Spatial Modelling of Maize Lethal Necrosis Disease in Bomet County, Kenya

Article · October 2017

3 authors, including:

Osunga Michael Otieno
Jomo Kenyatta University of Agriculture and 
2 PUBLICATIONS 0 CITATIONS

Felix Mutua
Jomo Kenyatta University of Agriculture and 
10 PUBLICATIONS 1 CITATION

Some of the authors of this publication are also working on these related projects:

Using factor analysis to categorize and map ecological variables influencing maize lethal necrosis disease in Bomet County, Kenya View project
Spatial Modelling of Maize Lethal Necrosis Disease in Bomet County, Kenya

Michael Osunga1,*, Felix Mutua1, Robinson Mugo2

1Department of Geomatics Engineering and Geospatial Information Systems, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya
2SERVIR-Eastern and Southern Africa, Regional Centre for Mapping of Resources for Development, Nairobi, Kenya
*Corresponding author: mikeotieno87@gmail.com

Abstract Maize lethal necrosis (MLN) is a disease that attacks maize crops with significant impacts on both food security and nutrition security on smallholder farmers in Kenya. The study used spatial regression analysis to model MLN severity on sampled farm fields in Bomet County, Kenya. The modelling analysis integrated spatial information based on derived crop mask, on-site derived MLN disease severity index at an optimal maize growing season and phenological stage. Relevant ecological variables derived spatially including temperature, rainfall, soil moisture and slope were identified and fed into a spatial regression model. Significant ecological variables were weighted and used as basis for generating spatially explicit MLN severity index map. MLN affected farms have spatial dependence with MLN severity becoming less correlated the further away from each MLN affected farm field. The ecological variables have negative influence on MLN severity except for temperature. Soil moisture, rainfall and slope are the most significant determinants of MLN severity index in Bomet (all <p 0.05), with high MLN severity areas identified in Chebunyo, Sigor and Kipereres. This study would help in MLN epidemiological surveillance and in developing site-specific control measures and interventions. The spatial model used in this study could be replicated and up-scaled to other MLN prone areas in Kenya and in Africa coupled with other statistically significant spatiotemporal ecological variables to fully understand and ascertain MLN disease outbreak.

Keywords: maize lethal necrosis, disease severity index, spatial regression


1. Introduction

Maize lethal necrosis (MLN) is a disease that attacks maize crops. MLN disease was first reported in Kansas in the year 1976 [1,2]. It then spread to Hawaii where it was reported in the year 1990 in the town of Kauai [3]. In Africa, it was first reported in the year 2011, the month of September in Bomet County where it spread into Chepalungu, Narok and Naivasha districts [4]. The disease was reported in all the provinces in Kenya with the exception of North eastern province [4]. It has been reported in Rwanda [5], Tanzania [6], Uganda [7], Democratic Republic of Congo [8], Ethiopia [9], Burundi [10] and South Sudan [9]. Maize crops are susceptible to MLN at all phenological stages [6]. The rate of MLN infection and damage is high, affecting yields and causing loss of crops [11] therefore having significant impacts on both food security and nutrition security of farming families in Kenya. MLN disease tetrahedron in East Africa lies within the interaction amongst pathogens, disease hosts and vectors with the environment [12], as illustrated in Figure 1.

Figure 1. MLN disease tetrahedron in East Africa
Ecological variables have direct and indirect effects on MLN as they tend to influence pathogen infection, affect the development of MLN insect vectors; contribute to more conducive environment for insect vectors’ breeding and development. Pathogens include maize chlorotic mottle virus (MCMV) and sugarcane mosaic virus (SCMV) that contribute to MLN [13,14,15,16]. The disease host includes the maize crop itself [14] and alternative hosts like sorghum, finger millet, sugarcane, Napier grass, kikuyu grass and wild grass [15]. MLN is transmitted mechanically and spread by several insect vectors such as aphids, leafhoppers [4,16,17], thrips, [3,18] and beetle species of the family chrysomelidae [17] namely; corn flea beetle, cereal flea beetle, flea beetle, western, northern and southern corn rootworms.

Researchers have shown that MLN occurrence tends to be influenced by rainfall, temperature, soil moisture and slope. The amount of damage caused by MCMV pathogen is strongly correlated with high precipitation levels [19]. MLN was restricted to areas which received less than 50mm of rain. MLN epidemics in Kansas, USA, followed years with above-normal rainfall [20]. MCMV appeared viable under conditions as low as 0 mm to 813mm [21]. Temperatures between 28 to 31 degrees favor MCMV infections [22]. In growth chambers, high temperatures favored rapid spread of MCMV. MCMV was profound in annual mean temperatures ranging between 11.6-23.9 degrees and mean temperature of coldest quarter ranging between 9.6 to 22.2 degrees Celsius [21]. Low moisture content in the soil results in maize plants becoming more susceptible to MLN. MLN is more severe during periods of soil moisture stress [23,24]. Population of corn rootworms vary according to whether samples are from upland or lowland crop fields [25]. Significant lower populations of thrips are observed during rainy and cold seasons than during hot and dry seasons [26,27,28]. Heavy rainfall negatively affects thrips populations by killing larvae and suppressing adult flight [28,29,30]. Rainfall positively influencess thrips population [26]. Thrips are significantly influenced by increased temperature [31]. Increased temperatures throughout spring result in greater thrps activity and population growth [28,32,33]. Mean precipitation and temperature are strongly associated with population dynamics of cereal aphids [34]. Aphid population density is positively affected by rainfall [35]. An average temperature of 18.06 degrees (Maximum 22.81 and Minimum 13.31) provides conductive conditions for aphid incidence [36]. Maximum temperature is positively correlated with aphid population and minimum temperature is negatively correlated with aphid population [37]. Slope explains a significant variation in aphid population density within the crop fields [38]. Leafhopper population has a significant and positive correlation with rainfall and a negative association with temperature [39]. A forward stepwise regression analysis detected a negative contribution of total precipitation to the estimated density of leaf hoppers [40]. The population of leaf hoppers is correlated with rainfall and temperature [41]. Varying day-night temperatures accelerated larval development of corn rootworms due to direct temperature effects [42]. Corn rootworm adults prefer to lay eggs in moist rather than in dry soils [43]. Females lay eggs near the soil surface if soil moisture is high [43].

The first study to explore the landscape ecology and epidemiology of MCMV and MLN across Africa was conducted by [44] in order to understand the spatiotemporal distribution of MCMV and MLN risk in Africa. An ecological niche model based on MLN incidence point data using genetic algorithm for rule-set prediction model (GARP) was used [44]. Inputs included 12 bioclimatic variables. MCMV and MLN-positive incidences across the region corresponded to a variety of temperature and precipitation regimes in the semi-arid and sub-humid tropical sectors of central and eastern Africa. The study provided views on distribution and epidemiology of MCMV and MLN across Africa [44]. Satellite imagery combined with field-based information on MLN infection rates has been used to map MLN severity levels in Bomet County, Kenya [45]. MLN severity levels were mapped using in-situ data set from the field as training data in Random Forest. Results from this study indicated possibility of using LANDSAT and Random Forest classification to monitor spatial distribution of disease infestation in small scale and fragments agro-ecological landscapes [45]. MLN severity has also been mapped using Landsat data [46]. Random Forest classifier was optimized using variables selection on spectral indices and bands. Using only most relevant spectral indices, three MLN severity classes could be mapped [46]. The study showed the possibility of mapping maize diseases using spectral vegetation indices and optimized machine learning algorithms like Random Forest. The results could be used to understand linkages between MLN and underlying ecological factors for better knowledge on MLN disease propagation and spread in Kenya [46].

The above-mentioned MLN studies [44,45,46] used models that did not explicitly account for spatial dependence and autocorrelation which often arises when observations are collected from points located in space. The objectives of this study are to spatially derive ecological variables leading to MLN, to investigate spatial dependence, autocorrelation and significance of each ecological variable with respect to MLN severity and to generate a spatially explicit MLN severity index map. This research is conducted in the month of March, 2017, over affected farmlands in Bomet County based on MLN field data at optimal maize growing season and phenological stage.

2. Data and Methods

2.1. Study Area

Bomet County is located in the southern part of the Rift Valley of Kenya at latitude 0.8015° S and longitude 35.3027° E (Figure 2). The county has a population of 730,129 inhabitants and a geographical coverage of approximately 1,997.9 km² with an elevation of 6,437 meters above sea level. Its main economic activity is agricultural and livestock farming. Administratively it is subdivided into 8 sub counties.
2.2. Data

The table below shows the list of spatially derived ecological variables used for the study, their sources and spatial resolutions.

<table>
<thead>
<tr>
<th>Ecological variables</th>
<th>Variable name</th>
<th>Units</th>
<th>Resolution and source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Rainfall</td>
<td>Millimeters</td>
<td>5km (CHIRPS)</td>
<td></td>
</tr>
<tr>
<td>2 Temperature</td>
<td>Degrees Celsius</td>
<td>5km (MODIS)</td>
<td></td>
</tr>
<tr>
<td>3 Slope</td>
<td>Degree slope inclination</td>
<td>30m(SRTM)</td>
<td></td>
</tr>
<tr>
<td>4 Soil moisture</td>
<td>Topographic wetness index</td>
<td>30m(Landsat,USGS)</td>
<td></td>
</tr>
<tr>
<td>5 Maize crop + MLN alternative hosts mask</td>
<td>Acreage</td>
<td>30m(Landsat,USGS)</td>
<td></td>
</tr>
</tbody>
</table>

CHIRPS, climate hazards group infrared precipitation with station data; SRTM, shuttle radar topography mission; MODIS, moderate resolution imaging spectroradiometer; USGS, United States geological survey.

Temperature data was sourced from MODIS while rainfall data was sourced from CHIRPS. Slope was sourced and derived [47] from SRTM. Multi-spectral, radiometrically corrected, 30m spatial resolution Landsat data sets that have spectral bands ranging in the visible bands to infra-red bands were utilized in this research. Radiometric calibration was done for each spectral band by converting image digital numbers to surface reflectance. Topographical wetness was derived from radiometrically corrected Landsat data. Tasseled cap wetness transformations were applied to spectral bands based on the coefficients [48]. The values were then standardized by a rescaling factor creating a range between 0 to 1 where 1 represents areas of high topographical wetness value and 0 areas with low wetness value. Random Forest classification algorithm [45,46] was used in differentiation of variations in spectral signatures between crops and non-crops hence distinguishing maize crops and MLN alternative hosts from other vegetation cover classes (Figure 3).

\[
X_{new} = \left( X - X_{min} \right) / \left( X_{max} - X_{min} \right)
\]

**Formula 2** Variable standardization formula
Where $X$ is the value of the pixel being standardized, $X_{\text{min}}$ is the minimum value while $X_{\text{max}}$ is the maximum value of the layer. Each variable was resampled to 100 meters spatial resolution.

2.3. Methods

MLN disease severity is the percentage of relevant host tissues covered by MLN symptom or lesion or damaged by the disease [50]. Severity results from the number and size of the lesions. MLN disease severity is an indicator of damage caused by the disease.

$\text{MLN disease severity} = \left( \frac{\text{sum of all disease rating}}{\text{total number of rating} \times \text{maximum disease grade}} \right) \times 100.$

**Formula 1** Crop disease severity formula

A standardized 1-5 MLN disease rating developed by CIMMYT [51].was used to derive MLN disease severity. MLN survey was conducted in electronic format using handheld GPS equipment [51] with an aim of collecting MLN disease ratings over sampled crop fields that would be used to derive MLN disease severity. Sampling enabled choosing which subjects to measure while ground referencing enabled finding and measuring the subjects in question [52]. Ground reference MLN data would generally not be collected for large portions of the entire project area therefore sampling was used [53]. The criteria considered for evaluating the suitability of MLN ground reference data was based on [53]. The criteria ensured that; the data collection method is systematic and representative of the entire area, the method has an element of randomness to avoid selection bias, a sufficient number of reference samples are utilized to provide an appropriate sample density and the reference data is reasonably contemporary with respect to the acquisition date of the spatial ecological variables. Proportionate stratified sampling technique was applied which randomly distributed MLN points across the sampling zone and across the spatially derived crop mask (Figure 3). The sampled MLN points with respective MLN severities were used to extract ecological variables from respective geospatial layers in Table 1 and subsequently used for spatial regression analysis.

The spatial regression model [54] involved; choosing a neighbourhood criterion, creating spatial weight matrix, examining spatial autocorrelation, applying weights matrix and predicting using the spatial regression model. Spatial error model was appropriate for this study particularly in correcting for spatial autocorrelation due to the use of spatial data [54]. The model included spatially correlated errors due to unobservable features or omitted variables associated with location. The error term $\epsilon$ had the spatial structure therefore incorporating spatial effects through error term.

$y = x\beta + \epsilon$

$\epsilon = \lambda W \epsilon + \xi$

**Formula 4** Spatial error model formula

Where:
- $W$ is the spatial weight matrix
- $x$ is the predictor variable
- $y$ is the response variable; MLN disease severity
- $\epsilon$ is vector of error terms, spatially weighted using the weights matrix ($W$)
- $\xi$ is vector of uncorrelated error terms
- $\beta$ is the spatial error coefficient

$\lambda$ is the spatial dependence parameter.

The spatial error model controlled spatial autocorrelation in the residuals, thus it controlled autocorrelation in both MLN severities and the ecological variables [55]. Moran’s $I$ statistic was used to test if MLN severity across the sampled farm fields had spatial dependence [54].

$I = \frac{N}{S_o} \left( e^T W e / e^T e \right)$

**Formula 5** Moran’s $I$

$S_o$ is a standardization factor that corresponds to the sum of weights for the non-zero cross-products: $S_o = \sum \sum w_{ij}$. For row-standardized weights $S_o$ would equal $N$, so $I = \left( e^T W e / e^T e \right)$. The spatial weight matrix provided information about which MLN affected farm fields were considered neighbors and also how their MLN severities are related to each other [54]. The spatial weight matrix was defined as $W$ with elements $w_{ij}$ indicating whether observations $i$ and $j$ are spatially close. The spatial weight matrices were row-standardized which meant the weights were summed up to one on each row. The weight matrix was constructed based on distance where units with distance $d_{ij}$ received a weight that is inversely proportional to the distance between the units and 0 if they were beyond a certain distance band $D$ [54].

$w_{ij} = \begin{cases} 1/d_{ij} & \text{if the distance between } i \text{ and } j < D, \\ 0 & \text{otherwise} \end{cases}$

**Formula 6** Spatial weight matrix based on distance

MLN severity field points were split into a 80:20 sample with 120 training points and 80 test points. The spatial model was built on the 80% training data and subsequently the model built was used to predict MLN disease severity on test data. Through predicting a fitted model to a raster data, MLN severity prediction was derived with spatial error model object using raster object with predictor ecological variables [56]. By calculating accuracy measures and error rates, the prediction accuracy of the spatial regression model was determined. A correlation between actual MLN disease severity and predicted MLN disease severity formed the basis for measurement of accuracy. Neither $R^2$ nor adjusted $R^2$ was utilised in assessing the spatial regression model therefore Akaike Information Criteria (AIC) was used [57].

**Figure 4.** Derived MLN field points
3. Results

Based on distance between MLN affected farm fields, the spatial regression model predicted MLN disease severity for each farm field based on distance where units within a specified radius were assigned a spatial weight. It is evident that MLN severity becomes less correlated the further away from each affected farm field. Varying nearest neighbour distances 1,3,4,7,21,79,81,83,84,89,97 and 100 were noted to have significant (p<0.05) correlations with MLN severity.

Moran’s I test tested if the ecological variables have spatial dependence. The test confirmed that there is significant (p<0.05) spatial dependence among ecological variables. This implies that MLN affected farm fields that are close to each other influence each other more in as far as MLN severity is concerned. It is observed that MLN severities collected are not independent, but rather positively spatially dependent, which means that MLN severity from one farm location tend to exhibit values similar to those from nearby farm locations. Simply, farms located nearby tend to have similar MLN severities than those separated by larger distances. A normal Moran’s I test plot should have a 45 degrees line i.e. a perfect prediction. This is not the case as a result of spatial dependence. Figure 6 suggests that an actual MLN severity value $x$ would be predicted to value $x1$ using the spatial values derived from the ecological variables.
MLN disease severity was used as the response variable while temperature, rainfall, slope and soil moisture were used as predictor variables. A spatial regression model was used to establish geostatistical significance between MLN disease severity and the four ecological response variables. The probability level of each variable was set at $p < 0.05$ with soil moisture, rainfall and slope as the most significant ecological variables that explained MLN disease severity. From spatial error model estimates, soil moisture, rainfall and slope variables have a negative influence on MLN severity while temperature has a positive influence, with one unit increase in temperature increasing MLN severity by 0.0004%, one unit increase in rainfall decreasing MLN severity by 0.03%, one unit increase in soil moisture decreasing MLN severity by 0.4245 and lastly one unit increase in slope decreasing MLN severity by 0.0492%. Collinearity test reveals that none of the ecological variables is correlated with the other. This is explained by the values of variance inflation factor (VIF); which quantifies the severity of collinearity, computed for each variable. The values are lower than 10 for each predictor variable. In addition the average VIF value is 1.611, which is closer to 1 indicating that collinearity is not a problem for the model.

MLN severity prediction was derived with spatial error model object using raster object with three most significant predictor ecological variables. The raster data was re-classified into four MLN severity classes using the Jenks natural break algorithm, which maximizes the variance in the data for subsequent classification. Figure 7 shows per pixel MLN severity map over the study region derived from amalgamating and weighting the three most significant ecological variables. High severity areas are illustrated in dark brownish colors, while low severity areas, colored light brownish, illustrate low MLN severity areas. High MLN severity areas were found to be in Chebunyo, Sigor and Kipreres.

Residuals from the spatial error model were mapped in order to look for geographical patterns that, if they exist, this would violate the assumption of independent errors and potentially affect both the estimate of the model coefficients and their standard errors. Residuals are the differences between MLN severity predicted by the model and actual observed MLN severity. The mapped residuals explain the degree at which the residual value for any MLN affected farm field correlate with the mean residual value for its nearest MLN affected farm field. The mapped residuals from the spatial error model display evidence of positive spatial autocorrelation. The geographical pattern shows significant correlation between the residual value at any one farm field and that of its nearest neighboring farm field. This is confirmed by the global Moran $I$ test for regression residuals which is significant ($p<0.05$).

The spatial error model has an Akaike’s Information Criterion of -641.6, R squared at 80.78%, a mean absolute percentage error (MAPE) of 51.52%, minimum-maximum accuracy of 95.06% and a correlation between actual MLN severity and predicted MLN severity at 92.62%.

5. Discussions and Conclusions

MLN is attributed to three ecological variables namely soil moisture, rainfall and slope. This study has demonstrated the strengths and merits of monitoring and predicting MLN from ‘space’ using spatially explicit ecological variables and spatial modelling techniques. From spatial
regression analysis, significance of soil moisture, rainfall and slope as key variables for MLN disease severity suggests that monitoring these ecological variables may be important in regard to predicting MLN outbreaks. From the spatial error model estimates, one unit increase in temperature increases MLN severity by 0.0004%, one unit increase in rainfall decreases MLN severity by 0.03%, one unit increase in soil moisture decreases MLN severity by 0.4245 and lastly one unit increase in slope decreases MLN severity by 0.0492%. On average, MLN attacks maize crops at 27.5 degrees Celsius, 85.6 millimetres rainfall, at an altitude of 1888.5 meters above sea level and at soil moisture levels of -1085.9 based on the topographical wetness index. From the spatial model, 441,394 acres of cropped fields had no MLN severity; 133,333 acres of crop fields had low MLN severity, 64,388 acres moderate severity while 9323 acres of cropped fields had high MLN severity. High MLN severity is largely experienced in Chebunyo, Sigor and Kipreres sub-counties and therefore MLN control measures should focus in these areas. The ecological variables have spatial dependence in that MLN affected farm fields that are close to each other influence each other more in as far as MLN severity is concerned.

Remote sensing coupled with spatial analysis is a good and readily available tool for MLN severity analysis as compared to conventional ground surveys. With the advancement in geostatistical models such as spatial regression, MLN can be attributed to its cause factors and spatial patterns established. Moreover, such studies help in epidemiological surveillance which is vital in developing any MLN control strategies. With the use of spatial analysis and remote sensing, such studies could be geared towards site-specific MLN control measures. MLN severity map can be utilized by multi-stakeholders to guide on location-based interventions, with emphasis on high severity areas. Farmers could be advised to plant maize at favourable MLN-free model-derived environmental conditions. At high-to-medium MLN severities, farmers are advised to break from planting maize for at least two planting seasons so as to avoid contamination. It is at these locations characterised by high-to-medium MLN severities that farmers are also advised to use MLN control measures such as pesticides to ward off MLN insect vectors, planting alternative crops such as sorghum, beans and millet and early planting at low-to-zero model-derived environmental conditions.

This study relied on MLN disease severity data as the response variable. Further research is required based on MLN presence and absence or any other relevant response variable using spatial and ecological models that account for spatial dependence and autocorrelation in order to generate comparable MLN risk maps for Bomet County. The study only shows 2017 MLN severity map derived from spatial error model hence there is need for spatiotemporal analysis to show MLN trends from the initial year of observation to date, for a clearer understanding of MLN. The study lacks human behavioural variables such as farm practices e.g. use of pesticides as a MLN severity mitigation measure. Therefore, there is need to do further research on farm practices as variables leading to MLN. Further studies should also investigate the role of other potential spatially derived ecological variables with relevant significance of statistical levels, better spatial and temporal resolutions in monitoring and predicting of MLN. The model employed in this study could be up-scaled and replicated to other MLN prone areas both in Kenya and in eastern southern Africa.

Acknowledgements

This paper is part of the results of the first author’s research project for fulfillment of the requirements for the award of a master’s degree in geospatial information systems and remote sensing from the Department of Geomatics Engineering and Geospatial Information Systems, Jomo Kenyatta University of Agriculture and Technology, Kenya. The authors are grateful to the Regional Center for Mapping of Resources for Development (RCMRD) and SERVIR Eastern and Southern Africa for the support (masters’ fellowship). Many thanks to the County government of Bomet for the field work support.

References

Journal of Geosciences and Geomatics


