Assessing dengue vector abundance in Penang Island by cluster analysis

N A Hashim1,2,6, A H Ahmad3, A Talib4 and Suwarno5
1School of Food Science and Technology, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia
2Laboratory for Pest, Disease and Microbial Biotechnology (LAPDiM), Central Laboratory, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia
3School of Biological Sciences, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia.
4School of Distance Education, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia.
5Biology Department, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Darussalam Banda Aceh, Indonesia.
E-mail: aida.hashim@umt.edu.my

Abstract. Dengue is one of the deadliest mosquito-borne diseases in the world. Aedes albopictus and Aedes aegypti are the two responsible vectors for the disease. In this study, self-organizing map (SOM) was applied for ordination, clustering and mapping of the Ae. aegypti and Ae. albopictus abundance with their breeding container sizes. It was found that the abundance of vector related with the size of breeding container. Regardless of urbanization level, Ae. albopictus was more abundant in medium size containers, while Ae. aegypti was found more abundant in large containers. This finding suggested that for control efforts, eliminating medium and large breeding containers will significantly reduce Aedes population in Penang Island.

1. Introduction
Dengue fever (DF) and dengue haemorrhagic fever (DHF) are caused by the dengue virus which belonging to genus Flavivirus, that consists of four serotypes (DEN 1, DEN 2, DEN 3 and DEN 4) [1]. The geographical spread of dengue is increasing. In 1950s only five countries documented with the disease. But now, there are more than 100 countries around the world reporting the incidence of DF and DHF [2]. Dengue viruses are transmitted from viremic to susceptible human beings by various mosquitoes of subgenus Stegomyia, notably Ae. aegypti and Ae. albopictus. Both species have been known to bite hosts during daytime and breed in and around human habitation. Aedes aegypti remains the principal vector for dengue haemorrhagic fever with Ae. albopictus being regarded as a secondary vector [3]. Aedes albopictus has been incriminated as the vector responsible for dengue and dengue haemorrhagic fever epidemics in several locations including Hawaii, Japan, Indonesia, Southern China, Thailand, Singapore, and Malaysia [4].

From the past 20 years the number of cases reported in Malaysia showed an upward trend [5, 6]. The reasons for this increase is due to the period of rapid urbanization and population growth (both

6 To whom any correspondence should be addressed.
local and foreign migration to cities), a different lifestyle (using and throwing non-biodegradable containers), rapid transportation and poor living conditions (poor water supply and poor scavenging services at squatter areas). All these gave rise to increased breeding areas for Aedes mosquitoes and easy and rapid spread of the virus. The behaviour and ecology of vectors and hosts are the dominant factors in diseases transmission and many can be attributed to explosive population growth and poverty [7]. The rapid human population growth and increased urbanisation have led to substandard housing, inadequate water supply, and waste management systems and consequently an abundance of mosquito breeding sites. Water storage and inadequate water disposal can provide habitat for mosquitoes, particularly in rapidly expanding urban areas. The habit of people of adding water to containers is one of the most important factors determining dengue endemicity [8].

The control of dengue fever is depending on managing populations of its vectors [9]. No effective vaccine or drug treatment for dengue fever is yet available; therefore management of disease has relied on vector control measures, such as reduction of breeding sites and use of insecticides. Such measures have succeeded in eradicating mosquitoes in some regions, but have proved difficult to maintain in the long term [10]. The knowledge of breeding habitat for Aedes mosquito is very important for Aedes larvae control. The two vector species are usually found in and around human habitations. They breed in artificial and natural containers and receptacles which hold clean and clear water. Aedes aegypti is one of the most mosquito vectors for arboviruses, because it is highly anthropophilic and thrives close to humans [11] preferring to live indoors [11,12]. It is common in urban areas especially in the most densely populated districts [13]. This species breeds in domestic and peridomestic water containers [14] that contain clean water [12]. On the other hand, Ref. [15] stated that Ae. albopictus is a highly adaptable species because it seems to be able to recolonize tree holes in forests after transported to a new region, thus making it hard to control.

The self-organizing map (SOM) [16] is a type of neural networks that uses unsupervised learning. It does not require target values to train the neural network as in supervised learning. While the neural networks based on supervised learning typically find out functional input-output relationship, the self-organizing map finds the patterns inherent in the input data. The SOM consists of a set of neurons spatially structured in a regular lattice, usually in a two-dimensional grid. Therefore, the SOM can be used as a data compression technique by mapping nonlinear and high-dimensional data (which is typical in modern automated processes where many variables are routinely collected) into a lower-dimensional grid. Additionally, it allows identifying groups of observations with similar characteristics (clusters) [16].

Although many papers on self-organizing maps have been published, very few studies have dealt with the use of SOM for mosquito population study particularly Aedes. The SOM was customized for prioritization of endemic zones of filariasis and malaria which can enable in efficient targeting of the risk areas for control operations in Changlang District, Arunachal Pradesh [17]. Previously SOM was used for visual classification of malaria-endemic zones into high, medium and low in the different districts of Manipur, India [18]. This approach was also adopted for classification and clustering of Anopheline sequences [19]. The present study applied SOM on the abundance of Aedes mosquitoes in their breeding habitat.

2. Materials and methods

2.1. Study sites

The Southwest District in Penang Island, Malaysia was selected as the study site because it is a common location affected by dengue. In this study, four localities with different geographical settings were selected. They are Pantai Jerjak (urban residential area), Bayan Lepas (urban residential/industrial area), Batu Maung (suburban residential area) and Balik Pulau (rural area).

2.2. Sampling methods

Monthly entomologic surveys were conducted for each study area between 0900 h and 1500 h by three two-person collection teams. Each survey consisted of the inspection of indoor and outdoor areas of each house visited for water-holding containers. This study was conducted for one year. All types of water-filled (wet) containers that could act as potential breeding sites and were positive for
mosquito larvae/pupae were recorded. The containers were categorized into three sizes: small (capacity < 1 L), medium (1 L < capacity < 15 L) and large (capacity > 15 L). The *Aedes* immatures (pupae and larvae) found during surveys were collected using pipette or sieves and placed into labeled zip-lock plastic bags. The samples were transported to the laboratory at the School of Biological Sciences, Universiti Sains Malaysia on the same day they were collected for further processing. In the laboratory, water samples were examined and the number of individuals for each immature stage (1st, 2nd, 3rd, 4th, and pupa) present was recorded. The 3rd and 4th instars larvae were identified under a dissecting microscope, Olympus CX41 (Olympus, Tokyo, Japan) to the species level using identification keys [20]. The pupae were reared until the emergence of adults after which the adults were identified to species, sexed and subsequently killed.

Data analysis was performed using the Statistical Package for Social Science (SPSS for Windows, version 22.0). Only the data for positive water-holding containers were used for statistical analysis. Descriptive statistics, which include the percentages, means, standard deviations and standard error of means, were used to analyse the characteristics of the positive water-holding containers (size, container type and the immature composition). Other descriptive statistics and inferential statistics like one-way ANOVA test, Kruskal-Wallis test and Chi-square test were also used for the test of statistical significance. The non-supervised Artificial Neural Networks (ANN) was also used to depict breeding patterns. This method is designed to identify patterns based on similarities between inputs. The most popular non-supervised ANN, self-organizing map (SOM) was applied for ordination, clustering and mapping the positive container sizes and *Ae. aegypti* and *Ae. albopictus* prevalence in the study areas. The data were mapped by U-matrix and K-means partitioning using the SOM Toolbox in MATLAB 7.0.

### 3. Results

In each study site, small containers were the most abundant, while large containers were least abundant. Figure 1 shows the percentage of positive containers by their sizes in the four study areas. The highest percentage (55.1%) of small containers was observed in Balik Pulau (BP) whereas the lowest (42.9%) was in Batu Maung (BM). Medium containers were most abundant in Batu Maung (41.8%) and least abundant in Pantai Jerjak (PJ) (28.5%). Bayan Lepas (BL) had the highest percentage of large positive containers (18.1%) and the lowest percentage (11.1%) was found in Balik Pulau. The chi square test revealed that there was a significant difference in the number of container in relation to size in each area (p < 0.05).

![Percentage of positive containers by size in four study areas](image_url)

**Figure 1.** Percentage of positive containers by size in four study areas (PJ = Pantai Jerjak; BL = Bayan Lepas; BM = Batu Maung; BP = Balik Pulau).
The prevalence of the immature stages of *Ae. albopictus* in the four areas of Southwest district of Penang Island is summarised in Table 1. Regardless of the number of breeding sites found and urbanization level of the study site, statistical analysis results showed that no significant differences in the mean number of immature of each stage among four study areas (p > 0.05). In contrast to *Ae. aegypti*, it showed significant differences in the mean number of immature of each stage in different study areas (Table 2). The number of *Ae. aegypti* immature were higher in more urbanized areas such as Pantai Jerjak and Batu Maung, but less abundance in the areas with higher vegetation as in Bayan Lepas and Balik Pulau. Kruskal-Wallis tests revealed that there were significant differences between the mean numbers of individuals of each immature stage from the four study areas (p < 0.05).

### Table 1. The number of immature stages (larvae and pupae) *Aedes albopictus* found in positive containers in four study sites.

| Study Areas | Positive Containers (N) | Total immature | Mean immature | SE | Kruskal-Wallis |
|-------------|-------------------------|----------------|---------------|----|---------------|
| **I & II instar** |                         |                |               |    |               |
| Pantai Jerjak | 288                     | 7139           | 24.79         | 4.23 | $\chi^2 = 0.20$ |
| Bayan Lepas  | 375                     | 7726           | 20.60         | 4.37 | df = 3        |
| Batu Maung   | 373                     | 7032           | 18.85         | 2.91 |               |
| Balik Pulau  | 470                     | 12474          | 26.54         | 6.23 |               |
| **III & IV instar** |                     |                |               |    |               |
| Pantai Jerjak | 288                     | 5951           | 20.66         | 3.23 | $\chi^2 = 0.62$ |
| Bayan Lepas  | 375                     | 8288           | 22.10         | 5.15 | df = 3        |
| Batu Maung   | 373                     | 6693           | 17.94         | 1.70 |               |
| Balik Pulau  | 470                     | 8719           | 18.55         | 2.03 |               |
| **Pupae**    |                         |                |               |    |               |
| Pantai Jerjak | 288                     | 1516           | 5.26          | 0.62 | $\chi^2 = 1.55$ |
| Bayan Lepas  | 375                     | 2621           | 6.99          | 1.21 | df = 3        |
| Batu Maung   | 373                     | 2437           | 6.53          | 0.65 |               |
| Balik Pulau  | 470                     | 2706           | 5.76          | 0.63 |               |
| **Total Immature** |                     |                |               |    |               |
| Pantai Jerjak | 288                     | 14606          | 50.72         | 7.52 | $\chi^2 = 0.53$ |
| Bayan Lepas  | 375                     | 18635          | 49.69         | 10.47 | df = 3        |
| Batu Maung   | 373                     | 16162          | 43.33         | 4.54 |               |
| Balik Pulau  | 470                     | 23899          | 50.85         | 8.38 |               |

** significance at p < 0.05

Figures 2a, 2b and 2c show the visualization of the self-organizing map (SOM) on prevalence of *Ae. albopictus* and *Ae. aegypti* immature stages from 1506 positive containers in Southwest District areas based on size. Figure 2a is a U-matrix visualization (Unified Distance Matrix) of the self-organizing map of the *Aedes* breeding container distribution by sizes. Hexagons having letters (S, M or L) indicate locations of map units and are coloured according to the medians of the surrounding hexagon. Hexagons without letters show the distances between two neighbouring map units. The axis of a U-matrix has no intrinsic biological meaning; it shows the distance between data vectors and SOM weight vectors. The colour scale for distance is shown as colour bar on the right.

The self-organizing map of the container distribution is shown in three areas (the lower left corner and the lower right corner) that have different hues from the rest of the map. By drawing borderlines separating the yellow/red areas and surrounding blue areas, we obtained three major categories of container sizes.
Table 2. The number of *Aedes aegypti* immature (larvae and pupae) from positive containers in four study sites.

| Study Areas | Positive containers (N) | Total immature | Mean immature | SE | Kruskal-Wallis |
|-------------|-------------------------|----------------|---------------|----|----------------|
| **I & II instar** | | | | | |
| Pantai Jerjak | 288 | 967 | 3.36 | 0.79 | $\chi^2 = 19.70$ |
| Bayan Lepas | 375 | 536 | 1.43 | 0.42 | df = 3 |
| Batu Maung | 373 | 905 | 2.43 | 0.86 | $p = 0.00^{**}$ |
| Balik Pulau | 470 | 146 | 0.31 | 0.21 | |
| **III & IV instar** | | | | | |
| Pantai Jerjak | 288 | 794 | 2.76 | 0.67 | $\chi^2 = 18.64$ |
| Bayan Lepas | 375 | 673 | 1.79 | 0.45 | df = 3 |
| Batu Maung | 373 | 1047 | 2.81 | 0.95 | $p = 0.00^{**}$ |
| Balik Pulau | 470 | 144 | 0.31 | 0.18 | |
| **Pupae** | | | | | |
| Pantai Jerjak | 288 | 186 | 0.65 | 0.13 | $\chi^2 = 14.82$ |
| Bayan Lepas | 375 | 219 | 0.58 | 0.15 | df = 3 |
| Batu Maung | 373 | 464 | 1.24 | 0.49 | $p = 0.00^{**}$ |
| Balik Pulau | 470 | 68 | 0.14 | 0.09 | |
| **Total Immature** | | | | | |
| Pantai Jerjak | 288 | 1947 | 6.76 | 1.42 | $\chi^2 = 20.35$ |
| Bayan Lepas | 375 | 1428 | 3.81 | 0.94 | df = 3 |
| Batu Maung | 373 | 2416 | 6.48 | 2.24 | $p = 0.00^{**}$ |
| Balik Pulau | 470 | 358 | 0.76 | 0.46 | |

** significance at $p < 0.05$

The top cluster (marked ‘‘S’’ for small size container) was identified as a class of containers with capacity less than 1 litre and this cluster contains 51% small containers out of 1506 breeding containers in Southwest District. The lower right corner (marked ‘‘M’’ for medium size container) was identified as a class of container with capacity more than 1 L but less than 15 L. It contains 34% medium containers out of 1506 *Aedes* breeding containers. The lower-left corner (marked ‘‘L’’ for large size container) was identified as a class of container with capacity more than 15 L. Within this class, 15% containers out of 1506 breeding containers were found in Southwest District. Thus, Figure 2b was constructed with three prominent clusters. Borderlines between the classes are highlighted. The top right corner of the cluster map correspond with small containers (blue), bottom left corner corresponds with large containers (green) and bottom right corner corresponds with medium container (red).

Figure 2c shows the component planes of U matrix (1), overall number of *Aedes* immature stages (2-5), *Ae. albopictus* immature stages (6-8), *Ae. albopictus* immature stages (9-11) and map clustered by container sizes (12). The 12 component planes (subplots) are linked by position: in each subplot, the hexagons in a certain position correspond to the same map unit. Each hexagon represents one map unit, and its colour (scaled on the colour bar) indicates the value of the component in that unit (the number of *Aedes* immature of the corresponding synonymous container size). Analogous hexagons on different images correspond to the same map unit. Red colour indicated high immature abundance, yellow indicates moderate abundance and blue indicate low immature abundance. A comparison of images of component planes and the clustered map shows that *Aedes* immatures abundance of 1st & 2nd instar larvae (2), 3rd & 4th instar larvae (3), pupae (4) and total immatures (5) mainly occur in the class of medium container.
Figure 2. Distribution of immature stages in Southwest District, Penang on the self-organizing map (SOM) according to container capacity and clustering of the trained SOM, (a) U-matrix (b) clustered map. Container size/capacity: S = small, M = medium, L = large (c) Component planes of data visualization of *Aedes* immature stages abundance in Southwest District, Penang concerning container size/capacity. Colour bars on the right of component plans show the scale for immature abundance (component plans 2-11). Red = high abundance, yellow = moderate abundance and blue = low abundance. In component plan 12, container capacity: S= small, M=medium, L= large. Colour bars for component plan 1 and 12 indicate distance.
**Aedes albopictus** immatures abundance of 1<sup>st</sup> & 2<sup>nd</sup> instar larvae (6), 3<sup>rd</sup> & 4<sup>th</sup> instar larvae (7), pupae (8) also mainly occur in the class of medium container indicated that *Ae. albopictus* immatures contribute strongly to the abundance of *Aedes* mosquito abundance in Southwest district. In contrast, *Ae. aegypti* immature abundance of 1<sup>st</sup> & 2<sup>nd</sup> instar larvae (9), 3<sup>rd</sup> & 4<sup>th</sup> instar larvae (10) and pupae (11) occurred mainly in the class of presumed large containers. Thus the result from unsupervised ANN clustering using SOM for Southwest district indicated that the highest *Ae. albopictus* immature abundance was in medium containers. Even though small containers were more abundant in the study areas, the result revealed that medium containers could support a higher number of *Ae. albopictus* immatures than small containers. The highest *Ae. aegypti* immatures abundance were detected in large container which found with high abundance indoor.

4. Discussion

Breeding site characteristics such as size is one of the factors that contributed to the success of mosquito oviposition and immature development. The size of container ensures the habitat stability which is crucial for mosquito aquatic stage. Medium and Large container can sustain higher amount of water and can support high number of mosquito immatures for a long period of time. Larger container also provides wider water edge/line for mosquito oviposition compared to small containers. A study of *Ae. albopictus* and *Ae. subalbatus* oviposition showed that more than 75% of the eggs were laid within 16 mm of the waterline [21].

In the present study, analysis of immature stage densities in three sizes of containers revealed that all immature stages of *Ae. albopictus* were found with the highest densities in medium containers, whereas *Ae. aegypti* immature stages were found with the highest densities in the large containers for each study area. The larger containers can support higher immature densities. Although analysis of containers confirmed that most of the small containers were utilized by *Ae. albopictus*, the containers can only support low immature densities. Medium containers can hold a higher volume of water and eventually support a higher density of *Ae. albopictus* immature. Large water tanks were 99% more productive, in terms of numbers of mainly *Ae. aegypti* and *Ae. scutellaris* (Walker) immature, then small containers which were three times more abundant [22]. A study in Charters Towers and Mingela/Ravenswood found that rainwater tanks (categorized as large container) were the key container for *Ae. aegypti* comprising 13.5% to 29.6% of positive containers but supporting 60.4% to 63.0% of immatures collected [23]. Therefore, immature density was correlated with container capacity and water volume.

In Penang, residents prefer to keep large container to store water. Thus, during surveys, the large containers is easy to spot on and check for *Aedes* infestation. While medium and small containers usually can be found in perimeter of the house whether kept by residents or discarded. The size of breeding container highly related to the abundance of vectors. *Aedes albopictus* was more abundant in medium-size containers while *Ae. aegypti* was found more abundant in large containers regardless of urbanization level. Thus for control efforts, eliminating medium and large breeding containers will significantly reduce *Aedes* population in Penang Island and subsequently will help to decrease dengue infection.

**References**

[1] Gubler D J 1997a Dengue and dengue haemorrhagic fever: Its history and resurgence as a global public health problem In Gubler D J and Kuno G (eds) *Dengue and Dengue Haemorrhagic Fever* (Cambridge, UK: University Press) 1-22
[2] Guha-Sapir D and Schimmer B 2005 *Emerging Themes in Epidemiology* 2 1-10
[3] Halstead S B 1994 *Singapore Medical Association Newsletter* 24 5-6
[4] Rai K S 1991 *Annual Review of Entomology* 36 459-484
[5] KKM (Kementerian Kesihatan Malaysia) 2002 *Laporan Tahunan 2002 Cawangan Penyakit Bawaan Vektor* (Kuala Lumpur Malaysia)
[6] MOH (Ministry of Health) 2009 *Annual Report 2009* (Kuala Lumpur Malaysia Vector-borne Disease Branch Ministry of Health Malaysia)
[7] Reiter P 2007 *Vector-Borne and Zoonotic Diseases* 7 261–73
Acknowledgments
The authors highly indebted to all members of mosquito survey team for their help. This work was supported by Postgraduate Research Grant Scheme (PRGS) Universiti Sains Malaysia, Grant no 1001/PBIOLOGY/841011.