Piecewise linear modelling and change-point analysis of COVID-19 outbreak in Malaysia

P Phang¹, N A Abdul Taib², R Safii³ and J Labadin⁴

¹Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia
²Department of Community Medicine and Public Health, Faculty of Medicine and Health Science, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

¹E-mail: npfang@unimas.my

Abstract. In Malaysia, COVID-19 were first detected as imported cases on 25 January and as local infection on 4 February 2020. A surge of positive cases ensued by March 2020 which led to a series of countrywide containment and mitigation measures known as Movement Control Order (MCO). We study the direct effects of MCO on the course of epidemic by analyzing the cumulative and daily infection cases of COVID-19 up to 31 December 2020 in Malaysia and its states using piecewise linear regression and segment neighborhoods algorithm of change-point analysis, respectively. Through piecewise regression on nationwide cases, MCO were likely to almost flatten the epidemic curve in just one month after it was first initiated. While for stateswide cases, the average length of series of concave downward is six months before it turn to concave upward, indicating the period of which deceleration of new cases can be expected. However, the starting of this wave of COVID-19 can be relatively vary for three months in different states and federal territories. Together with change-point analysis on daily cases, the statewise epidemic phases could be subdivided into two to four regimes, whereby the majority of phase transitions fall in April and last quarter of 2020. Overall, the statistical modelling shows that the immediate effect of MCO appears to be effective.

1. Introduction

Coronavirus Disease 2019 (COVID-19) is an acute respiratory illness caused by a novel strain of coronavirus, namely Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-COV-2), believed to be derived from bats [1–2]. The disease first emerged in December 2019 in Wuhan city of Hubei Province; the epicenter of COVID-19 outbreak in China. Within three months, the disease has spread interminably on a global scale prompting the World Health Organization (WHO) to declare it as a pandemic on March 2020 [3]. As of 23 February 2021, there have been approximately 111,400,000 laboratory-confirmed cases of COVID-19 reported worldwide with 2,470,772 fatalities according to WHO [4].

Malaysia initially joined the extensive list of countries affected by the novel coronavirus when three imported cases of COVID-19 were detected on 25 January 2020 amongst Chinese nationals who visited Malaysia via Singapore on 23 January 2020 [5]. By the beginning of February, the first local infection was identified on 4 February 2020, in which the victim had been travelling to Singapore for business purposes [6–7]. Two days later, Malaysia announced its first case of COVID-19 infected via local
transmission concerning a 40-year-old Malaysian female [7]. A surge of positive cases ensued by March 2020 and eventually the first reported death case was confirmed on 17 March 2020 [8]. In response to the rapid growth of cases in Malaysia, a country-wide lockdown formally known as the Movement Control Order (MCO) was implemented on the very next day. Subsequently, several different stages of MCO have been enforced as the country’s containment and mitigation strategies [9].

Prior modelling studies in regards to the Malaysian COVID-19 situation have been focusing on deterministic ordinary differential equation (ODE) models vis-à-vis a statistical approach, in which several authors have applied the compartmental framework while investigating the transmission dynamics of COVID-19 in Malaysia [9–15]. Five separate researchers [10–13] and [15] have developed the classic, compartmental Susceptible-Infectious-Recovered (SIR) model and predicted the infection peak period to be in the middle of March 2020 [10–13]. Two group of authors [9] and [14] have incorporated the incubation period and included an additional Exposed compartment in order to utilize the Susceptible-Exposed-Infectious-Recovered (SEIR) framework in their models.

At the time of writing, only two group of authors [16–17] have opted for the statistical approach in studying the transmissibility and transmission pattern of Malaysian COVID-19, as per our knowledge. Amiruzzaman et al. [16] employed the logistic growth curve to model the Malaysian COVID-19 infection and suggested that the first stage of MCO had a more substantial impact in terms of the virus containment. Erde et al. [17] implemented a time series forecasting of COVID-19 infected cases by deploying the auto regressive integrated moving average (ARIMA) model. Findings have demonstrated that the MCO measures could stabilize the growth rate of COVID-19 infection in Malaysia.

Due to the dearth of studies, our work aims at comprehending the various stages of MCO in Malaysia using piecewise linear regression for the cumulative COVID-19 positive cases and change-point analysis on the daily cases in Malaysia. State-wide cumulative and daily cases for all thirteen states and three federal territories from 13 March to 31 December 2020 are included in the analysis. We aim at giving a coarse description on spatio-temporal variability of the COVID-19 spread in Malaysia through statistical analysis.

It is commonly expected that time lags exist in between the dates of epidemic mitigation measures are initiated and the abrupt change of disease spreading rate. These time lags could be caused by observation delay, which is a combination of incubation period and delayed test results [18]. Consequently, it is difficult to predetermine the exact date when the MCO shows an effect. Hence, we perform the time series structural change analysis by first partitioning the cumulative infection data into several linear segments with different slopes through piecewise linear regression model, so as to analyze the trajectory of infection curve of COVID-19 in Malaysia in terms of its infection growth rate and the emergence of phases of transition. As each phase could share common statistical properties such as mean and variance across its observations, we then perform the change-point analysis of daily infection cases to further assess the division of the cumulative infection curve obtained from piecewise regression.

Similar to the construction of piecewise linear trend model in [19], our simple inspection of cumulative infection data enables us to better understand the association of growth patterns of COVID-19 cases and the date when various stages of MCO were implemented in Malaysia. This could present a timely evaluation of the effectiveness and impact of MCO in combating the spread of COVID-19 in entire Malaysia and also different states, as the first or second breakpoint after the steepest slope segment in piecewise linear regression model is greatly related to the date of the containment measures were put in place. We make use of change-point analysis as the second approach in quantifying the start and end of one period transitions into another. This is important to verify whether the segmentation of piecewise regression makes sense by taking into account its statistical properties over segments.

In the following, we briefly give the scope of data used. We propose to model the logarithmic scale of cumulative confirmed cases using piecewise regression as well as to study the statistical properties of segmentation in daily infection cases through change-point analysis. Results show that several transition phases could be derived based on both methods, which would in turn help in evaluating the effect of containment measures on the COVID-19 outbreak in Malaysia.
2. Methods

2.1 Data source

We focus on cumulative COVID-19 infection cases for entire Malaysia as well as all 13 states and 3 federal territories. We obtain the COVID-19 data for entire Malaysia from https://ourworldindata.org/coronavirus-source-data maintained by “Our World in Data”. For statewise level, data published by Ministry of Health Malaysia on the cumulative number of COVID-19 infected cases from 13 March to 31 December 2020 were analyzed.

We first give the evolutions of the cumulative infected cases in entire Malaysia in figure 1, which also depicts several stages of Movement Control Order (MCO) (see table 1) implemented by the federal government of Malaysia in response to the COVID-19 pandemic in the country.

![Figure 1. COVID-19 outbreak data (+) for Malaysia, from 25 January till 31 December 2020.](image)

Even though the nation’s first detection of COVID-19 was reported on 25 January 2020, the two states (Sabah and Sarawak) in the eastern part of Malaysia remained COVID-free until 13 and 12 March, respectively. Hence, we focus on the cumulative and daily confirmed cases of COVID-19 since 13 March 2020 in all thirteen states and three federal territories in Malaysia, until 31 December 2020 which marks the end of one of the stages of MCO.

| Table 1. Stages of Movement Control Order |
|------------------------------------------|
| Stages of mitigation measures | Date |
|-----------------------------------|------|
| MCO (Movement Control Order) | 18 – 31 March 2020 |
| MCO First Extension | 1 – 14 April 2020 |
| MCO Second Extension | 15 – 28 April 2020 |
| MCO Third Extension | 29 April – 3 May 2020 |
| CMCO (Conditional MCO) | 4 May – 9 June 2020 |
| RMCO (Recovery MCO) | 10 June – 31 August 2020 |
| RMCO (Recovery MCO) Extension | 1 September – 31 December 2020 |

2.2 Piecewise linear regression model

It is well known that the early stages of epidemics typically exhibit exponential growth \( I(t) = \alpha e^{\beta t} \), which is the solution to simple first-order ordinary differential equation

\[
\frac{dI(t)}{dt} = \beta I(t), \quad I(0) = \alpha
\]
where \( f(t) \) is a time-dependent variable for infected cases, \( \alpha \) is the initial infected number and \( \beta \) is the epidemic growth rate. Equation (1) implies that the change in the cumulative of infected cases is proportional to the current cumulative cases. In the context of modelling the relationship between the variables by fitting a linear equation, we take \( \ln \) on both sides of \( f(t) = \alpha e^{\beta t} \) and let \( \beta \equiv \hat{\beta} \), \( \alpha \equiv \hat{\alpha} \) to arrive at

\[
\ln \hat{f}(t) = \hat{\alpha} + \hat{\beta}t \tag{2}
\]

where \( \hat{f}(t) \) is the predicted cumulative number of reported infected cases at day \( t \), \( \hat{\alpha} \) is the estimated intercept at vertical axis and \( \hat{\beta} \) is the estimated slope coefficient. The left-hand side of equation (2) is in natural logarithmic scale such that the estimated slope \( (\hat{\beta}) \) naturally measures the growth rate of the epidemic at day \( t \). It is easy to note that differentiating both sides of equation (2) and dropping hat symbols, we recover exponential model (1). Also, the slope \( \hat{\beta} \) can be used to estimate the epidemic doubling time, \( \Delta t = (\ln 2)/\hat{\beta} \). Hence, we carry out linear regression for the logarithmic transformation of the cumulative infected curves of Malaysia along with all states and federal territories. The estimated slope of linear regression gives the growth rate \( \hat{\beta} \) for the infection. The growth rate is important in measuring the spreading speed of an epidemic. If the rate is positive (resp. negative), then the disease can (resp. cannot) invade a population, where the zero rate is indeed a disease threshold.

However, as linear regression assumes a uniform growth rate over entire period of time, it may not realistic in modelling epidemic in presence of interventions. Due to the non-linear behavior of the time-dependent cumulative case data, there is no single linear function appears to fit with the cumulative infected cases of either Malaysia or all states and federal territories in logarithm scale. Hence, we employ piecewise (a.k.a. segmented) linear regression for constructing best-fit linear function onto different ranges of \( t \). A segmented linear model is defined by slope parameters and breakpoints where the slope of the linear function changes between two adjacent segments, which may be represented by

\[
\ln \hat{f}(t) = \hat{\alpha} + \hat{\beta}_1 t + \sum_{j=1}^{m} \hat{\beta}_{j+1}(t - \psi_j)u(t - \psi_j) \tag{3}
\]

where \( \hat{\beta}_j \) is the slope of segment \( j \) ( \( j = 1, 2, \ldots, m+1 \) ), \( \psi_j \) is the location of slope changes (i.e. breakpoint), \( u(t - \psi_j) \) is an indicator function equal to one when \( t \geq \psi_j \) and 0 otherwise, and \( m \) is the number of breakpoints.

Having said that, in piecewise curve fitting, the number of breakpoints is typically unknown. As the curves of the cumulative infection generally manifest transitions from exponential regime to moderate growth and then to flatter curve, we start our curve fitting procedure with minimal breakpoints of two and the initial guess of these breakpoints is determined visually. As R-squared measures how successful the fit is in explaining the variation of the data in which a value closer to 1 indicating a better fit, whenever a desirable adjusted R-squared value that is greater than 0.99 is obtained, then the segmented linear model is chosen as the best-fit piecewise function. Otherwise, the piecewise curve fitting procedure is re-carried out by adding an additional number of initial guess of breakpoint on the portion of fit curve that does not match well with the data points. The respective best-fit piecewise linear functions, together with its cumulative infection curve in semi-logy plots, for entire Malaysia is shown in figure 2, while for all states and federal territories, the plots are displayed in figure 3 in descending order of magnitude in total cases as of 31 December 2020.

### 2.3 Change point analysis

Apart from partitioning the cumulative infection curves into several segments with different slopes (and intercepts), we examine the daily cases of COVID-19 in all states and federal territories, through change-point analysis of its statistical properties (either/both mean or/and variance). In this paper, we choose to reserve the terminology “change point” for an instance in time where the statistical properties before and after this time point differ, whereas the terminology “break point” represents the transition between
different slopes of piecewise linear regression model, despite the fact that they are usually interchangeable in time series structural analysis. On account of this, we are interested to give a simple comparison for the abrupt variations from the perspectives of aggregated (cumulative) cases and the individual (daily) cases. That is, to what extent, the shift of the epidemic phase in terms of its disease growth rate and its basic statistical properties of mean or variance of daily cases agree with each other.

Rather than finding the optimal positioning and potential number of change points for daily infected cases in mean and/or variance, we decide the number of change points based on the number of breakpoints in our fitted piecewise regression model for cumulative infection curves. For this purpose, we employ the segment neighborhood algorithm [20] in Changepoint package [21] by presetting the number of change-point to search for. We run the algorithm for detecting the abrupt changes in following three statistical properties, namely mean, variance as well as combination of both mean and variance. Only one of these that returns change-points identified closest to the breakpoints in its piecewise linear regression counterpart will be used for comparison purpose.

3. Analysis for cumulative confirmed cases and daily cases of COVID-19

3.1 Analysis of cumulative confirmed cases in entire Malaysia
At nationwide level, the growth rate in the first segment ($S_1=0.1016$) in figure 2 reveals the first wave of COVID-19 in Malaysia, which can be attributable to the imported cases from foreign nationals entering Malaysia through Singapore and China [5]. During this period, the local government had to conduct extensive health screening with thermal scanners and check-ups at all entry points into the country. The implication of this containment measures can be exemplified in the reduction of the slope on second segment ($S_2=0.0130$) whereby the source of the bulk of infected cases during this time has been shifted from imported cases to local transmission cases instead.

![Figure 2](image)

Figure 2. The cumulative infection curve (+) in semi-log plot and its best-fit piecewise linear regression (blue line). The blue squares give the breakpoints of piecewise linear segment and the slopes of linear segment are given by $S_1$, $S_2$, ..., $S_6$.

Apparently, the largest slope is found to be on third segment ($S_3=0.1924$) which lies in between 29 February and 23 March. Since exponential laws will show up as straight lines in logarithm scale, this segmented period of time could be regarded as the most significant exponential regime of the epidemic curve for entire Malaysia. The rapid growth of COVID-19 cases in this period was highly associated to a religious mass-gathering which took place in a Malaysian mosque, from 27 February to 1 March, that kicked off the second wave of COVID-19 infections in the country [22]. The Malaysian Health Ministry also declared it as the first local transmission cluster in Malaysia [23]. To combat this second wave, the Movement Control Order (MCO) was initiated on 18 March 2020. From our piecewise linear segment model, we found that the epidemic growth rate became moderate after 23 March and further reduced to almost flattening from 20 April onwards. However, from 21 September onwards, the entire
Malaysia could be considered as entering the third wave of COVID-19 with the slope of linear segment $S_6 = 0.0255$.

3.2 Analysis of cumulative confirmed cases in 13 states and 3 federal territories

In statewise level, the semi-log plots of cumulative infection cases for all states and federal territories, together with their fitted piecewise linear regression models are portrayed in figure 3. Unlike those nations with the order of magnitude of cumulative cases in $10^5$ to $10^6$ that experiencing the transition from exponential to power-law regime before flattening of the epidemic curve [24], the cumulative cases for all states and federal territories in Malaysia up to 31 December 2020 are in the orders ranging from tens to tens of thousands. Therefore, they can be fit well by linear functions with adjusted R-squared all close to 1.

As $m$ breakpoints will split the data into $m+1$ segments, the slopes for these segments (i.e. $S_1, S_2, ..., S_m, S_{m+1}$) together with its breakpoints (BP) are given in table 2. With slope in each line segment, the doubling period of COVID-19 cases can be estimated. For instance, at its third breakpoint (BP3) on 5 September 2020, the cumulative cases in Sabah is 433 and the estimated slope of line segment $S_4$ is 0.0641. By calculating $(\ln 2)/0.0641$, we obtain the estimated doubling period as 10.81 days. As a matter of fact, Sabah recorded 877 cumulative cases on 15 September 2020, which was just slightly higher than two times of 433. This shows that the resulting segmentation from piecewise regression can be useful in capturing the spread of the COVID-19 in all states and federal territories.
As we list the states and federal territories in decreasing order of magnitude, the number of breakpoints estimated is apparently not related to its order of magnitude. We highlight by bold font the greatest slope among the linear segments in table 2 and find that all (except Labuan) have the highest slope in its first linear segment ($S_1$), in which these steepest segments are all (except Labuan) ended before or on 3 April 2020. General speaking, this early stage of the pandemic often depicts the worst scenario before the interventions first show an effect. This suggests that the nationwide interventions from the federal government managed to bring down the statewide slopes by a significant amount in just roughly two weeks since the implementation of MCO on 18 March 2020.

Table 2. Estimation results for breakpoints (BP) and its corresponding slope of linear segments (S) for cumulative infection curves in semi-logy scale of all states and federal territories.

| State (Abbr.) | $S_1$ | BP1 | $S_2$ | BP2 | $S_3$ | BP3 | $S_4$ | BP4 | $S_5$ | BP5 | $S_6$ |
|---------------|--------|------|--------|------|--------|------|--------|------|--------|------|--------|
| Sabah (SBH)   | 0.3367 | Mar-19 | 0.0269 | Apr-16 | 0.0024 | Sep-05 | 0.0041 | Nov-04 | 0.0115 |
| Selangor (SGR) | 0.1087 | Apr-03 | 0.0100 | Jun-01 | 0.0011 | Oct-07 | 0.0324 |        |        |
| KL (KUL)      | 0.1133 | Apr-03 | 0.0191 | Jun-05 | 0.0013 | Nov-02 | 0.0269 |        |        |
| N.Sembilan (NSS) | 0.1550 | Mar-27 | 0.0307 | May-18 | 0.0013 | Oct-21 | 0.0413 | Nov-25 | 0.0158 |
| Johor (JHR)   | 0.2704 | Mar-20 | 0.0699 | Apr-10 | 0.0011 | Oct-07 | 0.0060 | Nov-22 | 0.0429 |
| Penang (PNG)  | 0.2614 | Mar-21 | 0.0334 | Apr-07 | 0.0011 | Oct-07 | 0.1373 | Oct-20 | 0.0212 |
| Perak (PRK)   | 0.8296 | Mar-16 | 0.2101 | Mar-26 | 0.0306 | Apr-09 | 0.0005 | Oct-09 | 0.0232 |
| Kedah (KDH)   | 0.2294 | Mar-23 | 0.0015 | Jul-25 | 0.0202 | Sep-30 | 0.2106 | Oct-07 | 0.0068 |
| Labuan (LBN)  | 0.0662 | Apr-06 | 0.0031 | Oct-01 | 0.0013 |        |        |        |        |
| Pahang (PHG)  | 0.8458 | Mar-16 | 0.0832 | Apr-11 | 0.0076 | May-29 | 0.0002 | Sep-13 | 0.0010 |
| Melaka (MLK)  | 1.1513 | Mar-15 | 0.0800 | Apr-17 | 0.0017 | Oct-15 | 0.0067 | Dec-22 | 0.0570 |
| Sarawak (SWK) | 0.5790 | Mar-17 | 0.0941 | Apr-06 | 0.0262 | Apr-27 | 0.0023 | Oct-22 | 0.0147 |
| Kelantan (KTN) | 0.5247 | Mar-18 | 0.0837 | Apr-01 | 0.0003 | Sep-25 | 0.0032 | Nov-10 | 0.0350 |
| Terengganu (TRG) | 0.4565 | Mar-20 | 0.0639 | Apr-10 | 0.0003 | Sep-24 | 0.0146 | Nov-03 | 0.0063 |
| Putrajaya (PJY) | 0.2457 | Mar-24 | 0.0513 | Apr-23 | 0.0009 | Oct-01 | 0.0134 | Nov-13 | 0.0074 |
| Perlis (PLS)  | 0.8047 | Mar-15 | 0.0300 | Apr-13 | 0.0001 | Aug-03 | 0.1216 | Aug-07 | 0.0031 |

Thereafter, the spread rates can be considered as moderate or even slow for a window of more than 4 months. To further investigate these phase transitions, we illuminate the concave upward slope increasing line segments as compared to its previous line segment by underlining those slope increases in table 2. On this account, we discover that all states and federal territories have concave downward slope decreasing line segments for first two segments. The average length of series of downward slope decreasing line segments in all piecewise linear regression is 6.6 months before it turn to concave upward. However, the lengths of these segments could range from earliest 134th days (or 25 July) for Kedah to latest 234th days (or 2 November) for Kuala Lumpur, indicating the starting of the third wave of COVID-19 can be relatively vary for three months in different states and federal territories.

We then compare the maximum slope (denoted as $S_{\text{max}}$, which are the slopes in bold font in table 2) and the slope after the latest breakpoint (denoted as $S_{\text{cur}}$, which are indeed the slope of last segments in each row of table 2) in figure 4. Specifically, the vertical axis, $S_{\text{max}}$ gives the epidemic growth rate at its worst scenario for a particular state or federal territory. Meanwhile, the horizontal axis, $S_{\text{cur}}$ provides the ongoing epidemic growth rate on the cutoff date (31 December 2020) of this analysis. By carrying out distance-based clustering, we find that states and/or federal territories within the same cluster tend to have similar maximum growth rate. It is interesting to note that several geographically neighboring states are close enough in their relative positions in figure 4. These include Selangor and Kuala Lumpur, Terengganu and Kelantan, as well as Perak and Pahang. Also, all six states (namely Selangor, Kuala Lumpur, Johor, Perak, Penang and Melaka) with concave upward slope in their respective last segment appear to be at the rightmost position of this plot.

We could make use of horizontal axis in figure 4 for short term prediction of statewide COVID-19 trajectories. If state A is to the right of state B, then A has a greater potential surge in COVID-19 numbers after the cutoff date. For instance, Johor reported the highest daily cases with 1069 on 27 January 2021 [25], which is more than twice as many as its highest daily case in the year before. However, in view of the rapid changing growth rate of COVID-19 [26], we unavoidably underestimated the infection cases in certain state in such a simple plot of figure 4. As an example, Sarawak has been continuing record
three digits daily cases since 11 January 2021 [27] as opposed to its highest recorded daily cases which was just 32 throughout the whole year 2020. We find that Sarawak is at the leftmost position in our plot and this could indicate the delay in the outbreak for the Sarawak state.

![Figure 4. The clustering plot of maximum slope ($S_{\text{max}}$) and slope after the latest breakpoint ($S_{\text{cur}}$) in piecewise linear regression model for cumulative confirmed cases of all states and federal territories.](image)

3.3 Analysis of daily infection cases in 13 states and 3 federal territories
As the cumulative infection cases are highly non-stationary, we further investigate the possible underlying causes of the piecewise linear segmentation by carrying out segment neighborhoods algorithm of analysis of change-point whereby fundamental shifts in statistical properties such as mean and variance for daily infected cases are found. Three types of change-point are sought, namely abrupt

![Figure 5. Plot of the COVID-19 daily infected cases along with horizontal (or vertical) blue lines for segment neighborhoods change-points, in which the change-point is denoted as the first observation of the new segment. The analysis is for changes in either both mean and variance (meanvar), mean (mean) or variance (var) for all states and federal territories.](image)
change in both mean and variance (denoted as “meanvar”), as well as abrupt change in mean (denoted as “mean”) or in variance (denoted as “var”). Among these three types of change-point, one of these that best tally with its corresponding breakpoints in piecewise regression are chosen. Figure 5 gives the statewise COVID-19 daily infection cases with its best type of change-point.

The time points (i.e. specific dates) where such change-points detected by segment neighborhoods algorithm are listed in table 3. For ease of comparison, we underline the change-points in table 3 that fall within 14 days if compared with its corresponding breakpoint given in table 2. Except for Penang and Perak, all states and federal territories have at least two change-points that tally with its breakpoints. In certain states, the second best type of change-points which gives the same number of change-points tally with its breakpoints is also given in the column “Other” in table 3. Overall, the change in mean and variance (i.e. meanvar) gives the better results in which it suits 7 out of 16 states and federal territories, especially for states which have higher order of magnitude. Despite the fact that the change in mean is very sensitive to outliers, five states have the abrupt change in mean in their daily infection cases that best tally with its breakpoints.

Table 3. The change-points (CP) detected on daily infection cases of all states and federal territories.

| State   | Type   | CP1  | CP2  | CP3  | CP4  | CP5  | Other |
|---------|--------|------|------|------|------|------|-------|
| Sabah   | meanvar| Apr-19| Aug-23| Sep-01| Oct-06|     |       |
| Selangor| meanvar| Jun-08| Sep-27| Nov-19|       |     |       |
| KL      | meanvar| Jun-07| Oct-08| Nov-12|       |     |       |
| N.Sembilan| var   | Oct-24| Nov-08| Dec-08|       | mean|       |
| Johor   | meanvar| Apr-19| Sep-04| Sep-24| Nov-21|     |       |
| Penang  | mean   | Oct-12| Nov-06| Nov-07| Dec-31| var |       |
| Perak   | meanvar| Jun-19| Jul-19| Aug-11| Sep-29| var |       |
| Kedah   | mean   | Oct-01| Oct-05| Oct-07| Oct-08| var |       |
| Labuan  | meanvar| Apr-18| Jun-06| Oct-04|       |     |       |
| Pahang  | meanvar| Jun-14| Jul-23| Jul-27| Sep-02| Dec-03| var   |
| Melaka  | var    | Apr-18| Apr-20| Dec-16|       |     |       |
| Sarawak | mean   | Mar-31| Apr-04| Apr-27| Oct-24| Nov-14|       |
| Kelantan| mean   | Apr-01| Nov-16| Dec-05| Dec-06| Dec-26| var   |
| Terengganu| var     | Apr-10| Oct-03| Nov-07| Dec-12|     |       |
| Putrajaya| mean   | Apr-21| Oct-01| Nov-15| Dec-16|     |       |
| Perlis  | var    | Apr-15| Jul-29| Aug-06| Aug-08|     |       |

We then give the visual connection between breakpoints and change-points through a connected dot plot in figure 6, which is produced by plotting the underlined change-points in table 3 and its corresponding breakpoints in table 2. The positions of pairs of dots illustrate the significant changes in both cumulative and daily infection cases for a particular state or federal territory. For instance, COVID-19 cases trajectory in Sabah could be concurrently segmented into three time windows in which the associated regime switching occurs in the mid of April and early September 2020, respectively. On the other hand, both cumulative and daily cases in Perak and Penang underwent only one significant structural change simultaneously in early October. In a nutshell, for all states and federal territories, the period between 13 March and 31 December 2020 can be subdivided into two to four epidemic phases, whereby the majority of the epidemic phase transitions fall in the month of April and last quarter of the year 2020.
Figure 6. The connected dot plot for displaying the coincidence between breakpoint (blue dot) and change-point (red dot). Whenever the breakpoint and change-point occur exactly the same date, only red dot is shown.

4. Conclusion
We have utilised piecewise linear regression and change-point analysis to respectively analyse COVID-19 cumulative and daily infected cases in Malaysia for all its states and federal territories. It is clear from figures 2 and 3 that the piecewise linear regression provides reasonably good fit to cumulative infected cases datasets. These, together with the change-point plots in figure 5, discover structural changes in epidemic trajectories with potentially relevant implication for effectiveness of various stages of MCO implemented in the year of 2020.

For nationwide cumulative confirmed cases, through our piecewise regression, we found evidence that the MCO were likely to almost flatten the epidemic curve in just one month after it was first initiated. While for state level, from table 2, we concluded that almost all states and federal territories had concave downward segment slopes in their first two segments. This indicates that the public health interventions contribute to deceleration of the spread of COVID-19 in Malaysia.

Although there was a promising sign that MCO could curb the spread of COVID-19 in Malaysia until third quarter of 2020, the clustering plot in figure 4 is useful for early alerts to some states (or federal territories) that have potential surge of number of infected cases on the cutoff date (31 December 2020). This short-term forecasting demonstrates its promising performance in capturing rapid changing growth rate of COVID-19 in certain state, for instance, Johor.

For daily infection cases, table 3 revealed that more than half of the change-points detected match its corresponding breakpoints in piecewise regression. It appears that almost all structural changes detected by using both approaches agree and give similar estimates for the timing for two to four abrupt transitions for each state and federal territories, as displayed by connected dot plot in figure 6.

The segmentations of COVID-19 outbreak data enable us to quantify the time window required by containment and mitigation measures like MCO to bring the epidemic growth rates from rapid to moderate and to flattening the cumulative curve. This statistical modelling can be used to obtain the direct effects of public health interventions such as MCO on the course of the epidemic but also it can be utilised in deterministic models in identifying the exact point (date) to conduct calibration for parameters of concerned.

Acknowledgments
The authors would like to thank the Ministry of Higher Education for their funding support of this research project under grant number GL/F08/KPT/02/2020 awarded to J. Labadin.
References

[1] Dallavilla T, Bertelli M, Morresi A, Bushati V, Stuppia L, Beccari T, Chiurazzi P, Marceddu G 2020 Bioinformatic analysis indicates that SARS-CoV-2 is unrelated to known artificial coronaviruses. *Eur Rev Med Pharmacol Sci.* 24(8):4558-64.

[2] Ji W, Wang W, Zhao X, Zai J, Li X 2020 Cross-species transmission of the newly identified coronavirus 2019-nCoV. *Journal of Medical Virology* 92(4):433-40.

[3] World Health Organization (WHO). “WHO Director-General’s opening remarks at the media briefing on COVID-19,” 2020.

[4] World Health Organization (WHO). “WHO Coronavirus Disease (COVID-19) Dashboard,” 2021.

[5] Pung R, Chiew CJ, Young BE, Chin S, Chen MI, Clapham HE, Cook AR, Maurer-Stroh S, Toh MP, Poh C, Low M 2020 Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. *The Lancet.* 395(10229):1039-46.

[6] Shah AU, Safri SN, Thevadas R, Noordin NK, Abd Rahman A, Sekawi Z, Ideris A, Sultan MT 2020. COVID-19 outbreak in Malaysia: Actions taken by the Malaysian government. *International Journal of Infectious Diseases* 97:108-16.

[7] Elengoe A 2020 COVID-19 outbreak in Malaysia. *Osong Public Health and Research Perspectives* 11(3):93.

[8] Tang KH 2020 Movement control as an effective measure against Covid-19 spread in Malaysia: an overview. *Journal of Public Health.* 1-4.

[9] Altahir AA, Mathur N, Thiruchelvam L, Abro GE, Radzi SS, Dass SC, Gill BS, Sebastian P, Zulkifli SA, Asirvadam VS 2020. Modeling the Impact of Lock-down on COVID-19 Spread in Malaysia. *bioRxiv*.

[10] Salim N, Chan WH, Mansor S, Bazin NE, Amaran S, Faudzi AA, Zainal A, Huspi SH, Khoo EJ, Shithil SM 2020. COVID-19 epidemic in Malaysia: Impact of lock-down on infection dynamics. *medRxiv*.

[11] Mohd MH, Sulayman F 2020 Unravelling the myths of R0 in controlling the dynamics of COVID-19 outbreak: A modelling perspective. *Chaos, Solitons & Fractals.* 138:109943.

[12] Arifin WN, Chan WH, Amaran S, Musa KI 2020 A Susceptible-Infected-Removed (SIR) model of COVID-19 epidemic trend in Malaysia under Movement Control Order (MCO) using a data fitting approach. *medRxiv*.

[13] Kamel Ariffin MR, Gopal K, Krishnarajah I, Ilias IS, Adam MB, Sham NM, Arasan J, Rahman NH, Dom NS 2020. Coronavirus Disease 2019 (COVID-19) Infectious trend simulation in Malaysia: a mathematical epidemiologic modelling study. *Available at SSRN 3586699*.

[14] Gill BS, Jayaraj VJ, Singh S, Mohd Ghazali S, Cheong YL, Md Iderus NH, Sundram BM, Aris TB, Mohd Ibrahim H, Hong BH, Labadin J 2020 Modelling the effectiveness of epidemic control measures in preventing the transmission of COVID-19 in Malaysia. *International Journal of Environmental Research and Public Health* 17(15):5509.

[15] Yatim SA, Kim P 2020 COVID-19: The impact of lockdowns, contact tracing, and other control measures on the transmission of COVID-19 in Malaysia. *Letters in Biomathematics*.

[16] Amiruzzaman M, Abdullah-Al-Wadud M, Nor RM, Aziz NA 2021 Evaluation of the effectiveness of movement control order to limit the spread of COVID-19. *Annals of Emerging Technologies in Computing (AETiC)*, Print ISSN. 2516-0281.

[17] Edre MA, ZA MA, Jamalludin AR 2020 Forecasting Malaysia COVID-19 Incidence based on Movement Control Order using ARIMA and Expert Modeler. *IIUM Medical Journal Malaysia* 19(2).

[18] Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, Priesemann V 2020 Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* 369(6500).

[19] Jiang F, Zhao Z, Shao X 2020, Time series analysis of COVID-19 infection curve: A change-point perspective. *Journal of Econometrics*.
[20] Auger IE, Lawrence CE 1989 Algorithms for the optimal identification of segment neighborhoods. *Bulletin of Mathematical Biology* 51(1):39-54.

[21] Killick R, Eckley I 2014 changepoint: An R package for changepoint analysis. *Journal of Statistical Software* 58(3):1-9.

[22] Saidan MN, Shbool MA, Arabeyyat OS, Al-Shihabi ST, Al Abdallat Y, Barghash MA, Saidan H 2020 Estimation of the probable outbreak size of novel coronavirus (COVID-19) in social gathering events and industrial activities. *International Journal of Infectious Diseases* 98:321-7.

[23] Ahmad NA, Chong ZL, Abd Rahman S, Ghazali MH, Nadzari EE, Zakiman Z, Redzuan S, Md Taib S, Kassim MS, Wan Mohamed Noor WN 2020 First local transmission cluster of COVID-19 in Malaysia: public health response. *International Journal of Travel Medicine and Global Health* 8(3):124-30.

[24] Verma MK, Asad A, Chatterjee S 2020 COVID-19 pandemic: power law spread and flattening of the curve. *Transactions of the Indian National Academy of Engineering* 103-8.

[25] S. Birruntha, “Covid-19: Johor records highest new cases at 1,069 | The Malaysia Reserve,” 2021. https://themalaysianreserve.com/2021/01/27/covid-19-johor-records-highest-new-cases-at-1-069/ (accessed Feb, 9, 2021).

[26] Reddy T, Shkedy Z, van Rensburg CJ, Mwambi H, Debbas P, Zuma K, Manda S 2021 Short-term real-time prediction of total number of reported COVID-19 cases and deaths in South Africa: a data driven approach. *BMC Medical Research Methodology* 21(1):1-1.

[27] M. Ten, “Covid-19: Sarawak records highest single day tally of 153 cases | The Borneo Post,” 2021. https://theborneopost.com/2021/01/11/covid-19-sarawak-records-highest-single-day-tally-of-153-cases/ (accessed Feb, 9, 2021).