Spatial Mapping and Modeling of Reported Dengue Incidences in Luzon

Nicole M. Hernandez, Carla Katrine P. Lucero, Dianne C. Yumol
and Jericho Manuel A. Zulueta
De La Salle University Integrated School, Manila

Shirlee R. Ocampo
De La Salle University, Manila

Abstract: Dengue, the most rapidly spreading mosquito-borne viral infection, has significantly spread worldwide in recent decades - flourishing both in affluent and impoverished locations of tropical and subtropical countries. In 2012, the Philippines ranked fourth out of the ten Association of the Southeast Asian Nations (ASEAN) countries in having the highest number of dengue cases. The following study intends to analyze the spatial distribution of dengue incidences across all Luzon provinces in 2018. It aims to determine significant correlates that affect dengue incidences, map the incidence rate of dengue cases, and explore the clustering of recorded dengue cases. Poisson and Negative Binomial (NB) regression analyses and Multiple Linear Regression Models (MLRM) were applied to determine the significant correlations affecting dengue incidence rates. Simultaneously, spatial mapping was utilized to visualize and detect clustering in the provinces through dengue count, incidence ratios, and standard incidence ratios (SIR). MLRM and NB showed that rainfall and poverty incidence are significant correlates of dengue counts and incidence, and Nueva Ecija and Tarlac were observed to be provinces with distinct dengue count and SIR greater than 1, as well as provinces found in clusters. With the provided results, health organizations can provide health programs and allocate more funds in areas with SIR greater than 1 to prevent dengue spreading.

Key Words: spatial mapping; regression modeling; dengue incidence; correlation; Aedes aegypti

1. INTRODUCTION

Dengue is the most quickly spreading mosquito-borne virus in the world. According to WHO (2020a), dengue's global incidence rate has increased in recent decades, wherein about half of the world's population is now at risk, with 100-400 million infections occurring each year. In 2016, more than half of the 375,000 reported cases in the Western Pacific region were solely from the Philippines.

In the first half of 2019, the Philippines experienced a dengue outbreak which accounted for 146,062 dengue cases and 600 casualties. By the end of 2019, the total cases went up to 429,409 - which is higher compared to 241,707 cases revealed during 2018 (WHO, 2020b). Studies on dengue incidence usually delve into environmental factors in a particular region. However, factors such as dengue count per province have not been recorded in previous papers, making it crucial to address other contributing factors affecting dengue incidence.

Therefore, the following study aims to identify the significant correlates that affect dengue incidences across Luzon's provinces in 2018, map the percentages of dengue incidence rates in Luzon, and explore the clustering of recorded dengue cases to identify high-risk areas on a broad geographical scale. The study will focus on the spatial mapping and regression modeling of dengue across all provinces of Luzon.

2. METHODOLOGY

The complete available DENC data set from the provinces of Luzon in 2018 was requested and utilized from the Department of Health (DOH) Regional Offices by coordinating with their epidemiology units. Data for the variables relative humidity (RH), rainfall (RF), and temperature (TMIN, TMAX, and TMEAN) were obtained from PAG-ASA Synoptic Stations; and population (POP) and poverty incidence (POVINC) were retrieved from DOH and the Philippine Statistics Authority; all of which were summarized upon collection. Microsoft Excel, Statistica, SAS, and GeoDa were utilized to perform Standardized Incidence Ratio (SIR), MLRM, Poisson and NB regression models, and Spatial Mapping respectively.

2.1. Theoretical Framework and Analysis

2.1.1. Standardized Incidence Ratio (SIR)

SIR determines if the occurrence of a disease in a relatively small population is high or low. It predicts if the number of observed cases in a particular geographic area, i = 1,2,..., m, is higher or lower than expected (Natividad et al., 2019).

The common risk, r, or incidence rate of a region is computed by

r=yN
where,
y= total count of the disease, and
N= total population exposed to risk in the region
SIR is obtained by:
SIRi = yiei
where,
yi= observed number of cases
ei= expected number of cases
The expected number of cases, ei, can be calculated using indirect standardization
ei=rNi
where,
Ni= the population in the region
r = common risk

2.1.2. Multiple Linear Regression Model

MLRM distinguishes the relationship between two or more independent variables and a dependent variable. It determines if changes in the independent variables can alter the dependent variable and approximate the variables' association. The dependent variable Yi is considered to be a function of p-1independent variables, X1i, X2i, ...,X(p-1)i,and their association can be written as:

Yi=0+1X1i+2X2i+...+βp-1X(p-1)i+i i=1,2,...,n
where,
Yi= the value of the dependent variable of observation i,
X1i, X2i, ...,X(p-1)i= the values of p-1predictors or independent variables, and
i = corresponds to the random error.

2.1.3. Poisson Regression Model

The Poisson regression model is used when Yiis a count variable. It assumes that the response variable is following Poisson distribution with parameter , Yi~Poisson(λ).

E[Y]=e0+1X1+...+p-1Xp-1=eX'β
where,
Y is the dependent variable,
is the p x 1vector of regression parameters
Xis a p x 1vector of independent variables.

2.1.4 Negative Binomial Regression Model

The NB model has the same mean structure as Poisson regression and is considered a generalized Poisson regression model. It contains an extra parameter to model overdispersion - narrowing the confidence intervals. NB is expressed as:

fY(y)=Γ(y+r-12-p)Γ(r-12-p)Γ(y+1)(r-12-pμ+r-12-p)r-12-p(1-r-12-pr-12-p)y   y=0,1,2,...
where,
i= exp{Xiβ}
r= dispersion parameter
(.)= usual gamma function

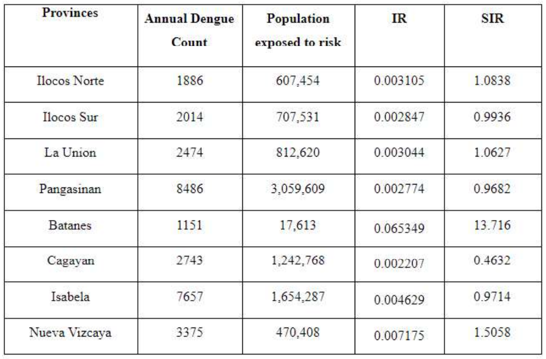
The following distribution has a mean equal to i and a variance equal to i+rip.

3. RESULTS AND DISCUSSION

3.1. Descriptive Statistics and Standardized Incidence Ratio (SIR)

The SIRs of dengue for the provinces of Luzon for 2018 were computed and summarized. In Table 3.1, 18 provinces and one district have an SIR greater than 1, implying that more dengue cases were observed than expected. Batanes had the highest relative risk of infection, followed by Apayao and Quirino. Meanwhile, the remaining 17 provinces and three districts experienced lesser cases of dengue. Clustering can be observed in provinces shaded in the same color, implying that the nearness of the values between dengue cases, incidence rates, and SIR of nearby provinces display a high dengue risk.

Table 3.1. Dengue Count and the Population Exposed to Risk in Each Province with Corresponding Incidence Ratios and Standard Incidence Ratios



Provinces	Annual Dengue Count	Population exposed to risk	IR	SIR
Ilocos Norte	1886	607,454	0.003105	1.0838
Ilocos Sur	2014	707,531	0.002847	0.9936
La Union	2474	812,620	0.003044	1.0627
Pangasinan	8486	3,059,609	0.002774	0.9682
Batanes	1151	17,613	0.065349	13.716
Cagayan	2743	1,242,768	0.002207	0.4632
Isabela	7657	1,654,287	0.004629	0.9714
Nueva Vizcaya	3375	470,408	0.007175	1.5058

Table 3.1. (cont.) Dengue Count and the Population Exposed to Risk in Each Province with Corresponding Incidence Ratios and Standard Incidence Ratios

Quirino	2137	196,119	0.010896	2.2869
Aurora	917	221,966	0.004131	1.5923
Bataan	2409	804,530	0.002994	1.1541
Bulacan	6476	3,515,504	0.001842	0.7100
Nueva Ecija	7062	2,268,553	0.003113	1.1999
Pampanga	6329	2,772,276	0.002283	0.8799
Tarlac	6657	1,420,364	0.004687	1.8065
Zambales	940	864,368	0.001087	0.4192
Batangas	2,757	2,887,957	0.000955	0.5288
Cavite	6,948	4,051,031	0.001715	0.9500
Laguna	5,647	3,258,735	0.001733	0.9598

Table 3.1. (cont.) Dengue Count and the Population Exposed to Risk in Each Province with Corresponding Incidence Ratios and Standard Incidence Ratios

Masbate	286	926,118	0.000309	0.5253
Sorsogon	512	823,469	0.000622	1.0577
Abra	1,095	244,800	0.004473	1.2543
Apayao	1,866	122,978	0.015173	4.2549
Benguet	1,344	832,635	0.001614	0.4526
Ifugao	851	209,611	0.004060	1.1385
Kalinga	1,582	219,077	0.007221	2.0250
Mountain Province	315	154,814	0.002035	0.5706
NCR-D1	29,618	2,899,200	0.001991	0.9060
NCR-D2	13,367	5,274,265	0.002534	1.1536
NCR-D3	5,802	3,015,062	0.001924	0.8759
NCR-D4	4,678	2,292,509	0.002041	0.9288

Table 3.1. (cont.) Dengue Count and the Population Exposed to Risk in Each Province with Corresponding Incidence Ratios and Standard Incidence Ratios

Quezon	3,841	2,202,097	0.001744	0.9661
Rizal	6,850	3,132,745	0.002187	1.2111
Marinduque	168	238,317	0.000705	0.2883
Occidental Mindoro	1,155	507,630	0.002275	0.9304
Oriental Mindoro	1,891	878,330	0.002153	0.8804
Palawan	3,787	1,170,905	0.003234	1.3225
Romblon	573	297,808	0.001924	0.7868
Albay	306	1,362,255	0.000225	0.3821
Camarines Norte	551	606,997	0.000908	1.5442
Camarines Sur	1,829	2,028,696	0.000902	1.5337
Catanduanes	53	269,469	0.000197	0.3346

Figure 3. Standard Incidence Ratios of Dengue Count and Population per Province

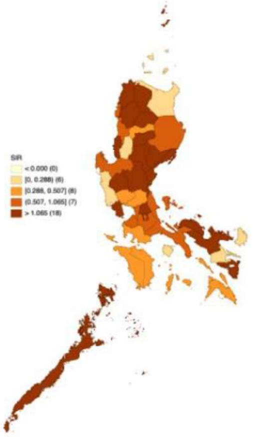


Figure 1. Plot Map for the Accumulated Dengue Count per Province

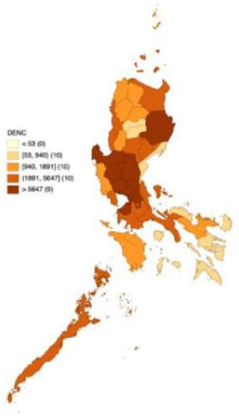


Figure 2. Map of the Incidence Rate of Dengue Count and Population per Province

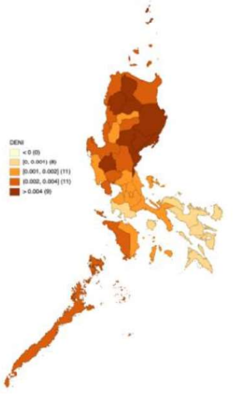


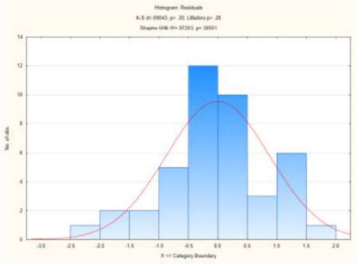
Figure 1 shows the map for the variable DENC, wherein provinces around the center of Luzon are reported to have a high dengue count in 2018.

Clustering is observed, revealing that dengue cases occurring in these areas may be due to spreading. In Figure 2, dengue incidences were derived from dividing DENC by the total population in each province, and it revealed that provinces in the northeastern part of Luzon had high incidence rates. Lastly, Figure 3 shows the SIR of dengue cases and the total population of each province, where provinces located in Region 2 and 3 are observed to have high reported SIRs of dengue. Despite Batanes having the highest SIR, Cagayan did not have a high SIR in 2018. However, provinces that have a clustering of SIR greater than 1 were observed in Central Luzon.

3.2. Multiple Linear Regression Model

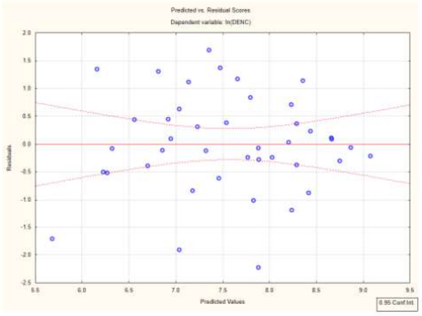
MRLM was applied to dengue cases and incidences but did not satisfy all assumptions. Thus, three transformations were utilized - namely arcsin(y), ln(y), and ln(1-yy). Only ln(y) for DENC satisfied all the assumptions. Stepwise selection procedure was utilized in selecting variables that fit the model. Figure 4 shows that all p-values are greater than 0.05, thus, the residuals are normally distributed.

Figure 4. Normality Test



The homogeneity of variances was checked through a scatterplot of residuals versus predicted values. No clear pattern should be in the distribution or the data would be heteroscedastic. From Figure 5, since there is no cone-shaped pattern, the assumption has been satisfied.

Figure 5. Predicted vs Residual Scores



The next assumption checked was the independence within the residuals through the Durbin-Watson test. In Table 4 (see appendix), as the serial correlation is approximately 0, there is no correlation between the observations. Additionally, since the Durbin-Watson statistic falls between 1.5 and 2.5, there is no first order correlation.

Lastly, there should be no multicollinearity, which occurs when the independent variables are highly correlated with each other. Since the tolerance limits on Table 5 (see appendix) are greater than 0.01, there is no multicollinearity. The errors are also normally distributed with mean 0 and constant variance.

The following output was generated following a 5% level of significance with % = 0.39828405. Table 3.2.1 shows the simplified summary of results while Table 3 (see appendix) shows the full summary.

Table 3.2.1. Simplified Summary of Results of MLRM for ln(dengue count)

Variables	b*	Stnd. Err. of b*	b	Stnd. Err. of b	t(39)	p-value	e^b
Intercept			15.25959	4.760606	3.20539	0.002826	
RF	-0.35336	0.149695	-0.00381	0.001615	-2.36053	0.023786	0.70232
POVINC	-0.35716	0.143864	-0.08164	0.032885	-2.48261	0.017840	0.69966

The summary of results shows that there are two significant dengue correlates: POVINC ($r = -0.35716$, $p < 0.05$) and RF ($r = -0.35336$, $p < 0.05$). A possible explanation for the negative correlation is when POVINC decreases in a certain area, people would flock there for more job opportunities, resulting in more chances for dengue to spread.

Moreover, studies showed that heavy rainfall can possibly lessen dengue fever transmission by reducing the survival rate of mosquitoes. Thammapalo et al. (2005) analyzed the independent effects of rainfall in Thailand, and they suggested that dengue cases have a slightly negative correlation with the precipitation in some provinces.

3.3. Poisson Regression Model

The GENMOD procedure was utilized to obtain the Poisson regression model because it can fit a wide range of generalized linear models. The log(pop) function was specified as an offset to account for possible different observation periods. Additionally, the DSCALE function was utilized due to overdispersion. The simplified results are seen in Table 3.3.1. while Table 6 (see appendix) shows the full analysis. Table 3.3.1. shows that the POP variable has a $p<0.05$ ($p=0.0184$), which indicates that POP is a significant correlate of DENC as expected because the bigger the population, the greater the chance of getting infected.

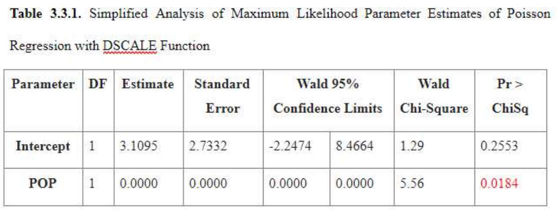


Table 3.3.1. Simplified Analysis of Maximum Likelihood Parameter Estimates of Poisson Regression with DSCALE Function

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	3.1095	2.7332	-2.2474	8.4664	1.29	0.2553
POP	1	0.0000	0.0000	0.0000	0.0000	5.56	0.0184

3.4. Negative Binomial Regression

Due to the overdispersion in Poisson regression, NB was used. The full results are in Table 7 (see appendix). On Table 3.4.1., the independent variables RF (p=.0028) and POVINC (p=.0007) show a p<0.05, meaning that RF and POVINC are statistically significant predictors of DENC.

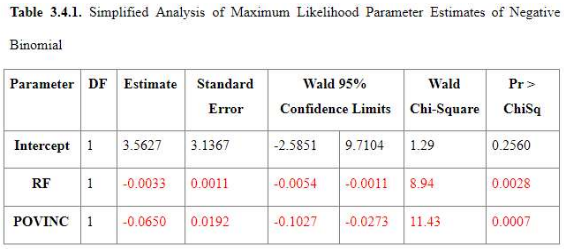


Table 3.4.1. Simplified Analysis of Maximum Likelihood Parameter Estimates of Negative Binomial

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	3.5627	3.1367	-2.5851	9.7104	1.29	0.2560
RF	1	-0.0033	0.0011	-0.0054	-0.0011	8.94	0.0028
POVINC	1	-0.0650	0.0192	-0.1027	-0.0273	11.43	0.0007

Based on Table 5 (see Appendix), the NB model is,

$\log\{E[Y]\} = 3.5627 - 0.0033\ RF + 4.3856\ TMAX + 4.0906\ TMIN - 8.5210\ TMEAN - 0.0038\ RH - 0.0650\ POVINC$

while the model with the significant correlates is,

$\log\{E[Y]\} = 3.5627 - 0.0033\ RF - 0.0650\ POVINC$

where E[Y] is the expected DENC per province.

Results showed that the provinces with the highest expected DENC are Tarlac (E[Y] = 2.055) and Nueva Ecija (E[Y] = 2.249), which are in areas with clustering of SIR greater than 1. RF and POVINC may be attributed as factors for high clustering in Central Luzon as the two provinces have previously recorded about 7,000 raw dengue cases with a poverty incidence of 7.73 and 6.57, respectively. Although some provinces have high DENC, POVINC must also be considered since it is an important factor in the NB model. Thus, Tarlac and Nueva Ecija are the only provinces with notable expected DENC due to high DENC and POVINC values.

The results of this study are consistent with a 2008 study by Sia-Su about the relationship of climatic factors and dengue prevalence in Metro Manila. Through MLRM, the results show that there is a significant correlation between rainfall and dengue incidence ($r = 0.614$, $p < 0.05$). It proved that dengue incidence in the region varies as rainfall patterns change. Moreover, this study explored the statistical procedures Poisson and NB regressions which led to an additional significant correlate, POVINC, of dengue counts and incidences in the Luzon provinces.

4. CONCLUSIONS

The Philippines remains one of the countries reporting high dengue cases globally. In practice, dengue surveillance relies mainly on disease reporting units. However, the illness limits reporting accuracy, leading to a multitude of unrecorded and under-reported dengue episodes. In this paper, significant dengue correlates were obtained and adapted into a MLR and Poisson Regression model. However, due to overdispersion in Poisson Regression, NB was used instead. Both MLRM and NB results revealed that POVINC and RF significantly correlate to DENINC. Although both showed similar significant correlates, the NB model has higher adjusted predictors.

DENC, IRs, and SIRs were mapped in Luzon provinces. The heat maps display hotspots in a specific area which are vital for observing clusters. Results show that Nueva Ecija and Tarlac have SIR higher than 1 and are found in high-risk cluster areas. These are accredited to RF and POVINC, as the NB model has revealed that the following provinces have the highest expected DENC.

With that, it is essential to monitor these provinces as they can be possible sources of dengue spreading in 2018. Although the data is from 2018, the derived results can approximate future dengue trends. Health organizations can use these results to provide health programs and allocate more funds in areas with SIR higher than 1. Visayas and Mindanao provinces should also be tested and modeled to ensure the Filipinos' safety from dengue and include them in programs.

It is also recommended to create a better estimate of the dengue counts by adding more independent variables other than meteorological factors. Additionally, the utilization of heat maps for data visualization, which focuses on new fields like data science, can help determine factors contributing to the spread of dengue.

Considering the COVID-19 pandemic, the statistical procedures used can be executed in determining factors correlating to the increase in COVID-19 cases. Given the availability of data, researchers can secure copies for statistical procedures; and adding variables such as population density, number of hospitals, and Intensive Care Units (ICUs) can be utilized to predict COVID-19 incidences and create a better-fitting model for predictions.

5. ACKNOWLEDGMENTS

First and foremost, the researchers would like to praise and thank God, the Almighty, who has granted countless blessings, knowledge, and opportunities to the researchers. They would like to thank Him for his guidance during the construction of the research paper.

Furthermore, they would also like to acknowledge the assistance of the epidemiology units that provided data for statistical analyses despite the ongoing COVID-19 pandemic. In addition, they would also like to thank them for their patience and understanding with the flurry of emails delivered by the researchers - enabling them to secure data for their research.

The researchers are indebted to the guidance and patience of their research adviser, Ms. Shirlee Ocampo from the Mathematics and Statistics Department of De La Salle University, in the formulation of this research paper and in answering all of the researchers' questions and never getting tired from it. Without her encouragement and advice, the researchers would not have been able to improve on the flaws of the research paper. They would also like to thank Mr. Angelo Alberto for his counsel and Practical Research lectures. During the duration that he was with the researchers, he brought out the best in them through his constructive criticism and enlightening insights.

Last but not the least, the researchers would like to acknowledge each other's efforts during the research process. Their providence of support and appreciation for each other led to the finishing of their research paper.

6. REFERENCES

Natividad, J.M., Necesito, R.M., Ocampo, S., & Leong, R.N. (2019). Bayesian conditional autoregressive model for mapping human immunodeficiency virus incidence in the national capital region of the Philippines. 62nd ISI World Statistics Congress, pp. 291-299.

Sia Su, G. L. (2008). Correlation of climatic factors and dengue incidence in Metro Manila, Philippines. AMBIO: A Journal of the Human Environment, 37(4), 292–294. doi:10.1579/0044-7447(2008)37[292:cocfad]2.0.co;2

Thammapalo, S. , Chongsuwiwatwong, V., McNeil, D., & Geater, A. (2005). The climatic factors influencing the occurrence of dengue hemorrhagic fever in Thailand. Southeast Asian Journal of Tropical Medicine and Public Health; 36(1):191-6. PMID: 15906666.

World Health Organization. (2020a, March 2). Dengue and severe dengue. World Health Organization. https://www.who.int/en/news-room/fact-sheets/detail/dengue-and-severe-dengue

World Health Organization. (2020b, January 16). Update on the Dengue situation in the Western Pacific region. World Health Organization. https://www.who.int/docs/default-source/wpro---documents/emergency/surveillance/dengue/dengue-20200116.pdf?sfvrsn=fc80101d_28