Determination of best fit model for the distribution and crop loss associated with Bacterial wilt of tomatoes

Ogeya Fredrick Odoyo

Master of Science

(Research Methods)

A dissertation submitted to the Department of Horticulture in the Faculty of Agriculture in partial fulfillment of the requirements for the award of the degree of Masters of Science in Research Methods of the Jomo Kenyatta University of Agriculture and Technology

2016

DECLARATION

This dissertation is my original work and has not been submitted to any other University for examination.

Signature ……………………………………. Date …………………………..

Ogeya Fredrick Odoyo   **Registration Number:**   **AG332-1930/2014**

DECLARATION BY SUPERVISORS

This dissertation has been submitted for examination with our approval as Supervisors:

Signature ………………………………………… Date …………………………

Prof. Losenge Turoop, Department of Horticulture, Jomo Kenyatta University of Agriculture and Technology

Signature ………………………………………… Date …………………………

Dr. Washingon Otieno, Plantwise Programme Executive, CAB International

CABI, Kenya

Signature ………………………………………… Date …………………………

Dr. Joseph Kyalo Mung'atu, Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology

DEDICATION

To all who inspired me to treasure education.

ACKNOWLEDGEMENTS

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AC   Area of one plant of tomato

AIC   Akaike Information Criterion

ANOVA   Analysis of Variance

BW   Bacterial Wilt

CL   Crop loss

Exp   Exponential

GLM   Generalized Linear Model

HL   Hosmer and Lemshow

IC   Infected crops

LA   Available land in acres for crop

LL   Log-likelihood

ML   Maximum Likelihood

OLS   Ordinary Least Square

PDF   Probability Density Function

POMS   Plant wise Online Management System

TC   Total crops in the farm,

ABSTRACT

Tomato is the second leading crop in Kenya in terms of production and value after potato. It is widely used as vegetable across the world. However, tomato varieties are attacked by bacterial wilt which devastates farmers. The bacterium is soil borne and persists in contaminated soils. It can be vector transmitted and has wide host's range of over 50 plant species making it difficult to control. Bacterial wilt attack on tomato farm results to losses of more than 90%. Its manifestation changes with varying conditions and farm management practices. Bacterial wilt infects roots and stems of many plants that are considered alternative hosts. This enables it to continue spreading across the tomato growing regions. The study used logistic model to determine the distribution, of bacterial wilt.

The study used secondary data obtained from the plant clinics on reported cases of bacterial wilt. In this study, disease incidents and distribution were inferred from cases presented by farmers to plant clinics. The response variables used were presence of bacterial wilt and estimated crop loss in the farm. The explanatory variables used in the model were weather data (minimum daily temperature, maximum daily temperature, minimum relative humidity, maximum relative humidity and precipitation), development stage of tomato and agro ecological zones (AEZ). Data from the year 2012 to 2013 obtained from Plantwise Online Management System (POMS) on tomato crop was used. R-Statistical Programming Package was used for the logistic analysis.

The results showed that relative humidity and minimum temperature significantly influenced bacterial wilt incidence and distribution as inferred from cases presented by farmer to plant clinics. In the analysis, agro-ecological zones, LH2, LM3, LM4, LM7, UM2 and UM4 significantly influence tomato losses in the farm due to attack by bacterial wilt. In the AEZ, the coefficient estimates are positive showing an increase in the bacterial wilt incidents. The disease incidents as presented cases brought to clinics by farmers varied in each county. Kirinyaga showed the highest incidents of the disease followed by Nakuru and Embu at 20.6%, 20.2% and 19.1% respectively. Bacterial wilt was found to be present in all the counties irrespective of difference in AEZ.

CHAPTER 1

INTRODUCTION

1.1 Overview and purpose of the study

The study aimed at using binary logistic model tool for estimating the bacterial wilt incidence and distribution among the tomato farmers in Kenya. In this study the disease incidence and distribution as presented was inferred from cases of crop anomalies brought to plant clinics by farmers. The study borrowed binary logistic application in health sector, education, social science and mathematics to enrich this study of determining best fit model for the distribution and crop loss associated with Bacterial wilt of tomatoes. In this study, I reviewed the state of bacterial wilt incidence and distribution. The problem of analyzing binary data collected in surveys and the need of logistic model is also presented. In chapter two, previous studies that applied binary logistic models and techniques used in parameter estimation are reviewed. Limitations of ordinary least square approaches in data analysis in the context of binary response data are also discussed. Chapter 3 gives details on the source of data used in this study. Logistic model results and discussions are presented in chapter 4. In chapter 5, conclusions on the application of binary logistic models and recommendations are discussed.

1.2 Background information of the study

Globally, the trends in food production has been on a decline due to the incidences of pest and diseases which are causing around 800 million people to lack enough food and at least 10% of food produced is lost to diseases (Strange & Scott, 2005).

Incidence of pests and diseases have contributed significantly to the reduction in food production in most agricultural regions (Oerke, 2006) and diseases are major problems for small scale farmers in the production of tomatoes and other crops (Yadessa, Bruggen, & Ocho, 2010). Tomato is the second leading crop in Kenya in terms of production and value after potato (Geoffrey, Hillary, Antony, Mariam, & Mary, 2014). It is widely used as vegetable across the world (Wani, 2011).

In Kenya, tomato is majorly grown in the open field, but use of greenhouses has been adopted in the recent past in most regions (Geoffrey et al., 2014). The protected production in greenhouses is essential for continuous production of tomato even during the adverse weather condition, though it can

act as optimal condition for rapid multiplication of many pathogens (Buschermohle & Grandle, 2012). The incidences of pest and diseases in greenhouse production are a major problem in tomato growing counties in Kenya viz Kiambu, Kajiado, Laikipia and Kirinyaga, (KARI, 2005).

In most cases poor practices in the farm such as hygiene, irrigation methods, continuous farming and cultural practices lead to increase in the incidence of pests and diseases in the farms (Abawi & Widmer, 2000). Among the diseases, bacteria wilt is a disease of economic importance in tomato production in the tropics (Lebeau et al., 2011) and a major constraint to tomato production (Hayward, 1991).

Bacterial wilt caused by *Ralstonia solanacearum* formerly known as *Pseudomonas solanacearum* affects all varieties of tomato (Afroz et al., 2009). The bacterium is soil borne and persists in contaminated soils and can be vector transmitted (Wani, 2011). It has a wide host's range of over 50 plant species (Maji & Chakrabartty, 2014). The pathogen is difficult to control and devastates farmers as it can cause losses of more than 90% when it infests tomatoes (Ajanga, 1987). Though it is persistent and resistant to most available control measures, its manifestation changes with varying conditions and management practices (Prior, Bart, Leclercq, Darrasse, & Anais, 1996). The use of disease free seedling is reported to delay disease onset and subsequent severity (Miller & Crosier, 2015). According to Chen *et al*. (2009), inducing resistance through gene silencing is reported to control bacterial wilt in tomato. Nutritional soil amendment with silicon is also known to control bacterial wilt in tomato (Ayana et al., 2011). However, continuous use of silicon in disease management will results into low soil pH which negatively affect crop production (Ogbodo, 2013). Despite the cultural management practices such as use of organic matter, control of bacterial wilt is still a challenge to farmers in the tropics.

Ralstonia Solanacearum can infect roots and stems of many plants which are considered as alternative hosts (Wenneker et al., 1999). This enables it to continue spreading across the tomato growing regions. The interaction of various environmental factors such as weather (temperature and rainfall), soil type and tillage practices determine the infection, distribution and survival of the pathogen. In order to understand the contributions and interactions of these factors the use of models is important.

Currently, the field of plant pathology embraces the use of statistical models to estimate the relationship between disease components to a number of environmental farm practices and host factors (Contreras-Medina, 2009). Various models are applied in understanding disease dynamics in crops. These models are useful for predicting diseases progression, development, distribution and epidemics (Calonnec, Cartolaro, & Chadoeuf, 2009). Linear regression models, multiple regression models and non-linear regression are commonly used in understanding disease epidemics.

Linear models are used in continuous data in sectors such as agriculture and social sciences (Wagner, 2013). They help in studying the relationship between the dependent variable and independent variable(s) (Wagner, 2013). In situations where the response variable is binary, it is extended to generalized linear model (GLM) which is efficient for nonlinear covariates (Hastie & Tibshirani, 1990). Binary response variable does not give a direct interpretation of the parameters as in the case of linear model since in binary the response estimate lies between 0 and 1. When linear regression model is used in a dichotomous (binary) dependent variable, linearity assumptions: homoscedasticity, zero mean and independent error terms are violated (Poole & O'Farrell, 1970), therefore logistic model is useful for the binary response (Haberman & Sinharay, 2010).

Logistic regression models were proposed for the first time in 1970s as the alternative technique used to model categorical response variable against other explanatory variables (Peng & Harry, 2002). The technique has been widely used in the fields of health sector in disease epidemiology, education and mathematics. Logistic model is efficient and gives various ways of coefficient interpretation and parameter estimation which are more intuitive (Babtain, 2015) and its efficiency in estimating probabilities, odds and odds ratios (Hosmer & Lemeshow, 1997). Since this study has a dichotomous response variable it suffices as the probability estimates lies between 0 and 1.

In an experimental design study, most cases the treatments are unbalanced and this results into inaccurate findings. However, logistic models are widely applied in unbalanced data and accurate estimates are obtained (Cnaan, Laird, & Slasor, 1997). However, application of binary logistic model has not been used to study distribution of diseases in various regions presumably due to difficulties in obtaining adequate data (Mila, Carriquiry, & Yang, 2004).

Understanding the distribution of bacterial wilt (BW) and crop loss associated with it helps in establishing appropriate measures and methods of managing the disease. Binary logistic model is the best model for estimating diseases occurrence and distribution while the crop loss is modeled using Poisson distribution. This type of model is useful in count data (Cameron & Trivedi, 2003).

Persistence of bacterial wilt of tomato and losses incurred as reported by farmers at the plant clinics led to a necessity to study the bacterial wilt distribution and incidences from the inferred cases in selected counties where clinics are operating. The study used logistic model to identify the distribution of bacterial wilt from inferred cases as reported by farmers at clinics. The purpose of this study is therefore to (i) determine the distribution of bacterial wilt of tomato in thirteen counties in Kenya, (ii) model the distribution of bacterial wilt under different agro-ecological zones and (iii) assess the crop losses attributed to bacterial wilt.

1.3 Statement of the problem

Tomato is the second leading vegetable crop in Kenya in terms of production and value after potato (Geoffrey et al., 2014). It is widely used as vegetable across the world (Wani, 2011). Tomato is attacked by bacterial wilt that is a problem in most regions. The persistence of bacterial wilt in the farms is due to poor practices in the farm such as hygiene, irrigation methods, continuous farming and cultural practices (Abawi & Widmer, 2000). Bacterial wilt lowers yield and quality of tomatoes (Oerke, 2006). It is a disease of economic importance in tomato production in the tropics (Lebeau et al., 2011) and a major constraint to tomato production (Hayward, 1991).

In Kenya, tomato is majorly grown in the open field which exposes tomato to attack by bacterial wilt, but use of greenhouses has been adopted in the recent past in most regions (Geoffrey et al., 2014). However, incidence of bacterial wilt in greenhouse production is still a problem in tomato growing counties in Kenya viz Kiambu, Kajiado, Laikipia and Kirinyaga, (KARI, 2005).

Bacterial wilt persistence in the soil and its wide range of alternative hosts facilitate its spread to regions which were not initially affected (Kelman, 1998). To determine the distribution, application of logistic model is effective. The study has a dummy response variable of presence of bacterial wilt as reported by farmers at the clinic 1= bacterial wilt present and 0 = No bacterial wilt.

Policy makers need the information on factors influencing bacterial wilt distribution for quick combat of the disease to help farmers produce quality tomatoes. The results obtained are important to farmers for effective planning to increase tomato production and yield. They would also make informed decision on correct methods of hygienic farming to control bacterial wilt and for more tomato production.

1.4 Objectives of the Study

1.4.1 General Objective

The overall objective of the study was to apply binary logistic model in determining the disease distribution and crop loss caused by bacterial wilt on tomatoes in selected counties in Kenya.

1.4.2 Specific Objectives

The specific objectives of the study were:

1. To determine the distribution of bacterial wilt of tomato in thirteen Counties in Kenya using descriptive statistics.
2. To model the distribution of bacterial wilt using binary logistic models under different ecological zones.
3. To assess the crop losses attributed to bacterial wilt using the logistic models.

1.5 Research question

The research questions which the study answered were:

1. Is the pattern of the distribution of bacterial wilt of tomato the same under different ecological zones?
2. Is bacterial wilt a major contributor to tomato crop loss under different agro-ecological zones?
3. Is the logistic model effective in explaining the distribution of bacterial wilt and mapping the crop losses associated with the disease?

1.6 Justification of the study

Bacterial wilt is a diseases of particular concern in tomato growing areas, as they reduce yield and quality of tomatoes (Oerke, 2006). Bacterial wilt is a disease of economic importance in tomato production in the tropics (Lebeau et al., 2011) and a major constrain to farmers (Hayward, 1991).

Bacterial wilt caused by *Ralstonia solanacearum* affects all varieties of tomato (Afroz et al., 2009). The bacterium is soil borne and persists in contaminated soils and can be vector transmitted (Wani, 2011). It has a wide host's range of over 50 plant species (Maji & Chakrabartty, 2014). The pathogen is difficult to control and it can cause losses of more than 90% when it infests tomatoes (Ajanga, 1987). Though it is persistent and resistant to most available control measures, its manifestation also changes with varying conditions and management practices (Prior et al., 1996). The use of disease free seedling is reported to delay disease onset and subsequent severity (Miller & Crosier, 2015).

Bacterial wilt can infect roots and stems of many plants which are considered as alternative hosts (Wenneker et al., 1999). This enables it to continue spreading across the tomato growing regions. The interaction of various environmental factors such as weather (temperature and rainfall), AEZ and development stage of tomato determine the infection, distribution and survival of bacterial wilt. To understand the contributions and interactions of these factors the use of logistic model becomes important. Logistic model was therefore used to study the distribution due to the dummy nature of response variable of 0 and 1 for presence of bacterial wilt and absent of bacterial wilt respectively (Mhamad, 2011).

1.7 Scope of the study

Bacterial wilt has been a devastating disease to farmers due to its persistence in the affected regions. In Kenya, most farmers grow tomatoes and bacterial wilt attacks them in the farm lowering the yield and quality. The study used secondary data on reported cases by farmers to the plant clinics to determine distribution and crop loss associated with bacterial wilt of tomatoes in Kenya. Plant clinics are set in thirteen counties in Kenya viz. Nakuru, Trans-Nzoia, Embu, Machakos, Nyeri, Kirinyaga, Kajiado, Kiambu, Bungoma, Marakwet, Narok, Tharakanithi and west Pokot. These counties are under different agro-ecological zones and different climatic conditions. The study used data set from 1980 tomato farmers for the year 2012 and 2013.

1.8 Assumptions made in the study

It was assumed that all wilted tomatoes reported or brought to the plant clinics were as result of bacterial wilt. Irrigation, farm management practices, source of tomato seeds were held constant across the ecological zones under the area of study

CHAPTER 2
LITERATURE REVIEW

2.1 Tomato production

Vegetables are one of the key income generating crops to the economy in developing countries, they are recognized for their nutritional values and income generation for small scale farmers (Lenne' & Spence, 2005). Tomato is a vegetable crop which is classified as one of the most important crop for generating income and building the economy (Domis et al., 2002). In Kenya vegetables farming have proved effective and shown huge potential in terms of growth rate and demand leading to growth of economy and creation of job opportunity to small scale farmers and the locals (Philip & Jaffee, 2004)

In Kenya horticulture has become effective and is steadily improving and among the crop exports, horticulture has accounted for two-thirds of all the exports (Philip et al., 2004). Tomato is one of the major horticulture crop grown in the country in both open and closed fields, though use of protected fields was adopted in the recent past (Geoffrey et al., 2014). According to Geoffrey et al., (2014), tomato is the second leading crop in Kenya in terms of production and value after potato. Tomato is financially attractive vegetable crop to both the small scale farmers in rural and peri-urban dwellings (Singh, & Regmi, 2013).

2.2 Constraints to tomato production

Key challenges faced by farmers in tomato production is pest and diseases (Lange & Bronson, 1981) as well as marketing (KHCP, 2011). Incidence of pests and diseases have contributed significantly to the reduction in food production in most agricultural regions (Oerke, 2006) and diseases is a major problem for small scale farmers in the production of tomatoes and other crops (Yadessa, Bruggen, & Ocho, 2010).

Diseases remain the biggest challenge to the tomato growing farmers globally and at least 10% of food produced is lost to diseases (Strange & Scott, 2005). Tomato is grown both outdoors and under glasses for both fresh market consumption and processing. It requires protection from a variety of pests including pathogens, weeds and diseases. Among the diseases, bacterial wilt has been a devastating disease to farmers and it affects various varieties of tomato (Afroz et al., 2009). The bacterium is soil borne and persists in contaminated soils for a long time and can be vector transmitted (Wani,

2011).Tomatoes offer a good condition for the stay of the bacterium pathogen and whenever tomato is grown it acts as a host for bacterial wilt pathogen since it offers shelter, food and production site for multiplication (Lange & Bronson, 1981).

2.3 Control of bacterial wilt

Though it is persistent and resistant to most available control measures, its manifestation changes with varying conditions and management practices (Prior et al., 1996). Attempts have been done in controlling the disease. The use of disease free seedling is reported to delay disease onset and subsequent severity (Miller & Crosier, 2015). According to Chen *et al*. (2009), inducing resistance through gene silencing is reported to control bacterial wilt in tomato. Nutritional soil amendment with silicon is also known to control bacterial wilt in tomato (Ayana et al., 2011). However, continuous use of silicon in disease management results into low soil pH which negatively affect crop production (Ogbodo, 2013).

Good intercropping of tomato with non host crops and a pre-planting soil amendment with urea was observed to reduce the disease incidents (Michel & Hartman, 1997). Despite the cultural management practices such as use of organic matter, control of bacterial wilt is still a challenge to farmers in the tropics.

2.4 Overview of the Logistic models

Binary logistic is an extension of linear regression due to its qualitative dummy covariate as response variable (Agresti, 2002). It measures the relationship between the outcome of one response variable and one or many explanatory variable. The response variable must be a dummy variable with a probability of success coded one and probability of failure coded zero. This will enable for investigation of the relationship between the response and explanatory variables. Logistic regression analysis is the technique of fitting a model to give the probability and odds ratio of an outcome on the binary response variable. It is more useful when using a binary response rather than just fitting a value of the response and the explanatory variables.

Logistic models is widely used in various areas to analyze most non-normal distributions. It is efficient in modeling dichotomous response variable and it was proposed in the 1970s to be used instead of Ordinary Least Square (OLS) to overcome the limitation of OLS (Peng & Harry, 2002). The statistical

problem of using the OLS to fit data that has non-normal distribution response variables violates the assumptions of linear regression models such as heteroscedasticity (Constant variance of error term), normality and linearity which always arise due to the nature of binary response which is either success or failure (Williams, 2014). The general linear regression model equation, which is the universal set containing simple regression and multiple regression as complementary subsets (Nau, 2014), is represented as;

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \varepsilon_i \qquad \qquad \varepsilon_i \sim N(0, \sigma^2) \qquad\qquad\qquad (2.1)$$

where Y is the response variable; X_i, are explanatory variables i=1, 2, 3,..., k; β_0 and β_i are the regression coefficients, representing the parameters of the model for a specific population; and ε_i is a stochastic term which is interpreted as resulting from the effect of unspecified explanatory variables or a totally random element in the relationship specified. Equation 2.1 expresses the form in which other distributions are based when link transformation is applied on the exponential families (Peng & Harry, 2002).

2.5 Binary data and the use of linear regression

The binary response variable is one in which the response outcome is either success or failure, and always coded as a dummy variable (success=1 and failure =0). This shows that the response variable has a probability between zero and one (Berry et al., 2015). However, use of least square regression model to fit the data of binary response variable is confronted by two major problems; conceptual in nature and statistical in nature.

2.5.1 Conceptual problem

In the process of fitting binary data in OLS regression, probability estimate which is above the maximum limit (one) or below minimum limit (zero) will be obtained, though according to the definition of probability, it should not be below zero or above one (Dixon & Koehler, 2001). Fitting binary data in a scatter plot results in two parallel lines on the ceiling (maximum limit) and on the bottom (minimum limit). The ceiling has a value of one and at the bottom a zero value. Fitting a straight line in this data will give a predicted value of response variable which may go below zero and

above one as shown in figure 1 (Koehler, 2001). The line fitted is either positive or negative depending on the coefficient of the covariates measured.

SLR of Joint Damage vs. Temperature

$$Y = 2.90476 - 3.74E\text{-}02X$$
$$R\text{-}Sq = 31.4 \%$$
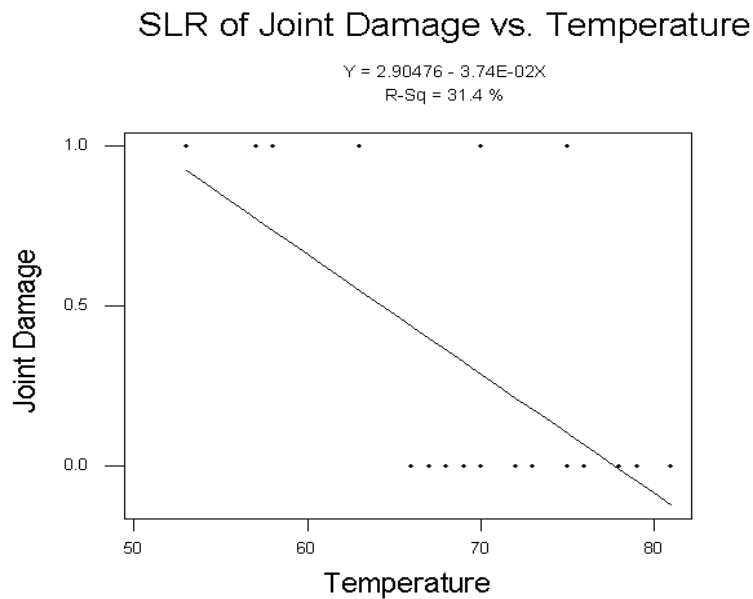
Figure 1: Fitted line plot (Source:(Dixon & Koehler, 2001).

However, to avoid the problem in linear line, non-linear curve is appropriate in binary response variable to ensure that the curve is not beyond one and zero resulting into a S-curve (Traditional & Methods, 2001; Mhamad, 2011) and the non-linear relationship will be the same as the S-curve as shown in Figure 2. This curve is positive if the coefficients are greater than zero and negative when they are less than zero.
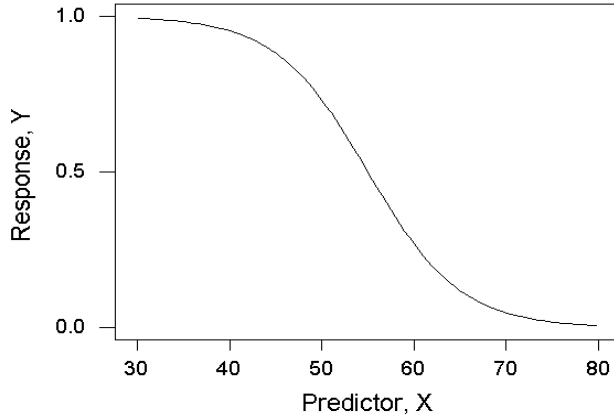
Figure 2: Representation of relationship between variables by logistic curve (Koehler K., Meeker, 2001)

The logistic regression above depends on the expression E.q 2.2, instead of the ordinary least square regression model that assumes for linearity. To develop the strict binary regression model, we begin by setting the notation used to describe the model. In the case of binary response and observe *n* independent pairs of (x_i, y_i), i= 1, 2,…, n, $X_i' = (x_{0i}, x_{1i}, …, x_{ki})$, $x_{0i} = 1$, denotes a vector K+1 for the assumed fixed covariates for the ith subject and $y_i = 0,1$ denotes an observation of the outcome for the response random variable Y_i. When using the logistic model we assume that p($Y_i = 1| X_i$) = $\pi(X_i)$, where, $\pi(X_i)$ = exp(g(X_i)) / (1+ exp(g(X_i))) and g(X_i) = $X_i'\beta$. Therefore, the parameter estimates are obtained by maximum likelihood (ML) and are denoted by $\hat{\beta}' = (\hat{\beta}_0, \hat{\beta}_{1,}, \hat{\beta}_k)$.

$$\Pr (Y_i=1 \mid X_i) = \frac{e^{g(Xi)}}{1+ e^{g(Xi)}} \qquad\qquad 2.2$$

2.5.2 Probabilistic problem

The statistical problem of using the least squares regression analysis on a binary response variable value, is violation of assumptions of linear regression models such as; heteroscedasticity (Constant variance of error term), normality and linearity. These problems always arise due to the nature of binary response which only take value one or zero (Williams, 2014). In the case of bacterial wilt attack, it is either presence of bacterial wilt in a sample or no bacterial wilt

In a binary variable, only two *Y* values and only two residuals exist for any single *X* value. For any value X_i the predicted probability equals to $b_o+b_1X_1$ and e_i is the residual term which is obtained when Y=1 or Y=0. Therefore, the residuals take the value of:

$e_i = 1- (b_o-b_1X_1)$ 2.3a When Y_i is equal to one

$e_i = 0- (b_o-b_1X_1)$ 2.3b When Y_i is equal to zero

These two equations 2.3a and 2.3b suggest the distribution has two values and the error term will never be normal at each level of the explanatory variable. According to Razali & Wah (2011), Kolmogorov-Smirnov (KS) statistic is used to test for normality of data of some ordered *n* and data points, $x_1<x_2...<x_n$ and shows significant difference when the statistic value is bigger.

Shapiro- Wilk (S-W) is also used to test for normality of data. It is not based on a graphical test but it is still most preferred to test for normality of data, assuming that the data is independent and identically distributed observations and is arranged in ascending order (Razali & Wah, 2011).

2.6 Some studies which applied logistic models

A study by Mila et al.(2004), used logistic models in agricultural sectors to investigate prevalence of soybean *sclerotinia* stem rot in the North-central region of United States. The factors used in the model were weather condition, tillage practices, region and soil type. The study used logistic regression model to investigate factors associated with the disease prevalence. The study also used Poisson model to estimate the incidence of disease on plants using the same explanatory variables.

According to Williams (2014), a comparison study on ordinary least square model and logistic model on qualitative response variable suggested that ordinary least square model is not appropriate. The main purpose of the study was to obtain the model suitable for the binary response outcome. He used both OLS and logistic model and concluded that logistic regression model is the best for the categorical response.

In one instance, (Calonnec et al., 2009), conducted a study on development of powdery mildew epidemics on vines. Their main aim was to identify features of spatiotemporal spread of powdery mildew in the fields. Logistic regression technique was used to highlight the features of the disease. The model developed managed to predict changes in powdery mildew over time.

2.7 Over dispersion and under dispersion of model parameters

Over dispersion and under dispersion is essential in model fitting as it ensures that the parameters fitted are of importance and do not show correlation. Over dispersion in logistic models majorly works in the principle of depending on the mean and variance of the observed variables (Rodriguez, 2013). Wrong measurement or omission of important covariates in the model gives a positive correlation and results into a wrong description of probability of success (Follmann & Lambert, 1989), but can be corrected by introducing the random effect in the model. The variation maybe introduced through geographical regions or individuals in a continuous response variable, though this may not be easy (Browne, et. al., 2003). According to Rodriguez (2013) under dispersion occurs when the variation is less than the expected variables. Various ways of testing over dispersion such as Pearson statistic is essential in modeling to avoid fitting a wrong model (Dean, 1992).

2.8 Over parameterization of the logistic model

Parameterization is a process of deciding or setting the parameters necessary for a complete or relevant specification on model of interest. According to Whittaker et al., (2010), hydrological models always uses more parameters causing over parameterization and poor prediction of the model through use of Soil and Water Assessment Tool (SWAT). However, when few covariates were fitted, the model prediction was accurate. Use of conditional distribution in an unbalanced data and small sample size may give unreliable results due to inclusion of many parameters in the model (Forster et al., 1995). Over parameterization is a problem in most studies and it results into unrealistic results on the model predictions. Therefore, only parameters of interest should be incorporated in the model to obtain a best fit model (Whittaker et al., 2010).

2.9 Orthogonal effect of data

In most cases, data is analyzed without confirming its orthogonal state. Before fitting any model or carrying out any analysis on non-orthogonal data, we have to appreciate the difference on the two set of data and kind of output we may get. Non-orthogonal data is complex to interpret, though under sampling or over sampling is carried out to ensure that the data is orthogonal (Kotsiantis, Kanellopoulos, & Pintelas, 2006). Though, there are controversial issues on handling unbalanced data (Kotsiantis et al., 2006), general linear mixed models is still efficient to analyze the unbalanced data and longitudinal data (Cnaan et al., 1997).

2.10 Diagnostics of the fit of model

In model development, diagnostic tests are important. This helps in choosing the best model and use appropriate model for a given study. The diagnostic test helps in detecting various aspect of the data variables to be investigated such as outlier response, extreme points and quantifying their effect on the model (Pregibon, 1981). Model diagnostics is used in various aspects including the residuals and deviance.

2.10.1 Residuals and Deviance

Effect of residuals in the model is important in discriminating models and selecting a suitable model with least deviance value (Landwehr, Pregibon, & Shoemaker, 1984). Residuals is measured by either using deviance or partial residuals. Deviance calculation is important in obtaining the best model by considering the model with least deviance value (Landwehr et al., 1984). The model building was obtained by running the iteration to select the best model. The model with the least Deviance and AIC is then refitted. This is done through backward processing to eliminate variables that add no impact in the model.

2.11 Logistic Regression Transformation

2.11.1 Logarithmic transformation

Logarithmic function are important in improving the fit of the logistic regression model for binary data by transforming the explanatory variables and the methods are based on consideration of the distributions of these variables and the outcome group (Kay, 1987). The transformations required are the functions of the explanatory variables which appear as the log distributions. The transformation of the data helps in solving the problem of heteroscedasticity, skewness; enable simplifying the model and converting multiplicative model to additive model for efficient interpretation of the output.

2.11.2 The Logit (Logged Odds)

Odds ratio is the probability value of an event occurring (π_i) and event not occurring ($1- \pi_i$). When the odd is less than one, then the event is less likely to occur than it is not to occur. When the odds is greater than one then there is high likely that the event occur than it is not to occur. According to Rodriguez (2013) the odds ratio is:

$$\text{Odds} = \frac{\pi_i}{1 - \pi_i} \qquad\qquad\qquad\qquad\qquad\qquad\qquad 2.4$$

Since : $\qquad\qquad 0 \leq \pi \leq 1$

Then: $\qquad\qquad 0 \leq \text{Odds} \leq \infty$

where p_i is the probability of success and $1-\pi_i$ is the probability of failure.

Therefore transformation of odds using log-odds remove the restriction on the probability and model is transformed as a linear function of the covariates of the explanatory variables (Rodriguez, 2013). Thus;

$$\text{logit}(\pi) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right)$$

Based on this, the relationship between logit (log odds) and explanatory variable is a linear relationship.

$$\text{log odds} = \ln(\pi) = b_o + \sum_{i=1}^{n} b_i x_i \qquad\qquad\qquad\qquad 2.5$$

2.12 Interpretation of logistic regression parameters

The impact of explanatory variables on the response presence of bacterial wilt has several interpretations. It can be interpreted in terms of logged odds (logit), odds and Odds ratio.

2.12.1 Interpretation in terms of logged odds (logit)

Logit uses parameters obtained from logistic regression model. It shows the change in the predicted logged odds for a unit change in the explanatory variables used in the model. The method is similar to the interpretation linear regression parameters, though in logistic regression, units change in explanatory variable is explained by the logged odds of the explanatory variable to obtain the resulting impact on the response variable.

2.12.2 Interpretation in terms of odds

This interpretation is useful by transforming logistic regression parameter so that the explanatory affects the odds and not the logged odds. It is achieved by taking exponential on the logit to enable explanatory variables have an effect on the odds as follows:

$$\ln(\frac{\pi_i}{1-\pi_i}) = \sum_{i=1}^{k} \beta_i x_i$$

$$\therefore \exp(\ln(\frac{\pi_i}{1-\pi_i})) = \exp(\sum_{i=1}^{k} \beta_i x_i)$$

$$\text{Odds} = (\frac{\pi_i}{1-\pi_i}) = \exp(\sum_{i=1}^{k} \beta_i x_i) \qquad\qquad 2.6$$

In the case of linear regression, when β_i's are negative then it shows a decrease in the response variables and a positive explains an increase. However, in the binary logistic models, exponential values for the parameter which exceeds one indicate an increase in Odds but values which are less than one or equal to zero explains that odds is decreasing.

2.12.3 Interpretation in terms of Odds Ratio

This is obtained through determination of the ratio of odds. It is not equivalent to the Odds since this is the ratio of probabilities (Glas, Lijmer, Prins, Bonsel, & Bossuyt, 2003). Odds ratio is obtained by comparing the relationship between a specific level of explanatory variable (X_i) and after adding one to the previous explanatory variable ($X_i +1$), this implies that $X_i^{'} = X_i +1$, it is from this where odds ratio will be computed. Odds ratio is obtained by the formula (2.6), It may also be obtained by obtaining $\exp(\beta)$. This helps in explaining the estimate of factors that affect the distribution of bacterial wilt and crop loss associated with bacterial wilt.

CHAPTER 3
MATERIALS AND METHODS

3.1 Study Area

The study was conducted using data from 57 plant clinics in 13 counties in Kenya viz. Nakuru, Trans-Nzoia, Embu, Machakos, Nyeri, Kirinyaga, Kajiado, Kiambu, Bungoma, Marakwet, Narok, Tharakanithi and west Pokot County. These counties are not delimited according to weather patterns. Agro-ecological zones (AEZ) were therefore used in considering weather aspects. Some AEZ overlap across counties and variations in the AEZs led to heterogeneity of environmental factors and farming activities.

3.2 Study design and data collection

The regions where the clinics are located were purposively selected according to the general plant health problems that farmers experience. Within the counties, sites for clinic location were randomly selected from a pool of several other possible alternatives. Different counties formed one level of interest and individual clinics were considered as another level.

The data collected at the clinics were both qualitative and quantitative viz. county, demographic information of farmers, type of crop, development stage of disease, parts of crop affected, area under crop, crop loss, distribution of disease in the farm, type of pest and disease, classification of diseases and methods of controlling diseases. The study used secondary data obtained from the plant clinics queries. The response variables used were presence of bacterial wilt and estimated crop loss. The explanatory variables used in the model were weather data, (minimum daily temperature, maximum daily temperature, minimum relative humidity, maximum relative humidity and precipitation), development stage of crops and agro ecological zones.

Data on the presence of bacterial wilt and crop loss were obtained from plant clinic records captured as farmers visit with diseased plant samples. The data from 2012 to 2013 used in the study was organized into various variables and filtered. The filtering entailed identification of entries that were not coded well and outlier in the data. The outliers were checked using the box plots and excluded in the analysis.

3.3 Distribution of bacterial wilt in the counties.

Disease distribution was determined in the thirteen counties according to plant clinic locations. Comparative analysis using cross-tabulation was employed using data from each county, taking into consideration that the counties have varied ecological zones. The disease incidents as inferred from clinic queries were compared across the counties and the statistical difference in the parameter tested was performed using Chi-square tests.

3.4 Modeling distribution of bacterial wilt

The study used binary logistic regression model with a dummy variable response outcome for either disease presence =1 or disease absence = 0 to investigate the distribution of bacterial wilt and other factors which influence disease distribution such as farming practices, AEZ, type of crop and weather. The Binary logistic model for the disease presence in the ith field was presented using generalized linear models (Koehler K., Meeker, 2001).

$$Y = \eta + \xi = \mathbf{X}' + e, \qquad\qquad\qquad 3.1$$

Eq. 3.1 has the following components:

i. a random component, denoted ξ

ii. a linear relationship between the dependent variable and its predictors. The estimate or predicted values of the prediction is denoted $\mathbf{\eta}$

iii. a link function captures the form of relationship between the dependent variable and predictors expected value. Log link ($\mathbf{\mu}$)= exp ($\mathbf{\eta}$) or $\mathbf{\eta}$= log ($\mathbf{\mu}$)

The impacts of predictor variables are multiplicative for k parameters. That is I = 1,2,3,…,k:

$$Y = \exp(b_0)\, \exp(b_1 X_1)\, \exp(b_2 X_2)\ldots \exp(b_k X_k)$$

where Y denotes the dependent variable, the presence of bacterial wilt in tomato sample taken to the plant clinic between the year 2012 to 2013. In the binary regression, the logit link function was used.

$$p(Y_{ij}=1|X_{ij}) = \pi_{ij} = \frac{Exp(\beta_{0j} + \beta_{1j}X_{ij})}{1 + Exp(\beta_{0j} + \beta_{1j}X_{ij})} \qquad 3.2$$

where i=1,2,...,N represent the sample size while j=1,2,...,k, represent the Independent variables while β's represent the parameters.

The exponent parameters (ξ) in equation (3.2) can be expressed in a matrix format as;

3.4.1 Selection of logistic models for distribution of bacterial wilt of tomato

The logistic models fitted for the bacterial wilt incidence using a log link function of exponential family, binomial (Appendix 2A, 2B and 2C). The binary logistic was fitted using the "glm" function in model 1, 2 and 3 respectively. They were developed using main effect parameters and the interactions of the main effects for model 1 and 2. Agro-ecological zones and weather parameters were correlated, possibly leading to multi-co-llinearity effect. Model 3 was developed using only the AEZ parameter and development stage of crops to investigate the dominance stage in the bacterial wilt incidence.

Figure 1: Selected three models using the "glm" function for binary logistic model on bacterial wilt

model1<-glm(PresenceBW ~ Minimum Temperature+ Maximum Temperature+ Precipitation +Minimum Relative Humidity + Maximum Relative Humidity+ Dev. Stage Seedling +Dev. Stage Intermediate+Dev. Stage Flowering+ Dev. Stage Fruiting +Dev. Stage Mature +Dev.Stage Post Harvest+ Minimum Temperature:Precipitation+ Maximum Temperature:Precipitation+ Minimum Relative Humidity: Precipitation+ Maximum Relative Humidity: Precipitation, data=BWdata, family="binomial")

summary(model1)

model2<glm(PresenceBW~MinTemp+MaxTemp+Precipitation+MinRelHumidity+MaxRelHumidity+ Precipitation:MinRelHumidity+Precipitation:MaxRelHumidity+Precipitation:MinTemp+Precipitation: MaxTemp,data=BWdata, family="binomial")

summary(model2)

model3<-glm(Presence BW~AEZ+ Dev.Stage Seedling + Dev.Stage Intermediate + Dev.Stage Flowering + DevStageFruiting + Dev.Stage Mature + Dev.Stage Post Harvest, data= BWdata, family="binomial")

summary(model3)

3.5 Crop losses Assessment

Poisson regression analysis was used to estimate the tomato crop loss. These were number of tomato plant in the farm that were reported by farmers to have been affected by bacterial wilt. Farmers assessed their farms and estimated tomato crop loss.

In fitting the poisson model for the estimated crop losses incurred. farmers estimated the crop loss due to bacterial wilt in the farm. Data was filtered for only tomato cases reported to have been attacked by the bacterial wilt. This reduced whole data set of tomato farmers from 1980 to a sample size of 267 as these were the reported cases of tomato crops prescribed by clinic doctors to be infested by bacterial wilt.

Crop loss was estimated using the standard unit of meter squared since farmers have different land sizes for growing tomato. In obtaining best model the variation in land size was extrapolated to one unit (M^2) for an accurate model estimate. Extrapolation of land size was effective in removing discrepancy due to variability of land size. The number of tomato plants infected as reported by farmers was recorded then a total number of tomato crops in the farm estimated. Total crop in square meters farm was calculated as in (Eq. 3.3). The proportion of crop loss was then calculated by obtaining the ratio of crops infected in the crop population and then extrapolating to a standard unit of one meter square. The tomato losses was estimated using AEZ and weather variables for each county to estimate the regions that were most affected.

$$CL = \left(\frac{\left(\left(\frac{IC}{TC} \right) * 100\% \right)}{LA} \right) \qquad \qquad 3.3$$

where

CL= crop loss

IC= Infected crops

TC= Total crops in the farm, but TC$= \dfrac{FarmM^2}{AC}$

AC= Area of one plant of tomato in meters square

LA = Available land in meters square for tomato crop

The study used poisson regression model to estimate crop loss in tomato. The response variable was crop loss and the independent variables were similar to the parameters used in determination of bacterial wilt disease distribution. Following Mila et al. (2004), Poisson model is written as:

$$p(Y=y \mid \mu) = \frac{\lambda^y e^{-\lambda}}{(y)!} \qquad \qquad 3.4$$

$$p(Y=y \mid \mu) = \frac{[\exp(\Sigma\beta X)]^y * \exp[-\exp(\Sigma\beta X)]}{y!},$$

where, $\lambda = \exp(\sum \beta X)$ this ensured that λ is non-negative integer because Poisson distribution is defined for only positive values (Mila et al., 2004). Hence, the estimated Poisson model is specified as:

Estimated tomato loss (y) = f(weather data, (minimum daily temperature, maximum daily temperature, minimum relative humidity, maximum relative humidity and precipitation), development stage of crops and agro ecological zones.)

3.6 Data and analysis

Data was first filtered to remove any outliers and wrong entry in each variable of data set. Box plot was used in checking skewness and outliers in the data. From the box plot, any data point outside the range of upper and lower whiskers was categorized as outliers. Skewness of variables was confirmed by establishing the middle value of box plot to divide the first quarter and upper quarter of box plot into two equal parts. The filtered data was analyzed using R Statistical Programming Package and Statistical Package for Social Scientist (SPSS, version 19). In the analysis both descriptive and inferential statistics were used. Descriptive statistics was used to estimate the percentages of farms attacked by bacterial wilt in each county while logistic models (inferential statistics) was used to estimate the bacterial wilt distribution and crop loss associated with bacterial wilt.

CHAPTER FOUR

RESEARCH FINDINGS

4.1 Distribution of Bacterial wilt in thirteen counties in Kenya

4.1.1 Proportion of bacterial wilt incidents in clinic queries by county

The results in this study showed that bacterial wilt is present in twelve counties. Among the thirteen counties where clinics are set, in Elgeyo Marakwet, farmers did not report any case of bacterial wilt of tomato in any clinic most probably due to other factors such as good farming practices or limited access to plant clinics. Bacterial wilt distribution a cross counties were significant (χ^2 = 202.079, p-value<0.001). Kirinyaga county had the highest bacterial wilt attack of reported cases of the disease followed by Nakuru and Embu at 20.6%, 20.2% and 19.1% respectively. In Narok, West Pokot, Tharaka-Nithi, Kajiado and Kiambu Counties, the cases of disease attack reported was 0.4%, 2.2%, 2.2%, 2.2% and 2.6% respectively as shown in (Fig. 1).
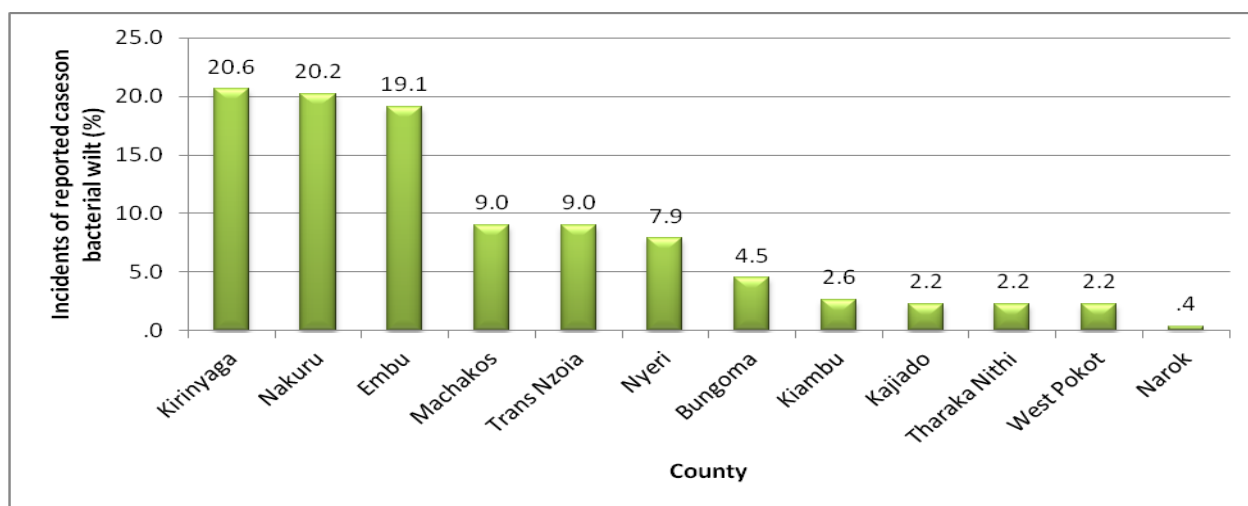
Figure 1: Reported cases of bacterial wilt of tomato at the plant clinics in thirteen counties in Kenya

Bacterial wilt cases on tomato were presented by farmers to plant clinics at various development stages. In the twelve counties, sample cases reported were showing signs of wilting presumably due to bacterial wilt infestation at fruiting and flowering stages. No reported cases of bacterial wilt at post harvest stage as shown in Table 1. During flowering and fruiting stage farmers pay much attention to tomatoes this is because at this stages of growth most wilting cases were reported to the clinic. No

24

samples were presented to the clinics at seedlings stage having bacterial wilt but reports on infestation at the intermediate stage indicate a high level of incidence. This is partly due to possible transplanting healthy seedlings from clean nurseries in already infested fields.

Table 1: Summarized data on reported cases of bacterial wilt presented relative to crop development stage

Development stage	No (%)	Yes (%)
Seedling	92.5%	7.5%
Intermediate	65.9%	34.1%
Flowering	64.0%	36.0%
Fruiting	59.6%	40.4%
Mature	83.1%	16.9%
Post Harvest	100.0%	-

4.2.1 Trend of cases of bacterial wilt reported to plant clinics

The trend in the incidence of cases reported by farmers showed that weather parameters used in the model influences the disease incidences. It was found that between January and March, temperature, precipitation and humidity was low but on a increasing trend (Fig.2). This showed that trend in the disease incidence was also observed to be higher in those months. However, between April and June the number of disease incidence reported was declining and increasing from June to November. December showed a decline and this may be attributed by the fact that during this month, farmers don't harvest tomato and rain is also reducing. Weather parameters have similar pattern with disease incidence. However, this could be enhanced if other farm practices that were not within the scope of this study would have been considered.
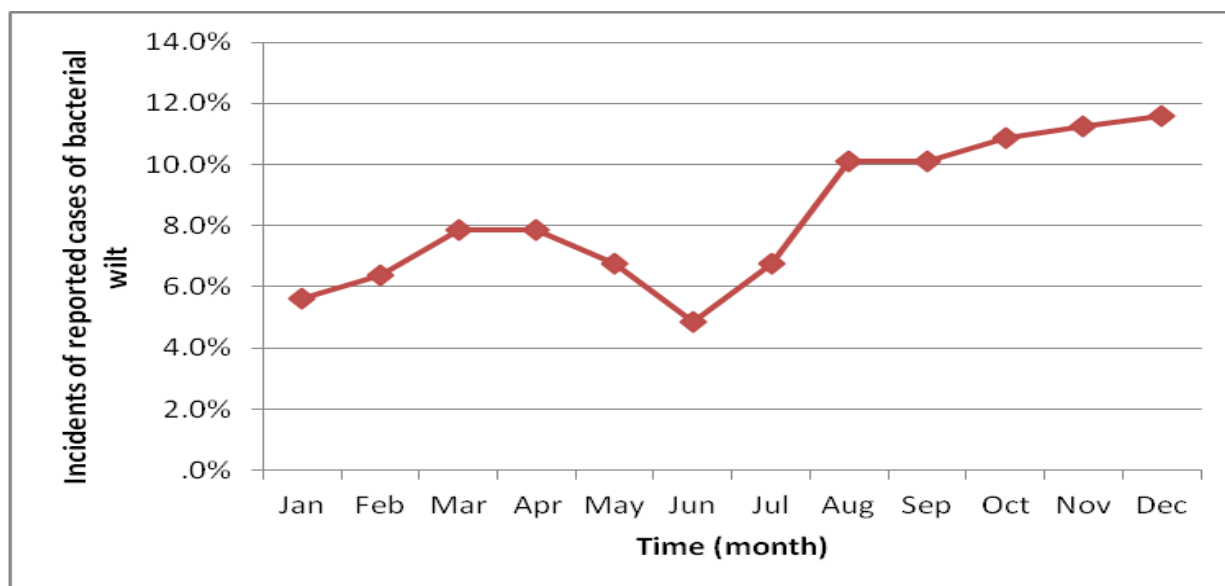
Figure 2: Bacterial wilt incident cases reported by farmers to the plant clinics over time

4.2 Modeling the distribution of bacterial wilt using the logistic models

4.2.1 Normality assumption and exploratory analysis for weather data from January to December

A range of temperature 27 ^{0}C to 32 ^{0}C is essential for optimum tomato production but this also influences the spread of *R. Solanacearum* (Mew, 1977). In the month of January to December the minimum temperatures (14 to 16 ^{o}C) was recorded which was slightly skewed (Fig. 3). However, the upper quartile range of temperature fluctuated over the months with narrow whiskers in May indicating low temperature variation.

The average monthly maximum temperature ranged from 24 ^{0}C to 28 ^{0}C, the maximum temperature increasing from January to March then falling until June, over the period of rainy season and then temperatures rose steadily from July to December.
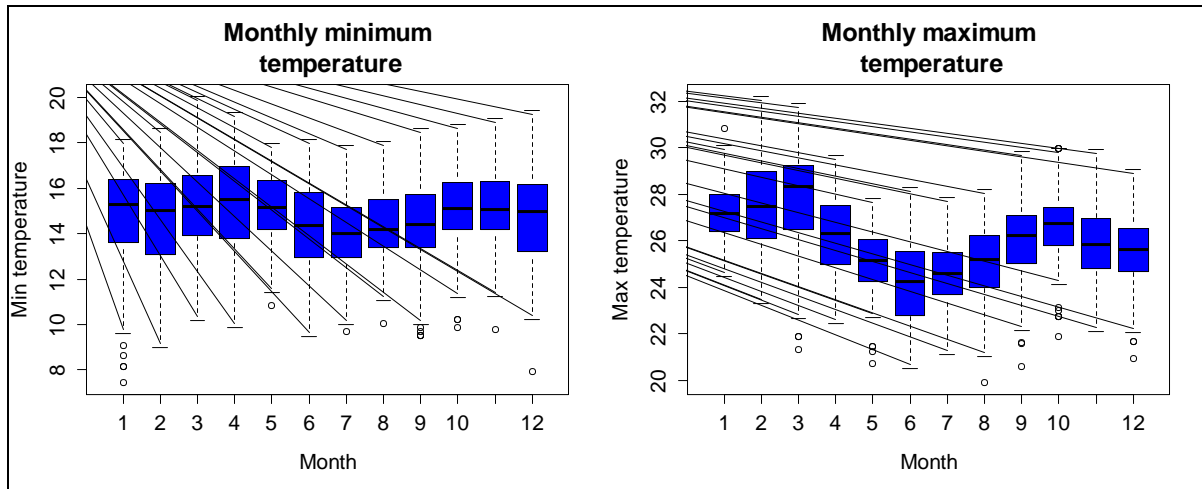
Figure 3:Paired box plot of monthly minimum and maximum temperatures (^{o}C) from January to December

Relative humidity was also a component of weather data used. The minimum monthly relative humidity ranges between 30% to 52%. For the month of January to March, minimum relative humidity was almost constant at 35% while during the month of April to June humidity increases steadily. However, between July to December it was falling. Monthly minimum relative humidity was normality distributed.

The monthly maximum relative humidity ranged between 90% and 98%, which showed less fluctuation as compared to the minimum monthly relative humidity. In the month of January to March, relative humidity was almost constant and this period rainfall is low. Then between April to May the humidity was increasing and falling between June to September and a slight increase for October to December. The distribution for the maximum relative humidity for the month of March, April and May was negatively skewed (Fig, 4).
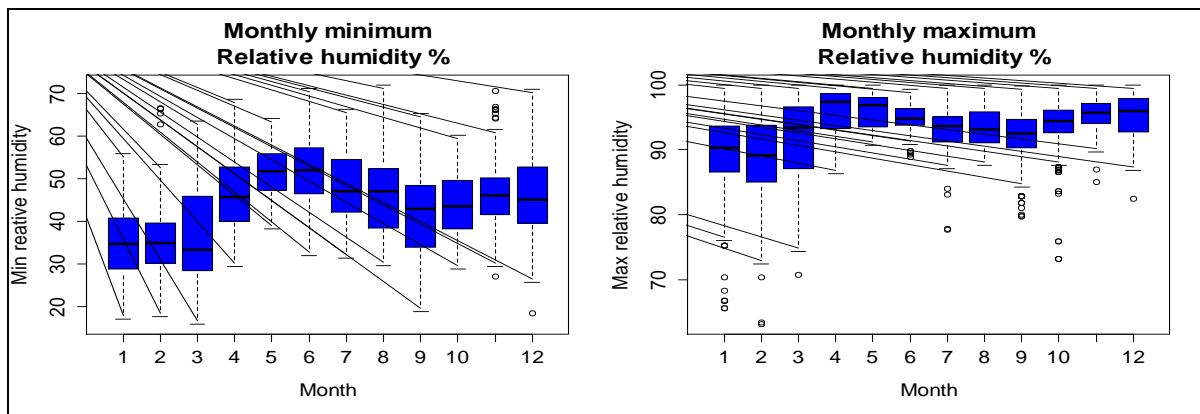
Figure 4: Paired box plot of monthly minimum and maximum relative humidity (%) from January to December

Monthly and daily precipitation was observed across the year. The average monthly precipitation was below 5mm and this would be attributed by diverse agro-ecological zones. In areas like Kirinyagathe annual rainfall is about 60mm. However, the daily precipitation showed a great fluctuation on the rainfall received in each day indicating that there were some days that there was no rainfall recorded. Majority of the days recorded rainfall below 20 mm (Fig. 5) Rainfall pattern in the selected regions was not normally distributed, this might have been caused by the change in the weather pattern in Kenya.
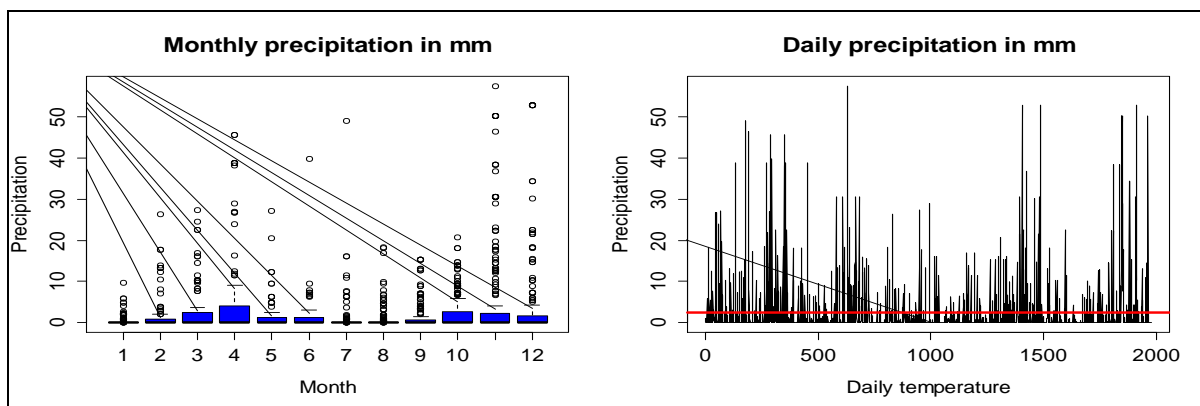
Figure 5: Paired box plot of monthly and daily precipitation (mm) from January to December

4.2.2 Results of model 1, 2 and 3 on z-statistic with appended likelihood ratio p-values

4.2.2.1 Model 1

In Table 2 below, the fitted logistic model 1 based on the selected parameters using "glm" function (Appendix 2A). The interaction terms of minimum temperature and precipitation; maximum temperature, minimum relative humidity and maximum relative humidity had no significant impact on the incidence of BW on tomato. The main effect on climatic factors also did not show any significant difference on bacterial wilt incidence as reported by farmers. However, variation in development stage showed an significant impact on bacterial wilt attack in tomato at various stages of development. The study also found out that some variables contributed negatively to the bacterial attack on tomato. For instance, minimum temperature had an estimate of -1.94E-02. This implies that as temperature decreases the attack of bacterial wilt in tomato will decline.

Table 2: Parameter estimates of the logistic regression model used to explain the incidence of bacterial wilt using weather parameters and development stage of infestation

	Estimate	Std. Error	z-value	pr(>|z|)
Intercept	-4.77E+00	2.19E+00	-2.175	0.02962
Minimum Temperature	-1.94E-02	5.80E-02	-0.334	0.73842
Maximum Temperature	5.98E-02	6.22E-02	0.962	0.33607
Precipitation	-4.73E-01	4.38E-01	-1.081	0.27962
Minimum Relative Humidity	2.17E-02	1.18E-02	1.835	0.06653
Maximum Relative Humidity	6.16E-03	1.66E-02	0.37	0.71111
Dev. Seedling	-4.00E-02	2.64E-01	-0.152	0.87948
Dev. Intermediate	4.41E-01	1.53E-01	2.871	0.00409
Dev. Flowering	3.41E-01	1.46E-01	2.343	0.01911
Dev. Fruiting	-8.07E-02	1.44E-01	-0.56	0.57555
Dev. Mature	-3.89E-01	1.83E-01	-2.132	0.033
Dev. Post Harvest	-1.25E+01	3.12E+02	-0.04	0.96799
Minimum Temp*Precipitation	1.57E-02	1.23E-02	1.275	0.20216
Maximum Temp*Precipitation	1.86E-04	1.08E-02	0.017	0.9863
Min Relative Humidity*Precipitation	-2.60E-04	1.77E-03	-0.147	0.88331
Max Relative Humidity*Precipitation	2.41E-03	4.20E-03	0.572	0.56698

[a] Significant codes: 0'***', 0.001'**', 0.01'*', 0.05'.', 0.1' ', 1

[b] Null deviance: 1566.2 on 1979 degrees of freedom; Residual deviance: 1520.9 on 1964 degrees of freedom; AIC: 1552.9; Number of Fisher Scoring iterations: 13

[c] Min Relative Humidity= Minimum Relative Humidity, Max Temperature= Maximum Temperature, Dev= Development stages of plant when disease incidents recorded (no, 0; yes, 1), Std. Error= Standard Error.

4.2.2.2 Model 2

In Table 3 the study found that minimum relative humidity was significant and had a positive coefficient estimate on bacterial wilt incidence (p= 0.0295). The intercept represent the base line value

of the response variable assuming no contribution of any explanatory variables. From the analysis it was found that the intercept was negative an indication of a decline on the bacterial wilt cases reported

Table 3: Parameter estimates of the logistic regression model used to explain the incidence of bacterial wilt by using weather parameters

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
Intercept	-5.844243	2.053974	-2.845	0.00444**
Min Temperature	0.009965	0.054479	0.183	0.85486
Max Temperature	0.080334	0.058211	1.38	0.16757
Precipitation	-0.001271	0.009585	-0.133	0.89454
Minimum Relative Humidity	0.024293	0.011166	2.176	0.02958*
Maximum Relative Humidity	0.007095	0.015756	0.45	0.65251

[a] Significant codes: 0'***', 0.001'**', 0.01'*', 0.05'.', 0.1' ', 1

[b] Null deviance: 1566.2 on 1979 degrees of freedom; Residual deviance: 1554.0 on 1974 degrees of freedom; AIC: 1566

4.2.3.3 Model 3

The findings in Table 4 presents the fitted model based on the selected binary logistic model on the z-statistics using "glm" function. The study showed that agro-ecological zones (AEZ) have no significant impact on bacterial wilt attack in tomato. This could be attributed by poor farming practices that most farmers adopt to reduce cost of production. Farmers always recycle seeds and continuously plant tomato in the same farm for many years. AEZ though not significant, a reduction in the disease cases reported by farmers. However, LM4, UM1 and UM3 had positive coefficient estimate indicating an increase in the BW attack on tomatoes with respect to LH1

Table 4: Parameter estimates of the logistic regression model used to explain the incidence of bacterial wilt in relation to agro-ecological zones and crop development stage

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-1.86369	0.63634	-2.929	0.0034**
AEZ LH2	-0.63965	0.74033	-0.864	0.38758
AEZ LH3	-0.76433	0.77983	-0.98	0.32702
AEZ LH4	-0.24353	1.23843	-0.197	0.84411
AEZ LM3	-0.16672	0.82976	-0.201	0.84075
AEZ LM4	0.28389	0.64871	0.438	0.66166
AEZ LM5	-0.51012	0.7186	-0.71	0.47778
AEZ LM6	-1.09581	0.77797	-1.409	0.15897
AEZ LM7	-0.65875	1.21909	-0.54	0.58895
AEZ UH1	-13.97764	0.91752	-0.019	0.98451
AEZ UH2	-1.38677	1.20158	-1.154	0.24845
AEZ UM1	0.3291	0.69903	0.471	0.63779
AEZ UM2	-0.13272	0.65456	-0.203	0.83932
AEZ UM3	0.04208	0.63655	0.066	0.94729
AEZ UM4	-0.28786	0.65021	-0.443	0.65797
AEZ UM5	-14.06598	0.95072	-0.024	0.98101
Dev. Seedling	0.02005	0.26245	0.076	0.93911
Dev. Intermediate	0.41146	0.15474	2.659	0.00783**
Dev. Flowering	0.33014	0.14509	2.275	0.02288*
Dev. Fruiting	-0.05577	0.14382	-0.388	0.69821
Dev. Mature	-0.35736	0.18223	-1.961	0.04988*
Dev. PostHarvest	-13.43874	1.43525	-0.026	0.97904

[a]Significant codes: 0'***', 0.001'**', 0.01'*', 0.05'.', 0.1' ', 1

[b]AEZ= Agro-ecological zone, LH= Lower upper land, LM= Lower middle land, UH= Upper highland, UM= Upper middle land, Dev= Development stages of plant when disease incidents recorded (no, 0; yes, 1).

4.2.3 Re-fitted model adjusted for over parameterization

Models in Tables 2, 3 and 4 above were over parameterized based on the z-values. The parameters were re-fitted considering the high AIC values and low value of residual deviance. The best-fit model was selected through the backward iteration process obtaining the model with the low residual deviance. Resulting model gives the model with parameters that have significant influence on the disease incidence. In the refitted model for model 1a, showed that the intercept, fruiting and maturity stage of tomato had a negative estimate. This is an implication of decline in the bacterial wilt attack on tomato (Table 5)

4.2.4.1 Refitting of models

Table 5: Model 1a, 2a and 3a output are refitted models after readjusting for over parameterization to improve estimation of bacterial wilt incidence

Model 1a				
	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-2.820777	0.338807	-8.326	< 2e-16***
Min Relative Humidity	0.017943	0.006915	2.595	0.00946**
Dev. Flowering	0.362961	0.141947	2.557	0.01056*
Dev. Fruiting	-0.043275	0.140261	-0.309	0.75768
Dev. Intermediate	0.474264	0.150113	3.159	0.00158**
Dev. Mature	-0.345616	0.178042	-1.941	0.05223.
Dev. Post Harvest	-12.568472	311.375236	-0.04	0.9678
Model 2a				
Intercept	-5.261978	1.263135	-4.166	3.1e-05 ***
Min Relative Humidity	0.026624	0.007814	3.407	0.000656 ***
MaxTempeature	0.08502	0.040427	2.103	0.035460 *
Model 3a				
				-
Intercept	-2.035	0.1074	18.944	< 2e-16 ***
Dev. Intermediate	0.5039	0.1454	3.464	0.000531 ***
Dev. Flowering	0.3426	0.1406	2.436	0.014847 *
Dev. Mature	-0.3531	0.1776	-1.988	0.046790 *
Dev. Post Harvest	-12.5594	310.6138	-0.04	0.967747

[a] Significant codes: 0'***', 0.001'**', 0.01'*', 0.05'.', 0.1' ', 1

[b] Min Relative Humidity= Minimum Relative Humidity, Max Temperature= Maximum Temperature, Dev= Development stages of plant when disease incidents recorded

[c] Model 1a, 2a, and 3a has different residuals deviance and AIC. They were developed from model 1, 2 and 3 through backward iteration process to take care of over parameterization. Model 1a has residuals deviance (1532.0) and AIC (1546), model 2a has residuals deviance (1554.2) and AIC (1560.2) and model 3a has a residuals deviance (1538.8) and AIC (1548.8).

4.3 Poisson Model for the assessment of reported crop loss due to bacterial wilt

4.3.1 Model 4

The parameters used in assessing tomato losses due to bacterial wilt were weather data (minimum temperature, maximum temperature, precipitation, minimum relative humidity and maximum relative humidity) and development stages of the disease on the tomato crop. From the analysis, minimum temperature, minimum relative humidity, maximum relative humidity, development stages; seedling, intermediate, flowering and fruiting had impact on tomato loss due to bacterial wilt prevalence (Table 6). From the study minimum relative humidity had a negative coefficient and showed that a unit increase in the minimum relative humidity would lead to a decrease of BW attack.

Table 6: Parameter estimates of the Poisson regression model of disease prevalence as relates to weather and crop development stage

Full model 4				
	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	2.3271553	0.3360046	6.926	4.33e-12***
Min Temperature	0.0279995	0.0090071	3.109	0.00188**
Max Temperature	0.0001011	0.0098366	0.01	0.991803
Precipitation	0.0023042	0.0016014	1.439	0.150193
Min Relative Humidity	-0.0051192	0.0018794	-2.724	0.006454**
Max Relative Humidity	0.0087269	0.0025266	3.454	0.000552***
Dev. Seedling	0.3707622	0.0371957	9.968	< 2e-16***
Dev. Intermediate	0.0611297	0.0244035	2.505	0.012247*
Dev. Flowering	0.1085326	0.0241847	4.488	0.0000072**
Dev. Fruiting	0.0446686	0.0232861	1.918	0.055079
Dev. Mature	-0.0352253	0.0330898	-1.065	0.287086.
Refitted model 4 to adjust for over parameterization				
Intercept	2.30406	0.231598	9.949	< 2e-16 ***
Min Temperature	0.028855	0.005981	4.824	1.40e-06 ***
Min Relative Humidity	-0.00465	0.001344	-3.46	0.000540 ***
Max Relative Humidity	0.008567	0.002502	3.424	0.000616 ***
Dev. Seedling	0.378026	0.036761	10.283	< 2e-16 ***
Dev. Intermediate	0.070608	0.023482	3.007	0.002640 **
Dev. Fruiting	0.048927	0.022948	2.132	0.032999 *
Dev. Flowering	0.112032	0.023107	4.848	1.24e-06 ***

[a] Significant codes: 0'***', 0.001'**', 0.01'*', 0.05'.', 0.1' ', 1

[b] Min Relative Humidity= Minimum Relative Humidity, Max Temperature= Maximum Temperature, Dev= Development stages of plant when disease incidents recorded

[c] Full model and refitted model based on the AIC process of selecting significant variables. Residual deviance of full model is 1459.7 and AIC of 2873.3 while the refitted model has a residual deviance (1462.4) and AIC (2870.1).

4.3.2 Model 5

The analysis showed that in upper midlands, the disease incidents as inferred from data queries brought to plant clinics was low compared to low midlands. This would be attributed by the low temperatures in the upper midlands (Table 7). Agro-ecological zones were found from the analysis to have significant influence on bacterial wilt incidents thereby resulting into losses of tomato due to infestation of BW. Various AEZ, LH2, LM3, LM4, LM7, UM2 and UM4 had a positive significant impact on average tomato loss with reference to LH1. It was an indication that bacterial wilt could be found across various AEZ

Table 7: Parameter estimates of the Poisson regression model on disease prevalence in relation to agro-ecological zones and crop loss as reported by farmers

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	3.20545	0.11625	27.574	< 2e-16***
AEZ LH2	0.35803	0.13252	2.702	0.006899**
AEZ LH3	-0.01087	0.14734	-0.074	0.941193
AEZ LH4	-0.0274	0.2349	-0.117	0.907146
AEZ LM3	0.4122	0.14221	2.898	0.00375**
AEZ LM4	0.45297	0.11844	3.824	0.000131***
AEZ LM5	0.20312	0.13111	1.549	0.121319
AEZ LM6	-0.01087	0.14734	-0.074	0.941193
AEZ LM7	0.70657	0.18307	3.86	0.000114***
AEZ UH2	-0.0274	0.2349	-0.117	0.907146
AEZ UM1	-0.01466	0.12914	-0.114	0.909621
AEZ UM2	0.3632	0.11998	3.027	0.002468**
AEZ UM3	0.10799	0.11783	0.916	0.359423
AEZ UM4	0.27902	0.11967	2.332	0.019721*

[a]Significant codes: 0'***', 0.001'**', 0.01'*', 0.05'.', 0.1' ', 1

[b]AEZ= Agro-ecological zone, LH= Lower upper land, LM= Lower middle land, UH= Upper highland, UM= Upper middle land.

[c] Model 5 has residuals deviance (1414.6) and AIC (2834.3) at 253 degree of freedom and null deviance (1628.7).

CHAPTER FIVE

DISCUSSION OF RESULTS

This chapter discusses the results obtained from modeling bacterial wilt in tomato.

5.1 Interpretation of the fitted binary logistic models

5.1.1 Model 1

In Model 1, weather data and development stage was used as independent variables. From the results it was observed bacterial wilt cases were reported to the plant clinics at intermediate, seedling development stage, flowering stage and maturity stage. Minimum relative humidity had significant impact on bacterial wilt incidence. However, minimum and maximum temperature, rainfall and maximum relative humidity did not show significant influence on BW incidence according to the reported cases of disease.

The intercept value of Model 1 was -4.773 indicated a reduction in incidence of bacterial wilt incidence (Table 2). Minimum temperature and rainfall showed a reduction in the presence of bacterial wilt though they were not significant. However, maximum temperature, minimum relative humidity, and maximum relative humidity had a positive impact in the presence BW as inferred cases from clinic queries. It was found that maximum relative humidity followed by maximum temperature had a greater impact (Table 2). From model 1a, it was observed that the intercept increased from -4.773 to -2.821 an indication of more attack of bacterial wilt. The null deviance explained more incidents of BW with the residual deviance of 1532 and AIC 1546. From the study, it was found that flowering, intermediate, and maturity stage, farmers reported more cases of bacterial wilt attack on tomatoes (Table 5). Temperature and precipitation did not significantly influence the bacterial wilt incidents as reported by farmers to the clinics. This would have been attributed by the fact that small scale farmers only plant during rainy season and solely depends on rainfall. Therefore, during dry season tomatoes are not in the field. However, minimum relative humidity was significant (z=2.595, p=0.00946). Daily temperature and precipitation influenced the distribution of bacterial wilt as it leads to its spread due to the runoff from the infested farms. According to Mila et al. (2004), temperature and precipitation increased the risk in bacterial wilt incidence as the it can tolerate up to a temperature of 32 ^{0}C. A study

by Mew et al. (1977) also observed that bacterial wilt can survive under varied range 26 ^{0}C to 32 ^{0}C temperature that tomato may scorch out while BW still survive.

In Table 5, there were some variable with negative coefficient on presence of bacterial wilt. The negative estimates indicated that on the reported cases by farmers these variables led to a decline in bacterial wilt incidents. The results indicated that at fruiting and maturity stage the bacterial wilt incidents were observed to be decreasing. However, at flowering and intermediate stage, the bacterial wilt incidents were observed to have increased. The Minimum relative humidity showed that with a unit increase in the temperature, bacterial wilt incidents increased by 0.0179 (z=2.595, p=0.00946).

5.1.3 Model 2

In Model 2 only weather variable was used. In the analysis, though maximum and minimum temperature had a positive coefficient of estimate, they were not significantly influencing BW incidences in tomato according to the reported cases. This may be attributed to other factors such as good farming practices and type of irrigation method used by farmers. A study carried out by Mila et al. (2004), determined temperature and precipitation as important factors to consider when investigating bacterial wilt incidence as it requires optimal condition to survive. In Table 2, it was found that minimum relative humidity increased the level of bacterial wilt incidents by 0.0242 and was significant (z= 2.176, p=0.02958) as from the clinic data queries. Maximum temperature, minimum temperature, maximum relative humidity and precipitation were not significant. The results showed that daily precipitation was fluctuating with some days recording no rain. Maximum relative humidity and minimum temperature had low coefficient estimates.

Considering the weather variables, the intercept showed that when no other factor was considered, the disease incidence reported reduced by 5.844 (Table 3). The refitted Model 2a was selected by comparing the AIC in the iterated models and the model with the least AIC. The variables used were maximum temperature and minimum relative humidity. The refitted Model 2a had an intercept of -5.262, an indication that bacterial wilt incidents was reduced by the magnitude that is slightly lower than the value when all variables are used (Table 5).

The results on Model 2a showed that from the clinic queries on bacterial wilt of tomato, maximum temperature was significant on the presence of bacterial wilt (z= 2.103, p= 0.035) and minimum

relative humidity was also significant (z= 3.407, p=0.00066) as shown in Table5. It was observed that the two variables increased the estimate of BW incidents that is an indication that most farmers reported more cases of the disease to plant clinics.

From the study, the mean maximum monthly temperature was observed to be in the range of 24 ^{0}C to 28 ^{0}C from January to December (Figure 4.1a). The temperature achieved throughout the year was advantageous for the survival of *Ralstonia solanacearum* as it can survive for long and in varied temperature over 21^{0}C. A study by Persley (1986), showed that *R. solanacearum* can survive in the soil for long time and within the range of 28 ^{0}C to 32 ^{0}C.

5.1.4 Model 3

In this model, agro-ecological zones were used to estimate the reported cases of bacterial wilt in tomato. Various counties were sub-grouped into various AEZ. However, the AEZ may be similar to other regions which are spatially different. The AEZ is based on the weather pattern of regions and since they are demarcated based on the weather data, there is correlation between the weather variables and AEZ. It therefore prompts for the separation of these two variables to avoid multicollinearity effect that results into biased standard error of the estimates.

The AEZ did not show any significant to the bacterial wilt incidents as per the reported cases (Table 4). This may be due to farm practices farmers adopted. A study by Michel et al. (1997), showed that various farm practices such as intercropping and soil amendment significantly reduces the population of BW of tomato. Although AEZ did not affect bacterial wilt incidents significantly in this study, the coefficient of estimates were negative. The intercept of the model was -1.897, an indicator of bacterial wilt incidents reduced by 1.897. The effect of AEZ on bacterial wilt showed a reduction and an increase in disease incidents. This showed that under AEZ with positive coefficient estimate, cases of disease incidents increased while the AEZ with negative coefficient estimates showed that cases of the bacterial wilt incidents reported by farmers reduced (Table 4). These estimates majorly depend on the farmers who visited the clinic and a practical study conducted may conform to what other researchers found. A study in Nigeria on agro-ecological effect of bacterial wilt on cassava showed significant difference of the disease attack under different environmental conditions (Ngeve & Nukenine, 2002).

Model 3a was refitted using both the development stage and AEZ. It showed that development stages significantly attribute to the bacterial wilt incidents as reported by farmers to the plant clinics.. Intermediate, flowering and mature stage was significant. However, at mature stage BW reported incidents decreased by 0.353 while flowering and intermediate stage showed an increase in the incidents of bacterial wilt with 0.343 and 0.504 respectively (Table 5).

5.2 Interpretation of the fitted poisson models

It was found that minimum temperature, minimum relative humidity and maximum relative humidity had impact on tomato loss due to attack by bacterial wilt. High and low relative humidity areas was found to influence the disease incidents resulting into tomato losses in the farm (Geoffrey et al., 2014). These losses were majorly observed at different stages of tomato development as reported by farmers. Farmers majorly recorded losses at seedling stage, intermediate stage, flowering and fruiting stage (Table 6). The intercept revealed a mean loss of tomato of 2.304. The coefficient of Model 4 and refitted Model 4a were slightly different. The optimal model obtained by refitting the parameters showed the best estimates since it had the least AIC of 2870.1 compared to 2873.3 AIC of the model fitted with all the parameters. Minimum relative humidity reduced the rate of crop loss by 0.0046 and (z= 3.460, p= 0.00054). This maybe attributed by canopy and poor farming practices most farmers do and overcrowded tomato plant in the farms. The overcrowding of tomato and its vegetative stems may form a canopy creating humid environment leading to rise bacterial incidence (Geoffrey et al., 2014).

Bacterial wilt was reported to be present in the lower elevated regions due to the warm temperature (Ajanga, 1987) and it leads to production of high quality tomato seeds in the high altitude regions where the BW is not widespread. However, since most farmers are small scale and cannot afford good seeds, they use the low quality seeds that are used over period leading to consistency in BW persistent in the farms.

Agro-ecological zones were used in Model 5. The AEZ which were significant in estimating crop losses due to attack by bacterial wilt were LH2, LM3, LM4, LM7, UM2 and UM4 (Table 8) and from the study, it was found that the risk of tomato crop infected by bacterial wilt increased. These AEZ are warm and due to their warm states they encourage the spread of bacterial wilt and its reoccurrence in the farm as it can survive at high temperatures of 27 ^{0}C to 32 ^{0}C (Deberdt et al., 1999).

Farmers in these regions recycle their tomato seeds making it difficult to eradicate the disease. According to Ajanga (1987), most of the own produced seeds in the lower elevated regions get into the market and farmers purchase them due to their low income status resulting into the spread of the BW. Bacterial wilt was observed to be spreading in all the regions and this would have been led by recycling own produced seeds. Hayward (1991), reported that BW is widely distributed in tropical, sub-tropical and some warm temperate regions leading to crop loss. The model coefficient estimates of different AEZ had a positive and negative marginal effect, the negative coefficient estimates indicated a diminishing marginal return of crop loss as a result of bacterial wilt while the positive coefficient estimates indicated an increase in tomato crop loss. When good farm practices are followed these losses would probably reduce due to reduction in bacterial wilt infestation on tomatoes.

CHAPTER SIX
CONCLUSIONS AND RECOMMENDATIONS

6.1 Introduction

In this chapter, the presentation of conclusions and recommendations for further research arising from the study.

6.2 Conclusion

The study applied binary logistic to model the distribution and crop loss associated with bacterial wilt of tomatoes within the scope of logistic models. Binary logistic model was used because the response variable was a dummy. The study found out that bacterial wilt attack to tomato was reported in all the thirteen counties where clinics are located in Kenya. Among those counties, The highest attack was reported in Kirinyaga followed by Nakuru and Embu. The least percentage of farms attacked were reported in Narok, West Pokot, Tharaka-Nithi, Kajiado and Kiambu.

Inferring from the binary logistic model, it was found that, distribution of bacterial wilt is not necessarily influenced by agro-ecological zones. The agro-ecological zones of a region is dependant on weather pattern of the region. Therefore, bacterial wilt distribution was determined independently based on weather data and agro-ecological zones. The study found out that, agro-ecological zones had no impact on the bacterial wilt distribution. This is due to poor farming practices within the tomato growing regions.

However, using the weather data, it was found that relative humidity and temperature had an impact on the bacterial wilt attack. Bacterial wilt can persist in the soil for a long time leading to frustration of farmers in managing. In cold regions, bacterial wilt does not express itself and remains in dormant state till a favorable temperature is attained. This may also results into bacterial wilt not identified at some stage in crop development. In tomato development stage, farmers reported more cases of bacterial wilt attack during intermediate, flowering and maturity stage. Farmers are always attentive to any anomalies in their crops especially incidents of any wilting in tomato as it is widely used for food and for commercial purpose. However, at some stages in tomato growth such as seedling, fruiting, and post harvest stage, the bacterial wilt cases reported had no impact in presence of bacterial wilt

Estimated tomato loss was determined using poisson model using the weather, agro-ecological zones and development stage. This helped to estimate the impact of these variables on the tomato losses associated with bacterial wilt. The study found out that, minimum temperature and maximum relative humidity increased the tomato loss associated by bacterial wilt. However, minimum relative humidity showed a decrease of tomato loss due to attack by bacterial wilt. The study also found that the development stages which farmers mostly reported attack on tomato were at seedling, intermediate, fruiting and flowering. They showed an increase in the impact of bacterial wilt attack on tomato. This study assumed that wilting was caused by bacterial wilt and irrigation, management practices and source of tomato seeds were held constants.

6.3 Recommendation

The study findings indicated that bacterial wilt is distributed in 12 out of 13 the counties under study. The distribution of bacterial wilt could have been attributed by factors that were assumed to be constant in all the counties. However, in some regions, the bacterial wilt attacks reported were low. The disease distribution presumably would be due to poor farming practices of farmers

Tomato farmers should be advised using the findings from the study by the policy makers to reduce bacterial distribution and crop loss associated by bacterial wilt. This would help farmers to improve the tomato yield and encounter the attack by bacterial wilt that has been reported in the 12 out of 13 counties in Kenya. Good training of farmers using the findings on the external factors that influence distribution of bacterial wilt in the tomato growing regions would lead to reduction of bacterial wilt and more production of tomatoes in the farm. The other external factors such as farming practices, irrigation and soil type can be further explored for more studies.

REFERENCE

Abawi, G. ., & Widmer, T. . (2000). Impact of soil health management practices on soilborne pathogens, nematodes and root diseases of vegetable crops. *Applied Soil Ecology*, *15*(1), 37–47. doi:10.1016/S0929-1393(00)00070-6

Afroz, A., Khan, M. R., Ahsan, N., & Komatsu, S. (2009). Comparative proteomic analysis of bacterial wilt susceptible and resistant tomato cultivars. *Peptides*, *30*, 1600–1607. doi:10.1016/j.peptides.2009.06.005

Agresti, A. (2002). Categorical Data Analysis (pp. 1–701). doi:10.1086/431086

Ajanga, S. (1987). Status of Bacterial Wilt of Potato in Kenya, 338–372.

Ayana, G., Fininsa, C., Ahmed, S., & Wydra, K. (2011). Effects of Soil Amendment on Bacterial Wilt Caused By Ralstonia Solanacerum and Tomato Yields in Ethiopia, *51*(1). doi:10.2478/v10045-011-0013-0

Babtain, A. A. (n.d.). Using Binary Logistic Regression and Probit Analysis To Students ' Characteristics with the Identification Results of Mental Ability Test : A Comparative Study, 1–13.

Berry, W. D., Demeritt, J. H. R., Esarey, J., Florida, W. D. B., & Emory, J. E. (2015). in Binary Logit and Probit Testing for Interaction Is a Product Term Essential ? Models : State University, *54*(1), 248–266.

Browne, W. J., Subramanian, S. V, Jones, K., & Goldstein, H. (2003). Variance partitioning in multilevel logistics models with over-dispersion. *Working Paper*, *168*(3), 599–613. Retrieved from http://www.maths.nott.ac.uk/personal/pmzwjb/materials/billcorr1.pdf

Buschermohle, M. J., & Grandle, G. F. (2012). W017-Controlling the Environment in Greenhouses Used for Tomato Production. *The University of Tennessee Agricultural Extension Service*.

Calonnec, a, Cartolaro, P., & Chadoeuf, J. (2009). Highlighting features of spatiotemporal spread of powdery mildew epidemics in the vineyard using statistical modeling on field experimental data. *Phytopathology*, *99*(4), 411–422. doi:10.1094/PHYTO-99-4-0411

Cameron, a C., & Trivedi, P. K. (2003). Essentials of Count Data Regression. *A Companion to Theoretical Econometrics*, 331–348. doi:10.1002/9780470996249.ch16

Chen, Y. Y., Lin, Y. M., Chao, T. C., Wang, J. F., Liu, A. C., Ho, F. I., & Cheng, C. P. (2009). Virus-induced gene silencing reveals the involvement of ethylene-, salicylic acid- and mitogen-activated protein kinase-related defense pathways in the resistance of tomato to bacterial wilt. *Physiologia Plantarum*, *136*, 324–335. doi:10.1111/j.1399-3054.2009.01226.x

Cnaan, A., Laird, N. M., & Slasor, & P. (1997). Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat.Med.*, *16*, 2349–2380. doi:10.1002/(SICI)1097-0258(19971030)16:20<2349::AID-SIM667>3.0.CO;2-E [pii]

Contreras-Medina, L. (2009). Mathematical modeling tendencies in plant pathology. *African Journal of ...*, *8*(25), 7399–7408. Retrieved from http://www.ajol.info/index.php/ajb/article/view/77754

Dean, C. B. (1992). Testing for Overdispersion in Piosson and Binomial Regression Models. *Journal of the American Statistical Association*, *87*(418), 451–457.

Deberdt, P., Quénéhervé, P., Darrasse, a., & Prior, P. (1999). Increased susceptibility to bacterial wilt in tomatoes by nematode galling and the role of the Mi gene in resistance to nematodes and bacterial wilt. *Plant Pathology*, *48*, 408–414. doi:10.1046/j.1365-3059.1999.00346.x

Follmann, D. a, & Lambert, D. (1989). Generalizing logistic regression by nonparametric mixing. *Journal of the American Statistical Association*, *84*(405), 295–300. doi:10.2307/2289877

Forster, J. J., Mcdonald, J. W., Smith, P. W. F., Journal, S., Statistical, R., Series, S., … Smitht, P. W. F. (1995). Monte Carlo Exact ConditionaTl ests forL og-lineara nd LogisticM odels, *58*(2), 445–453.

Geoffrey, S. K., Hillary, N. K., Antony, K. M., Mariam, M., & Mary, M. C. (2014). Challenges and Strategies to Improve Tomato Competitiveness along the Tomato Value Chain in Kenya. *International Journal of Business and Management*, *9*(9), 205–212. doi:10.5539/ijbm.v9n9p205

Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J., & Bossuyt, P. M. M. (2003). The diagnostic odds ratio: A single indicator of test performance. *Journal of Clinical Epidemiology*, *56*, 1129–1135. doi:10.1016/S0895-4356(03)00177-X

Haberman, S. J., & Sinharay, S. (2010). The Application of the Cumulative Logistic Regression Model to Automated Essay Scoring. *Journal of Educational and Behavioral Statistics*, *35*(5), 586–602. doi:10.3102/1076998610375839

Hastie, T. J., & Tibshirani, R. (1990). Generalized additive models. *Statistical Science*. doi:10.1016/j.csda.2010.05.004

Hayward, A. C. (1991). Bacterial Wilt Caused By. In *Annu. Rev. PhYlOpathol.* (pp. 65–87).

Hayward, A. C. (1991). Biology and epidemiology of bacterial wilt caused by pseudomonas solanacearum. *Annual Review of Phytopathology*, *29*, 65–87. doi:10.1146/annurev.py.29.090191.000433

Hosmer, D. W., Hosmer, T., Le Cessie, S., & Lemeshow, S. (1997). A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in Medicine*, *16*(May 1995), 965–980. doi:10.1002/(SICI)1097-0258(19970515)16:9<965::AID-SIM509>3.0.CO;2-O

Kay, B. Y. R. (1987). Transformations of the explanatory variables in the logistic regression model for binary data, *74*(3), 495–501.

Kelman A (1998) One hundred and one years of research on bacterial wilt. In: Prior P, Allen C, Elphinstone J. (eds). Bacterial Wilt: Molecular and Ecological Aspects, pp. 1-5. INRA Editions, Paris, France. Kelman A., Person L.H., 1961.

Kenya Agricultural Research Institute (KARI). 2005. KARI annual report 2005. Nairobi, Kenya.

Kenya Horticulture Competitiveness Project (KHCP) report, 2011. Fruits and vegetable retail tracker June – August 2011 report. USAID. http://www.fintrac.com/cpanelx_pu/kenya%20khcp/10_00_3937_KHCP_Horticulture%20Retail%20Audit_August%202011.pdf . Viewed 28th August 2015

Koehler K., Meeker, S. and C. (2001). Binary Response and Logistic Regression Analysis.

Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Handling imbalanced datasets : A review. *Science*, *30*, 25–36. doi:10.1007/978-0-387-09823-4_45

Landwehr, J. M., Pregibon, D., & Shoemaker, A. C. (1984). Graphical Methods for Assessing Logistic Regression Models, *79*(385), 61–71.

Lange, W. H., & Bronson, L. (1981). Insect pests of tomatoes +6212.

Lebeau, A., Daunay, M.-C., Frary, A., Palloix, A., Wang, J.-F., Dintinger, J., … Prior, P. (2011). Bacterial wilt resistance in tomato, pepper, and eggplant: genetic resources respond to diverse strains in the Ralstonia solanacearum species complex. *Phytopathology*, *101*(January), 154–165. doi:10.1094/PHYTO-02-10-0048

Maji, S., & Chakrabartty, P. K. (2014). Biocontrol of bacterial wilt of tomato caused by Ralstonia solanacearum by isolates of plant growth promoting rhizobacteria. *Australian Journal of Crop Science*, *8*(2), 208–214.

Mew, T. W. (1977). Effect of Soil Temperature on Resistance of Tomato Cultivars to Bacterial Wilt. *Phytopathology*, *77*, 909. doi:10.1094/Phyto-67-909

Michel, V., Wang, J. F., Midmore, D., & Hartman, G. (1997). Effects of intercropping and soil amendment with urea and calcium oxide on the incidence of bacterial wilt of tomato and survival of soil-borne Pseudomonas solanacearum in Taiwan. *Plant Pathology*, *46*, 600–610. Retrieved from http://onlinelibrary.wiley.com/doi/10.1046/j.1365-3059.1997.d01-45.x/full

Mila, a L., Carriquiry, a L., & Yang, X. B. (2004). Logistic regression modeling of prevalence of soybean sclerotinia stem rot in the north-central region of the United States. *Phytopathology*, *94*, 102–110. doi:10.1094/PHYTO.2004.94.1.102

Miller, A. J. H., & Crosier, W. F. (2015). Association of Official Seed Analysts Society of Commercial Seed Technologists Pathogenic Associates of Tomato Seed : Their Prevalence , Relation to Field Disease and Elimination.

Nau, R. (2014). Notes on Linear Regression. *Notes*, (c), 1–18.

Ngeve, J. M., & Nukenine, E. N. (2002). Genotype × environment effects on severity of cassava bacterial blight disease caused by, 763–770.

Oerke, E.-C. (2006). Crop losses to pests. *The Journal of Agricultural Science*, *144*(01), 31–43. Retrieved from href="http://dx.doi.org/10.1017/S0021859605005708

Ogbodo, E. N. (2013). Impact of the use of Inorganic Fertilizers to the Soils of the Ebonyi State Agro-Ecology , South-Eastern Nigeria ., *3*(7), 33–39.

Peng, C.-Y. J., & Harry, T.-S. (2002). Logistic Regression Analysis and Reporting.

Persley G. J, Batugal P., G. D. (1986). Bacterial Wilt Disease in Asia and the South Pacific. *Australian Centre for International Agricultural Research CI.P.O. Box 1571, Canberra, A.C.T. 2601*, (13), 126–143.

Philip, Jaffee, S., & Okello, J. (2004). Exporting out of Africa: Kenya's horticulture success story. *Scaling up Poverty Reduction: A Global Learning Process Conference*.

Poole, M. a, & O'Farrell, P. N. (1970). The assumptions of the linear regression model. *Transactions of the Institute of British Geographers*, *52*, 145–158.

Pregibon, D. (1981). Logistic Regression Diagnostics ', *9*(4), 705–724.

Prior, P., Bart, S., Leclercq, S., Darrasse, a, & Anais, G. (1996). Resistance to bacterial wilt in tomato as discerned by spread of Pseudomonas (Burholderia) solanacearum in the stem tissues. *Plant Pathology*, *45*, 720–726. doi:10.1046/j.1365-3059.1996.d01-9.x

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk , Kolmogorov-Smirnov , Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, *2*(1), 21–33.

Rodriguez, G. (2013a). Models for Count Data With Overdispersion. *Princeton Statistics*, 1–7.

Rodriguez, G. (2013b). Models for Count Data With Overdispersion. *Princeton Statistics*, (September 2007), 1–7.

Singh, R., & Regmi, K. R. (2013). A Value Chain Analysis of Vegetables: A Case Study of Palpa District, Nepal. *Economic Affairs*, *58*(2), 135–146. Retrieved from http://www.ndpublisher.in/ndpjournal.php?j=EA\nhttp://search.ebscohost.com/login.aspx?direct=true&db=ecn&AN=1442041&site=ehost-live

Strange, R. N., & Scott, P. R. (2005). Plant disease: a threat to global food security. *Annual Review of Phytopathology*, *43*(Figure 1), 83–116. doi:10.1146/annurev.phyto.43.113004.133839

Wagner, H. H. (2013). Rethinking the linear regression model for spatial ecological data. *Ecology*, *94*(11), 2381–2391.

Wani, A. H. (2011). An overview of the fungal rot of tomato, *9*, 33–38.

Wenneker, M., Verdel, M. S. W., Groeneveld, R. M. W., Kempenaar, C., Van Beuningen, a. R., & Janse, J. D. (1999). Ralstonia (Pseudomonas) solanacearum race 3 (biovar 2) in surface water and natural weed hosts: First report on stinging nettle (Urtica dioica). *European Journal of Plant Pathology*, *105*(biovar 2), 307–315. doi:10.1023/A:1008795417575

Whittaker, G., Confesor Jr., R., Di Luzio, M., & Arnold, J. G. (2010). Detection of overparameterization and overfitting in an automatic calibration of SWAT. *Transactions of the ASABE*, *53*(5), 1487–1499.

Williams, R. (2014). Logistic Regression , Part I : Problems with the Linear Probability Model (LPM), 1–11. Retrieved from http://www3.nd.edu/~rwilliam/stats2/l81.pdf

Yadessa, G. B., van Bruggen, a. H. C., & Ocho, F. L. (2010). Effects of different soil Amendments on bacterial wilt caused by ralstonia solanacearum and on the yield of tomato. *Journal of Plant Pathology*, *92*(2), 439–450.

APPENDICES

Appendix 1: Data exploration using box plot for weather data

```
===============================================================
==Exploratory analysis using box plot for the parameter which influence BW incidence ======
===============================================================
par(mfrow=c(1,2))
boxplot(BWdata$MinTemperature~BWdata$Month,col="blue",    xlab="Month",    ylab="Min
temperature", main="Daily minimum\n temperature")
boxplot(BWdata$MaxTemperature~BWdata$Month, col="blue", xlab="Month", ylab="Max
temperature", main="Daily maximum\n temperature")
par(mfrow=c(1,2))
boxplot(BWdata$PercentMinRelativeHumidity~BWdata$Month,  col="blue",  xlab="Month",
ylab="Min reative humidity", main="Daily minimum\n Relative humidity %")
boxplot(BWdata$PercentMaxRelativeHumidity~BWdata$Month,  col="blue",  xlab="Month",
ylab="Max relative humidity", main="Daily maximum\n Relative humidity %")
par(mfrow=c(1,2))
plot(BWdata$Precipitation_mm, type="l")
boxplot(BWdata$Precipitation_mm~BWdata$Month,col="blue",xlab="Month",
ylab="Precipitation", main="Daily precipitation in mm")
abline(plot(BWdata$Precipitation_mm, type="l",ylab="Precipitation",xlab="Daily temperature",
main="Daily precipitation in mm"), h=2.374, col="red", lwd=3)
```

Appendix 2: Modeling the incidence of bacterial wilt

2A. Model 1

```
============================================================================
====  model for BW incidence using Climatic data and development stage======
============================================================================

model1<-glm(PresenceBW~MinTemp+MaxTemp+Precipitation+MinRelHumidity+
MaxRelHumidity+DevStageSeedling+DevStageIntermediate+DevStageFlowering+DevStageFru
iting+DevStageMature+DevStagePostHarvest+MinTemp:Precipitation+MaxTemp:Precipitati
on+MinRelHumidity:Precipitation+MaxRelHumidity:Precipitation,data=BWdata,
family="binomial")

summary(model1)
```

2B. Model 2

```
model2<-
glm(PresenceBW~MinTemp+MaxTemp+Precipitation+MinRelHumidity+MaxRelHumidity+Precipi
tation:MinRelHumidity+Precipitation:MaxRelHumidity+Precipitation:MinTemp+Precipita
tion:MaxTemp,data=BWdata, family="binomial")
summary(model2)
```

2C. Model 3

```
model3<-glm(PresenceBW~AEZ+   DevStageSeedling   +   DevStageIntermediate   +
DevStageFlowering + DevStageFruiting + DevStageMature + DevStagePostHarvest,
data= BWdata, family="binomial")
summary(model3)
```

Appendix 3: Selected R codes

Data management

```
library(foreign, pos=4)
BWdata<-read.csv(file.choose(), header=T)
BWdata
summary(BWdata)
MinTemp<-(BWdata$MinTemperature)
MaxTemp<-(BWdata$MaxTemperature)
Precipitation<-(BWdata$Precipitation_mm)
MinRelHumidity<-(BWdata$PercentMinRelativeHumidity)
MaxRelHumidity<-(BWdata$PercentMaxRelativeHumidity)
CropAffected<-(BWdata$CropAffected)
AEZ<-(BWdata$Agro.ecological.zones)
County<-factor(BWdata$FarmerCounty,        levels=c(1,2,3,4,5,6,7,8,9,10,11,12,13),labels=c("Embu",
"Kirinyaga", "Machakos", "Nakuru", "Bungoma","Elgeyo Marakwet","Kajiado", "Kiambu", "Narok",
"Nyeri","Tharaka Nithi", "Trans Nzoia", "West Pokot"))
AEZ<-factor(BWdata$Agro.ecological.zones,
lables=c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16),levels=c("LH1","LH2","LH3","LH4","LM3","LM4",
"LM5","LM6","LM7","UH1","UH2","UM1","UM2","UM3","UM4","UM5"))
PresenceBW<-factor(BWdata$PresenceBacterialWilt, levels=c(0,1), labels=c("No","Yes"))
DevStageSeedling=factor(BWdata$DevStageSeedling, levels=c(0,1), labels=c("No","Yes"))
DevStageIntermediate=factor(BWdata$DevStageIntermediate, levels=c(0,1), labels=c("No","Yes"))
DevStageFlowering=factor(BWdata$DevStageFlowering, levels=c(0,1), labels=c("No","Yes"))
DevStageFruiting=factor(BWdata$DevStageFruiting, levels=c(0,1), labels=c("No","Yes"))
DevStageMature=factor(BWdata$DevStageMature, levels=c(0,1), labels=c("No","Yes"))
DevStagePostHarvest=factor(BWdata$DevStagePostHarvest, levels=c(0,1), labels=c("No","Yes"))
```

Model Selection codes

Model 1

Initial model specification

```
model1<-glm(PresenceBW~MinTemp+MaxTemp+Precipitation+MinRelHumidity+
MaxRelHumidity+DevStageSeedling+DevStageIntermediate+DevStageFlowering+DevStageFru
iting+DevStageMature+DevStagePostHarvest+MinTemp:Precipitation+MaxTemp:Precipitati
on+MinRelHumidity:Precipitation+MaxRelHumidity:Precipitation,data=BWdata,
family="binomial"(link="logit"))
summary(model1)
model1.b<-step((model1), direction="backward")
```

Model selection using backward method

```
modelrefit<-
glm(PresenceBW~MinRelHumidity+DevStageFlowering+DevStageFruiting+DevStageIntermedi
ate+DevStageMature+DevStagePostHarvest,data=BWdata,
family="binomial"(link="logit"))
summary(modelrefit)
```

Model 2

Initial model specification

```
model2<-
glm(PresenceBW~MinTemp+MaxTemp+Precipitation+MinRelHumidity+MaxRelHumidity+Precipi
tation,data=BWdata, family="binomial"(link="logit"))
summary(model2)
model2.b<-step((model2), direction="backward")
```

Model selection using backward method

```
modelrefit2<-glm(PresenceBW~MinRelHumidity+ MaxTemp
, data=BWdata, family="binomial"(link="logit"))
summary(modelrefit2)
```

Model 3

Initial model specification

```
model3<-
glm(PresenceBW~AEZ+DevStageSeedling+DevStageIntermediate+DevStageFlowering+DevStag
eFruiting+DevStageMature+DevStagePostHarvest, data=BWdata,
family="binomial"(link="logit"))
summary(model3)
model3.b<-step((model3), direction="backward")
```

Model selection using backward method

```
modelrefit3<-
glm(PresenceBW~DevStagePostHarvest+DevStageFlowering+DevStageMature+DevStageInterm
ediate,data=BWdata, family="binomial"(link="logit"))
summary(modelrefit3)
```

Poisson Model for the bacterial wilt prevalence

Mode 4

```
model4<-
glm(Cropattacked~MinTemp+MaxTemp+Precipitation+MinRelHumidity+MaxRelHumidity+
DevStageSeedling+DevStageIntermediate+DevStageFlowering+DevStageFruiting+
DevStageMature+DevStagePostHarvest, data=CropAffected, family="poisson"
(link = "log"))
summary(model4)
model4.b<-step((model4), direction="backward")
```

Model selection using backward method

```
modelrefit4<-
glm(Cropattacked~MinTemp+MinRelHumidity+MaxRelHumidity+DevStageSeedling+
DevStageIntermediate+DevStageFruiting+DevStageFlowering, data=CropAffected,
family="poisson"(link="log"))
summary(modelrefit4)
```

Mode 5

```
model5<-glm(Cropattacked~AEZ,data=CropAffected, family="poisson"(link="log"))
summary(model5)
```