

Developing dengue index through the integration of crowdsourcing approach (X-Waba)

Lyes Mokraoui*, Norzailawati Mohd Noor and Alias Abdullah

Department of Urban and Regional Planning, Kulliyah of Architecture and Environmental Design, International Islamic University Malaysia, Jalan Gombak, 53100, Kuala Lumpur, Malaysia

lyes.mokraoui@gmail.com

Abstract. This research aims at improving the existing dengue indices by developing an algorithm that would use the variables affecting positively the vector's lifecycle and monitoring them daily to generate an improved Dengue Index that would help forewarn on the high possibility of a dengue outbreak. Researchers attempted to identify the factors influencing the behaviour of the mosquito carrier of the virus in the epidemiological context by generating indices based on the number of mosquitos caught in a given number of traps set within a small urban area or based on the number of mosquitoes found in a household. The House (premises) Index (HI), the Container Index (CI) and the Breteau Index (BI) have been described as not effective in predicting dengue outbreaks. The main drawback of these methods is the fact that they do not consider other variables associated with the vector's lifecycle, landing habit and geographical extents. The developed index would use crowdsourcing data as an additional tool for the citizens to get involved in providing spatial information and specific attributes for more accurate predictions. Using the data published by the ministry of health Malaysia in the years of 2014 and 2015 for the state of Selangor, the federal territories of Kuala Lumpur and Putrajaya compared with other data; namely, the temperatures, rainfall and moon cycles. Our findings using the time series method of the improved dengue index show a correlation with the dengue cases time series. The crowdsourcing app would in the future further enhance the identification of the hot spots with high dengue fever probabilities. In conclusion, displaying spatially on a map such forecasts approximately 50 days before the occurrence of the outbreak would be beneficial for authorities to carry out preventive measures.

1. Introduction

The ongoing urban developments in the world is resulting in the high number of dengue outbreaks in the past years; this research on current and historical data is an attempt to predict where outbreaks will occur by studying existing cases and probable expansions of the outbreaks using Geographic Information Systems (GIS) techniques by enhancing the existing BI. Dengue is an alarming urban health related diseases Malaysia has been facing since the first outbreak in 1962. Dengue is the result of the effect of a growing urban environment. Putrajaya, the new administrative capital of Malaysia with its modern urban planning is no exception to the outbreaks [1]. In the Urban health studies



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

context, Acevedo-Garcia, D. (2001) in their report titled "Why Urban Health Matters" by the World Health Organization (WHO) [2], it is stated that cities growth tend to promote unhealthy lifestyles, such as "convenient" diets, sedentary behaviour, smoking, and the consumption of alcoholic substances [2]. Other effects such as the availability in these environments of man-made containers, transportation of recycled tires and other goods are responsible for the rapid growth of the dengue epidemic globally from two (2) countries in the fifties to 100 in 2014 [3]. "Mosquito" is the common name for *Aedes*. the Dengue Fever (DF) vector. It is the origin of 9 vector borne diseases and they are known as reported by WHO report [4] including zika, yellow fever and the Nile fever. Insects biting habits increase with the moon light and mosquitoes might not be an exception. Mokraoui [5] suggests that there is a correlation between the moon phases and the dengue outbreaks as the full moon and third quarter of the moon phase are the ones where the vector is the most active [5]. Little research in Malaysia was done through a report done by WHO [6] advising to conduct such research to confirm the existence of the correlation between moon phase and malaria. The temperature and rainfall patterns are proven to affect the vectors' reproduction cycle. In order to collect more current data on the vector activity, Crowdsourcing will contribute through Volunteered Geographic Information. Citizens would help to provide information on the location of such as water containers locations with larvae. In an attempt to establish the epidemiologic and entomologic correlations of the variables affecting the outbreaks, researchers have developed three indices; the house (premises) index (HI), container index (CI) and Breteau index (BI). However, these indices do not reflect potential dengue transmission nor the areas with greater risks [7]. In this research we aimed in enhancing the temporal and spatial properties of the index by using a crowdsourcing App data and other variables that affect positively the landing and biting habit of the vector. In this research the scope would include:

- a. Data Analysis of the dengue outbreaks for the years of 2014 until 2015.
- b. The developing a formula to enhance the existing dengue indexes and generating time series which will be compared with the actual historical dengue data taking in consideration the land use, temperatures, rainfall and moon phases.
- c. The development of a crowdsourcing app to enhance the existing dengue indices to provide
- d. and give access to a wider spatial content.

The crowdsourcing method will be used as a surveillance system to track pandemics and make rapid decisions consequently reducing the spread of communicable diseases. It is suggested that new technologies [8] could be developed moving towards citizens science (citizens who act as observers), by engaging the general public. The Center of Social Media Innovations for Communities (COSMIC) in Singapore has developed a prototype of an integrated social media-based system called Mo-Buzz to enhance predictive surveillance; encourage civic engagement and health communication. The system is using a predictive algorithm and computer simulations (based on weather, vector and human data) to predict dengue outbreaks. In collaboration with the Colombo Municipal Council (CMC) and the University of Colombo, the Mo-Buzz system was deployed in July 2013, Sri Lanka. Mobile phones might potentially transform the fields of epidemiology and health communication from being complementary to synergistic entities. [9] Thus, awareness and population involvement could be developed and implemented using smartphones.

2. Study area

The research study area is based on the data from Selangor inclusive of the federal territories of Kuala Lumpur and Putrajaya – Figure 1. The state of Selangor, which extends along the west coast of peninsular Malaysia, is located between 2°35' and 3°60' N and between 100°45' and 102°00' E. Selangor, is the most rapidly developing state in Malaysia, has an area of approximately 800,000 ha and borders the state of Perak to the north, the state of Negeri Sembilan to the south, and the state of Pahang to the east [10].



Figure 1. Location map of Selangor State in Peninsular Malaysia

The location of the research will be concentrated in the state of Selangor and its federal territories which are Kuala Lumpur and Putrajaya. This choice was made taking in consideration that the statistics show that Selangor is the state with the highest number of dengue cases between the year 2000 and 2014 [11] and has the highest rate of smart phone penetration in the country [12]. The state of Selangor is located on the west coast of peninsular Malaysia. It contains the federal territories of Kuala Lumpur and Putrajaya. Selangor also contains two cities, Shah Alam and Petaling Jaya. Thus, the state is the most populated in the country with a total of more than 6.1 million population excluding Kuala Lumpur and Putrajaya which is more than 20% of the overall population in Malaysia.

3. Materials and Methods

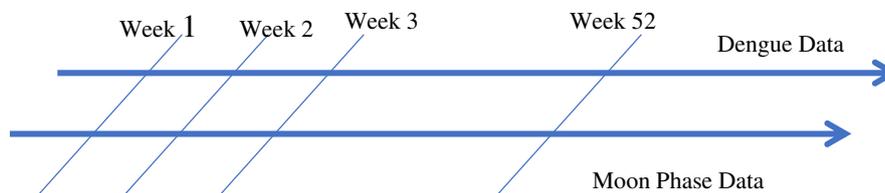
The dengue data used for this research are of the years of 2014 and 2015 as they are the most complete, data from prior years are not continuous and have several gaps of non-available data. Table 1 describes the data used and the source of the data. Throughout the research project, Microsoft Excel and FME Desktop were used in order to geocode data, calculate statistics and generate the time series of the dengue outbreak data and the index time series.

Table 1. Required Data for the index calculation

Tabular data (soft copy)	Source of Data
Dengue cases by address by week	www.data.gov.org
Moon Phase data	http://www.timeanddate.com/moon/phases/malaysia/kuala-lumpur
Temperature data	Weather Department Malaysia
Rainfall data	Weather Department Malaysia
Land use data	Town Planning Department Malaysia

3.1 Collecting the Data

All the data used in this research are in softcopy and structured in a proper standardized way. Little adjustment was required and the exception of the historical dengue data from the ministry of health available from the Malaysian government open data initiative website www.data.org.my. The data used for this research was restricted to the years of 2014 and 2015 as they were the years with the most complete data for the dengue outbreaks. As the dengue data is provided on weekly basis starting on Monday, all the remaining data dates had to be adjusted accordingly to be aligned with these data. The moon phase data was downloaded from www.timeanddate.com website. The data is freely available for download. As the moon phase data would negligibly vary from state to state in Malaysia. Data for Kuala Lumpur as a base for the research. All the data, had to be temporally uniform, thus the moon cycle data would have to be converted into weeks of the year. Processing the data required tedious arrangement of the element and uniformizing the temporal data (see Figure 2). The moon phase data is textual, the processing of the data will consist of converting the starting of each phase date into a week of the year and appending numerical values to the different phases which are namely new moon, 1st quarter, full moon and third quarter.

**Figure 2.** Standardizing the moon phase data

3.2. Crowdsourcing (X-Waba)

The crowdsourcing app for volunteered geographical information is called X-Waba and available on Google Play for android phones. The App was developed using PhoneGap which is provided by Adobe as a free app development service (free for one app). It is based on Apache Cordova mobile application development framework. X-Waba's developed interface was designed with the simplicity of use as primary objective. It collects the coordinates of the location automatically from the phone's GPS and the users would just have to choose a few options before submitting the form to a server for further backend processing (Figure 3).

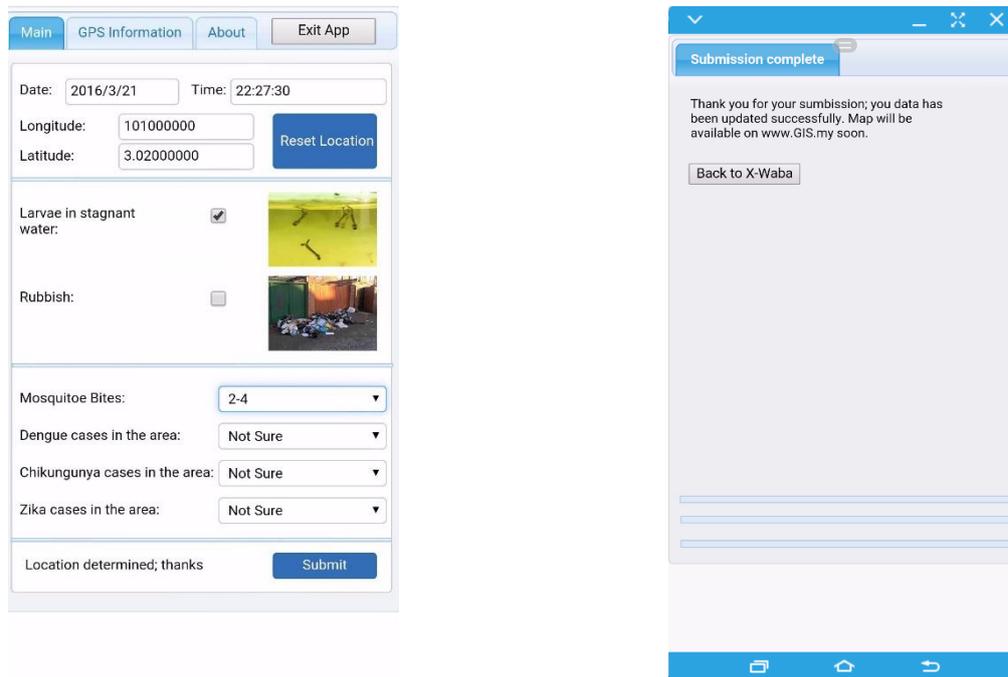


Figure 3. X-Waba crowdsourcing app interface for android phones.

4. Dengue Index

Currently there are three indices used to determine the mosquitoes populations and the risks of dengue outbreak; they are namely Container Index (CI), House Index (HI) and Breteau index (BI). These entomological indexes are based on the calculation of the number of mosquitoes trapped on the number of traps within a certain area. By definition, Bi is the number of positive containers per 100 houses [13] however, the index does not reflect potential dengue transmission nor the areas with greater risks [7].

- The House (premises) Index (HI) or Aedes Index: Percentage of houses or premises positive for Aedes larvae.
- Container Index (CI): percentage of water-holding containers positive for Aedes larvae.
- Breteau Index (BI): number of positive containers per 100 houses in a specific location.

Enhancing the Dengue Index

The developed index will be generating the scalable dengue index outbreak. The index will be based on the Breteau index the key measure to quantify the severity of the dengue outbreak prediction. The proposed Dengue Index would be calculated based on additional four factors. The factors have been classified into: Temporal and Spatial factors. These factors are namely:

- Land use (Spatial Factor)
- Rainfall (Temporal and Spatial Factor)
- Temperature (Temporal and Spatial Factor)
- Moon Phase (Temporal Factor)
- Crowdsourced data (Temporal and Spatial Factor)

A Boolean value of “YES” or “NO” for each factor. A value of 1 if the criteria is satisfied and a value of 0 if it is not. As example, we will give a value of 1 if the temperature is above 28 degrees Celsius or 0 if it is less. The crowdsourced data would be the spatial component of the index in order to locate areas with higher dengue outbreak probabilities. This is solely relying on the civic engagement. The proposed index formula would be based on the calculation from all active temporal and spatial factors calculated on a scale of 0 to 100. Where zero would be the indicative of a no risk of contracting DF or

any A. aegypti based disease and 100 being the most critical in the event if all the factors are available. The proposed formula would give a 20% credit to each factor (value based on the predetermined Boolean value of the factor Yes/No). There are fixed parameters and temporal/spatial parameters (Figure 4).

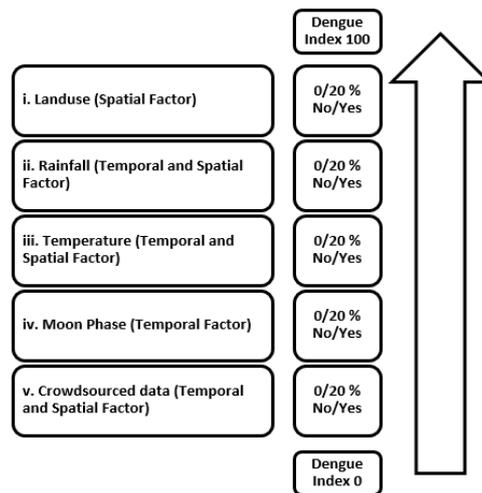


Figure 4. Calculation of Dengue Index

The index will be generated based on the following formula:

$$I = Lu + Pr + Te + Mp + Cs \quad (1)$$

where:

Lu: Spatial location with a value of 20% on residential and 0 for other locations

Pr: Spatial location during rainfall where raining period exceeds 25 days with a value of 0 during periods of rain and 20% during periods of drought.

Te: Temperature component with a value of 0 when the temperature is below 28° C and 20 when temperatures are 28° C and above.

Mp: Moon Phase value will take the following values:

New Moon Phase:0

First Quarter Phase:10%

Full Moon and Third Quarter:20%

Cs: Crowdsourced Data, values will vary as follow:

No data: 0

Uncollected rubbish: 5%

Mosquito Bites: 5%

Mosquito Larvae: 5%

Known Dengue case in the area: 5%

In the event if no spatial data is available from crowdsourcing, the data will be represented as time series only. The goals of representing the data in time series analysis are to identify the phenomenon in a sequence of observations, and attempt to forecast future dengue outbreaks. The patterns will be observed from the time series data. Once patterns established, we will be able to interpret and integrate it with current data. We can then extrapolate to identify pattern to predict outbreaks. In this research,

temperature, rainfall and moon phases along with dengue data will be the ones represented as time series. The analysis based on searching for clusters of temperature, rain and dengue outbreaks. The data will be compared to detect if there is any relationship between the variables and the dengue outbreak phenomenon. The method used for the analysis of the time series will be “Exploratory” Data Analysis (EDA). This method is made by summarizing the data sets with a visual method to determine the existence of a correlation.

4.1 Generating the Index

The dengue index has been generated using Microsoft Excel as per the formula. The formula would give a 20% credit to each factor. The trend line has been also generated using Microsoft Excel. The method used is “Moving Average” with a period Value of “30”. The trendline evens out fluctuations in data to show a pattern or trend more clearly. The period which has been set to 30, as the number of days (a month) will be used by this method by taking 30 days data/points and average them to be represented as a point in the line. A moving average trendline uses this equation:

$$F_t = \frac{A_t + A_{t-1} + \dots + A_{t-n+1}}{n} \quad (2)$$

The index calculation has been carried out with the following values to be “True” or equal to 1. Crowdsourcing data has not been taken in consideration as no data from that time period was available. All the Crowdsourcing data (Represented as Cs in the formula) has been set to 0.

Appendix A respectively shows the time series with the overlapping trendline representing the dengue index for the year 2014 and 2015. The observation that we have made is that the graphs representing the number of outbreak sites replicates the period of the trendline but with a phase shift of approximately 30 days.

In Figure 5 and Figure 6 Appendix A we have taken a few values to estimate the time between the highest amplitude of the trendline in T_Index1 which occurs in day 202 and the actual amplitude of the dengue outbreak data T_Outbreak1 which is estimated at day 241. By subtracting T_Outbreak1 from T_Index1. The results are shown in Table 2.

Table 2. Days laps calculation of dengue index and outbreaks observations

	T_Outbreak(n) Days	T_Index(n) Days	Dt(n) Days
Dt1	241	202	39
Dt2	344	290	54
Dt3	135	65	70
Dt4	241	195	46
	Average (days)		52.25

5. Results and Discussion

We have calculated 4 different days laps made form observation, we would then calculate an average day laps which is the mean of the pre-calculated. In average, the dengue index could predict an outbreak roughly about 52 days before it occurs. However, the amplitude of the outbreak on the Dengue Index trendline does not match the one derived from the actual data. These differences of days lapsed would depend on the factors used to calculate the index. The most important is the temperature and the rainfall. The longer are the dry seasons, the more dengue outbreak sites would be created by the mosquito. Rain from our analysis affects the mosquito’s life cycle but not as much as the temperature. This is due to the availability of artificial breeding sites which provide water for the larvae to leave without predators. The amplitude can be adjusted with crowdsourcing. The App would provide the sufficient information to confirm the dengue index results and would further enhance the results. In order to collect crowdsourcing data, we have developed an App which we named X-WABA

and posted on Google Play since August 2016, there is little interest so far from the public as there is no advertising made. X-WABA app has been developed only for Android devices.

6. Recommendations for future research

The research time frame which consists of 2 years is in our opinion not enough to finalize this research. We would suggest applying the same methodology in data to come to confirm the validity of the index. The formula of the enhanced dengue index is flexible and would allow to change the ceiling temperatures and rainfall depending on the county's weather where it is used by adapting the variables on the local criteria's. To further confirm the influence of the vector's landing habit by the moon phases, we propose that a research on this variable to be conducted thoroughly by comparing the existing data from the dengue outbreaks in the region with the moon phases and establish if there is a relationship between the dengue outbreak phenomenon and the moon phase variables. In Malaysia, we would suggest that in order to establish this relationship, the ministry of health to adjust the dengue data acquisition from the Gregorian calendar weeks to the moon calendar week. By the same way, locations of the location of the outbreaks are mainly in residential areas, we would advise the ministry of health to consider widening the data to the working place or the schools and keeping two sets of data. This will help to pinpoint the true origin of the outbreak. A person can carry the virus contracted from the factory and take it home. In short period of time, the same person would infect neighbours at his/her residential neighbourhood with the help of the vector. The new infected person would then without knowing it carries the virus itself to a new location with no dengue cases prior to this. Once the index calculated and localized spatially. We also suggest to use remote sensing to depict the temperatures to generate the index and display it spatially in the form of a heat map layer using the point density method.

7. Conclusion

The proposed enhanced dengue Index requires regular updates through daily data feeds. The challenge is to encourage citizens to be involved in providing data for crowdsourcing. Gathering other data such as temperature, rainfall and moon phases can be achieved easily through existing weather services. The enhanced dengue index takes in consideration temporal and spatial factors rather than information gathered in small areas. By widening the range of data accumulated from crowdsourcing this would be an achievement in the urban health sector. By using X-Waba could be used by governments to control and predict outbreaks. The crowdsourcing system can be beneficial for the government to point faster the locations prone to dengue. By using VGI alerts from the citizens. However, such ambitious project will only be successful with the continuous advertising and backup of the government.

The dengue index will be a good way to quantify the probability of an outbreak. Given the fact that all the data provided in by the ministry of health in Malaysia are post-outbreak, such initiative will help to inform the population about the high risks of the outbreaks.

Moon being a factor for mosquitoes activities is not conclusive however a detailed study of this factor deserves further consideration. The importance of such confirmation is critical in fighting against the dengue fever. Knowing that the vector activity increases during full moon and the third quarter of the moon would be beneficial to all researchers in the region.

References:

- [1] Mulligan K, Elliott S J and Schuster-Wallace C 2012 The place of health and the health of place: Dengue fever and urban governance in Putrajaya, Malaysia *Health Place* **18** 613–20
- [2] WHO 2010 *Why urban health matters*
- [3] WHO 2015 *WHO | Dengue and severe dengue* (World Health Organization)
- [4] WHO 2014 *WHO | Vector-borne diseases* (World Health Organization)
- [5] Mokraoui L and Noor N M 2016 Developing System On Urban Health Using Crowdsourcing Method: A Case Study Of Dengue Outbreak
- [6] Slooff R, Verdrager J and Organization W H 1972 *Anopheles balabacensis balabacensis* baisas

- 1936 and malaria transmission in South-Eastern areas of Asia
- [7] Dibo M R, Chierotti A P, Ferrari M S, Mendonça A L and Neto F C 2008 Study of the relationship between *Aedes (Stegomyia) aegypti* egg and adult densities, dengue fever and climate in Mirassol, state of São Paulo, Brazil *Mem. Inst. Oswaldo Cruz* **103** 554–60
- [8] Salawu A, Safie M, Bello A A and Bernad S C 2013 Geo Visualization of Disease Outbreak Using Geospatial Crowd Sourcing Web Application *Int. J. Res. Rev. Appl. Sci.* **16** 370–6
- [9] Lwin M O, Vijaykumar S, Fernando O N N, Cheong S A, Rathnayake V S, Lim G, Theng Y-L, Chaudhuri S and Foo S 2014 A 21st century approach to tackling dengue: Crowdsourced surveillance, predictive mapping and tailored communication *Acta Trop.* **130** 100–7
- [10] Abdullah S A and Nakagoshi N 2007 Forest fragmentation and its correlation to human land use change in the state of Selangor, peninsular Malaysia *For. Ecol. Manage.* **241** 39–48
- [11] Mudin R N 2015 Dengue Incidence and the Prevention and Control Program in Malaysia 5
- [12] MCMC 2015 Hand Phone Users Survey 2014 *Communications* 48
- [13] Wee L K, Weng S N, Raduan N, Wah S K, Ming W H, Shi C H, Rambli F, Ahok C J, Marlina S, Ahmad N W, McKemy A, Vasan S S and Lim L H 2013 Relationship between rainfall and *Aedes* larval population at two insular sites in Pulau Ketam, Selangor, Malaysia *Southeast Asian J. Trop. Med. Public Health* **44** 157

Appendix A:

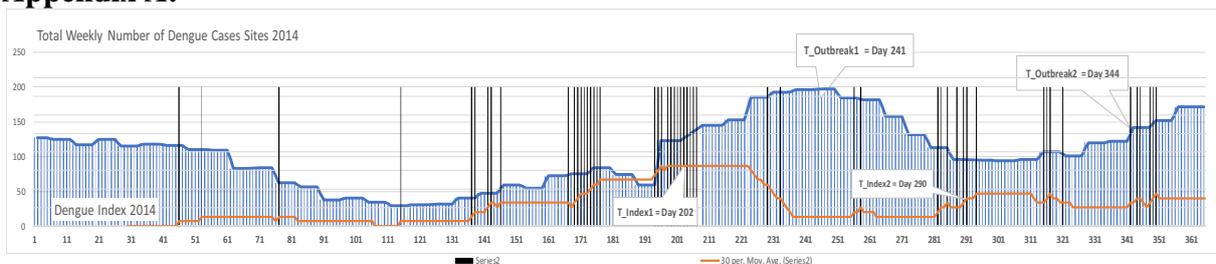


Figure 5. Correlation of Dengue Index and Number of Outbreak Cases - 2014 Data

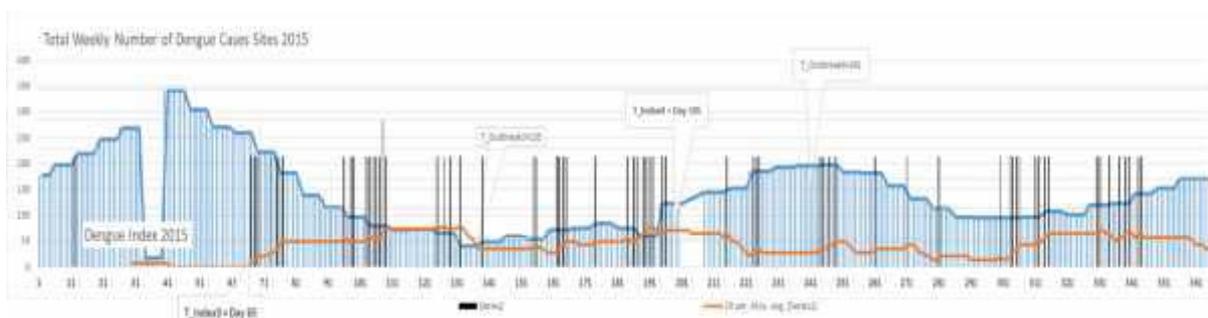


Figure 6. Correlation of Dengue Index and Number of Outbreak Cases - 2015 Data