

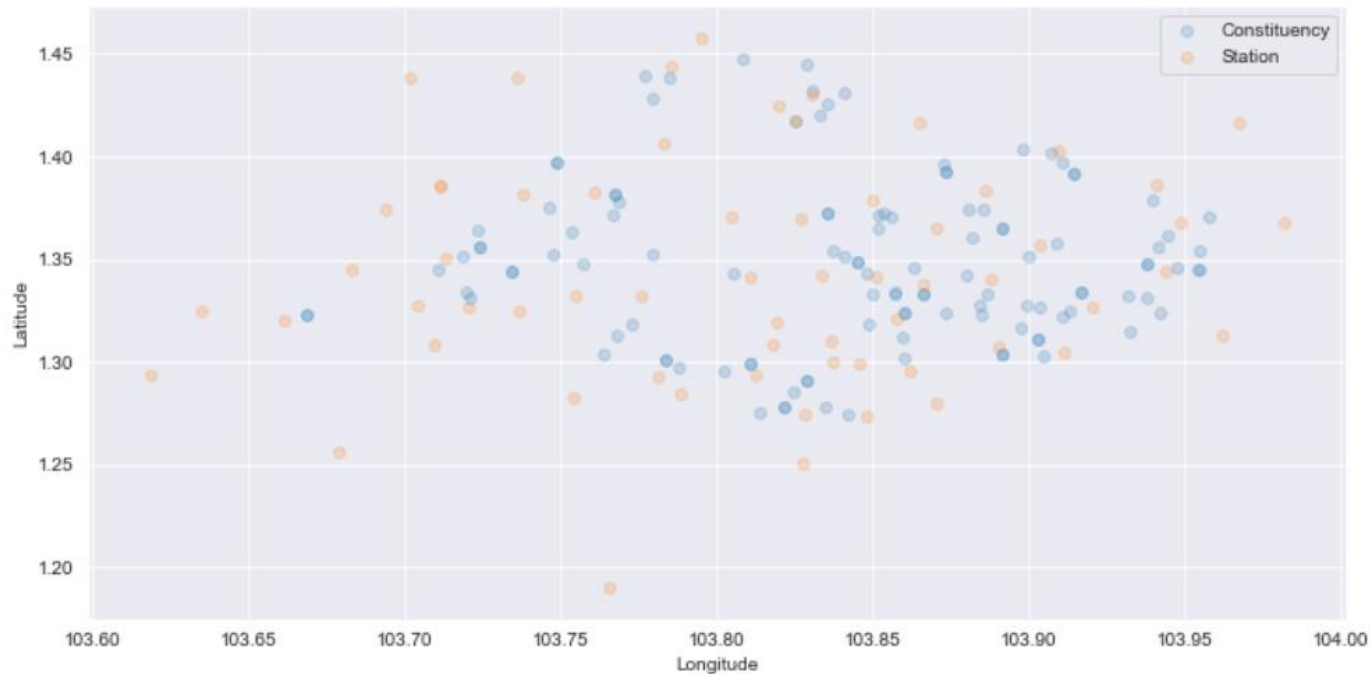
Weather Data

- Constituency rainfall and temperature extracted from historical NEA data:
 - Constituency LAT, LONG coordinates retrieved from OneMap API

```
request_str = "https://developers.onemap.sg/commonapi/search?searchVal=" + \
    cur_str + "&returnGeom=Y&getAddrDetails=N&pageNum=1"

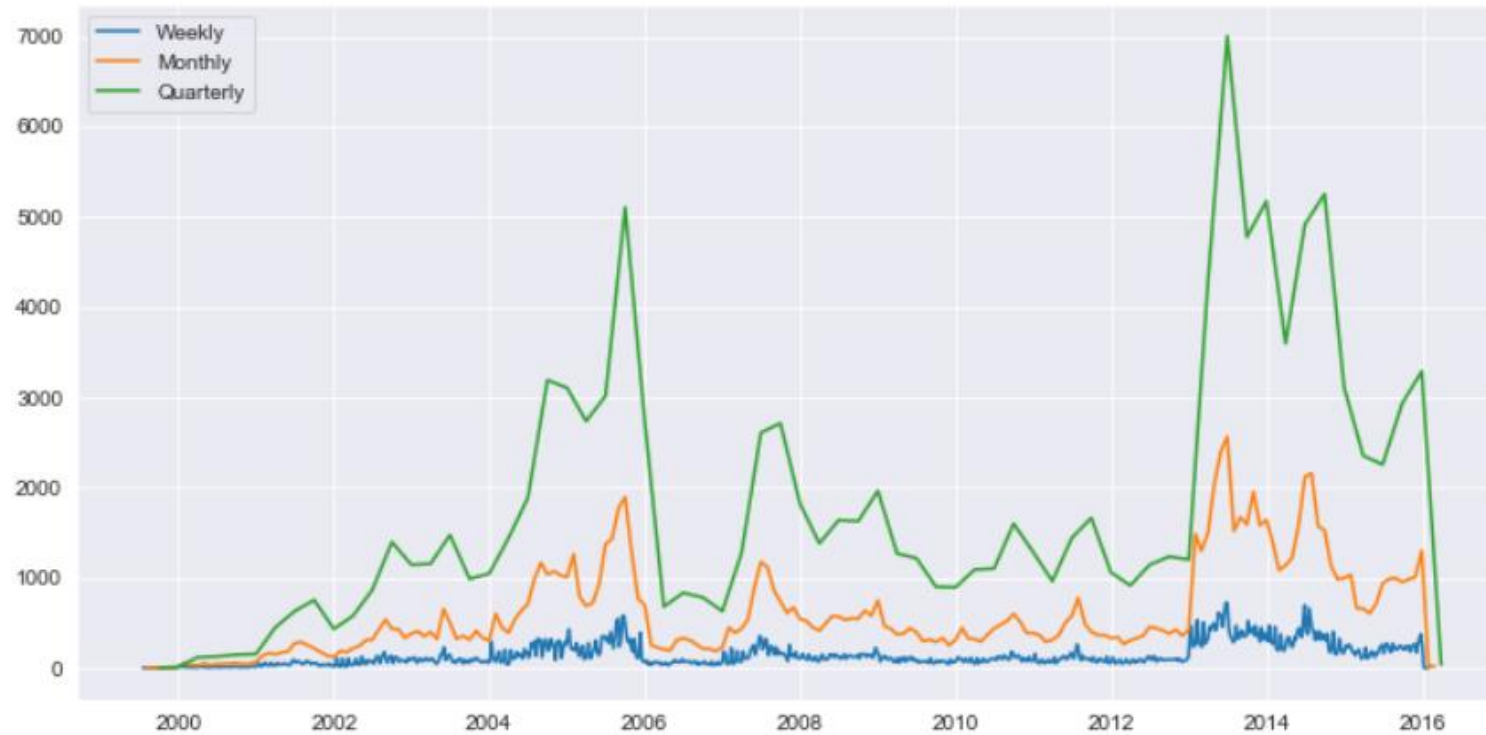
page = requests.get(request_str)
```

- Weather station LAT, LONG coordinates retrieved from HTML code of NEA website
- Each constituency mapped to the nearest weather station



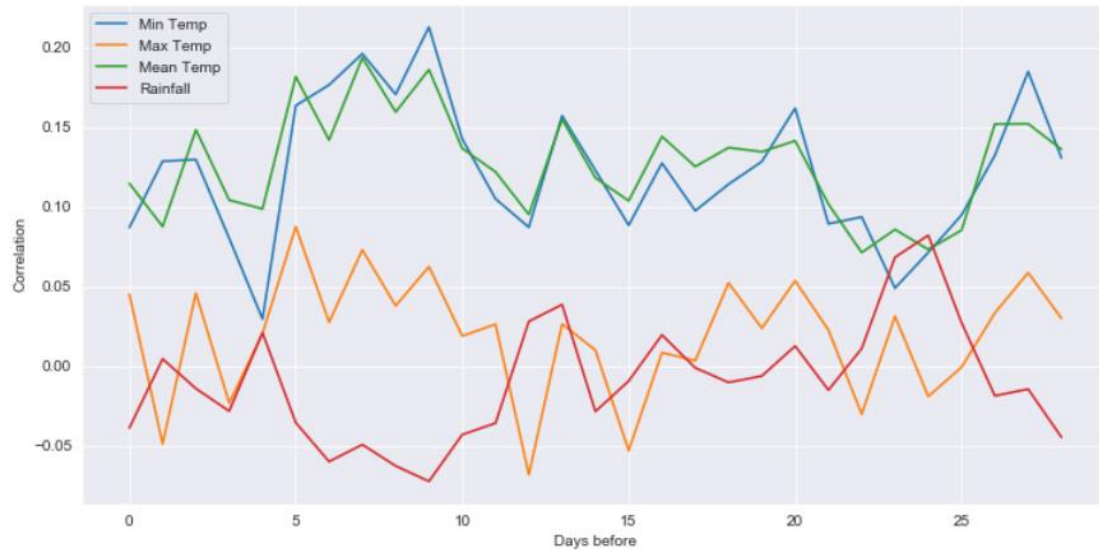
Dengue Count Data

- Dengue data contain information on the number of dengue cases in each constituency



Correlation

- Correlation study of dengue count:
 - Uses past daily temperature (max, min, mean) and rainfall
 - Studies correlation of monthly counts of each constituency with rainfall and temperature
- Dengue count more correlated with temperature than rainfall, agreeing with previously published work
- Correlation study will help feature selection



The influence of climate variables on dengue in Singapore

Article in International Journal of Environmental Health Research · May 2011

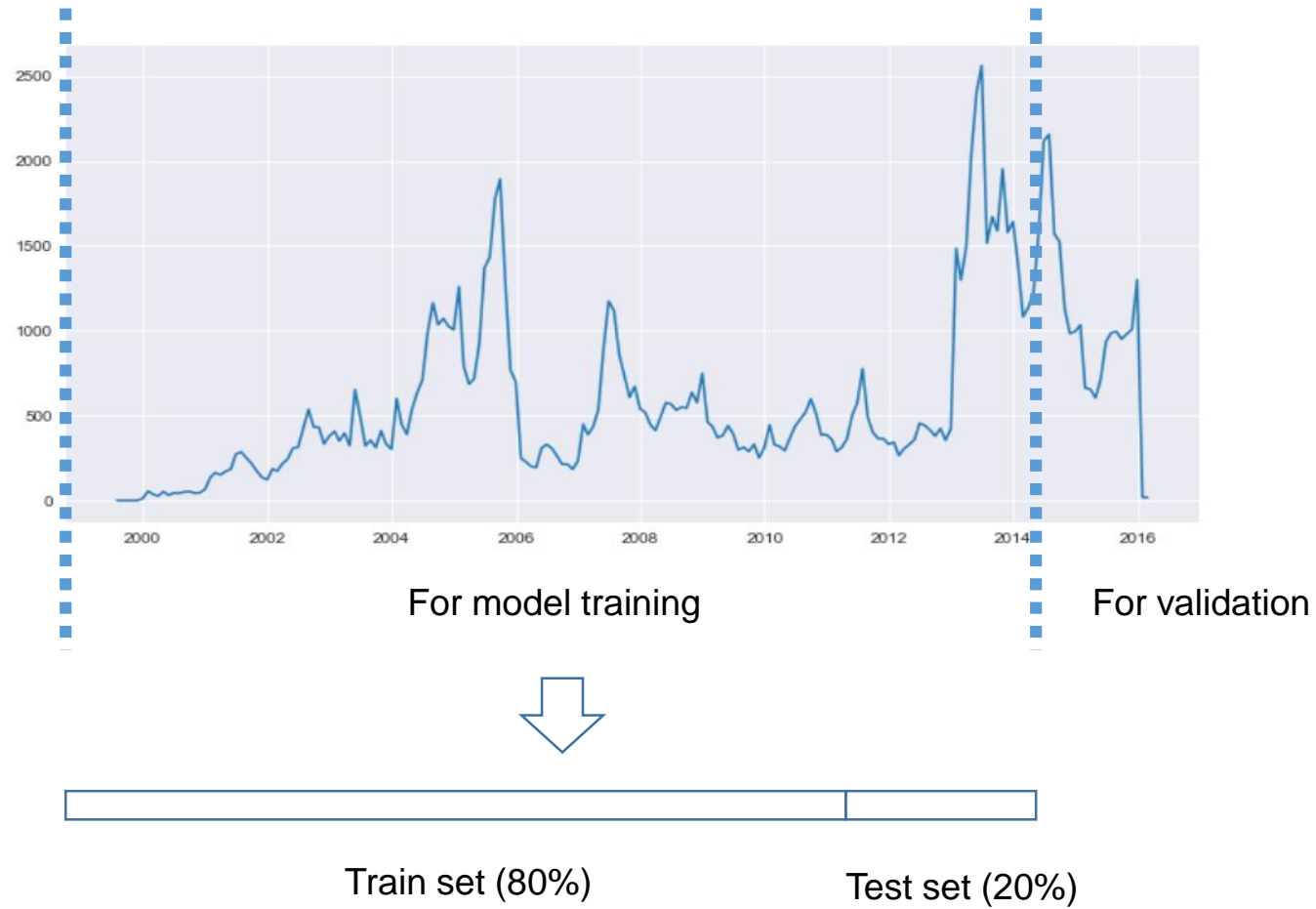
DOI: 10.1080/09603123.2011.572279 · Source: PubMed

22.2–184.6% in the number of dengue cases. For the minimum temperature, we observed that for the same variation, there was an average increase of 26.1–230.3% in the number of the dengue cases from April to August. The precipitation and the relative humidity, after analysis of correlation, were discarded in the use of Poisson Regression Model because they did not present good correlation with the dengue cases. Additionally, the relative risk of the occurrence of the cases of the disease

Modelling Approach

- Aim: to build a regression model to forecast the dengue count for the constituency 4 months in the future
- Data sample: Dengue count for a constituency for a month + accompanying features
- Features: average monthly min temperature, dengue counts for the constituency from 4 months+ back, long & lat coordinates, month of the year
- 2 models based on gradient boosting and linear regression (elastic net)
- A baseline forecaster that uses count from 4 months back as prediction

Model Training and Validation



Results

Gradient boosting regressor outperforms Elastic Net and the baseline

GB

Mean Absolute Error: 4.371479061802911
Mean Squared Error: 71.15480878948036
Root Mean Squared Error: 8.435330982805615

ENet

Mean Absolute Error: 5.3794420243894585
Mean Squared Error: 97.24077577937466
Root Mean Squared Error: 9.861073764016506

Baseline

Mean Absolute Error: 5.7688540646425075
Mean Squared Error: 122.66405484818804
Root Mean Squared Error: 11.075380573514757

