Determining the Correlation between Abaca (Musa textilis) Distribution Patterns and Bunchy Top Disease Prevalence in the Caraga Region through Spatial Point Pattern and Statistical Analysis

Hazel Anne F. Animo
Department of Geodetic Engineering, College of Engineering and Geosciences, Caraga State University

Mitche R. Tajale
Department of Geodetic Engineering, College of Engineering and Geosciences, Caraga State University

Arnaldo C. Gagula
Department of Geodetic Engineering, College of Engineering and Geosciences, Caraga State University

Elizabeth P. Parac
Department of Plant and Soil Science, College of Agriculture and Agro-industries, Caraga State University

https://doi.org/10.5109/7323312

出版情報：Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES). 10, pp.532-537, 2024-10-17. International Exchange and Innovation Conference on Engineering & Sciences
バージョン：

Determining the Correlation between Abaca (*Musa textilis*) Distribution Patterns and Bunchy Top Disease Prevalence in the Caraga Region through Spatial Point Pattern and Statistical Analysis

Hazel Anne F. Animo[1], Mitche R. Tajale[1], Arnaldo C. Gagula[1,3], Elizabeth P. Parac[2]
[1]Department of Geodetic Engineering, College of Engineering and Geosciences, Caraga State University
[2]Department of Plant and Soil Science, College of Agriculture and Agro-industries, Caraga State University
[3]Caraga Center for Geo-informatics, Caraga State University
Corresponding author email: acgagula@carsu.edu.ph

Abstract: *Abaca (Musa textilis), a versatile and strong natural fiber endemic to the Philippines, presents an eco-friendly alternative to wood in pulp and paper production and is preferred over plastics and synthetic materials. Despite its benefits, the cultivation of Abaca faces significant challenges from Banana Bunchy Top Disease (BBTD), threatening the industry's viability. This study employs Spatial Point Pattern Analysis to investigate the distribution of Banana Bunchy Top Virus (BBTV), utilizes logistic regression to identify environmental factors affecting disease spread, and creates a hotspot map pinpointing municipalities at higher risk. The logistic regression analysis identifies precipitation, maximum and minimum temperatures, and wind speed as significant predictors of BBTV transmission in abaca plants. The hotspot analysis revealed that out of the 19 municipalities with abaca crops, Butuan City, Las Nieves, Bayugan, Esperanza, and Sibagat are high-risk areas for BBTV infection, underscoring the need for targeted disease management strategies in these regions.*

Keywords: Abaca, Banana Bunchy Top Disease, Spatial Point Pattern, Logistic Regression, Hotspot

1. INTRODUCTION

The Abaca plant (*Musa textilis*), a near relative of the banana and a member of the Musaceae family, is extensively dispersed in the Philippines' humid tropics. It is a feasible alternative to wood as a natural pulp and paper manufacturing resource. It is preferred over plastic and other synthetic materials, making it the "strongest natural fiber in the world," according to the Philippine Fiber Industry Development Authority [1]. Most of the time, mature rootstock pieces are planted for propagation at the beginning of the rainy season. Generally rich, loose, loamy soils with good drainage are preferable for the plants' growth [2].

Abaca bunchy top disease (ABTD), which is brought on by the BBTV carried by the brown aphid, *Pentalonia nigronervosa*, has significantly reduced agricultural productivity and caused enormous harm to abaca plantations [3]. ABTD belongs to the genus Babuvirus and is classified within the Nanoviridae family [4]. Bunchy-top disease in Abaca Plantation in Caraga has affected the production of abaca in the region.

Spatial Point Pattern Analysis plays a crucial role in understanding the relationship between variables, as one variable can positively or negatively impact another variable's existence and spatial distribution [5]. Over time, point process statistics for spatial structures has become a well-established discipline, gaining popularity among researchers [6]. Some approaches besides spatial point patterns include epidemiological and behavioral models, such as the study of Khan et al. for investigating disease dynamics [7]. Within plant ecology, spatial point-pattern analysis has a rich history and relies on an extensive collection of test statistics called summary statistics [8]. These statistics assess and describe point patterns' statistical properties and spatial structure [9].

In the Philippines, the Bunchy Top Disease (BTD) is a significant threat to the cultivation of Abaca (*Musa textilis*). The relationship between the distribution of the Abaca (*Musa textilis*) plants and the prevalence of the BTD remains ambiguous. Though studies and literature on various disciplines about Abaca (*Musa textilis*) and BTD already exist in the region, no studies still investigate the spatial relationship between them, thus posing a significant knowledge gap.

We aim to determine and explain the significant implications of the correlation between the Abaca (*Musa textilis*) distribution pattern and the BBTD in the Caraga region through Spatial Point Pattern and Statistical Analysis. Specifically, to examine the spatial arrangement of Abaca plants and identify any significant patterns with the occurrence of BBTD, identify the potential environmental factors influencing the spatial dynamics of the BBTD, and generate a prediction map depicting the potential spread of BBTD in Caraga region.

2. MATERIALS AND METHODS

2.1 Study Area

Caraga Region (Fig. 1) is a geographical area located in the northeastern part of Mindanao, with a total land area of approximately 2,147,845 hectares. It comprises five provinces: Agusan del Norte, Agusan del Sur, Surigao del Norte, Surigao del Sur, and Dinagat Islands. Its diverse topography includes a mountainous coastline along the Philippine Sea, such as the Diwata Mountain Range, and vast river systems, most notably the Agusan River Basin. The region is rich in valuable natural resources, including minerals like gold and nickel, and it is home to diverse and thriving ecosystems. Moreover, agriculture is crucial to the economy of the region, with a focus on the production of rice, corn, and coconuts.

2.2 Data Collection

The research began by gathering a comprehensive dataset comprising spatial and non-spatial information on the study region's abaca points, BBTV, and municipal boundaries. Simultaneously, non-spatial data was carefully compiled, capturing relevant environmental factors believed to impact virus transmission of BBTV.

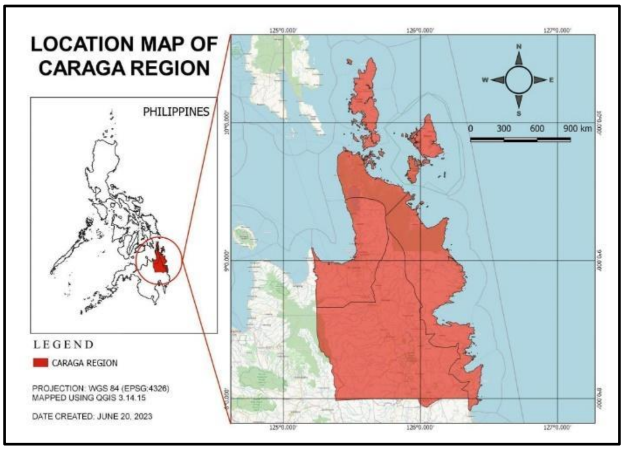

Fig. 1. Map of the Study Area

Three hundred ninety-seven (397) abaca location points were surveyed and analyzed in a molecular biology laboratory using the PCR method. Table 1 shows the identification of abaca locations and the counts of total and infected abaca points. This information was obtained by overlaying Municipal boundaries on a map and employing the spatial join tool in GIS.

Table 1. Number of Abaca location and BBTV-infected Abaca locations

Municipal Name	Number of Abaca location points	Number of BBTV-infected Abaca location points
Gigaquit	21	0
Lianga	22	0
Mainit	22	0
San Agustin	20	0
San Miguel	24	0
Sison	20	0
Tubay	20	0
Kitcharao	18	1
Buenavista	20	2
Talacogon	20	4
Butuan City	13	6
RTR	7	6
Sibagat	21	7
Cabadbaran City	20	10
Las Nieves	20	10
Tandag City	20	11
San Luis	30	12
Esperanza	21	13
Prosperidad	38	15
Total	**397**	**97**

Meanwhile, Table 2 shows the variables used for regression analysis. Precipitation, maximum and minimum temperature, wind speed and direction, river networks, road networks, building footprints, SAR DEM, and population were the nine publicly available environmental datasets that were thought to have a plausible and significant relationship with BBTV. These datasets were selected as covariates in a logistic regression model.

Table 2. Covariates used for Regression Analysis

Datasets	Sources
Precipitation Maximum Temperature Minimum Temperature Wind Speed and Direction	National Aeronautics and Space Administration (NASA)
River Networks Road Networks Building Footprints Synthetic Aperture Radar Elevation Model	Caraga Center for Geo-informatics (CCGeo)
Population Data	Philippine Statistics Authority (PSA)

2.3 Spatial Point Pattern Analysis

The research design incorporates spatial point pattern analysis techniques to analyze the distribution patterns of BBTV using Ripley's K Function in ArcGIS software. It assesses whether there is clustering, dispersion, or randomness in the distribution of BBTV values, considering both the values of the variable at individual locations and the spatial relationships between them.

Once the spatial data of abaca points and BBTV cases within the study area is acquired, the next step involves formatting the data appropriately. This involves structuring the information into separate layers, such as point shapefiles for both BBTV occurrences and abaca plant locations, along with a polygon shapefile outlining the municipal boundaries of the Caraga region. Subsequently, a new polygon shapefile was generated, encompassing only the municipalities where the abaca point samples are situated. This new shapefile contains information regarding the number of BBTV cases within each municipal polygon.

2.4 Logistic Regression Analysis

2.4.1 Data Preparation

The weather covariates from NASA were acquired in point shapefiles with corresponding spatial and non-spatial values using the GIS software, these point shapefiles were loaded and projected into WGS 84 UTM Zone 51N. Subsequently, the Inverse Distance Weighting (IDW) interpolation technique was used to get the estimated values for the entire area, like Metran et al. [10] did in their study on mapping and estimating road traffic emissions. After acquiring the estimated values, data from both uninfected and infected abaca samples were added to the interface. Utilizing the "extract multi values to points" tool, the weather attributes, elevation, and population of each point where the uninfected and infected abaca samples were obtained.

On the other hand, Euclidean distance was used to obtain the distance from each point to the roads, rivers, and built-up areas. The same process involves the "extract multi values to points" tool to get the distance values of each point. Shown in Figure 2 are the maps of predictors together with the location of abaca location points.

2.4.1 Logistic Regression Model

The IBM SPSS Statistics software analyzes the correlation between the BBTV and the nine independent variables. Binary logistic regression was utilized, with the binary outcome variable representing the presence or absence of BBTV infection. First, the corresponding values of 397 points from nine independent variables were added to the software, and a new table was created containing the values 0 for uninfected and 1 for infected. This table is used as the dependent variable in the regression analysis.

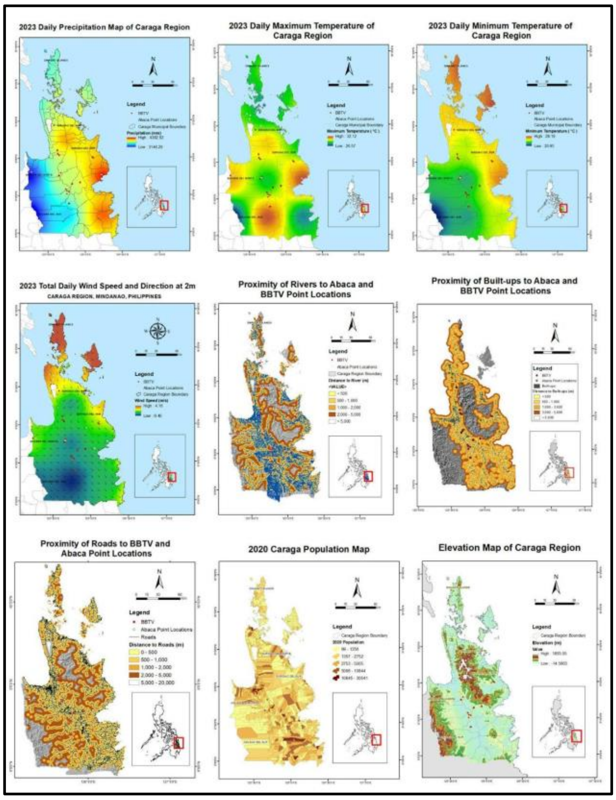

Fig. 2. Maps of BBTV and Abaca Location Points with the Predictors

Initially, all covariates were entered into a binary logistic regression model. A significance level screening was undertaken, wherein variables with a significance level (p-value) below 0.25 were retained. This relatively lenient threshold allowed for an inclusive selection process in the initial screening phase. The refinement process continued using multinomial logistics. The remaining variables were entered again, and p-values equal to or greater than 0.05 were iteratively removed from the model. This step ensured that only statistically significant predictors, with p-values below the conventional threshold of 0.05, were retained in the final model.

The coefficients of the final model were interpreted to understand the relationships between predictors and the probability of BBTV infection. Subsequently, the model was utilized to predict the likelihood of BBTV infection at various locations based on the values of the significant predictors.

2.4 Data Analysis and Interpretation

The regression analysis results identified which abaca points were uninfected and infected. Those predicted data points were extracted in ArcGIS software, and a hotspot analysis was performed to know which abaca points have a higher risk of being infected with BBTV. Shapefiles delineating high-risk and no-risk areas were obtained, establishing distinct spatial boundaries for analysis. Zonal statistics were then computed within these delineated zones using Geographic Information System (GIS) software, allowing for the calculation of summary statistics such as mean, median, standard deviation, and range for each significant environmental parameter. The resultant zonal statistics facilitated a quantitative comparison between high-risk and no-risk areas, providing insights into the spatial distribution and variability of environmental conditions across different risk areas. Finally, a comparative analysis was conducted to identify significant differences in environmental parameters between high-risk and no-risk areas.

3. RESULTS AND DISCUSSIONS

3.1. Abaca Location Points and BBTV Distribution Mapping

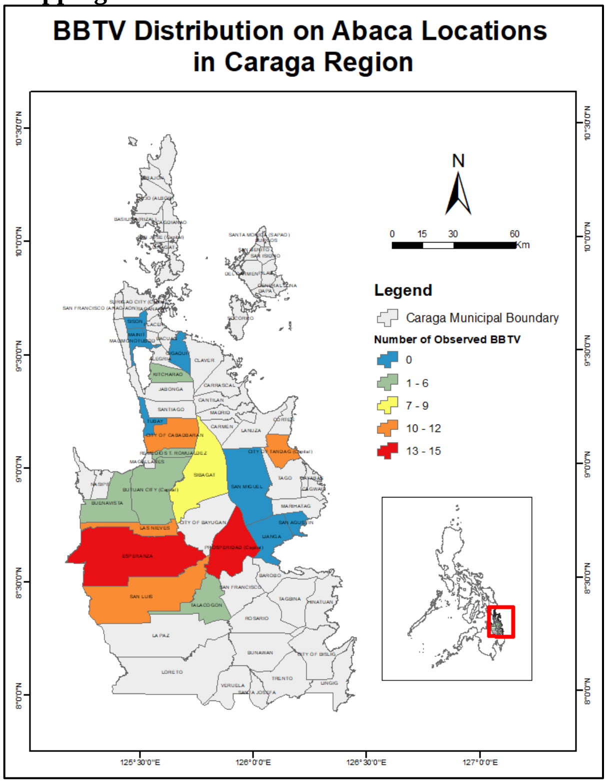

Fig. 3. Map of BBTV Distribution on Abaca Location Points in Caraga Region

Figure 3 shows the mapped distribution of Abaca point locations in the Caraga Region. Only the highlighted municipal boundaries were considered since they have existing Abaca plantations. Those not highlighted represent 0 abaca point locations and are hence disregarded as quadrats as they can skew statistical analyses and may not accurately represent the potential for disease spread.

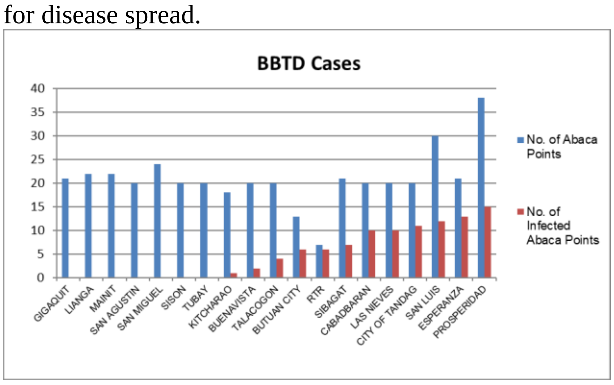

Fig. 4. Number of BBTD Cases on Abaca Location Points in the Caraga Region

Figure 4 shows the number of infected abaca points across 19 municipalities with existing abaca location points. The data presents a significant disparity in the number of infected points across the selected municipalities; while some areas appear to have no infections, Prosperidad (Capital) and San Luis show higher infection rates. Also, Kitcharao, Buenavista, and Talacogon municipalities exhibit moderate infection levels. The figure above shows that municipalities like Butuan City and Remedios T. Romualdez report higher infection rates despite having fewer abaca points. These regional disparities indicate that urban factors could contribute to the rate of BBTV infection. This initial observation is essential for providing insights into the various independent factors that could affect the spread of BBTV.

3.2. Spatial Point Pattern Analysis using Ripley's K Function

Figure 5 shows the K function plot used to analyze the spatial distribution of points. It shows that the red line lies above the blue line and is outside the confidence envelope, signifying significant clustering. Further, this indicates that the data points tend to cluster together spatially, implying a non-random distribution pattern within the dataset clustering phenomenon.

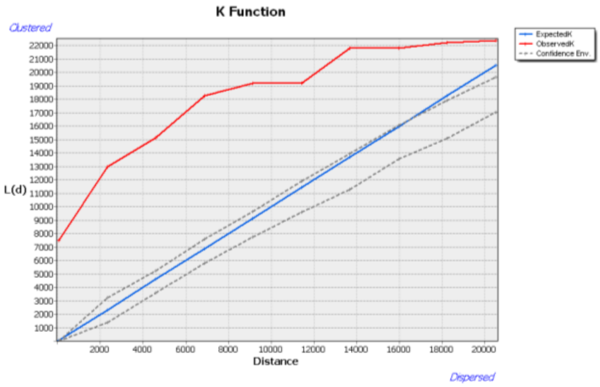

Fig. 5. Ripley's K function result

3.3 Logistic Regression Analysis

Table 3 shows the results of the Multinomial Logistic Regression analysis on nine covariates that were identified to have a potential influence on the occurrence of BBTV in abaca plants. It is to be considered that the significance threshold set for this analysis is $p < 0.25$. Precipitation with a p-value of 0.007 shows a significant positive correlation with BBTV infection, implying that increased rainfall marginally raises the likelihood of infection. Minimum temperature with a p-value of 0.006, though negatively associated, also significantly influences BBTV infection. On the other hand, maximum temperature, though not significant at a traditional 0.05 level ($p = 0.061$), still shows a strong potential relationship considering the initial threshold $p < 0.25$. Similarly, on the borderline of significance, Built-ups, Roads, and Wind Speed with significant values of p=0.098, p=0.078, and p=0.030, respectively, indicate a slight increase in infection risk.

On the other hand, variables such as River Proximity, Population, and Elevation, with their significant values p=0.506, p= 0.710, and p=0.781, respectively, do not meet the set significance level of $p < 0.25$. Hence, these three variables show negligible coefficients and are to be excluded from further backward stepwise regression analysis.

Table 3. Multinomial Logistic Regression of Covariates considering p<0.25 significance.

	B	S.E	Wald	df	Sig.
Precipitation	0.012	0.005	7.187	1	0.007
MaxTemp	1.509	0.806	3.504	1	0.061
MinTemp	-4.799	1.742	7.591	1	0.006
River	0.000	0.000	0.443	1	0.506
Built_ups_	0.000	0.000	2.736	1	0.098
Population	0.000	0.000	0.138	1	0.710
Elevation	0.000	0.002	0.077	1	0.781
Roads	0.001	0.000	3.115	1	0.078
WindSpeed	5.812	2.678	4.709	1	0.030
Constant	16.695	6.735	6.145	1	0.013

Table 4. Multinomial Logistic Regression of six Covariates of p < 0.05 significance

	B	S.E	Wald	df	Sig.
Precipitation	0.011	0.004	7.568	1	0.006
MaxTemp	1.256	0.614	4.187	1	0.041
MinTemp	-4.197	1.385	9.184	1	0.002
WindSpeed	4.850	2.083	5.422	1	0.020
Built_ups	0.000	0.000	3.352	1	0.067
Roads	0.001	0.000	2.834	1	0.092
Constant	15.617	6.069	6.622	1	0.010

Based on Table 4, Precipitation with a p-value of 0.006 meets the set significance level of p<0.25 and stands out as a crucial factor, indicating that it can be a significant predictor of BBTV infection. The coefficient (B=0.011) suggests that for each unit increase in precipitation, there is a 1.1% rise in the likelihood of BBTV infection. The narrow confidence interval for the odds ratio (Exp(B)=1.011) confirms the reliability of this finding. Additionally, maximum temperature (p=0.041) and wind speed ($p = 0.020$) also emerges as a significant predictor, implying that with higher values of these variables increase the likelihood of BBTV infection. Contrarily, minimum temperature demonstrates a strong negative relationship with BBTV infection (p=0.002), indicating that higher minimum temperatures significantly decrease the odds of infection. Also, factors such as Built-ups and Roads exhibit p-values slightly above the significance threshold (p=0.067 and p=0.092, respectively), hinting a potential but non-significant or has the least influence on BBTV infection.

With this, the proximity to roads covariate is considered insignificant based on the backward stepwise regression. It exhibited the farthest value above the significance level among the variables considered. Given the result, it shall be excluded from further analysis.

Table 5. Multinomial Logistic Regression (Backward Stepwise Method)

Model	Effect(s)	Effect Selection Tests		
		Chi-Square	df	Sig.
Step 0	<all predictors entered>	.		
Step 1	Built_ups (removed)	1.953	1	0.162

Table 5 shows the multinomial logistic regression analysis results using a backward stepwise method. Step 0 indicated the included significant predictors from the previous analysis, excluding Roads. Step 1 presents the removal of the Built_ups predictor, measured by a chi-square value of 1.953 with 1 degree of freedom. Also, the p-value (0.162) indicates that excluding the said predictor does not significantly worsen the model fit, hinting that it may not be an essential predictor.

Table 6. Multinomial Logistic Regression of the remaining Covariates at $p < 0.05$

	B	S.E	Wald	df	Sig.
Precipitation	.014	.004	12.984	1	<.001
MaxTemp	1.291	.577	5.006	1	.025
MinTemp	-4.674	1.317	12.590	1	<.001
WindSpeed	5.264	1.949	7.292	1	.007
Constant	15.317	5.892	6.757	1	.009

Still, precipitation appears as a crucial factor (p<0.001), with each incremental increase linked to a 1.4% rise in infection odds, emphasizing the role of rainfall patterns in BBTV transmission. Maximum temperatures (p=0.025) hint that temperature also impacts virus transmission. Furthermore, a higher minimum shows a robust negative relationship with BBTV infection (p<0.001), explicitly suggesting a decrease in infection odds. Wind speed also demonstrates significance (p=0.007), associating higher speed with a substantial increase in the likelihood of BBTV infection. These results underline the multifaceted nature of BBTV transmission and offer further insights targeted for effective disease management strategies in abaca cultivation nationwide.

3.3. Model Validation

Table 7. Goodness-of-fit tests

	Chi-Square	df	Sig.
Pearson	274.255	286	0.681
Deviance	317.150	286	0.099

As indicated in Table 7, the Pearson and Deviance goodness-of-fit tests yield p-values above the significance level of 0.05, indicating a failure to reject the null hypothesis and suggesting that the observed data fits the expected distribution.

3.4. Covariates Analysis Results and Interpretation of Infections

Out of the 19 municipalities with existing abaca location points, only 3 were identified as high-risk areas for BBTV infection, as shown in Figure 6. These municipalities are Las Nieves, Sibagat, and Prosperidad. Out of 300 abaca locations with no detected BBTV occurrence, fourty-four (44) abaca locations were predicted to be infected by BBTV, twenty (20) abaca points in the Municipality of Las Nieves, specifically in Barangay Bonifacio. In the Municipality of Sibagat, twenty-one (21) points were pinpointed at heightened risk of BBTV infection, and three (3) from the Municipality of Prosperidad. This is based on the results of regression analysis coupled with GIS analysis through statistical analysis.

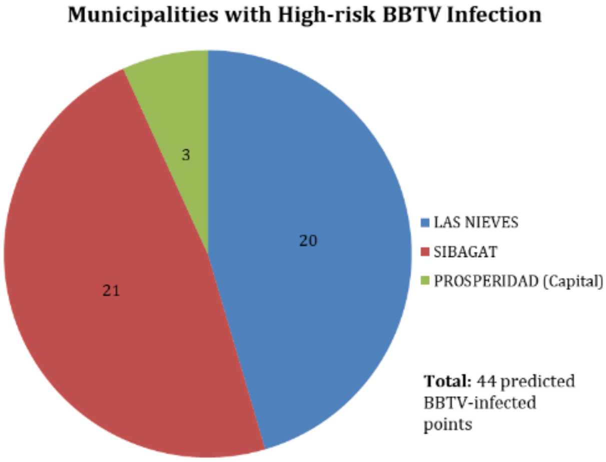

Fig. 6. Municipalities with high-risk BBTV infection using statistical analysis.

3.4. Hotspot Mapping

Another prediction tool was hotspot mapping. Figure 7 represents the geographical distribution of BBTV risk across the Caraga region. Areas delineated on the map and marked as hotspots signify regions with a higher risk of BBTV incidence, designating areas where the virus is not only present but possibly thriving. The created hotspot region includes 55 barangays, which are part of five (5) municipalities, namely Butuan City, Las Nieves, Bayugan, Esperanza, and Sibagat were identified as high-risk areas based on the mapping and hotspot analysis of the 44 predicted points that are predicted to be infected by BBTD in consideration of the four (4) predictors which are the Precipitation, Minimum and Maximum Temperature, and Wind speed. Seven (7) barangays were identified as high-risk areas in Butuan City, fourteen (14) barangays in Las Nieves, thirteen (13) from the City of Bayugan, eight (8) from Esperanza, and thirteen (13) barangays from the municipality of Sibagat. Though situated in different municipalities, these barangays exhibit notably high Z-scores, which indicates that they possess the unique values of the identified predictors (Precipitation, Windspeed, Minimum, and Maximum Temperature) that can substantially contribute to the increase in the likelihood of getting infected by BBTV.

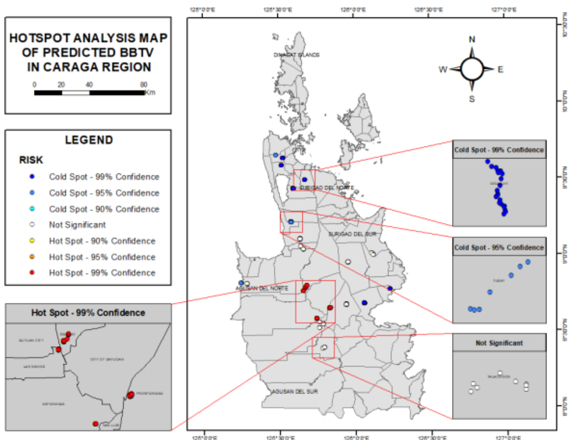

Fig 7. Hotspot Map of Predicted BBTV in Caraga Region

4. CONCLUSIONS AND RECOMMENDATIONS

This study successfully mapped and analyzed the spatial distribution of Banana Bunchy Top Virus (BBTV) across the Caraga region's abaca plantations.

Out of the data of confirmed locations and the BBTV occurrences, an analysis of GIS using Ripley's K function shows significant clustering behavior. This indicates that the data points cluster spatially. Multinomial Logistic Regression Analysis results state that covariates precipitation, maximum and minimum temperatures, and wind speed were identified predictors with $p < 0.05$, $p<0.25$, $p < 0.05$ significance. The model was validated using Pearson and Deviance goodness-of-fit tests, which show p-values above the significance level of 0.05, indicating a failure to reject the null hypothesis and suggesting that the observed data fits the expected distribution.

Binary Logistic Regression highlights three municipalities, namely, Properidad, Las Nieves, and Sibagat, with 3, 21, and 20 predicted abaca locations to be infected by BBTV. The hotspot map reveals five (5) municipalities with a high probability of risk based on the 44 recently predicted abaca locations to be infected by BBTV. These municipalities are Butuan City, Las Nieves, Bayugan, Esperanza, and Sibagat.

By identifying specific areas with higher infection rates, the map allows for targeted interventions and resource allocation to manage and mitigate the spread of BBTV. Binary logistic regression and hotspot mapping not only highlight the regions most affected by the disease but also enhance the understanding of the spatial dynamics of BBTV and support more effective disease management strategies to protect abaca production in the region. Effective management strategies, including targeted eradication of infected plants and vector control practices, should be prioritized in these regions to safeguard banana crops and the livelihoods of local farming communities.

The researchers also suggest incorporating additional environmental variables into their analysis. This recommendation stems from the recognition that the spread of BBTD is influenced by many factors, many of which may have yet to be accounted for in the initial selection of independent variables.

5. REFERENCES

[1] The Philippine Consulate General Website: https://www.vancouverpcg.org/vpcg-resources/yamang-pinoy/abaca/the-philippine-abaca-industry/ (accessed 31.03.23).

[2] J. M. Nuñez, Social Impacts of the Abaca Bunchy Top Disease and Adaptive Strategies of Farm Households: A Case in Leyte, Philippines, Annals of Tropical Research. 35 (2) (2013) 74–82.

[3] M. B. Dale, Spatial Pattern Analysis in Plant Ecology, (2001) 283-284.

[4] D. Stoyan, F. J. Rodríguez-Cortés, J. Mateu, W. Gille, Mark variograms for spatio-temporal point processes, Spatial Statistics. 20 (2017) 125–147.

[5] G. L. W. Perry, B. P. Miller, and N. J. Enright, A comparison of methods for the statistical analysis of spatial point patterns in plant ecology, Plant Ecology 187 (1) (2006) 59–82

[6] Routledge Website: https://www.routledge.com/Handbook-of-Spatial-Point-Pattern-Analysis-in-Ecology/Wiegand-Moloney/p/book/9780367576233 (accessed 21.06.23)

[7] M. U. R. Khan, M. R. Arefin, and J. Tanimoto, Investigating vaccination behavior and disease dynamics of a time-delayed two-strain epidemic model: An evolutionary approach, Proceedings of International Exchange, and Innovation Conference on Engineering & Sciences (IEICES). 8 (2022) 147-154.

[8] A. D. Raymundo and I. Pangga, Simulation Modeling of Bunchy Top Epidemics in a Changing Climate, Journal of Environmental Science and Management. 14 (2) (2011)

[9] M. D. Anhalt, R. P. P. Almeida, Effect of Temperature, Vector Life Stage, and Plant Access Period on Transmission of Banana bunchy top virus to Banana, Phytopathology. 98(6) (2008) 743–748.

[10] B. M. Metran, J. P. Elumba, M. C. Alvarado, M. M. Bermoy, Mapping and Estimation of Road Traffic Emissions in Caraga Region using Sentinel 5P Satellite Images, Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES). 9 (2023) 6-12.