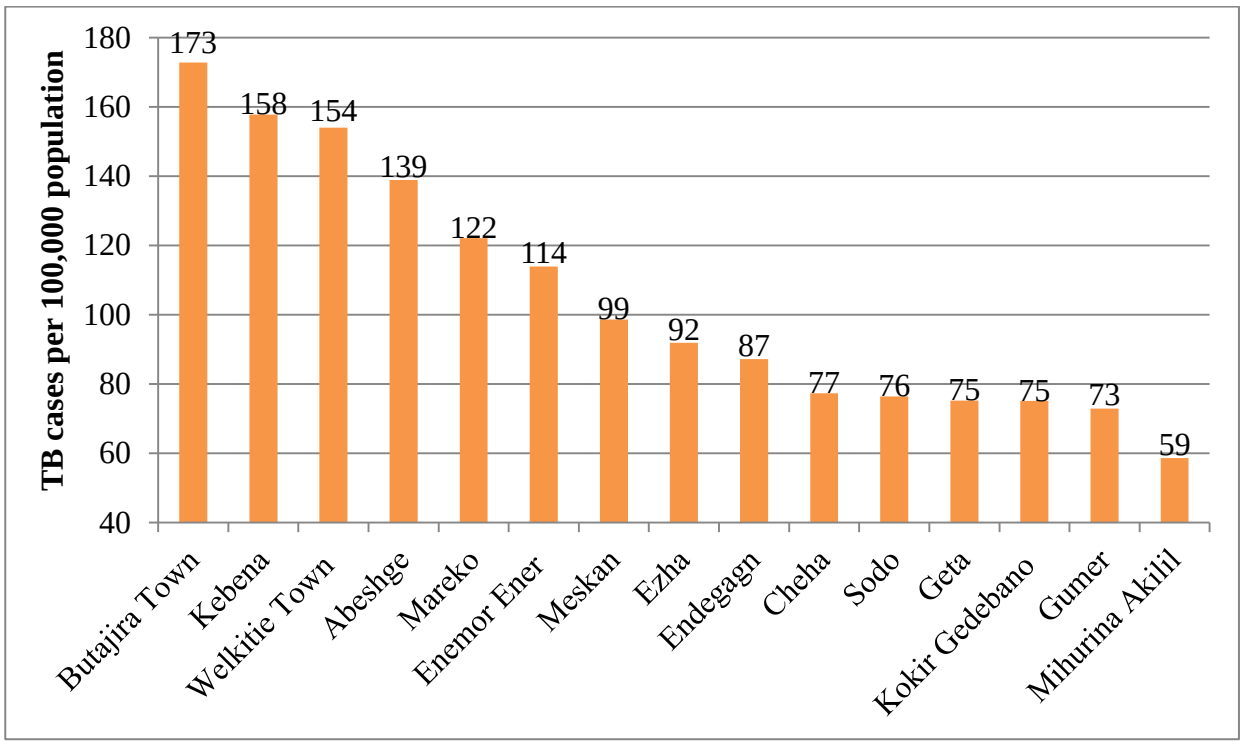


Figure 9 TB prevalence rates by districts in Gurage Zone, Southern Ethiopia

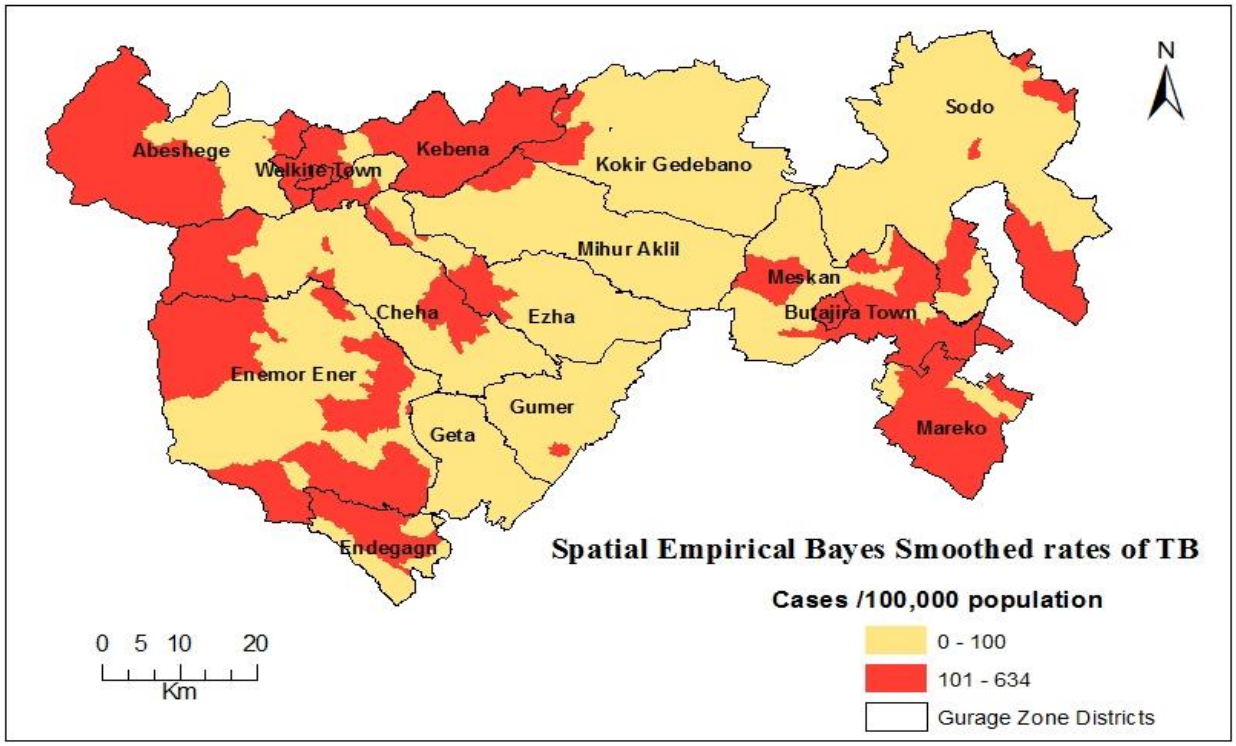


Figure 10 The SEBS rates of TB in Gurage Zone, Southern Ethiopia

5.2.3 The best-fitted models

The best geostatistical kriging models were decided to be: 1) Pentaspherical semivariogram, local detrending, true anisotropy and altitude and latitude covariates for modeling with 40% of spatial dataset, 2) Rational Quadratic semivariogram, local detrending, true anisotropy and altitude and latitude covariates for modeling with 60% of spatial dataset, and 3) K-Bessel semivariogram, global detrending, true anisotropy and altitude and latitude covariates for modeling with 80% of spatial dataset. The detrending pattern of the models changed from local to global as the size of spatial dataset increased. Moreover, the models predictive accuracies also improved as the size of spatial dataset increased, which was indicated by 0 MSE, 1 RMSSE, and ASE approached RMSE (i.e., the variability in prediction is correctly assessed) (**Table 7**).

Table 7 Comparison of cross-validation statistics for TB spatial datasets in Gurage Zone, Southern Ethiopia, 2017

Cross-validation statistics	Ordinary cokriging models		
	With 40% dataset	With 60% dataset	With 80% dataset
MSE	0	0	0
RMSSE	1	1	1
RMSE	89	88	87
ASE	93	87	87

5.2.4 Sensitivity analyses outputs

The parameter values for nugget and range from the input geostatistical model sources were 8123.85 and 72891.45 for the model with 40% spatial dataset, 7178.77 and 78808.04 for the model with 60% spatial dataset, and 7210.46 and 78767.27 for the model with 80% spatial dataset, respectively. Five random nugget and range values that were found within 10% of the input models' nugget and range values were calculated for each dataset and used as input parameters. There were only small fluctuations in the prediction outputs for the corresponding input parameters, indicating more accurate predictive performance of the models (**Table 8**).

Table 8 The semivariogram sensitivity analyses results for TB spatial datasets in Gurage Zone, Southern Ethiopia, 2017

Model	Random Parameter	Prediction	Standard Error	Nugget	Range
Modeling with 40%	Nugget	79.00	22.95	7898.75	72891.45
Dataset	Nugget	78.99	23.34	8170.68	72891.45
	Nugget	79.01	22.66	7702.39	72891.45
	Nugget	79.01	22.35	7467.51	72891.45
	Nugget	79.02	22.47	7572.18	72891.45
	Range	78.99	23.18	8062.43	76769.04
	Range	78.65	23.19	8068.52	79894.51
	Range	78.66	23.17	8050.92	72346.21
	Range	78.66	23.17	8051.90	72671.44
	Range	78.65	23.19	8068.39	79821.99
Modeling with 60%	Nugget	67.35	21.45	6995.20	78808.04
Dataset	Nugget	67.63	20.94	7360.37	78808.04
	Nugget	67.57	21.05	7284.92	78808.04
	Nugget	67.66	20.90	7425.62	78808.04
	Nugget	67.68	21.24	7669.06	78808.04
	Range	67.44	21.16	7018.50	80198.91
	Range	83.57	21.03	6920.55	73192.36
	Range	67.55	21.00	7076.75	85175.07
	Range	77.54	20.82	6979.91	77256.03
	Range	67.55	21.00	7076.00	85106.29
Modeling with 80%	Nugget	77.35	20.49	7017.76	78767.27
Dataset	Nugget	77.46	20.86	7356.95	78767.27
	Nugget	77.51	21.51	7865.18	78767.27
	Nugget	77.20	20.02	6603.89	78767.27
	Nugget	77.45	20.84	7332.31	78767.27
	Range	77.82	20.60	7212.61	80939.15
	Range	77.82	20.60	7212.40	80888.19
	Range	74.71	19.43	7172.59	72233.82
	Range	74.71	19.44	7179.62	73658.34
	Range	77.82	20.61	7221.91	83206.47

5.2.5 The estimated spatial risk of TB distribution

The ordinary cokriging models with 40%, 60% and 80% of the spatial datasets estimated high risk of TB at northwest, western, southwest and southeast parts of the Gurage Zone. The risk distribution crossed between high and low at west-central parts**.** Moreover, the models estimated high uncertainties of prediction at border areas of the zone, the magnitude of which decreased as the spatial dataset increased (**Figure 11**).

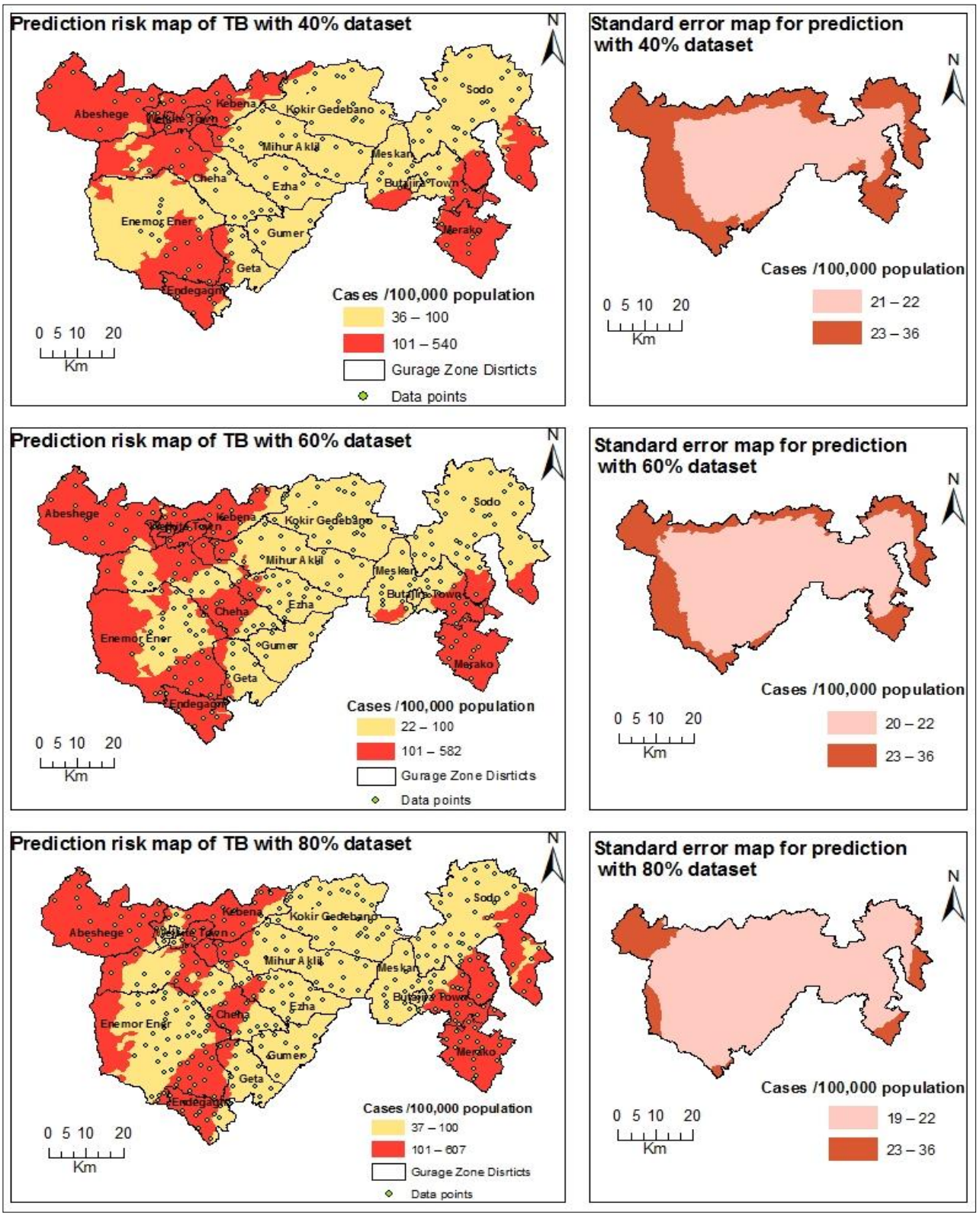


Figure 11 The prediction risk maps of TB and associated standard error maps in Gurage Zone, Southern Ethiopia

5.2.6 Comparison between estimated and actual spatial risk of TB distribution

The three estimation models identified areas for high risk of TB at locations that were closely similar to the actual high-risk areas, with reasonable predictive accuracies. These locations included northwest, western, southwest and southeast parts of the Gurage Zone. The risk distribution crossed between high and low at west-central parts (see **Figures 10 and 11 above**).

5.3 Specific objective 3 (Paper III): Ecological factors affecting spatial distribution of TB

5.3.1 Characteristics of TB cases

From 16,618 TB cases diagnosed at all DOTS-providing health facilities in Gurage Zone from 2007 to 2016, 4.9% were excluded from the final analyses due to incomplete addresses or being outside of the zone. Out of the 15,805 cases included in the final analyses 55.3% were males. A high proportion (85.8%) of the cases belonged to the working age group. About 93.2% of the cases were newly diagnosed. Pulmonary TB comprised more than two-thirds of all the cases. About 85.3% of the cases were diagnosed at the health centers in the zone **(Table 9)**. Higher prevalence of TB was observed in urban areas compared to the rural areas during 2007 to 2016 **(Figure 12)**.

Table 9 Demographic and clinical characteristics of TB cases in Gurage Zone, Southern Ethiopia, 2007-2016

Variables	Number	Percent
Sex		
Male	8742	55.3
Female	7063	44.7
Age (in years)		
≤ 14	1447	9.2
15-34	6888	43.6
35-64	6665	42.2
≥ 65	805	5.1
DOTS-providing facility		
Health center	13480	85.3
Hospital	2325	14.7
Category of TB		
Newly diagnosed	14736	93.2
Retreatment	1069	6.8
Type of TB		
PTB+	5197	32.9
PTB-	6049	38.3
EPT	4559	28.8

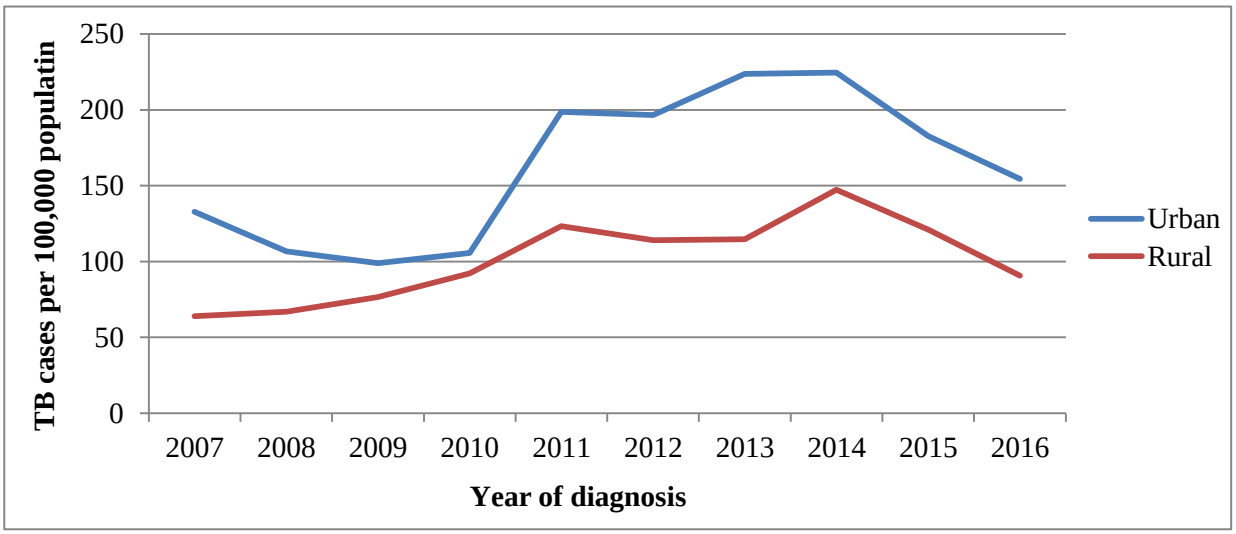


Figure 12 Trend of TB prevalence in urban and rural areas of Gurage Zone, Sothern Ethiopia

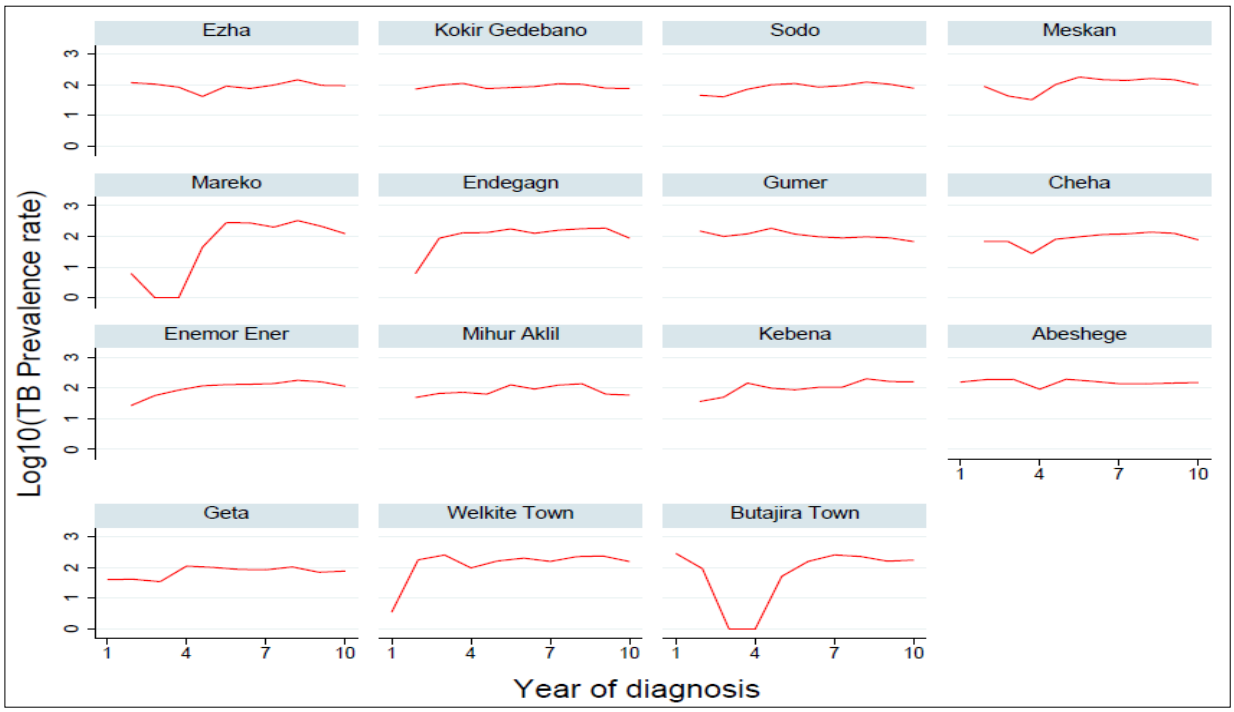
5.3.2 Description of the panel data

The panel data included 425 kebeles that were observed for the period 2007 to 2016. There was higher variability in rainfall and temperature within kebeles compared to between kebeles. The variability in NDVI and population density was observed to be higher between kebeles compared to within kebeles. A higher variability in prevalence rate of TB was observed within kebeles compared to between kebeles during the study period (**Table 10**). This variability was lower at district level compared to the variability at kebele level (**Figure13)**.

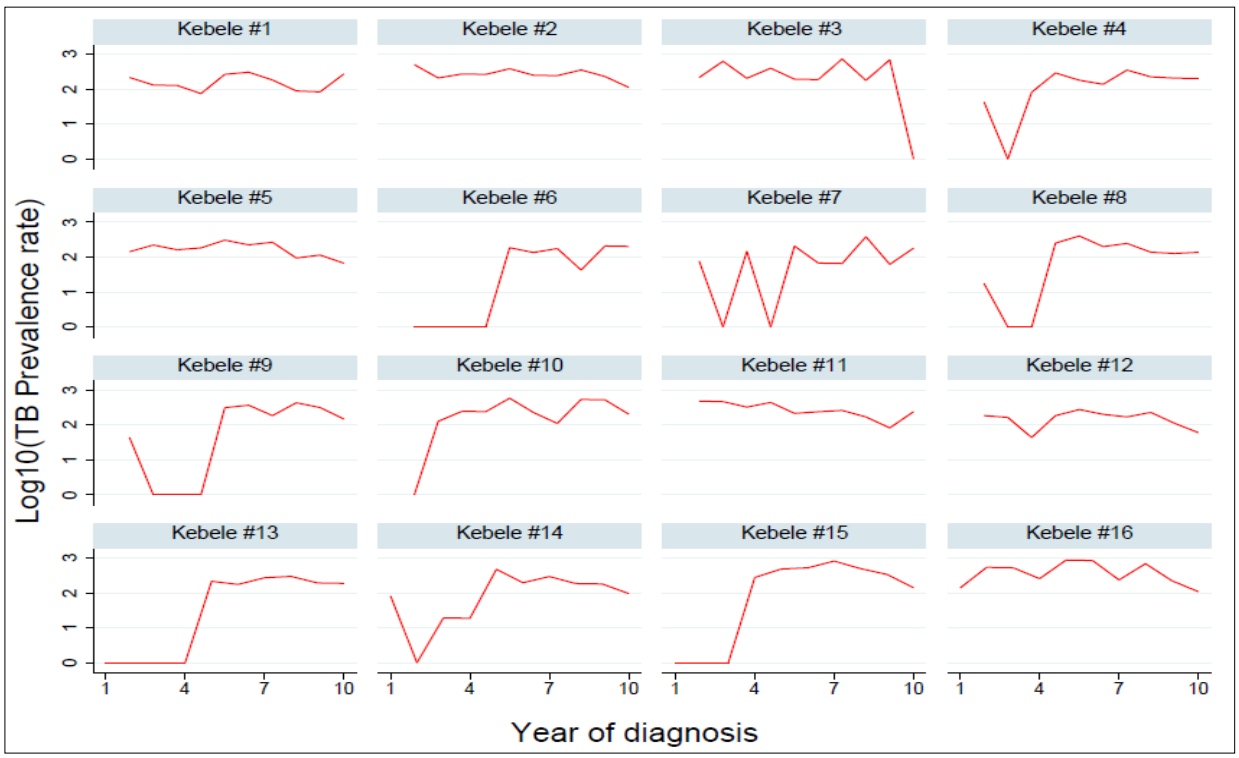
Table 10 Summary of panel data used for the study in Gurage Zone, Southern Ethiopia, 2007-2016

Variable	Variation	Mean	Std. Dev.	Min	Max	Observation*
Rainfall	Overall	89.9	21.5	31.9	147.9	N = 4250
	Between		14.7	59.7	121.0	n = 425
	Within		15.7	36.5	126.4	T = 10
Temperature	Overall	18.9	2.1	9.2	69.0	N = 4250
	Between		1.8	14.5	25.6	n = 425
	Within		1.0	11.8	64.7	T = 10
NDVI	Overall	5493.4	806.0	2718.8	7017.8	N = 4250
	Between		772.8	3159.6	6643.8	n = 425
	Within		231.9	4813.8	6199.2	T = 10
Population density	Overall	332.1	444.2	31.5	7300.6	N = 4250
	Between		441.6	35.8	6467.7	n = 42
	Within		52.1	-441.6	1802.1	T = 10
TB prevalence rate	Overall	111.9	132.0	1.0	1820.0	N = 4250
	Between		82.5	1.0	887.6	n = 425
	Within		103.2	-774.6	1106.1	T = 10

*****N = total number of observations, n = the number of kebeles, T = observation per kebele



A) Trend of $\log_{10}$(TB prevalence rate) at district level



B) Trend of $\log_{10}$(TB prevalence rate) at kebele level

Figure 13 Variation of TB prevalence rate at A) districts and B) sample kebeles in Gurage Zone, Southern Ethiopia

5.3.3 Ecological factors affecting the spatial distribution of TB

The fixed-effects spatial autocorrelation model revealed that both the spatial autoregressive coefficient (ρ) in the spatial lag of TB prevalence rate and the spatial autocorrelation coefficient (λ) in the spatial error terms were statistically significant, indicating that the TB prevalence rate observed in a given kebele was affected by both TB prevalence rate and unobserved factors in the neighboring kebeles. By controlling the spatial effects, a 1°C rise in temperature was associated with an increase in the number of TB prevalence rate by 0.72. By controlling the spatial effects, a 1 person per square kilometer increase in population density was related to an increase in the number of TB prevalence rate by 1.19 (**Table 11**).

Table 11 Outcome of fixed-effects spatial autocorrelation model for the spatial association between ecological factors and TB prevalence rate in Gurage Zone, Southern Ethiopia, 2007-2016

Variable	**Coefficient**	**Std. Err.**	**Z**	**P>|Z|**	**[95% Conf. Interval]**	
Rainfall	0.07	0.08	0.88	0.378	-0.080	0.212
Temperature	0.72	0.32	2.28	0.022	0.102	1.345
NDVI	0.23	0.34	0.69	0.491	-0.432	0.901
Population density	1.19	0.17	6.93	<0.001	0.855	1.529
ρ	0.83	0.01	66.52	<0.001	0.800	0.849
λ	-0.70	0.03	-20.62	<0.001	-0.769	-0.636

5.4 Summary of the results by specific objectives

Specific objective 1 (Paper I): The prevalence of TB varied from 70.4 to 155.3 cases per 100,000 population in the Gurage Zone during 2007 to 2016. Eleven purely spatial and three space-time clusters were detected (P-value < 0.001). The clusters were concentrated in border areas of the Gurage Zone.

Specific objective 2 (Paper II): The predictive accuracies of ordinary cokriging models have improved with the true anisotropy, altitude and latitude covariates, the change in detrending pattern from local to global, and the increase in size of spatial dataset. The risk of TB was estimated to be higher at western, northwest, southwest and southeast parts of the study area, and crossed between high and low at west-central parts**.**

Specific objective 3 (Paper III): The TB prevalence rate observed in a given kebele was influenced by both TB prevalence rate and unobserved factors in the neighboring kebeles in the Gurage Zone. By controlling the spatial effects, a 1°C rise in temperature was associated with an increase in the number of TB prevalence rate by 0.72, and a 1 person per square kilometer increase in population density was related to an increase in the number of TB prevalence rate by 1.19.

6. DISCUSSION

6.1 Specific objective 1 (Paper I): Spatial and space-time clustering of TB

This study aims to identify the location, size and risk of purely spatial and space-time clusters for high occurrence of TB in Gurage Zone, Southern Ethiopia during 2007 to 2016. The clusters with high likelihood of TB occurrence were detected in border areas of the zone. The possible explanation for this could be that there were frequent cross-border population movements from the neighboring border areas of Jimma, Yem, Hadiya, Silte, West Shewa and East Shewa zones for economic and social reasons, which could favor the disease transmission in these areas. This is true according to other studies (90). Therefore, future TB prevention and control efforts in these areas should include strengthening health infrastructure, staff capacity building, considering early diagnosis and treatment of symptomatic cases, and increasing community awareness. Furthermore, establishing a neighborhood collaboration network may also help reduce the disease burden in these areas.

Three space-time clusters that were persistent for five years were detected in the zone. This might be due to the uniform implementation of TB prevention and control activities in the zone without targeting high-risk geographical areas (9). On the other hand, the high occurrence of TB could be due to the community-based TB intervention through TB REACH program during 2011 to 2015. Thus, application of GIS and spatial statistical techniques to identify purely spatial and space-time clusters for high occurrence of TB can be recommended for optimal utilization TB resources (49).

6.2 Specific objective 2 (Paper II): Estimating spatial risk of TB distribution

This study has underscored that the geostatistical kriging approach can be applied to estimate the spatial risk of TB distribution in settings where spatially limited data are available. The estimation models indicated that there was spatial heterogeneity in the risk of TB distribution in the Gurage Zone, indicating the disease did not affect all of the communities in the area with the same severity. The risk was higher in northwest, western, southwest and southeast parts of the zone. However, the risk distribution interlocked between high and low at west-central parts. Evidences have revealed that differences in underlying socioeconomic, climatic and geographic conditions, and uneven allocation of public health resources could contribute for the spatial heterogeneity in the risk of TB distribution (11, 18, 91). Moreover, the cross-border population

movements from the neighboring border areas could also facilitate the high transmission of TB, especially at border areas of the zone (90, 92, 93). Therefore, the estimated risk map of TB may help local public health authorities prioritize locations that required immediate interventions.

This study revealed that cokriging with altitude and latitude were the best geostatistical models, which suggested that including these covariables improved the predictive accuracies of the models. This reflects that geographical factors can affect the risk distribution of TB in the Gurage Zone. Previous studies have highlighted that the geographical factors had explicit impacts on the risk distribution of TB (11, 18, 53, 91). Thus, impacts of geographical factors on TB prevention and control should be evaluated, and interventions should be formulated based on geographical features.

This study has also practical implications to TB prevention and control programs in low-income countries, where obtaining spatially complete TB data is difficult. The recent advancements in geostatistical modeling techniques and increasing availability of public health data from the national prevalence surveys, demographic and health system surveys and health facilities will be the good opportunities for epidemiologists working in such settings to predict the spatial risk of TB distribution and associated prediction uncertainty at non-surveyed locations (11, 94, 95). The resulting prediction risk map may allow them measure burden of the disease at all locations, identify high-risk geographical areas for targeted interventions, and evaluate the impacts of intervention programs. This will be useful for optimal utilization of the scarce public health resource.

6.3 Specific objective 3 (Paper III): Ecological factors affecting spatial distribution of TB

This study has applied the spatial panel data modeling techniques to estimate the effects of ecological factors on spatial distribution of TB prevalence rate in Gurage Zone, Southern Ethiopia using panel data from 425 kebeles and the period 2007 to 2016. A spatial panel data model is more appropriate when longitudinal data from multiple spatial units exhibit spatial autocorrelation (96). The model corrects for the deviations that are caused by the spatial interaction between kebeles and unobserved spatial heterogeneity by introducing a spatial weight matrix and a spatial heterogeneity into statistical modeling (89, 97).

This study revealed that the TB prevalence rate observed in a given kebele was affected by both the TB prevalence rate and unobserved factors in the surrounding kebeles, suggesting that there has been sustained transmission of TB within the communities. This finding is similar to previous studies (9, 12, 87, 88). Therefore, it can be recommended that neighboring kebeles should cooperate in the design and implementation of TB prevention and control strategies.

The spatial panel data analysis has revealed that there was a positive association between temperature and TB prevalence rate during the study period. Previous studies support this finding (98). The relationship could be explained by the fact that higher temperatures promote the survival and replication of MTB and improve its activity (53). This implies that the burden of TB will be higher in the coming decades since there is a rapid increment in global temperature. Thus, TB prevention and control strategies should take temperature variations into consideration.

This study has also elucidated that there was a positive association between population density and TB prevalence rate in the study area. The reason for this could be that higher population density increases the risk of personal contact and facilitates the transmission of TB (9). Therefore, policymakers and public health authorities should consider spatial locations with higher population density for TB prevention and control interventions.

6.4 Internal validity and generalizability

6.4.1 Internal validity

Internal validity refers to the ability of the study to measure what it intends to measure correctly. It can be evaluated based on selection bias, information bias, confounding, and chance (99).

Selection bias

Selection bias refers to the distortion that arises from the processes by which individuals are selected into the study population (99). For the three specific objectives (Papers I-III) of this study, TB patients' information was retrieved from all the unit TB registers at each DOTS-providing health facilities in the study area. The TB patients were linked to their home address using geo-codes from CSA. This could help understand the true spatial epidemiology of TB in each kebele. The problem of ecologic bias (i.e., inferring results from aggregated data to individuals) was minimized by carrying out the analysis at the kebele level, the smallest administrative unit in the study area. The scan statistic, which was used in the cluster analyses,

has avoided a pre-selection bias by including all TB cases and areas without pre-determining high and low rates (Paper I).

Information bias

Information bias refers to the distortion that occurs from inaccuracies in the measurement of individual characteristics, and incorrect classification (99). This thesis has used secondary data that were collected from different sources. The TB patients' information was obtained from all the unit TB registers at each DOTS-providing health facilities in the study area, linked to the correct home address, and geo-coded. Information bias depends on the capacity of health care workers to classify TB and complete the information in the register like address and date of diagnosis. The data consistency was checked by using the information from the unit TB registers. The population data were obtained from the CSA. Training was provided for supervisors and data collectors on issues like, the field methods, data extraction and record keeping (Papers I-III). Appropriate statistical methods were employed (Papers I-III). Altitude data were extracted from ASTER Global Digital Elevation Model V2 (Paper II). The NDVI data were derived from the MODIS imagery, and the temperature and rainfall data from the Meteorological Agency of Ethiopia (Paper III).

Confounding

Confounding refers to the distortion in the interpretation of findings due to the association of third variable with the exposure and outcome of interest obscuring whether the exposure is a cause of the outcome (99). The third variable does not lie in a causal pathway between the exposure and outcome. Confounding can underestimate, overestimate or change the direction of association. The ordinary cokriging models and spatial panel data regression were employed to control for the potential cofounders in Papers II and III, respectively. The confounding effects of ecological level factors, like unemployment rate and literacy rate were not controlled in the spatial analyses since it was not possible to access these variables at the kebele level. This could bias the results (Papers I-III).

Chance

Chance refers to the random distortion of the observed association between the exposure and outcome of interest due to sampling variability (99). In Papers I and III, the role of chance was assessed by performing test of statistical significance at a P-value ≤ 0.05 or a 95% confidence interval not crossing the null hypothesis. The cross-validation technique was also used to decide the best set of input parameters for the geostatistical kriging models in Paper II. Moreover, all TB cases diagnosed at each DOTS-providing health facilities were included in the study for each kebele and for each year, which could improve precision of the studies' results (Papers I-III).

6.4.2 Generalizability

Generalizability refers to the extent to which the study findings can be generalized to the larger population or other similar situations (100). The findings of this thesis can be used to understand the spatial epidemiology of TB in geographical locations that have similar socioeconomic, health services and other characteristics to the area of these studies. Moreover, the spatial epidemiological methods used in these studies can also be applied in other settings to discover geographical areas that need more attention, measure the disease burden at locations where data is unavailable, identify neighborhood factors affecting TB distribution, explore the transmission pattern, and evaluate the impacts of interventions.

6.5 Strengths and limitations

6.5.1 Strengths

This study has several strengths. A high resolution spatial data were used to assess the spatial nature of TB distribution at the smallest administrative level, which might reduce the role of ecological bias (Papers I-III). The study covered a wider geographical area containing urban and rural areas (Papers I-III). Errors related to the geo-coding of cases were avoided by linking each case to the correct home address using geo-codes from the CSA (Papers I-III). The spatial data included all forms of TB, so findings did reflect the total burden of TB in the study area (Papers I-III). A longer study period data were used to evaluate the changes in spatial pattern of TB and its ecological level determinants (Papers I and III). Multiple methods were used to detect location, size and severity of purely spatial and space-time clusters for high occurrence of TB in the study area (Paper I). The use of a geostatistical kriging approach to estimate communicable disease's predicted risk surface is a new trend in public health and it is robust. The method was

employed to estimate the spatial risk of TB distribution in the study area using limited spatial datasets (Paper II). The spatial panel data analysis, which could be more appropriate method to analyze space-time data, was used to examine the relationship between ecological factors and TB prevalence rate (Paper III).

6.5.2 Limitations

This study has also limitations. The study did not include TB patients who would remain undiagnosed due to different reasons, and those diagnosed and treated at health facilities outside the study area. These could affect the nature of TB distribution by underestimating the prevalence (Papers I-III). The modifiable area unit problem might arise due to data aggregation at the kebele level. The results of the analysis could vary if the geographic boundaries of kebeles were modified. This might potentially influence the validity of the results (Papers I-III). The denominator population numbers could be affected by uneven population growth across the study area since the numbers were projected from the 2007 census (Papers I-III). The confounding effects of factors, like unemployment rate, literacy rate and age dependency ratio stigma, level of awareness about TB, HIV prevalence, migration, health service coverage and performance of health system were not controlled in the spatial analyses since it was difficult to access these variables at the kebele level (Papers I-III). The Kulldroff's scan statistic used circular spatial scanning windows and space-time cylinders with circular spatial bases which could not detect irregular shaped clusters, and could include a few non-significant locations (Papers I).

7. CONCLUSIONS

The following conclusions were made based on the findings of the study:

- The spatial and space-time clusters for high occurrence of TB were mainly concentrated at border areas of the Gurage Zone.
- The prevalence rate of TB in a given kebele was influenced by both the prevalence rate of TB and unobserved factors in its neighboring kebeles in the Gurage Zone, indicating sustained transmission of the disease within the communities.
- The spatial risk of TB distribution between kebeles in the Gurage Zone was partly explained by spatial variations in temperature, population density, altitude, and latitude.
- The geostatistical kriging approach can be applied to estimate the spatial risk of TB distribution in data limited settings.

8. RECOMMENDATIONS

1. For TB prevention and control programs

- The neighboring kebeles should cooperate in the design and implementation of TB prevention and control strategies to interrupt the chain of disease transmission between the communities.
- TB prevention and control efforts at high-risk geographical areas should include strengthening health infrastructure, staff capacity building, early diagnosis and treatment of symptomatic cases, and increasing community awareness.
- The designing of locally effective TB prevention and control strategies should consider spatial locations with higher temperature and population density.

2. For policy implementation

- Policymakers should consider applying the spatial epidemiological methods to assess spatial inequalities in TB distribution, and identify geographical areas where progress is lagging and greater public health attention is needed.
- Planning TB prevention and control strategies based on data obtained from regional averages may mask local variations within regions. Thus, TB data from health facilities, demographic and health surveillances, national surveys and others should be used to estimate the spatial risk of TB distribution at kebele level, the smallest administrative unit used for healthcare planning.

3. For further research

- Further studies are required to evaluate the effectiveness of geographically targeting TB prevention and control interventions using the inputs from spatial analysis and mapping.
- Further studies should be conducted to estimate the local spatial effects of ecological factors at each kebele using geographically weighted regression.

9. REFERENCES

1. World Health Organization. Global tuberculosis report. Geneva, Switzerland: World Health Organization; 2017.

2. Tadesse S. Stigma against tuberculosis patients in Addis Ababa, Ethiopia. PLoS One 2016; 11(4): e0152900.

3. Kamenju P, Aboud S. Tuberculosis-HIV co-infection among patients admitted at Muhimbili National Hospital in Dares Salaam, Tanzania. Tanzan J Health Res 2011;13.(1):25-31.

4. Baker MG, Venugopal K, Howden CP. Household Crowding and Tuberculosis Copenhagen, Denmark. World Health Organization Regional Office for Europe; 2011.

5. Getahun B, Ameni G, Biadgilign S, Medhin G. Mortality and associated risk factors in a cohort of tuberculosis patients treated under DOTS programme in Addis Ababa, Ethiopia. BMC Infect Dis 2011; 11:127.

6. Federal Ministry of Health. National TB/Leprosy Control Program Report. Addis Ababa, Ethiopia: Ministry of Health; 2010.

7. Deribew A, Abebe G, Apers L, Abdissa A, Deribe F, Woldemichael K, et al. Prevalence of pulmonary TB and spoligotype pattern of Mycobacterium tuberculosis among TB suspects in a rural community in southwest Ethiopia. BMC Infect Dis 2012, 12(1):54.

8. Yassin MADDG, Olivia T, Markos P, Aschalew M, Shargie EB, Mesay RK, et al. Innovative community-based approaches doubled tuberculosis case notification and improved treatment outcome in Southern Ethiopia. PLoS One 2013; 8(5):e63174.

9. Dangisso MH, Datiko DG, Lindtjorn B. Spatio-temporal analysis of smear-positive tuberculosis in the Sidama Zone, Southern Ethiopia. PLoS One 2015; 10(6):e0126369.

10. Tadesse T, Demissie M, Berhane Y, Kebede Y, Abebe M. The clustering of smear-positive tuberculosis in Dabat, Ethiopia: a population based cross sectional study. PloS One 2013; 8(5):e65022.

11. Li XX, Wang LX, Zhang H, Jiang SW, Fang Q, Chen JX, et al. Spatial variations of pulmonary tuberculosis prevalence co-impacted by socio-economic and geographic factors in People's Republic of China, 2010. BMC Public Health 2014; 14 257.

12. Shaweno D, Shaweno T, Trauer JM, Denholm JT, McBryde ES. Heterogeneity of distribution of tuberculosis in Sheka Zone, Ethiopia: drivers and temporal trends. Int J Tuberc Lung Dis 2017; 21(1):79-85.

13. Federal Ministry of Health. Health sector transformation plan-I: Annual performance report. Addis Ababa, Ethiopia: Ministry of Health; 2017.
14. Federal Ministry of Health. First Ethiopian national population-based tuberculosis prevalence survey. Addis Ababa, Ethiopia: Ministry of Health; 2011.
15. Tadesse T, Demissie M, Berhane Y, Kebede Y, Abebe M. Two-thirds of smear-positive tuberculosis cases in the community were undiagnosed in Northwest Ethiopia: Population-based cross-sectional study. PloS One 2011; 6(12):e28258.
16. Goswami ND, Hecker EJ, Vickery C, Ahearn MA, Cox GM, Holland DP, et al. Geographic information system-based screening for TB, HIV, and syphilis (GIS-THIS): a cross-sectional study. PLoS One 2012; 7:e46029.
17. Maciel E, Pan W, Dietze R, Peres R,Vinhas S, Ribeiro F, et al. Spatial patterns of pulmonary tuberculosis incidence and their relationship to socio-economic status in Vitoria, Brazil. Int J Tuberc Lung Dis 2010; 14:1395-1402.
18. Li XX, Wang LX, Zhang J, Liu YX, Zhang H, Jiang SW, et al. Exploration of ecological factors related to the spatial heterogeneity of tuberculosis prevalence in P. R. China. Glob Health Action 2014; 7:23620.
19. Daniel TM. The history of tuberculosis. Respir Med 2006; 100(11):1862-1870.
20. Lönnroth K, Jaramillo E, Williams BG, Dye C, Raviglione M. Drivers of tuberculosis epidemics: The role of risk factors and social determinants. Soc Sci Med 2009; 68(12):2240-2246.
21. Kiers A, Klarenbeek A, Mendelts B, Van Soolingen D, Koeter G. Transmission of Mycobacterium pinnipedii to humans in a zoo with marine mammals. Int J Tuberc Lung Dis 2008; 12:1469-1473.
22. Alexander KA, Laver PN, Michel AL, Williams M, Helden PD, Warren RM, et al. Mycobacterium tuberculosis complex pathogen, M. mungi. Emerg Infect Dis 2010; 16:1296-1299.
23. Lee RB, Chatterlee D, Lee RF. Rapid structural characterization of the arabinogalactan and lipoarabinomannan in live mycobacterial cells using 2D and 3D HR-MAS NMR: structural changes in the arabinan due to ethambutol treatment and gene mutation are observed. Glycobiology 2005; 15(2):139-51.

24. Healthcare.net. How tuberculosis is transmitted? Accessed on: July 18, 2017; Available at: http://respiratory-lung.health-cares.net/tuberculosis-transmission.php.

25. World Health Organization. Global tuberculosis control: Epidemiology, strategy, financing: WHO report 2009 Geneva: Swtzerland: World Health Organization; 2009.

26. Wanyeki I, Olson S, Brassard P, Menzies D, Ross N, Behr M, et al. Dwellings, crowding, and tuberculosis in Montreal. Soc Sci Med 2006; 63:501-511.

27. Ryan KJ, Ray CG. Sherris Medical Microbiology. 4th ed, McGraw Hill: New York; 2004.

28. Govot-Revol V, Innes IA, Hackforth S, Hinks T, Lalvani A. Regulatory T cells are expanded in blood and disease sites in patients with tuberculosis. Am J Resp Crit Care Med 2006; 173: 803-810.

29. Centers for Disease Control and Prevention. Interactive core curriculum on tuberculosis. Centers for Disease Control and Prevention. Accessed on: June 10, 2017; Available at: http://www.cdc.gov/th/webcourses/CoreCurr/TB_Course/Menu/frameset_internet.htm.

30. Jensen PA, Lambert LA, Jademarco MF, Ridzon R. Centers for Disease Control and Prevention: Guidelines for transmission of Mycobacterium tuberculosis in healthcare settings. MMWR Recomm Resp 2005; 54(RR-17): 1-141.

31. American Thotacic Society and Centers for Desease Control and Prevention. Diagnostic standards and classification of tuberculosis in adults and childen. Am J Respir Crit Care Med 2000; 161:1376-1395.

32. Centers for Disease Control and Prevention. Tuberculosis elimination: the difference between latent tuberculosis infection and active tuberculosis desease. Centers for Disease Control and Prevention. Accessed on: May 7, 2017; Available at: http:/cdc.gov/tb/pubs/tbfactsheets/LTBandActiveTB.pdf.

33. Paton NL, Chua YK, Farnest A, Chee CR. Randomized controlled trial of nutritional supplementation in patients with newly diagnosed tuberculosis and wasting. Am J Clin Nutr 2004; 80:450-465.

34. Ddungu H, Johnson II. Smieja M, Manavia-Kizza H. Digital clubbing in tuberculosis-relationship to HIV infection. Extent of disease and hypoalbuminemia. MBC Infet Dis 2006; 6: 45.

35. Centers for Disease Control and Prevention. Surveillance reports' reported tuberculosis in the United States: Centers for Disease Control and Prevention 2005. Accessed on: May 18, 2016; Available at: http:/www.cdc.gov/tb/surv/surv2005/default.htm.

36. Frieden TR, Sterlino TR, Munsiff SS, Watt CJ, Dye C. Tuberculosis. Lancent 2003; 362: 887-899.

37. Wang IY, Hsueh PR, Wang SK, Jan IS, Lee LN, Liaw YS, et al. Disseminated tuberculosis: a 10-year experience in a medical center. Medicine (Baltimore) 2007; 86(1):39-46.

38. Grant A, Gothard P, Thwaites G. Managing drug resistant tuberculosis. British Med J 2008; 28:337.

39. World Health Organization. World Health Organization report on global tuberculosis control. Geneva, Switzerland: World Health Organization; 2013.

40. Elliott P, Wartenberg D. Spatial epidemiology: Current approaches and future challenges Envir Health Perspectives 2004; 112(9):998-1006.

41. Clements ACA, Firth S, Dembelé R, Garba A, Touré S, Sacko M, et at. Use of Bayesian geostatistical prediction to estimate local variations in Schistosoma haematobium infection in western Africa. Bull World Health Organ 2009; 87:921-929.

42. Slater H, Michael E. Mapping, Bayesian Geostatistical Analysis and Spatial Prediction of Lymphatic Filariasis Prevalence in Africa. PLoS One 2013; 8(8):e71574.

43. Pullan RL, Gething PW, Smith JL, Mwandawiro CS, Sturrock HJW, Gitonga CW, et al. Spatial modelling of soil-transmitted helminth infections in Kenya: a disease control planning tool. PLoS Negl Trop Dis 2011; 5(2):e958.

44. Tonnang HEZ, Kangalawe RYM, Yanda PZ. Predicting and mapping malaria under climate change scenarios: the potential redistribution of malaria vectors in Africa. Malar J 2010; 9:111.

45. Noor AM, Clements ACA, Gething PW, Moloney G, Borle M, Shewchuk T, et al. Spatial prediction of Plasmodium falciparum prevalence in Somalia. Malaria J 2008; 7:159.

46. Brownstein JS, Freifeld CC, Chan EH, Keller M, Sonricker AL, Mekaru SR, et al. Information technology and global surveillance of cases of 2009 H1N1 influenza. New England J Med 2010; 552(18):1731-1735.

47. Heimer R, Barbour R, Shaboltas AV, Hoffman IF, Kozlov AP. Spatial distribution of HIV prevalence and incidence among injection drugs users in st petersburg: Implications for HIV transmission. AIDS 2008; 22(1):123-130.

48. Ge E, Zhang X, Wang X, Wei X. Spatial and temporal analysis of tuberculosis in Zhejiang Province, China, 2009-2012. Infec Dis Poverty 2016; 5:11.

49. Tiwari N, Adhikari CM, Tewari A, Kandpal V. Investigation of geo-spatial hotspots for the occurrence of tuberculosis in Almora district, India, using GIS and spatial scan statistic. Int J Health Geogr 2006; 5:33.

50. Teklegiorgis K, Tadesse K, Mirutse G, Terefe W. Level of data quality from health management information systems in a resources limited setting and its associated factors, eastern Ethiopia. South Afr J Info Manag 2016; 17(1):612.

51. Luo W, Taylor MC, Parker SR. A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. Int J Climatol 2008; 28:947-959.

52. Osei FB. Current Statistical Methods for Spatial Epidemiology: A Review. Austin Biom and Biostat 2014; 1(2):7.

53. Sun W, Gong J, Zhou J, Zhao Y, Tan J, Ibrahim AN, et al. A spatial, social and environmental study of tuberculosis in China using statistical and GIS technology. Int J Environ Res Public Health 2015; 12:1425-1448.

54. Barter DM, Agboola SO, Murray MB, Barnighausen T, et al. Tuberculosis and poverty: the contribution of patient costs in Subsaharan Africa – A systematic review. BMC Public Health 2012; 12:980.

55. World Health Organization. Global tuberculosis report. Geneva, Switzerland: World Health Organization; 2015.

56. Kwan CK, Ernst JD. HIV and tuberculosis: A deadly human syndemic. Clin Microbiol Rev 2011; 24(2):351-376.

57. Maher D, Harries A, Getahun H. Tuberculosis and HIV interac-tion in Subsaharan Africa: Impact on patients and programmes; implications for policies. Trop Med Int Health 2005; 10: 734-742.

58. Harries A, Dye C. Tuberculosis. Annals Trop Med Parasitol 2006; 100(5):415-431.

59. Sengupta S, Pungrassami P, Balthip Q. Social impact of tuberculosis in Southern Thailand: Views from patients, care providers and the community. Int J Tuberc Lung Dis 2006; 10:1008-1012.

60. Tanimura T, Jaramillo E, Weil D, Raviglione M, Lonnroth K. Financial burden for tuberculosis patients in low- and middle-income countries: A systematic review. Eur Respir J 2014; 43:1763-1775.

61. Tadesse T, Demissie M, Berhane Y, Kebede Y, Abebe M. Long distance travelling and financial burdens discourage tuberculosis DOTs treatment initiation and compliance in Ethiopia: a qualitative study. BMC Public Health 2013; 13:424.

62. Kemp JR, Mann G, Nhlema SB, Salaniponi FML, Squire SB. Can Malawi's poor afford tuberculosis services? Patient and household costs associated with a tuberculosis diagnosis in Lilongwe. Bull World Health Organ 2007; 85:580-585.

63. Mesfin MM, Newell JN, Madeley JR. Cost implications of delays to tuberculosis diagnosis among pulmonary tuberculosis patients in Ethiopia. BMC Public Health 2010; 10:173.

64. World Health Organization. Eliminating the financial hardship of TB via universal health coverage and other social protection measures :World Health Organization; 2013. Accessed on: May 13, 2016; Available at: http://www.who.int/tb/uhc_socialprotection/en/.

65. Federal Ministry of Health. Implementation guideline for GeneXpert MTB/RIF Assay in Ethiopia. Addis Ababa, Ethiopia: Ministry of Health; 2014.

66. Federal Ministry of Health. Guideline for clinical and programmatic management of TB, Leprosy and TB/HIV in Ethiopia. Addis Ababa, Ethiopia: Ministry of Health; 2012.

67. Banteyerga H. Ethiopia's health extension program: improving health through community involvement. MEDICC Rev 2011; 13(3):46-49.

68. Dagnaw AM, Tiruneh BZ. Laboratory diagnostic systems used in the diagnosis of tuberculosis in Ethiopia: A systematic review. J Medical Laboratory and Diagnosis 2014; 5(2):14-21.

69. Federal Ministry of Health. Tuberculosis, leprosy and TB/HIV prevention and control program manual. Addis Ababa, Ethiopia: Ministry of Health; 2012.

70. Central Statistical Agency. The 2007 population and hhousing census of Ethiopia: Statistical report for Southern Nations, Nationalities and Peoples' Region. Addis Ababa, Ethiopia: Central Statistical Agency; 2007. Accessed on: January 21, 2016; Available at: http://www.csa.gov.et/census-report/complete-report/census-2007.html?start=10.

71. Balabanova D, McKee M, Mills A. Good health at low cost' 25 years on. What makes a successful health system? London: LSHTM; 2011.

72. Shargie E, Lindtjørn B. Determinants of treatment adherence among smear-positive pulmonary tuberculosis patients in Southern Ethiopia. PLoS Med 2007; 4:0280-0287.
73. Southern Nations Nationalities and Peoples' Regional State Health Department. The 2016 health report. Awassa, Ethiopia: Regional Health Department; 2017.
74. Berhane Y, Wall S, Kebede D, Emmelin A, Enquselassie F, Byass P, et al. Establishing an epidemiological field laboratory in rural areas-potential for public health research and intervention: the Butajira rural health program 1987-99. Ethiop J Health Dev 1999; 13:1-47.
75. Federal Ministry of Health. Revised national strategic plan: Tuberculosis, TB/HIV, DR-TB, and leprosy prevention and control 2013/14-2020. Addis Ababa, Ethiopia: Ministry of Health; 2017.
76. Levin KA. Study design VI- Ecological studies. Evidence-Based Dentistry 2003; 7:60-61.
77. Anselin L, Syabri I, Kho Y. GeoDa: An Introduction to spatial data analysis. Geogr Anal 2006; 38(1):5-22.
78. Liu Y, Li X, Wang W, Li Z, Hou M, He Y, et al. Investigation of space-time clusters and geospatial hot spots for the occurrence of tuberculosis in Beijing. The Int J Tuberc lung dis: the official J Int Union against Tuberc Lung Dis 2012; 16(4):486-491.
79. Moran PAP. Notes on continuous stochastic phenomena. Biometrika 1950; 37:17-23.
80. Kulldorff M. SaTScanTM User Guide for version 9.2; 2010.
81. Getis A, Ord JK. The analysis of spatial association by use of distance statistics. Geogr Anal 1992; 24:189-206.
82. Kim SY, Yi SJ, Eum YS, Choi HJ, Shin H, Ryou HG, et al. Ordinary kriging approach to predicting long-term particulate matter concentrations in seven major Korean cities. Environ Health Toxicol 2014; 29:e2014012.
83. Luo W, Taylor MC, Parker SR. A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. Int J Climatol 2008; 28:947-959.
84. Konak A. A kriging approach to predicting coverage in wireless networks. Int J Mobile Network Design and Innovation 2009; 3:2.
85. Yalcin E. Cokriging and its effect on the estimation precision. J South African Inst Mining and Metallurgy 2005; 105:223-228.

86. Aretouyap Z, Nouck PN, Nouayou R, Kemgang FEG, Toko ADP, Asfahani J. Lessening the adverse effect of the semivariogram model selection on an interpolative survey using kriging technique. SpringerPlus 2016; 5:549.

87. Tadesse S, Enqueselassie F, Hagos S. Spatial and space-time clustering of tuberculosis in Gurage zone, southern Ethiopia. PLoS One 2018; 13(6):e0198353.

88. Tadesse S, Enqueselassie F, Gebreyesus SH. Estimating the spatial risk of tuberculosis distribution in Gurage zone, southern Ethiopia: a geostatistical kriging approach. BMC Public Health 2018; 18:783.

89. Belotti F, Hughes G, Mortari AP. Spatial panel data models using Stata. The Stata J 2017; 17(1):139-180.

90. Kozińska M, Zientek J, Augustynowicz-Kopeć E, Zwolska Z, Kozielski J. Transmission of tuberculosis among people living in the border areas of Poland, the Czech Republic, and Slovakia. Pol Arch Med Wewn 2016; 126 (1-2):32-40.

91. Dangisso MH, Datiko DG, Lindtjørn B. Accessibility to tuberculosis control services and tuberculosis programme performance in southern Ethiopia. Glob Health Action 2015; 8:29443.

92. Regassa N, Yusufe A. Gender Differentials in migration impacts in Southern Ethiopia. Anthropologist 2009; 11(2):129-137.

93. Boru CG, Shimels T, Bilal AI. Factors contributing to non-adherence with treatment among TB patients in Sodo Woreda, Gurage Zone, Southern Ethiopia: A qualitative study. J Infect Public Health 2017; 10:527-533.

94. Ibrahim S, Hamisu I, Lawal U. Spatial pattern of tuberculosis prevalence in Nigeria: A comparative analysis of spatial autocorrelation indices. Am J Geogr Infor System 2015; 4(3):87-94.

95. Gething P, Atkinson P, Noor A, Gikandi P, Hay S, Nixon M. A local spacetime kriging approach applied to a national outpatient malaria dataset. Comput Geosci 2007; 33(10):1337-1350.

96. Wang H, Du Z, Wang X, Liu Y, Yuan Z, Liu Y, et al. Detecting the association between meteorological factors and hand, foot, and mouth disease using spatial panel data models. Int J Infect Dis 2015; 34:66-70.

97. Rao HX, Zhang X, Zhao L, Yu J, Ren W, Zhang XL, et al. Spatial transmission and meteorological determinants of tuberculosis incidence in Qinghai Province, China: A spatial clustering panel analysis. Infect Dis Poverty 2016; 5:45.
98. Khaliq A, Batool SA, Chaudhry MN. Seasonality and trend analysis of tuberculosis in Lahore, Pakistan from 2006 to 2013. J Epidemiol Glob Health 2015; 5:397-403.
99. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia, PA: Lippincott Williams and Wilkins; 2008.
100. Polit DF, Beck CT. Generalization in quantitative and qualitative research: Myths and strategies. Int J Nursing Studies 2010; 47:1451-1458.

10. ANNEXES

Annex I: Original Papers I-III

Paper I

PLOS ONE

RESEARCH ARTICLE

Spatial and space-time clustering of tuberculosis in Gurage Zone, Southern Ethiopia

Sebsibe Tadesse[1]*, Fikre Enqueselassie[2], Seifu Hagos[2]

1 Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia, **2** School of Public Health, College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia

* sbsbtadesse90@gmail.com

OPEN ACCESS

Citation: Tadesse S, Enqueselassie F, Hagos S (2018) Spatial and space-time clustering of tuberculosis in Gurage Zone, Southern Ethiopia. PLoS ONE 13(6): e0198353. https://doi.org/10.1371/journal.pone.0198353

Editor: Mohammad Ali, Johns Hopkins Bloomberg School of Public Health, UNITED STATES

Received: October 30, 2017

Accepted: May 17, 2018

Published: June 5, 2018

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Introduction

Spatial targeting is advocated as an effective method that contributes for achieving tuberculosis control in high-burden countries. However, there is a paucity of studies clarifying the spatial nature of the disease in these countries. This study aims to identify the location, size and risk of purely spatial and space-time clusters for high occurrence of tuberculosis in Gurage Zone, Southern Ethiopia during 2007 to 2016.

Materials and methods

A total of 15,805 patient data that were retrieved from unit TB registers were included in the final analyses. The spatial and space-time cluster analyses were performed using the global Moran's *I*, Getis-Ord G_i^* and Kulldorff's scan statistics.

Results

Eleven purely spatial and three space-time clusters were detected (P <0.001).The clusters were concentrated in border areas of the Gurage Zone. There were considerable spatial variations in the risk of tuberculosis by year during the study period.

Conclusions

This study showed that tuberculosis clusters were mainly concentrated at border areas of the Gurage Zone during the study period, suggesting that there has been sustained transmission of the disease within these locations. The findings may help intensify the implementation of tuberculosis control activities in these locations. Further study is warranted to explore the roles of various ecological factors on the observed spatial distribution of tuberculosis.

Introduction

Ethiopia remains among high Tuberculosis (TB) endemic countries in the world, with an estimated annual incidence of 177 per 100,000 population [1]. TB places an extraordinary public health, financial and social burden in the country [2, 3]. It is one of the most important infectious diseases responsible as a leading cause of death and second cause of hospital admission [4, 5]. The patients face various levels of isolation and rejection, including loss of employment, reduced education opportunities, vulnerability to disability and divorce or spoiled marriage prospects [6]. Moreover, co-infection with Human Immunodeficiency Virus and the emergence of resistance to numerous anti-TB agents are recognized as increasing problems in the country [7, 1].

During the past two decades, Geographic Information Systems (GIS) and spatial statistics were used to detect spatial and space-time clustering of TB in many developing countries [8–12]. Several studies have been conducted at the national, regional and district levels to investigate the epidemiology of TB in Ethiopia [13–16]. Although these studies reveal helpful information for TB control programs, spatial context such as clustering patterns of the disease have rarely been taken into account. Spatial and space-time cluster analyses of TB may help public health officials discover the high-risk geographical areas and population groups that require targeted interventions [8–12]. Lack of such information may contribute for the partial effectiveness of TB control programs in Ethiopia, where the national TB control program implements a uniform approach to allocate resource across the regions.

This study aims to identify the location, size and risk of purely spatial and space-time clusters for high occurrence of TB in Gurage Zone, Southern Ethiopia during 2007 to 2016. The information may contribute to more effective budget allocation, active search for symptomatic patients, drug distribution, recruitment of skilled human resources, guiding the design of vaccination programs, community awareness creation through public health advocacy, and identifying factors behind the spread of the disease in high-risk areas.

Methods and materials

Study area

The study was conducted in the Gurage Zone, Southern Ethiopia, which is located between 7˚ 76' and 8˚45' N latitude and 37˚46' and 38˚71' E longitude (**Fig 1**).The zone is divided into 13 districts and two town administrations (at Butajira and Wolkite).There are 403 rural and 20 urban kebeles (the smallest administrative units with a population of 5,000 on average) in the zone. About 84% of the populations live in the rural areas [17]. The zone has an average population of about 1,453,531 during 2007 to 2016 [18].

The zone has a total of 6 hospitals, 70 health centers, 414 health posts and 92 clinics that are involved in the prevention and control of TB. The health posts and clinics provide health education, identify suspects, refer patients to health centers and support treatment through trained health extension workers from the community. Health centers perform sputum microscopy, treatment and the referral of smear-negative and extra-pulmonary cases to hospitals for further management, while hospitals render diagnosis, treatment and inpatient care services [19].

The zone is selected for conducting this study mainly for two reasons. First, the zone has a geographic diversity characterized by mountain ranges and valleys. This may impact a geographic access to TB diagnosis and treatment centers [20–21]. Hence, it is assumed that TB cases are clustered in specific-geographic locations, and implementing a uniform TB control measures may not be effective in mitigating the problem. Second, according to the Southern

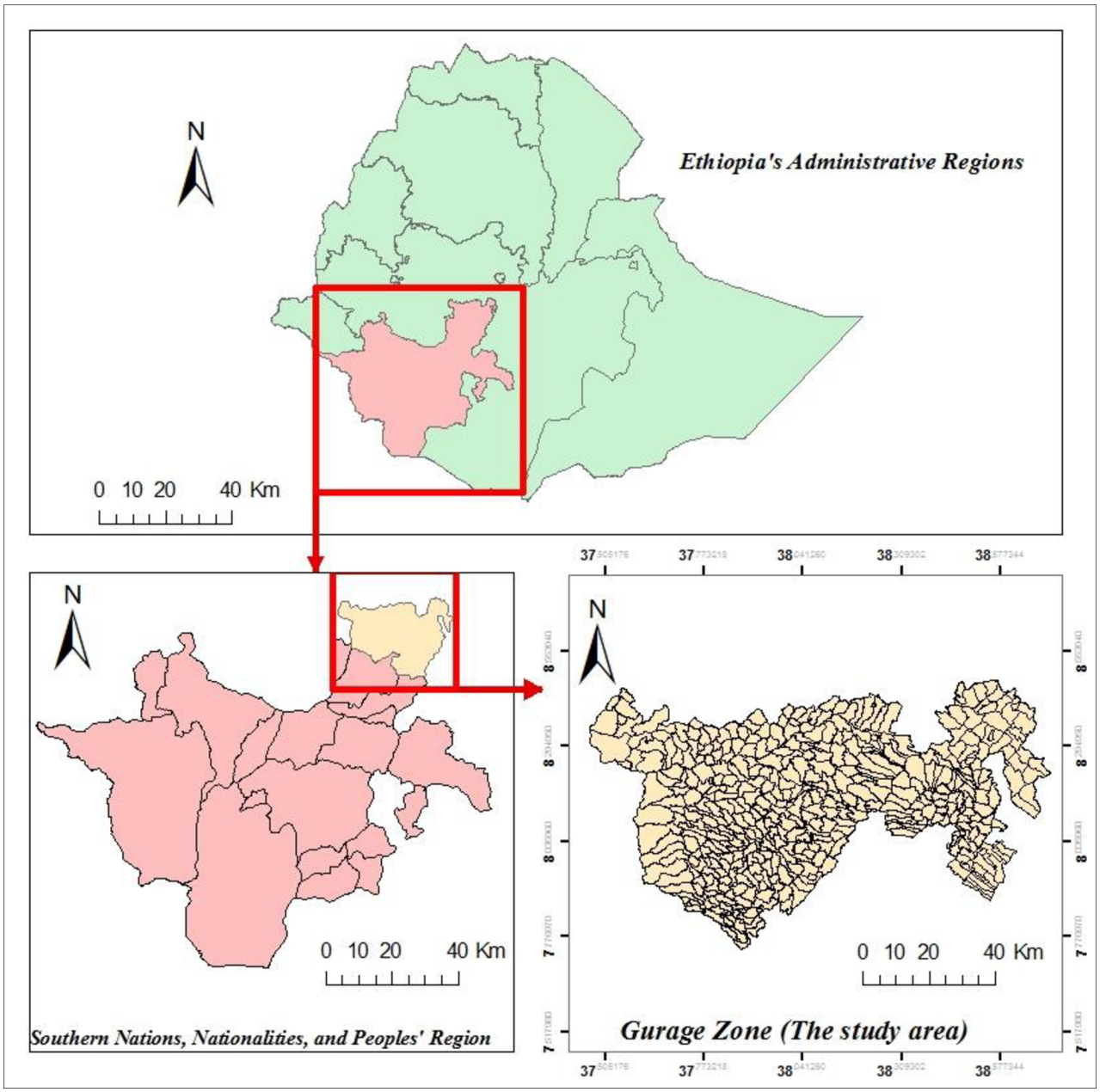


Fig 1. Map of the study area (Gurage Zone).

https://doi.org/10.1371/journal.pone.0198353.g001

Nations, Nationalities, and Peoples' Regional State Health Department Report of 2016 the zone is among the highest hit areas by TB in the region [19].

Data sources

The study was conducted from June to September, 2017. The list of DOTS-providing health facilities were obtained from the Gurage Zone Health Department database. Trained data collectors retrieved the patient information on sex, age, address, TB type, patient category and date of treatment started from unit TB registers at DOTS-providing health facilities in the Gurage Zone during 2007 to 2016. The patients' addresses with similar names but from

different geo-locations were linked to their correct geo-locations to prevent duplication. TB cases were diagnosed using pathogen detection, X-ray, and pathologic diagnosis according to the diagnosis criteria recommended by the national TB diagnosis and treatment guideline of Ethiopia [22]. The population and geo-location data for each kebele in the Zone were obtained from the Central Statistical Agency of Ethiopia (CSA).

Data quality control. The training of data collectors and supervisors emphasized issues such as data extraction format, field methods, and record keeping. The data were double entered and checked page-by-page by year, district, kebele and health facilities against unit TB registers for consistency and completeness throughout the entire data collection.

Data management and processing. Data entering, validating, cleaning and coding were employed using MS Excel (MicroSoft, Redmond, WA, USA). In the study area TB data were reported by the Basic Management Units. The reports might include cases outside of the administrative catchment or miss cases from their catchment enrolled in neighboring health facilities. The reason for this could be that the patients could cross the administrative boundaries for seeking health services by reason of access, preference and quality of care. Data aggregation based on the correct address of patients could help understand the true spatial nature of TB burden in the study area. Therefore, the patient data were aggregated at kebele level for spatial analyses in this study. Kebele centroids were used to represent a geographically weighted central location as coordinates.

Spatial autocorrelation analyses

The global Moran's *I* statistic was run in ArcGIS 10.2to examine the presence of spatial clustering of TB in the whole study area. The value of Moran's *I* is calculated based on the deviation from the mean of two neighboring values. The following equation is used to calculate the Moran's *I* statistic [8]:

$$I = \frac{n}{S_0}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{i,j}z_i z_j}{\sum_{i=1}^{n}z_i^2}$$

where z_i is the deviation of a prevalence of TB for kebele *i* from its mean $(X_i - \bar{X})$, z_j is the deviation of a prevalence of TB for kebele *j* from its mean $(X_j - \bar{X})$, $\omega_{i,j}$ is the spatial weight between kebele *i* and *j*, *n* is the total number of kebeles, and S_0 is the aggregate of all the spatial weights: $S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{i,j}$. The z_I-score for the statistic is computed as: $z_I = \frac{I-E[I]}{\sqrt{V[I]}}$, where $E[I] = -1/(n-1)$ and $V[I] = E[I^2]-E[I]^2$. The spatial relationships among kebeles were conceptualized by calculating the spatial weights from the input file containing the prevalence rate of TB for each kebele (the number of TB cases divided by the population of a given year and multiplied by 100,000) and the geo-coordinates data. A first order queen polygon continuity weights matrix, which defines the neighbors as those with either a shared border or vertex, was used for spatial weights [10]. The spatial weighing matrix was constructed in Geographical data analysis tool (GeoDa) by using the kebele-level polygon shape file Statistical significance for high occurrence of TB was decided when the Z-score $\geq$1.96 and a P-value $\leq$ 0.05.

Purely spatial and space-time cluster analyses

The Kulldroff's scan statistic was performed in SaTScan 9.2 to identify the location, size and severity of purely spatial and space-time clusters using the number of TB cases, the population for each kebele, the year of TB diagnosis and the geo-coordinates data as input files (**S1** and **S2 Tables)** [23]. The discrete Poisson model was used with the assumption that the number of TB cases at each location was Poisson distributed with a known population at risk. Scan circles of

various sizes were used to identify the purely spatial clusters for high occurrence of TB. The upper limit for the maximum cluster size was set to 50% of the population at risk, which allowed small and large clusters to be detected. To identify space-time clusters for high occurrence of TB, a cylindrical window with the circular geographic base representing to the space and height to time was used. The size of the window was limited to 50% of the expected number of TB cases, and the time was set to the time period from 2007 to 2016. In both cases, the likelihood ratio was computed to measure a Relative Risk (RR) of TB occurrence within the cluster when compared to the risk outside using Monte Carlo simulations. The maximum number of replications for Monte Carlo simulation was set to 99,999. The cluster with the maximum Log Likelihood Ratio (LLR) was defined as the most likely cluster. The P-value was created using the combination of approximation [23]. A standard of 'no geographical overlap' was selected to report secondary clusters. Statistical significance was reported when a P-value was ≤ 0.05.

The Getis-Ord G_i^* statistic was also implemented in ArcGIS 10.2 to identify the locations of clusters for high occurrence of TB. The G_i^* statistic performs the spatial analysis by looking at each kebele within the context of a neighboring kebele. The local sum for a kebele and its neighbors is proportionally compared to the sum of all kebeles. When the local sum is much different than the expected local sum and that difference is too large to be the result of random chance, a statistically significant Z-score result. The following equation is used to compute the G_i^* statistic [8]:

$$G_i^* = \frac{\sum_{j=1}^{n} \omega_{i,j} x_j - \bar{X} \sum_{j=1}^{n} \omega_{i,j}}{S \sqrt{\frac{\left[n \sum_{j=1}^{n} \omega_{i,j}^2 - \left(\sum_{j=1}^{n} \omega_{i,j} \right)^2 \right]}{n-1}}}$$

where x_j is the prevalence of TB for kebele j, $\omega_{i,j}$ is the spatial weight between kebeles i and j, n is the total number of kebeles, $\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n}$, and $S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2}$. Therefore, the G_i^* statistic is a Z-score. The spatial relationships among kebeles were conceptualized by calculating the spatial weights from the input file containing the prevalence rate of TB for each kebele and the geo-coordinates data. A first order queen polygon continuity weights matrix, which defines the neighbors as those with either a shared border or vertex, was used for spatial weights [10]. Statistical significance for high occurrence of TB was decided when the $G_i^* \geq 1.96$ and a P-value ≤ 0.05.

Ethical considerations

The study protocol was reviewed and approved by the Research and Ethical Committee (REC) of the School of Public Health, and the Institutional Review Board (IRB) of the College of Health Sciences, Addis Ababa University. Since the study used data from the retrospective review of unit TB registers during 2007 to 2016, both the REC and IRB waived the requirement for informed consent from the patients. For this reason, informed consent was not obtained from the patients. A letter of support was obtained from the Gurage Zone Department of Health to obtain information from all districts and health facilities. The anonymity of cases was kept by using pseudo identification. Medical records were stored in a secure place to help maintain the confidentiality of the clinical information of cases.

Results

Demographic characteristics of TB cases

A total of 16,618 TB cases were diagnosed between 2007 and 2016. Of these, 4.9% were excluded from the final analyses due to incomplete addresses or being outside of the study

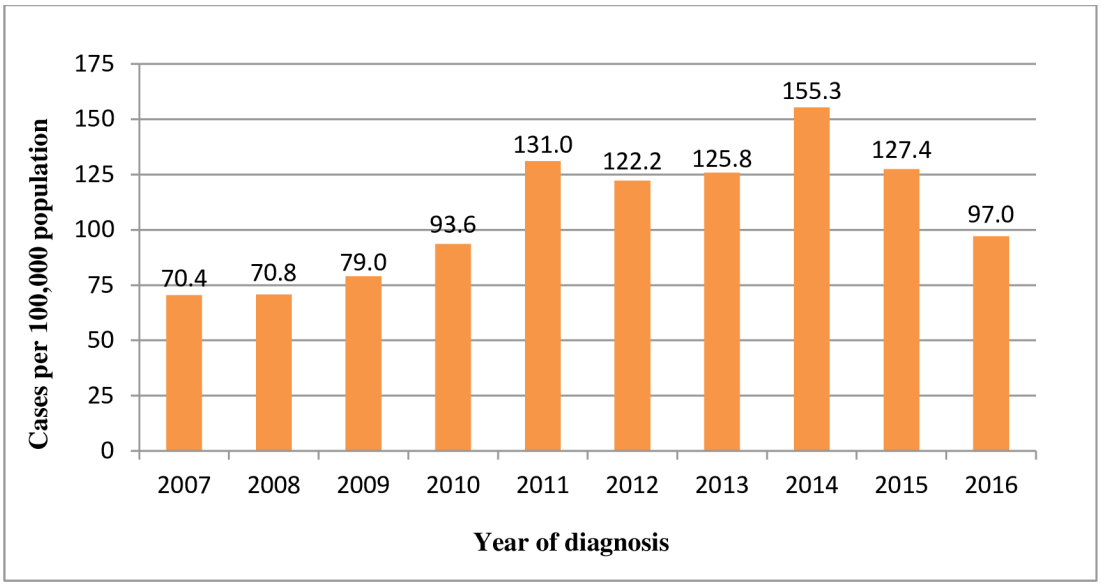


Fig 2. Trend of TB prevalence in Gurage Zone, Sothern Ethiopia, 2007–2016.

https://doi.org/10.1371/journal.pone.0198353.g002

area. Out of the 15,805 cases included in this study55.3% were males. The mean age with a standard deviation of the cases was 34.0±16.4. About 93.2% of the cases were newly diagnosed, while 6.8% were retreatment cases. Fifteen percent of the cases were from urban areas. The prevalence of TB varied from 70.4 to 155.3 cases per 100,000 population during 2007 to 2016 (Fig 2).

Spatial autocorrelation

The global Moran's *I* statistic was significant for each year, implying that there were clusters in the distribution of TB in Gurage Zone, Southern Ethiopia (Table 1).

Purely spatial clusters

The purely spatial cluster analyses identified the clusters for high occurrence of TB at peripheral areas of the geographic zone. The most likely cluster with 291 TB cases (71.19 expected cases) was detected at southwest of Abeshege district. The size of the cluster was within a

Table 1. Global spatial autocorrelation of TB distribution in Gurage Zone, Southern Ethiopia, 2007–2016.

Year	Moran's I	Z-score	P-value	Pattern
2007	0.266568	12.3	<0.001	Clustered
2008	0.143781	6.8	<0.001	Clustered
2009	0.194124	9.6	<0.001	Clustered
2010	0.288589	13.4	<0.001	Clustered
2011	0.191062	9.1	<0.001	Clustered
2012	0.161519	7.7	<0.001	Clustered
2013	0.116817	5.4	<0.001	Clustered
2014	0.130803	6.1	<0.001	Clustered
2015	0.174448	8.2	<0.001	Clustered
2016	0.098785	4.6	<0.001	Clustered

https://doi.org/10.1371/journal.pone.0198353.t001

Table 2. Purely spatial clusters for high occurrence of TB in Gurage Zone, Southern Ethiopia, 2007–2016.

Cluster type	Cluster Year	Cluster center/radius	Observed cases	Expected cases	LLR	RR	P-value
Most likely cluster	2007–2016	(8.27 N, 37.58 E) / 4.45 km	291	71.19	192.87	4.16	<0.001
Secondary cluster	2007–2016	(8.12 N, 38.53 E) / 16.50 km	3325	535.65	135.75	1.39	<0.001
2nd secondary	2007–2016	(7.85 N, 37.82 E) / 9.97 km	1365	998.14	64.99	1.40	<0.001
3rd secondary	2007–2016	(8.33 N, 37.88 E) / 12.32 km	1253	942.62	49.53	1.36	<0.001
4th secondary	2007–2016	(8.05 N, 37.64 E) / 14.69 km	685	460.81	49.21	1.51	<0.001
5th secondary	2007–2016	(7.97 N, 38.06 E) / 3.98 km	340	213.02	32.51	1.61	<0.001
6th secondary	2007–2016	(8.24 N, 38.06 E) / 0 km	36	8.93	23.14	4.04	<0.001
7th secondary	2007–2016	(8.14 N, 38.31 E) / 1.56 km	218	136.27	20.91	1.61	<0.001
8th secondary	2007–2016	(8.18 N, 37.85 E) / 0 km	58	23.55	17.86	2.47	<0.001
9th secondary	2007–2016	(8.41N, 38.24 E) / 0 km	44	17.22	2.56	14.52	<0.001
10th secondary	2007–2016	(8.32 N, 38.55 E) / 0 km	140	86.04	14.29	1.63	<0.001

https://doi.org/10.1371/journal.pone.0198353.t002

radius of 4.45 km. People within this cluster had 4.16 times higher risk of TB infection than those outside the cluster. Eleven significant secondary clusters for high occurrence of TB were also detected. Most of these locations belong to the districts of Gumer, Endegagn, Enemor Ener, eastern Abeshege, Kebena, Welkite Town, northern Ezha, northwestern and eastern Cheha, Meskan, Butajira Town, northern Merako, southwest Sodo (Table 2, Fig 3).

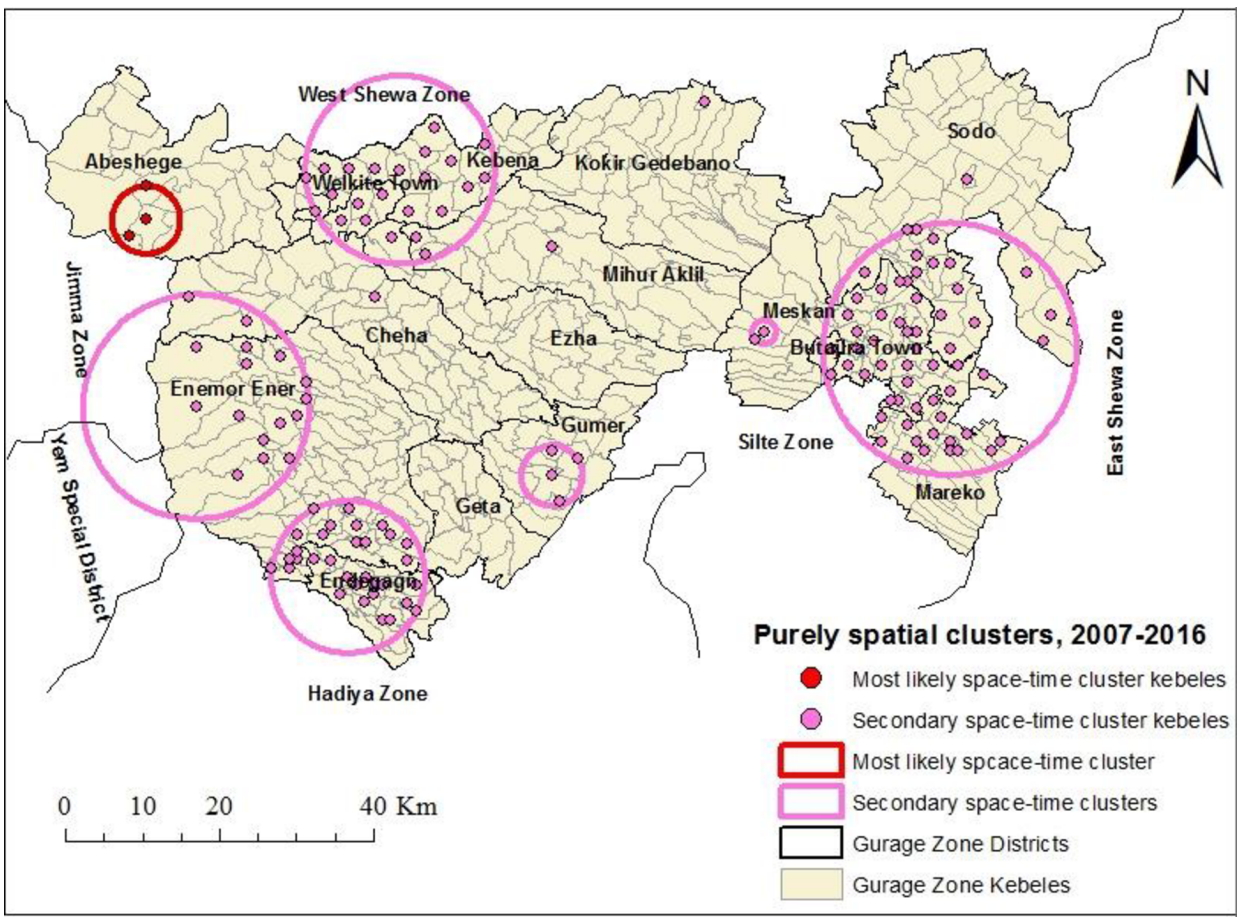


Fig 3. Significant purely spatial clusters for high occurrence of TB in Gurage Zone, Southern Ethiopia, 2007–2016.

https://doi.org/10.1371/journal.pone.0198353.g003

The nature of the clusters for high occurrence of TB in the study area was evaluated for each year from 2007 to 2016. There were considerable spatial variations in the risk of TB by year (Table 3, Fig 4). Nearly consistent results were obtained by the G_i^* statistic (Fig 5).

Table 3. Annual purely spatial clusters for high occurrence of TB in Gurage Zone, Southern Ethiopia, 2007–2016.

Cluster type	Cluster Year	Cluster center/radius	Observed cases	Expected cases	LLR	RR	P-value
Most likely cluster	2007	(8.15 N, 38.17 E) / 32.07 km	635	406.44	118.53	2.91	<0.001
Secondary cluster	2007	(8.29 N, 37.51 E) / 14.61 km	86	24.92	47.64	3.71	<0.001
Most likely cluster	2008	(8.36 N, 37.49 E) / 14.45 km	101	22.78	75.66	4.85	<0.001
Secondary cluster	2008	(8.28 N, 37.99 E) / 23.45 km	325	215.73	32.70	1.78	<0.001
2nd secondary	2008	(8.23 N, 38.49 E) / 4.54 km	44	12.91	23.40	3.53	<0.001
3rd secondary	2008	(7.94 N, 38.07 E) / 0 km	14	2.13	14.56	6.65	<0.001
4th secondary	2008	(7.84 N, 37.83 E) / 7.78 km	68	34.01	13.78	2.08	<0.001
5th secondary	2008	(8.14 N, 38.31 E) / 1.56 km	25	8.04	11.55	3.17	<0.004
Most likely cluster	2009	(8.43 N, 37.56 E) / 47.88 km	382	194.53	92.08	2.50	<0.001
Secondary cluster	2009	(7.89 N, 37.84 E) / 6.97 km	85	39.00	21.26	2.28	<0.001
2nd secondary	2009	(8.22 N, 38.51 E) / 7.02 km	61	26.13	17.44	2.42	<0.001
3rd secondary	2009	(8.14 N, 38.31 E) / 1.56 km	28	9.20	12.53	3.10	<0.001
4th secondary	2009	(8.33 N, 38.20 E) / 11.49 km	100	60.88	11.28	1.71	<0.005
5th secondary	2009	(7.97 N, 38.11 E) / 9.16 km	71	39.17	10.90	1.87	<0.006
Most likely cluster	2010	(7.86 N, 37.73 E) / 12.99 km	137	54.72	46.24	2.68	<0.001
Secondary cluster	2010	(7.97 N, 38.11 E) / 5.98 km	89	31.16	36.90	2.99	<0.001
2nd secondary	2010	(8.25 N, 38.51 E) / 10.06 km	111	52.84	25.61	2.20	<0.001
3rd secondary	2010	(8.06 N, 38.46 E) / 3.13 km	60	21.73	23.24	2.85	<0.001
4th secondary	2010	(8.13 N, 38.30 E) / 0 km	23	6.54	12.56	3.56	<0.001
5th secondary	2010	(8.31 N, 37.96 E) / 4.93 km	35	14.97	9.85	2.38	<0.017
Most likely cluster	2011	(8.13 N, 38.64 E) / 23.49 km	507	267.67	103.35	2.23	<0.001
Secondary cluster	2011	(8.36 N, 37.49 E) / 14.45 km	109	45.68	32.59	2.47	<0.001
2nd secondary	2011	(8.03 N, 37.83 E) / 0 km	26	7.19	14.71	3.65	<0.001
3rd secondary	2011	(7.84 N, 37.83 E) / 1.56 km	27	7.84	14.32	3.48	<0.001
4th secondary	2011	(8.17 N, 38.25 E) / 7.07 km	72	39.60	10.94	1.85	<0.006
Most likely cluster	2012	(8.00 N, 38.58 E) / 27.86 km	525	319.81	69.98	1.91	<0.001
Secondary cluster	2012	(8.12 N, 37.76 E) / 4.58 km	60	21.56	23.39	2.84	<0.001
2nd secondary	2012	(8.26 N, 37.73 E) / 5.93 km	93	57.16	9.80	1.66	<0.018
Most likely cluster	2013	(8.06 N, 38.47 E) / 11.50 km	380	231.48	46.66	1.80	<0.001
Secondary cluster	2013	(7.86 N, 37.75 E) / 2.20 km	42	12.21	22.33	3.49	<0.001
2nd secondary	2013	(8.28 N, 38.53 E) / 4.96 km	49	21.85	12.62	2.28	<0.001
3rd secondary	2013	(8.27 N, 37.58 E) / 4.45 km	25	8.55	10.44	2.95	<0.001
Most likely cluster	2014	(8.00 N, 38.50 E) / 4.96 km	151	52.54	63.04	3.00	<0.001
Secondary cluster	2014	8.33 N, 37.85 E) / 3.30 km	48	11.27	33.11	4.33	<0.001
2nd secondary	2014	(8.10 N, 37.70 E) / 11.50 km	144	87.60	15.87	1.68	<0.001
3rd secondary	2014	(8.19 N, 37.94 E) / 3.98 km	46	22.48	9.53	2.07	<0.024
Most likely cluster	2015	(8.06 N, 38.51 E) / 17.92 km	481	334.77	34.58	1.57	<0.001
Secondary cluster	2015	(8.18 N, 37.63 E) / 40.94 km	843	688.91	25.27	1.38	<0.001
2nd secondary	2015	(7.82 N, 37.89 E) / 3.13 km	43	18.53	11.88	2.35	<0.003
Most likely cluster	2016	(8.29 N, 37.51 E) / 9.92 km	66	23.11	26.96	2.94	<0.001
Secondary cluster	2016	(7.91 N, 37.86 E) / 3.98 km	57	21.01	21.32	2.78	<0.001
2nd secondary	2016	(8.00 N, 38.58 E) / 25.06 km	343	251.21	18.24	1.47	<0.001
3rd secondary	2016	(8.33 N, 37.88 E) / 12.32 km	148	95.13	13.49	1.61	<0.001
4th secondary	2016	(8.05 N, 37.64 E) / 7.78 km	26	9.82	9.22	2.68	<0.030

https://doi.org/10.1371/journal.pone.0198353.t003

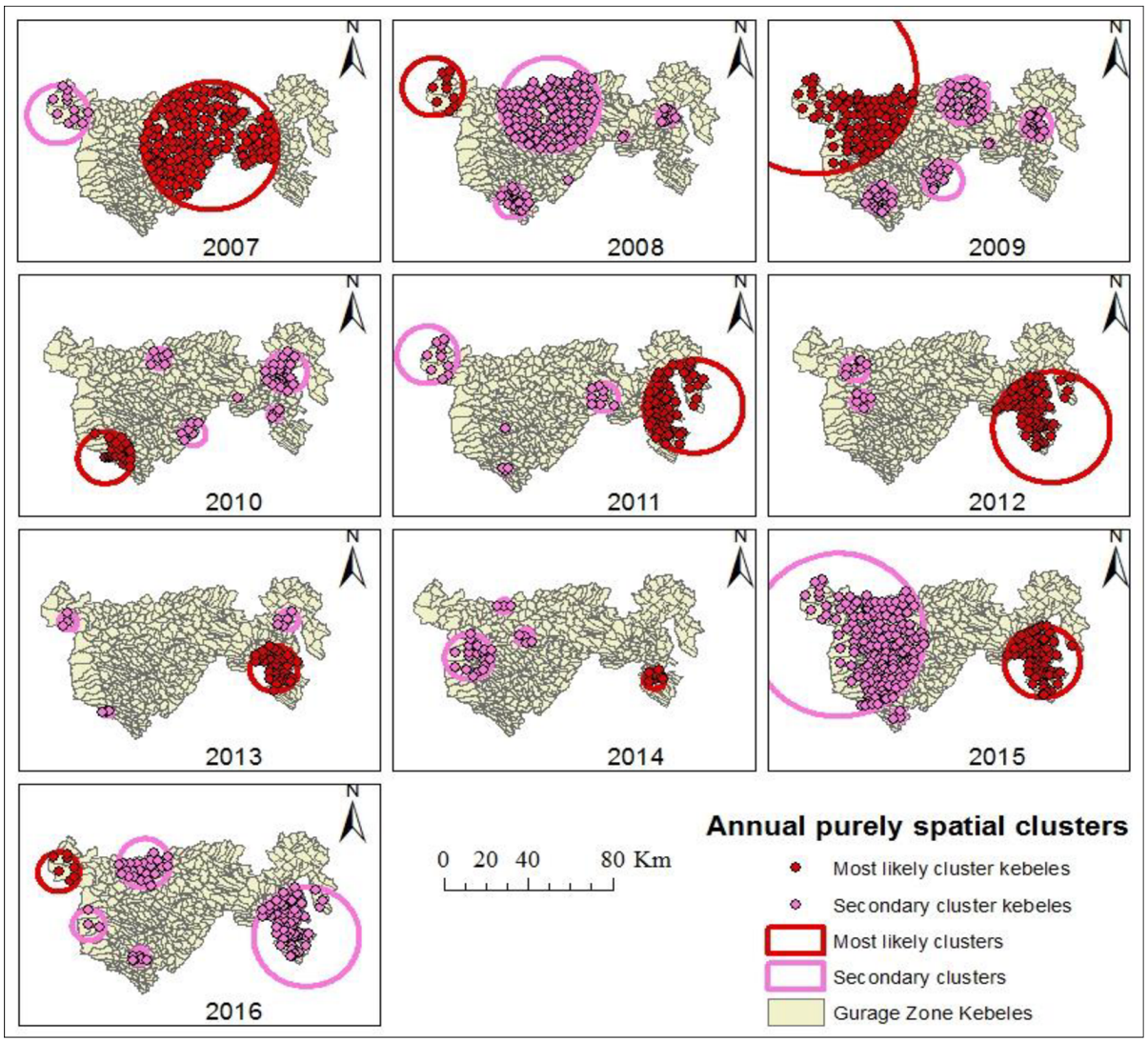


Fig 4. Annual purely spatial clusters for high occurrence of TB identified by using SaTScan statistic in Gurage Zone, Southern Ethiopia, 2007–2016.

https://doi.org/10.1371/journal.pone.0198353.g004

Space-time clusters

The space-time cluster analysis identified the clusters for high occurrence of TB at peripheral areas of the study area. The most likely cluster with 2,502 observed cases (1,353.51 expected cases) was detected for the period 2011 to 2015.The cluster covered Meskan, Butajira Town, Merako and southwest of Sodo areas within a radius of 17.92 km. People living within this cluster had 2.01 times higher risk of TB infection than those outside the cluster. A secondary cluster covering Endegagn, Enemor Ener, Cheha, Abeshege, Welkite Town, weastern Kebena, northern Ezha and western Mihur Aklil areas was detected for the period 2011 to 2015. The second secondary cluster was detected at central Gumer district for the period 2007 to 2011 (Table 4, Fig 6).

Discussion

This study aims to identify the location, size and risk of purely spatial and space-time clusters for high occurrence of TB in Gurage Zone, Southern Ethiopia during 2007 to 2016. The

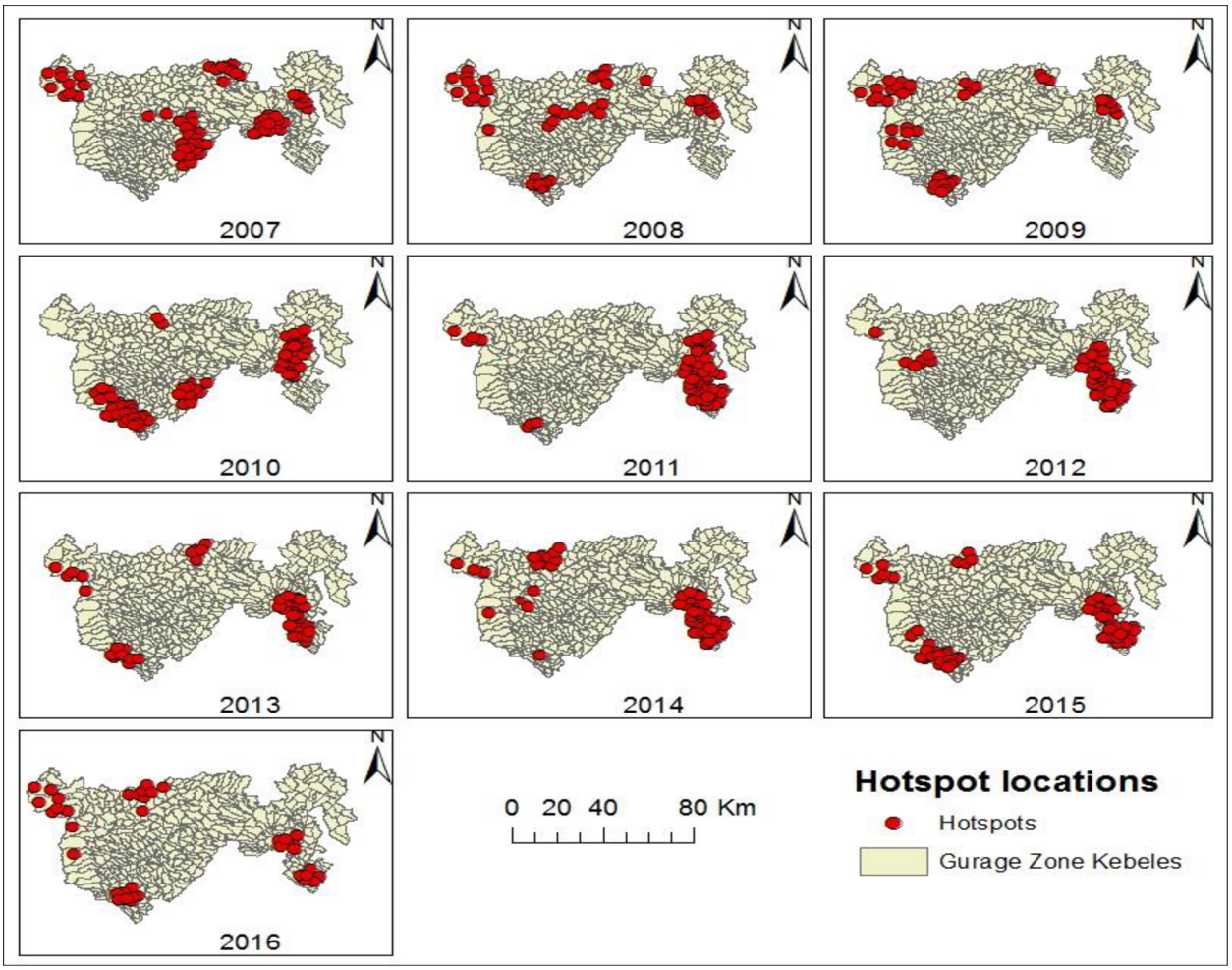


Fig 5. Spatial locations of significant hotspots of TB identified by using Getis-Ord G_i^* statistic in Gurage Zone, Southern Ethiopia, 2007–2016.

https://doi.org/10.1371/journal.pone.0198353.g005

clusters with high likelihood of TB occurrence were detected in border areas of the zone. The possible explanation for this could be that there were frequent cross-border population movements from the neighboring border areas of Jimma, Yem, Hadiya, Silte, West Shewa and East Shewa zones for economic and social reasons, which could favor the disease transmission in these areas. This is true according to other studies [24]. Therefore, future TB prevention and control efforts in these areas should include strengthening health infrastructure, staff capacity building, considering early diagnosis and treatment of symptomatic cases, and increasing community awareness. Furthermore, establishing a cross-border collaboration network may also help reduce the disease burden in these areas.

Table 4. Space-time clusters for high occurrence of TB in Gurage Zone, Southern Ethiopia, 2007–2016.

Cluster type	Cluster Year	Cluster center/radius	Observed cases	Expected cases	LLR	RR	P-value
Most likely cluster	2011–2015	(8.06 N, 38.51 E) / 17.92 km	2502	1353.51	435.61	2.01	<0.001
Secondary cluster	2011–2015	(8.12 N, 37.64 E) / 39.95 km	4117	3076.12	202.81	1.46	<0.001
2nd secondary	2007–2011	(7.97 N, 38.06 E) / 3.98 km	197	99.21	37.64	2.00	<0.001

https://doi.org/10.1371/journal.pone.0198353.t004

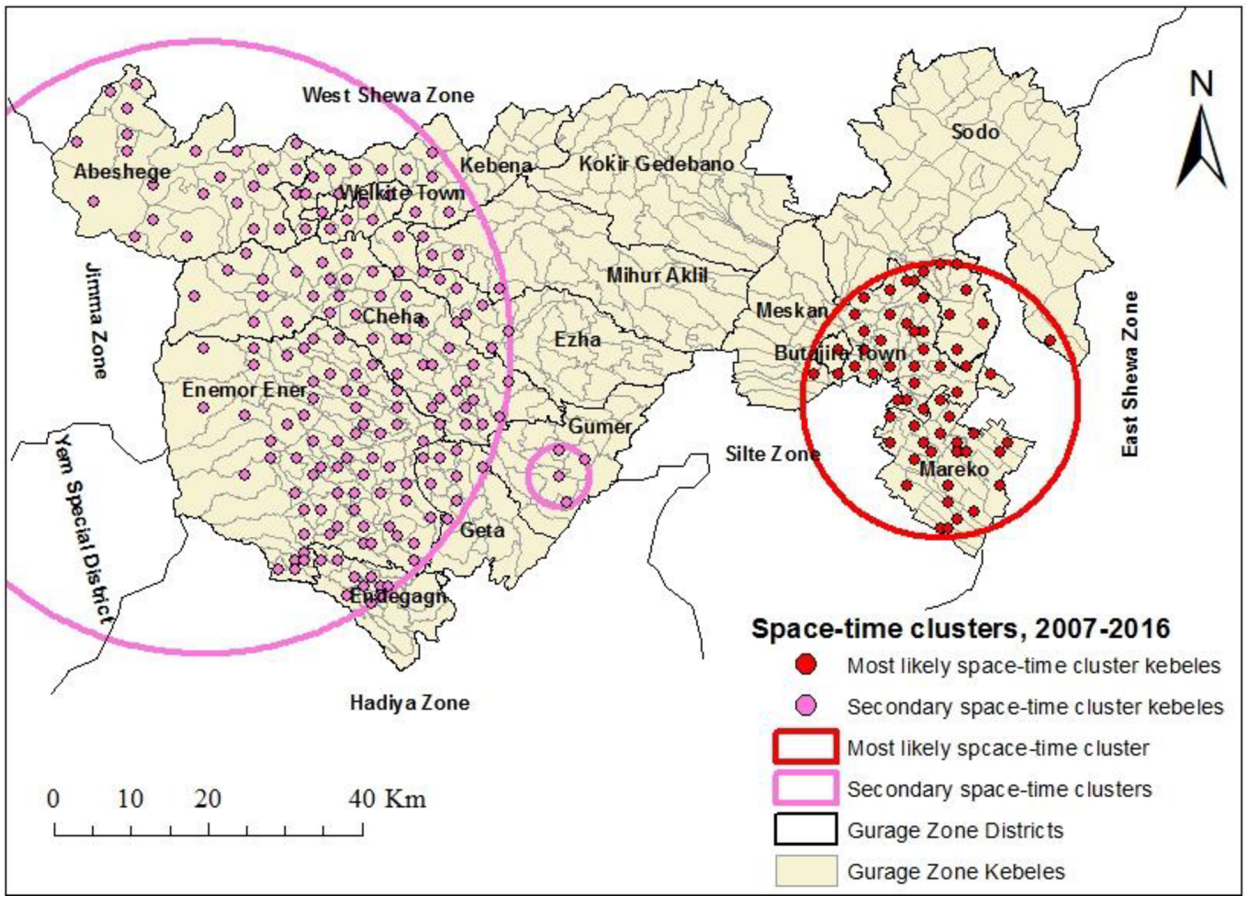


Fig 6. Significant space-time clusters for high occurrence of TB in Gurage Zone, Southern Ethiopia, 2007–2016.

https://doi.org/10.1371/journal.pone.0198353.g006

Three space-time clusters that were persistent for five years were detected in the zone. This might be due to the uniform implementation of TB prevention and control activities in the zone without targeting high-risk geographical areas [10]. Thus, application of GIS and spatial statistical techniques to identify purely spatial and space-time clusters for high occurrence of TB can be recommended for optimal utilization TB resources [25].

This study used the Kulldroff's scan and Getis-Ord G_i^* statistics to detect statistically significant clusters for high occurrence of TB in the Gurage Zone, Southern Ethiopia. The Kulldorff's scan statistic is widely used in the field of public health to detect purely spatial and space-time clusters of infectious diseases, including TB [9–11]. The method searches for disease clusters without prior hypothesis on their location, size or time period. Moreover, Monte-Carlo randomization technique for hypothesis testing gives the empirical joint distribution of the statistics and hence accounts for the correlation among the statistics, providing a P-value after taking into account multiple testing [26]. In this and previous studies Kulldroff's scan and Getis-Ord G_i^* statistics have generated comparable results in identifying geographic areas where unusually high rates of TB occurred [10, 27]. However, some inconsistencies were observed which may be due to the fact that the methods used varying assumptions in determining clusters. The Kulldroff's scan statistic calculated the maximum likelihood ratio of TB cases with relation to the underlying population in the area to detect larger clusters [23], whereas the Getis-Ord G_i^* statistic examined each distance-defined grouping of values to

identify more localized clusters [8]. Since there is no one gold standard method to detect disease clusters in spatial analyses it would be better to use more than one method at a time to cross-validate the results.

This study has some limitations. The study did not include TB patients who would remain undiagnosed, and those residents who were diagnosed and treated at health facilities outside the Gurage Zone. These could affect the nature of TB distribution by underestimating the prevalence. Another limitation of this study was that the Kulldroff's scan statistic used circular spatial scanning windows and space-time cylinders with circular spatial bases which could not detect irregular shaped clusters, and could include a few non-significant locations. The annual projected population based on the 2007 census was used to provide up-to-date denominator population numbers which could be affected by uneven population growth across the kebeles. Besides, the confounding effects of covariates, like age and sex were not controlled in the spatial analyses since it was not possible to access the geo-coordinates data for the individual patients at the unit TB registers. This could also bias the results.

The strength of this study was that multiple methods were used to detect spatial clusters for high occurrence of TB in the study area. A high resolution spatial data were used to examine the spatial and space-time clustering of TB at the smallest administrative unit, which might be the best approach for TB control planning. A longer study period data were used to evaluate the changes in spatial and space-time clustering of TB. The study covered a wider geographical area containing urban and rural areas. Errors related to the geo-coding of cases were avoided by linking each case to the correct home address using geo-codes from the CSA. The spatial data included all forms of TB to explore the high-risk geographical locations which require more focused public health attention.

Conclusions

This study showed that TB clusters were mainly concentrated at border areas of the Gurage Zone, suggesting that there has been sustained transmission of the disease within these locations. The findings may help intensify the implementation of TB control activities in these locations. Further study is warranted to explore the roles of various ecological factors on the observed spatial distribution of TB.

Supporting information

S1 Table. Data for spatial cluster analyses.
(XLSX)

S2 Table. Data for space-time cluster analyses.
(XLSX)

Acknowledgments

Authors would like to thank the Gurage Zone Department of Health, the data collectors and the personnel at the health facilities in the study area.

Author Contributions

Conceptualization: Sebsibe Tadesse, Fikre Enqueselassie, Seifu Hagos.

Data curation: Sebsibe Tadesse, Fikre Enqueselassie.

Formal analysis: Sebsibe Tadesse, Fikre Enqueselassie, Seifu Hagos.

Investigation: Sebsibe Tadesse, Seifu Hagos.

Methodology: Sebsibe Tadesse.

Software: Sebsibe Tadesse.

Supervision: Fikre Enqueselassie, Seifu Hagos.

Writing – original draft: Sebsibe Tadesse, Fikre Enqueselassie, Seifu Hagos.

References

1. World Health Organization. Global tuberculosis report. Geneva: Switzerland: World Health Organization; 2017.
2. Federal Ministry of Health. First Ethiopian national population-based tuberculosis prevalence survey. Addis Ababa, Ethiopia: Federal Ministry of Health; 2011.
3. Barter DM, Agboola SO, Murray MB, Bärnighausen T. Tuberculosis and poverty: the contribution of patient costs in Subsaharan Africa -a systematic review. BMC Public Health 2012; 12:980. https://doi.org/10.1186/1471-2458-12-980 PMID: 23150901
4. Ali E, Woldie M. Reasons and outcomes of admissions to the medical wards of Jimma University Specialized Hospital, Southwest Ethiopia. Ethiop J Health Sci 2010; 20(2):113–120. PMID: 22434969
5. Melaku YA, Sahle BW, Tesfay FH, Bezabih AM, Aregay A, Abera SF, et al. Causes of death among adults in Northern Ethiopia: Evidence from verbal autopsy data in health and demographic surveillance system. PLoS One 2014; 9(9): e106781. https://doi.org/10.1371/journal.pone.0106781 PMID: 25188025
6. Tadesse S. Stigma against tuberculosis patients in Addis Ababa, Ethiopia. PLoS One 2016; 11(4): e0152900. https://doi.org/10.1371/journal.pone.0152900 PMID: 27054714
7. Tadesse S, Tadesse T. HIV co-infection among tuberculosis patients in Dabat, northwest Ethiopia. J Infect Dis Immun 2013; 5(3):29–32.
8. Liu Y, Li X, Wang W, Li Z, Hou M, He Y, et al. Investigation of space-time clusters and geospatial hot spots for the occurrence of tuberculosis in Beijing. Int J Tuberc lung dis 2012; 16 (4):486–491. https://doi.org/10.5588/ijtld.11.0255 PMID: 22325066
9. Tadesse T, Demissie M, Berhane Y, Kebede Y, Abebe M. The clustering of smear-positive tuberculosis in Dabat, Ethiopia: a population based cross sectional study. PloS One 2013; 8(5):e65022. https://doi.org/10.1371/journal.pone.0065022 PMID: 23717686
10. Dangisso MH, Datiko DG, Lindtjorn B. Spatio-temporal analysis of smear-positive tuberculosis in the Sidama Zone, Southern Ethiopia. PLoS One 2015; 10(6):e0126369. https://doi.org/10.1371/journal.pone.0126369 PMID: 26030162
11. Zhao F, Cheng S, He G, Huang F, Zhang H, Xu B, et al. Space-time clustering characteristics of tuberculosis in China, 2005–2011. PLoS One 2013; 8(12):e83605. https://doi.org/10.1371/journal.pone.0083605 PMID: 24367604
12. Gómez-Barroso D, Rodriguez-Valín E, Ramis R, Cano R. Spatio-temporal analysis of tuberculosis in Spain, 2008–2010. Int J Tuberc Lung Dis 2013; 17(6):745–751. https://doi.org/10.5588/ijtld.12.0702 PMID: 23676156
13. Shargie EB, Yassin MA, Lindtjorn B. Prevalence of smear-positive pulmonary tuberculosis in a rural district of Ethiopia. Int J Tuberc Lung Dis 2006; 10(1):87–92. PMID: 16466043
14. Tadesse T, Demissie M, Berhane Y, Kebede Y, Abebe M. Two-thirds of smear-positive tuberculosis cases in the community were undiagnosed in Northwest Ethiopia: population based cross-sectional study. PLoS One 2011; 6(12):e28258. https://doi.org/10.1371/journal.pone.0028258 PMID: 22164256
15. Demissie M, Zenebere B, Berhane Y, Lindtjorn B. A rapid survey to determine the prevalence of smear-positive tuberculosis in Addis Ababa. Int J Tuberc Lung Dis 2002; 6(7):580–584. PMID: 12102296
16. Shargie EB, Morkve O, Lindtjorn B. Tuberculosis case-finding through a village outreach programme in a rural setting in southern Ethiopia: community randomized trial. Bull World Health Organization 2006; 84(2):112–119.
17. Southern Nations Nationalities and Peoples Regional Investment Department. Investment opportunities. Investment Department. Accessed on: November 6, 2016. Available from: http://www.southinvest.org.et/publications/Von2009/Investment%20reviesd%20Guide%20line%20in%20English%202005.pdf.
18. Central Statistical Agency. The 2007 Population and Housing Census of Ethiopia: Statistical Report for Southern Nations, Nationalities and Peoples' Region. Addis Ababa, Ethiopia: Central Statistical

Agency, 2007. Available at: http://www.csa.gov.et/census-report/complete-report/census-2007.html?start=10; Accessed on: April 2017.

19. Southern Nations Nationalities and Peoples Regional Health Department. The -2016 health report. Awassa, Ethiopia; 2017.
20. Dangisso MH, Datiko DG, Lindtjorn B. Accessibility to tuberculosis control services and tuberculosis programme performance in southern Ethiopia. Glob Health Action 2015; 8:29443. https://doi.org/10.3402/gha.v8.29443 PMID: 26593274
21. Balabanova D, McKee M, Mills A. Good health at low cost 25 years on. What makes a successful health system? London: LSHTM; 2011.
22. Federal Democratic Republic of Ethiopia Ministry of Health. Guidelines for clinical and programmatic management of TB, leprosy and TB/HIV in Ethiopia. Addis Ababa, Ethiopia: Ministry of Health; 2012.
23. Kulldorff M. SaTScanTM User Guide for version 9.2; 2010.
24. Kozińska M, Zientek J, Augustynowicz-Kopeć E, Zwolska Z, Kozielski J. Transmission of tuberculosis among people living in the border areas of Poland, the Czech Republic, and Slovakia. Pol Arch Med Wewn 2016; 126 (1–2):32–40. PMID: 26842376
25. Tiwari N, Adhikari CMS, Tewari A, Kandpal V. Investigation of geo-spatial hotspots for the occurrence of tuberculosis in Almora district, India, using GIS and spatial scan statistic. Int J Health Geogr 2006; 5:33. https://doi.org/10.1186/1476-072X-5-33 PMID: 16901341
26. Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. PLoS Med 2005; 2:e59. https://doi.org/10.1371/journal.pmed.0020059 PMID: 15719066
27. Ge E, Zhang X, Wang X, Wei X. Spatial and temporal analysis of tuberculosis in Zhejiang Province, China, 2009–2012. Infect Dis Poverty 2016; 5:11. https://doi.org/10.1186/s40249-016-0104-2 PMID: 26906041

Paper II

Tadesse et al. BMC Public Health (2018) 18:783
https://doi.org/10.1186/s12889-018-5711-3

BMC Public Health

RESEARCH ARTICLE Open Access

Estimating the spatial risk of tuberculosis distribution in Gurage zone, southern Ethiopia: a geostatistical kriging approach

Sebsibe Tadesse[1*], Fikre Enqueselassie[2] and Seifu Hagos Gebreyesus[2]

Abstract

Background: In low-income countries it is difficult to obtain complete data that show spatial heterogeneity in the risk of tuberculosis within-and-between smaller administrative units. This may contribute to the partial effectiveness of tuberculosis control programs. The aim of this study was to estimate the spatial risk of tuberculosis distribution in Gurage Zone, Southern Ethiopia using limited spatial datasets.

Methods: A total of 1601 patient data that were retrieved from unit tuberculosis registers were included in the final analyses. The population and geo-location data were obtained from the Central Statistical Agency of Ethiopia. Altitude data were extracted from ASTER Global Digital Elevation Model Version 2. Aggregated datasets from sample of 169(40%), 254(60%) and 338(80%) kebeles were used to estimate the spatial risk of TB distribution in the Gurage Zone by using a geostatistical kriging approach. The best set of input parameters were decided based on the lowest prediction error criteria of the cross-validation technique. ArcGIS 10.2 was used for the spatial data analyses.

Results: The best semivariogram models were the Pentaspherical, Rational Quadratic, and K-Bessel for the 40, 60 and 80% spatial datasets, respectively. The predictive accuracies of the models have improved with the true anisotropy, altitude and latitude covariates, the change in detrending pattern from local to global, and the increase in size of spatial dataset. The risk of tuberculosis was estimated to be higher at western, northwest, southwest and southeast parts of the study area, and crossed between high and low at west-central parts.

Conclusion: This study has underlined that the geostatistical kriging approach can be applied to estimate the spatial risk of tuberculosis distribution in data limited settings. The estimation results may help local public health authorities measure burden of the disease at all locations, identify geographical areas that require more attention, and evaluate the impacts of intervention programs.

Keywords: Geostatistical kriging, Risk of tuberculosis, Semivariogram model, Spatial heterogeneity

Background

Tuberculosis (TB) continues to place an extraordinary public health, financial and social burden on those afflicted by the disease and their families, and on government. In 2016, there were an estimated 10.4 million incident cases and 1.7 million deaths worldwide [1]. The global distribution of the disease is skewed heavily toward low-and-middle income countries, which accounted for about 87% of all estimated incident cases. Ethiopia is a low-income country in east Africa that remains highly afflicted by TB and is ranked among the list of 14 countries with high burden of TB, Human Immunodeficiency Virus (HIV)-associated TB (TB/HIV) and Drug Resistant TB (DR-TB) [1].

It is difficult to obtain spatially complete data on TB in Ethiopia [1, 2]. About 36% of the estimated TB cases were not notified to the national TB program in 2016. The national TB prevalence surveys were conducted through sampling a limited number of locations due to logistical and financial limitations [3, 4]. Moreover, other regional reports did also not show continuous spatial distribution and burden of the disease within-and-between smaller geographical

* Correspondence:
[1]Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia
Full list of author information is available at the end of the article

locations [5, 6]. This presents a substantial obstacle to measure burden of the disease, identify high-risk geographical locations, evaluate the impacts of intervention programs and allocate public health resources.

Spatial interpolation is the process of estimating values for a variable of interest at unmeasured locations using data from the surrounding locations [7]. All the spatial interpolation models share a common underlying assumption of the closer values are more related than the distant ones [8]. These models are divided into two categories, deterministic and geostatistical models [9]. The deterministic models estimate between measured values using mathematical formulas that vary the smoothness of the estimated surface. The spatial correlation of the data is not considered in the estimation. Consequently, deterministic models do not estimate the uncertainty of predictions. Conversely, geostatistical models consider the spatial correlation of the dataset. Rather than giving interpolation weights based on arbitrary formulas, the semivariogram models are used to found out weights from the observed data [10]. The weights dictate how each measured value contributes to the interpolated value at unmeasured location. Geostatistical models produce prediction estimates and associated prediction errors at all unmeasured locations. Kriging is the most robust and widely used geostatistical method of interpolation in many fields of science [11]. It is known as the optimal interpolation method because it minimizes the mean square error of predictions and is statistically unbiased (i.e., estimated values and measured values agree on average) [12]. Recently, there have been several applications of kriging in the area of public health for estimating the predicted risk surface of infectious diseases, such as TB [3], malaria [11], cholera [13], helminths [14] and schistosomiasis [15].

The aim of this study was to estimate the spatial risk of TB distribution in Gurage Zone, Southern Ethiopia. Three geostatistical kriging models were fitted with 40, 60 and 80% of spatially aggregated TB dataset co-impacted by geographical factors. The estimated risk map may help local authorities as a guide for planning, budgeting and resource mobilization. The graphical demonstration may be a good tool for advocacy since stakeholders can easily identify the spatial structure of the disease by watching over the map and may be interested to implement geographically targeted interventions. Furthermore, the study findings may also contribute for the growing body of geostatistical research on TB.

Methods

Study area

This study was conducted in the Gurage Zone in southern Ethiopia, which is located between 7°76′ and 8°45’ N latitude and 37°46′ and 38°71′ E longitude (Fig. 1). The zone has 13 districts and two town administrations (at Butajira and Wolkitie). It covers an area of about 5932 km^2. There are 403 rural and 20 urban kebeles (the smallest administrative units with a population of 5000 on average) in the zone. There were a total of 1,542,131 populations in 2016, about 84% of which live in the rural areas [16].

There are a total of 6 hospitals, 70 health centers, 414 health posts and 92 clinics in the zone that provide TB prevention and control services [16]. The clinics and the health posts provide community education, identify and refer presumptive TB cases to health facilities for further investigation, give *Bacillus Calmette-Guérin* vaccination, contact locating and screening, trace and link lost to follow up cases, and support treatment adherence through female health extension workers. The health centers carry out all activities as health posts and clinics, and additionally provide intensified case finding, sputum microscopy services, provide isoniazid preventive therapy for eligible persons, diagnose and manage adverse drug reactions and other complications, carry out TB/HIV collaborative activities, refer smear negative presumptive TB, extra-pulmonary TB and DR-TB patients to higher level facilities, provide support to health post staff, keep patient records and manage medicines stocks, plan and implement TB infection control. Health centers additionally provide Directly Observed Treatment-Short courses (DOTS) services for patients with DR-TB referred by treatment initiation centers. The hospitals carry out activities as health centers, and additionally provide referral services and admission care for seriously ill TB patients. Selected hospitals provide diagnosis and treatment for DR-TB patients, including inpatient care. The GeneXpert machines are installed at hospital laboratories. Private health facilities are also engaged in TB diagnosis, treatment and/or referral of presumptive TB and DR-TB cases depending on their capacity [17].

Data sources

The list of health facilities providing DOTS services were obtained from the Health Department database of Gurage Zone. All TB patients who were residents of the zone and registered at the health facilities during January to December, 2016 were included in the study. The patient data were retrieved from the unit TB registers from June to September, 2017. The patients’ addresses were checked for duplication and linked to their true geo-locations. The data on geo-location and population of each kebele in the zone were accessed from the Central Statistical Agency of Ethiopia (CSA). Altitude of each kebele was extracted from ASTER Global Digital Elevation Model V2 [18].

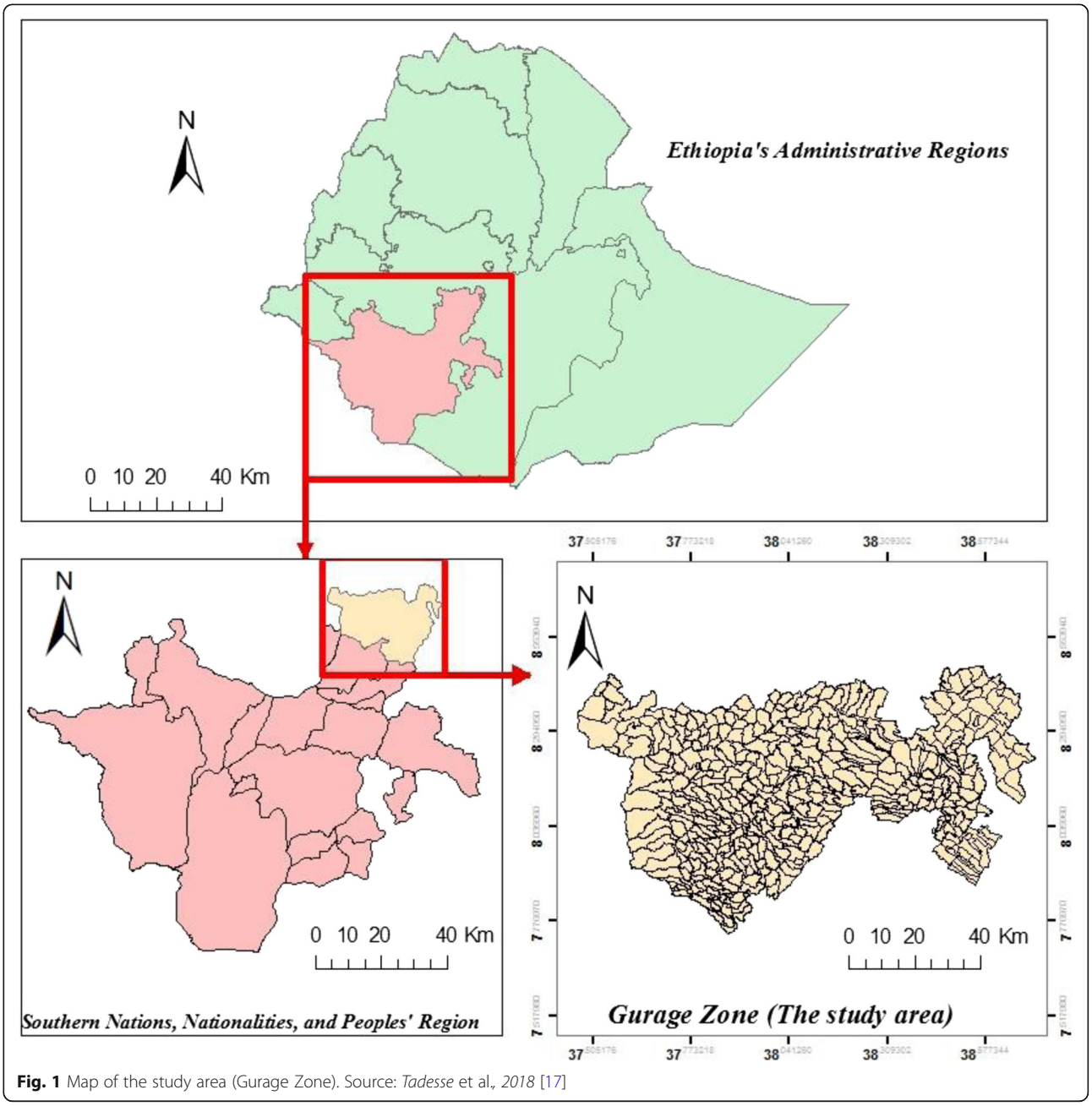


Fig. 1 Map of the study area (Gurage Zone). Source: *Tadesse* et al., *2018* [17]

TB diagnosis and case definition

The diagnostic criteria of the national TB diagnosis guideline of Ethiopia were used to diagnose the TB cases [19].

Smear-positive pulmonary TB (PTB+): is diagnosed when at least 2 initial sputum smear examinations are positive for Acid Fast Bacilli (AFB) or 1 smear-positive result for AFB and culture-positive result for *M. tuberculosis* or 1 smear-positive result for AFB and radiographic abnormalities indicative of active TB, in addition to a clinician's judgment. The regional laboratory carries out external quality assurance on all slides, and provides a feedback to the health facility providing DOTS services.

Smear-negative pulmonary TB (PTB-): is diagnosed when there are symptoms evocative of TB, at least 3 smear-negative initial results for AFB, lack of response to antibiotics, smear-negative and radiological abnormalities indicative of pulmonary TB, and judgment of a clinician.

Extra-pulmonary TB (ETB): is diagnosed when a specimen from an extra-pulmonary site is culture positive or histo-pathological abnormality from a biopsy, and strong clinical evidence indicative of active ETB.

However, because of insufficient laboratories for histo-pathological or culture examinations, most of the health facilities diagnose ETB based on a clinician's judgment.

Newly diagnosed case of TB: is a patient who has never taken anti-TB drugs or taken for less than a month.

Retreatment TB case: is a patient who has previous treatment failure, or relapse or default.

Data quality control

Supervisors and data collectors were trained on the field methods, data extraction and record keeping. The data completeness and consistency was checked page-by-page by health facilities, kebele, and district against unit TB registers.

Data management and processing

Data were entered, validated, cleaned, and coded using MS Excel (MicroSoft, Redmond, WA, USA). The patients' data were linked to their true address using CSA codes, and aggregated at kebele level. The aggregated dataset from a total of 423(100%) kebeles were used to examine the actual spatial risk of TB distribution. Then, the aggregated TB datasets co-impacted by geographic factors from sample of 169(40%), 254(60%) and 338(80%) kebeles were used to estimate the spatial risk of TB distribution and associated standard error by using a geostatistical kriging approach (Additional file 1). A simple random sampling technique was used to select the spatial sample kebeles. Geographically weighted central locations were represented by the kebele centroids as coordinates. A table containing the number of TB cases, the population, the coordinates and the prevalence rates (the number of TB cases divided by the population of a given year and multiplied by 100,000) were prepared, and were joined to ArcGIS 10.2.

Spatial smoothing

Spatial Empirical Bayes Smoothing (SEBS) method was employed in Geographic Data analysis tool (GeoDa) in order to overcome small areas variance instability, which is due to variations in population size as well as few cases of TB in some areas [20]. The population for each kebele was used as a base variable and number of TB cases was used as an event. A queen weights matrix that defines the neighboring kebeles as those with either a shared border or vertex was used for spatial weights [20]. The SEBS method was not applied for the datasets that were used for spatial prediction since the geostatistical kriging would result smoothed estimates by using a weighted linear combination of the known measured values.

Ordinary kriging

Varieties of kriging have been developed, such as ordinary, universal, simple and indicator. Ordinary kriging was preferred to other types of kriging because it predicts an estimate for unsampled kebele by assuming a constant mean in the local neighborhood of each estimation kebele, which is a characteristic of focal diseases like TB. Besides, it is a good geostatistical method to model data that exhibit spatial trend [21]. It uses a semivariogram model to measure spatial autocorrelation between pairs of prevalence rates as follows [10]:

$$\gamma(\mathrm{h}) = \frac{1}{2\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{Z}(\mathrm{x})-\mathrm{Z}(\mathrm{x}+\mathrm{h}))^2 \quad (1)$$

where n is the total number of pairs of sample kebeles, $Z(x)$ and $Z(x+h)$ are the prevalence rates at any two kebeles x and $x+h$ separated by distance h. Calculations of $\gamma(h)$ are repeated for *2 h, 3 h, 4 h, ..., kh*. The models of spatial autocorrelation commonly exhibit similar characteristics, which are called the sill, range, and nugget. The sill is the maximum variability between pairs of prevalence rates. The separation distance at which the sill is reached is termed the range and represents the maximum distance beyond which prevalence rates are spatially independent. The nugget effect refers to the situation in which the difference between prevalence rates taken at sampling kebeles that are close together is not zero. It represents spatial sources of variation at distances smaller than the sampling interval (i.e. spatial variations of prevalence rates at village level, which is a spatial subset of kebele) or measurement error (e.g. passive case detection).

As described in detail previously [12], an unknown prevalence rate $\hat{Z}_u$ at kebele u is estimated as a weighted-linear combination of n known samples as follows:

$$\hat{Z}_u = \sum_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{W_i}\mathrm{Z_i} \quad (2)$$

where

$$\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{W_i} = 1$$

The optimal weights which produce the minimum estimation error in eq. (2) can be determined by using the following simultaneous equations:

$$\begin{array}{ccccccc} \mathrm{W_i}\gamma(\mathrm{h}_{1,1}) & +\cdots & +\mathrm{W_n}\gamma(\mathrm{h}_{1,\mathrm{n}}) & + & \lambda = & \gamma(\mathrm{h}_{1,\mathrm{u}}) \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ \mathrm{W_1}\gamma(\mathrm{h}_{\mathrm{n},1}) & +\cdots & +\mathrm{W_n}\gamma(\mathrm{h}_{\mathrm{n},\mathrm{n}}) & + & \lambda = & \gamma(\mathrm{h}_{\mathrm{n.u}}) \\ \mathrm{W_1} & +\cdots & +\mathrm{W_n} & & = & \gamma(\mathrm{h}_{\mathrm{n.u}}) \end{array} \quad (3)$$

where $\gamma(h_{i,\ j})$ is a semivariogram model which is a function of distance $h_{i,\ j}$ between prevalence rates i and j,

and λ is the Lagrange Multiplier to minimize the kriging error. The correlation between prevalence rates i and j is expected to decrease as their separation distance $h_{i,\ j}$ increases. The optimal weights in eq. (2) are calculated as follows:

$$\begin{bmatrix} W_1 \\ \vdots \\ W_n \\ \lambda \end{bmatrix} = \begin{bmatrix} \gamma(h_{1,1}) & \cdots & \gamma(h_{1,n}) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(h_{n,1}) & \cdots & \gamma(h_{n,n}) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma(h_{1,u}) \\ \vdots \\ \gamma(h_{n,u}) \\ 1 \end{bmatrix} \quad (4)$$

Therefore, ordinary kriging produces an unbiased estimate with minimum variance.

Ordinary cokriging

Ordinary cokriging is an extension of ordinary kriging method that uses both the spatial autocorrelation for prevalence rate (i.e. the main variable of interest) and the spatial cross-correlations between prevalence rate and geographic variables (i.e. altitude, latitude and longitude) to make estimations of the prevalence rates at unsampled kebeles. The development of the ordinary cokriging system is identical to the development of ordinary kriging system. The mathematical formulation of ordinary cokriging has been described in detail by Yalcin [22].

In this study both ordinary kriging and ordinary cokriging models were tested for the three categories of datasets, and ordinary cokriging models were selected as the best-fitted ones.

Model selection

In this study the effects of the different types of semivariogram models (i.e., stable, spherical, circular, tetraspherical, pentaspherical, Gaussian, exponential, rational quadratic, K-Bessel, hole effect and J-Bessel), detrending (i.e., neighborhood, global and local), anisotropy (i.e., false and true) and geographic covariates (i.e., longitude, latitude and altitude) on the predictive performance of kriging were checked by using a cross-validation technique. The technique leaves and adds each sample points in the dataset turn by turn to provide pairs of predicted and measured values that can be compared to evaluate the model's performance. A total of 528 geostatistical kriging models were generated for each category of spatial dataset (i.e., 40, 60 and 80%) (Additional file 2). The final models for each category of the spatial dataset were decided based on the lowest total error, obtained by sorting values of Root-Mean-Square Error (RMSE), absolute value of Mean-Standardized Error (MSE), Root-Mean-Square-Standardized Error (RMSSE) and absolute value of the difference of Average-Standard Error (ASE) from RMSE in ascending order, and then ranking and summing up the ranks. All these errors are expressed by eqs. (5)-(8) below [23]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[Z^*(x_i) - Z(x_i)]^2} \quad (5)$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{Z^*(x_i) - Z(x_i)}{\sigma^2(x_i)}\right] \quad (6)$$

$$RMSSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{Z^*(x_i) - Z(x_i)}{\sigma^2(x_i)}\right]^2} \quad (7)$$

$$ASE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\sigma^2(x_i)} \quad (8)$$

where $\sigma^2(x_i)$ is the kriging variance for location x_i, and $Z^*(x_i)$ and $Z(x_i)$ are the predicted and the sampled values at the location x_i, respectively.

Sensitivity analyses

The Semivariogram Sensitivity tool, which is found under the Geostatistical Analyst toolbox of ArcGIS 10.2, was used to perform sensitivity analyses on the predicted values and associated Standard Errors (SE) by varying the nugget and range within a percentage of the original values. The outputs of the analyses were a table indicating which parameter values were used and what the resulting predicted and standard error values were. Small fluctuations in the output with small changes in the input parameter values indicate more confident predictions which can be used to make decisions.

Risk measurement

The prevalence rate was used as a proxy variable to estimate the risk of TB in the study area. The estimated risk surface was categorized as *a low risk area* where the prevalence rate was below and equal to 100 cases per 100,000 population, and *a high risk area* where the prevalence rate was above 100 cases per 100,000 population [20].

Results

Patient characteristics

A total of 1626 TB cases were diagnosed during January to December, 2016. Only 1.6% of them were excluded from the final analyses because of incomplete addresses or being outside of the study area. Out of 1601 cases included in this study 57.5% were males and 42.5% were women, yielding a male to female ratio of 1.3:1. The mean age with a standard deviation was 36 ± 17 years for all cases, 34 ± 16 years for males and 38 ± 17 years for females. The majority, 89.6%, of the cases were newly diagnosed, while 10.4% were retreatment cases. Of the cases 41.2% were PTB+, 31.9% PTB- and 26.9% ETB. Residentially, 86.6% were from rural areas.

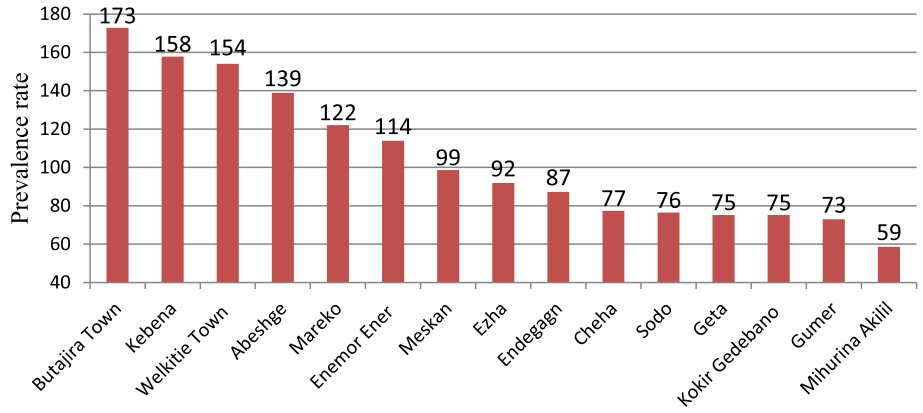


Fig. 2 TB prevalence rates by districts in Gurage Zone

The actual spatial risk of TB distribution

The risk distribution of TB varied from 59 to 173 cases per 100,000 population across districts of the Gurage Zone (Fig. 2). Moreover, the smoothed rates of TB varied from zero to 634 cases per 100,000 population across kebeles of the zone. High risk of TB was observed at northwest, western, southwest and southeast parts. The risk distribution crossed between high and low at west-central parts (Fig. 3).

The best-fitted models

The best geostatistical kriging models were decided to be: 1) Pentaspherical semivariogram, local detrending, true anisotropy and altitude and latitude covariates for modeling with 40% of spatial dataset, 2) Rational Quadratic semivariogram, local detrending, true anisotropy and altitude and latitude covariates for modeling with 60% of spatial dataset, and 3) K-Bessel semivariogram, global detrending, true anisotropy and altitude and latitude covariates for modeling with 80% of spatial dataset. The detrending pattern of the models changed from local to global as the size of spatial dataset increased. Moreover, the models predictive accuracies also improved as the size of spatial dataset increased, which was indicated by 0 MSE, 1 RMSSE, and ASE approached RMSE (i.e., the variability in prediction is correctly assessed) (Table 1).

Sensitivity analyses outputs

The parameter values for nugget and range from the input geostatistical model sources were 8123.85 and 72,891.45 for the model with 40% spatial dataset, 7178.77 and 78,808.04 for the model with 60% spatial dataset, and 7210.46 and 78,767.27 for the model with

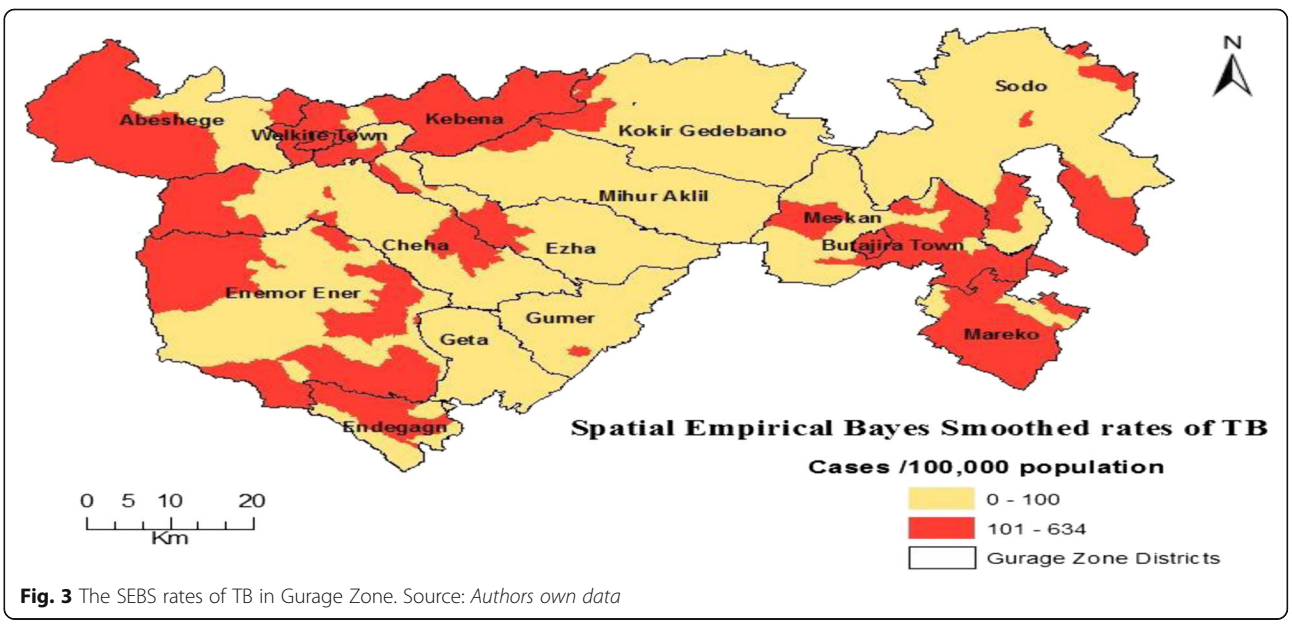


Fig. 3 The SEBS rates of TB in Gurage Zone. Source: *Authors own data*

Table 1 Comparison of cross-validation statistics for TB spatial datasets in Gurage Zone, Southern Ethiopia, 2017

Cross-validation statistics	Ordinary cokriging models		
	With 40% dataset	With 60% dataset	With 80% dataset
MSE	0	0	0
RMSSE	1	1	1
RMSE	89	88	87
ASE	93	87	87

80% spatial dataset, respectively. Five random nugget and range values that were found within 10% of the input models' nugget and range values were calculated for each dataset and used as input parameters. There were only small fluctuations in the prediction outputs for the corresponding input parameters, indicating more accurate predictive performance of the models (Table 2).

The estimated spatial risk of TB distribution

The ordinary cokriging models with 40, 60 and 80% of the spatial datasets estimated high risk of TB at northwest, western, southwest and southeast parts of the Gurage Zone. The risk distribution crossed between high and low at west-central parts**.** Moreover, the models estimated high uncertainties of prediction at border areas of the zone, the magnitude of which decreased as the spatial dataset increased (Fig. 4).

Comparison between estimated and actual spatial risk of TB distribution

The three estimation models identified areas for high risk of TB at locations that were closely similar to the actual high-risk areas, with reasonable predictive accuracies. These locations included northwest, western, southwest and southeast parts of the Gurage Zone. The risk distribution crossed between high and low at west-central parts (see Figs 3 and 4 above).

Discussion

This study has underscored that the geostatistical kriging approach can be applied to estimate the spatial risk of TB distribution in settings where spatially limited data are available. The estimation models indicated that there was spatial heterogeneity in the risk of TB distribution in the Gurage Zone, indicating the disease did not affect all of the communities in the area with the same severity. The risk was higher in northwest, western, southwest and southeast parts of the zone. However, the risk distribution interlocked between high and low at west-central parts. Evidences have revealed that differences in underlying socioeconomic, climatic and geographic conditions, and uneven allocation of public health resources could contribute for the spatial heterogeneity in the risk of TB distribution [3, 24, 25].

Table 2 The semivariogram sensitivity analyses results for TB spatial datasets in Gurage Zone, Southern Ethiopia, 2017

Model	Random Parameter	Prediction	SE	Nugget	Range
Modeling with 40% dataset	Nugget	79.00	22.95	7898.75	72,891.45
	Nugget	78.99	23.34	8170.68	72,891.45
	Nugget	79.01	22.66	7702.39	72,891.45
	Nugget	79.01	22.35	7467.51	72,891.45
	Nugget	79.02	22.47	7572.18	72,891.45
	Range	78.99	23.18	8062.43	76,769.04
	Range	78.65	23.19	8068.52	79,894.51
	Range	78.66	23.17	8050.92	72,346.21
	Range	78.66	23.17	8051.90	72,671.44
	Range	78.65	23.19	8068.39	79,821.99
Modeling with 60% dataset	Nugget	67.35	21.45	6995.20	78,808.04
	Nugget	67.63	20.94	7360.37	78,808.04
	Nugget	67.57	21.05	7284.92	78,808.04
	Nugget	67.66	20.90	7425.62	78,808.04
	Nugget	67.68	21.24	7669.06	78,808.04
	Range	67.44	21.16	7018.50	80,198.91
	Range	83.57	21.03	6920.55	73,192.36
	Range	67.55	21.00	7076.75	85,175.07
	Range	77.54	20.82	6979.91	77,256.03
	Range	67.55	21.00	7076.00	85,106.29
Modeling with 80% dataset	Nugget	77.35	20.49	7017.76	78,767.27
	Nugget	77.46	20.86	7356.95	78,767.27
	Nugget	77.51	21.51	7865.18	78,767.27
	Nugget	77.20	20.02	6603.89	78,767.27
	Nugget	77.45	20.84	7332.31	78,767.27
	Range	77.82	20.60	7212.61	80,939.15
	Range	77.82	20.60	7212.40	80,888.19
	Range	74.71	19.43	7172.59	72,233.82
	Range	74.71	19.44	7179.62	73,658.34
	Range	77.82	20.61	7221.91	83,206.47

Moreover, the cross-border population movements from the neighboring border areas could also facilitate the high transmission of TB, especially at border areas of the zone [26–30]. Therefore, the estimated risk map of TB may help local public health authorities prioritize locations that required immediate interventions.

This study revealed that cokriging with altitude and latitude were the best geostatistical models, which suggested that including these covariables improved the predictive accuracies of the models. This reflects that geographical factors can affect the risk distribution of TB in the Gurage Zone. Previous studies have highlighted that the geographical factors had explicit impacts on the risk distribution of TB [3, 24, 25, 31, 32].

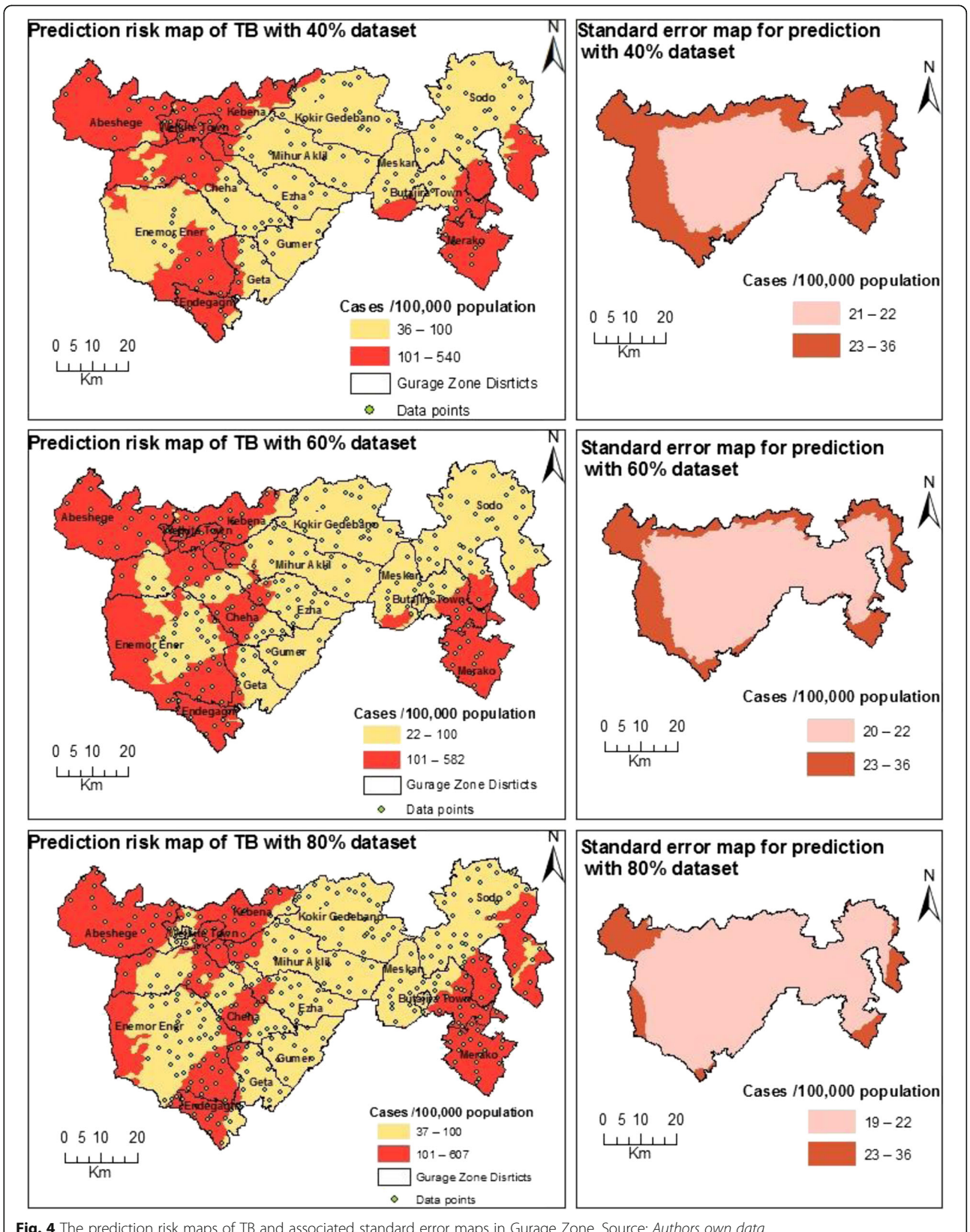


Fig. 4 The prediction risk maps of TB and associated standard error maps in Gurage Zone. Source: *Authors own data*

Thus, impacts of geographical factors on TB prevention and control should be evaluated, and interventions should be formulated based on geographical features.

In the present study the prediction standard error values were relatively higher in the western, northwest, southwest and southeast parts than in the others. This could be due to the fact that the spatial distribution of TB risk was higher in the western, northwest, southwest and southeast parts of the study area and kriging method underestimated the higher values [3]. The other reason could be that the majority of the spatial sample points were sparser in areas with higher prediction standard error [3]. The lack of data point beyond the borders of the study area could also explain the higher prediction standard error [12]. Therefore, it would be better to take sample data points with better spatial distribution beyond or on the boundaries of these locations in order to obtain more accurate and stable kriging surface estimates.

This study has also practical implications to TB prevention and control programs in low-income countries, where obtaining spatially complete TB data is difficult. The recent advancements in geostatistical modeling techniques and increasing availability of public health data from the national prevalence surveys, demographic and health system surveys and health facilities will be the good opportunities for epidemiologists working in such settings to predict the spatial risk of TB distribution and associated prediction uncertainty at non-surveyed locations [3, 8, 11]. The resulting prediction risk map may allow them measure burden of the disease at all locations, identify high-risk geographical areas for targeted interventions, and evaluate the impacts of intervention programs. This will be useful for optimal utilization of the scarce public health resource.

This study has some limitations. The estimated risk of TB might be underestimated in some areas because the study did not include those patients who would remain undiagnosed for the disease, and those diagnosed and treated at health facilities outside the study area. The modifiable areal unit problem might arise due to the spatial unit of data aggregation. However, the spatial unit of analysis used in this study was the finest resolution available, kebele, which was also the spatial unit used for healthcare planning in the study area. The denominator population numbers could be affected by uneven population growth across the study area since the numbers were projected from the 2007 census [33].

Conclusion

This study has underlined that the geostatistical kriging approach can be applied to estimate the spatial risk of tuberculosis distribution in data limited settings. The estimation results may help local public health authorities measure burden of the disease at all locations, identify geographical areas that require more attention, and evaluate the impacts of intervention programs.

Additional files

Additional file 1: The aggregated datasets for the study. (XLSX 68 kb)
Additional file 2: The geostatistical kriging models. (XLSX 232 kb)

Abbreviations

AFB: Acid Fast Bacilli; ASE: Average Standard Errors; CSA: Central Statistical Agency of Ethiopia; DOTS: Directly Observed Treatment-Short courses; DR-TB: Drug Resistant TB; ETB: Extra-pulmonary TB; MSE: Mean Standardized Error; PTB-: Smear-negative pulmonary TB; RMSE: Root-Mean-Square Error; RMSSE: Root-Mean-Square Standardized Error; TB: Tuberculosis; TB/HIV: Human Immunodeficiency Virus (HIV)-associated TB; PTB +: Smear-positive pulmonary TB.; SEBS: Spatial Empirical Bayes Smoothing; SE: Standard Errors

Acknowledgements

Authors would like to thank the Gurage Zone Department of Health, the data collectors and the personnel at the health facilities in the study area.

Funding

Funding for this study was provided by the University of Gondar and the Addis Ababa University, Ethiopia.

Availability of data and materials

The datasets for the current study are included within the article and its additional files.

Authors' contributions

ST conceived the study, collected, analyzed and prepared the draft manuscript. FE, as a primary advisor for ST, advised and supervised the conception, the conduct as well as the analysis and the write-up of the manuscript. SH advised and supervised the conception, the analysis and the write-up of the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The study protocol was reviewed and approved by the Research and Ethical Committee of the School of Public Health, and the Institutional Review Board of the College of Health Sciences, Addis Ababa University. Because of the retrospective nature of the study, informed consent from patients was not required. A letter of support was obtained from the Gurage Zone Department of Health to obtain information from all districts and health facilities. Personal identifiers of the cases were coded prior to analysis, and medical records were kept in a secure place to help maintain the confidentiality of the clinical information of cases.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia. [2]School of Public Health, College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia.

Received: 15 March 2018 Accepted: 14 June 2018
Published online: 25 June 2018

References

1. World Health Organization. Global tuberculosis report. Geneva, Switzerland: World Health Organization; 2017.
2. Teklegiorgis K, Tadesse K, Mirutse G, Terefe W. Level of data quality from health management information systems in a resources limited setting and its associated factors, eastern Ethiopia. South Afr J Info Manag. 2016;17(1):612.

3. Li XX, Wang LX, Zhang H, Jiang SW, Fang Q, Chen JX, Zhou XN. Spatial variations of pulmonary tuberculosis prevalence co-impacted by socio-economic and geographic factors in People's republic of China, 2010. BMC Public Health. 2014;14:257.
4. Federal Ministry of Health. First Ethiopian national population-based tuberculosis prevalence survey. Addis Ababa, Ethiopia: Federal Ministry of Health; 2011.
5. Hamusse SD, Demissie M, Lindtjorn B. Trends in TB case notification over fifteen years: the case notification of 25 districts of Arsi zone of Oromia regional state, Central Ethiopia. BMC Public Health. 2014;14:304.
6. Shargie EB, Yassin MA, Lindtjorn B. Prevalence of smear-positive pulmonary tuberculosis in a rural district of Ethiopia. Int J Tuberc Lung Dis. 2006;10(1):87–92.
7. Goovaerts P. Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson kriging. Int J Health Geogr. 2005;4:31.
8. Ibrahim S, Hamisu I, Lawal U. Spatial pattern of tuberculosis prevalence in Nigeria: a comparative analysis of spatial autocorrelation indices. Am J Geogr Infor System. 2015;4(3):87–94.
9. Piazza AD, Conti FL, Viola F, Eccel E, Noto LV. Comparative analysis of spatial interpolation methods in the Mediterranean area: application to temperature in Sicily. Water. 2015;7:1866–88.
10. Luo W, Taylor MC, Parker SR. A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. Int J Climatol. 2008;28:947–59.
11. Gething P, Atkinson P, Noor A, Gikandi P, Hay S, Nixon M. A local space-time kriging approach applied to a national outpatient malaria dataset. Comput Geosci. 2007;33(10):1337–50.
12. Konak A. A kriging approach to predicting coverage in wireless networks. Int J Mobile Network Design and Innovation. 2009;3:2.
13. Ali M, Goovaerts P, Nazia N, Haq MZ, Yunus M, Emch M. Application of poisson kriging to the mapping of cholera and dysentery incidence in an endemic area of Bangladesh. Int J Health Geogr. 2006;5:45.
14. Sturrock HJW, Gething PW, Clements ACA, Brooker S. Optimal survey designs for targeting chemotherapy against soil-transmitted helminths: effect of spatial heterogeneity and cost-efficiency of sampling. Am J Trop Med Hyg. 2010;82(6):1079–87.
15. Guimaraes RJ, Freitas CC, Dutra LV, Felgueiras CA, Drummond SC, Tibirica SH, Oliveira G, Carvalho OS. Use of indicator kriging to investigate schistosomiasis in Minas Gerais state, Brazil. J Tropical Med. 2012;2012:837428.
16. Southern Nations, Nationalities, and Peoples' Regional State Health Department. The 2016 health report. Awassa, Ethiopia; 2017.
17. Tadesse S, Enqueselassie F, Hagos S. Spatial and space-time clustering of tuberculosis in Gurage zone, southern Ethiopia. PLoS One. 2018;13(6): e0198353.
18. Federal Ministry of Health. Revised national strategic plan: Tuberculosis, TB/HIV, DR-TB, and leprosy prevention and control 2013/14-2020. Addis Ababa, Ethiopia: Federal Ministry of Health; 2017.
19. USEROS. ASTER global digital elevation model version 2. Sioux Falls, SD: Geological Survey Earth Resources Observation Science Center; 2017.
20. Federal Ministry of Health. Guidelines for clinical and programmatic management of tuberculosis, leprosy and TB/HIV in Ethiopia. Addis Ababa, Ethiopia: Federal Ministry of Health; 2012.
21. Dangisso MH, Datiko DG, Lindtjorn B. Spatio-temporal analysis of smear-positive tuberculosis in the Sidama Zone, Southern Ethiopia. PLoS One. 2015;10(6):e0126369.
22. Kim SY, Yi SJ, Eum YS, Choi HJ, Shin H, Ryou HG, Kim H. Ordinary kriging approach to predicting long-term particulate matter concentrations in seven major Korean cities. Environ Health Toxicol 2014;29:e2014012.
23. Yalcin E. Cokriging and its effect on the estimation precision. J South African Inst Mining and Metallurgy. 2005;105:223-28.
24. Aretouyap Z, Nouck PN, Nouayou R, Kemgang FEG, Toko ADP, Asfahani J. Lessening the adverse effect of the semivariogram model selection on an interpolative survey using kriging technique. SpringerPlus 2016;5:549.
25. Li XX, Wang LX, Zhang J, Liu YX, Zhang H, Jiang SW, Chen JX, Zhou XN. Exploration of ecological factors related to the spatial heterogeneity of tuberculosis prevalence in P. R. China. Glob Health Action. 2014; 7:23620.
26. Dangisso MH, Datiko DG, Lindtjorn B. Accessibility to tuberculosis control services and tuberculosis programme performance in southern Ethiopia. Glob Health Action. 2015;8:29443.
27. Regassa N, Yusufe A. Gender Differentials in Migration Impacts in Southern Ethiopia. Anthropologist 2009;11(2):129-37.
28. Carter B, Rohwerder B. Rapid fragility and migration assessment for Ethiopia: Rapid literature review. Birmingham: GSDRC, University of Birmingham; 2016.
29. Kozinska M, Zientek J, Augustynowicz-Kopec E, Zwolska Z, Kozielski J. Transmission of tuberculosis among people living in the border areas of Poland, the Czech Republic, and Slovakia. Pol Arch Med Wewn. 2016;126(1-2):32-40.
30. Hanson K, Jack W. Incentives could induce Ethiopian doctors and nurses to work in rural settings. Health Affairs 2010; 29(8):1452-60.
31. Boru CG, Shimels T, Bilal AI. Factors contributing to non-adherence with treatment among TB patients in Sodo Woreda, Gurage Zone, Southern Ethiopia: A qualitative study. J Infect Public Health 2017; 10:527-33.
32. Sun W, Gong J, Zhou J, Zhao Y, Tan J, Ibrahim AN, Zhou Y. A spatial, social and environmental study of tuberculosis in China using statistical and GIS technology. Int J Environ Res Public Health 2015; 12:1425-48.
33. Central Statistical Agency. The 2007 Population and Housing Census of Ethiopia: Statistical Report for Southern Nations, Nationalities and Peoples' Region. Addis Ababa, Ethiopia: Central Statistical Agency; 2007. Available at: http://www.csa.gov.et/census-report/complete-report/census-2007.html?start=10. Accessed Apr 2017.

Paper III

Ecological factors affecting spatial distribution of tuberculosis in Gurage Zone, Southern Ethiopia

Sebsibe Tadesse[1@], Fikre Enqueselassie[2], Seifu Hagos Gebreyesus[2]

[1]Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia

[2]School of Public Health, College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia

[@] Corresponding Author: sbsbtadesse90@gmail.com

Abstract

Background: The spatial and space-time distribution of tuberculosis is nonrandom and cases are clustered at specific-geographical locations in Ethiopia. However, there is a dearth of studies clarifying the effects of underlying ecological factors on the observed heterogeneities. This study aims to estimate the effects of ecological factors on spatial distribution of tuberculosis prevalence rate in Gurage Zone, Southern Ethiopia.

Methods: The study data were obtained from unit tuberculosis registers at health facilities in Gurage Zone, the rainfall and temperature data from Meteorological Agency of Ethiopia, the population and geo-location data from Central Statistical Agency of Ethiopia, and the Normalized Difference Vegetation Index data from Moderate Resolution Imaging Spectroradiometer imagery. The spatial panel data analysis was used to examine the relationship between ecological factors and tuberculosis prevalence rate. The data included 425 kebeles and the period 2007 to 2016.

Results: The tuberculosis prevalence rate observed in a given kebele was influenced by both tuberculosis prevalence rate and unobserved factors in the neighboring kebeles in the Gurage Zone. By controlling the spatial effects, a 1°C rise in temperature was associated with an increase in the number of tuberculosis prevalence rate by 0.72, and a 1 person per square kilometer increase in population density was related to an increase in the number of tuberculosis prevalence rate by 1.19.

Conclusion: This study highlighted that the designing of locally effective tuberculosis prevention and control strategies should consider the variations in tuberculosis prevalence rate in neighboring areas, temperature and population density.

Keywords: Ecological factors, Spatial autocorrelation, Spatial panel data analysis, Spatial weight matrix, Tuberculosis distribution

Background

Despite the intensified efforts to control Tuberculosis (TB), Ethiopia has remained among the global high-burden countries for the disease. According to the 2016 World Health Organization report, there were an estimated 499 new cases and 71 deaths in the country every day [1]. The disease has placed devastating economic impacts on the patients, families and healthcare system due to direct household expenditures, income and productivity losses, disabilities related to lung damage and resource depletion at health facilities, slowing down the countries desire toward becoming a middle-income country by 2025 [2-4]. Furthermore, the resurgence of co-infection with Human Immunodeficiency Virus and the emergence of drug resistance have more complicated the disease burden in the country [1, 5].

Spatial epidemiological studies reveal that the spatial and temporal distribution of TB is nonrandom and cases are clustered at specific-geographic locations in Ethiopia [6-9]. It is known that the tendency of TB clustering at a given spatial location can be affected by both the spatial and temporal heterogeneity of underlying ecological factors, like climatic, socioeconomic and environmental factors [10-12]. However, most previous studies did not consider both the spatial and temporal effects of these factors on TB distribution in Ethiopia [13-16]. This might result in a loss of information by ignoring the heterogeneity in both time and space. Recently, spatial panel data models are becoming useful tools in the analysis of space-time data of infectious diseases, including TB [17-19]. They are more informative, contain more variation, less collinearity among the variables and more degrees of freedom, and hence improve the efficiency of estimation [18, 19].

In this study spatial panel data modeling techniques were used to estimate the effects of ecological factors on spatial distribution of TB prevalence rate in Gurage Zone, Southern Ethiopia. The study will fill a critical gap in understanding the role of ecological factors on TB

distribution, and contributes for the growing body of spatial epidemiological research on TB in developing countries. Moreover, such studies may also help policymakers in developing evidence-based interventions.

Methods

Study area

This study was carried out in the Gurage Zone in southern Ethiopia, which is located between 7°76' and 8°45' N latitude and 37°46' and 38°71' E longitude **(Fig 1)**. The zone covers a geographic area of about 5,932 km^2. It is divided into 13 districts, with town administrations at Butajira and Wolkitie. Based on the 2007 census, the zone has a total population of 1,279,646 (622,078 males and 657,568 females) in 2007. About 84% of the population live in the rural areas [20].

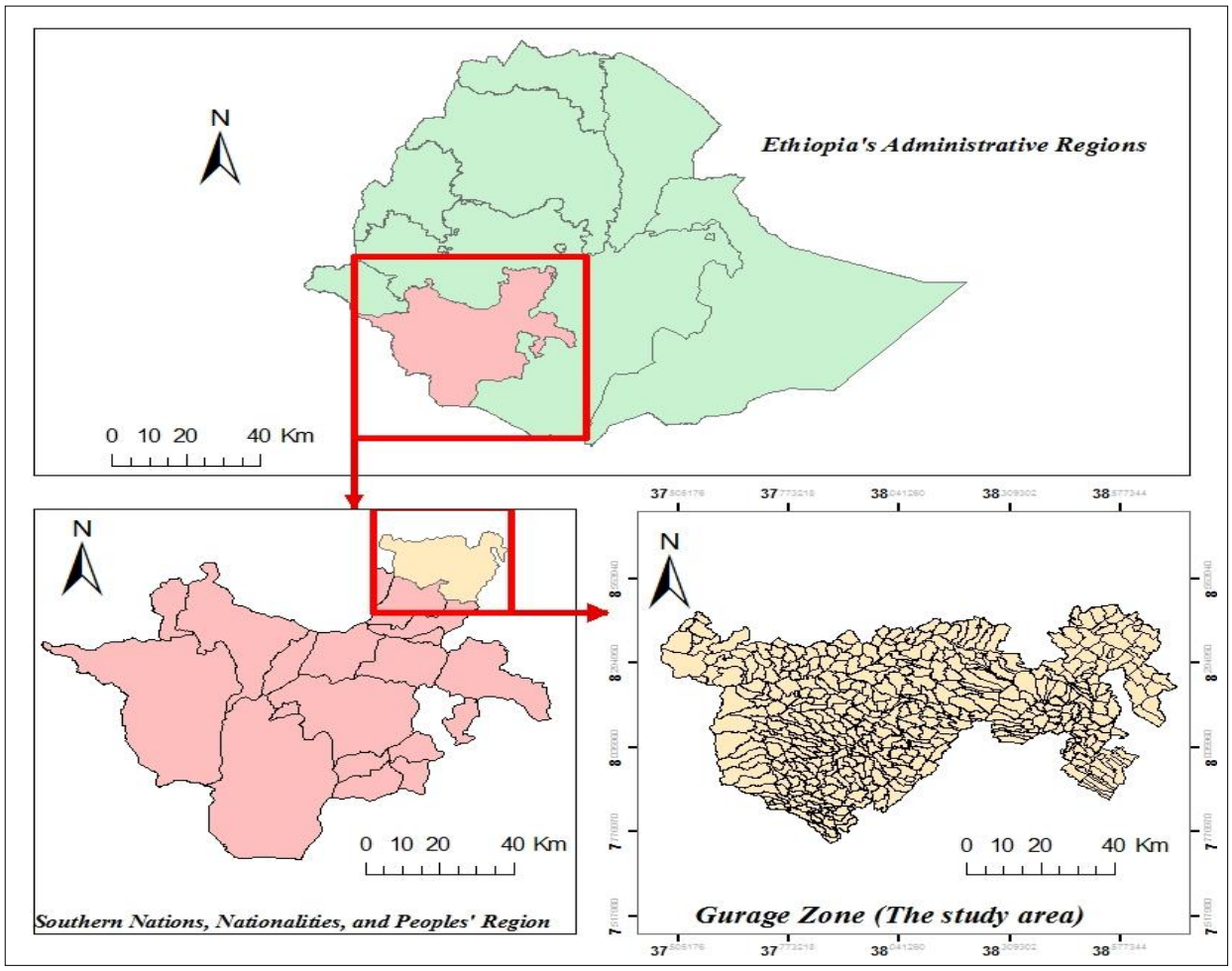


Fig 1 Map of the study area (Gurage Zone). Source: *Tadesse et al., 2018*[7, 8].

There area total of 6 hospitals, 70 health centers, 414 health posts and 92 clinics in the zone that provide TB preventive and control services [20]. The health centers and hospitals provide Directly Observed Treatment-Short courses (DOTS) services, while the health posts and clinics are involved in health education, active case finding, refer patients to health centers and support treatment adherence through female health extension workers. The health centers conduct sputum smear microscopy, treatment and the referral of smear-negative and extra-pulmonary cases to hospitals for further investigation, while hospitals provide diagnostic, treatment and inpatient care services.

According to the national TB diagnosis and treatment guideline of Ethiopia, TB patients were diagnosed using pathogen detection, X-ray, clinical diagnoses and physical examination. Patients who had never had treatment for TB or who had taken anti-TB drugs for less than one month were categorized as newly diagnosed cases, and while treatment failure, relapse and default were categorized as retreatment cases [21].

Study design

An ecological study design was used because TB cases were aggregated at kebele level (the smallest administrative units with an average of 5,000 people) to calculate the prevalence rate of TB. The study explored the relationship between the prevalence rate of TB and ecological factorsat kebele level. The kebele was chosen as a unit of analysis since it was the finest resolution used for health care planning in the study area.

Data collection

The data collection was carried out from June to September, 2017. Thestudy included all TB patients who were residents of the Gurage Zone and registered at DOTS-providing health facilities in the zone during 2007 to 2016. Trained data collectors retrieved the patient information on sex, age, address, TB type, patient category and date of treatment started from the unit TB registers. Incomplete treatment starting dates were replaced by the mid-dates between the start dates of the adjacent registered patients. The Normalized Difference Vegetation Index (NDVI) data reflecting the surface vegetation coverage were derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) imagery [22]. The temperature (in °C) and rainfall (in mm) data were obtained from the Statistical Data Sharing Office of Meteorological Agency of Ethiopia. The kebele-level polygon shape file obtained from the Central Statistical

Agency of Ethiopia was used to compute area of each kebele and spatial weighting matrix. The annual projected population data based on the 2007 census were used for calculations of the population density (the population of a given year divided by the area of a given kebele in square kilometer) and the prevalence rate of TB (the number of TB cases of a given kebele divided by the population of a given year and multiplied by 100,000) [23].

Data quality control

Training on field methods, data extraction and record keeping was given for data collectors and supervisors. The data were double entered and checked page-by-page by year, district, kebele and health facilities against unit TB registers for consistency and completeness.

Data management and processing

The MS Excel (MicroSoft, Redmond, WA, USA) was used for data entry, validation, cleaning and coding. The patients' addresses with similar names but from different kebeles were linked to their actual kebeles to prevent duplication. The data were aggregated at kebele level for spatial analyses. The spatial weighting matrix, which was needed to perform the spatial panel regression analyses, was constructed in Geographic Data analysis tool (GeoDa) by using the kebele-level polygon shape file. A first order queen polygon contiguity weights matrix, which defines the neighbors as those with either a shared border or vertex, was used for spatial weights [6] (**Additional file 1**). The zero TB prevalence ratesin the spatial panel dataset were replaced by one in order to make the data structure strongly balanced (i.e., all kebeles have data for all years). All the study variables were log transformed to improve estimation (**Additional file 2**).

Spatial panel data models

The previous study has shown that there were statistically significant similarities in TB prevalence rates between neighboring kebeles in the Gurage Zone for each year during 2007 to 2016 [7, 8], suggesting that the traditional ordinary least squares regression shouldincorporate the spatial effects in the analysis to improve estimation.Therefore, this study used spatial panel data modeling techniques to estimate the effects of ecological factors (i.e., rainfall, temperature, NDVI, and population density) on spatial distribution of TB prevalence rate.

Four spatial panel data models have been compared to select the best fit one for the data analyses. These are:

1. **Spatial lag or Spatial Autoregressive model (SAR).** This model explains the interaction among prevalence rate of TB as:

   $y_t = \rho W y_t + X_t\beta + \partial + \varepsilon_t$ ..... (1)

   where y_t denotes the prevalence rate of TB, ρ denotes the spatial autoregressive (or lag) coefficient reflecting the severity of spatial interdependence in the distribution of TB prevalence between neighboring kebeles (the normal range is 0 to 1, and a high value indicated strong neighborhood effect), W denotes the spatial weight matrix describing the spatial proximity between kebeles, X_t denotes the matrix of ecological variables, β denotes the regression coefficients, ∂ denotes the kebele-specific fixed effects whose omission could bias the estimates, and ε_t denotes the error terms.

2. **Spatial Error Model (SEM).** This model specifies the interaction among error terms as:

   $y_t = X_t\beta + \partial + \varphi_t$, with $\varphi_t = \lambda W\varphi_t + \varepsilon_t$ ..... (2)

   where φ_t denotes the spatially autocorrelated error term and λ denotes the coefficient of spatial autocorrelation in error terms.

3. **Spatial Durbin Model (SDM).** This model includes spatially lagged values of TB prevalence rates and spatially weighted ecological variables from neighboring kebeles as independent variables. It reads as:

   $y_t = \rho W y_t + X_t\beta + WZ_t\theta + \partial + \varepsilon_t$ ..... (3)

   where WZ_t represents the weighted average effect of the neighboring kebeles on the ecological variables and θ represents the coefficient of spatial dependence between the ecological variables.

4. **Spatial Autocorrelation model (SAC).** This model combines the SAR model with autoregressive errors, and is also known as SARAR model. It is specified as:

   $y_t = \rho W y_t + X_t\beta + \partial + \varphi_t$, with $\varphi_t = \lambda W\varphi_t + \varepsilon_t$ ..... (4)

Specification tests

The Hausman test has revealed that the SDM with spatial fixed-effects specification is appropriate when compared to the SDM with random-effects specification ($chi^2 = 107.95$, $P < 0.001$). Next, the SAR and SEM models are compared against SDM since they are nested in SDM [24]. The SDM is preferred to the SAR ($chi^2 = 93.66$, $P < 0.001$) and the SEM ($chi^2 = 143.84$, $P < 0.001$) models. Finally, since the SDM and the SAC model are non-nesting models, the Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are used to compare them. The SAC model is proved to be the best fit model for explaining the variations in TB prevalence in the different kebeles with AIC = 8585.79 and BIC = 8630.27 against AIC = 8732.05 and BIC = 8795.59 for SDM. The spatial panel data analyses and the specification tests were executed in Stata 14 (Stata Corp, College Station, Texas) using the XSMLE (estimating spatial panel data using Maximum Likelihood Estimator) command [24].

Results

Characteristics of TB cases

From 16,618 TB cases diagnosed at all DOTS-providing health facilitiesin Gurage Zone from 2007 to 2016, 4.9% were excluded from the final analyses due to incomplete addresses or being outside of the zone. Out of the 15,805 cases included in the final analyses 55.3% were males. The majority, 85.8%, of the cases belonged to the working age group. About 93.2% of the cases were newly diagnosed. Pulmonary TB comprised more than two-thirds of all the cases. The majority, 85.3%, of the cases were diagnosed at the health centers in the zone **(Table 1)**. Higher prevalence of TB was observed in urban areas compared to the rural areas during 2007 to 2016 **(Fig 2)**.

Table 1 Demographic and clinical characteristics of TB cases in Gurage Zone, Southern Ethiopia, 2007-2016

Variables	Number	Percent
Sex		
Male	8742	55.3
Female	7063	44.7
Age (in years)		
≤14	1447	9.2
15-34	6888	43.6
35-64	6665	42.2
≥65	805	5.1
DOTS facility		
Health center	13480	85.3
Hospital	2325	14.7
Category of TB		
Newly diagnosed	14736	93.2
Retreatment	1069	6.8
Type of TB		
Pulmonary, smear positive	5197	32.9
Pulmonary, smear negative	6049	38.3
Extra pulmonary	4559	28.8

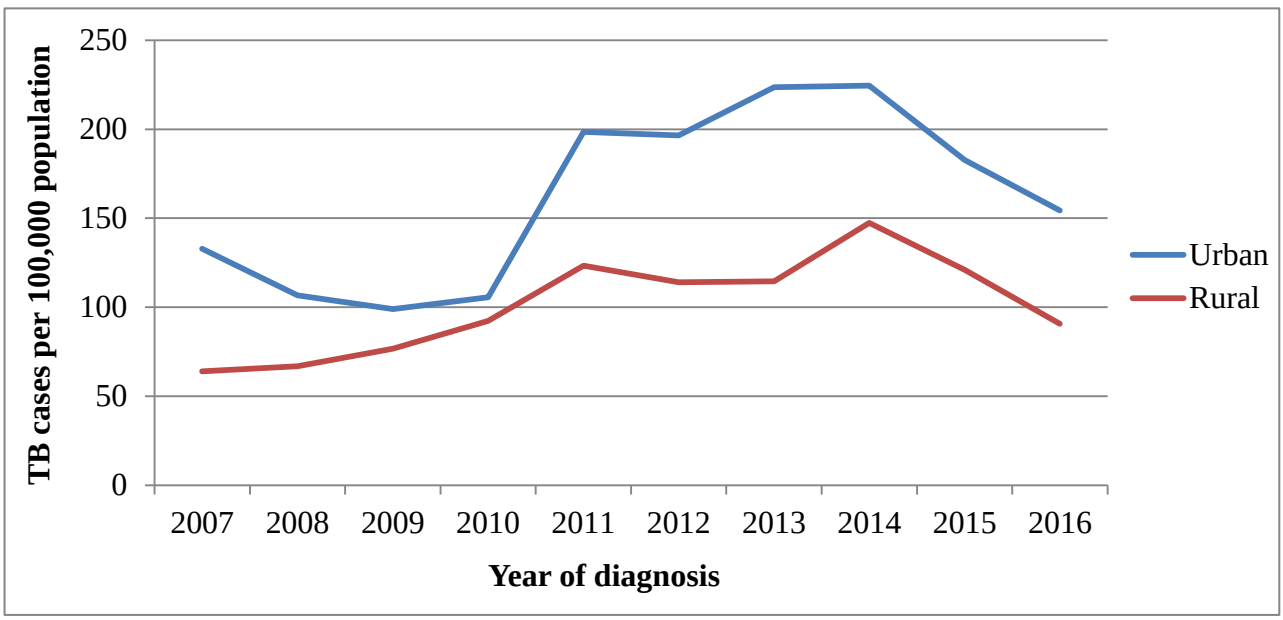


Fig 2 Trend of TB prevalence in urban and rural areas of Gurage Zone, Sothern Ethiopia, 2007-2016

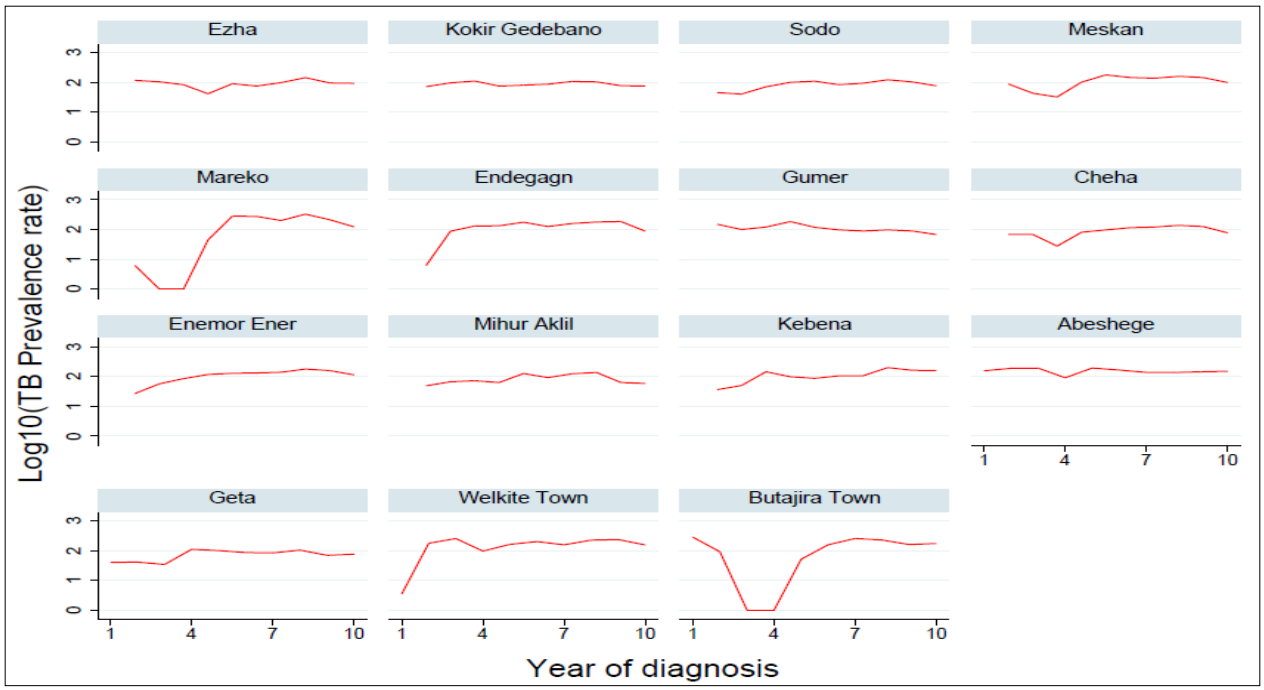
Description of the panel data

The panel data included 425 kebeles that were observed for the period 2007 to 2016, making a total of 4,250 observations. The average rainfall of the study area was 89.9 mm (range: 116 mm). There was higher variability of rainfall within kebeles (range: 89.9 mm) compared to between kebeles (range: 61.3 mm) during the study period. The average temperature of the study area was 18.9°C (range:59.8°C). The variability of temperature was higher within kebeles (range: 52.9°C) compared to between kebeles (range: 11.1°C). The study area has an average NDVI of 5,493.4. A higher variability in NDVI was observed between kebeles (range: 3,484.2) compared to within kebeles (range: 1,385.4) during the study period. The average population density of the study area was 332.1 persons per square kilometer. There was higher variability of population density between kebeles (range: 6,431.9 persons per square kilometer) compared to within kebeles (range: 2,243.7 persons per square kilometer). The average TB prevalence rate of the study area was 111.9 cases per 100,000 population. A higher variability in prevalence rate of TB was observed within kebeles (range: 1,880.7 cases per 100,000 population) compared to between kebeles (range: 886.6 cases per 100,000 population) during the study period (**Table 2**). This variability was lower at district level compared to the variability at kebele level (**Fig 3**).

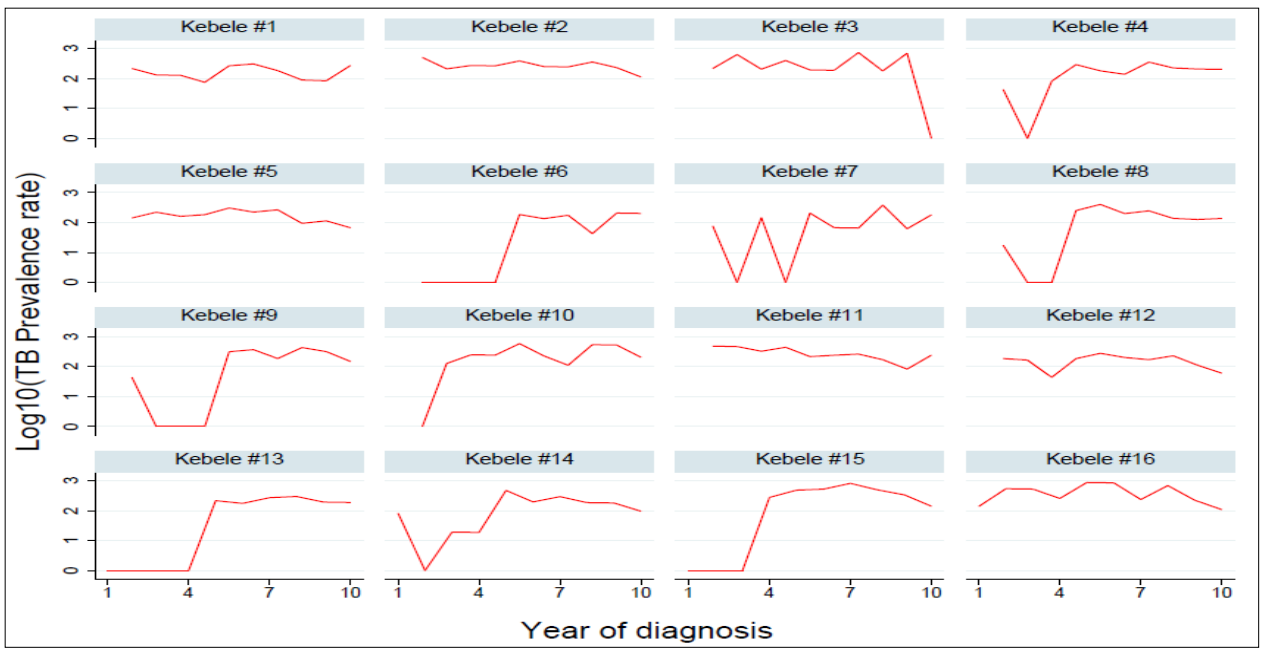
Table 2 Summary of panel data for the study in Gurage Zone, Southern Ethiopia, 2007-2016

Variable	Variation	Mean	Std. Dev.	Min	Max	Observation*
rainfall	Overall	89.9	21.5	31.9	147.9	N = 4250
	Between		14.7	59.7	121.0	n = 425
	Within		15.7	36.5	126.4	T = 10
Temperature	Overall	18.9	2.1	9.2	69.0	N = 4250
	Between		1.8	14.5	25.6	n = 425
	Within		1.0	11.8	64.7	T = 10
NDVI	Overall	5493.4	806.0	2718.8	7017.8	N = 4250
	Between		772.8	3159.6	6643.8	n = 425
	Within		231.9	4813.8	6199.2	T = 10
Population density	Overall	332.1	444.2	31.5	7300.6	N = 4250
	Between		441.6	35.8	6467.7	n = 42
	Within		52.1	-441.6	1802.1	T = 10
TB prevalence rate	Overall	111.9	132.0	1.0	1820.0	N = 4250
	Between		82.5	1.0	887.6	n = 425
	Within		103.2	-774.6	1106.1	T = 10

*****N = total number of observations, n = the number of kebeles, T = observation per kebele



A) Trend of $\log_{10}$(TB prevalence rate) at district level



B) Trend of $\log_{10}$(TB prevalence rate) at kebele level

Fig 3 Variation of TB prevalence rateat A) districts and B) sample kebeles in Gurage Zone, Southern Ethiopia, 2007-2016. Source: *Authors own data*

Ecological factors affecting the spatial distribution of TB

The fixed-effects SAC model revealed that both the spatial autoregressive coefficient (ρ) in the spatial lag of TB prevalence rate and the spatial autocorrelation coefficient (λ) in the spatial error terms were statistically significant indicating that the TB prevalence rate observed in a given kebele was affected by both TB prevalence rate and unobserved factors in the neighboring kebeles. By controlling the spatial effects, a 1°C rise in temperature was associated with an increase in the number of TB prevalence rate by 0.72. By controlling the spatial effects, a 1 person per square kilometer increase in population density was related to an increase in the number of TB prevalence rate by 1.19 (**Table 3**).

Table 3 Results of fixed-effects SAC model for the spatial association between ecological factors and TB prevalence rate in Gurage Zone, Southern Ethiopia, 2007-2016

Variable	Coefficient	Std. Err.	Z	P>|Z|	[95% Conf.	Interval]
Rainfall	0.07	0.08	0.88	0.378	-0.080	0.212
Temperature	0.72	0.32	2.28	0.022	0.102	1.345
NDVI	0.23	0.34	0.69	0.491	-0.432	0.901
Population density	1.19	0.17	6.93	<0.001	0.855	1.529
ρ	0.83	0.01	66.52	<0.001	0.800	0.849
λ	-0.70	0.03	-20.62	<0.001	-0.769	-0.636

Discussion

This study has applied the spatial panel data modeling techniques to estimate the effects of ecological factors on spatial distribution of TB prevalence rate in Gurage Zone, Southern Ethiopia using panel data from 425 kebeles and the period 2007 to 2016. A spatial panel data model is more appropriate when longitudinal data from multiple spatial units exhibit spatial autocorrelation [18]. The model corrects for the deviations that are caused by the spatial interaction between kebeles and unobserved spatial heterogeneity by introducing a spatial weight matrix and a spatial heterogeneity into statistical modeling [19, 24].

This study revealed that the TB prevalence rate observed in a given kebele was affected by both TB prevalence rate and unobserved factors in the surrounding kebeles, likely suggesting that there has been sustained transmission of TB within the communities. This finding is similar to previous studies [6-8, 25]. Therefore, it can be recommended that neighboring kebeles should cooperate in the design and implementation of TB prevention and control strategies.

The spatial panel data analysis has revealed that there was a positive association between temperature and TB prevalence rate during the study period. Previous studies support this finding [26]. The relationship could be explained by the fact that higher temperatures promote the survival and replication of *Mycobacterium tuberculosis* and improve its activity [27]. This implies that the burden of TB will be higher in the coming decades since there is a rapid increment in global temperature. Thus, TB prevention and control strategies should take temperature variations into consideration.

This study has also elucidated that there was a positive association between population density and TB prevalence rate in the study area. The reason for this could be that higher population density increases the risk of personal contact and facilitates the transmission of TB [6]. Therefore, policymakers and public health authorities should consider spatial locations with higher population density for TB prevention and control interventions.

There are some limitations in this study. The prevalence rate of TB might be underestimated in some kebeles because the study did not include those patients who were diagnosed and treated at health facilities outside the Gurage Zone, and those who would remain undiagnosed for the disease due to different reasons. The uneven population growth across kebeles could affect the denominator population numbers because they were projected from the 2007 census [23]. Although spatial panel data modeling technique has the advantage of controlling for spatial dependency and unknown spatial heterogeneity, it does not show the local spatial effects of individual covariates. Hence, the geographically weighted regression modeling would be recommended to explore the local spatial effects of covariates in the future [18]. Another limitation is that other ecological factors, such as age dependency ratio, level of school enrollment, income, wind speed, humidity, air quality and hours of sunshine were not evaluated in this study [27-29].

Conclusion

This study has demonstrated that the designing of locally effective TB prevention and control strategies should consider the variations in TB prevalence rate in neighboring areas, temperature and population density.

Declarations

Abbreviations

AIC: Akaike's Information Criterion; BIC: Bayesian Information Criterion; DOTS: Directly Observed Treatment-Short courses; GeoDa: Geographic Data analysis tool; MODIS: Moderate Resolution Imaging Spectroradiometer; NDVI: Normalized Difference Vegetation Index; SAC: Spatial Autocorrelation model; SAR: Spatial Autoregressive model; SARAR: Spatial Autoregressive model with Spatial Autocorrelation in errors; SDM: Spatial Durbin Model; SEM: Spatial Error Model; TB: Tuberculosis; XSMLE: Estimating spatial panel data using Maximum Likelihood Estimator

Acknowledgement

Authors would like to thank the Gurage Zone Health Department, the data collectors and the personnel at the health facilities in the study area.

Funding

Funding for this study was obtained from the Addis Ababa University and the University of Gondar, Ethiopia.

Availability of data and materials

The datasets for the study are included within the article and its additional files.

Authors' contributions

ST conceived the study, collected, analyzed and prepared the draft manuscript. FE, as a primary advisor for ST, advised and supervised the conception, the conduct as well as the analysis and the write-up of the manuscript. SH advised and supervised the conception, the analysis and the write-up of the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The study protocol was reviewed and approved by the Research and Ethical Committee of the School of Public Health, and the Institutional Review Board of the College of Health Sciences, Addis Ababa University. Informed consent from patients was not required since the study used retrospective data. A letter of support was obtained from the Gurage Zone Health Department to obtain information from all districts and health facilities. Personal identifiers of the cases were coded prior to analysis, and medical records were kept in a secure place to help maintain the confidentiality of the clinical information of cases.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interest.

Authors' information

[1]Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia

[2]School of Public Health, College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia

References

1. World Health Organization. Global tuberculosis report. Geneva, Switzerland: World Health Organization; 2017.
2. Federal Ministry of Health. First Ethiopian national population-based tuberculosis prevalence survey. Addis Ababa, Ethiopia: Ministry of Health; 2011.
3. Barter DM, Agboola SO, Murray MB, Barnighausen T. Tuberculosis and poverty: the contribution of patient costs in Subsaharan Africa: a systematic review. BMC Public Health 2012; 12:980.
4. Tadesse S. Stigma against tuberculosis patients in Addis Ababa, Ethiopia. PLoS One 2016; 11(4):e0152900.
5. Tadesse S, Tadesse T. HIV co-infection among tuberculosis patients in Dabat, northwest Ethiopia. J Infect Dis Immun 2013; 5(3):29-32.
6. Dangisso MH, Datiko DG, Lindtjorn B. Spatio-temporal analysis of smear-positive tuberculosis in the Sidama Zone, Southern Ethiopia. PLoS One 2015; 10(6):e0126369.
7. Tadesse S, Enqueselassie F, Gebreyesus SH. Estimating the spatial risk of tuberculosis distribution in Gurage zone, southern Ethiopia: a geostatistical kriging approach. BMC Public Health 2018; 18:783.
8. Tadesse S, Enqueselassie F, Hagos S. Spatial and space-time clustering of tuberculosis in Gurage zone, southern Ethiopia. PLoS One 2018; 13(6):e0198353.
9. Tadesse T, Demissie M, Berhane Y, Kebede Y, Abebe M. The clustering of smear-positive tuberculosis in Dabat, Ethiopia: a population based cross sectional study. PloS One 2013; 8(5):e65022.
10. Li XX, Wang LX, Zhang H, Jiang SW, Fang Q, Chen JX, et al. Spatial variations of pulmonary tuberculosis prevalence co-impacted by socio-economic and geographic factors in People's Republic of China, 2010. BMC Public Health 2014; 14:257.

11. Li XX, Wang LX, Zhang J, Liu YX, Zhang H, Jiang SW, et al. Exploration of ecological factors related to the spatial heterogeneity of tuberculosis prevalence in P. R. China. Glob Health Action 2014; 7:23620.
12. Maciel EL, Pan W, Dietze R, Peres RL, Vinhas SA, Ribeiro FK, et al. Spatial patterns of pulmonary tuberculosis incidence and their relationship to socio-economic status in Vitoria, Brazil. The Int J Tuberc Lung Dis: the official J Int Union against Tuberc Lung Dis 2010; 14 (11):1395–402.
13. Tadesse T, Demissie M, Berhane Y, Kebede Y, Abebe M. Two-thirds of smear-positive tuberculosis cases in the community were undiagnosed in northwest Ethiopia: population based cross-sectional study. PLoS One 2011; 6(12):e28258.
14. Shargie EB, Yassin MA, Lindtjorn B. Prevalence of smear-positive pulmonary tuberculosis in a rural district of Ethiopia. Int J Tuberc Lung Dis 2006; 10(1):87-92.
15. Demissie M, Zenebere B, Berhane Y, Lindtjorn B. A rapid survey to determine the prevalence of smear-positive tuberculosis in Addis Ababa. Int J Tuberc Lung Dis 2002; 6(7):580-584.
16. Getahun B, Ameni G, Biadgilign S, Medhin G. Mortality and associated risk factors in a cohort of tuberculosis patients treated under DOTS program in Addis Ababa, Ethiopia. BMC Infect Dis 2011; 11:127.
17. Xiao G, Xu C, Wang J, Dongyang Yang D, Wang L. Spatial-temporal pattern and risk factor analysis of bacillary dysentery in the Beijing-Tianjin-Tangshan urban region of China. BMC Public Health 2014; 14:998.
18. Wang H, Du Z, Wang X, Liu Y, Yuan Z, Liu Y, et al. Detecting the association between meteorological factors and hand, foot, and mouth disease using spatial panel data models. Int J Infect Dis 2015; 34:66-70.
19. Rao HX, Zhang X, Zhao L, Yu J, Ren W, Zhang XL, et al. Spatial transmission and meteorological determinants of tuberculosis incidence in Qinghai Province, China: a spatial clustering panel analysis. Infect Dis Poverty 2016; 5:45.
20. Southern Nations, Nationalities, and Peoples' Regional State Health Department. The 2016 health report. Awassa, Ethiopia; 2017.
21. Federal Ministry of Health. Guidelines for clinical and programmatic management of TB, leprosy and TB/HIV in Ethiopia. Addis Ababa, Ethiopia: Ministry of Health; 2012.

22. USGS Land Processes Distributed Active Archive Center (LP DACC). Accessed on: 2017; Available from: https://lpdaac.usgs.gov/products/modis_products_table.
23. Central Statistical Agency. The 2007 population and housing census of Ethiopia: statistical report for Southern Nations, Nationalities and Peoples' Region. Addis Ababa, Ethiopia: Central Statistical Agency; 2007. Accessed on: 2017; Available from: http://www.csa.gov.et/census-report/complete-report/census-2007.html?start=10.
24. Belotti F, Hughes G, Mortari AP. Spatial panel data models using Stata. The Stata J 2017; 17(1):139-180.
25. Shaweno D, Shaweno T, Trauer JM, Denholm JT, McBryde ES. Heterogeneity of distribution of tuberculosis in Sheka Zone, Ethiopia: drivers and temporal trends. Int J Tuberc Lung Dis 2017; 21(1):79-85.
26. Khaliq A, Batool SA, Chaudhry MN. Seasonality and trend analysis of tuberculosis in Lahore, Pakistan from 2006 to 2013. J Epidemiol Glob Health 2015; 5:397-403.
27. Sun W, Gong J, Zhou J, Zhao Y, Tan J, Ibrahim AN, et al. A spatial, social and environmental study of tuberculosis in China using statistical and GIS technology. Int J Environ Res Public Health 2015; 12:1425-1448.
28. Cao K, Yang K, Wang C, Guo J, Tao L, Liu Q, et al. Spatial-temporal epidemiology of tuberculosis in Mainland China: An analysis based on Bayesian Theory. Int J Environ Res Public Health 2016; 13:469.
29. Koh GCKW, Hawthorne G, Turner AM, Kunst H, Dedicoat M. Tuberculosis incidence correlates with sunshine: an ecological 28-year time series study. PLoS One 2013; 8(3): e57752.

Additional files

Additional file 1.The spatial weighting matrix. (dta)

Additional file 2.The panel data for the study. (dta)

Annex II: Data collection format

The following data collection format was used to retrieve patient information from the Unit TB Registers at DOTS-providing health facilities in Gurage Zone, Southern Ethiopia.

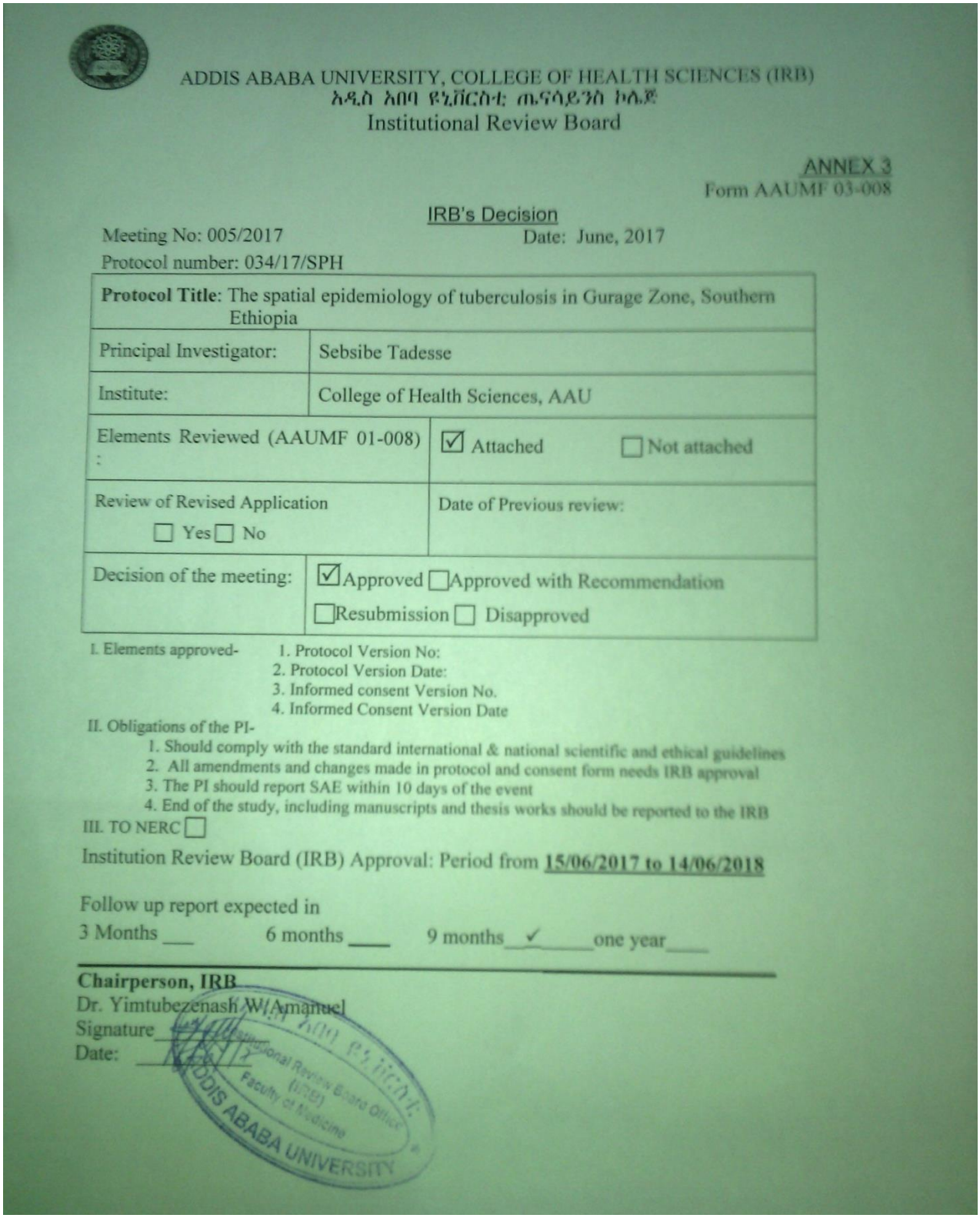
Name of data collector: _____

Date of data collection: _____

Code number: _____

Unit TB No.	Name of the patient	Address of the patient			Sex (M/F)	Age	TB classification (P/Pos ,P/neg, EP)	Treatment started (DD/MM/YY)
		District	Kebele	Phone No.				

Annex III: Ethical approval



ADDIS ABABA UNIVERSITY, COLLEGE OF HEALTH SCIENCES (IRB)
አዲስ አበባ ዩኒቨርስቲ፡ ጤናሳይንስ ኮሌጅ
Institutional Review Board

ANNEX 3
Form AAUMF 03-008

IRB's Decision

Meeting No: 005/2017 Date: June, 2017

Protocol number: 034/17/SPH

Protocol Title: The spatial epidemiology of tuberculosis in Gurage Zone, Southern Ethiopia		
Principal Investigator:	Sebsibe Tadesse	
Institute:	College of Health Sciences, AAU	
Elements Reviewed (AAUMF 01-008) :	☑ Attached	☐ Not attached
Review of Revised Application ☐ Yes ☐ No	Date of Previous review:	
Decision of the meeting:	☑ Approved ☐ Approved with Recommendation ☐ Resubmission ☐ Disapproved	

I. Elements approved-
1. Protocol Version No:
2. Protocol Version Date:
3. Informed consent Version No.
4. Informed Consent Version Date

II. Obligations of the PI-
1. Should comply with the standard international & national scientific and ethical guidelines
2. All amendments and changes made in protocol and consent form needs IRB approval
3. The PI should report SAE within 10 days of the event
4. End of the study, including manuscripts and thesis works should be reported to the IRB

III. TO NERC ☐

Institution Review Board (IRB) Approval: Period from **15/06/2017 to 14/06/2018**

Follow up report expected in
3 Months ___ 6 months ___ 9 months ✓ one year ___

Chairperson, IRB
Dr. Yimtubezenash W/Amanuel
Signature
Date:

Institutional Review Board Office (IRB) Faculty of Medicine ADDIS ABABA UNIVERSITY

Declaration

I, the under signed, declared that this is my original work, has never been presented in this or any other University, and that all the resources and materials used for the dissertation have been fully acknowledged.

Name: **Sebsibe Tadesse Onke**
Signature: _____
Place: **School of Public Health, College of Health Sciences, Addis Ababa University**
Date of submission: _____

This dissertation has been submitted for examination with my approval as University Supervisor.

Name: **Prof. Fikre Enqueselassie Gashe (PhD)**
Signature: _____
Date:_____