CONONAVIRUS SPATIAL BIG DATA ANALYSIS FOR DEVELOPING SPECTRUM MODELS OF DIFFERENT STAGES OF THE OUTBREAK TO PREDICT THE TRENDS OF THE SPECTRUM

Arun Kumar Verma (1), Anjul Verma (2), Ranbir Nandan (3), Aditi Verma (4)
[1]Defence Research Development Organisation, Ministry of Defence, New Delhi, India
[2]Business School, University of Liverpool, Liverpool L69 3BX, United Kingdom
[3]Geology Department, BN College, Patna University, Patna, India
[4]Qualcomm India Pvt Ltd, Hyderabad-500081, India
Email: arun@vidyadaan.org; amrevlujna@gmail.com ; nandanrranbir@gmail.com; aditi.verma.004@gmail.com

KEYWORDS: Population mortality, Severity of the outbreak, Knowledge classifier, Variability factor, Southeast Asian region.

ABSTRACT: The outbreak of the novel coronavirus disease (covid-19) emerged from Wuhan, Hubei province of China, spread exponentially in geo-spatiotemporal way to the new geographical locations covering more than 210 countries, which caused more than 25.416 million people of the global population infected and 0.851 million deaths (as on 30 August 2020). The global spatial spreading of spectrum of the outbreak due to large-scale migrations has seriously threatened the human health and life of the people by attacking societies at their core as global health crisis and the human crisis. This posed the challenges to control the severity of the coronavirus spectrum by creating a unique health response system to suppress further spreading of the outbreak as measures. In the southeast Asian region, the first case of coronavirus was reported in Thailand on 13 January 2020, which was followed by South Korea on 20 January 2020, and Vietnam and Taiwan on 22 January 2020 prior to reach Hong Kong and Singapore on 23 January 2020. Malaysia reported the first coronavirus case on 25 January 2020, which further spread to Philippines on 30 January 2020 prior to reach Indian Sub-continent on 31 January 2020. During this period, coronavirus spread geo-spatially in more than 100 countries situated at latitudes between 64˚N and 35˚S. National lockdowns were imposed by the respective governments of these countries as measures to control the severity of the spectrum of the outbreak in addition to medical response system such as tracing, testing and treatment. This stressed the role of geo-spatial big data analysis for understanding the spectrum of the outbreak and its effect on population mortality, which determines the severity of the outbreak. Knowledge classifiers as coronavirus predictive spectrum models of 31 countries such as Finland, Denmark, Ireland, Poland, Czech Republic, Romania, Israel, Egypt, Portugal, Brazil, Saudi Arabia, Russia, Serbia, Mexico, Panama, Columbia, Peru, Argentina, Japan, Peru, Hong Kong, Vietnam, South Korea, Singapore, Malaysia, Philippines, Thailand, Taiwan, India, Indonesia and Myanmar, that lie below the latitude 64°N, based on 5 days moving average new coronavirus cases from 15 February to 15 May 2020, describes six different stages of spectrum such as complete recoverable stage, recoverable stage, safe stage, stabilizing stage, critical stage and beyond the critical stage of the outbreak. In this paper, polynomial models of knowledge classifiers have been used for developing models of six different stages of the spectrum of the outbreak, for predicting the trends of the spectrum of any geo-spatial locations. Further, predictive spectrum models of six different stages of the outbreak have been used predict the trends to new geo-spatial locations of Indian subcontinent. In this paper, the impact of the latitude on population mortality and its variability have been described for determining the factor for the severity, based on population mortality data from 15 April to 30 August 2020 for 31 countries situated at latitudes between 64°N and 35°. This shows that the variability factor of population mortality are determining factor for the severity of the coronavirus outbreak, whereas lower population mortality observed for countries situated at latitudes between 35°N and 35°S supporting Vitamin-D.

1. INTRODUCTION

The outbreak of the 2019 novel coronavirus disease (Covid-19) spread geo-spatially in more than 210 countries of the globe causing more than 25.416 million people of the global population infected and 0.851 million deaths (as on 30 August 2020). The exponential increase in spreading of corona virus spectrum in spatiotemporal way to the new geographical locations has seriously threatened the human health and life of the people as well as posed the challenges for countries to control the severity of the outbreak (Coronavirus,2020). The spatial spreading of coronavirus spectrum due to large-scale migration from Hubei province of China caused the outbreak in the southeast Asian region covering the latitude between 38°N to 6°S. The first case of coronavirus was reported in Thailand on 13 January 2020, which was followed by South Korea on 20 January 2020, and Vietnam and Taiwan on 22 January 2020 prior to reach Hong Kong and Singapore on 23 January 2020. Malaysia reported the first coronavirus case on 25 January 2020, which further geo-spatially spread to Philippines on 30 January 2020 prior to reach Indian Sub-continent on 31 January 2020. National lockdowns were imposed by the respective governments of the southeast Asian countries as measures to control the severity of the spectrum of the outbreak (BBC,2020). The Hong Kong, Vietnamese and South Korean governments imposed national lockdowns as measures to control the exponential rise of the spectrum of coronavirus from 8, 13 and 20 February 2020 respectively, after 16, 22 and 31 days of the first reported coronavirus case. The governments of Singapore, Malaysia, Philippines, Thailand, Taiwan and India imposed these measures from 6, 13, 15, 20, 24 and 25 March 2020 respectively, whereas, the Indonesian government imposed a national lockdown from 15 March 2020. The Myanmarese government executed a national lockdown on 13 March 2020, prior to the arrival of first coronavirus case on 27 March 2020 (BBC, 2020).

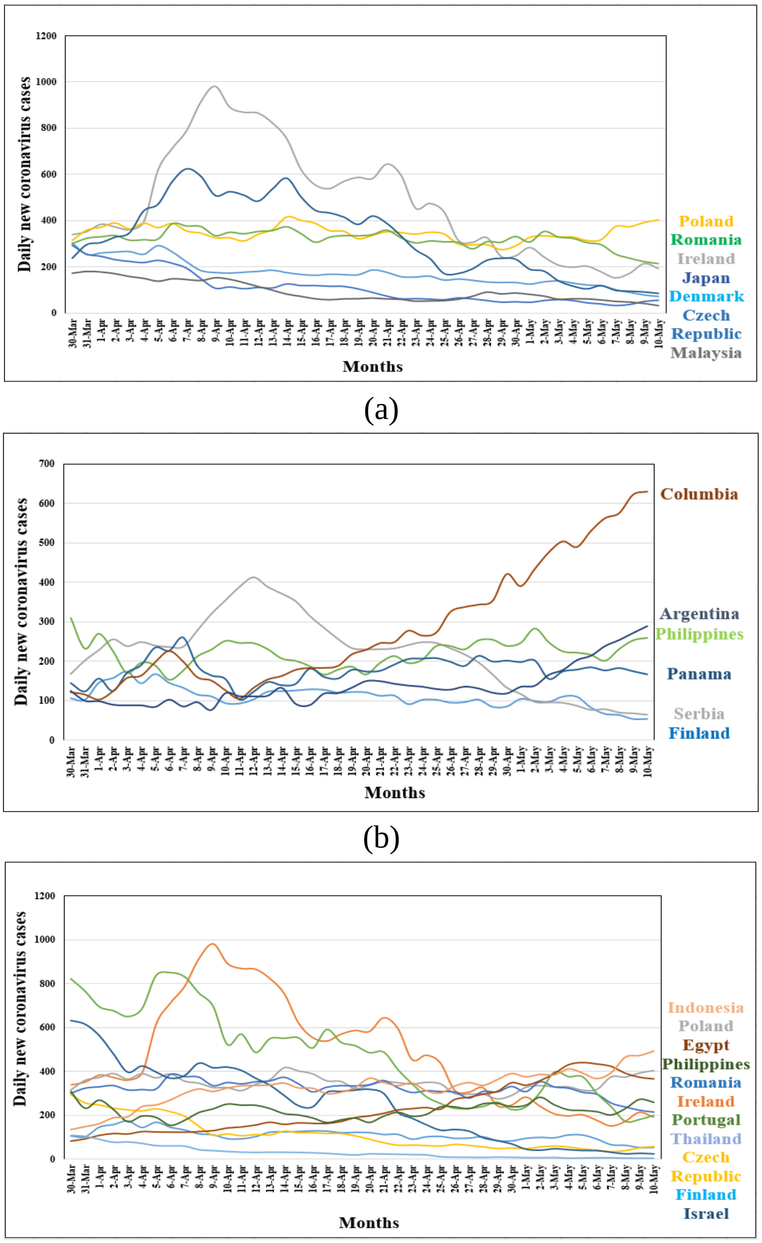
There are marked variations in the spectrum of daily new cases of coronavirus between different countries situated at different latitudes below 64°N of the hemisphere (Coronavirus,2020). Spatial big data predictive analysis of daily new coronavirus cases carried out for 28 countries that lie below the latitudes of 60°N including the southeast Asian region, resulted into the development of predictive spectrum models for six different stages of the spectrum of the outbreak such as complete recoverable stage, recoverable stage, safe stage, stabilising stage, critical stage and beyond the critical stage as knowledge classifier (Verma et al., 2020). These predictive spectrum models as knowledge classifier can be used to predict the trends of the spectrum of coronavirus for new geographic locations. Higher population mortality from coronavirus observed in northern latitude exhibiting the population mortality with decreasing north south gradient based on mortality data of 02 April 2020 (Panarese and Shahini, 2020). In Rhodes et al. (2020), population mortality from coronavirus between different countries situated at latitudes below 64°N showed marked variations with relatively low population mortality at latitudes below 35°N based on mortality data of 15 April 2020. Spatial big data analysis on population mortality carried out for 28 countries based on mortality data from 15 April to 08 June 2020 supports lower population mortality for countries situated at latitudes between 35°N and 35°S (Verma et al.,2020).

In this paper, polynomial models of knowledge classifiers have been used for developing models of six different stages of the spectrum of the outbreak, for predicting the trends of the spectrum of any geo-spatial locations. Further, predictive spectrum models of six different stages of the outbreak have been used predict the trends to new geo-spatial locations of Indian subcontinent. In this paper, the impact of the latitude on population mortality and its variability have been described based on population mortality data from 15 April to 30 August 2020 for 31 countries situated at latitudes between 64°N and 35°S.

2. SPATIAL BIG DATA ANALYSIS FOR KNOWLEDGE CLASSIFIER

The spectrum of 5-day moving average of daily new coronavirus cases for different countries acts as knowledge classifier for predicting the trend of the spectrum of coronavirus for any geographic locations due to the expected similarity in the spectrum. Figure 1 (a) depicts the characteristics of the spectrum of coronavirus for Ireland and Japan. The spectrum increased exponentially and attained the peak at approximately 1,000 and 600 cases, prior to decreases following a log-normal spectrum distribution. Similar characteristics of the spectrum are observed for Malaysia, Denmark and Czech Republic, attaining peak value between 200 and 300 cases. Figure 1(b) depicts the spectrum of coronavirus cases for Serbia and Finland, attaining peak values of 400 and 200 cases respectively, prior to decrease following a log-normal spectrum distribution. Figure 1(c) depicts similar spectrum distribution model for Ireland, Portugal and Israel after attaining maximum value of daily new cases of approximately

1,000, 800 and 600 respectively, whereas Finland, Czech Republic and Thailand followed a similar spectrum of coronavirus after attaining peak values ranging from 100 to 300 cases. These spectrum distributions are known as knowledge classifier.



(a)

(b)

(c)

Figure 1: Coronavirus spectrum distribution models as a knowledge classifier.

3. BIGDATA PREDICTION MODELS BASED ON KNOWLEDGE CLASSIFIER

The empirical models developed based on the coronavirus spectrum to successfully control the outbreak and recover from the first outbreak can be used for any other geographical locations for predicting the recovery from the expected spectrum by extrapolating the models developed by polynomial regression techniques. The envelope of the spectrum of empirical models shall vary with the different phases, such as Beyond the Critical Stage (Phase5), Critical Stage

(Phase4), Stabilising Stage (Phase3), Safe Stage (Phase2), Recoverable Stage (Phase1), and Complete Recovery Stage (Phase0) of controlling the outbreak. The polynomial spectrum distribution models developed for predicting the trend of the spectrum and predicting recovery from the outbreak, based on 5-day moving average of coronavirus data for different countries with varying maximum coronavirus spectrum and its characteristics for different stages of the spread of the outbreak is expressed as:

$$y(x) = A_0 + A_1 x + A_2 x^2 + A_3 x^3 + A_4 x^4 \quad (1)$$

where x is the number of days, $y(x)$ is daily new coronavirus cases, and A_0, A_1 , A_2 , A_3 and $A_4$are the coefficients for different stages of the outbreak for different countries. Table 1 describes the coefficients of different models developed for different countries in the present work, which are Beyond the Critical Stage (Phase5), Critical Stage (Phase4), Stabilising Stage (Phase3), Safe Stage (Phase2), Recoverable Stage (Phase1), and Complete Recovery Stage (Phase0)for controlling the outbreak.

Table 1: Coefficients of spatial big data predictive models.

Stages	Country	Model Coefficients					Max Spectrum
		A_0	A_1	A_2	A_3	A_4	
Complete Recovery	Serbia	61.04	-22.28	2.6674	-0.0714	0.0006	412
	Japan	124.00	-45.49	5.1747	-0.1425	0.0011	623
Recovery	Vietnam	2.1623	1.8872	-0.131	0.0029	-0.00002	12
	Taiwan	7.4031	2.5238	-0.198	0.0047	-0.00004	22
	Hong Kong	-10.05	10.544	-0.654	0.0137	-0.00009	54
	South Korea	74.429	8.9373	-0.681	0.0138	-0.00009	115
Safe Stage	Ireland	219.82	-59.82	6.7487	-0.1845	0.0015	980
	Singapore	222.56	-64.11	4.414	-0788	0.0004	1097
	Israel	-140.1	80.599	-3.144	0.0405	-0.0002	648
	Portugal	-151.7	95.967	-3.079	0.0258	0.00001	851
	Czech Republic	-23.73	42.837	-2.135	0.0367	-0.0002	295
	Denmark	-34.29	30.989	-1.164	0.0159	-0.00008	302
	Malaysia	72.181	17.17	-0.935	0.0167	-0.0001	180
	Thailand	-19.55	25.68	-1.533	0.0314	-0.0002	127
Stabilising Stage	Finland	22.512	7.193	-0.130	-0.0000	0.000007	174
	Panama	-21.14	14.61	-0.440	0.0069	-0.00005	260
	Romania	-43.18	22.48	-0.205	-0.0055	0.00007	387
	Argentina	-1.534	7.482	-0.056	-0.0047	0.00008	>290
	Philippines	-82.06	34.33	-1.345	0.0214	-0.0001	>320
	Poland	15.264	6.805	1.0701	-0.0417	0.0004	>417

	Indonesia	68.763	-11.48	1.594	-0.0425	0.0004	>493
Critical Stage	Egypt	-0.134	7.4195	-0.367	0.0118	-0.0001	>440
	Colombia	-18.29	11.646	-0.335	0.0041	0.00002	>630
	Mexico	-30.37	19.207	-1.370	0.055	-0.0005	>1825
	Saudi Arabia	102.57	-13.13	0.5756	0.0212	-0.0003	>2025
Beyond the Critical Stage	Peru	-74.90	34.724	-3.055	0.1405	-0.0014	>3525
	India	-80.4	23.035	-0.297	0.0222	-0.00007	>3835
	Brazil	-240.5	121.11	-6.981	0.2224	-0.0015	>11100
	Russia	-131.6	69.939	-8.056	0.4085	-0.004	>11100

4. SPECTRUM MODELS OF DIFFERENT STAGES OF THE OUTBREAK

The empirical spectrum models for predicting the trends of different stages of the outbreak have been for developed based on coefficients of predictive models of different countries given in Table-1 as knowledge classifier, such as Beyond the Critical Stage (Phase5), Critical Stage (Phase4), Stabilising Stage (Phase3), Safe Stage (Phase2), Recoverable Stage (Phase1), and Complete Recovery Stage (Phase0). The polynomial spectrum models developed for predicting the trend of the spectrum for different stages of the outbreak is expressed as:

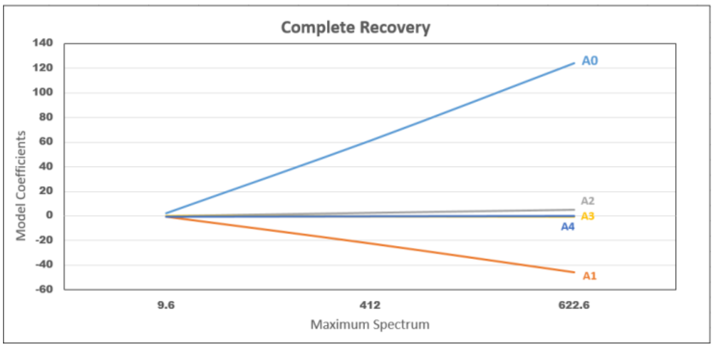
$$y(x) = A_0 + A_1 x + A_2 x^2 + A_3 x^3 + A_4 x^4 \quad (2)$$

where x is the number of days, $y(x)$ is daily new coronavirus cases, and A_0, A_1, A_2 , A_3 and A_4 are the coefficients for different stages of the outbreak and given in Table 2.

Table 2 : Coefficients of different stages of the outbreak

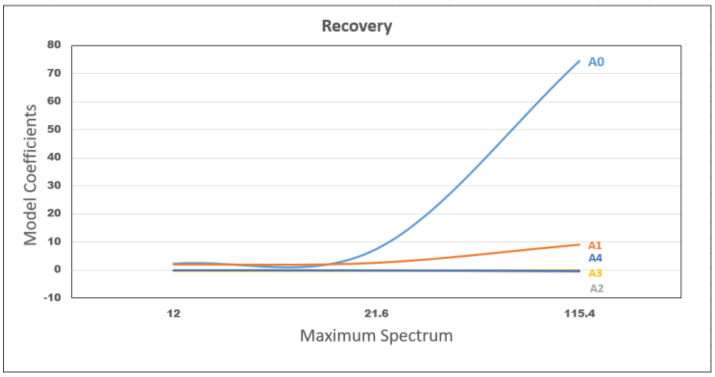
Complete Recovery	**Equations**	**x2**	**x1**	**constant**
A0	$y = 2.0336x^2 + 52.841x - 52.766$	2.0336	52.841	-52.766
A1	$y = -0.8587x^2 - 18.92x + 18.994$	-0.8587	-18.92	18.994
A2	$y = -0.0437x^2 + 2.7256x - 2.6091$	-0.0437	2.7256	-2.6091
A3	$y = -0.0007x^2 - 0.0674x + 0.0663$	-0.0007	-0.0674	0.0663
A4	$y = -5E-05x^2 + 0.0007x - 0.0007$	-5.00E-05	0.0007	-0.0007
Recovery	**Equations**	**x2**	**x1**	**constant**
A0	$y = 30.893x^2 - 87.437x + 58.707$	30.893	-87.437	58.707
A1	$y = 2.8885x^2 - 8.0287x + 7.0275$	2.8885	-8.0287	7.0275
A2	$y = -0.2074x^2 + 0.5547x - 0.4786$	-0.2074	0.5547	-0.4786
A3	$y = 0.0037x^2 - 0.0092x + 0.0084$	0.0037	-0.0092	0.0084
A4	$y = -2E-05x^2 + 2E-05x - 3E-05$	-2.00E-05	2.00E-05	-3.00E-05
Safe Stage	**Equations**	**x2**	**x1**	**constant**

A0	y = -7.8657x2 + 9.1193x - 14.723	-7.8657	9.1193	-14.723
A1	$y = 4.1343x^2 - 6.9721x + 30.654$	4.1343	-6.9721	30.654
A2	$y = -0.1155x^2 + 0.2827x - 1.7892$	-0.1155	0.2827	-1.7892
A3	$y = 0.0004x^2 - 0.0031x + 0.035$	0.0004	-0.0031	0.035
A4	$y = 1E-05x^2 - 4E-05x - 0.0002$	1.00E-05	-4.00E-05	-0.0002
Safe Stage	**Equations**	**x2**	**x1**	**constant**
A0	y = 2.74x + 217.08		2.74	217.08
A1	y = -4.289x - 55.536		-4.289	-55.536
A2	y = -2.3347x + 9.0834		-2.3347	9.0834
A3	y = 0.1057x - 0.2902		0.1057	-0.2902
A4	y = -0.0011x + 0.0026		-0.0011	0.0026
Stabilizing Stage	**Equations**	**x2**	**x1**	**constant**
A0	$y = 10.809x^2 - 76.084x + 87.787$	10.809	-76.084	87.787
A1	$y = 0.2236x^2 + 6.749x + 0.2206$	0.2236	6.749	0.2206
A2	$y = 0.2724x^2 - 1.1272x + 0.7247$	0.2724	-1.1272	0.7247
A3	$y = -0.0097x^2 + 0.0359x - 0.0263$	-0.0097	0.0359	-0.0263
A4	$y = 9E-05x^2 - 0.0003x + 0.0002$	9.00E-05	-0.0003	0.0002
Stabilizing Stage	**Equations**	**x2**	**x1**	**constant**
A0	$y = 33.507x^2 - 136.71x + 90.587$	33.507	-136.71	90.587
A1	$y = -11.286x^2 + 47.985x - 26.037$	-11.286	47.985	-26.037
A2	$y = 0.4531x^2 - 1.529x + 0.7396$	0.4531	-1.529	0.7396
A3	$y = -0.0067x^2 + 0.016x - 0.0064$	-0.0067	0.016	-0.0064
A4	$y = 5E-05x^2 - 8E-05x + 6E-05$	5.00E-05	-8.00E-05	6.00E-05
Beyond the critical	**Equations**	**x2**	**x1**	**constant**
A0	$y = -25.83x^2 + 74.344x - 124$	-25.83	74.344	-124
A1	$y = 15.715x^2 - 47.969x + 64.262$	15.715	-47.969	64.262
A2	$y = -0.4206x^2 + 0.1492x - 1.8163$	-0.4206	0.1492	-1.8163
A3	$y = -0.017x^2 + 0.148x - 0.0443$	-0.017	0.148	-0.0443
A4	$y = 0.0003x^2 - 0.0019x + 0.0008$	0.0003	-0.0019	0.0008
Critical Stage	**Equations**	**x2**	**x1**	**constant**
A0	$y = 3.0407x^2 - 27.28x + 24.105$	3.0407	-27.28	24.105
A1	$y = 1.6672x^2 - 0.7752x + 6.5275$	1.6672	-0.7752	6.5275
A2	$y = -0.2615x^2 + 1.3184x - 1.4248$	-0.2615	1.3184	-1.4248
A3	$y = 0.0134x^2 - 0.0694x + 0.0678$	0.0134	-0.0694	0.0678
A4	$y = -0.0001x^2 + 0.0007x - 0.0007$	-0.0001	0.0007	-0.0007



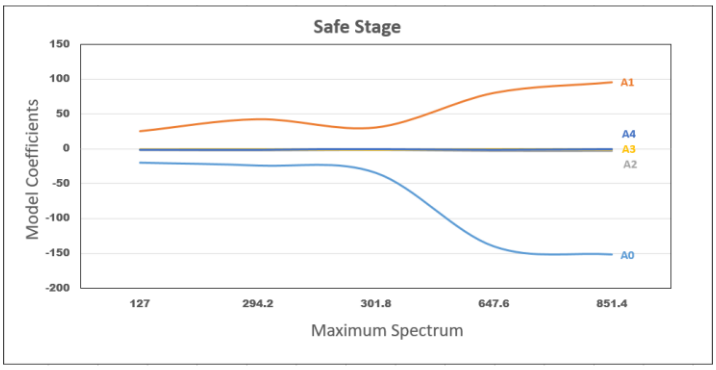
Complete Recovery
Model Coefficients
Maximum Spectrum
140
120
100
80
60
40
20
0
-20
-40
-60
9.6
412
622.6
A0
A2
A3
A4
A1

(a)



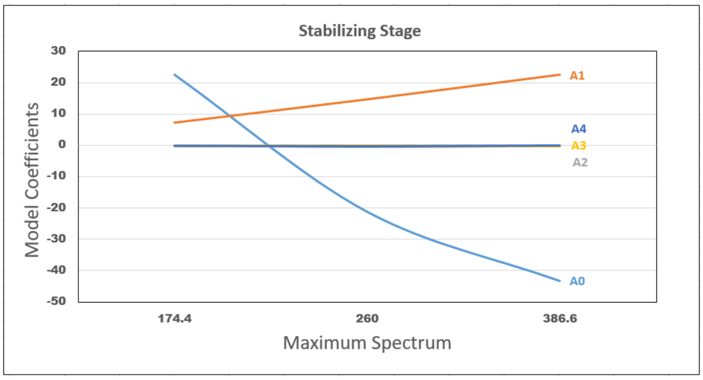
Recovery
Model Coefficients
Maximum Spectrum
80
70
60
50
40
30
20
10
0
-10
12
21.6
115.4
A0
A1
A4
A3
A2

(b)



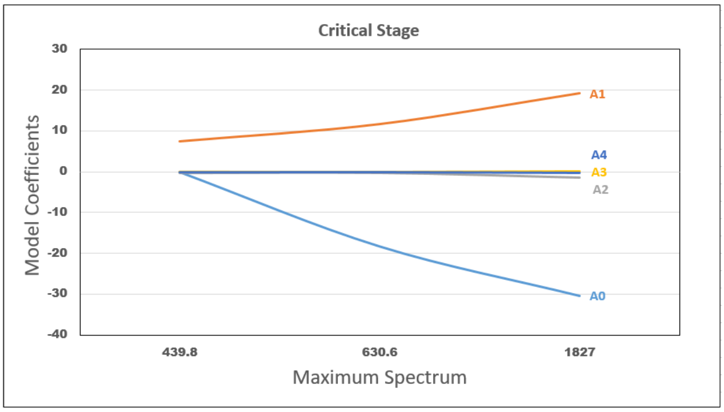
Safe Stage
Model Coefficients
Maximum Spectrum
150
100
50
0
-50
-100
-150
-200
127
294.2
301.8
647.6
851.4
A1
A4
A3
A2
A0

(c)



Stabilizing Stage
Model Coefficients
Maximum Spectrum
30
20
10
0
-10
-20
-30
-40
-50
174.4
260
386.6
A1
A4
A3
A2
A0

(d)



(e)

Figure 2: Variation of Spectrum Model Coefficients with Spectrum

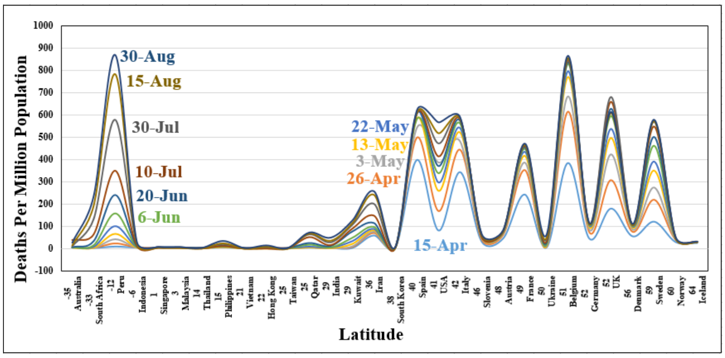
These coefficients vary with maximum expected peak of spectrum of the outbreak and follow polynomials such as

$$y\ (x{:}max) = \text{Constant} + X_1\,x + X_2x^2 \qquad (3)$$

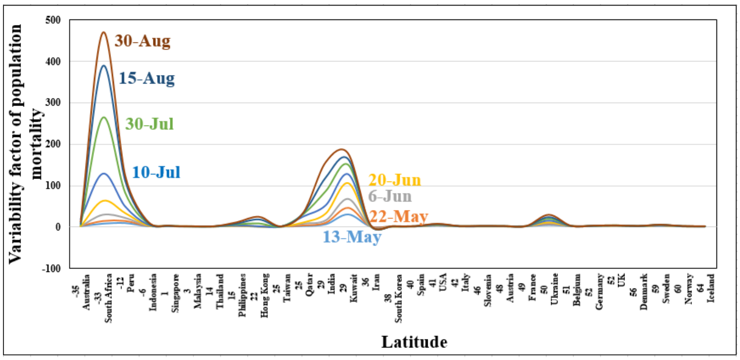
Figure 2(a) to 2(e) depicts the variation of the coefficients of spectrum models for different stages of the outbreak. It shows the different variations of the coefficients for different stages of the outbreak and helps in predicting the trends of the spectrum.

5. POPULATION MORTALITY

Figure 3(a) depicts the relatively lower population mortality for countries situated below the latitude 38˚N based on population mortality data of 30 countries from 15 April to 30 August 2020 supporting vitamin D factor , whereas Figure 3 (b) depicts the significant variability factor of population mortality for countries that lie below the latitude 38˚N, which determines the severity of the outbreak (Verma et al., 2020).



(a)



(b)

Figure 3: Variation of Population Mortality and its Variability Factor with Latitude

6. CONCLUSION

The present study describes the spectrum models for predicting the trends of the spectrum for different stages of the coronavirus outbreak, which can be used by the decision or policy maker to design the geo-spatial measures to control the outbreak. Significant temporal variability factor of population mortality from 13 May to 30 August 2020 determines the severity of the outbreak for the countries situated at the latitude between 64 $^{\circ}$N and 35°S, whereas population mortality depends on the latitude of the countries showing relatively lower populating mortality below the latitude 38 $^{\circ}$N.

7. REFERENCES

BBC, 2020.Coronavirus: The world in lockdown in maps & charts/ Source: Oxford Covid-19 Government Response Tracker /BBC Research. Available online at: https://www.bbc.com/news/world-52103747 (Last access date: 10 July 2020).

Coronavirus, 2020. Global Covid-19 Data for All Countries. Available online at: https://www.worldmeters.info/coronavirus (Last access date: 10 July 2020).

Jonathan, M.R., Sreedar, S., Eamon, L.& Rose, A.K., 2020. Low population mortality from COVID-19 in countries south of latitude 35 degree north supports vitamin D as a factor determining severity. Aliment. Pharm. Therap. Vol. 51, pp 1438-1439.

Panarese A. & Shahini E., 2020. Covid-19 and vitamin D. Aliment. Pharm. Therap. Vol.51, pp 993-995.

Verma, A.K., Verma, A., Verma A., 2020. Coronavirus Spatial Big Data Predictive Analysis for the Southeast Asian Region. Defence S & T Technical Bulletin, Ministry of Defence, Malaysia. Vol.12 (2), pp 344- 366.