Spatial cluster analysis of COVID-19 in Malaysia (Mar-Sep, 2020)

Sami Ullah,1 Nurul Hidayah Mohd Nor,1 Hanita Daud,1 Nooraini Zainuddin,1 Muhammad Salim Javed Gandapur,2 Iftikhar Ali,3 Alamgir Khalil4

1Department of Fundamental & Applied Sciences, Universiti Teknologi Petronas, 32610 Seri Iskandar, Perak, Malaysia; 2IIC University of Technologi, Kingdom of Cambodia; 3Project Wing (PSDP), Planning Division & Special Initiatives, Govt of the Pakistan, Islamabad, Pakistan; 4Department of Statistics, University of Peshawar, Peshawar, Pakistan

Abstract

Coronavirus disease 2019 (COVID-19) is the current worldwide pandemic as declared by the World Health Organization (WHO) in March 2020. Being part of the ongoing global pandemic, Malaysia has recorded a total of 8639 COVID-19 cases and 121 deaths as of 30th June 2020. This study aims to detect spatial clusters of COVID-19 in Malaysia using the Spatial Scan Statistic (SaTScan™) to guide control authorities on prioritizing locations for targeted interventions. The spatial analyses were conducted on a monthly basis at the state-level from March to September 2020. The results show that the most likely cluster of COVID-19 occurred in West Malaysia repeatedly from March to June, covering three counties (two federal territories and one neighbouring state) and moved to East Malaysia in July covering two other counties. The most likely cluster shows a tendency of having moved from the western part to the eastern part of the country. These results provide information that can be used for the evidence-based interventions to control the spread of COVID-19 in Malaysia.

Introduction

A cluster of pneumonia cases infected with a coronavirus was reported in Wuhan, a city in Hubei province of China in December 2019 (Huang et al., 2020; Zhu et al., 2020). The Chinese authorities officially announced on 7 January 2020 that the aetiological agent was a novel coronavirus, now known as SARS-CoV-2, which is the cause of the acute respiratory coronavirus disease 2019 or COVID-19 (Zhou et al., 2020). The World Health Organization (WHO) declared COVID-19 outbreak, first as a Public Health Emergency of International Concern on 30 January 2020 and subsequently as a worldwide pandemic on 11 March 2020 (WHO, 2020, 2020a). Population density and intensity of social contacts are the main drivers for the propagation and amplification of this virus (Rocklöv & Sjödin, 2020).

COVID-19 was confirmed to have reached Malaysia in January 2020, when it was detected on travellers from China arriving via Singapore on 25 January 2020, following the first outbreak in Hubei Province, China (Reuters Staff, 2021). The number of reported cases remained first low (fewer than 30 cases) and were largely confined to imported cases until localized clusters began to emerge in March 2020; the largest cluster was linked to a religious gathering held in Sri Petaling, Kuala Lumpur in late February and early March 2020, leading to massive spikes of local cases and export of the infection to neighbouring countries by travel (Ng, 2021). Within a few weeks, Malaysia had recorded the largest cumulative number of confirmed COVID-19 infections in Southeast Asia (The Strait Times, 2021), breaching the number of 2000 active cases by the end of March 2020.

Few studies have been conducted focusing some of the characteristics of COVID-19 pandemic in Malaysia, such as impact of preventive measures on the infection dynamics (Khor et al., 2020; Salim et al., 2020), effect on mental health (Shanmugam et al., 2020) and behavioural changes in the general population (Hanafiah & Wan, 2020; Koh et al., 2020). However, spatial cluster analysis of COVID-19 cases in the country seems not to have been carried out so far. This would be useful by guiding control programs on prioritizing locations for targeted interventions, rapid testing and allocation of other limited resources.

One of the statistical methods for detecting spatial/geographical clusters of a disease is the Spatial Scan Statistic (Kulldorff, 1997), which is a freely available software called SaTScan™.
It is widely used in public health and epidemiology (Amin et al., 2019; Omrodor et al., 2019; Souza et al., 2019; Amin & Rivera, 2020; Desjardins et al., 2020; Ullah et al., 2020). There are several other cluster analysis algorithms in the literature, such as DBSCAN (Ester et al., 1996). It is understood that none of the cluster analysis methods is better than all the others. Each algorithm has some advantages for specific applications. We used SaTScan™ as it has the advantage of detecting cluster location, while also evaluating its statistical significance. This approach supplement and assist creation of basic maps of disease rates by relying on a variety of data models to determine whether the observed spatial or space-time patterns of a disease are due to chance or if they are randomly distributed. A useful function of SaTScathen™ that assist decision making is the possibility to integrate model outputs and application of geographic information systems (GIS) for a map-visualization.

This study aimed to identify statistically significant spatial clusters of COVID-19 cases at the state-level in Malaysia during the period from March to September 2020 using scan statistics under the discrete Poisson model. Spatial analysis was carried out for each month to identify the location of significant clusters in each month of the pandemic period.

**Materials and methods**

**Study area and data**

This study was conducted at the state-level in Malaysia. The total terrestrial area of Malaysia is divided into thirteen states and three federal territories, separated by the South China Sea into two regions, East Malaysia and West Malaysia. The geographical map of the study area is shown in Figure 1. West Malaysia shares a land and maritime border with Thailand and maritime borders with Singapore, Vietnam and Indonesia. East Malaysia shares land and maritime borders with Brunei and Indonesia and a maritime border with The Philippines and Vietnam. The monthly data on COVID-19 cases were collected for each of the sixteen counties (thirteen states and three federal territories) during the period from March to September 2020 from the Ministry of Health Malaysia (2020). These data are currently updated on a daily basis while the confirmed cases are spatially aggregated at the state-level on a monthly basis. The population counts for the respective states with respect to the second quarter 2019 were obtained from Demographics Statistics Malaysia (Statista, 2020).

**Spatial scan statistic**

Spatial scan statistics scans the region under study by a circular moving window defining a set of zones. The circular scanning window is centred on each county’s centroids with a radius size continuously varying from zero up to a fixed value. The circle includes the counties whose centroids lie within the circle. Hence, a large number of overlapping circles of different sizes are obtained that mutually cover the total study area. Each circle denotes a possible candidate cluster.

When using this approach with the Poisson model, the case counts, \( x_i \) in each circular zone \( z \) are assumed to be distributed under the null hypothesis \( H_0 \) of spatial randomness (Eq. 1):

\[
x_i \sim \text{Pois}(\lambda_0)
\]

The likelihood functions \( L_0 \) with the parameter restricted under the null hypothesis \( H_0 \) and \( L_1 \) with parameter unrestricted under the alternative hypothesis are then computed for each circular zone \( z \). In each zone \( z \), \( L_0 \) and \( L_1 \) have different parameters, from the heterogeneous population distribution. The method searches for zone \( z \) that maximizes the likelihood ratio (LR) as in Eq. 2:

\[
\text{LR}(z) = \frac{L_1}{L_0}
\]

![Geographical map of the study area (Wikipedia, 2020).](image-url)
For each possible zone, the Log-likelihood ratio (LLR) is calculated and its statistical significance tested by Monte Carlo simulations (Kulldorff, 2018). For each potential cluster, a large number of replications of datasets (usually 999 or 9999) using the restricted parameters under H0 is generated. The LLR is computed and a distribution is obtained. If the potential cluster has an LLR higher than, for instance, 95% of the datasets generated under the null hypothesis, the cluster is said to be statistically significant at 0.05 confidence level. The zone with the maximum LLR value and with the statistical significance is considered to be the most likely cluster, that is, the cluster least likely to have occurred by chance. The other zones with statistically significant LLR values are defined as the secondary clusters.

In this study, the spatial scan statistics was applied under the discrete Poisson Model using circular windows and with the default setting of maximum spatial window sizes \( i.e. <50\% \) of the total population of the study area (Kulldorff, 2018). The clusters having a P-value of less than 0.05 were considered to be the significant clusters.

## Results

Potential spatial clusters of COVID-19 cases in Malaysia were detected during each month from March to September 2020 (Table 1). In March, most likely clusters were seen in the three states (Negeri Sembilan, Putrajaya, Kuala Lumpur), and a secondary cluster in one state (Selangor). In April, the same states were detected as most likely clusters with Sarawak as the 1st secondary cluster, and Selangor as the 2nd secondary cluster. In May, the most likely clusters were again detected in same states, this time with a secondary cluster in only one state (Selangor). In June, the picture was again the same with respect to the most likely clusters, but this time the secondary cluster was in Melaka State. In July, the most likely cluster was detected in East Malaysia, covering two counties (Sarawak and Labuan), while two secondary clusters were detected in West Malaysia, each covering a single county (Kuala Lumpur and Johor). In August, the Most likely cluster was seen in Kuala Lumpur, 1st secondary cluster in two states (Perlis, Kedah), and the 2nd secondary cluster in a single county (Labuan). In September, the most likely cluster had moved back to the Eastern state Saba, while the secondary cluster was seen in the western state Kedah. All the clusters detected in March, April, and May were statistically significant at P<0.05. However, the secondary clusters in June (Melaka), in July (Johor), in August (Labuan) and in September (Kedah) did not reach statistical significance at this level (Table 1).

The resulting geographical locations of COVID-19 clusters in the study area for each month from March to September are shown in MAPS IN Figures 2-8.

## Discussion

The results show that the potential spatial clusters of COVID-19 tend to move from the Western part to the Eastern Part of the country. The most likely cluster has occurred repeatedly in each consecutive month from March until June, covering three states (Negeri Sembilan, Putrajaya, and Kuala Lumpur) in the western part of the country. The strong spread of COVID-19 in these states can be attributed to the cramped living conditions of foreign worker residences in factories and immigration detention centres (CodeBlue, 2020; Malaysiakini, 2020). The large number of COVID-19 can also be linked to a religious gathering at the Seri Petaling Mosque in Kuala Lumpur (BH Online, 2020; TheStar, 2020). However, in July, the most likely cluster moved from West Malaysia to East Malaysia. In August it moved to the West Malaysia for one month and then again back East Malaysia in September. The spike in COVID-19 cases in the eastern part of the country may be due to the expatriates from Indonesia crossing the border illegally. From the spatial point of view, the clusters were inconsistent during the whole study period i.e., they moved from one place to another without producing a trend. The control measure of the government, such as movement control order (MCO) in the highly effected areas, and events of social gathering in other

### Table 1. Spatial clusters of COVID-19 cases in Malaysia during each month from March to September.

<table>
<thead>
<tr>
<th>Month</th>
<th>Cluster type</th>
<th>State</th>
<th>No. of states</th>
<th>Observed cases (no.)</th>
<th>Expected cases (no.)</th>
<th>LLR</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>March</td>
<td>Most likely</td>
<td>Negeri Sembilan, Putrajaya, Kuala Lumpur</td>
<td>3</td>
<td>637</td>
<td>256</td>
<td>230.282</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>1st secondary</td>
<td>Selangor</td>
<td>1</td>
<td>704</td>
<td>554</td>
<td>23.816</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>April</td>
<td>Most likely</td>
<td>Negeri Sembilan, Putrajaya, Kuala Lumpur</td>
<td>3</td>
<td>1189</td>
<td>29</td>
<td>900.917</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>1st secondary</td>
<td>Sarawak</td>
<td>1</td>
<td>351</td>
<td>279</td>
<td>9.361</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>2nd secondary</td>
<td>Selangor</td>
<td>1</td>
<td>727</td>
<td>648</td>
<td>5.788</td>
<td>0.116</td>
</tr>
<tr>
<td>May</td>
<td>Most likely</td>
<td>Negeri Sembilan, Putrajaya, Kuala Lumpur</td>
<td>3</td>
<td>1157</td>
<td>168</td>
<td>1627.381</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>1st secondary</td>
<td>Selangor</td>
<td>1</td>
<td>478</td>
<td>364</td>
<td>20.799</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>June</td>
<td>Most likely</td>
<td>Negeri Sembilan, Putrajaya, Kuala Lumpur</td>
<td>3</td>
<td>566</td>
<td>73</td>
<td>893.267</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>1st secondary</td>
<td>Melaka</td>
<td>1</td>
<td>28</td>
<td>23</td>
<td>0.598</td>
<td>0.983</td>
</tr>
<tr>
<td>July</td>
<td>Most likely</td>
<td>Sarawak, Labuan</td>
<td>2</td>
<td>109</td>
<td>31</td>
<td>70.402</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>1st secondary</td>
<td>Kuala Lumpur</td>
<td>1</td>
<td>40</td>
<td>19</td>
<td>9.706</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>2nd secondary</td>
<td>Johor</td>
<td>1</td>
<td>51</td>
<td>40</td>
<td>1.672</td>
<td>0.601</td>
</tr>
<tr>
<td>August</td>
<td>Most likely</td>
<td>Kuala Lumpur</td>
<td>1</td>
<td>111</td>
<td>20</td>
<td>113.008</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>1st secondary</td>
<td>Perlis, Kedah</td>
<td>2</td>
<td>120</td>
<td>27</td>
<td>99.452</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>2nd secondary</td>
<td>Labuan</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2.251</td>
<td>0.369</td>
</tr>
<tr>
<td>September</td>
<td>Most likely</td>
<td>Sabah</td>
<td>1</td>
<td>1460</td>
<td>226</td>
<td>2148.832</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>1st secondary</td>
<td>Kedah</td>
<td>1</td>
<td>151</td>
<td>126</td>
<td>2.504</td>
<td>0.314</td>
</tr>
</tbody>
</table>

LLR, Log-likelihood ratio.
Figure 2. COVID-19 cluster map for March.

Figure 3. COVID-19 cluster map for April.
Figure 4. COVID-19 cluster map for May.

Figure 5. COVID-19 cluster map for June.
Figure 6. COVID-19 cluster map for July.

Figure 7. COVID-19 cluster map for August.
areas might have caused the spatial clusters of COVID-19 to change their locations in this way. However, the geographical cluster locations identified deserve increased public health and clinical attention as there is important to control the spread of COVID-19 in the country. Some important patterns in the spatial clusters of COVID-19 cases were identified. These patterns are of high significance for epidemiologists to find the environmental and social factors that possibly foster the spread of COVID-19 and other infectious diseases in an area.

Conclusions

This study contributes to the ongoing COVID-19 surveillance efforts by highlighting the potential spatial clusters for targeted interventions to mitigate further spread. The presence of spatial clusters of COVID-19 at the state-level in Malaysia during each month from March to September 2020 provides clues for policy-makers for better designing and implementing the control measures at the local level. With knowledge of the most likely cluster showing a tendency of moving from West Malaysia toward East Malaysia in September, opens the possibility to intervene pre-emptively. Our analysis also provides a good understanding of the temporal pattern of the spatial clusters of COVID-19 in the study area which should be useful for the surveillance of COVID-19 and other infectious diseases.

References


