Modelling the impacts of droughts on tuberculosis over East African Community (EAC)

By

Enock Mwizerwa (enock.mwizerwa@aims.ac.rw)
African Institute for Mathematical Sciences (AIMS), Rwanda
Supervised by: Dr. Oladapo Michael Olagbegi
South African National Bioinformatics Institute, University of the Western Cape, Cape Town, South Africa

June 2022

AN ESSAY PRESENTED TO AIMS RWANDA IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE IN MATHEMATICAL SCIENCES

DECLARATION

This work was carried out at AIMS Rwanda in partial fulfilment of the requirements for a Master of Science Degree.

I hereby declare that except where due acknowledgement is made, this work has never been presented wholly or in part for the award of a degree at AIMS Rwanda or any other University.

Student: Enock Mwizerwa, June 2022

Supervisor: Dr. Oladapo Michael Olagbegi, June 2022

ACKNOWLEDGEMENTS

I am delighted to acknowledge my supervisor Dr. Oladapo Michael Olagbegi from the University of Western Cape, for his endless commitment, encouragement, inspiration, guidance, and supervision of this research. May Almighty bless him and keep guiding his steps. My appreciation also goes to the staff of AIMS Rwanda especially, my tutor Mr. Joël Pascal Ntwali N'konzi, for the support, guidance, and advice he gave me during the essay phase. May God compensates him. Furthermore, I am thankful to my colleagues for the spirit of cooperation we were having. Above all, I am grateful to the Almighty for the protection throughout this project.

DEDICATION

With great pleasure,
I dedicate this work

To all AIMS staff for their support
To my family and friends

Thanks!!!!!!!

Abstract

Climate change is anticipated to raise global temperatures by more than 1.50 degrees Celsius by the end of the century. Extreme weather conditions such as droughts are expected to become more common as the global average temperature rises, creating a greater risk to human health. While previous studies have shown that precipitation and relative humidity are associated with tuberculosis (TB), impact of droughts on TB in Africa is sparsely studied. This study therefore examined the associations of drought indices with number of TB cases and incidence in East African Community (EAC).

Standardized Precipitation Index (SPI) and Standard Precipitation and Evapotranspiration Index (SPEI) were calculated from climate data using SPI and SPEI packages in R. Categories of drought indices were categorized based on SPI and SPEI categories in literature; TB incidence and number of TB cases were separately compared across categories of drought based on SPI and SPEI. The study further investigated the associations of TB incidence and number of TB cases with drought indices (SPI and SPEI), precipitation and potential evapotranspiration (PET) in EAC using a panel data modelling.

The data contains the 6 countries in EAC for a period of 19 years starting from 2000 to 2018. Panel models were tested and the random effect model was selected as the best model that fits the relationship. Kruskal Wallis test indicated no significant difference ($p > 0.05$) in TB incidence and number of TB cases across categories of drought based on SPI and SPEI. TB incidence was positively associated with PET ($\beta = 1.971$, p $= 0.010$) but no significant association was found between number of TB cases and drought indices. The study's outcome suggests that TB incidence increased with increased PET. This further highlights the importance of monitoring extreme weather events with the view of mitigating their effects on TB associated mortality and morbidity.

Keywords: Drought indices, Panel Data, Random Effects Model.

List of Tables

List of Abbreviations

AIDS	Acquired Immunodeficiency Syndrom
DRC	Democratic Republic of Congo
EAC	East African Community
FEM	Fixed Effects Models
HIV	Human Immunodeficiency Virus
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
REM	Random Effects Model
SPI	Standardized Precipitation Index
SPEI	Standardized Potential Evapotranspiration Index
TB	Tuberculosis
WHO	World Health Organization.
MDR-TB	Multi-Drug resistance Tuberculosis.
IPCC	Intergovernmental Panel on temperature change
MTB	Mycrobaterium tuberculosis.
NTLP	National Tuberculosis and Leprosy Programme.

Contents

1. Introduction

1.1 Background of Study

Tuberculosis (TB) is a transmissible disease that is caused by mycobacterium tuberculosis (MTB) bacteria that affects the lungs in most cases. This micro-organism spreads when a susceptible individual comes in contact with respiratory droplets generated through coughing, sneezing, or speaking by a contagious person (WebMD, 2020). In 2020 TB was the 13^{th} leading cause of death worldwide and the second leading infectious killer after COVID-19 (above HIV/AIDS) (World Health Organization(WHO), 2021).

TB kills numerous people around the world with 1.5 million deaths in 2020, among which 214000 also had HIV. In the same year, the prevalence of TB was estimated to 10 million people worldwide, with 5.6 million men, 3.3 million women, and 1.1 million children falling sick TB globally. In 2020 still, 30 countries accounted for 86% of new TB cases. Eight of them regard for two-thirds of the overall, with India is leading the count, followed by China, Indonesia, the Philippines, Pakistan, Nigeria, and Republic of South Africa (WHO, 2019).

Africa, which has 11% of the world's population, carries 29% of the world the burden of TB cases and 34% of associated deaths. The challenges of eradicating the disease have never been greater (E.Chaisson et al., 2008).

Drought is defined as a prolonged lack of precipitation.The rainfall deficiency lead to a decrease in soil moisture and groundwater, decreased stream flow, crop damage, and a general water shortage (Tsakiris et al., 2007).

Drought can last for months or times, or could also be declared once in or as many as fifteen days (MOHAMMED, 2017). Drought has considerable impacts on the affected region's health as well as the native frugality. Annual dry seasons within the tropics vastly increase the chances of a drought developing. Drought conditions will be greatly exacerbated by warm epochs because of rapid evaporation of water vapor (Kidder and Worsley, 2004).

Drought could be a continual point of the climate in most parts of the earth. Still, these regular droughts become more extreme and more unpredictable because of temperature changes (Oceanic and Administration, 2002).

Droughts occur in most all countries. However, the frequency, severity, and length vary from country to country and from region to region. Over-exploitation of natural resources, weather variability, and climate change mostly account for drought. IPCC (2015) reported that extreme weather and climate events such as droughts and floods have significant impacts on human health in Africa .

The dust, dry conditions, and backfires that always accompany drought will always have an adverse effect on human health. Pollen, bank and fluorocarbons as well as fire, dry soil and foliage increase the quantity of particulates floating in the air (Bryan et al., 2020). These substances irritate the cartilaginous tube passages and lungs, increasing the risk of acute metastatic infections like a tuberculosis and microorganism pneumonia (Centers for Disease Control and Prevention, 2010). An increase in TB cases is an immediate consequence of drought.

Essential climate factors associated with drought such as temperature, precipitation, and humid-

ity have been shown to impact the spread of mycobacterium tuberculosis with TB prevalence increasing as temperature rises (Gelaw et al., 2019). Air quality is fully affected by atmospheric pollution. Carbon monoxide induced bacillary reactivation will increase the occurrence of TB. Furthermore, large seasonal confines of TB generally occur in highland regions with temperate mountain climates and low seasonal average temperature in summer, which increases the likelihood of a high TB incidence in drought season as a result of hot temperature and pollution (CDC Data, 2020).

The East African Community (EAC) was created in 2001 as a trade union between the countries of Burundi, Kenya, Rwanda, United Republic of Tanzania with South Sudan, Democratic Republic of Congo (DRC) and Uganda. The mandate of the EAC is to form profitable relationship and political security across the region. The cooperation between the countries facilitates trading, trip, and migration, making the eventuality for the spread of MDR-TB between the nations. Presently, Tanzania, Kenya, and Uganda rank among high TB burden countries, with reportable TB prevalence rates of 292, 253, and 200 per a 100000 people. (Elsevier, 2021).

1.2 Problem Statement

Tuberculosis is one of the infectious diseases that poses high burden to African countries especially in the East African Region with Kenya, Uganda and Tanzania are being among the top ten countries with high TB cases worldwide (WHO, 2021). Even African countries are developing several TB control strategies in collaboration with the WHO, including the provision of global leadership on TB-related issues, It is important to sustain the development and monitoring of evidence-based policies, strategies, and standards for tuberculosis prevention and control

Consequently, it is pertinent to investigate the associations of extreme climate change events such as droughts with TB incidence within EAC since climate change or variability is known to affect the occurrence and spread of many diseases. Studies examining the effects of extreme climate conditions such as drought on the incidence of TB are scarce. Notably, African research on the relationship of climatic variables with incidence/prevalence of TB are very few. Therefore, a thorough understanding of drought's health consequences could provide evidence for early warning systems and drought adaptation for implementation by different governments and health authorities.

1.3 Research Objectives

The present study aims to investigate the impact of drought on TB incidence in the East African Community. More specifically, this study aims to:

1. Develop a statistical model for the association between TB incidence, cases and droughts indices (Standardized Precipitation index, Standardized Potential Evapotranspiration Index), precipitation and potential evapotranspiration using panel data modeling.

2. Compare the incidence of TB during drought and wet days in the EAC based on Standardized Precipitation index (SPI), Standardized Potential Evapotranspiration Index (SPEI)

classifications to understand the effects of drought on TB incidence in EAC

2. Literature Review

2.1 Drought Characteristics

According to the World Health Organization (2022), drought is a prolonged period of shortage of precipitation in the natural climate cycle that can occur anywhere in the world. It is a slow-onset disaster characterized by a lack of precipitation, resulting in a water shortage. Drought can have serious impact on health, agriculture, economies, energy and the environment.

Drought indices are used to monitor the drought conditions of a region. Various drought indices have been introduced in the past few decades, but some of them are region specific and have limitations of applicability in other climatic conditions (Jain et al., 2015), Mishra (2010) reported that many drought indices have been derived in recent decades. Commonly, a drought index could be a prime variable for assessing the impact of drought and shaping completely different drought characteristics, that embodies intensity, duration, severity, and spatial extent. The most commonly used period, the drought analysis could be a year, followed by a month. Although the yearly continuance is long, it also can be accustomed to abstract data on the regional behavior of droughts (Panu and Sharma, 2002).

There are 3 main drought indices. The palm drought index (PDSI) estimates relative dryness using readily available temperature and precipitation data (Dracup et al., 1980).Standardized Precipitation Evapotranspiration Index (SPEI) is an extension of the popular Standardized Precipitation Index (SPI). The SPEI is intended to determine drought by taking into account both precipitation and potential evapotranspiration (Vicente-Serrano et al., 2010). The Standardized Precipitation index (SPI) is the most used index to characterize drought on a variety of timescales. On short timescales, the SPI is closely associated with soil wet, whereas at longer timescales, the SPI may be associated with groundwater and reservoir storage. The precipitation utilized in SPI may be accustomed to calculating this percentage of average precipitation for a period of i months (B.McKee et al., 1993).

Palmer (1965) said that, drought event is defined as a period during which the SPI is continuously negative and hence reaches a value of -1.0 or less. The drought begins once the SPI first falls below zero and ends with the positive value of SPI. Drought intensity is usually determined based on the value of SPI of the SPI. 2.1 outlines the commonly used drought categories based on SPI/SPEI with the following categories:

Table 2.1: Drought Categories

SPI/SPEI Value	Drought Category
2.00+	Extremely wet
1.5 to 1.99	Severely wet
1 to 1.499	moderately wet
0.99 to -0.99	Near normal
-1.00 to -1.49	Moderate drought
-1.5 to -1.99	Severely drought
≤ -2.00	Extremely drought

The same as SPI, SPEI values that are below zero considered a result of the drought, and positive values of SPEI indicate that a no-drought event, a drought event is taken into consideration a quantity with negative SPEI values (Steinemann, 2003). The drought Duration (D) is that the number length throughout the SPEI is continuous negative starting from the SPEI values are adequate at -1 and end once the SPEI values persuade be positive. The drought severity (S) is that the cumulated SPEI values among the drought duration, that's printed by

$$S = -\sum_{i=1}^{D} SPEI_i$$

and intensity of drought is the quantitative relation of its severity to its amount. Many drought planners appreciate the pliability of SPI as a locality of the observance and early warning efforts, as a result of the SPEI calculation, supported the calculation of the SPI then the classification level of dryness follows the SPI classification projected by (B.McKee et al., 1993).

2.2 Tuberculosis Incidence

The report from WHO (2021) shows that TB is a disease typically affects the lungs but it can also affect totally different sites. It mostly attacks the adult people where 90% of people sick from it are adults with tons of cases among males than females.

The African region accounted for 24% of world cases and 32% of deaths. It fenced countries with a variety of the severest TB epidemics around the globe, notably those with a high prevalence of HIV among the overall population (WHO, 2020).

The Estimates of TB incidence in 2020 suggests that a total of approximately 10 million people were sick TB in 2020, with the best estimate being 9.9 million, or 127 cases per 100 000 population. Both figures' annual rates of decline slowed compared to the previous year (1.9 percent for the incidence rate and 0.87 percent for the absolute number of cases between 2019 and 2020), compared with 2.3 percent and 1.2 percent respectively between 2018 and 2019) (WHO, 2021). From survey drained African countries in 2012 by Migambi et al. (2012) was conducted in 73 clusters whereby all participants were aged fifteen years and better than were screened for TB by symptoms and chest X-ray(CXR) the 43779 people were listed throughout this survey, the result found from this survey shows that the TB incidence rate was 119.3 per 100000 population in Rwanda.

National Tuberculosis and Leprosy Programme (NTLP) (2020) reported that the TB incidence in the United Republic of Tanzania has fallen from 306 to 222 per 100000 population in 2020. The most TB incidence in this country is recorded in Dar es Salaam town whereby it's contributed to 15% of all TB incidence rates in the United Republic of Tanzania.

The Republic of Uganda was among thirty selected countries with a high burden of TB in 2019 with TB incidence rate of 200 per 100000 population (Arsenault et al., 2019).

Kenya conducted the last national TB prevalence survey in 1958 and has so relied on estimates from the UN agency to extrapolate TB incidence and case detection rates. At that point, the prevalence of TB was 3,142 per 100000 (Roelsgaard.E and Nyboe.J, 2016).

According to the Nationwide cross-section survey done in 2016 in Kenya where all participants were supposed to be tested for TB symptoms and Chest X-rays, the results found in this survey

show that TB prevalence in this country was 558 per 100000 adult population. In 2015 Kenya's TB incidence has significantly increased from 50 per 100000 population in 1990 to 233 per 100000 in 2015.

TB remains a major public health concern, with Kenya having a high burden of TB and TB/HIV. He stated that in 2020, the country recorded 72,943 TB cases, of which 8 percent (5,663) were children, and that the 2016 prevalence survey revealed that the country nearly missed 40 percent of the estimated cases, emphasizing the importance of involving everyone in the fight against TB. (Kenya, 2022).

The World Health Organization (WHO), world Fund alongside partners, and together with Makerere University supported the Ministry of Health National infectious disease and Hansen's disease Program (MOH/NTLP) to conduct the National infectious disease (TB) prevalence survey. The goal of the survey was to alert the National Tuberculosis and Leprosy Programme (NTLP) to realize a much better understanding of the burden of TB and thus on determine ways in which TB management within the country, the prevalence of TB in youth is 36 cases per 100000 population. HIV among patients with the infectious disease within the community was 27 which is less than HIV among TB patients seen at the health facilities. The key to notice, about 41,000 patients with TB aren't detected annually (Makerere-University and WHO., 2016).

Between 2000 and 2016, the prevalence of tuberculosis in Burundi was more than halved (290 per 100000 to 118). However, case detection is only about 60%. The calculable burden of multidrug resistant TB (MDR-TB) is 3.2% among new TB cases and Bastille Day among re-treatment cases. The prevalence of HIV amongst TB patients is 14% currently ninety-fifth of TB patients receive HIV testing (UNAIDS, 2013).

The Democratic Republic of Congo is among the thirty countries that bear 87% of the worldwide TB burden. World Health Organization estimates, 270 000 individuals fell sick with TB in 2018, which interprets into the associate incidence of 321 cases per 100000 population, where 31000 (12%) were people living with HIV (Tamuzi et al., 2021).

2.3 Impact of Drought on Tuberculosis Incidence in EAC

Different studies on the impact of drought suggested that drought can be related to respiratory and infectious diseases, drought might impact disease through exposure such as an increase of amount airborne dust or wildfire smoke.

According to research done by Alam et al. (2021) on the impact of drought characteristics on respiratory diseases by using two statistical methods which are the Generalized linear model(GLM) and Poisson regression model. The found results were showing that there is an association between drought duration and the severity of respiratory diseases.

The analytical results from research done by Rios et al. (2000) using mathematical modeling, show that the low rates in T.B. notifications over the amount 1971-1981 have been modified, halting in 1982 and reversing with high incidence from 1983 ahead. Because the features of summer in East Africa are similar to those of droughts, such as precipitation deficiency and high temperature, TB incidence will be high during drought season. As it was said by Harries et al. (2021) in their research they rightly point out that the number of mortality and incidence has increased from climate-sensitive diseases TB is one of the diseases that could be aggravated by

climate change.

Examples of the previous embody raised risk of displaced populations living in incommodious conditions because of severe weather events and raised risk of TB transmission. The latter, vulnerability to TB illness, will be raised by additional frequent severe weather events (such as floods), that disrupt access to preventive and therapeutic care. They continued by saying that one important pathway through that temperature change can have an effect on the TB pandemic is its impact on food security and nutrition. In several high TB burden countries like Asian countries, undernutrition remains the leading risk issue for TB. supported decades of knowledge, we all know that undernutrition is related to the exaggerated risk of TB disease incidence, exaggerated severity of TB disease, and exaggerated mortality (Sinha et al., 2021).

As it was revealed by Center for Deceases Control and prevention(CDC) (2020), Drought can cause an increase in infectious diseases, when rainfall falls, viruses, protozoa, and bacteria can pollute both groundwater and surface water, those with underlying chronic conditions are also at a higher risk. Acute respiratory and gastrointestinal illnesses are more easily transmitted from person to person when hand washing is hampered by a perceived or real lack of available water. When there is a lack of clean water, the risk of infectious disease rises.

Climate change has an impact on infectious diseases both directly and indirectly. Temperature, precipitation, humidity, and radiation all have an impact on the infectious disease by modulating pathogen, host, and transmission pathways. Droughts and floods, for example, have a direct impact on the prevalence and spread of infectious diseases. Climate change has an indirect impact on infectious diseases because it alters the ecological system, including the underlying surface and vegetation distribution (Wu et al., 2014).

3. Research Methodology and Materials

3.1 Research Data

This section goes through the data used in this study, including how it was cleaned and processed.

3.1.1 Panel Data Set. Climate variables and tuberculosis incidence panel data of East African Community countries were used. TB Incidence panel data were obtained from WHO's Global Health Observatory data (World Health Organization(WHO) data, 2022) and Climate data are also obtained from (The archive of the Centre of Environmental Data Analysis (CEDA), 2022), where Average monthly precipitation and potential evapotranspiration data were used to calculate standard precipitation evaporation index (SPEI) and standard precipitation index (SPI) for each country at timescale of 12 months using the SPEI and SPI packages in R.

A panel, or longitudinal, data set contains continual observations on similar units: people, households, firms, countries, or any set of entities that stay stable through time. Panel data allows to study cross-sectional effects, along variation across the entities, and time-series effects, along, variation across time (Das and Van Soest, 1999).

Notation:
In matrix form panel data can be represented as

$$Y_{it} = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{it} \\ \vdots \\ y_{iT} \end{bmatrix}, X_i = \begin{bmatrix} X_{11} & X_{21} & \dots & X_{k1} \\ X_{12} & X_{22} & \dots & X_{k2} \\ \vdots & \vdots & \dots & \vdots \\ X_{1t} & X_{2t} & \dots & X_{it} \\ \vdots & \vdots & \vdots & \vdots \\ X_{1T_i} & X_{2T_i} & \dots & X_{kT_i} \end{bmatrix}$$

where $i = 1, 2, \dots, N$ is index for an entity, $t = 1, 2, \dots, T$ is a period, X_{kT_i} is the k^{th} variable for entity i at time T and Y_{it} is endogenous variable for entity i at time t.

Kind of Panel Data

In general,there are two kinds of panel data which are:

3.1.2 Balanced panel data. Panel data is balanced if all panel individuals have the same number of observations over a given period of time t.

3.1.3 Unbalanced Panel Data. Panel data is said to be unbalanced if the number of time-series observations is different across the panel units at given period of time t (Mućk, 2011).

Panel data models provide information on individual behavior over time and across individuals. The data and models are cross-sectional and time-series in nature. It is critical to utilize this random and exogenous source of variation to identifying a causal relationship between health indicators and weather (Torres-Reyna and Oscar, 2007).

The climate variables and TB variables used in this study are described in table below:

Table 3.1: Description of Tb and climate variables

Variable	description	unit used in this study
SPI	logarithm of Standardized Precipitation Index	mm/day
SPEI	logarithm of standardized Potential Evapotranspiration Index	mm/day
TB Inc	logarithm of Tuberculosis Incidence	population
TB Case	logarithm of Tuberculosis Cases	population
PRE	logarithm of Precipitation	mm/day
PET	logarithm of Potential Evapotranspiration	mm/day

This research focused on 6 East African countries because of continuous research on drought and TB incidence for improving the strategies to reduce TB incidence and drought impacts over this region which are still lacking. The TB incidence and number of TB cases were the used separately as dependent variable while the drought indices (SPI,SPEI) and climate variables (PRE,PET) were the independent variables

Python was used for data visualization while R was used for analysis of the association between Drought indices and TB incidence and TB cases.

3.1.4 Area of the study. The East African Community is a regional intergovernmental organization of 7 partner states: The Democratic Republic of Congo, the Republics of Burundi, Kenya, Rwanda, South Sudan, Uganda. All countries have different drought characteristics because of their topographic features. Even tuberculosis incidence is different for each country Where Tanzania, Kenya, and Uganda are among the 30 countries with the highest burden of TB in the world, whereas Rwanda and Burundi have relatively low TB incidences compared with those other mentioned countries WHO (2019). South Sudan was not included in this study because climate data and TB data were not available for most of the period under review, probably because it is newly created country.

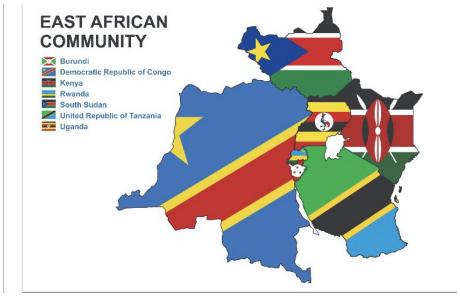


Figure 3.1: East African Community Map (Wandera and Niyibizi, 2018)

3.2 Research Methodology

This chapter discussed more in detail research methods and mathematical models that were used to analyze the data. Models used to examine the association between drought characteristics and TB incidence were the Pooled Ordinary Least Squares (OLS), Random effects, and Fixed effect models. This research was mainly investigating the impacts of Drought on Tuberculosis incidence. It was also based on the test hypothesis of the causal relationship between Drought characteristics and TB incidence while implementing the quantitative approach.

3.2.1 Panel Data Model. Panel data models provide a description of the individual behavior both across time and across individuals (Arellano and Honoré, 2001).
Generally, the equation of panel data model for $i = 1, 2, 3, \ldots, N$ individuals at $t = 1, 2, 3, \ldots, T$ period of time can be written as:

$$Y_{it} = \beta_0 + \sum_{j=1}^{k} \beta_j X_{jit} + \sum_{p=1}^{s} \gamma_p Z_{pi} + \varepsilon_{it} \tag{3.2.1}$$

where:
$i = 1, 2, \ldots, N$ represents individuals,
$t = 1, 2, \ldots, T$ is time period,
j represents observed explanatory variables,
p represents unobserved explanatory variables,
X stands for independent variables of interest,
β is a vector of the parameter of interest, The Z_p variables are unobserved and it is difficult to get the information about the $\sum \gamma_p Z_p$ component of the model. To get around this, introduce the term C_i for unobserved effect that represent the impact of the Z_p on Y_i. Now we have

$$C_i = \sum_{p=1}^{s} \gamma_p Z_{pi} \tag{3.2.2}$$

Hence we can rewrite Equation 3.2.1 as:

$$Y_{it} = \beta_0 + \sum_{j=1}^{k} \beta_j X_{jit} + C_i + \varepsilon_{it}$$

where
β_0 is the intercept

Y_{it} is the Independent variable for i individual at time t

X_{kit} is k^{th} explanatory variable for i individual at time t

C_i is time invariant unobserved variable for individual i

ε_{it} is Error term for i individual at time t

$\beta_1, \ldots, \beta_k$ are the structural parameters

They are different assumptions we have to consider for model accuracy:

A_1 The equation 3.2.1 is linear (Y is linearly associated with the covariate X) where $\mathbb{E}[\varepsilon_{it}] = 0$ and $\mathbb{E}[C_i] = 0$.

A_2 $\mathbb{E}[\varepsilon_i | X_i, C_i] = 0$.

A_3 $Var[\varepsilon_{it}] = \sigma^2 I$ i.e homoskedastic

A_4 Rank(X)= full rank

In vector form Equation 3.2.1 is:

$$Y = X\beta + C + \varepsilon$$

where: for balanced panel data set NT stands for total number of observations Y is $NT \times 1$ vector matrix of response variables, X is $NT \times k$ matrix of independent variables, β is $k \times 1$ vector of coefficient of parameters and ε is $NT \times 1$ vector of error terms

Depending on how the coefficients in the model differ for each cross-section in the panel, we have different types of panel data models. There are three most popular types of model (pooled Ordinary Least Squares (OLS), Fixed effects and Random effects) which were used in this research.

3.2.2 Pooled Ordinary Least Squares Model. Pooled OLS is one of the types of panel data model where, we consider that the unobserved variables are time-invariant and uncorrelated with the observed variables at any period of time.
By this type of model we consider that $C_i = \alpha$. In this case, Equation 3.2.1 can be written as

$$Y_{it} = X'_{it}\beta + \alpha + \varepsilon_{it},$$

where Y_{it} is the value of dependent variable for individual i at time t, β is the vector of parameters of the model, X_{it} is a vector of i covariate for the i^{th} individual at time t and ε is a vector of error terms

In addition of assumption A_1 to A_4 on Equation 3.2.1, this method is also based on the following assumptions:

A_1 $\mathbb{E}[\varepsilon^T \varepsilon] = 0$ i.e error terms are i.i.d. (independent and identically distributed).

A_2 $\mathbb{E}[X_i, \alpha] = 0$.

A_3 we assume that $C_i = \alpha \quad \forall_i$

A_4 No perfect multicollinearity which means that two or more regressors in multiple regression are not allowed to be strongly correlated.

In vector form, we can write

$$Y = X\beta + \varepsilon \tag{3.2.3}$$

Where
Y is $NT \times 1$ matrix of response

X is $NT \times K$ Matrix of explanatory variable

β is $K + 1 \times 1$ matrix of parameters

ε is $NT \times 1$ matrix of error terms with zero conditional mean.

Using the Sum of Squares Estimator(SSE) we can find the values of β where we can find SSE from the equation 3.2.3

$$SSE = \sum \varepsilon^T \varepsilon = (Y - X\beta)^T (Y - X\beta) \tag{3.2.4}$$

By taking $\frac{\partial SSE}{\partial \beta}$ to zero, we get :

$$\hat{\beta}^{Pooled} = (X^T X)^{-1} X^T Y = \begin{bmatrix} \hat{\beta}_1^{Pooled} \\ \vdots \\ \hat{\beta}_k^{Pooled} \end{bmatrix}$$

There are importance of using OLS modeling in analyzing the panel data such as: Pooled cross-sections can be useful for evaluating the impact of certain events or policy interventions.
Pooled OLS can be used to calculate unbiased and consistent parameter estimates even when time constant attributes are present.
And also there are some limitations of this model
There is violation of the conditional mean of error It can cause a problem of heteroskedasticity because of ignoring the unmeasured heterogeneity of inherent of C_i It is appropriate only when the individual fixed effect C_i is not varying with time (Vuko and Čular, 2014).

3.2.3 Fixed Effects Model(FEM). Fixed effects models are statistical models in which The independent variables are assumed to be fixed, and only the dependent variable changes in response to the levels of independent variables. It controls for all time-invariant differences between the individuals where the estimated coefficients of the fixed effects can not be biased due to omitted time-invariant characteristics (Bhargava et al., 1982).
For this model we assume that there is correlation between unobserved variables and observed variables i.e $\mathbb{E}[C_i|X_i] = g(X_i) = \alpha_i^*$. Another assumption is that $C_i = \alpha_i$ is not random and it

is time-invariant and the overall intercept is not (usually) included in the model. we can write general equation 3.2.1 of panel model as:

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \cdots + \beta_k X_{kit} + C_i + u_{it}, \tag{3.2.5}$$

where the C_i are unobserved time-invariant heterogeneities across the entities $i = 1, \ldots, n$ the aim of using this model is to estimate β_1, the effect on Y_i of a change in X_i holding constant C_i By letting $\alpha_i = \beta_0 + C_i$ we obtain

$$Y_{it} = \alpha_i + \beta_1 X_{1it} + \cdots + \beta_k X_{kit} + u_{it} \tag{3.2.6}$$

equation 3.2.6 is called fixed effects model with $i = 1, \ldots n$, and $t = 1, \ldots, T$. The α_i are entity-specific intercepts that capture heterogeneities across entities. we can also represent this model as :

$$Y_{it} = \beta_0 + \sum_{k=1}^{n} \beta_k X_{kit} + \sum_{p=2}^{s} \gamma_p D_{pi} + u_{it} \tag{3.2.7}$$

Model 3.2.7 has p different intercepts, one for each entity and $D_{2i}, \ldots, D_{pi}$ are dummy variables Where: β is the unknown coefficient for each entity.

X_{it} is representing the independent variable for i entity and t time

ε_{it} is the error term.

In order to use Fixed Effect Model we assume that:
A_1 $\mathbb{E}[\varepsilon_{it}|X_{it}] = 0$ which means the error term has zero conditional mean

A_2 X_{it}, ε_{it} are i.i.d for their joint distribution In general we can write the equation of FEM as:

$$Y_{it} = \alpha_i + \beta_1 X_{1it} + \beta_2 X_{2it} + \cdots + \beta_k X_{kit} + u_{it} \tag{3.2.8}$$

By using assumption on time invariant parameter we get:

$$\overline{Y}_i = \alpha_i + \beta_i \overline{X}_{1i} + \beta_2 \overline{X}_{2i} + \cdots + \overline{u}_i \tag{3.2.9}$$

where
$\overline{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$, $\overline{X}_{1i} = \frac{1}{T}\sum_{t=1}^{T} X_{1it}$, $\overline{X}_{ki} = \frac{1}{T}\sum_{t=1}^{T} X_{kit}$ and $\overline{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_i$
by subtracting 3.2.9 from 3.2.8 we get:

$$\begin{aligned} y_{it} - \overline{y}_i &= (\alpha_i - \alpha_i) + \beta_1(X_{1it} - \overline{X}_{1i}) + \cdots + \beta_k(X_{kit} - \overline{X}_{ki}) + (u_{it} - \overline{u}_i) \\ &= \beta_1(X_{1it} - \overline{X}_{1i}) + \cdots + \beta_k(X_{kit} - \overline{X}_{ki}) + (u_{it} - \overline{u}_i) \end{aligned} \tag{3.2.10}$$

Let $\tilde{Y}_{it} = y_{it} - \overline{y}_i$,$\tilde{X}_{1t} = X_{1it} - \overline{X}_{1i}$,$\tilde{X}_{kit} = X_{kit} - \overline{X}_{ki}$ and $\tilde{u}_{it} = u_{it} - \overline{u}_i$

Finally we can rewrite Equation 3.2.10 as:

$$\tilde{Y}_{it} = \beta_1 \tilde{X}_{1t} + \cdots + \beta_k \tilde{X}_{kit} + \tilde{u}_{it} \tag{3.2.11}$$

Note that we don't estimate directly fixed effects.
We can write Equation in vector form 3.2.11 as:

$$\tilde{Y} + \tilde{X}\beta + \tilde{u}$$

where,

$\tilde{Y}$ is $NT \times 1$ matrix

$\tilde{X}$ is $NT \times k$ matrix

β is $k \times 1$ vector

$\tilde{u}$ is $NT \times 1$ matrix

then

$$\tilde{\beta}^{FE} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\tilde{Y} = [\tilde{\beta}_1^{FE} \dots \tilde{\beta}_k^{FE}]^T \tag{3.2.12}$$

In this model, the OLS estimate of the parameter of interest β_1 is equal to the estimate obtained using equation 3.2.7 (Stock and Watson, 2003; Mućk, 2011). As the importance of FEM It does not require $cov(X, \varepsilon) = 0$ which is usually difficult to justify.
It is fully efficient for a large number of individuals and it allows us to control time-invariant omitted variables. even though it has some importance, it has also different limitations like: the fixed-effects model will not work well if there is unobserved heterogeneity due to unmeasured characteristics that do vary over time.
In Fixed Effects Modeling the random variables are treated as though they are not random.
It is hard to estimate the effect of a variable that is time-invariant. changing variables over time (Bhargava et al., 1982; Baltagi et al., 2003).

3.2.4 Random Effects Model(REM). The difference between this type and the fixed effects model is that the differences between individuals are random, drawn from the distribution which is constant parameters when we are applying this model we assume that C_i and X_i are uncorrelated this assumption help us to avoid the problem of bias that may occur during estimation of β (Baltagi et al., 2003). That is

$$\mathbb{E}[C_i|X_i] = 0$$

We also assume that there is no correlation and no auto-correlation between error terms.

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \cdots + \alpha_i + \varepsilon_{it}$$

Where $\alpha_i \sim iid(0, \sigma_\alpha^2)$, and $\varepsilon_{it} \sim iid(0, \sigma_\varepsilon^2)$, α_i is constant with time and homoscedastic across individuals .
Compact Notation

$$Y = X\beta + \nu \quad \text{where} \quad \nu = \alpha_i + \varepsilon_{it}$$

X_i is $T \times k$ matrix

β is $k \times 1$ matrix

ν is $T \times 1$ matrix

Y and ε are $T \times 1$ matrices. then

$$\hat{\beta}^{RE} = (X^T X)^{-1} X^T Y = \begin{bmatrix} \hat{\beta}_1^{RE} \\ \vdots \\ \hat{\beta}_k^{RE} \end{bmatrix}$$

The importance of using Random Effects Modeling are: Random effects rely on a strong assumption: multilateral resistance is normally distributed across countries, with a given standard deviation.
The Random Effects model tells us that multilateral resistance is important, but it doesn't tell us anything about its distribution.
And the limitations of Random Effects modeling are: Random effects require that unobserved heterogeneity obey some probability constraints, i.e. follow a particular distribution.
A random-effects model assumes that explanatory variables have fixed relationships with the response variable across all observations, but that these fixed effects may vary from one observation to another (DerSimonian and Kacker, 2007).

3.3 Diagnostic Tests

This section shows different statistical tests performed to see the behavior of the data and to see the relationships between the independent variables and individual fixed effects and also to choose the model that fit the data

3.3.1 F-Test. In general, F-test in regression compares the fits of various linear models. not like t-tests that may assess just one parametric statistic at a time, the F-test will assess multiple coefficients at the same time.

The F-test of the significance may be a specific sort of the F-test. It compares a model with no predictors to the model that you specify. A regression model without predictors is additionally called associate degree intercept-only model.

The hypotheses for the F-test of overall significance are as follows:

H 0: The fit of the intercept-only model and your model is equal.

H_a: The fit of the intercept-only model is significantly reduced compared to your model.

Mathematically we can find F-test as:

$$F_0 = \frac{\frac{(SSR_r - SSR_{ur})}{q}}{\frac{SSR_{ur}}{(n+(k+1))}},$$

where SSR_r stands for the sum of the squared residuals of the restricted model and SSR_{ur} is the same for the unrestricted model. We also have that n is the number of observations, k is the number of independent variables in the unrestricted model and q is the number of the number of coefficients being jointly tested (Blackwell, 2008).

3.3.2 Hausman Test. Hausman utilized in model choice and compare the estimators of the tested models. Hausman takes a look at what is often used beneath the null hypothesis one amongst the compared models offers consistent and economical results and also the different – consistent, however inefficient, and at a similar time beneath the choice hypothesis the primary the model should provide inconsistent results and also the second – consistent (Baltagi et al., 2003; Pace and LeSage, 2008). The general form of Hausman test statistics is:

$$H = (\hat{\beta}^{FE} - \hat{\beta}^{RE})^T [var(\hat{\beta}^{FE}) - var(\hat{\beta}^{RE})]^{-1} (\hat{\beta}^{FE} - \hat{\beta}^{RE}) \tag{3.3.1}$$

If H is significant i.e $P - value > 5\%$ we can not use the Random-effects model where the tests are developed, with the null and alternative hypotheses.
H_0: Fixed effects model

H_a: Random-effects model

3.3.3 Kruskal-Wallis test. The Kruskal-Wallis (Kruskal & Wallis, 1952) is statistical test that is used to assess the differences among more than two independently sampled groups on a single,and non-normally distributed continuous variable. Non-normally distributed data are suitable for the Kruskal-Wallis test. In contrast, the one-way analysis of variance (ANOVA), which is a parametric test, may be used for a normally distributed continuous variable. The Kruskal-Wallis test is the non parametric version of the one-way ANOVA(McKight and Najab, 2010). They are numerous assumptions for this test which are :

A_1 The dependent variable should be measured at the ordinal or continuous level.

A_2 The independent variable should consist of two or more categorical, independent groups. A Kruskal-Wallis H test is commonly used when there are three or more categorical, independent groups, but it can also be used when there are only two groups (a Mann-Whitney U test is more commonly used), but it can be used for just two groups (i.e., a Mann-Whitney U test is more commonly used for two groups).

A_3 We should have independence of observations, which means that there is no relationship between observations in each group or between groups.

Mathematically we can find this test as:

$$H = (N-1)\frac{\sum_{i=1}^{g} n_i(\overline{r}_i - \overline{r})^2}{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(r_{ij} - \overline{r})^2},$$

where:
N is the total number of observations across all groups,

g is the number of groups,

n_i is the number of observations in group i

r_{ij} is the rank of observation j from group i

$\overline{r}_i = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}$ is the average rank of all observations in group i

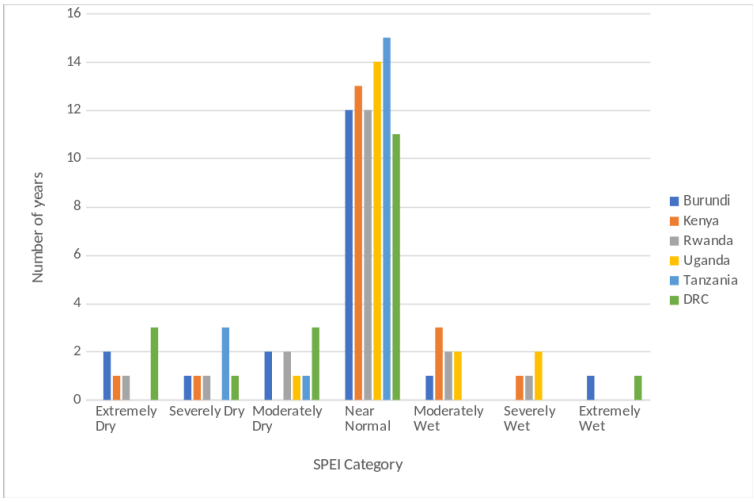
$\overline{r} = \frac{1}{2}(N+1)$ is the average of all the r_{ij}

4. Results and Discussion

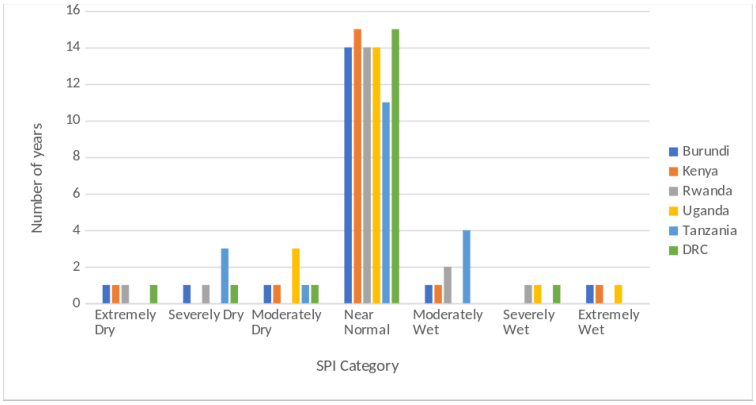
Results

This chapter illustrates and discusses the results obtained using some climatic variables and TB variables that we have discussed in the methodology. The results demonstrate two things. First, figures that show drought occurrence in each country of EAC at a different time period. The second one is the tables to summarize results obtained from performing different panel data models on the association between TB incidence, TB cases, and SPI, SPEI, PRE, and PET. The last one is the tables that summarize the results of Kruskal-Wallis test

4.1 Drought Occurrence In EAC



(a) distribution of the SPEI categories of drought in EAC.



(b) distribution of the SPI categories of drought in EAC.

Figure 4.1a shows the distribution of the SPEI categories of drought in EAC countries between 2000 and 2018. Extremely dry conditions were experienced in DRC for a total of 3 years, in

Burundi for a total of 2 years, and in Kenya and Rwanda for total of 1 year each. Findings further shows that severe dry conditions occurred in Tanzania for a total of 3 years and in Burundi, Kenya, Rwanda and DRC for a total of one year each. Whereas in Figure 4.1b there is the distribution of the SPI categories of drought in EAC countries between 2000 and 2018. Based on SPI drought categories, extremely dry conditions were experienced in Burundi, Kenya, Rwanda and DRC for total of 1 year each. Results further shows that severe dry conditions occurred in Tanzania (for a total of 3 years), Burundi, Kenya, Rwanda and DRC for a total of one year each.

4.2 Panel Data Modelling

In this section, we give the results obtained from OLS, FEM, and REM where PRE, PET SPI,and SPEI were treated as explanatory variables whereas TB cases were treated as the dependent variables. Normality test performed on the data showed that the data for all 6 variables were not normally distributed, hence they were subjected to logarithm transformation before being used in panel data modelling.

4.2.1 Pooled OLS for for TB_cases as response variable.

```
      Source |       SS           df       MS      Number of obs   =       114
-------------+----------------------------------   F(4, 109)       =     12.82
       Model |  12.0147669         4  3.00369172   Prob(F statistic) =   1.377e-08
    Residual |  25.5366201       109  .234280919   R-squared       =    0.3200
-------------+----------------------------------   Adj R-squared   =    0.2950
       Total |   37.551387       113  .332313159   Root MSE        =    .48403

------------------------------------------------------------------------------
 log_tb_case |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     log_pre |   1.717334   .4565494     3.76   0.000     .8124677      2.6222
     log_pet |   5.957737   .8636795     6.90   0.000     4.245952    7.669522
    log_spei |  -1.537605   .5981768    -2.57   0.012    -2.723172   -.3520378
     log_spi |    1.30538   .6916582     1.89   0.062    -.0654639    2.676224
       _cons |  -1.753948   1.272578    -1.38   0.171    -4.276157    .7682614
------------------------------------------------------------------------------
```

Pooled OLS was performed where PRE, PET, SPI, and SPEI were treated as explanatory variables whereas TB cases were treated as the dependent variables. From the above results we see that the (adjusted R-squared = 0.2950 or about 30%) which measures the fraction of the total variance in tuberculosis cases. The (F-test for regression = 12.82, p = 1.377e-08) that measure the joint significance of the model's parameters, thereby leading us to conclude that the model's coefficient estimates are significant at p-value $< 5\%$. All explanatory variables were significantly associated with tuberculosis cases whereas SPI was not significantly associated with TB cases. The TB cases were positively associated with PRE ($\beta = 1.7173, p = 0.000$), TB cases were positively associated with PET ($\beta = 5.9577, p = 0.000$), TB cases were positively associated

with SPI ($\beta = 1.30538, p = 0.062$) whereas TB cases were negatively associated with SPEI ($\beta = -1.537605, p = 0.012$).

4.2.2 Fixed Effects Model for TB_cases as response variable. Since we can not rely on Pooled OLS because of some limitations it is better to use another approach, we applied a fixed effect to see the association between TB cases with SPI, PRE, SPEI and PET.

```
Fixed-effects (within) regression               Number of obs      =       114
Group variable: country_id                      Number of groups   =         6

R-sq:                                           Obs per group:
     within  = 0.0699                                         min =        19
     between = 0.1878                                         avg =      19.0
     overall = 0.1858                                         max =        19

                                                F(4,104)           =      1.95
corr(u_i, Xb)  = 0.2352                         Prob > F           =    0.1073

------------------------------------------------------------------------------
 log_tb_case |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     log_pre |  -.0589282   .0757706    -0.78   0.439    -.2091841    .0913277
     log_pet |   1.416412   .9559817     1.48   0.141    -.4793354     3.31216
    log_spei |   .1415666   .0785468     1.80   0.074    -.0141947    .2973278
     log_spi |  -.0980416   .0884615    -1.11   0.270    -.2734642    .0773809
       _cons |   4.109573   .5838762     7.04   0.000     2.951725    5.267422
-------------+----------------------------------------------------------------
     sigma_u |  .58009531
     sigma_e |  .06003254
         rho |  .98940384   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(5, 104) = 1396.36                     Prob > F = 0.0000
```

From the above results, the (R-squared =0.1858 or about 18%) which measures the fraction of the total variance in tuberculosis cases. The (F-test for regression=1.952 with p=0.1073) that measure the significance of the model's parameters, thereby leading us to conclude that the model's coefficient estimates were not significant since p-value $> 5\%$. We see that the correlation between errors u_i and regressors were 0.2352, the (rho = 99%) which is measure of the variance due to errors u_i. All explanatory variables were not significantly associated with tuberculosis cases. Where TB cases were negatively associated with PRE ($\beta = -0.0589282, p = 0.439$), TB cases were positively associated with PET ($\beta = 1.416412, p = 0.141$), TB cases were positively associated with SPEI ($\beta = 0.1415666, p = 0.074$) and TB cases were negatively associated with SPI ($\beta = -0.0980416, p = 0.270$). In order to choose which model we can use between pooled OLS and Fixed Effect Model we have to perform F-test where if p-value of this test is less that 5% the fixed effects model has to be taken as good model unless Pooled OLS should be considered. The follow are results of F-test:

```
    F test for individual effects

data:  Log_TB_Case ~ Log_PRE + Log_PET + Log_SPEI + Log_SPI
F = 1396.4, df1 = 5, df2 = 104, p-value < 2.2e-16
alternative hypothesis: significant effects
```

From this result we see that the p-value of F-test was significant with 2.2e-16 based on this value we can conclude that Fixed Effect model is the better model than Pooled OLS.

4.2.3 Random Effects Model for TB_cases as response variable. Even if the F-test shows that Fixed Effects Model is better than Pooled OLS that based on its performance we are not going to conclude directly that it is model for prediction that is why we have to try the another competing model.

```
Random-effects GLS regression                   Number of obs     =       114
Group variable: country_id                      Number of groups  =         6
R-sq:                                           Obs per group:
     within  = 0.0697                                         min =        19
     between = 0.1904                                         avg =      19.0
     overall = 0.1885                                         max =        19

                                                Wald chi2(4)      =      8.53
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =    0.0739

------------------------------------------------------------------------------
 log_tb_case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     log_pre |   -.055628    .075194    -0.74   0.459    -.2030054    .0917495
     log_pet |   1.544031   .9181525     1.68   0.093     -.255515    3.343577
    log_spei |   .1400319   .0780574     1.79   0.073    -.0129577    .2930216
     log_spi |  -.0968096   .0879172    -1.10   0.271    -.2691241    .0755048
       _cons |    4.03353   .6258687     6.44   0.000      2.80685     5.26021
-------------+----------------------------------------------------------------
     sigma_u |   .6744387
     sigma_e |  .06003254
         rho |   .9921393   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

From the above results, the (R-squared $= 0.1885$ or about 19%) which measures the fraction of the total variance in tuberculosis cases. The (Ward chi2-test$= 1.952$, p$= 0.1073$) that measure the significance of the model's parameters leading us to conclude that the model's coefficient estimates are not significant since p-value $> 5\%$. We see that the correlation between errors u_i and regressors is 0 from assumption, (rho$=$ 99%) which is measure of the variance due to errors u_i. All explanatory variables are not significantly associated with tuberculosis cases

TB cases were negatively associated with PRE ($\beta = -0.055628, p = 0.459$),TB cases were positively associated with PET ($\beta = 1.544031, p = 0.093$), TB cases were positively associated with SPEI ($\beta = 0.1400319, p = 0.073$) and TB cases were negatively associated with SPI ($\beta = -0.0968096, p = 0.459$). In order to choose which model we can use between Fixed Effect Model and Random Effects Models we have to perform Hausman test where if p-value of this test is less that 5% the fixed effects model has to be taken as good model unless Random Effects Model should be considered. The following are results obtained from Hausman test:

```
     Hausman Test

data:  Log_TB_Case ~ Log_PRE + Log_PET + Log_SPEI + Log_SPI
chisq = 0.25075, df = 4, p-value = 0.9928
alternative hypothesis: one model is inconsistent
```

From this result we see that Random Effect Model is better that Fixed since p-value of the test is greater that 5%

4.2.4 Pooled OLS for for TB_Inc as response variable.

```
.  regress log_tb_inc log_pre log_pet log_spei log_spi

      Source |       SS           df       MS      Number of obs   =       114
-------------+----------------------------------   F(4, 109)       =     18.30
       Model |  3.52978718         4  .882446795   Prob > F        =     1.588e-11
    Residual |   5.2567665       109  .048227216   R-squared       =    0.4017
-------------+----------------------------------   Adj R-squared   =    0.3798
       Total |  8.78655368       113  .077757112   Root MSE        =    .21961

------------------------------------------------------------------------------
  log_tb_inc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     log_pre |   .4988034   .2071407     2.41   0.018     .0882573    .9093495
     log_pet |   2.873287   .3918594     7.33   0.000     2.096634    3.649939
    log_spei |  -.7613866   .2713983    -2.81   0.006    -1.299289   -.2234839
     log_spi |   .5733261   .3138117     1.83   0.070    -.0486385    1.195291
       _cons |  -.0819182   .5773806    -0.14   0.887    -1.226268    1.062431
------------------------------------------------------------------------------
```

Pooled OLS was performed where PRE, PET, SPI, and SPEI were treated as explanatory variables whereas TB Incidence were treated as the dependent variables. From the above results, the adjusted R-squared which measures the fraction of the total variance in response variable is 0.3798 or about 37% this number is small i.e we can rely on and conclude that OLS is good model for prediction.

The F-test for regression that measure the joint significance of the model's parameters has produced a test statistic of 18.30 with p-value of 1.588e-11 thereby leading us to conclude that

the model's coefficient estimates are joint significant at p-value $< 5\%$. Since the performance of this model is not enough for using it in prediction we need compare it to a competing model.

From results we see that if we apply Pooled OLS all explanatory variables are not significantly associated with response variables where PRE, PET, and SPEI are significantly correlated with TB incidence whereas SPI is not significantly associated with TB cases. By this model, we can see that TB incidence will be increased by 0.4988034 for 1 unit of precipitation, for potential evapotranspiration TB incidence will be increased by 2.873287 for 1 unit of PET, for SPI TB incidence will be increased by 0.5733261 for one unit of SPI whereas it will be decreased by 0.7613866 for 1 unit of SPEI.

4.2.5 Fixed Effects Model for TB_Inc as response variable. Since we can not rely on Pooled OLS because of some limitations it is better to use another approach, we applied a fixed effect to see the association between TB incidence with SPI, PRE,SPEI, and PET.

```
    . xtreg log_tb_inc log_pre log_pet log_spei log_spi,fe

Fixed-effects (within) regression               Number of obs      =       114
Group variable: country_id                      Number of groups   =         6

R-sq:                                           Obs per group:
     within  = 0.0600                                         min =        19
     between = 0.3554                                         avg =      19.0
     overall = 0.3243                                         max =        19

                                                F(4,104)           =      1.66
corr(u_i, Xb)  = -0.1326                        Prob > F           =    0.1652

------------------------------------------------------------------------------
  log_tb_inc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     log_pre |  -.1198523   .1152087    -1.04   0.301    -.3483155    .1086108
     log_pet |    2.17402   1.453564     1.50   0.138    -.7084525    5.056493
    log_spei |   .0260968   .1194299     0.22   0.827    -.2107371    .2629308
     log_spi |  -.0948705   .1345052    -0.71   0.482    -.3615994    .1718585
       _cons |   1.469497   .8877801     1.66   0.101    -.2910045    3.229998
-------------+----------------------------------------------------------------
     sigma_u |   .2330893
     sigma_e |  .09127911
         rho |  .86703569   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(5, 104) = 105.38                     Prob > F = 0.0000
```

In order to see the performance of model we need to check on different things which are:

From the above results, R-squared which measures the fraction of the total variance in response variable is 0.3243 or about 32% that is good number but it is not enough so that we can say

that it is good performance and conclude that FEM is good model for prediction.

The F-test for regression that measure the joint significance of the model's parameters has produced a test statistic of 1.66 with p-value of 0.1652 thereby leading us to conclude that the model's coefficient estimates are not joint significant since p-value $> 5\%$.For Fixed Effects Model we assume that $\mathbf{E}(u_i|X_i) \neq 0$, from the result we see that the correlation between errors u_i and regressors is -0.1326, rho which is measure of the variance due to errors u_i is 87%. From results we see that if we apply Fixed Effects Model all explanatory variables are not significantly associated with endogenous variables By this model, we can see that TB incidences will be decreased by 0.1198523 for 1 unit of precipitation, for potential evapotranspiration TB incidence will be increased by 2.17402 for 1 unit of PET, for SPEI TB incidence will be increased by 0.0260968 for one unit of SPEI whereas it will be decreased by 0.0948705 for 1 unit of SPI.

In order to choose which model we can use between pooled OLS and Fixed Effect Model we have to perform F-test where if p-value of this test is less that 5% the fixed effects model has to be taken as good model unless Pooled OLS should be considered.

The follow is the result of F-test:

```
F-test for individual effects

data:  Log_TB_Inc ~ Log_PRE + Log_PET + Log_SPEI + Log_SPI
F = 592.19, df1 = 5, df2 = 104, p-value < 2.2e-16
alternative hypothesis: significant effects
```

From this result we see that the p-value of this test is significant with 2.2e-16 based on this value we can conclude that Fixed Effect model is the better model than Pooled OLS.

4.2.6 Random Effects Model for TB_Inc as response variable. from result of the F-test we clearly see that Fixed Effects Model is better than Pooled OLS but that based on its performance we are not going to conclude directly that it is model for prediction that is why we have to try the another competing model.

```
.  xtreg log_tb_inc log_pre log_pet log_spei log_spi,re

Random-effects GLS regression                   Number of obs     =       114
Group variable: country_id                      Number of groups  =         6

R-sq:                                           Obs per group:
     within  = 0.0597                                         min =        19
     between = 0.3574                                         avg =      19.0
     overall = 0.3261                                         max =        19

                                                Wald chi2(4)      =     11.65
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =    0.0202

------------------------------------------------------------------------------
```

log_tb_inc	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
log_pre	-.1057407	.1154282	-0.92	0.360	-.3319757	.1204944
log_pet	1.971447	.7670248	2.57	0.010	.4681058	3.474788
log_spei	.0107835	.1219639	0.09	0.930	-.2282613	.2498283
log_spi	-.0820094	.1375311	-0.60	0.551	-.3515654	.1875466
_cons	1.553305	.5359404	2.90	0.004	.5028814	2.603729
sigma_u	.16041252					
sigma_e	.09127911					
rho	.75540562	(fraction of variance due to u_i)				

From results, if we apply Random Effects Model all explanatory variables are not significantly associated with endogenous variables except potential evapotranspiration (B = 1.97144, p = 0.010). By this model, we can see that TB incidence will be decreased by 0.1057407 for 1 unit of precipitation, for potential evapotranspiration TB incidence will be increased by 1.971447 for 1 unit of PET, for SPEI TB incidence will be increased by 0.0107835 for one unit of SPEI whereas it will be decreased by 0.0820094 for 1 unit of SPI. The Chi-square of the model was significant (Wald chi2(4) =11.65, p = 0.0202) and it explained about 33% of the variances of TB incidence (R squared = 0.3261). In order to choose which model we can use between Fixed Effect Model and Random Effects Models we have to perform Hausman test where if p-value of this test is less that 5% the fixed effects model has to be taken as good model unless Random Effects Model should be considered. The following are results obtained from Hausman test:

```
Hausman Test

data:  Log_TB_Inc ~ Log_PRE + Log_PET + Log_SPEI + Log_SPI
chisq = 0.87987, df = 4, p-value = 0.9274
alternative hypothesis: one model is inconsistent
```

From this result we see that Random Effect Model is better that Fixed since p-value of the test is greater that 5%

SPI Category	TB Cases		TB Incidence	
Burundi	**Median(IQR)**	**H-value(P-value)**	**Median(IQR)**	**H-value(P-value**
Extremely Dry	150000 (150000, 150000)	5.647(0.342)	211(211, 211)	4.051(0.542)
Severely Dry	12000(12000, 12000)		133(133, 133)	
Moderately Dry	14000(14000, 14000)		195(195, 195)	
Near Normal	13000(12000, 14500)		148(121, 193.25)	
Moderately wet	18000(18000, 18000)		270(270, 270)	
Severely Wet	-		-	
Extremely Wet	12000(12000, 12000)		138(138, 138)	
Kenya				
Extremely Dry	144000(144000, 144000)	**6.334(0.176)**	451(451, 451)	**3.987(0.408)**
Severely Dry	-		-	
Moderately Dry	231000(213000, 213000)		630(630, 630)	
Near Normal	212000(182000, 226000)		506(423, 610)	
Moderately wet	180000(180000, 180000)		534(534, 534)	
Severely Wet	-		-	
Extremely Wet	160000(16000, 1600000		319(319, 319)	
Rwanda				
Extremely Dry	8100(8100, 8100)	**3.789(0.435)**	93(93,93)	**1.863(0.761)**
Severely Dry	9000(9000, 9000)		102(102, 102)	
Moderately Dry	-		-	
Near Normal	7550(7175, 9375)		80.5(61.75, 100.25)	
Moderately wet	8350(8300, -)		81.5(80, -)	
Severely Wet	6800(6800, 6800)		83(83, 83)	
Extremely Wet	-		-	260
DRC				
Extremely Dry	158000(158000, 158000)	**6.918(0.140)**	327(327, 327)	**2.115(0.715)**
Severely Dry	168000(168000, 168000)		327(327, 327)	
Moderately Dry	173000(173000, 173000)		327(327, 327)	
Near Normal	218000(191000,247000)		327(324, 327)	
Moderately wet	-		-	
Severely Wet	163000(163000, 163000)		327(327, 327)	
Extremely Wet	-		-	
Tanzania				
Extremely Dry	-	**0.382(0.944)**	-	**1.921(0.589)**
Severely Dry	185000(147000, -)		510(269, -)	
Moderately Dry	189000(189000, 189000)		426(426, 426)	
Near Normal	178000(163000, 191000)		474(336, 506)	
Moderately wet	173500(144500, 198000)		369.5(261.5, 491)	
Severely Wet	-		-	
Extremely Wet	-		-	
Uganda				
Extremely Dry	-	**1.333(0.721)**	-	**1.019(0.797)**
Severely Dry	-		-	
Moderately Dry	65000(64000, -)		233(202, -)	
Near Normal	66500(65000, 77750)		215(201.75, 244)	
Moderately wet	-		-	
Severely Wet	71000(71000, 71000)		205(205,205)	
Extremely Wet	73000(73000, 73000)		203(203,203)	

Table 4.1: **Comparison of TB cases and incidence across categories of SPI drought index**

SPEI Category	TB Cases		TB Incidence	
Burundi	**Median(IQR)**	**H-value(P-value)**	**Median(IQR)**	**H-value(P-value**
Extremely Dry	13500 (12000, -)	**4.117(0.533)**	172(133, -)	**3.178(0.673)**
Severely Dry	14000(14000, 14000)		195(195, 195)	
Moderately Dry	15500(12000, -)		202(114, -)	
Near Normal	13000(12000, 13750)		148(123, 178)	
Moderately wet	18000(18000, 18000)		270(270, 270)	
Severely Wet	-		-	
Extremely Wet	12000(12000, 12000)		138(138, 138)	
Kenya				
Extremely Dry	144000(144000, 144000)	**5.623(0.229)**	451(451, 451)	**4.118(0.39)**
Severely Dry	231000(231000, 231000)		630(630, 630)	
Moderately Dry	-		-	
Near Normal	212000(184000,225500)		506(433, 588)	
Moderately wet	182000(180000, -)		534(380, -)	
Severely Wet	160000(160000, 160000)		319(319, 319)	
Extremely Wet	-		-	
Rwanda				
Extremely Dry	8100(8100, 8100)	**4.639(0.461)**	93(93, 93)	**4.036(0.544)**
Severely Dry	9000(9000, 9000)		102(102, 102)	
Moderately Dry	8600(7600, -)		98.5(96, -)	
Near Normal	7400(7125, 9100)		69(61.25, 98.75)	
Moderately wet	8350(8300, -)		81.5(80, -)	
Severely Wet	6800(6800,6800)		83(83, 83)	
Extremely Wet	-		-	
DRC				
Extremely Dry	168000(15800, -)	**8.834(0.065)**	327(327, 327)	**5.769(0.217)**
Severely Dry	179000(179000,179000)		327(327, 327)	
Moderately Dry	198000(19100, -)		327(327, 327)	
Near Normal	233000(204000, 254000)		326(323,327)	
Moderately wet	-		-	
Severely Wet	-		-	
Extremely Wet	163000(163000, 163000)		327(327, 327)	
Tanzania				
Extremely Dry	-	**0.304(0.859)**	-	**0.811(0.667)**
Severely Dry	185000(147000, -)		510(269, -)	
Moderately Dry	189000(189000, 189000)		426(426, 426)	
Near Normal	178000(163000, 195000)		452(327, 504)	
Moderately wet	-		-	
Severely Wet	-		-	
Extremely Wet	-		-	
Uganda				
Extremely Dry	-	**5.682(0.128)**	-	**5.299(0.151)**
Severely Dry	-		-	
Moderately Dry	65000(65000, 65000)		233(233,233)	
Near Normal	65500(64750, 70500)		219.5(205.75, 250)	
Moderately wet	81500(77000, -)		201(200, -)	
Severely Wet	72000(71000, -)		204(204, 204)	
Extremely Wet	-		-	

Table 4.2: **Comparison of TB cases and incidence across categories of SPEI drought index**

Table 4.2 shows the Kruskal Wallis test comparison of number of TB cases and incidence across categories of drought indices (extremely dry, severely dry, moderately dry, near normal, moderately wet, severely wet, and extremely wet) in EAC based on SPEI. The results showed no significant differences in TB cases and incidence across the categories of drought indices in each of the six countries of EAC. Table 4.1 shows the Kruskal Wallis test comparison of number of TB cases and incidence across categories of drought indices (extremely dry, severely dry, moderately dry, near normal, moderately wet, severely wet, and extremely wet) in EAC based on SPI. The results showed no significant differences in TB cases and incidence across the categories of drought indices in each of the six countries of EAC

Discussion

positive associations between both TB cases and TB incidence and independent variables, indicate that during the winter or rainy season the TB cases will be increased. This is because during winter there is vitamin D deficiency which will lead to the reduction of human immunity, This weakens the immune system and makes people more susceptible to tuberculosis.

As it was shown in a report reported by Fernandes et al. (2017) TB incidence and TB cases increase in a different seasons where in some regions TB Incidence and TB cases are high in winter while in other regions they increase in summer, coming to the results found from REM as the best model the negative association between precipitation and response variables this indicate that in EAC TB cases and Incidence will be high in summer, because poor air quality due to air pollution during summer the an enhanced increase in respiratory diseases including Tuberculosis. The association between potential evapotranspiration and TB incidence were positive and significant, this indicates that once evapotranspiration increase there will be an increase in CO2 in the atmosphere because once evaporation is taking place on green plant and we all know that plants have more CO2 because of photosynthesis process, this is a problem to the respiratory system so increase of PET can enhance the increase of TB cases and TB incidence once people have a respiratory problem. If they come in contact with contagious people they will be infected easily.

The association between standardized potential evapotranspiration index(SPEI) and TB cases and even TB incidence is positive according to the results from the Random Effects model this indicates that one SPI increases TB cases and TB will also increase according to the meaning of SPEI once it is increasing positively means that the precipitation is increasing and we saw that during the rainy season there will be vitamin D deficiency which will cause the decline in human immunity one people lost their body immunity to be infected will be easier. On other hand, if SPEI decreases to negative means that there is precipitation deficiency which will cause food shortage, poverty and malnutrition this will cause TB cases and incidence to increase because this will cause people to move from their region to find jobs and food.

Furthermore, impaired water quality due to extreme weather events such as flooding may lead to infectious diarrhea and parasitic infections that exacerbate undernutrition in people vulnerable to TB. As it was said by Sinha et al. (2021) in research done in India, based on decades of information, we all know that undernutrition is related to the augmented risk of TB incidence, augmented severity of TB, and augmented mortality. Granger food production in Bharat is vulnerable to

irregular rain, extreme climate events, hotter temperatures, a discount in productive land by saltwater infiltration thanks to rising ocean levels and cyclones, and reduced crop output.

This shows that some climate change events like drought have a great role in increasing TB cases in different regions backing on EAC also effects of climate change events are there whereas we have seen in the figure 4.1b, 4.1a there is the occurrence of drought in a different year and this drought event could affect precipitation, potential evapotranspiration over the area this will cause food scarcity which will force people to move from their region in order to get food as people are moving they have to speak, sneeze, coughing and so on if there are no policies that can help them to protect themselves this will risk them to be affected by TB and its cases will be increased. EAC is a region which has historical seen period of prolonged and extreme drought across the region causing food insecurity, frequent mortality (Secretariat and Secretariat, 2010). One of the known consequence of droughts and their associated risk factors are infectious diseases out break which are mostly increased by malnutrition, poor access of water, sanitation, hygiene, and population displacement. TB is a potential causative agent at such out break (Penrose et al., 2010). There were some limitations to this research some of them are different in social economic activities, different government policies are taken to control tuberculosis and droughts.

5. Conclusion and Recommendation

5.1 Conclusion

In EAC both tuberculosis cases and incidence have been affected by the drought, precipitation and Potential evapotranspiration. This challenge put EAC under pressure to look for long-term strategies to reduce the tuberculosis cases and incidence which may rise because of drought, PRE and PET change. The study of analyzing the impacts of drought indices, precipitation, and Potential evapotranspiration variables on tuberculosis cases and incidence variables was done using the Pooled, Fixed Effects and Random Effects models.

The results found using the Polled OLS model were not accurate to interpret the relationship between drought indices, precipitation, and Potential evapotranspiration variables on tuberculosis cases and incidence because its results were not well-fitted to the observed tuberculosis cases and incidence data due to the big difference between the mean and variance in the data other reason why we can't rely on the results of this model is because it ignores the individuality among panel units and observation time dependence. The Fixed Effects model results showed that there is no significant association between drought indices, precipitation, and Potential evapotranspiration variables and tuberculosis cases and incidence this means that the effect of drought, precipitation and potential evapotranspiration on tuberculosis cases and incidence is not reasonable in EAC from 2000 to 2018.

On other hand Random effects shows that potential evapotranspiration kept affecting tuberculosis cases and incidence in Some regions. Tuberculosis cases and incidence increase in winter and others summer due to the weather conditions that could be found in the region. Based on Random Effects results, tuberculosis cases and incidence were highly affected by the increase in evapotranspiration in EAC. Tuberculosis incidence and cases mostly depend on average rainfall during the season, where in winter tuberculosis cases and incidence are increasing this is because in winter there is vitamin D deficiency and also the displacement of people due to landslide, floods all of these problem may cause TB cases and incidence to increase once people get in contact while they are looking for place to live. In summer also Tuberculosis cases and incidence increase because of problem of poor air quality due to air pollution, banks food shortage which could cause malnutrition once people get in problem of malnutrition it will be easier for them to get infected because of lacking the human Immunity. To enable effective planning for a potentially more drought-prone EAC, inequity must be addressed, research on the health implications of drought should be enhanced, and better drought diplomacy is required to improve drought resilience under climate change.

5.2 Recommendation

All EAC countries used to meet with the challenge of applying Tuberculosis adaptation strategies due to a lack of knowledge and skills related to the issues of drought, precipitation and potential evapotranspiration. The long-term actions to address those issues include training programs for people and leaders to work together during implementation. Training might enable them to

enhance management methods to adjust for the effects of drought, precipitation, technology advances, relocate high number tuberculosis cases and incidence seasons or move to other locations. When all of the adaptation techniques outlined in conclusion are implemented, people will be able to manage the impact of drought and control the rise of tuberculosis cases and incidence over all East African Community. We recommend further studies to consider many climate variables to analyze the relationship between climate conditions and Tuberculosis.

References

Intekhab Alam, Shinji Otani, Abir Majbauddin, Qing Qing, Satomi Funaki Ishizu, Toshio Masumoto, Hiroki Amano, and Youichi Kurozawa. The effects of drought severity and its aftereffects on mortality in bangladesh. *Yonago Acta Medica*, 64(3):292–302, 2021.

Manuel Arellano and Bo Honoré. Panel data models: some recent developments. In *Handbook of econometrics*, volume 5, pages 3229–3296. Elsevier, 2001.

Arsenault, S. C. Roder-DeWan, & Kruk, and M. E. Measuring and improving the quality of tuberculosis care: A framework and implications from the lancet global health commission. journal of clinical tuberculosis and other mycobacterial diseases, 16, 100112. retrieved fromhttps://www.ncbi.nlm.nih.gov/pubmed/31497655. 2019.

Badi H Baltagi, Georges Bresson, and Alain Pirotte. Fixed effects, random effects or hausman–taylor?: A pretest estimator. *Economics letters*, 79(3):361–369, 2003.

Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 1–4. Springer, 2009.

Alok Bhargava, Luisa Franzini, and Wiji Narendranathan. Serial correlation and the fixed effects model. *The Review of Economic Studies*, 49(4):533–549, 1982.

Matthew Blackwell. Multiple hypothesis testing: The f-test. *Matt Blackwell Research*, pages 1–7, 2008.

Thomas B.McKee, Nolan J. Doesken, and John Kleist. The relationship of drought frequency and duration to time scales, https://www.droughtmanagement.info. 1993.

Kimberly Bryan, Sarah Ward, Liz Roberts, Mathew P White, Owen Landeg, Tim Taylor, and Lindsey McEwen. The health and well-being effects of drought: assessing multi-stakeholder perspectives through narratives from the uk. *Climatic Change*, 163(4):2073–2095, 2020.

CDC Data. Health implication of drought. Technical report, National Center for Environmental Health, 2020.

Center for Deceases Control and prevention(CDC). Health implications of drought. *Lung Diseases and Respiratory Health*, 2020.

Centers for Disease Control and Prevention. When every drop counts: protecting public health during drought conditions-a guide for public health professionals. *Centers for Disease Control and Prevention. Department of Health and Human Services: Atlanta, GA, USA*, 2010.

Jiachen Chen and W Kenneth Jenkins. Facial recognition with pca and machine learning methods. In *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, pages 973–976. IEEE, 2017.

Marcel Das and Arthur Van Soest. A panel data model for subjective information on household income growth,Lecture15PanelDataModelshttps://www.bauer.uh.edu. *Journal of Economic Behavior & Organization*, 40(4):409–426, 1999.

Rafael E De Hoyos and Vasilis Sarafidis. Testing for cross-sectional dependence in panel-data models. *The stata journal*, 6(4):482–496, 2006.

Rebecca DerSimonian and Raghu Kacker. Random-effects model for meta-analysis of clinical trials: an update. *Contemporary clinical trials*, 28(2):105–114, 2007.

John A Dracup, Kil Seong Lee, and Edwin G Paulson Jr. On the definition of droughts. *Water resources research*, 16(2):297–302, 1980.

Richard E.Chaisson, M.D., Neil A. Martinson, M.B., B.Ch., and M.P.H. *Tuberculosis in Africa*. The new England Journal of medicine, 2008.

Elsevier. Journal of clinical tuberculosis and other mycobacterial diseases, 2021.

Fernanda Monteiro de Castro Fernandes, Eder de Souza Martins, Daniella Melo Arnaud Sampaio Pedrosa, and Maria do Socorro Nantua Evangelista. Relationship between climatic factors and air quality with tuberculosis in the federal district, brazil, 2003-2012. *Brazilian Journal of Infectious Diseases*, 21:369–375, 2017.

Yalemzewod Assefa Gelaw, Weiwei Yu, Ricardo J Soares Magalhães, Yibeltal Assefa, and Gail Williams. Effect of temperature and altitude difference on tuberculosis notification: a systematic review. *Journal of global infectious diseases*, 11(2):63, 2019.

Gebremedhin Gebremeskel Haile, Qiuhong Tang, Siao Sun, Zhongwei Huang, Xuejun Zhang, and Xingcai Liu. Droughts in east africa: Causes, impacts and resilience. *Earth-science reviews*, 193:146–161, 2019.

AD Harries, L Martinez, and JM Chakaya. Tackling climate change: measuring the carbon footprint of preventing, diagnosing and treating tb. *Public health action*, 11(1):40–40, 2021.

IC Harris, PD Jones, et al. Cru ts4. 01: Climatic research unit (cru) time-series (ts) version 4.01 of high-resolution gridded data of month-by-month variation in climate (jan. 1901–dec. 2016). *Centre for Environmental Data Analysis*, 25, 2017.

Graeme D Hutcheson. Ordinary least-squares regression. *L. Moutinho and GD Hutcheson, The SAGE dictionary of quantitative management research*, pages 224–228, 2011.

IPCC. Droughts in asian least developed countries: Vulnerability and sustainability. weather and climate extremes. pages 7 8–23. Earth and Environmental Science, 2015.

Vinit K Jain, Rajendra P Pandey, Manoj K Jain, and Hi-Ryong Byun. Comparison of drought indices for appraisal of drought characteristics in the ken river basin. *Weather and Climate Extremes*, 8:1–11, 2015.

Kenya. *Documents on TB and asthma management, https://african.business/2022/03/apo-newsfeed/kenya-documents-on-tb-and-asthma-management-launched/*. World health organization, 2022.

David L Kidder and Thomas R Worsley. Causes and consequences of extreme permo-triassic warming to globally equable climate and relation to the permo-triassic extinction and recovery. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 203(3-4):207–237, 2004.

Irwin Law. Global tuberculosis report on east african community. 2013.

Makerere-University and WHO. Ministry of Health and partners review the National TB prevalence survey report. 2016.

Patrick E McKight and Julius Najab. Kruskal-wallis test. *The corsini encyclopedia of psychology*, pages 1–1, 2010.

Ben Mhenni and Shinoda. Assessment of drought frequency, severity, and duration and its impacts on vegetation greenness and agriculture production in mediterranean dryland: A case study in tunisia. natural hazards, 105(3), 2755–2776, https://doi.org/10.1007/s11069-020-04422-w. 2021.

Patrick Migambi, Michel Gasana, Claude Bernard Uwizeye, Eliane Kamanzi, Vedaste Ndahindwa, Nico Kalisvaart, and Eveline Klinkenberg. Prevalence of tuberculosis in Rwanda: Results of the first nationwide survey in 2012 yielded important lessons for TB control. 2012.

Ashok K. Mishra. Review of drought concepts. 2010.

Suleiman Yahaya MOHAMMED. Drought incidence and its effects on agricultural production in kano metropolis, kano state, nigeria. 2017.

Jakub Mućk. Econometrics of panel data. *Quantitative Economics*, pages 12–30, 2011.

NASA. Assessment of drought events, their trend and teleconnection factors over , east africa. *Theoretical and Applied Climatology*, 2014.

David J Nash, Jørgen Klein, Georgina H Endfield, Kathleen Pribyl, George CD Adamson, and Stefan W Grab. Narratives of nineteenth century drought in southern africa in different historical source types. *Climatic Change*, 152(3):467–485, 2019.

National Tuberculosis and Leprosy Programme (NTLP). Tb prevalence in Tanzania. 2020.

National Oceanic and Atmospheric Administration. In *"Warm Temperatures and Severe Drought Continued in April Throughout Parts of the United States; Global Temperature For April Second Warmest on Record*. National Oceanic and Atmospheric Administration, 2002.

R Kelley Pace and James P LeSage. A spatial hausman test. *Economics Letters*, 101(3):282–284, 2008.

W.C. Palmer. Meteorological drought.u.s.weather bureau res. paper no. 45, 58. 1965.

Panu and Sharma. Challenges in drought research: some perspectives and future directions. j. hydrol. sci.47. 2002.

Katherine Penrose, Marcia Caldas de Castro, Japhet Werema, and Edward T Ryan. Informal urban settlements and cholera risk in dar es salaam, tanzania. *PLoS Neglected Tropical Diseases*, 4 (3):e631, 2010.

M Hashem Pesaran. Estimation and inference in large heterogenous panels with cross section dependence. *Available at SSRN 385123*, 2003.

M Rios, JM Garcia, JA Sanchez, and D Perez. A statistical analysis of the seasonality in pulmonary tuberculosis. *European journal of epidemiology*, 16:483–488, 2000.

Roelsgaard.E and Nyboe.J. A tuberculosis survey in Kenya. Bull World Health Organ [Internet]. 1961 [cited 2018 Jul 6];25(6):851–70. Available from:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2555628/ViewArticleGoogleScholar. 2016.

EAC Secretariat and EAC Secretariat. Eac climate change policy. 2010.

Pranay Sinha, Knut Lönnroth, Anurag Bhargava, Scott K Heysell, Sonali Sarkar, Padmini Salgame, William Rudgard, Delia Boccia, Daniel Van Aartsen, and Natasha S Hochberg. Food for thought: addressing undernutrition to end tuberculosis. *The Lancet Infectious Diseases*, 21 (10):e318–e325, 2021.

James H Steiger. Tests for comparing elements of a correlation matrix. *Psychological bulletin*, 87(2):245, 1980.

Anne Steinemann. Drought indicators and triggers: a stochastic approach to evaluation 1. *JAWRA Journal of the American Water Resources Association*, 39(5):1217–1233, 2003.

Stock and Watson. Fixed effects model. 2003.

Jacques L Tamuzi, Ley M Muyaya, and Patrick Mbuesse. Hiv/tb co-infection prevalence and treatment outcomes in democratic republic of the congo from 2005 to 2020: Systematic review and meta-analysis. *Community Medicine and Health Education Research*, 2(1):17–35, 2021.

Eurosurveillance Editorial Team et al. Who publishes global tuberculosis report 2013. *Eurosurveillance*, 18(43):20615, 2013.

The archive of the Centre of Environmental Data Analysis (CEDA). fromhttps://catalogue.ceda.ac.uk/uuid/c26a65020a5e4b80b20018f148556681. *CEDA*, 2022.

Torres-Reyna and Oscar. Panel data analysis fixed and random effects using Stata (v. 4.2).Data and Statistical Services, Priceton University 112. 2007.

G Tsakiris, A Loukas, D Pangalou, H Vangelis, D Tigkas, G Rossi, A Cancelliere, et al. Drought characterization. *Drought management guidelines technical annex*, 58:85–102, 2007.

UNAIDS. HIV and Tuberculosis in Burundi. 2013.

UN_Secretariat_General. United nations convention to combat drought and desertification in countries experiencing serious droughts and/or desertification, particularly in africa. 1994.

Sergio M Vicente-Serrano, Santiago Beguería, and Juan I López-Moreno. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate*, 23(7):1696–1718, 2010.

Tina Vuko and Marko Čular. Finding determinants of audit delay by pooled ols regression analysis. *Croatian Operational Research Review*, pages 81–91, 2014.

Catherine Wandera and Alphonse Niyibizi. Potential benefits of adoption of laser materials processing in east africa's manufacturing industry. In *3rd DeKUT International Conference on STI&E NYERI, KENYA*, pages 125–141, 2018.

WebMD. Tuberculosis. *Lung Diseases and Respiratory Health*, 2020.

Halbert White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society*, pages 817–838, 1980.

WHO. *Tuberculosis report*. World health organization, 2019.

WHO. Global tuberculosis report. 2020.

WHO. *Tuberculosis report*. World health organization, 2021.

World Health Organization. *Drought report*. World health organization, 2022.

World Health Organization(WHO). Tuberculosis. *Lung Diseases and Respiratory Health*, 2021.

World Health Organization(WHO) data. https://apps.who.int/gho/data/node.main.1320?lang=en. *WHO's Global Health Observatory data*, 2022.

XiaoXu Wu, HuaiYu Tian, Sen Zhou, LiFan Chen, and Bing Xu. Impact of global change on transmission of human infectious diseases. *Science China Earth Sciences*, 57(2):189–203, 2014.