

## Preliminary Investigation of the Relationship between Weather Data and Hand-Foot-Mouth Disease Cases in Sarawak

Maryanne Nicholas

Department of Computational Science and Mathematics  
Faculty of Computer Science and Information  
Technology  
Universiti Malaysia Sarawak (UNIMAS)  
e-mail: maryanne\_880111@yahoo.com

Louis Anak Denis

Department of Computational Science and Mathematics  
Faculty of Computer Science and Information  
Technology  
Universiti Malaysia Sarawak (UNIMAS)  
e-mail: louisdenisikau@gmail.com

Jane Labadin

Department of Computational Science and Mathematics  
Faculty of Computer Science and Information Technology  
Universiti Malaysia Sarawak (UNIMAS)  
e-mail: ljane@fit.unimas.my

**Abstract**— This paper investigates the relationship between climate changes and the outbreak of Hand-Foot-Mouth Disease (HFMD) in Sarawak. This work is motivated by the statements made by some researchers that there may be a possibility that the outbreak of the HFMD could have a relationship with the weather. Two quantitative variables which consist of weather data and HFMD cases data are crucial for this research. The weather data comprises of the total amount of rainfall and the mean surface air temperature for the year 2000 until 2008. The evaluation of the variables' relationship is done by using estimated weather data using spline interpolation during the outbreak of HFMD in Sarawak from year 2000 until 2008. The results obtained indicate that there exists correlation between the climate changes and the outbreak of HFMD cases in Sarawak.

**Keywords**— *Hand-Foot-Mouth Disease (HFMD), spline interpolation, correlation, correlation coefficient*

### I. INTRODUCTION

Hand-Foot-Mouth Disease (HFMD) is most commonly caused by viruses from the *enteroviruses* group of the *Picornaviridae* family especially *Coxsackie A16* virus and *Enterovirus 71 (EV71)*. According to HFMD guidelines provided by the Health Ministry of Malaysia in 2007, in the temperate countries, most HFMD cases occur in the summer and fall and the incubation period is 4-6 days [1]. It is acknowledged that there is a trend for the outbreak of HFMD every year in Sarawak starting from the year 1997 [11].

[10] have already showed that the effects of global warming especially in rainfall season and air temperature have influenced the outbreak of HFMD in Tokyo. Due to this, the changing season in Malaysia especially in Sarawak, those are rainfall season and mean surface air temperature

need to be concerned and observed whether the climates can influence the outbreak of HFMD in Sarawak.

In another work, [7] have found that there exist a weak relationship between the cases of HFMD and the weather. However, the data that they used are weather data taken from five meteorological stations against HFMD cases obtained from the Sarawak General Hospital which came from various clinics in Sarawak. Therefore, one may argue since Sarawak is big then is the weather data from the five stations enough to represent the weather data for the whole of Sarawak. In this work, interpolation technique is used to answer such query.

### II. INTERPOLATION TECHNIQUE

Interpolation refers to the process of defining a function that takes on specified values at specified points. It is also used for inferring data values taken between the points or extrapolates values beyond the range of the available data. Interpolation is a mathematical estimation technique that is applied to a given value of a collected set of data to expand to the sites where no samples available or to predict the values for a whole surface based on the collected sample point values.

Thus, interpolation is the technique to be used to estimate the weather in Sarawak given the two weather data obtained from the Meteorological Department of Kuching which consists of the rainfall season (total amount of rainfall) and air temperature taken from five different stations in Sarawak.

### III. SPLINE INTERPOLATION AND CORRELATION ON BASE YEAR

Records of monthly total rainfall amount and records of monthly mean surface air temperature from year 2000 until 2008 were collected from the Malaysia Meteorological

Department of Sarawak. These two records were measured at five different stations in Sarawak which are located at the Kuching Airport, Sri Aman Airport, Sibu Airport, Bintulu Airport, and Miri Airport. Since Sarawak is a big state then the data obtained from the five stations may not be able to represent the weather for the entire Sarawak's region, unless they are interpolated.

The data obtained are monthly average only and thus we are not provided with the derivative values. Therefore, spline interpolation technique is chosen for this project as the slopes can be estimated using the data given and MatLab provides built-in function for this purpose.

#### A. Estimation of Rainfall Data

The source data obtained from the Meteorological Department of Sarawak are 108 sets of monthly data recorded for nine year for every five different stations. Therefore, the actual data are 540 data. The data (rainfall data) is taken from January until December of year 2000 until 2008.

In a particular year, there exist twelve months and for every month there consists of five different values (taken from the five stations) to represent the total rainfall in Sarawak for that particular month. Therefore, a vector of five values corresponding to a particular month is created. For example, the values taken from each station in the month of January for the year 2000 is represented as a vector:

$$R_{jan} = [902.9, 331.0, 640.3, 382.9, 193.0]$$

where

- $R_{jan}(1)$  – total of rainfall in Kuching in January 2000
- $R_{jan}(2)$  – total of rainfall in Sri Aman in January 2000
- $R_{jan}(3)$  – total of rainfall in Sibu in January 2000
- $R_{jan}(4)$  – total of rainfall in Bintulu in January 2000
- $R_{jan}(5)$  – total of rainfall in Miri in January 2000

Similarly, the vector for the rest of the months in a year which are February, March, April, May, June, July, August, September, October, November and December can be created. Every particular month has a different total amount of rainfall. Spline interpolation technique is applied in order to estimate the total amount of rainfall for every month for any location in Sarawak for year 2000. Here, each weather station where the weather data is collected is given a location number (1 to 5, respectively). Therefore, any random real number ranging from 1 to 5 is an estimated location number representing any location in Sarawak. For example, to estimate the total amount of rainfall for January in the year 2000 for random location of Sarawak is given by  $N$ . Using the built-in function in MatLab, this number  $N$  can be any real number ranging from 1 to 5. Since there are 12 months for a particular year, thus  $N$  is a vector of twelve elements.

In this study, since the aim is to investigate the relationship of weather data and the HFMD cases then our technique in estimating the values of  $N$  is by observing the best fitted line on a scatterplot for the base year (in this case

the year 2000). Scatterplots are the best way to start observing the relationship between two quantitative variables. It gives a good visual picture of the relationship between the two variables, and aids the interpretation of the correlation coefficient or regression model. If there is a strong association between two variables, knowing one helps a lot in predicting the other. But when there is a weak association, information about one variable does not help much in guessing the other [10].

At the same time the correlation coefficient is calculated to ensure that there exists strong relationship between the HFMD cases and the rainfall. The correlation coefficient  $r$  is a measure of linear relationship between two attributes or columns of data. The value of  $r$  can range from -1 to +1 and is independent of the units of measurement. A value of  $r$  near 0 indicates little correlation between attributes and a value near +1 or -1 indicates a high level of correlation.

It was found that the set of values  $N$  that satisfy the mentioned criteria above is

$$N = [3.4, 1.2, 4.9, 3.4, 3.8, 2.3, 4.0, 4.1, 4.1, 1.1, 1.5, 1.1]$$

Table 1 shows the estimated rainfall amount,  $M$  for each month using the values in  $N$ .

TABLE I ESTIMATION OF RAINFALL AMOUNT VALUE IN YEAR 2000

Month	Random Value, $N$	Estimation of Rainfall Amount Value, Mililitre (M)
January	3.4	603.1424
February	1.2	484.2052
March	4.9	191.3718
April	3.4	322.1368
May	3.8	97.7424
June	2.3	164.0557
July	4.0	186.6000
August	4.1	200.8607
September	4.1	530.6174
October	1.1	309.4070
November	1.5	437.0969
December	1.1	501.1742

Using the definition of interpolation, these values of  $N$  represents any location in Sarawak. The same twelve values  $N$  obtained earlier will be used to estimate whether there exists correlation between rainfall season and HFMD cases in Sarawak from year 2001 until 2008.

#### B. Estimation of Mean Surface Air Temperature Data

In similar fashion above, a vector of 12 elements is acquired to represent any location in Sarawak so that the mean surface air temperature has strong correlation with the HFMD cases for the year 2000. This vector,  $A$ , is found to be

$$A = [4.6, 4.7, 4.3, 4.9, 1.9, 1.0, 1.1, 1.5, 1.3, 1.9, 1.5, 1.6]$$

Table II represents the twelve random values for any location in Sarawak for the mean surface air temperature,  $B$ ,