SPATIO-TEMPORAL ANALYSIS AND MODELLING OF DENGUE INCIDENCES IN QUEZON CITY USING ORDINARY LEAST SQUARES AND SPATIAL REGRESSION

A.C. Blanco[1 2], B.J.J. Harder[1], I.A.R. Teh[1]

[1]Department of Geodetic Engineering, University of the Philippines Diliman, Quezon City 1101, Philippines, (acblanco@up.edu.ph)
[2]Philippine Space Agency, Diliman, Quezon City 1101, Philippines

Commission IV, WG IV/7

KEY WORDS: Dengue, OLS regression, Spatial regression, Moran’s I, Factor analysis.

ABSTRACT:

This research examined monthly dengue incidences per barangay (village) in Quezon City, Philippines over a period of six years (2010-2015) to determine the relative significance of environmental variables on dengue prevalence. The data were subjected to correlation analysis, spatial autocorrelation assessment, multiple factor analysis, ordinary least squares (OLS) and spatial regression. Local Indicators of Spatial Autocorrelation (LISA) Moran’s I cluster maps indicate significant (p=0.05, 0.01) High-High clustering and Low-Low clustering in the northern and southern parts of the city, respectively. Monthly total cases indicated increasing trend staring from May/June, peaking at around August/September, and declining afterwards to lower levels in November/December. This corresponds to the typical temporal rainfall pattern. Dengue cases were found to be positively correlated (α=0.05) with Population (R=0.84), Informal_Settlements (IS) (R=0.716), Very_Low_Density_Residential (VLDR) (R=0.512), Open_Spaces (OS) (R=0.339), Mean_Rainfall (RFM) (R=0.637), and Mean Elevation (EM) (R=0.498). Dengue incidence (DI) was negatively correlated with Mean Air Temperature (ATM) (R=-0.3 to 0.5). Based on factor analysis, the dengue incidences were closely related to these variables, though factors F1 and F2 accounted for only 28% of the data variability. Ordinary Least Square (OLS) regression analysis of DI with general land use (LU) classes (e.g., no subclasses in residential areas) identified only IS and OS, explaining 43% of the variability of DI, with IS having twice as much influence on DI compared to OS. When residential subclasses are considered, VLDR was added to the model, slightly increasing R-squared to 0.452. Considering, in addition, EM, RFM, and ATM, the R-squared improved to 0.589, with RFM and EM considered more influential on dengue compared to IS and OS. ATM was however removed due to multicollinearity. The use of Spatial Error regression (SER) and Spatial Lag regression (SLR) produced improved models relative to the OLS model with R-squared of 0.676 and 0.667, respectively. This indicates the importance of spatial dependence. This can be explained by the fact that mosquitos fly over considerably long distances, traversing across the different barangays. The SER model (AIC=1359.89; SE=26.9584) is slightly better than the SLR model (AIC=1366.05; SE= 27.3399).

1. INTRODUCTION

Dengue continues to be one of the leading vector-borne diseases particularly in South-East Asia. The Philippines is a dengue hotspot, ranking highly in number of cases and deaths compared to its neighboring countries. Inhabitants of urban areas are generally of high risk because of presence of larval habitats of the Aedes mosquitoes, high population density, and higher mobility of people (Marti et al., 2020). Despite multi-sectoral efforts to minimize the number of Dengue cases in Metro Manila and other part of the Philippines, the number of Dengue cases remains very high in many places including Quezon City. A significant number among these cases lead to fatalities. It is, therefore, important to further understand Dengue, especially the influence of environmental factors such as rainfall, air and surface temperatures, vegetation distribution, land use distribution, and other variables. This can potentially lead to enhanced measures against Dengue, potentially resulting to fewer cases and mortalities.

Protecting people from Dengue requires many things, including early diagnosis which is critical especially for severe cases. In most cases, doctors do a series of observations to be able to conclude that the illness is Dengue. There are quick and conclusive tests, but such are quite expensive. Recently, cheaper test kits have become available. Early diagnosis of Dengue as well as mitigating potential increases in cases can be facilitated by integrated reporting and sharing of Dengue occurrences as they occur. Comparing this with Dengue occurrence and outbreak likelihood models may assist doctors on early diagnosis. More importantly, such models can provide decision support on how to combat Dengue occurrences in present and future conditions considering scenarios and predictions on meteorological variables.

Geographic information systems have been utilized in several studies to assess dengue risk (Bohra and Andrianasolo, 2001, Bandari et al., 2008) and vulnerability (Zafar et al., 2021) utilizing sociocultural data and environmental data, many of which are derived from satellite remote sensing. Analysis of previous data would enable insights on the underlying environmental and socio-economic factors influencing the spatio-temporal patterns of Dengue cases. For example, Bandari et al. (2008) identified “housing pattern/densities, frequency of cleaning water storage containers, frequency of cleaning drainage/garbage, number of flower pot/home garden, mosquito protection measure/awareness, and storage of water” as significantly related to the number of dengue incidences. The resulting maps can pinpoint areas where interventions must be made as soon as possible. Collating, integrating, and analyzing Dengue and other related variables as Dengue cases occur is essential in preventing and mitigating outbreaks and epidemics.

The main objective of the study is to identify factors (e.g., land use land cover, meteorological) influencing the spatio-temporal prevalence of dengue cases in Quezon City, Metro Manila. This

This contribution has been peer-reviewed.

was done using regression analysis of Dengue incidence data on a per barangay (village) level. As Marti et al. (2022) found out, most of the studies on the relationships between dengue occurrence and landscape features (environmental variables) utilized data aggregated at a chosen spatial unit or geographic level.

2. STUDY SITE

This study was conducted using data for Quezon City, the largest highly urbanized city (HUC) in the National Capital Region (NCR) or Metro Manila. The city has an area of 161.12 sq. km., which is approximately 25% of the area of NCR. Quezon City has the largest total population among the 33 HUCs in the Philippines, with over 2.96 million based on the 2020 Census of Population and Housing. There are 142 barangays or villages. The city has an average population density of roughly 18,000 residents per sq. km. As shown in Figure 1, the city is composed of residential, commercial, and industrial areas.

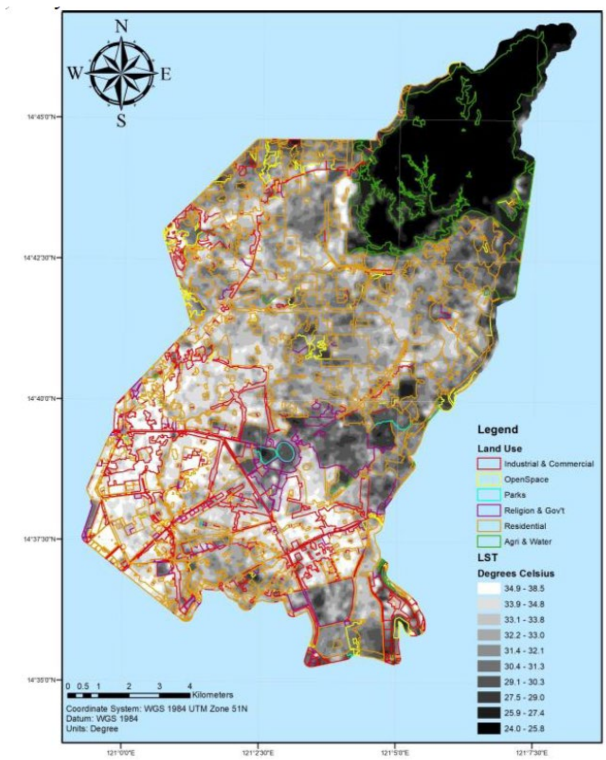

Figure 1. Land use and land surface temperature of Quezon City (Source of map: Alcantara et al., 2019)

Figure 2 shows the spatial variation of total dengue cases in 2015. Higher number of cases were found in the northern barangays or villages. There are areas where population densities are relatively higher. Relatively lower number of cases occurred in the southern barangays. In these areas, we can find mixes of residential, industrial and commercial, and institutional areas as shown in Figure 1.

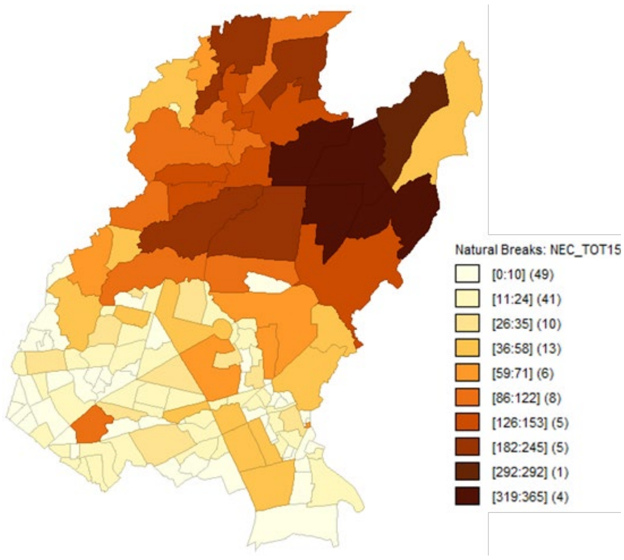

Figure 2. Total dengue cases per barangay in Quezon City in 2015.

3. METHODOLOGY

The methodology used in this study is illustrated in Figure 3. The methods used include correlation analysis, spatial autocorrelation, multiple factor analysis, ordinary least squares (OLS) regression, and spatial regression analysis.

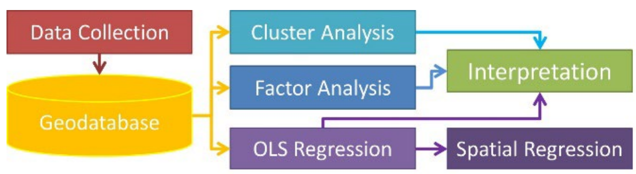

Figure 3. Methodological workflow

3.1 Data

The following data were used in this study (variable name is indicated in parentheses):

a. Total number of dengue cases in each barangay of Quezon City in 2015 (NEC_TOT15)
b. Total number of dengue cases for each month of 2015 in each barangay of Quezon City (NEC2_MMYY where MM is month and YY is year)
c. Barangay population in 2015 (POP_2015)
d. Normalized barangay population in 2015 (POP_NORM)
e. Informal Settlements (INF_SET)
f. Very Low Density Residential (VL_RESD)
g. Open Spaces (OPNSPCS)
h. Medium Density Residential (M_RESD)
i. Commercial Areas (COMM)
j. Mean Rainfall (RainMean)
k. Monthly Rainfall (MMM_RF where MMM represents month)
l. Mean Air Temperature (AirTempMean)
m. Monthly Air Temperature (MMM_AT where MMM represents month)
n. Mean Land Surface Temperature (LSTMean)
o. Maximum Land Surface Temperature (LSTMax)

https://doi.org/10.5194/isprs-archives-XLVIII-4-W6-2022-87-2023

p. Monthly Mean Land Surface Temperature (LST_MMM where MMM represents month)

The data on dengue cases were obtained from the Department of Health – National Epidemiology Center (DOH-NEC). LST was derived from the thermal band of Landsat images.

3.2 Cluster Analysis

Global Moran's I and Local Indicators of Spatial Autocorrelation (LISA) Moran's I were used to evaluate clustering. These are commonly used to identify the dependencies globally and locally present in geographical data. Global Moran's I measures global autocorrelation (i.e., the spatial dependencies in global way) and the index varies between -1 and +1. Values close to +1 indicate presence of spatial clusters. On the hand, Local Moran's I is commonly used in the detection of outliers or hotspots in geographic data set. High risk cluster regions can be identified using Local Moran's I. The value of the index is not limited to the interval -1 to +1. A strong local spatial cluster is identified through a large positive value of Local Moran's I (Anselin, 1995). In this study, the Cluster and Outlier Analysis (Anselin Local Moran's I) tool in ArcGIS was utilized to identify clusters of high values, clusters of low values, as well as spatial outliers.

3.3 Multiple Factor Analysis

Multiple factor analysis (MFA) was used to examine relationship between variables. Through MFA, a set of observations described by several groups of variables can be analyzed.

3.4 Regression Analysis

Ordinary Least Squares (OLS) Regression was first applied to the dengue cases, considering the total cases in a year as well as the cases during low incidence and high incidence months. In terms of the groupings of explanatory or independent variables, land use classes (e.g., residential areas of different densities) were first considered. In succeeding runs, the meteorological variables, specifically rainfall, air temperature, and land surface temperature (LST), were also considered. The analysis also considered the possible temporal lag effects on the influence of these meteorological variables on the incidence of dengue.

OLS regression can be severely limited since it is a global approach which does not consider potential spatial interactions or relationships that may exist depending on the processes being considered. In the case of dengue incidences, the vector can move from place to place. Therefore, we can expect that cases at nearby barangays may exhibit some sort of spatial dependence. To examine such, spatial regression analyses, namely, Spatial Lag and Spatial Error, were considered.

The correlogram indicates that the autocorrelation of Dengue Cases variable is zero at around 6 kilometers (Figure 4). Spatial weights, which are needed in spatial regression, were generated using threshold distances 6 km, 5 km, 4 km, 3 km, and 2.079 km, which is the minimum where there are enough neighbors to evaluate the statistic.

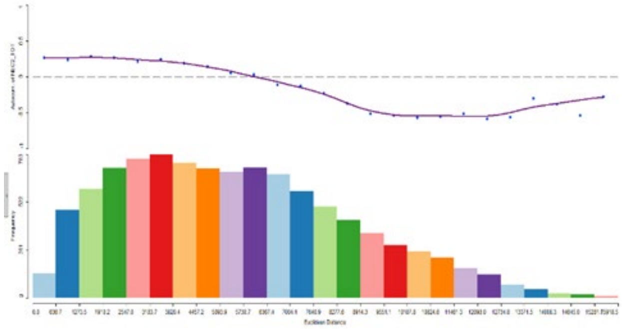

Figure 4. Correlogram indicating spatial process at work over varying spatial lags

4. RESULTS AND DISCUSSION

4.1 Clusters

The spatial distribution of dengue incidences manifested a North to South trend with northern barangays having experienced clusters of higher number of cases. Moran's I values [0.491 (2014) to 0.714 (2015)] indicated this tendency to cluster. Local Indicators of Spatial Autocorrelation (LISA) Moran's I cluster maps indicate significant (p=0.05) High-High clustering and Low-Low clustering in the north and south parts of the city, respectively (see Figure 5). However, these clusters/bands are separated by cluster/band of non-significance in the central parts of the city.

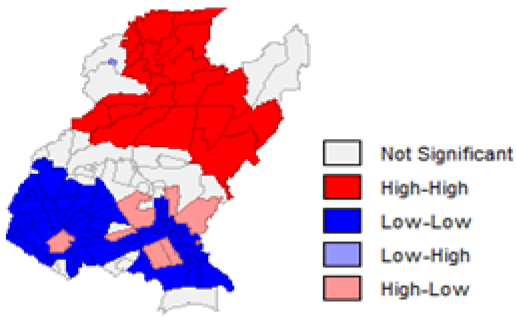

Figure 5. Spatial clusters based on LISA Moran's I of total cases in 2015.

Monthly total cases over six years indicated increasing trend staring from May/June, peaking at around August/September, and declining afterwards to lower levels in November/December. This corresponds to the typical temporal rainfall pattern. Spatially, based on local Moran's I, High-High and Low-Low clusters, in the north and south parts of the city respectively, started expanding and aggregating starting in June and contracted towards December (Figure 6).

The clustering of high-high values in the northern part of Quezon City and low-low clusters in the southern part of the city are persistent, whether yearly or monthly. It is noted that the spatial distribution of dengue cases in 2015 indicate very high values in areas with large informal settlers.

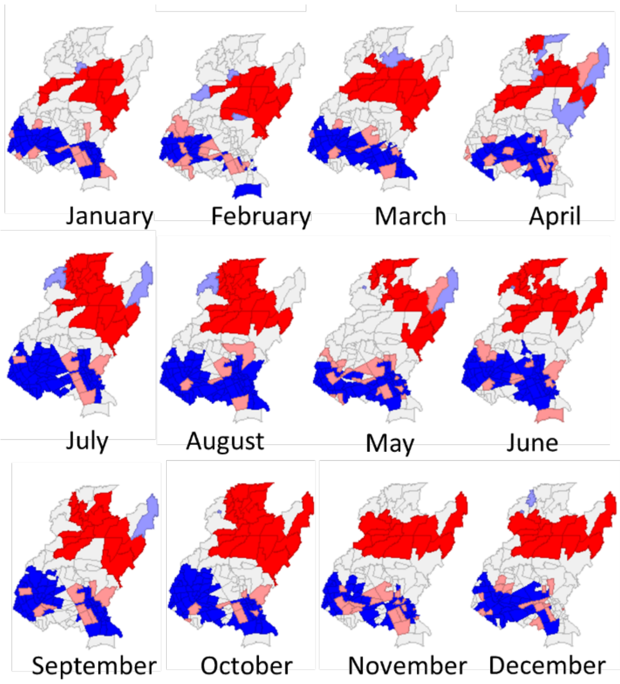

Figure 6. Spatial clusters based on LISA Moran's I of monthly total cases in 2015

4.2 MFA Results

The plot of factor loadings (Figure 7) provides information on the relationship among the variables, considering entire 2015 data set. Dengue cases for 2015 (NEC_TOT15) are directly correlated with Population (POP_2015, POP_NORM), Informal Settlements (INF_SET), Very Low Density Residential (VL_RESD) and Open Spaces (OPNSPCS). Medium Density Residential (M_RESD) and Commercial Areas (COMM) appear to be inversely related to Dengue cases.

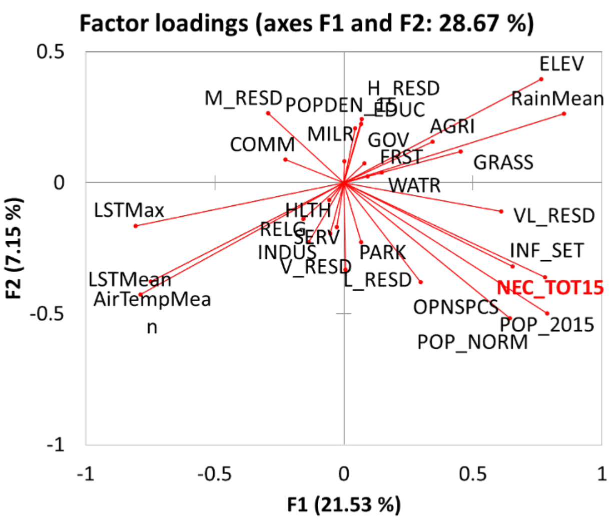

Figure 7. Factor loadings plot

Meteorological variables (RainMean, AirTempMean) appear to be weakly related to Dengue Cases. However, it should be noted that Factors F1 and F2 account for only 28.67% of the data variability. Hence, observations may not be fully conclusive and additional analyses are needed.

4.3 Regression Models

4.3.1 Models for 2015 Using Land Use Variables Only

Equation 1 is the OLS model considering 2015 total cases (NEC2_TOT15).

$$\textit{NEC2\_TOT15} = 18.2545 + 3.40894 * \textit{OPNSPCS} + 5.52552 * \textit{INF\_SET} + 3.28358 * \textit{VL\_RESD} \tag{1}$$

Three (3) significant Land Use variables, namely Open Spaces (OPNSPCS), Informal Settlements (INF_SET), and Very Low Density Residential Areas (VL_RESD), account for 65.9% (R-squared) of the variability of dengue cases in 2015. However, standard error (S.E.) is 41.8. It can be said that Open Spaces and Very Low Density Residential Areas have large enough spaces where water may pond. If these areas are not well maintained, they can serve as breeding grounds for mosquitoes. In Informal Settlements, sanitary practices, including water and waste management, are typically poor. Water may accumulate in containers, tyres, and other objects scattered around.

Upon evaluation, spatial dependence was found to be significant and therefore, spatial regression can be used. Spatial Lag regression (SL) is preferred based on spatial dependence diagnostics (i.e., Langrange Multiplier). Equation 2 is the SL Model for 2015 Using Land Use Variables Only:

$$\textit{NEC2\_TOT15} = 4.25566 + 0.503008 * \textit{W\_NEC2\_TOT15} + 2.26656 * \textit{OPNSPCS} + 4.54279 * \textit{INF\_SET} + 1.81258 * \textit{VL\_RESD} \tag{2}$$

where W_NEC2_TOT15 is the NEC2_TOT15 at the spatial lag considered.

The Spatial Lag regression improved the R-square to 0.71 from 0.659 obtained with OLS. S.E. is 38.6, which is lower that 41.8 from OLS. These indicate significant improvements and underscores the fact that the movement of vectors result to significant spatial dependence in dengue cases.

5.3.2 Models for May 2015 (Low Dengue Case)

Equation 3 is the OLS model for case of low dengue case (May 2015). It indicates that the contributions of the previous month's dengue cases (NEC2_0415) and rainfall (APR_RF) are significant. R-squared is 0.65. S.E. is significantly lower at 1.02.

$$\textit{NEC2\_0515} = -2.48342 + 0.064177 * \textit{INF\_SET} + 0.352647 * \textit{NEC2\_0415} + 0.276018 * \textit{APR\_RF} - 0.0340807 * \textit{MAY\_RF} \tag{3}$$

Based on diagnostics for spatial dependency, analysis should proceed to spatial error (SE) regression. Table 1 lists the different SE models for May 2015 which represents the low dengue case. Different spatial lags were used to general the spatial weights.

	2079 m	3000 m	4000 m	5000 m	6000 m
CONSTANT	-2.75459	-2.40935	-2.65555	-0.555201	-1.33082
INF_SET	0.0653806	0.0777196	0.0680193	0.0540399	0.066272
NEC2_0415	0.2452	0.28731	0.381428	0.67094	0.402536
APR_RF	0.30563	0.305177	0.180028	0.160818	0.248589
MAY_RF	-0.0373751	-0.0398926	*-0.0148306**	-0.0267819	-0.0370072
LAMBDA	-0.853233	-0.636094	0.580023	14.7081	6.57941
R-squared	**0.771867**	0.700874	0.603818	0.000000	0.289270
S.E.	**0.811829**	0.929604	1.06984	2.30418	1.43292
AIC	**363.46**	387.998	384.884	390.298	397.839

Table 1. SE Models for May 2015 (Low Dengue Case) Using Different Spatial Lag and Weights.

Note that the SE model for the shortest spatial lag performed best with highest R-squared value of 0.772 and lowest S.E. at 0.812. The AIC value is significantly the lowest. Based on flight range studies, most female Ae. aegypti may spend their usually fly an average of 400 meters only. The shortest lag distance used in this study is limited by the areas of the barangays. It might be possible that, if smaller spatial units are used, the model performance would improve since shorter lag distance comparable to the 400 meters can be used. However, in addition to the vectors, people move the virus between communities and places, so spatial lag can be significantly greater than 400 meters.

5.3.3 Models for September 2015 (High Dengue Case) Using All Variable Types

	OLS_S1	OLS_S2	OLS_S3
CONSTANT	430.649	571.315	-11.608
NEC2_0815	1.2378	1.28206	1.30283
INF_SET	0.517221	0.512069	0.479771
AUG_AT	-1.37287	---	---
AUG_RF	-0.118146	---	---
SEP_AT	---	-1.81117	---
SEP_RF	---	-0.216222	
LST_AUG	---	---	3.01199
LST_SEP	---	---	-2.29511
R-squared	0.891201	0.894165	**0.898082**
S.E.	6.53452	6.44489	**6.32449**
AIC	940.984	937.062	**931.707**

Table 2. OLS models for September 2015 considering previous month's weather conditions.

The OLS models in Table 2 indicate significant contribution of the previous month's dengue cases, meteorological condition and even land surface temperature (LST). The OLS model with LST variables included yielded the lowest S.E. and AIC values.

Among the land uses, Informal Settlement is the only type included in the models, indicating that such areas may be breeding grounds of mosquitos. Based on diagnostics for spatial dependency, analysis should proceed to spatial lag (SL) regression.

For high dengue case month (September), spatial lags model using 6-km threshold distance is the best model, incorporating the rainfall (RF) and LST of the previous month. It can be seen that the performance of the three models given in Table 3 are not significantly different based on R-squared, S.E. and AIC.

	SL_RF	SL_LST	SL_Combi
CONSTANT	-58.3262	-6.79409	-113.872
W_NEC2_0915	-0.524919	-0.187042	-0.317454
NEC2_0815	1.28829	1.35122	1.25994
INF_SET	0.514707	0.51866	0.518478
AUG_RF	-0.404695	---	0.109881
SEP_RF	0.706446	---	
LST_AUG	---	1.82658	2.89623
LST_SEP	---	-1.33582	
R-squared	0.902351	0.901424	0.902154
S.E.	6.08064	6.10946	6.08678
AIC	928.08	929.037	928.093

Table 3. SL Models for September 2015 (high dengue case) using meteorological variables.

5. CONCLUSIONS

The Informal Settlement variable was found significant in all regression models and even in factor analysis. This indicates that such areas serving as breeding grounds is highly possible. This underscores the importance of knowing the land use as it provides an idea of the activities (e.g., sanitary practices, water management) carried out in such areas.

Rainfall and temperature variables (air and land surface), including those that are temporally lagged, can improve the modelling of dengue cases. This is related to the life cycle of the mosquitoes and how the fecundity can be affected.

Modelling of dengue cases was enhanced with the use of spatial lags and temporal lags. This supports the fact that the movement of vectors and people promote spatial dependence, which is better explained using spatial regression.

REFERENCES

Alcantara, C. A., Escoto, J. D., Blanco, A. C., Baloloy, A. B., Santos, J. A., and Sta. Ana, R. R.: Geospatial Assessment and Modeling of Urban Heat Islands in Quezon City, Philippines Using OLS and Geographically Weighted Regression, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-4/W16, 85–92, https://doi.org/10.5194/isprs-archives-XLII-4-W16-85-2019, 2019.

Anselin L., 1995. Geographical Analysis 27 93

Bandari, K.P, Raju, P.L.N., and Sokhi, B.S., 2008. Application of GIS Modeling for Dengue Fever Prone Area based on Socio-cultural and Environmental Factors – A Case Study of Delhi City Zone. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*., Vol. XXXVII. Part B8

Bohra and Andrianasolo, 2001. Application of GIS in Modeling of Dengue Risk Based on Sociocultural Data: Case of Jalore, Rajasthan, India. Dengue Bulletin Vol. 25, 92-102

Marti, R., Li, Z., Catry, T., Roux, E., Mangeas, M., Handschumache, P., Gaudart, J., Tran, A., Demagistri, L., Faure, J-F., Carvajal, J.J., Drumond, B., Xu, L., Herbreteau, V., Gurgel, H., Dessay, N. and Gong, P., 2020. A Mapping Review on Urban Landscape Factors of Dengue Retrieved from Earth Observation Data, GIS Techniques, and Survey Questionnaires. Remote Sens. 2020, 12, 932; doi:10.3390/rs12060932

Zafar, S., Shipin, O., Paul, R.E., Rocklöv, J., Haque, U., Rahman, M.S., Mayxay, M., Pientong, C., Aromseree, S., Poolphol, P., et al., 2021. Development and Comparison of Dengue Vulnerability Indices Using GIS-Based Multi-Criteria Decision Analysis in Lao PDR and Thailand. Int. J. Environ. Res. Public Health 2021, 18, 9421. https://doi.org/10.3390/ijerph18179421