Pattern Analysis of Corona Virus Disease (COVID-19) -Outbreak in Malaysia

Article · January 2020

DOI: 10.5373/JARDCS/V12I2/S20201380

Nur Syarafina Mohamed
University of Kuala Lumpur

Pattern Analysis of Corona Virus Disease (COVID-19) - Outbreak in Malaysia

Nur Syarafina Mohamed, Malaysian Institute of Industrial Technology, Universiti Kuala Lumpur, Pasir Gudang, Johor, Malaysia,

Siti Nor Zawani Ahmmad, Malaysian Institute of Industrial Technology, Universiti Kuala Lumpur, Pasir Gudang, Johor, Malaysia,

Nor Afiqah-Aleng, Institute of Marine Biotechnology, Universiti Malaysia Terengganu (UMT), 21030 Kuala Nerus, Terengganu, Malaysia,

Hafizah Farhah Saipan Saipol, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, 54100 Kuala Lumpur,

Shazlyn Milleana Shaharuddin, Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, Malaysia,

Abstract---The ongoing Corona Virus Disease (COVID-19) outbreak is now declared as the pandemic by World Health Organization (WHO). This disease began in Wuhan, China in late 2019 and is widely spread now all over the world. Progressively, Malaysia has been the leading country in Southeast Asia for this outbreak with cases more than 2000 as on 26thMarch 2020. This article highlights the analysis of the outbreak pattern which follows the exponential growth regression line. Data is collected daily for 66 days starting from the 1st case defined on 25th January 2020. Regression line is used because it can describe the relationship between predictors and the outcome within the datasets that can be used for prediction purposes. Fitting the real data to the graph, an equation which follows the exponential growth model is obtained. The calculation of the relative error between the exact and the approximate data shows that the pattern follows the exponential growth model as it is compared with the quadratic regression line. This analysis can be particularly beneficial for the health authorities in preparing immediate and effective strategies to flatten the curve. Malaysia government *is currently working hard in flattening the curve by implementing Restricted Movement Order (RMO).*

Index Terms: COVID-19, Pandemic, Mathematical Modelling, Regression, Trend Line

I. Introduction

The Corona Virus Disease-(COVID-19) is now declared as a pandemic by World Health Organization (WHO). It is a disease caused by the infection of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), which is previously recognized as 2019 novel coronavirus (2019-nCoV) [1]. There are 465,915 cases announced as the confirmed cases followed by 21,031 of confirmed deaths all around the world, which consists of 200 countries, including Malaysia [1]. As of now, new cases are recorded daily and accumulated with the total cases observation. According to the recorded data, COVID-19 poses a huge threat to the global public health and economics, compared to the number of deaths associated with other types of coronaviruses, such as, Severe Acute Respiratory Syndrome Corona virus (SARS-CoV) and Middle East Respiratory Syndrome Corona virus (MERS-CoV [2,3].Several patterns emerge to examine the spread of the COVID-19. [3,4] investigated the outbreak pattern based on the effect of the temperature and specific humidity. [5] developed Artificial Intelligence-based model to predict the new and cumulative confirmed cases of COVID-19 in provinces across China. Meanwhile, [6] observed the trend analysis with two classes of growth; sub-exponential mainly in China and exponential growth for the remaining countries. The findings of previous researchers may serve as examples for other countries to handle and improve strategies to reduce and monitor the spread of pandemics. Nonetheless, different model should be applied depending on the country being studied, because they follow different growth curves (Weber et al, 2020). Thus, the aim of this paper is to develop and analyse the outbreak pattern in Malaysia. From the statistics given by WHO, Malaysia seems to be the leading country in Southeast Asia. To date, Malaysia has 2013 confirmed cases and 23 deaths come together with 215 cases of total recovered. The data in Table 1 shows the observed numbers for the last 64 days, [7]. This paper studies on the pattern analysis shown by the outbreak in Malaysia by using the regression analysis. Famously known with the idea of relative absolute error in (1),

DOI: 10.5373/JARDCS/V12I2/S20201380
**Corresponding Author Nur Syarafina Mohamed, Email: nursyarafina@unikl.edu.my*
Article History: Received: Mar 23, 2020, Accepted: June 26, 2020

$$\text{Relative Error} = \left| \frac{\text{Exact Value-Approximate Value}}{\text{Exact Value}} \right| \quad (1)$$

regression analysis is one of the prediction tools used in forecasting field. The idea is to fit the exact data with a suitable function used in modelling the equation. Data fitting is the process of fitting the data in finding the appropriate function by fitting the models to data and analysing the accuracy of the fit [8-10]. For this study, the Least Square method is used. The idea is to convert the linear Least Square to the exponential function and compare it with the quadratic least square.

Table 1. COVID-19 cases in Malaysia [2]

Day	***Total Cases***	***New Cases***	***Day***	***Total Cases***	***New Cases***
1	0	0	34	22	0
2	0	0	35	22	0
3	0	0	36	22	0
4	4	4	37	23	1
5	4	0	38	25	2
6	4	0	39	25	0
7	7	3	40	29	4
8	7	0	41	29	0
9	8	1	42	36	7
10	8	0	43	50	14
11	8	0	44	55	5
12	8	0	45	83	28
13	8	0	46	93	10
14	10	2	47	99	6
15	12	2	48	117	18
16	14	2	49	129	12
17	15	1	50	149	20
18	16	1	51	158	9
19	17	1	52	197	39
20	18	1	53	238	41
21	18	0	54	428	190
22	18	0	55	553	125
23	19	1	56	673	120
24	19	0	57	790	117
25	21	2	58	900	110
26	22	1	59	1030	130
27	22	0	60	1183	153
28	22	0	61	1306	123
29	22	0	62	1518	212
30	22	0	63	1624	106
31	22	0	64	1796	172
32	22	0	65	2031	235
33	22	0	66	2161	130

Details explanation is discussed in the next sections where section two will elaborate more on the derivation process followed by the results, discussions and conclusion.

II. Methodology

The Least Squares method is chosen because of its data fitting procedure to find a function, where fitting the models to data and analyzing the accuracy of the fit. For linear fitting, the Least Squares method is relatively simple, where it needs to employ matrix formation. The Least Squares method is also useful for comparisons between multiple regression models [11-14]. The classical regression model is defined by

$$y = h\left(x_1, x_2, ..., x_p + \varepsilon\right)$$

where y is the response variable, x_i is the predictor variable with $i = 1, 2, ..., p, p > 0$ is an integer constant and ε is the error term. The function $h\left(x_1, x_2, ..., x_p\right)$ describes the type of relationship between y and $x = \left(x_1, x_2, ..., x_p\right)$. The simplest form of regression is defined by the following linear regression model

$$y = a_0 + a_1 x_1 + a_2 x_2 + ... + a_p x_p + \varepsilon$$

where $a_0, a_1, ..., a_p$ are the regression parameter. The parameters $a = \left(a_0, a_1, ..., a_p\right)$ can be acquired from a multivariate process of calculus [15-16].

Derivation Process

The least squares method involves determining the best approximating models; linear and quadratic, by comparing the total least squares error [13]. Consider a set of data, $\left(x_i, y_i\right)$ where x is said to be exact if only y values have errors. The error is defined as

$$E_i = (a_0 + a_1 x_i) - y_i .$$

The strategy for fitting the "best" line through the data would be to minimize the sum of the residual error squares for all the available data given by

$$\min \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} ((a_0 + a_1 x) - y_i)^2 \qquad (2)$$

$$\min \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} ((a_0 + a_1 x + a_2 x^2) - y_i)^2 \qquad (3)$$

Differentiate (2) and (3) and solve them simultaneously,

$$\frac{\partial}{\partial a_0} \sum_{i=1}^{n} (a_0 + a_1 x_i - y_i)^2 = 0$$

$$\sum_{i=1}^{n} 2(\ (a_0 + a_1 x_i - y_i) \ = 0$$

$$2a_0 n + 2a_1 \sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} y_i = 0$$

$$a_0 n + a_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\frac{\partial}{\partial a_1} \left[\sum_{i=1}^{n} (a_0 + a_1 x_i - y_i)^2 \right] = 0$$

$$\sum_{i=1}^{n} 2(\ (a_0 + a_1 x_i - y_i) \ = 0$$

$$2a_0 \sum_{i=1}^{n} x_i + 2a_1 \sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} x_i y_i = 0$$

$$a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i .$$

DOI: 10.5373/JARDCS/V12I2/S20201380
*Corresponding Author Nur Syarafina Mohamed, Email: nursyarafina@unikl.edu.my
Article History: Received: Mar 23, 2020, Accepted: June 26, 2020

Then, $\sum_{i=1}^{n} a_0 + a_1 \sum_{i=-=1}^{n} x_i = \sum_{i=1}^{n} y_i \quad a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$.

Then, the general formula to find the parameters of a_0 and a_1 for a linear model can be described as

$$\begin{bmatrix} n & \sum_{i=1}^{n} x \\ \sum_{i=1}^{n} x & \sum_{i=1}^{n} x^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix} \tag{4}$$

Whereas, by using the same step as above for (3), a general formula to find the parameters of a_0, a_1 and a_2 for a quadratic model can be described as

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i^2 y_i \end{bmatrix}$$

Therefore, the general formula of finding the Least Squares models can be described

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \ldots & \sum_{i=1}^{n} x_i^m \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 \ldots & \sum_{i=1}^{n} x_i^{m+1} \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^4 & \sum_{i=1}^{n} x_i^{m+2} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{n} x_i^m & \sum_{i=1}^{n} x_i^{m+1} & \sum_{i=1}^{n} x_i^{m+2} \ldots & \sum_{i=1}^{n} x_i^{2m} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \ldots \\ a_m \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i^2 y_i \\ \vdots \\ \sum_{i=1}^{n} x_i^m y_i \end{bmatrix} \tag{5}$$

The values for $a_0, a_1, a_2 \ldots a_m$ can be obtained by solving the matrix system above by using any available methods as long as the system has unique solution. In this study, the focus is on linear and quadratic Least Squares models which are

$$f(x) = a_0 + a_1 x \tag{6}$$

and

$$f(x) = a_0 + a_1 x + a_2 x^2 \tag{7}$$

respectively where the value of a_0, a_1 and a_2 are obtained from (4) and (5) separately. The following algorithm explains how this method works.

Exponential Growth Model

With the current outbreak of the COVID-19, the Exponential Growth model is one of the functions used to identify the pattern of the infected people [17-21]. From (6), the derivation of the exponential growth model can be written as;

$$f(x) = e^{ax} \tag{8}$$

We neglect the value of a_0 as the first case is approximately 0.

Algorithm 5.1: Least Squares Method

Step 1: Identify formula from (5) and (8) for both quadratic and exponential
Step 2: Identify variables and data summation.
Step 3: Calculation of a_0, a_1 and a_2

Step 4: Generate the approximate functions for both Exponential and quadratic
Step 5: Calculate error by using (1).
Step 6: Model estimation

III. Results and Discussions

Table 2 shows the relative error calculated based on the difference between the approximate data of quadratic and exponential trend line with the exact data of COVID-19 in Malaysia. This analysis is important to study the trend and pattern of the cases occurred as it was tabulated randomly as per daily cases, as shown in Figure 1. The exact data is obtained from Table 1, while the approximate data is calculated based on the trendline obtained from the regression analysis of the graph i.e. Quadratic Trend Line (QTL) and Exponential Growth Trend Line (EGTL) as shown in Figure 2. Data from Table 1 are collected daily for 64 days, starting from the 1st case on 25thJanuary 2020. The x-variable denotes the day progress and the y-variable denotes by the cases. For the data fitting purpose, the data are collected up to64th day. The rest days are used for approximation purpose where we want to calculate its relative errors. The approximate data is calculated, and the error is evaluated using the equation in (1).

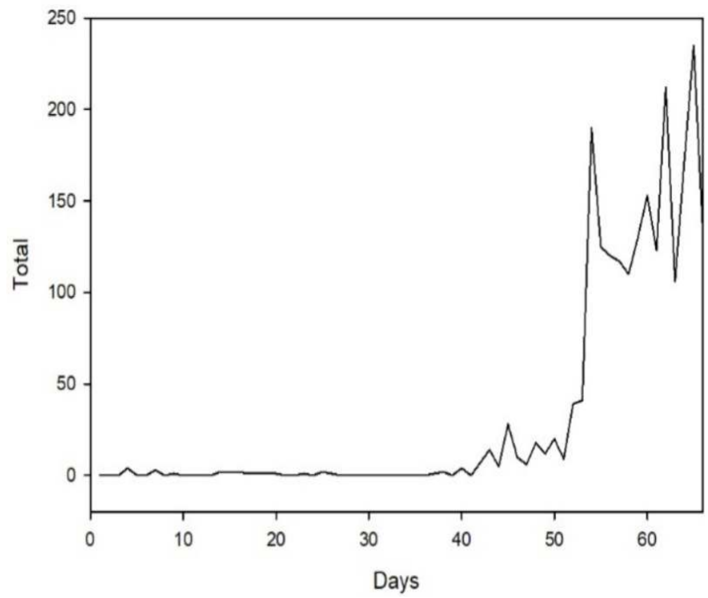

Figure 1. Daily Cases of COVID-19 in Malaysia

Table 2. The Relative Error between the Exact Data and Approximate Data using Quadratic and Exponential Trend Line

x	REQTL	REEGTL	x	REQTL	REEGTL
1	NaN	NaN	34	49.7484	1.4031
2	NaN	NaN	35	51.6189	1.7006
3	NaN	NaN	36	53.4893	2.0348
4	39.5143	0.6013	37	52.9964	2.2622
5	29.2266	0.5519	38	50.4827	2.3727
6	18.9389	0.4965	39	52.1287	2.7902
7	4.5150	0.6766	40	46.4955	2.6719
8	1.3637	0.6366	41	47.9144	3.1264
9	6.4621	0.6427	42	39.9353	2.7355
10	11.6059	0.5985	43	29.8564	2.0225
11	16.7497	0.5488	44	27.9813	2.0878
12	21.8936	0.4929	45	19.3750	1.2994
13	27.0374	0.4301	46	17.8416	1.3062
14	25.9450	0.4877	47	17.2366	1.4346
15	25.2167	0.5202	48	15.0904	1.3150

DOI: 10.5373/JARDCS/V12I2/S20201380
*Corresponding Author Nur Syarafina Mohamed, Email: nursyarafina@unikl.edu.my
Article History: Received: Mar 23, 2020, Accepted: June 26, 2020

16	24.6965	0.5378	49	14.0986	1.3596
17	25.8601	0.5153	50	12.6166	1.2957
18	26.8783	0.4893	51	12.2154	1.4330
19	27.7767	0.4599	52	10.2039	1.1928
20	28.5752	0.4267	53	8.7913	1.0398
21	30.8614	0.3558	54	5.4287	0.2747
22	33.1475	0.2760	55	4.5020	0.1086
23	33.6214	0.2292	56	3.9388	0.0237
24	35.7872	0.1338	57	3.5556	0.0199
25	34.4337	0.1193	58	3.2890	0.0332
26	34.7845	0.0553	59	3.0400	0.0507
27	36.6550	0.0617	60	2.8110	0.0712
28	38.5254	0.1931	61	2.6719	0.0545
29	40.3959	0.3408	62	2.4655	0.0858
30	42.2664	0.5068	63	2.3952	0.0397
31	44.1369	0.6933	64	2.2845	0.0242
32	46.0074	0.9029	65	2.1561	0.0303
33	47.8779	1.1384	66	2.1056	0.0242

From the analysis, the equations obtained from the data fitting process are,

$$f(x) = 326.6598 - 42.0566x + 0.9059x^2 \quad (9)$$

$$f(x) = e^{0.1167x} \quad (10)$$

where $f(x)$ is the total cases recorded daily and x is the n^{th} day of cases recorded. For example, we can predict the possible total cases in day 66, as follows

$$f(x) = e^{0.1167(66)} = 2213$$

Thus, from exact data in Table 1, the exact data showed on day 66 is 2161 where the relative error calculated is 0.0242. The errors can be seen from Table 2 for which REQTL stands for the relative error for QTL and REEGTL is for the relative error for EGTL. Thus, with this small error, this model can be used as the pattern for the cases growth in Malaysia. From Figure 2, the red line shows the exact data plotted where the black and blue lines are for EGTL and QTL respectively. Figure 2 shows the exact data line follow the EGTL as it shows the smaller error compared to QTL.

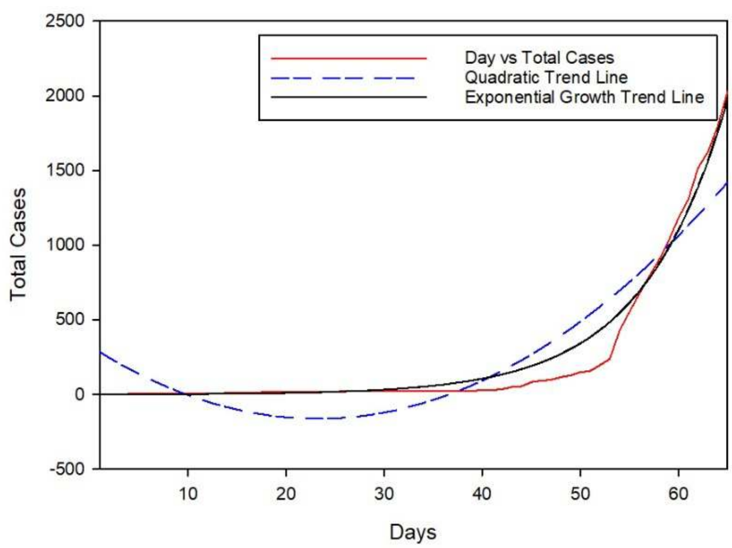

Figure 2. Daily Total with Fitted Trend Line of QTL and EQTL

IV. Conclusion

In conclusion, this paper discussed the pattern follows the data for COVID-19 cases in Malaysia. It was found that the pattern follows the exponential growth regression line model which can be used to predict the pattern of outbreak cases growth in Malaysia. The use of this model particularly beneficial for the health authorities to prepare immediate and effective strategies to flatten the curve. This model also enables the health authorities for a better understands regarding the outbreak pattern. To date, Malaysia starts the Restricted Movement Order (RMO) in order to flatten the curve. The limitation of this paper is it only covers the outbreak pattern without considering the effect after the RMO. This pattern shows a small error between the real and predicted data. This paper only covers the pattern without considering the outcome after the order. Even the graph is yet flattening and predicted to hit the third wave of infection, but the daily cases show slower progress.

V. Acknowledgment

The author appreciates all the comments from the reviewers and readers of this article.

References

[1] E. Gorbalenya, S. C. Baker, R. S. Baric et al., "The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2," Nature Microbiology 5 (2020) 536–544.

[2] WHO https://www.who.int/emergencies/diseases/novel-coronavirus-2019 27th march 2020

[3] Bukhari, Q. and Jameel, Y. (2020) Will Coronavirus Pandemic Diminish by Summer? SSRN, doi http://dx.doi.org/10.2139/ssrn.3556998.

[4] Sajadi, M. M., Habibzadeh, P., Vintzileos, A., Shokouhi, S., Miralles-Wilhelm, F. and Amoroso, A. (2020) Temperature, Humidity and Latitude Analysis to Predict Potential Spread and Seasonality for COVID-19. SSRN. Doi http://dx.doi.org/10.2139/ssrn.3550308.

[5] Hu, Z., Ge, Q., Li, S., Jin, L. and Xiong, M. (2020) Artificial Intelligence Forecasting of Covid-19 in China. arXiv.

[6] Weber, A., Ianelli, F. and Goncalves, S. (2020) Trend analysis of the COVID-19 pandemic in China and the rest of the world, arXiv.

[7] W. A. Bogoch, A.Thomas-Bachli, C. Huber et al., "Pneumonia of unknown etiology in Wuhan, China: potential for international spread via commercial air travel," Journal of Travel Medicine 27 (2020)

[8] J. T. Wu, K. Leung, G. M. Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study," Lancet 395 (2020) 689-697.

[9] U. Moyi, W. J. Leong, I. Saidu, "On the application of threeterm conjugate gradient method in regression analysis," *International Journal of Computer Applications 102, (2014) 1–4.*

[10] S. C. Chapra, R. P. Canale, *Numerical methods for engineers*, McGraw-Hill, 2014.

[11] N. S. Mohamed, M. mamat, M. Rivaie et al., "Estimating the unemployment rate using least square and conjugate gradient methods," *International Journal of Engineering & Technology, 7 (2018) 94-97.*

[12] J. S. Armstrong, R. Fildes, "Correspondence on the selection of error measures for comparisons amongforecasting methods," Journal of Forecasting*, 14 (1995) 67-71.*

[13] S. Makridakis, S. Anderson, R. Carbone et al., *The Forecasting Accuracy of Major Times Series Methods*, New York: John Wiley & Sons, 1984.

[14] S. Narula, J. Wellington, (1977). "An algorithm for the minimum sum of weighted absolute errors regression," *Communication in Statistics 6B (1977) 341-352.*

[15] T. M. Khoshgoftaar, B. B. Bhattacharyya, G. D. Richardson, G. D., "Predicting software errors during development using nonlinearregression models: a comparative study," *IEEE Transactionson Reliability 41 (1992) 390–395.*

[16] R. L. Burden, J. D. Faires, Numerical Analysis: Interpolation and Polynomial Approximation, Brooklyn: Cengage Learning, 2011.

[17] G. Yuan, Z. Wei, "New line search methods for unconstrained optimization," *Journal of the Korean Statistical Society 38 (2009) 29-39.*

[18] Q. Zhao, Y. Chen, D.S. Small, "Analysis of the epidemic growth of the early 2019-nCoV outbreak using internationally confirmed cases,". medRxiv, 2020.

[19] H. Xiong, H. Yan, "Simulating the infected population and spread trend of 2019-nCov under different policy by EIR model," unpublished.

DOI: 10.5373/JARDCS/V12I2/S20201380
Corresponding Author Nur Syarafina Mohamed, Email: nursyarafina@unikl.edu.my
Article History: Received: Mar 23, 2020, Accepted: June 26, 2020

[20] K. Mizumoto, K. Kagaya, G. Chowell, "Early epidemiological assessment of the transmission potential and virulence of 2019 Novel Coronavirus in Wuhan City: China, 2019-2020," unpublished.

[21] S. Zhao, Q. Lin, J. Ran et al., "Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak," *International Journal of Infectious Diseases 92 (2020) 214-217.*

Authors

First & Correspondence Author – Nur Syarafina Mohamed, PhD in Mathematics (Optimization), Malaysian Institute of Industrial Technology, Universiti Kuala Lumpur, Pasir Gudang, Johor, Malaysia, nursyarafina@unikl.edu.my.

Second Author – Siti Nor Zawani Ahmmad, PhD in Electrical Engineering, Malaysian Institute of Industrial Technology, Universiti Kuala Lumpur, Pasir Gudang, Johor, Malaysia, sitinorzawani@unikl.edu.my.

Third Author – Nor Afiqah-Aleng, PhD in Systems Biology, Institute of Marine Biotechnology, Universiti **Fourth Author** – Hafizah Farhah Saipan Saipol, PhD in Numerical Analysis and Parallel Computing System, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, hafizah.farhah@utm.my.

Fifth Author – Shazlyn Milleana Shaharuddin, PhD in Statistics, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, shazlyn@fmst.upsi.edu.my.