GEO-STATISTICAL DENGUE RISK MODEL
CASE STUDY OF LAHORE DENGUE OUTBREAKS 2011

Bilal Tariq
Arjumand Z. Zaidi
Institute of Space Technology (IST), Karachi
INTRODUCTION

Mosquito-borne diseases have been a serious threat to human health in many parts of the worlds, especially in developing countries of Asia and Africa. (Lemon et al. 2008)

According to World Health Organization report (WHO, 2006), dengue fever is present in at least 100 countries and due to this fever, approximately 40 percent of the world population (2.5 billion people) are at risk in tropics and sub-tropics regions.
Climatic change pays an important role on the mosquito breeding, especially dengue virus. The spread and extend of dengue fever may help to make connections between land use, climate and public health. (L. Rusch, Perry, 2011)

In Pakistan dengue fever has been known since 1994 when first case was registered in the southern part of Karachi and epidemics were reported from almost all part of the country.

In the Punjab province, especially in Lahore the number of cases are increasing since 2007. (BBC News, 13 Sep 2011)
The worst ever 2011 outbreak of dengue in Lahore occurred from March to December while in other parts of Punjab such outbreaks are mostly reported after the Monsoon season. The numbers reached a peak during the hot and rainy season (August to October). According to Punjab Disaster Management Authority (PDMA), total 17,330 dengue cases were registered in Lahore in 2011. By the end of the rainy season, these incidences returned back to a lower level.
OBJECTIVE

- To identify Land Cover/Land Use and its relationship with dengue fever outbreaks in Lahore.

- To identify the environmental factors that contributes to the dengue outbreak using Remote Sensing data.

- To correlate the identified environmental factors with dengue outbreak distribution pattern.

- To develop geo-statistical dengue risk model for identification of risk prone areas by linking environmental, demographic and land use / cover parameters with dengue cases.
**STUDY AREA**

Lahore is the capital city of Pakistani province Punjab. According to 1998 census, Lahore's population was 6,318,745 and population for year 2010 was estimated at 8,592,000 making it the second largest city of Pakistan after Karachi. It is located within the latitude and longitude of 31°32'59"N and 74°20'37"E. The northern side of the city lies on the bank of river Ravi and the eastern side borders with India.
**Methodological Workflow**

### Phase 1 (Data Collection)
- Satellite Imagery
  - Spot (5 m)
  - Landsat (30 m)
  - Google Earth
- Climatic Data
  - Medical Data
  - Population Data
  - Administrative Data

### Phase 2 (Data processing, extraction & Development of Geo-database)
- Image Classification
- LST Calculation
- NDVI Calculation
- NDWI Calculation
- Built-up area Digitization

### Phase 3 (Analysis)
- GIS Database
- Statistical Analysis (OLS and GWR)
- Regression Models
- Dengue Prediction Map
ENVIRONMENTAL FACTORS & DATA

- Topography (SRTM)
- Meteorological Data
- Land Surface Temperature (LST) (Landsat TM)
- Normalized Difference Vegetation Index (NDVI) (Spot-5)
- Normalized Difference Water Index (NDWI) (Spot-5)
- Land Cover/Land Use (LCLU) (Spot-5)
- Built-up area, population and population density
RESULTS & DISCUSSION
Dengue cases usually occur after rainfall and the life span of dengue is 2-3 weeks, a date wise comparison of the two variables cannot be possible. To find out the relationship between these two factors, dengue cases are aligned with rainfall data by shifting them 15 to 18 days back from their recorded dates of incidence. Following time intervals are selected for shifting of dengue cases according to the amount of rainfall.

<table>
<thead>
<tr>
<th>Rainfall</th>
<th>Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>less than 1 mm or equal</td>
<td>0-1 days</td>
</tr>
<tr>
<td>1.1-10 mm</td>
<td>2-4 days</td>
</tr>
<tr>
<td>10.1-40 mm</td>
<td>5-9 days</td>
</tr>
<tr>
<td>equal to 40.1 mm or greater</td>
<td>10-18 days</td>
</tr>
</tbody>
</table>
Correlation between dengue outbreaks & rainfall data

Data Gunj Baksh Town (U.C. Race Course)

\[ y = 0.4957x + 2.7237 \]
\[ R^2 = 0.6645 \]

Data Gunj Baksh Town (U.C. Shadman)

\[ y = 0.3057x \]
\[ R^2 = 0.7285 \]
Correlation between dengue outbreaks & rainfall data

**Samanabad Town ( U.C. Bahawalpur House)**

- Y = 0.1213x
- $R^2 = 0.7338$

**Samanabad Town ( U.C. Icchra)**

- Y = 0.3474x + 2.4003
- $R^2 = 0.8035$
Correlation between dengue outbreaks & rainfall data

Results of linear regression show statistically significant relationship between rainfall and dengue outbreaks in all samples union councils with $R^2$ ranging from 0.66 (minimum) to 0.8 (maximum)
LAND COVER LAND USE (LCLU)
TOPOGRAPHY

Legend
- District Boundary
- Town Boundary

Digital Elevation Model (in meters)
- 195.61 - 202.00
- 202.01 - 209.00
- 209.01 - 216.00
- 216.01 - 223.00
- 223.01 - 229.39

Dengue Cases
- 0 - 137
- 138 - 504
- 505 - 867
- 868 - 1438
- 1439 - 1853
RELATIONSHIP BETWEEN LST & DENGUE OUTBREAKS
RELATIONSHIP BETWEEN NDVI & DENGUE OUTBREAKS
**GEO-STATISTICAL RISK MODEL**

- In order to test the research hypothesis and to find out the relationship between dengue cases and study parameters and the direction of relations, the following regression analyses are conducted on the data.
  1. *Ordinary Least Squares (OLS) Regression*
  2. *Geographically Weighted Regression (GWR)*

- Ordinary Least Squares (OLS) is a global regression method. Geographically Weighted Regression (GWR) is a local spatial regression method that allows the relationships to model across the study area.
**OLS Regression between Dengue Cases and All Study Parameters**

- Average values of NDVI, NDWI, LST and maximum LST are used in the statistical analysis. Other parameters used in this analysis are 2008 population (in thousands), population density, UC area, and built area.

- Although many parameters do not have apparent linear relationship with dengue cases but a trial OLS regression analysis is conducted using OLS tool of ArcGIS selecting dengue cases as dependent variable and all other study parameters mentioned above as independent variables.

- Results show that although a high $R^2$ of 0.747 (Adjusted $R^2 = 0.733$) value is found, the overall model has problems like higher significance level ($P$-values $>0.05$) of regression coefficients.
OLS Regression between Dengue Cases and Built-up Area and Population Density

- Regression result shows that the Adjusted R-Squared value is 0.667, or 67%. This indicates that using population density and built-up area, the model is explaining 67% of the variation in dengue incidences.

- Also all explanatory variables are statistically significant but the significance of the Jarque-Bera statistics is making this model biased and hence undesirable. Also the Koeker Test is statistically significant that implies that the relationships between the dependent and some or all of the explanatory variables are non-stationary.
This means, that the explanatory variable (*built-up area* and *population density*) might be an important predictor of dengue cases at some areas but may result in a weak prediction in other locations.

Therefore, it is concluded that study parameters might have spatially varying relationships and therefore, are not suitable for OLS global model.

Geographically Weighted Regression analysis is a form of linear regression that can model spatially varying relationships between variables.
GWR BETWEEN DENGUE CASES AND ALL STUDY PARAMETERS

- Geographically Weighted Regression (GWR) tool of ArcGIS is run selecting dengue cases as dependent variable and all other study parameters as explanatory variables.

- The model fails to execute and results cannot be computed successfully.

- As suggested by the model error message, this may be due to either severe global or severe local multi-collinearity (redundancy among model explanatory variables).

- Different trail runs of the model are executed either by removing redundant variables from the model that have large Variance Inflation Factor (VIF-values > 7.5) in the previously run OLS model or trying other combinations of study parameters as explanatory variables.
GWR for Dengue Cases, Built Area and Population Density

- The model results show that the Adjusted R-Squared value is 0.774 (R² = 0.805). This indicates that using population density and built-up area as explanatory variables, the model can explain 77% of the variance in dengue incidences.

- The GWR output is a standard residual map that is an indicator of model performance. Residuals are the portion of the total variability of the observed data that is unexplained by the model or the part of the model under and over predictions.

- The results of Spatial Autocorrelation analysis show that the regression residuals are randomly distributed since the z-score (=0.17) is not statistically significant. The null hypothesis of complete spatial randomness is, therefore, not rejected that confirms the randomness of the residuals required for a well performing model.
Standard Residual Map of GWR Model
Spatial Autocorrelation Report

Moran's Index: -0.001072
z-score: 0.169452
p-value: 0.865441

Significance Level (p-value)
- 0.01
- 0.05
- 0.10
---
- 0.10
- 0.05
- 0.01

Critical Value (z-score)
- < -2.58
- -2.58 - -1.96
- -1.96 - -1.65
- -1.65 - 1.65
- 1.65 - 1.96
- 1.96 - 2.58
- > 2.58

Given the z-score of 0.17, the pattern does not appear to be significantly different than random.
SPATIAL DISTRIBUTION OF REGRESSION COEFFICIENT - U.C. BUILT-UP AREA
SPATIAL DISTRIBUTION OF REGRESSION COEFFICIENT – POPULATION DENSITY

Legend
- Union Council Boundary
- Geographically Weighted Regression Coefficient-2 Population Density
  - -4.0 - -7.1
  - -4.0 - 0.8
  - 0.9 - 1.9
  - 2.0 - 3.5
  - 3.6 - 6.6
  - 6.7 - 9.8

Projection/Datum: WGS84
Map Data Sources:
Spatial Data: Urban Unit
**Dengue Prediction Map**

- Geographical Weighted Regression model also be used to predict values of dependent variables for locations within the study area with some projected values of explanatory variables.

- The prediction map is prepared assuming projected values of 10 percent and 20 percent, respectively, for *built-up areas* and *population densities* in all union councils.

- The predicted dengue cases are shown in next slide based on projected values of explanatory variables (*built-up area* and *population density*).

- There are some negative predicted values that are not acceptable and indicate model inaccuracy in these areas in giving acceptable estimations. Areas with negative values are shaded out and not included in analysis since these predictions are not reliable.
**Dengue Prediction Map**

Predicted Dengue Cases - Projected Data

- **Dengue Cases**
  - 0 - 15
  - 16 - 48
  - 49 - 124
  - 125 - 212
  - 213 - 397
  - 398 - 1586

Observed Dengue Cases

Projection/Datum: WGS84

Kilometers
CONCLUSIONS

- The objective of this study is to develop a geo-statistical dengue risk map for Lahore city by identifying environmental, demographic and land-use/cover factors that significantly influence dengue outbreaks.

- This study analyzes the impact of these parameters on the distribution pattern of dengue outbreaks in Lahore city. The individual as well as synergistic impact of study parameters are evaluated and among them factors are selected that show significant influence on dengue outbreaks. Several factors including vegetation cover, land surface temperature, water, built-up area and population density are analyzed for their influence on the spread of the disease in the study area.
The model show statistically significant relationship between rainfall and dengue outbreaks in all samples union councils with $R^2$ ranging from 0.66 (minimum) to 0.8 (maximum).

Based on study results, it is concluded that study parameters are not suitable for OLS global model since no statistically strong model can be found using OLS.

GWR analysis is a form of linear regression that can model spatially varying relationships between variables. The GWR model shows that using population density and built-up area as explanatory variables, the model can explain 77% of the variance in dengue incidences.
Densely populated and heavily built areas are the most vulnerable ones that may provide suitable breeding grounds for virus to grow.

There is a fear that these areas may experience severe dengue outbreaks if effective measures are not taken in order to control the disease or to minimize its risk.

The geo-statistical dengue risk model developed in this study can be used to predict risk areas in the most vulnerable city Lahore that need special attention in order to effectively and efficiently manage and mitigate dengue outbreaks in future.
RECOMMENDATION

- It is suggested that a high resolution DEM be used for future work. It is important due to rapid urbanization and rapid changes in the ground elevation. Because the high resolution DEM provides precise and accurate results of study area.

- This research found that the LST data plays an important role towards the distribution pattern of dengue outbreaks. Due to the expiry of Landsat-5 (TM), it is proposed that the Landsat-8 or Aster Satellite data used to generate the LST data for future analysis.

- The incidences data need to be collected at the waged people level and Government should start spraying their respective areas where water stands after rainfall.
ACKNOWLEDGEMENT

I am thankful to Member SAR, Pakistan Space and Upper Atmospheric Research Commission (SUPARCO), Director Pakistan Metrological Department (PMD), Director General (DG) Punjab Disaster Management Authority (PDMA) and Director The Urban Unit for providing data for this research work.
REFERENCES


M Ilyas Khan. BBC News (13 September 2011). Islamabad


Thank You