
Denguenator

— Predicting Dengue outbreaks using
Human Mobility and Climatic factors —

Content

Introduction

Background

Data set

Preprocessing

Mapping different geographical areas

Time expanded contact networks

Hidden Markov Model

Metapopulation Model

Introduction

- Dengue is a major health concern in Sri Lanka, which causes about 30,000 cases and 200 deaths annually[1].
- Many countermeasures have been taken, but still is not possible to reduce to a significant level.
- The aim of this project is to build a predictive model of Dengue propagation in Sri Lanka based on human mobility and other climatic factors.



Background

Researches show that there is high probability of causing Dengue outbreaks with horizontal propagation[2]

- Human mobility is an important factor

Data Set

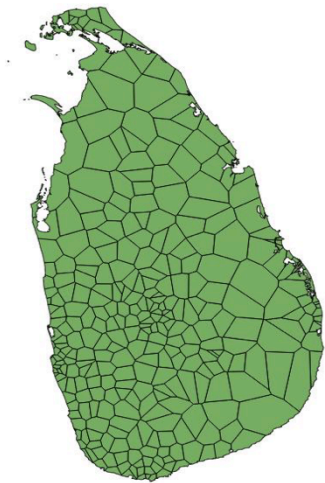
- Call Detail Records
- Population Data
- Meteorological Data
 - Rainfall Data
 - Temperature Data
- Past Dengue Cases



Pre processing

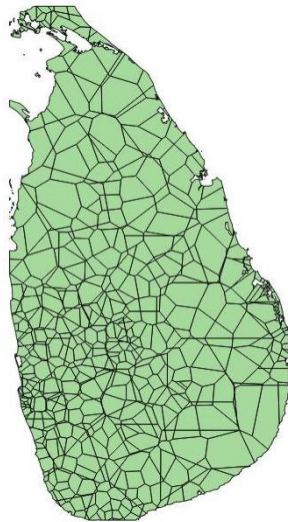
Data we have corresponds to different geographical areas...

HOW to map them together???



MOH areas

+



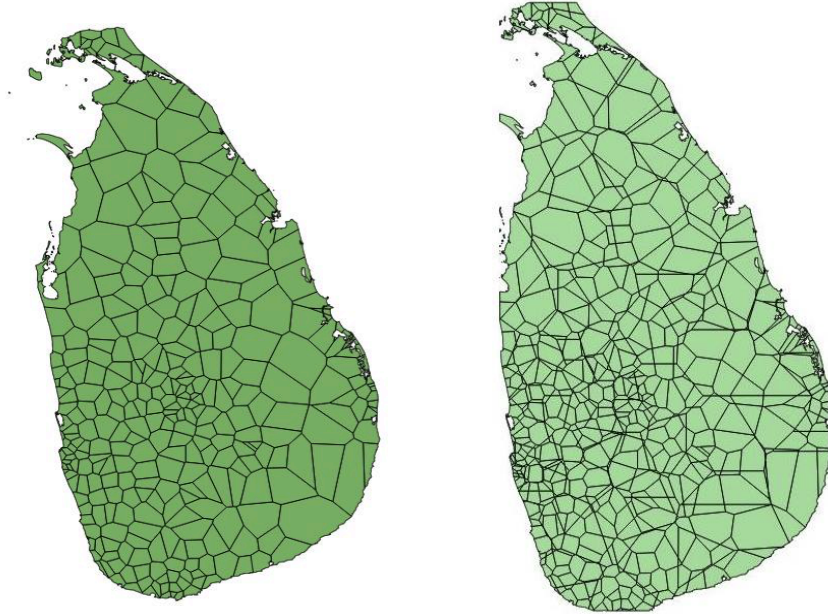
Weather Stations

=>

???

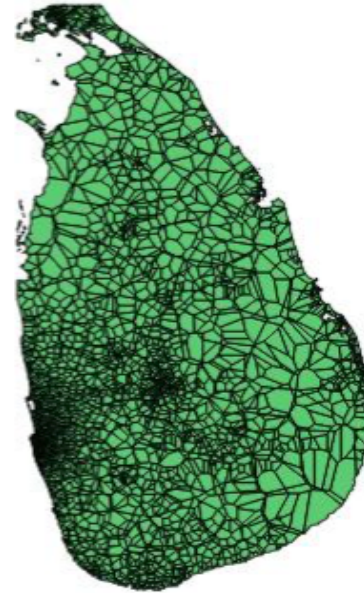
Mapping different geographical areas

- Creating Voronoi diagrams



Mapping different geographical areas

- Creating Voronoi diagrams
- Overlapping each other



MOH areas and cell towers overlapping map

Mapping different geographical areas

- Creating Voronoi diagrams
- Overlapping each other



- Assign each unit (towers or weather stations) to MOH level
- Calculate MOH level values

Mobility in MOH level

Get the list of the cell towers corresponding to a certain MOH area

Keep the fractional coverages of each of them

Using this fractional coverage values, get the mobility between MOH areas.



Mobility to MOH level

$$m_{MOH} = \sum_{i=1}^n \frac{1}{2} \left(\frac{1}{m_i} + \frac{1}{m_{i+1}} \right) \left(\frac{1}{m_i} - \frac{1}{m_{i+1}} \right) \ln \left(\frac{m_i}{m_{i+1}} \right)$$

Weather to MOH

Filling missing data

Rainfall data

Discarded weather stations which are missing more than 50% of the data

Inverse Distance Weighting (IDW) method[3]

$$R_i = \frac{\sum_{j=1}^n \frac{R_j}{d_{ij}^2}}{\sum_{j=1}^n \frac{1}{d_{ij}^2}}$$
$$R_i = \frac{\sum_{j=1}^n \frac{R_j}{d_{ij}^2}}{\sum_{j=1}^n \frac{1}{d_{ij}^2}}$$

Weather to MOH

Filling missing data

Temperature data

Discarded one weather station

Filled the missing values by the average temperature

Weather to MOH

Aggregating to weeks

Getting weather stations corresponding to a MOH

Getting fractional coverages

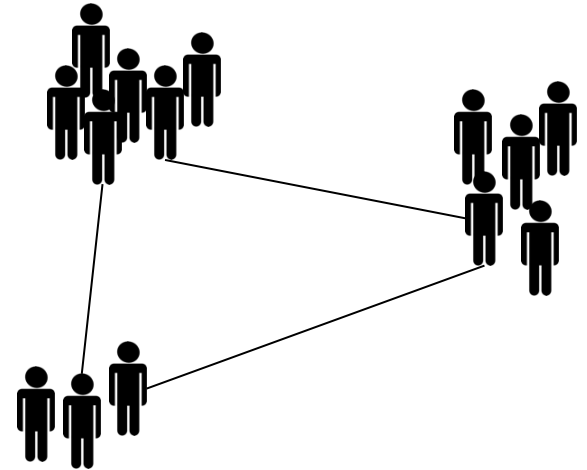
Using fr
level



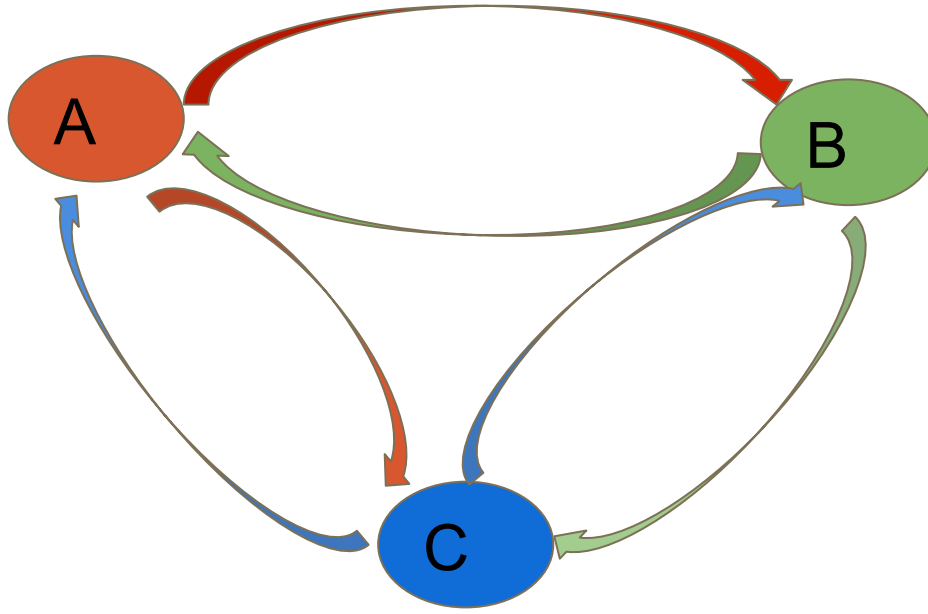
1e MOH

Methods to Predict Dengue Propagation

1. Time-expanded contact network
2. Metapopulation model

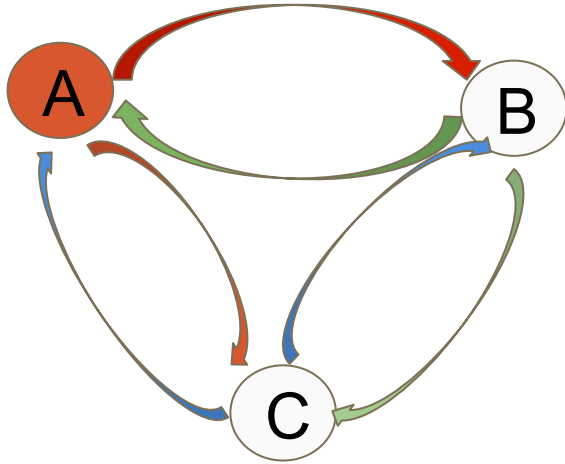


Time-expanded contact network

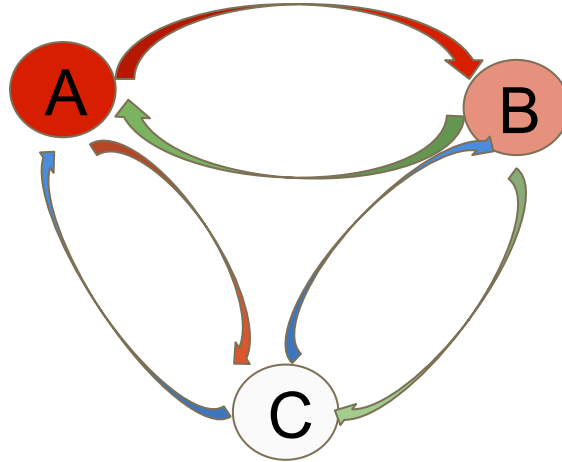


Time-expanded contact network

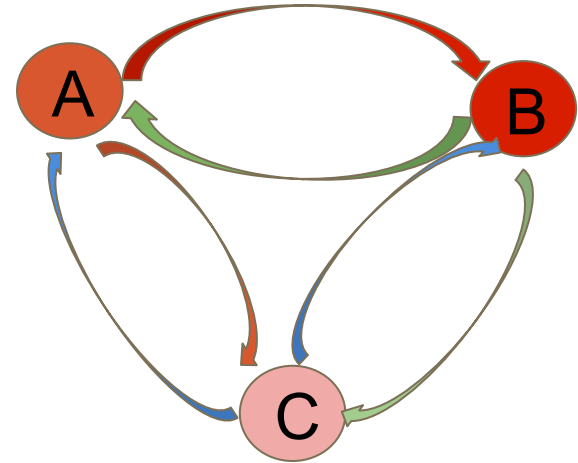
Consider three weeks :



Week 1



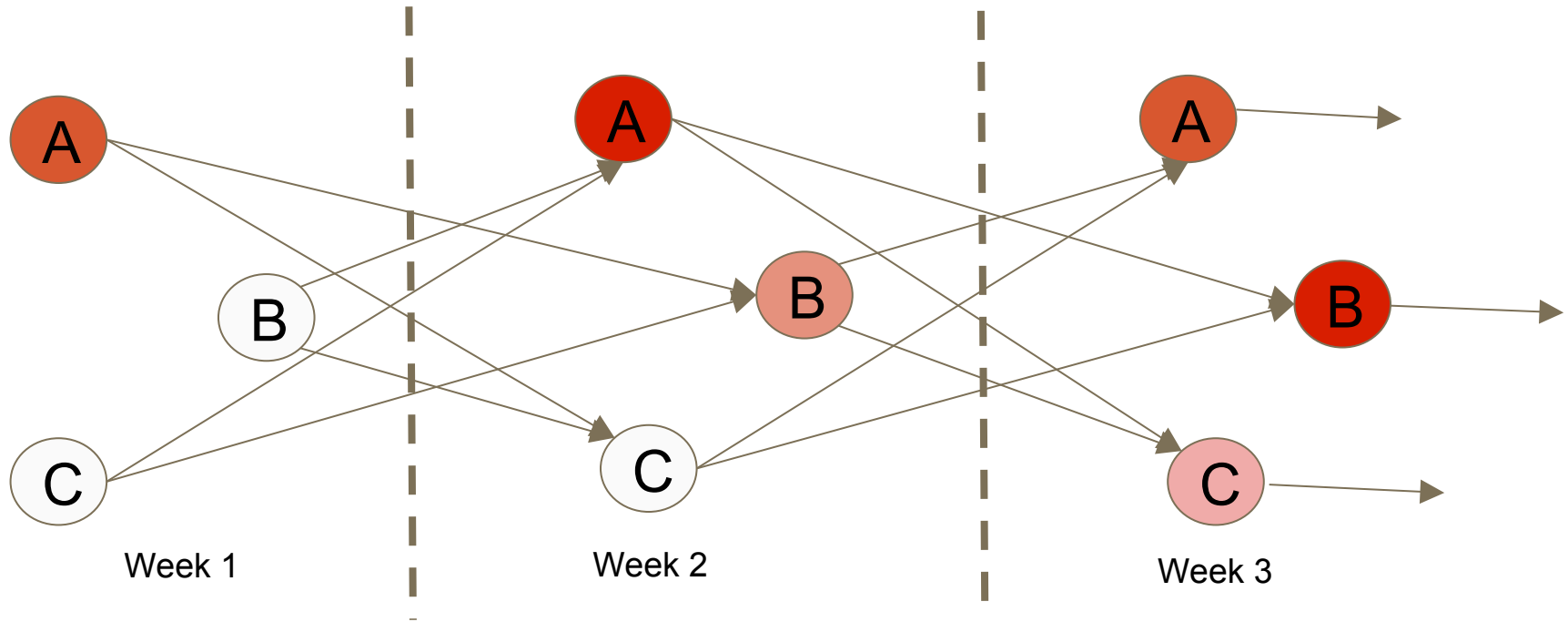
Week 2



Week 3

Time-expanded contact network

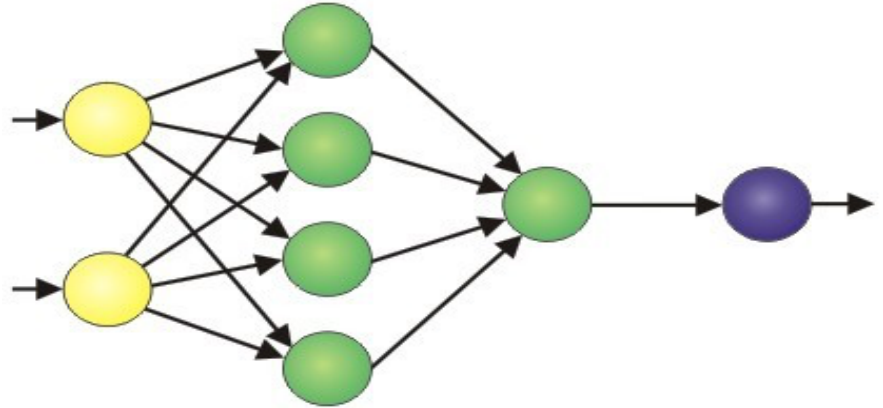
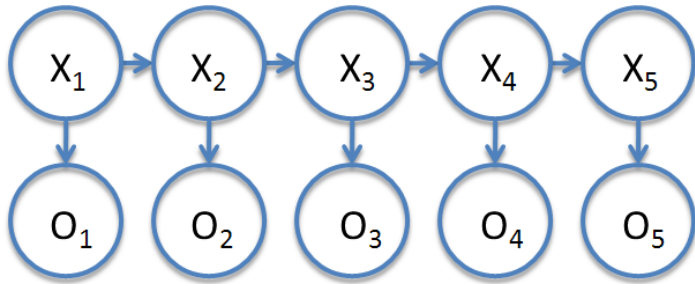
How to get connections between weeks :



Time-expanded contact network - implementation

Hidden Markov Models

Artificial Neural Networks

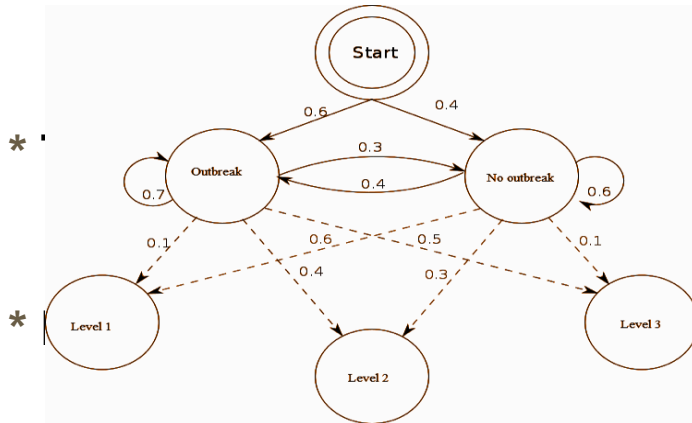


Hidden Markov Model

How HMM is suitable for this problem ?

There is a hidden state - Next week's dengue state

There are related observation - * This week's dengue states



Hidden Markov Model

Two Approaches :

Separate HMM for each neighbor

Single HMM for all the neighbors

- Existing algorithms do not support for multiple emission sequences for a single hidden sequence

- Thousands of observation states for a single hidden state that makes predictions unreliable

Hidden Markov Model

How to define neighbors :

- get all the inflow mobility values from all the MOH areas to considering MOH area
- get mobility per host for all the MOH areas by dividing the above values by the populations of MOHs
- get the average mobility per infected host by multiplying the above value by average dengue cases of MOHs
- mean of all the average mobility per infected host values is considered as the threshold value (eg: For Moratuwa MOH, threshold value is 0.375)

Hidden Markov Model

How to define states :

- Using discrete values

How to define discrete values :

- Using intervals

How to define intervals :

- Observed the data set to define the size of the intervals based on the weekly reported dengue cases (identified as 1 - 10)
- Used automated brute force approach to select the best value which gives the minimum RMSE value for predictions (identified as 4 for most cases)
- eg : 0-3 => 1, 4-7 => 2, 8-11 => 3, for value 6 state should be 2

Hidden Markov Model

MatLab - HMM toolbox is used

Hidden states sequence - start from week 2

Emission sequence (Mobility per infected host of neighbours, rainfall of candidate MOH, previous week dengue level of candidate MOH) - start from week 1

Separate sequence for each neighbour

Estimate the state transition matrix and observation emission matrix that can be considered as the model training step

Hidden Markov Model

How to get predictions :

- Using Viterbi algorithm
- The Viterbi algorithm is a dynamic programming algorithm for finding the most likely sequence of hidden states
- Give hidden state for a given observation, using observation emission matrix and state transition matrix

eg : Next week's dengue level of Moratuwa =

HMM_Viterbi (This week's dengue level, matrix_1, matrix_2)

Hidden Markov Model

How to merge predictions :

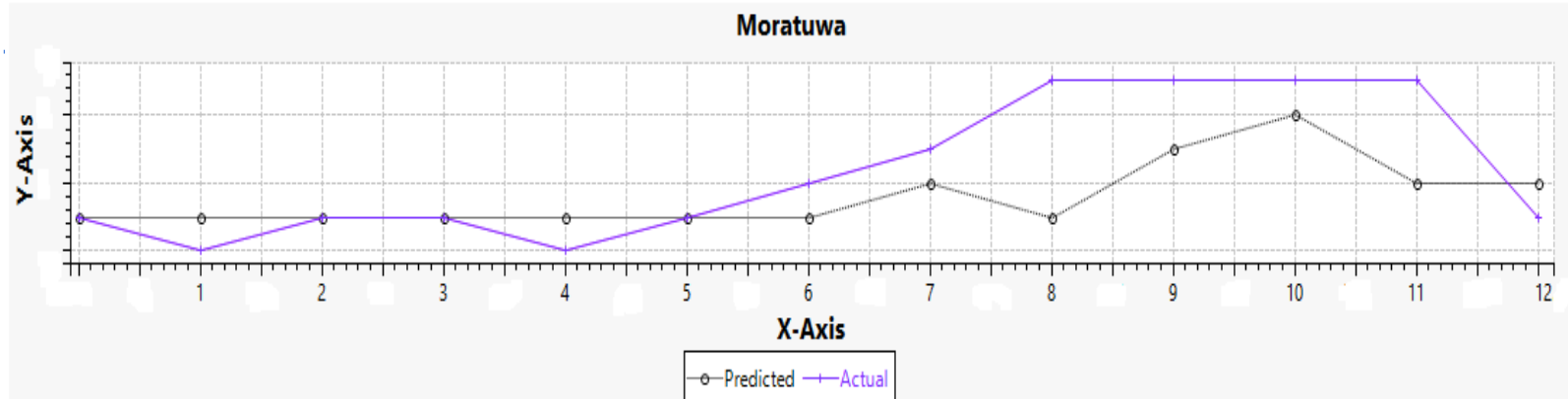
- get predictions for this week using previous week's data
- get error using actual value and prediction (error = actual value - prediction)
- define a weight function ($w = 1/\text{error}$)
- get weight functions for each neighbor (w_1, w_2, \dots, w_n)
- get predictions for next week from each neighbor (p_1, p_2, \dots, p_n)
- get final prediction (p)

$$p = (w_1 * p_1 + w_2 * p_2 + \dots + w_n * p_n) / (w_1 + w_2 + \dots + w_n)$$

Hidden Markov Model

Results for Moratuwa MOH area :

- 70% of data for training, 30% of data for testing
- Root mean square error = 5.65685



Hidden Markov Model

Many limitations.....

- Merging results from many models

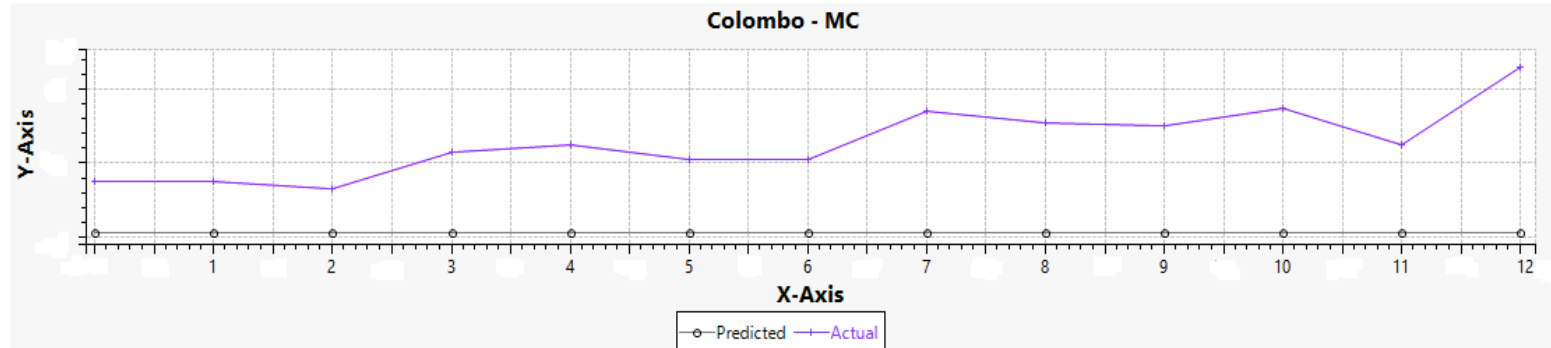
- Cannot get prediction for a mobility per infected host state which is greater than the maximum state in train sequence

- Limited set of training data

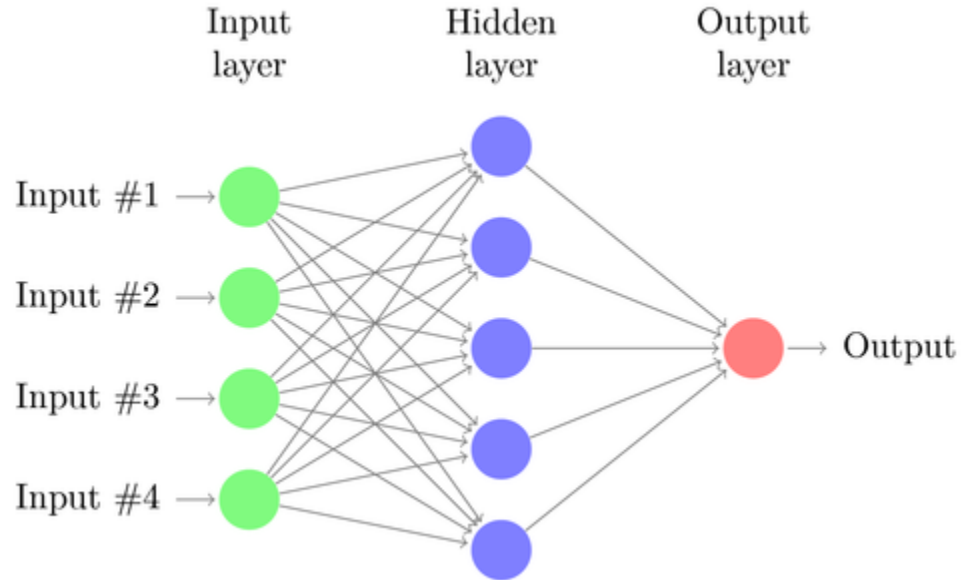
- Model is specifically tuned to observe mobility, therefore cannot get predictions for dengue hotspots like Colombo - MC

Hidden Markov Model

- Dengue level of Colombo - MC is comparatively very high than the neighbors
- Assume neighbors does not affect for the dengue propagation in Colombo - MC



Artificial Neural Network



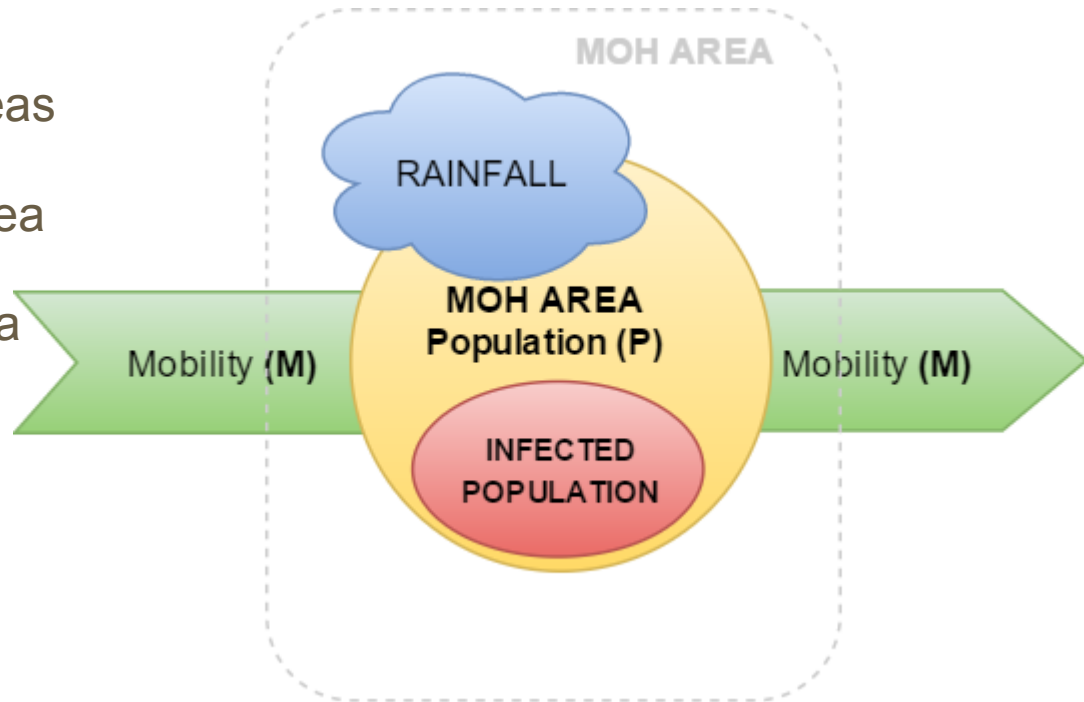
Artificial Neural Network

Data Set

Mobility data between MOH areas

Dengue cases of each MOH area

Rainfall data for each MOH area



Artificial Neural Network

Data preprocessing

Define the top **MOHs** with the highest average mobility per infected host of considering **MOH**

A - Average dengue cases of the considering **MOH**

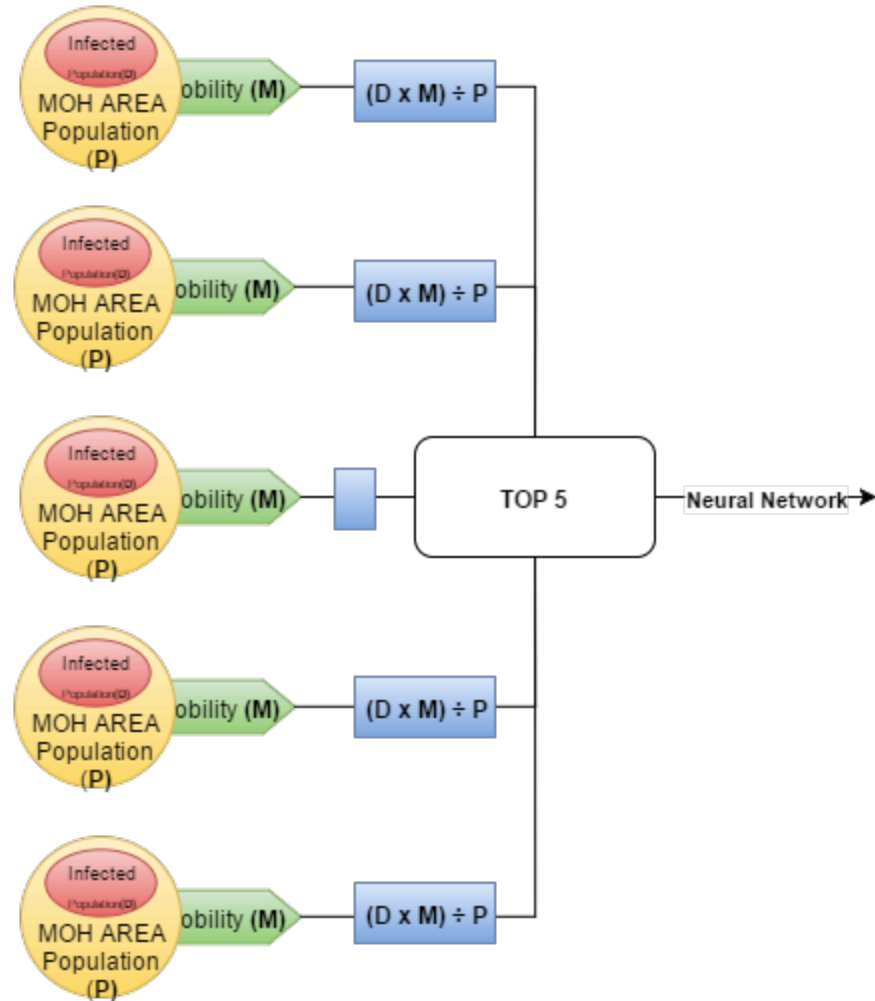
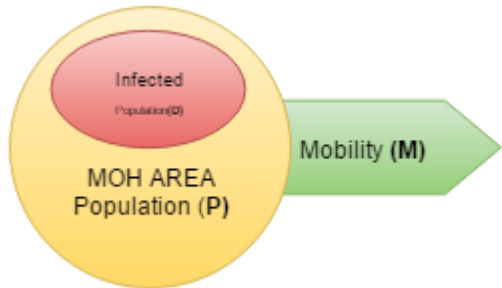
M - Mobility with the candidate **MOH**

P - Population of the considering **MOH**

Average mobility per infected host of considering **MOH** = $(M * A) / P$

Artificial Neural Network

One MOH Division



Artificial Neural Network

Week Number	MOH1	MOH2	MOH3	MOH4	MOH5	Last Week Rainfall	Last Week Dengue Cases	This Week Dengue Cases
1	23	12	43	3	54	10	21	78
2	32	22	51	21	32	22	12	22
3	11	12	32	32	12	13	32	22

Artificial Neural Network

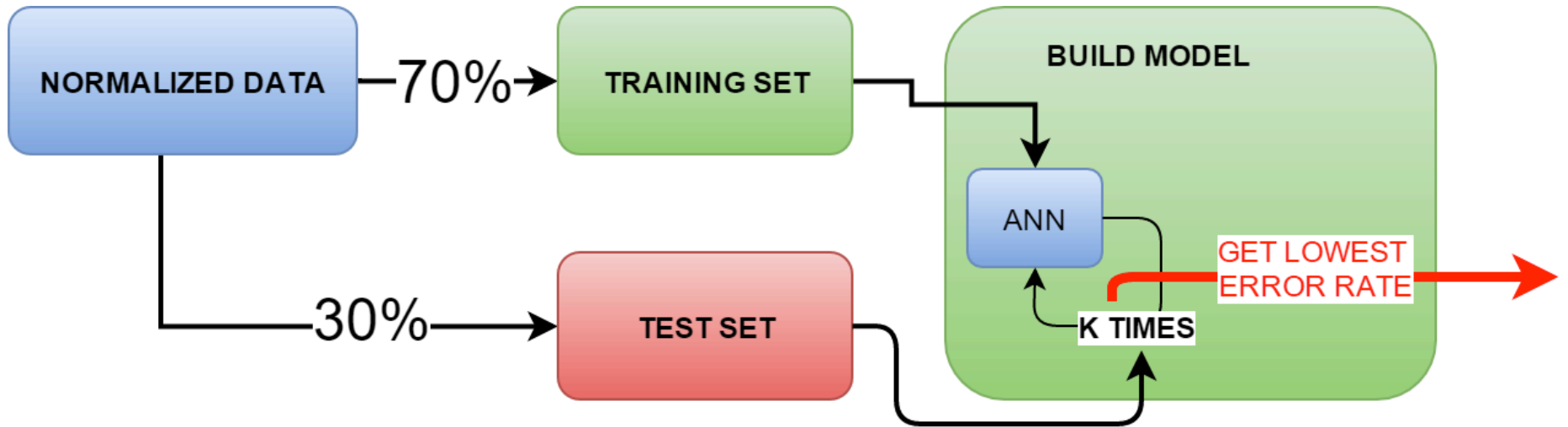
Data preprocessing

Data normalization

In Order to get convergent results the data set was normalized using the following method

$$\text{Normalized Values} = (\text{Value} - \text{Min})/(\text{Max}-\text{Min})$$

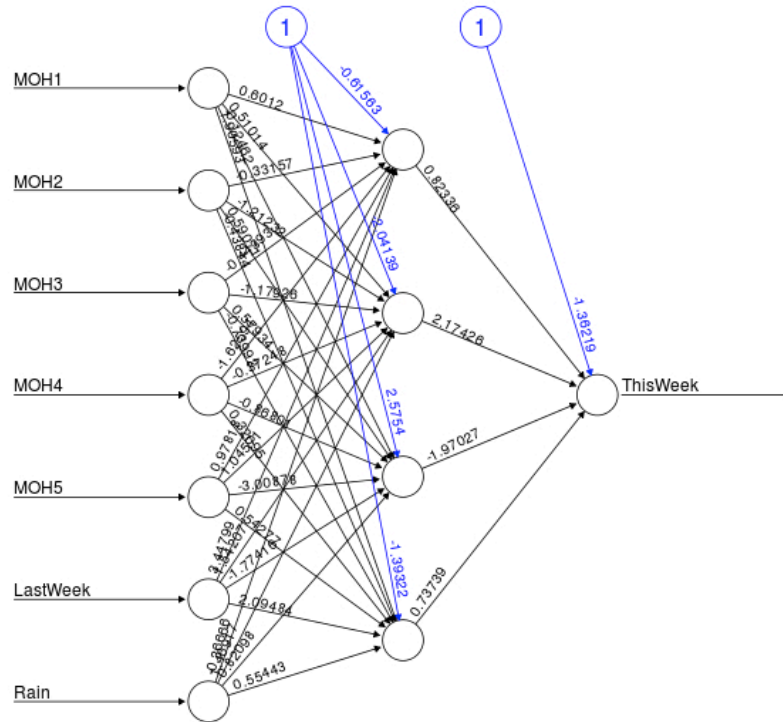
Artificial Neural Network



Artificial Neural Network

Colombo-MC-MOH

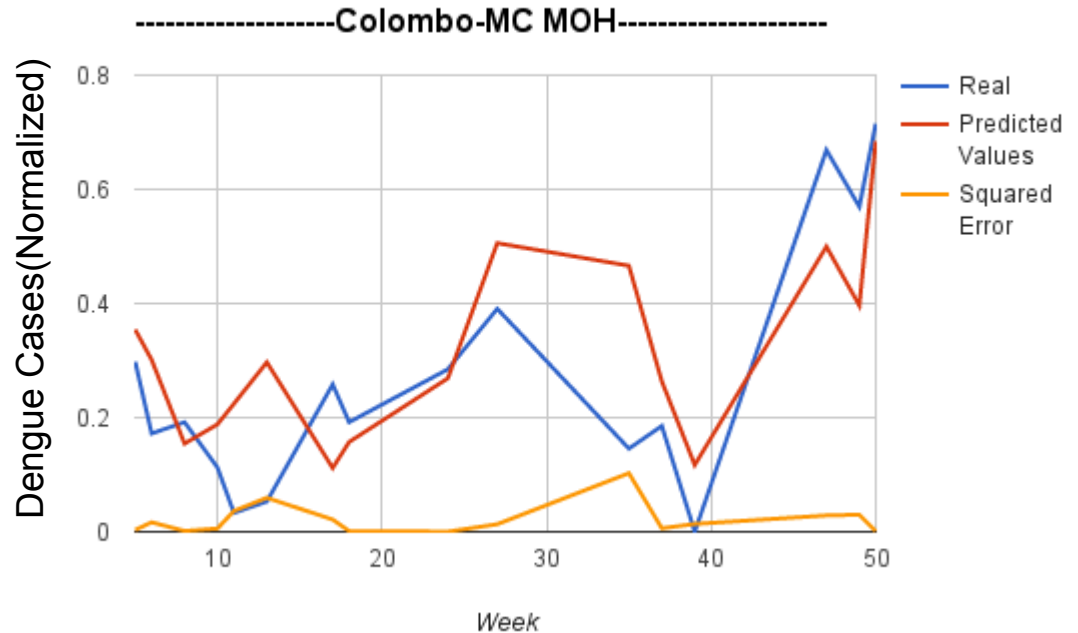
K=4



Artificial Neural Network

Colombo-MC-MOH

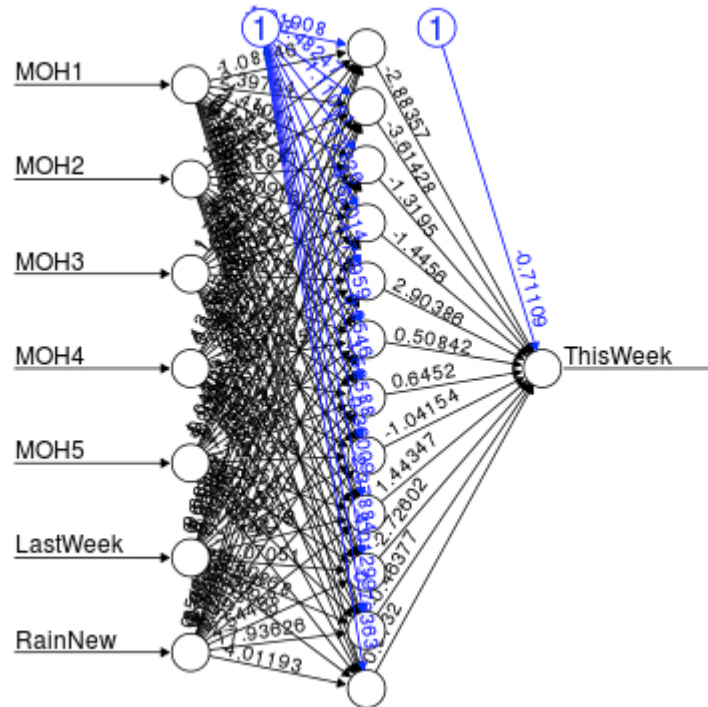
RMS Error
20.68674



Artificial Neural Network

Dehiwala MOH

K=12

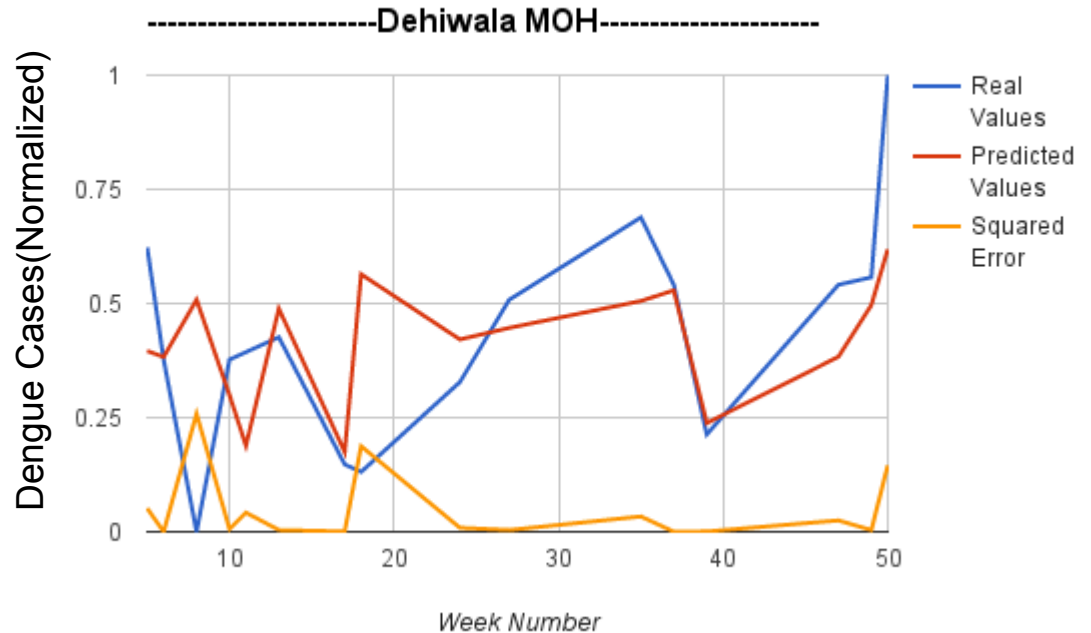


Artificial Neural Network

Dehiwala MOH

RMS Error

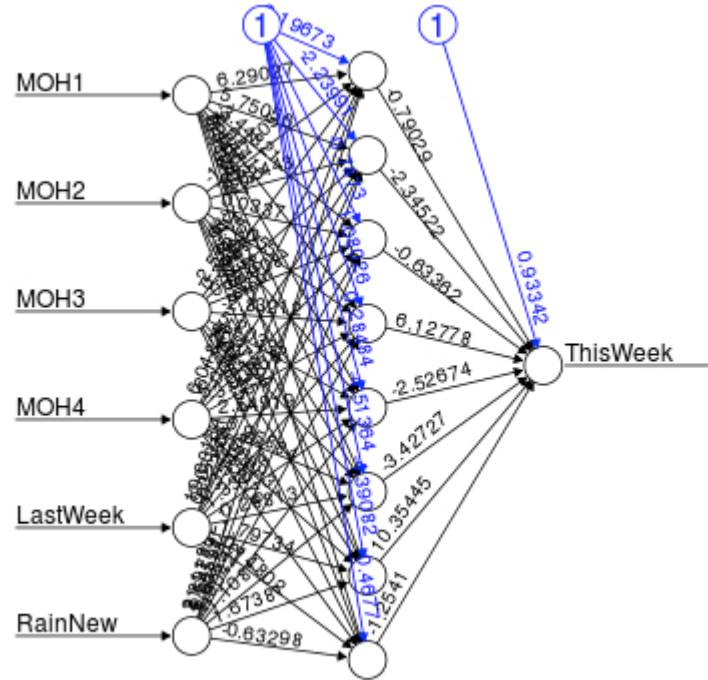
12.93261



Artificial Neural Network

Kolonnawa MOH

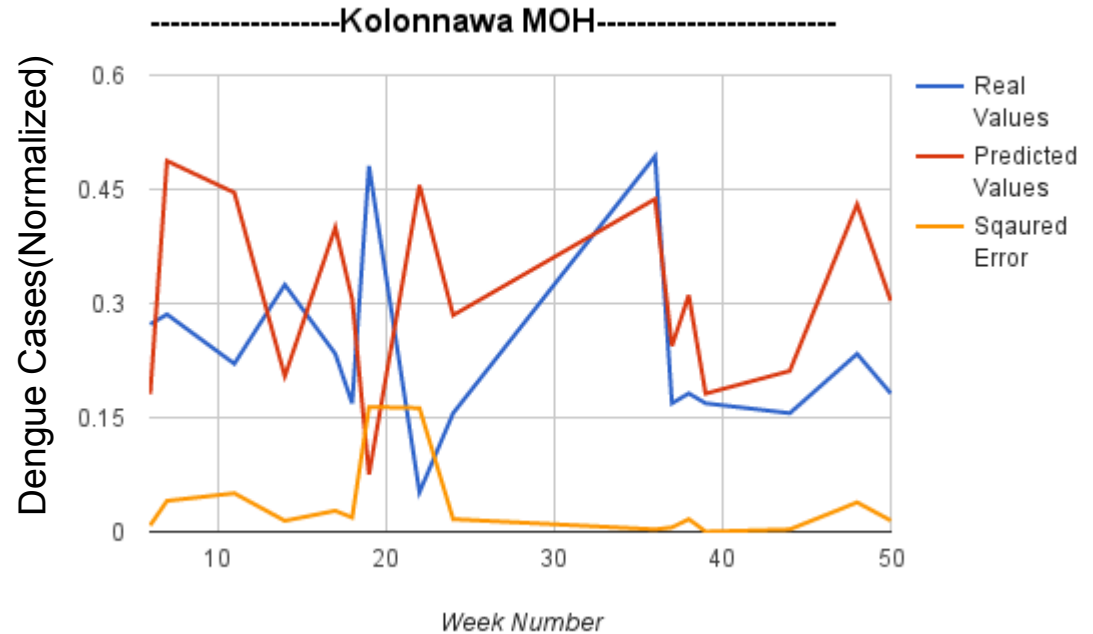
$K=8$



Artificial Neural Network

Kolonnawa MOH

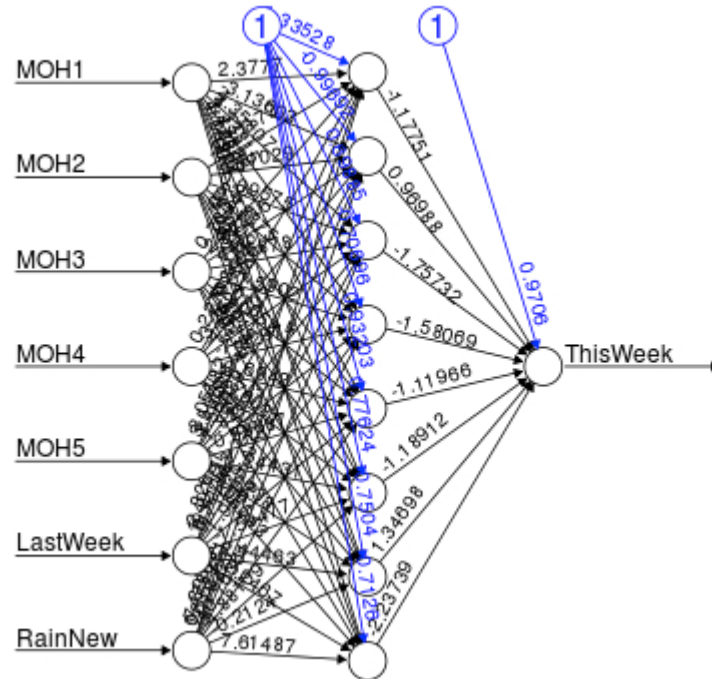
RMS Error
13.3945



Artificial Neural Network

Moratuwa MOH

K=8

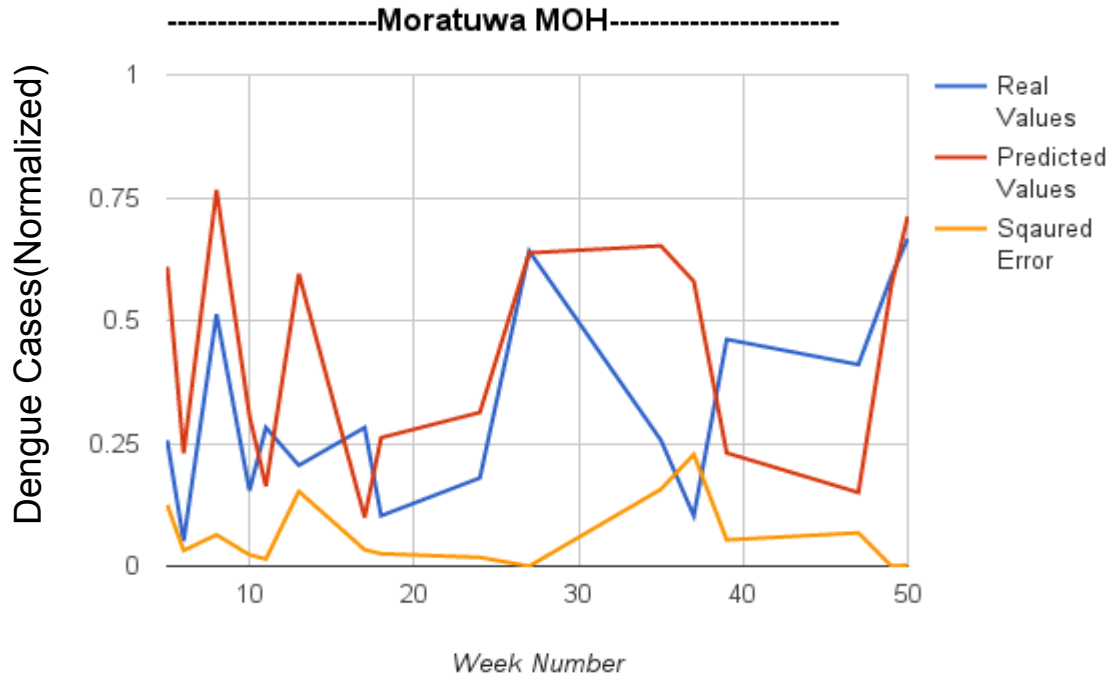


Artificial Neural Network

Moratuwa MOH

RMS Error

9.98



Artificial Neural Network

Limitations

Accuracy of weather data due to lack of weather stations

Lack of data points(52 per ANN)

Artificial Neural Network

Future Improvements

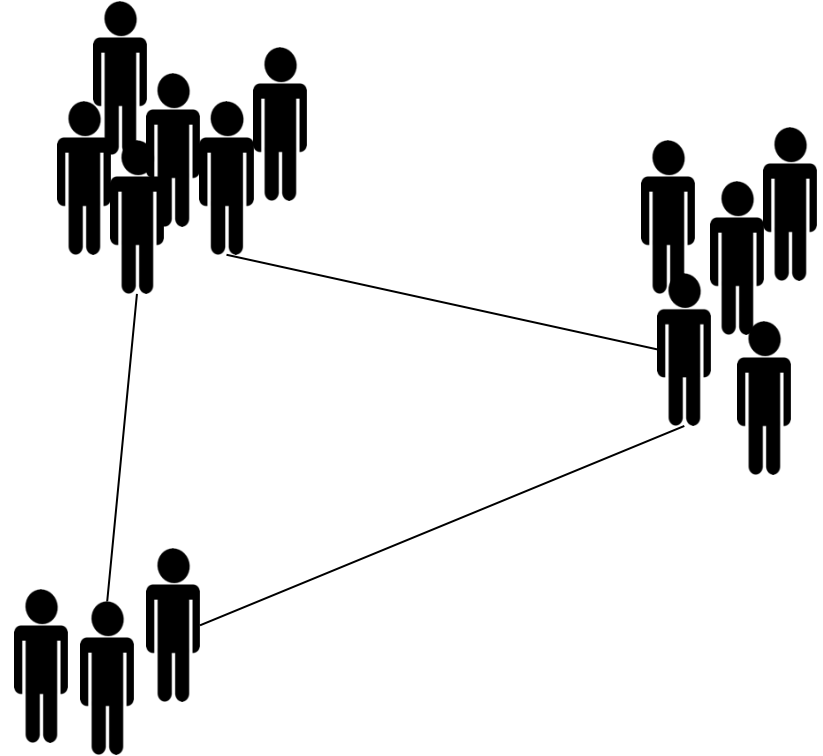
Train a model for the entire country

Metapopulation Model

Metapopulation.

Group of populations that are **separated by space** but consist of the **same species**.

These spatially separated populations interact as individual members **move** from one population to another.



Metapopulation Model

Metapopulation Model

Used to describe internal dynamics of a population

The concept of metapopulation model was introduced by Richard Levins (an American ecologist) in 1969. [Ref]

The Levins model

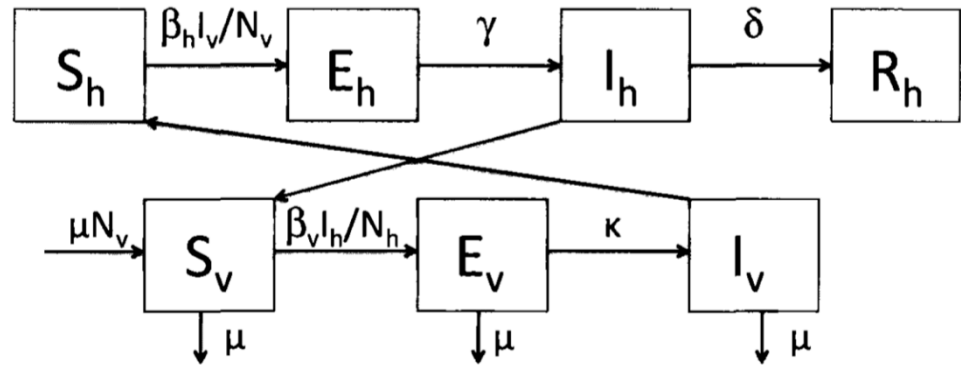
$$dp/dt = mp(1 - p) - ep$$

Metapopulation Model

Dengue Dynamics

Seven class vector host model

Introduced by Carlos Alan Torres in 2000 [50]



Metapopulation Model

$$\dot{S}_v = \mu N_v - \beta_v S_v \frac{I_h}{N_h} - \mu S_v$$

$$\dot{E}_v = \beta_v S_v \frac{I_h}{N_h} - \mu_v E_v - \kappa E_v$$

$$\dot{I}_v = \kappa E_v - \mu_v I_v$$

$$\dot{S}_h = -\beta_h S_h \frac{I_v}{N_v}$$

$$\dot{E}_h = \beta_h S_h \frac{I_v}{N_v} - \gamma E_h$$

$$\dot{I}_h = \gamma E_h - \delta I_h$$

$$\dot{R}_h = \delta I_h.$$

The Basic Model

- Basic model for a single population without the effect of stochastic migration.
- Calculate the rate of change for each types of vectors/hosts.
- Dimensions of each class is either hosts/day or vectors/day

Metapopulation Model

Model With Stochastic Migration

Basic model with two patches

Consider only the host migration

$$\dot{S}_{v1} = \mu N_{v1} - \beta_{v1} \left(P_{11} \frac{I_{h1}}{N_{h1}} + P_{21} \frac{I_{h2}}{N_{h2}} \right) S_{v1} - \mu S_{v1}$$

$$\dot{E}_{v1} = \beta_{v1} \left(P_{11} \frac{I_{h1}}{N_{h1}} + P_{21} \frac{I_{h2}}{N_{h2}} \right) S_{v1} - \mu E_{v1} - \kappa E_{v1}$$

$$\dot{I}_{v1} = \kappa E_{v1} - \mu I_{v1}$$

$$\dot{S}_{h1} = -\beta_{h1} P_{11} S_{h1} \frac{I_{v1}}{N_{v1}}$$

$$\dot{E}_{h1} = \beta_{h1} P_{11} S_{h1} \frac{I_{v1}}{N_{v1}} - \gamma E_{h1}$$

$$\dot{I}_{h1} = \gamma E_{h1} - \delta I_{h1}$$

$$\dot{R}_{h1} = \delta I_{h1}$$

$$\dot{S}_{v2} = \mu N_{v2} - \beta_{v2} \left(P_{12} \frac{I_{h1}}{N_{h1}} + P_{22} \frac{I_{h2}}{N_{h2}} \right) S_{v2} - \mu S_{v2}$$

$$\dot{E}_{v2} = \beta_{v2} \left(P_{12} \frac{I_{h1}}{N_{h1}} + P_{22} \frac{I_{h2}}{N_{h2}} \right) S_{v2} - \mu E_{v2} - \kappa E_{v2}$$

$$\dot{I}_{v2} = \kappa E_{v2} - \mu I_{v2}$$

$$\dot{S}_{h2} = -\beta_{h2} P_{22} S_{h2} \frac{I_{v2}}{N_{v2}}$$

$$\dot{E}_{h2} = \beta_{h2} P_{22} S_{h2} \frac{I_{v2}}{N_{v2}} - \gamma E_{h2}$$

$$\dot{I}_{h2} = \gamma E_{h2} - \delta I_{h2}$$

$$\dot{R}_{h2} = \delta I_{h2}$$

Metapopulation Model

Dengue Mobility

Probability of an infected person from patch “i” visiting patch “j”

$$\sum_{i=1}^n P_{ij} \frac{I_{hi}}{N_{hi}}$$

P_{ij} - Probability of a person from patch “i” visiting patch “j”

Past researches derived P_{ij} using the gravity model

In our study,

$$P_{ij} = \theta \frac{n_i^\alpha n_j^\beta}{d_{ij}^\gamma},$$

MNDB

$$P_{ij} = \frac{m_{ij}}{N_{hi}}$$

Metapopulation Model

λ - Host per capita infection rate (1/5.5 Days)

δ - Host per capita recovery rate (1/4 Days)

μ_v - Vector per capita birth/death rate (1/10.5 Days)

k - Vector per capita infection rate (1/5.5 Days)

μ_h - Host per capita birth/death rate (1/(3*10⁴)Days)

Metapopulation Model

Mosquito Population - N_{vi}

In past researches it is assumed to be 3 times the host population.[3][4]

No credible source of information in Sri Lanka.

Can vary from one MOH area to another.

Solution is to introduce a scaling factor “ θ ”

θ = Number of vectors per host

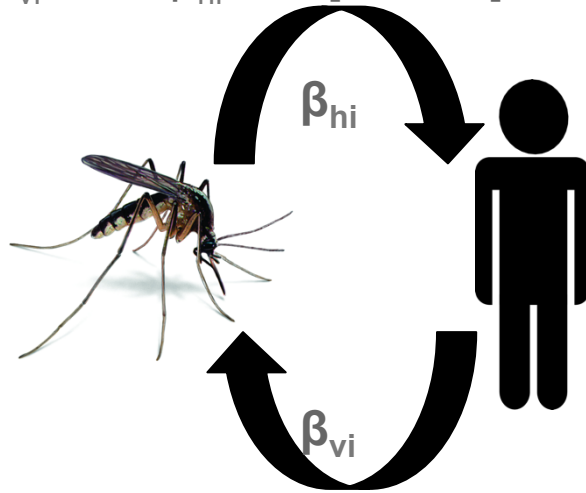
So that, $N_{vi} = \theta N_{hi}$

Metapopulation Model

Transmission Rates - β_{vi} And β_{hi}

Calculate best transmission rates per MOH area that fits for our data

In Peru the ranges for β_{vi} and β_{hi} are [0.2-0.4] and [0.25-0.45]. [3]



Metapopulation Model

Improvements to the model

Introduced θ

P_{ij} derived using MNBD

Removed the effect of μ_{hi} when
calculating initial susceptible
population

$$\dot{S}_{vi} = \mu_{vi}N_{vi} - \beta_{vi} \left(\sum_{i=1}^n P_{ij} \frac{I_{hi}}{N_{hi}} \right) S_{vi} - \mu_{vi}S_{vi}$$

$$\dot{E}_{vi} = \beta_{vi} \left(\sum_{i=1}^n P_{ij} \frac{I_{hi}}{N_{hi}} \right) S_{vi} - \mu_{vi}E_{vi} - \kappa E_{vi}$$

$$\dot{I}_{vi} = \kappa E_{vi} - \mu_{vi}I_{vi}$$

$$\dot{S}_{hi} = N_{hi} - \beta_{hi}P_{ii} \frac{I_{vi}}{N_{vi}} S_{hi} - \mu_{hi}S_{hi}$$

$$\dot{E}_{hi} = \beta_{hi}P_{ii} \frac{I_{vi}}{N_{vi}} S_{hi} - \lambda E_{hi} - \mu_{hi}E_{hi}$$

$$\dot{I}_{hi} = \lambda E_{hi} - \delta I_{hi} - \mu_{hi}I_{hi}$$

Metapopulation Model

The Challenge

How to calculate θ , β_{vi} and β_{hi} For each MOH area ?

Model I_{hi} values for each combination of θ , β_{vi} and β_{hi}

Calculate the root mean squared error (RMSE) of modeled values against actual I_{hi} (past dengue cases) values.

Pick the combination that has the minimum RMSE value

$$RMSE = \sqrt{\frac{\sum (I_{actual} - I_{model})^2}{n}}$$

Can be identified as an optimization problem

Metapopulation Model

Calculating minimum RMSE value

Brute Force approach

θ from 5 to 100 with a resolution of 5

β_{vi} and β_{hi} from 0 to 1 with a resolution of 0.01

Operations per MOH area $19 \times 100 \times 100 \times 37 \times 7 = 48.1\text{Mn}$

Takes around 140 seconds per MOH area

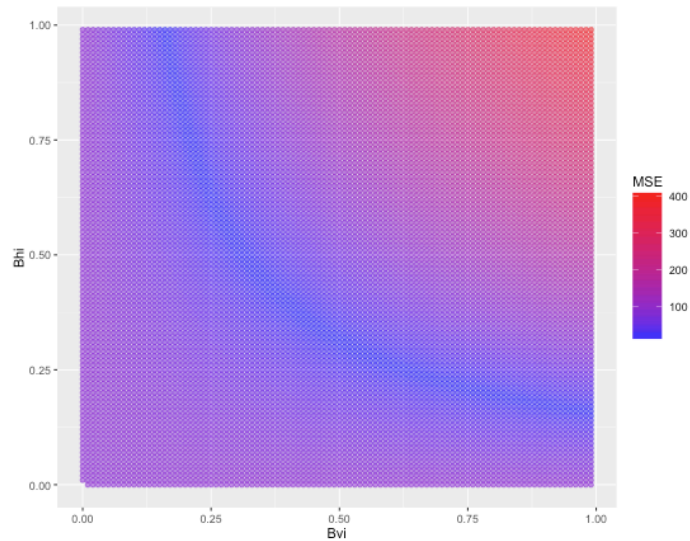
12 hours to calculate for the whole country

Simulated annealing

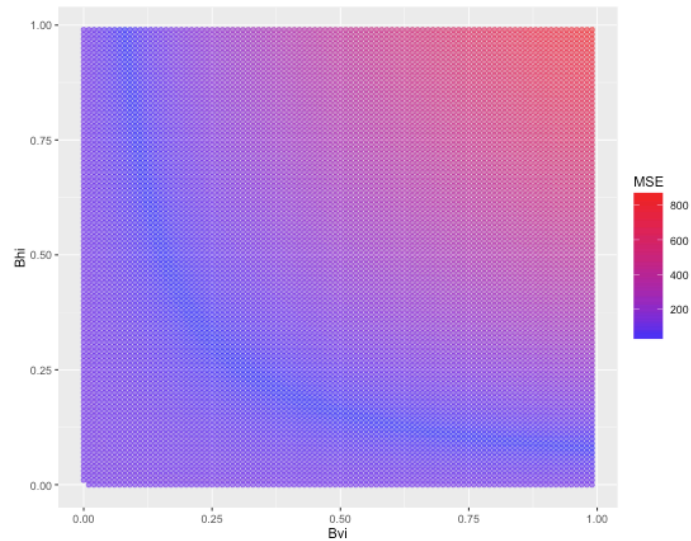
Genetic Algorithm

Metapopulation Model

RMSE density plots for β_{vi} and β_{hi} combinations



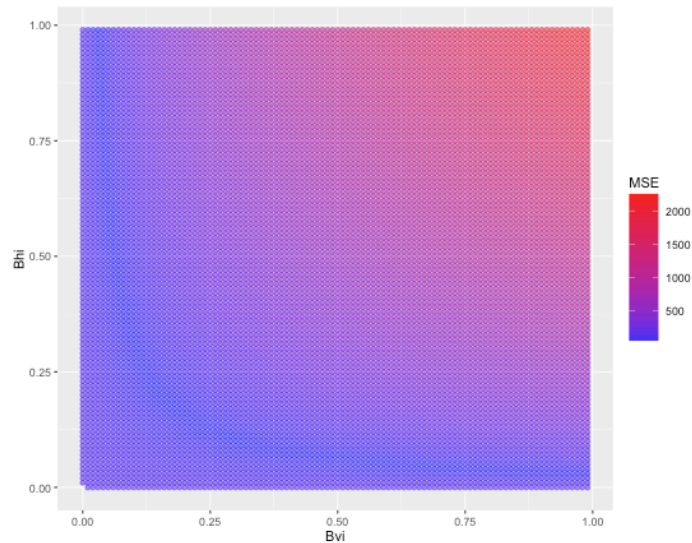
MC- Colombo $\theta = 10$



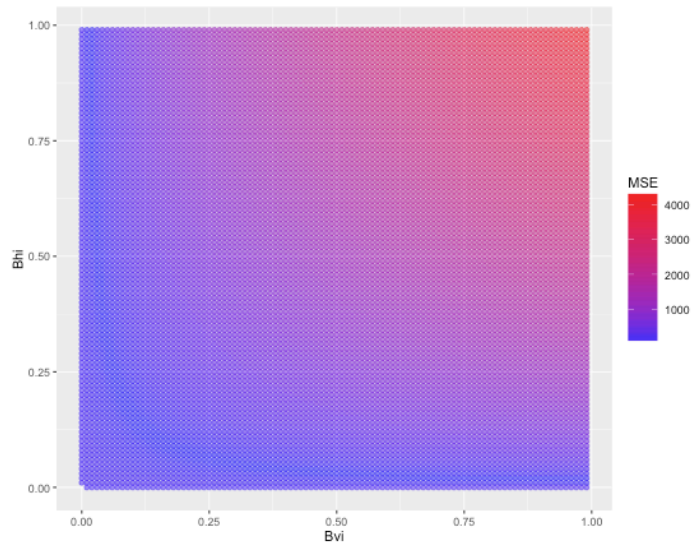
MC- Colombo $\theta = 20$

Metapopulation Model

MSE density plots for β_{vi} and β_{hi} combinations



MC- Colombo $\theta = 50$



MC- Colombo $\theta = 100$

Metapopulation Model

Fitted β_{vi} and β_{hi} and θ values

MOH	B _{vi}	B _{hi}	RMSE	Theta	Weekly average cases
MC - Colombo	0.14	0.56	8.935	20	75
Dehiwala	0.12	0.97	6.2741	60	28
Kollonnawa	0.52	0.9	10.8857	20	24
Kaduwela	0.68	0.94	4.9048	85	19
Moratuwa	0.73	0.68	3.6663	35	15
Panadura	0.27	0.81	4.7938	55	14

Above MOH areas has the highest averages of weekly reported dengue cases

Metapopulation Model

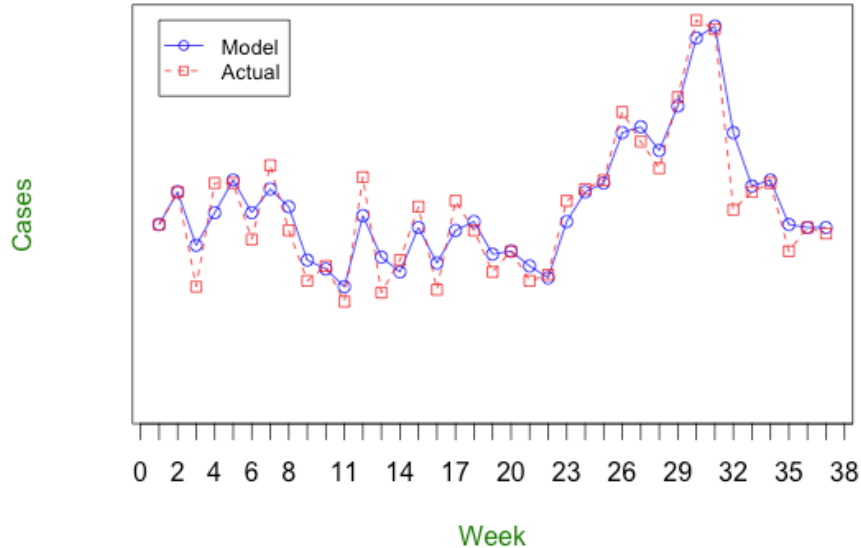
Fitted β_{vi} and β_{hi} and θ values

Calculated values for MOH areas which has low weekly average cases

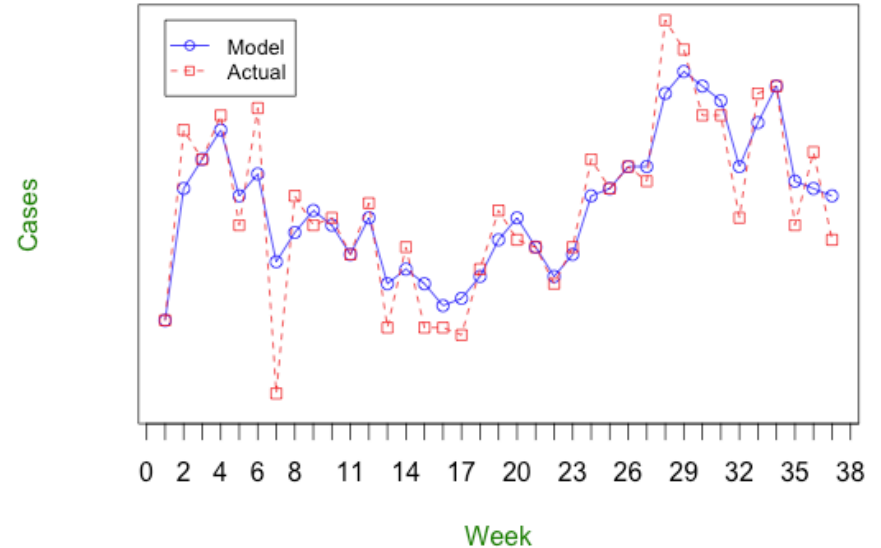
MOH	Bvi	Bhi	RMSE	Theta	Weekly average cases
Rajanganay	0	0	1.1094	10	3
Deraniyagala	0	0	1.0190	10	3
Beliatta	0	0	0.8320	10	2
Hasalaka	0	0	0.8660	10	2
Dambulla	0	0	0.8879	10	2

Metapopulation Model

Model fit for 2013 first 37 weeks



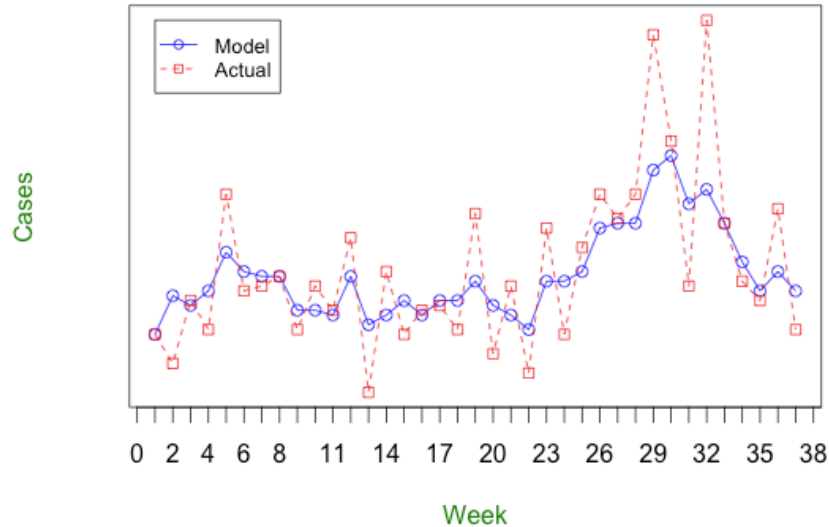
MC-Colombo MOH RMSE=11.3544



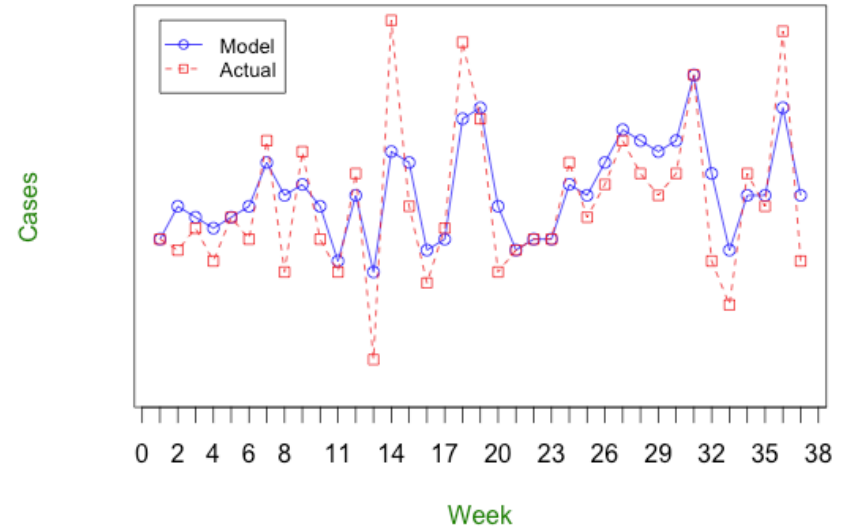
Dehiwala MOH RMSE = 6.2741

Metapopulation Model

Model fit for 2013 first 37 weeks



Kolonnawa MOH RMSE = 10.8857

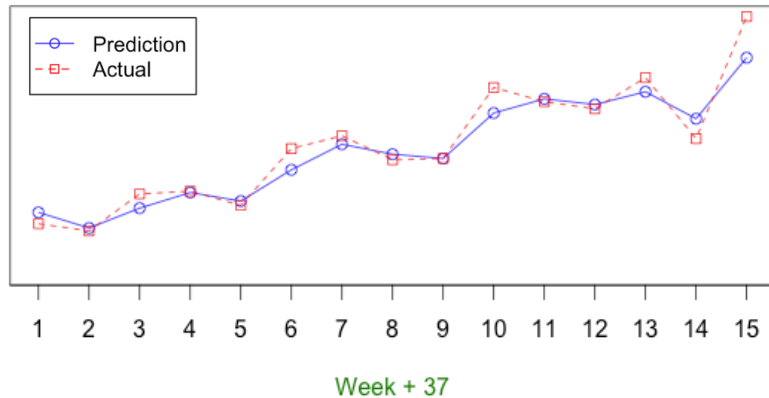


Kaduwela MOH RMSE=4.9048

Metapopulation Model

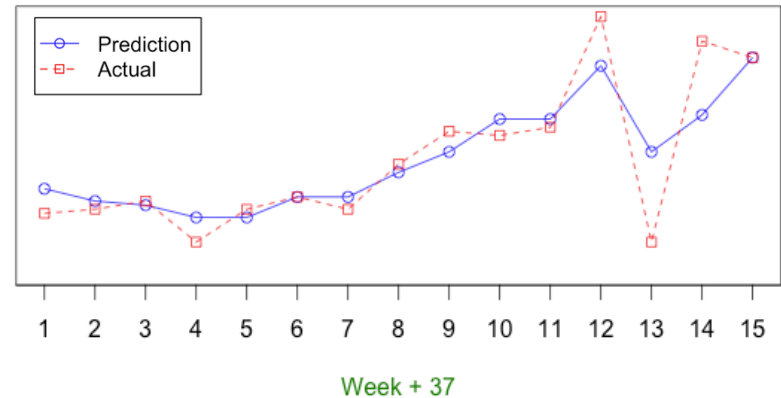
2013 predictions for weeks 38-52

Cases



MC-Colombo MOH, RMSE = 6.0906

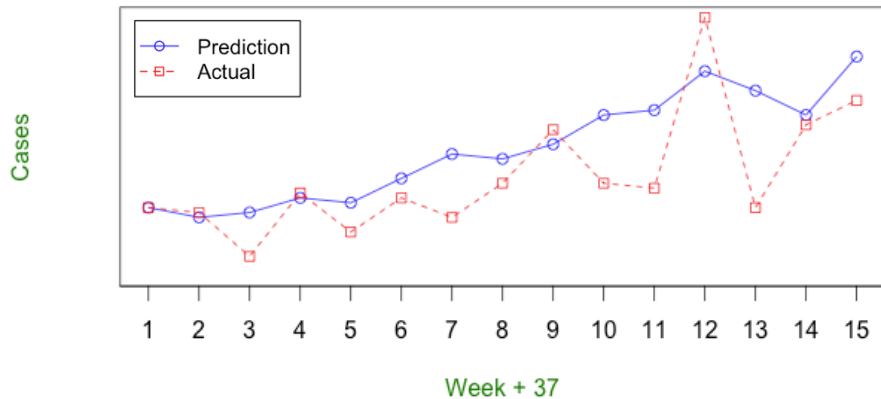
Cases



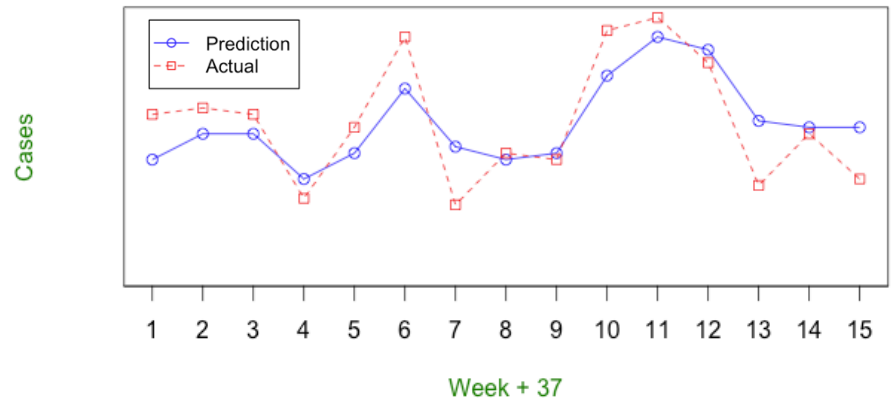
Dehiwala MOH, RMSE = 4.5804

Metapopulation Model

2013 predictions for weeks 38-52



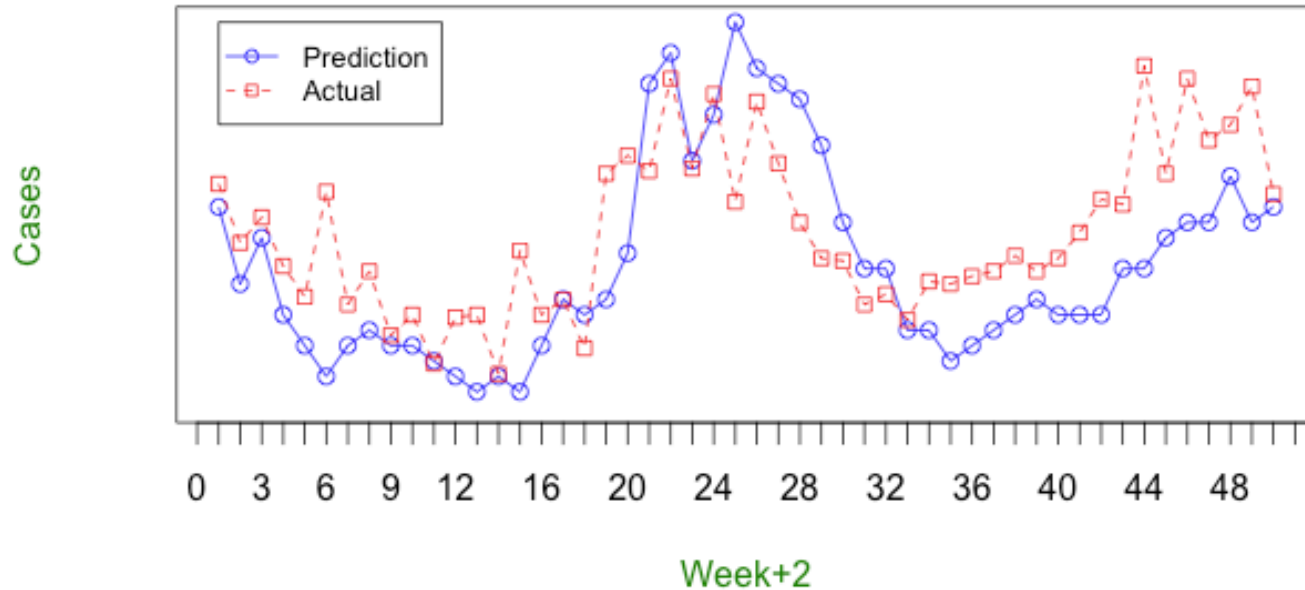
Kolonnawa MOH, RMSE = 5.4982



Kaduwela MOH, RMSE = 3.0159

Metapopulation Model

2014 predictions



MC-Colombo, RMSE = 30.8557

Metapopulation Model

Problems and challenges Faced

Lack of knowledge in epidemiology domain

No credible source for disease related data (rates/parameters etc.) in Sri Lanka

Not many researches to compare the results with

Metapopulation Model

Potential Enhancements/Future work

- Predict based on predictions

- Make θ a temperature-based parameter

- Fully automated learning and prediction system

- Visualization of the predictions

Conclusion

Metapopulation model outperforms ANN and HMM models

MOH area	Metapopulation Model	ANN	HMM
MC-Colombo	6.0906	20.6867	-
Dehiwala	4.5804	12.9326	19.5959
Kolonnawa	5.4982	13.3945	12.0
Kaduwela	3.0159	9.98	11.313

RMSE values of predictions

There is a significant effect on human travel on dengue propagation

Dengue disease dynamics can be explained using metapopulation model

Thank You!

Q & A

References:

- [1] - National Dengue Control Unit, Sri Lanka. <http://www.dengue.health.gov.lk/>. Accessed: 2015-09-07.
- [2] - What is Dengue and how is it treated? <http://www.who.int/features/qa/54/en/>. Accessed: 2015-09-10.
- [3] - Chen, Feng-Wen, and Chen-Wuing Liu. "Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan." *Paddy and Water Environment* 10.3 (2012): 209-222.
- [4] - Torre, Carlos Alan. *Deterministic and stochastic metapopulation models for dengue fever*. Arizona State University, 2009.

References

[5] - Sarzynska, Marta, Oyita Udiani, and Na Zhang. "A study of gravity-linked metapopulation models for the spatial spread of dengue fever." *arXiv preprint arXiv:1308.4589* (2013).