

# Clustering Pandemic COVID-19 and Relationship to Temperature and Relative Humidity Among the Tropic and Subtropic Region

Giarno

*Paotere Meteorological Maritime Station, Indonesian Meteorological, Climatological, and Geophysical Agency, Makassar 90163, Indonesia*

(Corresponding author's e-mail: [giarno97182@gmail.com](mailto:giarno97182@gmail.com))

*Received: 2 April 2020, Revised: 17 June 2021, Accepted: 25 June 2021*

## Abstract

The outbreak of Novel Corona Virus (COVID-19) has been spreading almost in all countries of the world and become a deadly pandemic. The infections and deaths vary from high in some countries and low in others. The weather conditions significantly affect life, including viruses. In low temperature and humidity the spreading of coronavirus is expected to be fast and massive, and on the other hand, high temperature and humidity decreases the virus. However, recent data of COVID-19 shows that in tropical region infection and deaths vary of which there is a need of thorough spreading analysis. The clustering of infections and mortality at the beginning of COVID-19 outbreak was group based on the country's profile similarity, and associated with the meteorological factors. The result shows that countries such as China, Spain, Italy and the United States have very severe attacks of COVID-19 infection. Furthermore, countries with the potential real threats of COVID-19 infections are Austria, Australia, Azerbaijan, Belgium, Bahrain, Brazil, Belarus, Canada, Switzerland, Czech Germany, Denmark, Dominican Republic, Algeria, Ecuador, Estonia, Egypt, Finland, France, Georgia, Croatia, Indonesia, Ireland, Israel, India, Iraq, Iran, Japan, Cambodia, South Korea, Kuwait, Lebanon, Sri Lanka, Lithuania, Monaco, Macedonia, Mexico, Malaysia, Nigeria, Netherlands, Norway, Nepal, New Zealand, Oman, Philippines, Pakistan, Qatar, Romania, Russia, Sweden, Singapore and Thailand. The threat of COVID-19 is not only in dry and humid sub-tropical countries, but it cannot be undermined the effect to some warm and humid tropical countries such as Brazil, Ecuador, Indonesia, Malaysia and the Philippines, which are massively infected, and the mortality rate compared to the population are very high. The study also found that dynamic humidity is a factor that must be considered, especially in the tropics.

**Keywords:** COVID-19, Weather, World, Tropic, Subtropic

## Introduction

The current COVID-19 outbreak is driven by a novel coronavirus (SARS CoV-2) that is spreading among people. The recent outbreak of novel coronavirus (2019-nCoV) that infected people in almost all countries was first noted in Wuhan [1,2], the capital of the Hubei province of Central China. It was declared by the WHO (World Health Organization) on January 30, 2020 as a pandemic disease and was officially called COVID-19 [3]. The symptoms of this virus include breathing difficulty, dry cough, fever, fatigue and bilateral lung infiltration in severe cases, similar to those caused by SARS and MERS infections [4].

Coronaviruses are enveloped positive-sense, non-segmented RNA viruses belonging to the Coronaviridae family and the Nidovirales order and widely distributed in humans and other mammals [2]. The COVID-19 geographically connected to the Huanan wild animal market, and the habits of its inhabitants who eat wild animals without being slaughtered and cooked properly. The 1<sup>st</sup> human infections were reported at the end of December 2019 in Wuhan, Hubei province in China when a cluster

of [3] pneumonia cases was identified as a novel coronavirus [5]. The analysis found that a third of the cases had direct exposure to the Huanan Seafood market. In-depth analysis found that the sequences of illness caused by COVID-19 which is 79.6 % identical to SARS-CoV, and 96 % identical at the whole-genome level to a bat coronavirus [5,6].

Almost a year before it happened, Fan *et al.* (2019) predicted that there would be more likely a coronavirus outbreak in the future that will originate from bats, and there is an increased probability that this will occur in China [7]. Bats are the 2<sup>nd</sup> largest order of mammals, and the only mammals with the capability of powered flight which contributed to a longer range of migration compared to land mammals, and are distributed worldwide [8]. Soon after the 1<sup>st</sup> case in Wuhan, the COVID-19 spreads from person to person was confirmed and had become an outbreak with more than 680,000 confirmed infections with 31,000 reported deaths worldwide as of March 29, 2020 [9]. The deaths and infections vary from high such as Italy and Spain, and some countries have a lot of infections but less mortality such as Japan, Korea and some Middle East nations.

The weather condition significantly affects life, including viruses. Chan *et al.* (2011) reported that infectivity on types of corona virus such as SARS-CoV is lost after exposure to heat at 56 °C for 15 min, albeit it will be longer on plastic [10]. Moreover, the ultraviolet light, alkaline (pH > 12), or acidic (pH < 3) conditions make the virus inactive [11]. The human corona viruses have been shown to survive in Phosphate-buffered saline (PBS) or cultured medium with 5 - 10 % Fetal Calf Serum (FCS) for several days [12-14], but they only survive a few hours after drying. Some studies reported that weather conditions and air pollution are related to the SARS outbreak [15-17]. The environment at different temperature and humidity condition is significant to the survival of the coronavirus family such SARS-CoV or SCoV [10,18]. Chan *et al.* (2011) found that the dried virus on smooth surfaces retained its viability for over 5 days at temperatures of 22 - 25 °C and relative humidity of 40 - 50 %. However, the virus rapidly dissipate at higher temperatures and relative humidity (e.g., 38 °C and > 95 % of relative humidity). The stability of coronavirus at low temperature and humidity may facilitate its transmission in the subtropical, and from the aforementioned, some Asian countries in the tropical region (such as Malaysia, Indonesia or Thailand) did not have major community outbreaks of SARS.

Many infectious diseases have a seasonal pattern in their occurrence and transmission globally, influenza is an example [19]. The influenza, SARS and MERS show significant seasonal fluctuation in temperate regions of the world, but nevertheless display less seasonality in tropical areas [20-22]. Since first identified in January 10, 2020, the disease caused by COVID-19 has been found in the sub-tropical region of Wuhan, China, then quickly proliferate almost all over the world. The United States, Italy, China, Spain and Germany are countries with the highest infections and more than 50,000 deaths on March 29, 2020 [8]. However, based on the proportion between the number of people infected and the number of deaths, Indonesia which ranks second to Italy, has 1,285 infections to 114 deaths, which is 11.2 %. This is actually even higher than Italy, which accounts for 11.0 %. So, considering that the population of Indonesia and other Asian countries in the tropics are very large and with minimal health facilities, it is necessary to study the characteristics of the spread of COVID-19 in this region.

Whereas, recent study shows that the distribution of outbreaks is related to latitude, temperature and humidity [22]. These environmental factors affect the variability of spreading COVID-19. Another data shows that there is an increase in the number of COVID-19 infections, and a high number of deaths in relatively hot and humid tropics such as Indonesia [9]. Therefore, it is necessary to present a cluster region based on similar cases and their relation to weather condition. We mainly compare the Tropics and Subtropic region on COVID-19 cases of infection and death in relation to the temperature and humidity at the beginning of the outbreaks. The grouping is very important so that it can be used as models of handling COVID-19 pandemic and minimizing death casualties.

## Materials and methods

The data used in this study are from the monitoring report of European Centre for Disease Prevention and Control [24]. The variables of the data include country, date, new case, death and population. For mapping purposes, the location of the capital city of a country is used as a reference to

determine whether an area is tropical or not. Data on capital positions of these countries are obtained from SimpleMaps.com [25]. Determining the tropical and sub-tropical regions is based on astronomical zone criteria. The tropical regions are in 30 °S - 30 °N and 30 - 66 °N/S. Meteorological data such as air temperature, relative humidity and wind speed are taken from National Center for Atmospheric Research (NCAR) [26].

**Table 1** Covid-19 cases at the beginning pandemic for each country.

Country	Day	Case	Death	Country	Day	Case	Death	Country	Day	Case	Death
Afghanistan	80	106	3	Denmark	90	2,201	65	Japan	90	1,693	52
Angola	8	4	0	Djibouti	11	15	0	Jordan	17	235	1
Albania	21	197	10	Dominica	81	719	28	Kuwait	87	235	0
Andorra	16	308	4	Ecuador	85	1,835	48	Laos	5	6	0
Antigua	10	7	0	Egypt	86	536	30	Latvia	23	305	0
Algeria	85	409	26	El Salvador	11	19	0	Lebanon	86	412	8
Argentina	23	745	19	Eritrea	8	6	0	Saudi Arabia	24	1,203	4
Armenia	81	424	3	Estonia	85	640	1	Philippines	86	1,075	68
Aruba	8	28	0	Eswatini	15	9	0	Poland	24	1,638	18
Australia	90	3,809	14	Ethiopia	16	16	0	Portugal	27	5,170	100
Austria	90	8,291	68	Faroe	10	155	0	Qatar	86	590	1
Azerbaijan	83	182	4	Fiji	10	5	0	Romania	88	1,452	29
Bahamas	12	11	0	Finland	85	1,218	9	Russia	85	1,264	5
Bahrain	89	473	4	France	90	37,575	2,314	Vatican	15	60	0
Barbados	12	26	0	Germany	90	52,547	389	Myanmar	7	8	0
Belarus	82	94	0	Ghana	16	141	5	Netherlands	90	9,762	639
Belgium	90	9,134	353	Gibraltar	10	56	0	New Zealand	81	476	1
Belize	6	2	0	Greece	88	1,061	32	Nigeria	79	97	1
Benin	13	6	0	Haiti	10	8	0	Norway	90	3,845	20
Bermuda	10	22	0	Honduras	16	110	2	Oman	84	152	0
Bhutan	16	4	0	Hungary	24	408	13	Pakistan	85	1,408	11
Bolivia	17	81	0	India	89	979	25	Slovakia	22	295	0
Brazil	87	3,904	114	Indonesia	83	1,155	102	Slovenia	24	691	9
Brunei	19	120	1	Iran	90	35,408	2,517	South Africa	22	1,187	2
Bulgaria	20	331	7	Iraq	88	506	42	South Korea	90	9,583	152
Burkina F	17	180	9	Ireland	88	2,415	36	Spain	90	72,248	5,690
Cambodia	81	103	0	Israel	87	3,619	12	Sudan	16	5	1
Cameroon	17	99	2	Italy	90	92,472	10,023	Sweden	90	3,447	102
Canada	90	5,386	60	Luxembourg	83	1,831	0	Switzerland	90	13,152	235
Cabo Verde	9	6	1	Madagascar	9	28	18	Taiwan	88	283	2
Chad	10	5	0	Malaysia	89	2,320	0	Thailand	83	1,245	6
Chile	25	1,909	6	Malta	20	139	27	Timor-Leste	8	1	0
China	90	82,342	3,306	Mauritania	15	5	0	Tunisia	21	227	6

Country	Day	Case	Death	Country	Day	Case	Death	Country	Day	Case	Death
Colombia	19	608	6	Mauritius	10	102	0	Turkey	16	7,402	108
Congo R.	14	19	0	Mexico	82	848	2	UK	90	17,089	1,019
Costa Rica	22	295	2	Moldova	20	231	16	United States	90	124,665	2,191
Croatia	87	657	5	Monaco	78	43	2	Venezuela	15	119	1
Cuba	15	119	3	Liberia	13	3	0	Vietnam	86	223	0

Cluster analysis is widely used in various fields including the classification of weather parameters [27-29] and the treatment of a pandemic [30-32]. This analysis uses the task of grouping similar objects in the same group. Clustering is a broad set of techniques for finding subgroups of observations within a data set. This is an unsupervised method because there isn't a response variable, which implies that one may investigate the relationships between the observations without being trained by a response variable. The K-means clustering is the simplest and the most commonly used clustering method for splitting a data set into a set of  $k$  groups. Moreover, this method was chosen because it is simple, guarantees convergence, can warm-start the positions of centroids, easily adapts to new examples, and generalizes to clusters of different shapes and sizes [33,34]. The K-means application can be used to detect climate change variables and predict weather [35].

The classification of observations into groups requires some methods for computing the distance or the (dis)similarity between each pair of observations, and the choice of distance measures is a critical step in clustering. If there are 2 elements ( $x, y$ ) and it will influence the shape of the clusters, then the classical methods for distance measures are Euclidean, Manhattan and Pearson correlation distance, which are defined as follow:

1. Euclidean distance:

$$d_{euc}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

2. Manhattan distance:

$$d_{man}(x, y) = \sum_{i=1}^n |x_i - y_i|$$

where,  $x$  and  $y$  are 2 vectors of length  $n$ .

3. Pearson correlation distance:

$$d_{cor}(x, y) = 1 - \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

The choice of distance measures is very important, as it has a strong influence on the clustering results. For most common clustering software, the default distance measure is the Euclidean distance. However, depending on the type of the data and the research questions, other dissimilarity measures might be preferred and you should be aware of the options.

K-means clustering is the most commonly used unsupervised machine learning algorithm for partitioning a given data set into a set of  $K$ -groups. In K-means clustering, each cluster is represented by its center (i.e, centroid) which corresponds to the mean of points assigned to the cluster. The basic idea behind K-means clustering consists of defining clusters, so that the total intra-cluster variation (known as total within-cluster variation) is minimized. The standard algorithm is the Hartigan-Wong algorithm

(1979), which defines the total within-cluster variation as the sum of squared Euclidean distances between items and the corresponding centroid:

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2$$

where  $x_i$  is a data point belonging to the cluster  $C_k$  and  $\mu_k$  is the mean value of the points assigned to the cluster  $C_k$ .

Each observation  $x_i$  is assigned to a given cluster such that the sum of squares (SS) distance of the observation to their assigned cluster centers  $\mu_k$  is minimized. Then, the total within-cluster variation as follows:

$$tot. withiness = \sum_{k=1}^k W(C_k) = \sum_{k=1}^k \sum_{x_i \in C_k} (x_i - \mu_k)^2$$

The total within-cluster sum of square measures the compactness (i.e goodness) of the clustering and we want it to be as small as possible.

The 1<sup>st</sup> step of K-means algorithm is to indicate the number of clusters (K) that will be generated in the final solution. The algorithm starts by randomly selecting K-objects from the data set to serve as the initial centers for the clusters. The selected objects are also known as cluster means or centroids. Next, each of the remaining objects is assigned to the closest centroid between the object and the cluster mean. After the assignment step, the algorithm computes the new mean of each cluster. The cluster or exactly centroid update is used to design this step. The centers have been recalculated and every observation is checked again to see if it might be closer. K-means algorithm can be summarized as follows:

- Specify the number of clusters (K) to be created.
- Select randomly K-objects from the data set as the initial cluster centers or means.
- Assigns each observation to their closest centroid, based on the Euclidean distance.
- Update the K-cluster centroid by calculating the new mean values of all the data points in the cluster.
- Iteratively minimize the total within sum of square.

Meteorological variables such as air temperature and humidity of each cluster are to be analyzed to see the relationship between meteorological factors and the spread of COVID-19. In the tropics with warm temperatures throughout the year have different variability in each cluster with regards to spreading of COVID-19.

## Results and discussion

The cluster analysis refers to a class of techniques assigned to classify a multivariate dataset into clusters whose members are more similar to one another. Simply, these work clusters are divided into 4 classes based on the days since the virus had been known in a country, number of positive cases, number of deaths, the average number of deaths in positive cases and density positive cases of the countries. Moreover, the days, since cases had been discovered against the number of positive cases in reference to the preparation of a country facing an outbreak.

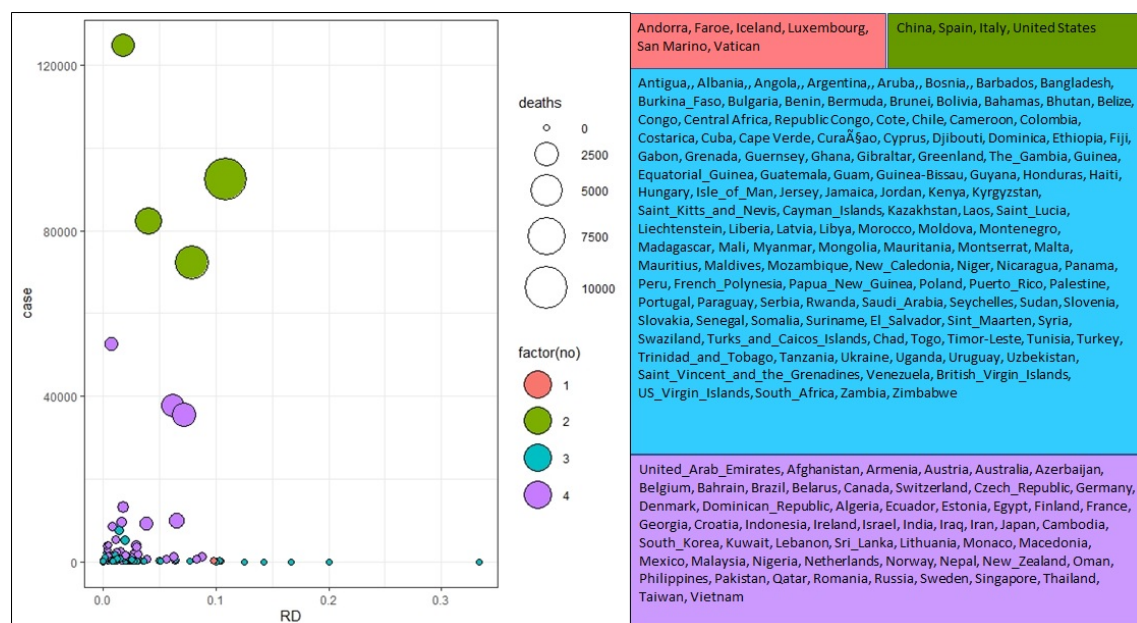
### Clustering results

A country that has high mobility allows the number of positive cases. If there is a minimal positive cases, the country may immediately impose restrictions so that there will be no massive transmission. Meanwhile, the number of deaths and the percentage of deaths in cases show how dangerous COVID-19 situation in the country. Additionally, population density variables need to be considered, with view in mind that population density makes it easier for viruses to spread rapidly. The clustering is carried out in

2 stages, namely globally and specifically equatorial. The application of cluster grouping following the tropical and subtropical regions, the results as follows:

The first, the clustering was done in all positively infected countries, without territorial division as shown in **Figure 1**. From the result, the 1<sup>st</sup> group consists of countries with a small population and territory. This group includes Andorra, Faroe, Iceland, Luxembourg, San Marino and Vatican. Meanwhile, the 2<sup>nd</sup> group includes the countries with large population and many victims died due to COVID-19 infection. Countries such as China, Spain, Italy and the United States have a very large number of fatalities due to infection with this virus. While, the 3<sup>rd</sup> group are countries where the number of cases is relatively low compared to the total population. Finally, there is a densely populated country with a relatively large number of positive cases, but the number of people who died compared to the population is relatively less. Looking into the dense population group, this last group will have a potential increase of the threat of COVID-19 in the future.

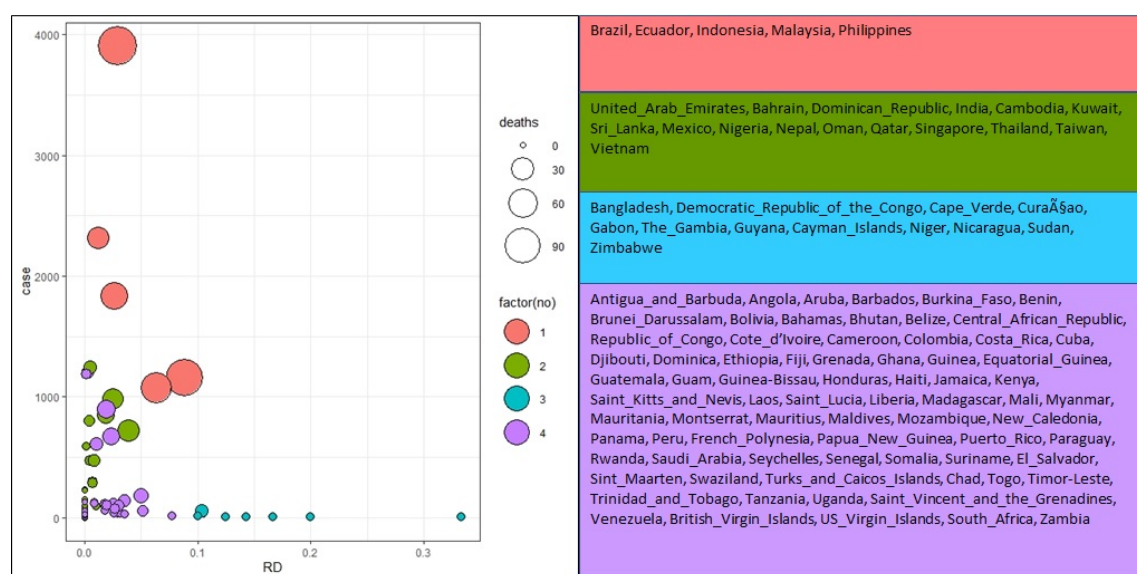
The 2<sup>nd</sup> group is very prominent in the cluster as seen in the 4 green circles in **Figure 1**. However, potential in the threat of virus in the 4<sup>th</sup> group is also very obvious, especially countries with positive cases of more than 1,000 such as Austria, Australia, Azerbaijan, Belgium, Bahrain, Brazil, Belarus, Canada, Switzerland, Czech Germany, Denmark, Dominican Republic, Algeria, Ecuador, Estonia, Egypt, Finland, France, Georgia, Croatia, Indonesia, Ireland, Israel, India, Iraq, Iran, Japan, Cambodia, South Korea, Kuwait, Lebanon, Sri Lanka, Lithuania, Monaco, Macedonia, Mexico, Malaysia, Nigeria, Netherlands, Norway, Nepal, New Zealand, Oman, Philippines, Pakistan, Qatar, Romania, Russia, Sweden, Singapore and Thailand. So much so that this group also includes France and Iran where the death toll is very large, therefore the countries in this group should be vigilant.



**Figure 1** The clustering in all positively infected countries. Picture in the left, the color on the circle shows the cluster and the size of the circle is the number of deaths from COVID-19. Whereas, the colors in the right picture shows a group, where red, dark red, yellow and light green indicate a group of clusters of countries 1 to 4.

Deeper analysis divided clusters into climate zones, the tropical and subtropical regions. Clusters in the tropics show that Brazil, Ecuador, Indonesia, Malaysia and Philippine are in the 1<sup>st</sup> group, where the population and people infected with COVID-19 are quite large as shown in **Figure 2**. Indonesia is the country in this group with the highest case-fatality rate of 8,831 %. The next group are countries with a large number of cases, but the death rate is still low. Included in this group are the United Arab Emirates, Bahrain, Dominican Republic, India, Cambodia, Kuwait, Sri Lanka, Mexico, Nigeria, Nepal, Oman, Qatar, Singapore, Thailand, Taiwan and Vietnam. Meanwhile, the 3<sup>rd</sup> and 4<sup>th</sup> group are densely populated, but the number of people infected with the virus is still low, or in countries where there is little population and the number of corona infections is rarely found.

Brazil is the most prominent country in the 1<sup>st</sup> group with the highest number of infected people, while Indonesia and the Philippines are the countries with the highest mortality rates in this group. Although the mortality rate in these 2 countries is not the highest, especially when compared to the 3<sup>rd</sup> group, the number of positive cases in Indonesia and the Philippines is much higher, so if not intervened the threat of COVID-19 virus victims will increase.

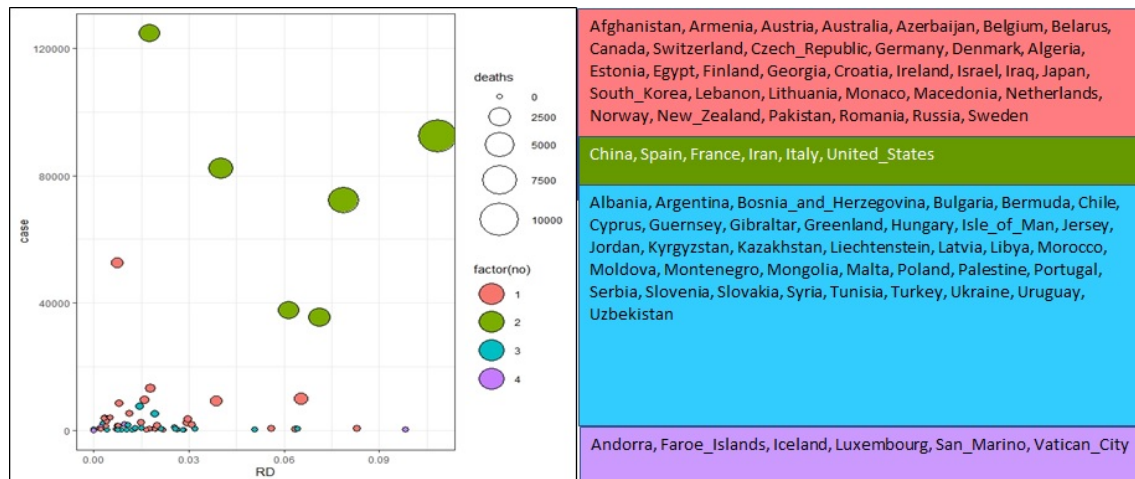


**Figure 2** The clustering in the tropical region. Picture in the left, the color on the circle shows the cluster and the size of the circle is the number of deaths from COVID-19. Whereas, the colors in the right picture shows a group, where red, dark red, yellow and light green indicate a group of clusters of countries 1 to 4.

For subtropics that are relatively cooler, the 1<sup>st</sup> group are countries with large populations and COVID-19 infected people, but has low mortality compared to the number of positive cases as depicted in **Figure 3**. This 1<sup>st</sup> group which started to be threatened by the danger of the corona virus includes Afghanistan, Armenia, Austria, Australia, Azerbaijan, Belgium, Belarus, Canada, Switzerland, Czech Republic, Germany, Denmark, Algeria, Estonia, Egypt, Finland, Georgia, Croatia, Ireland, Israel, Iraq, Japan, South Korea, Lebanon, Lithuania, Monaco, Macedonia, Netherlands, Norway, New Zealand, Pakistan, Romania, Russia and Sweden. While the 2<sup>nd</sup> group are countries that has a very large number of infections and many had died as a result of this virus. Included in this group of countries are China, Spain, France, Iran, Italy and United States. In the subtropics, the 3<sup>rd</sup> and 4<sup>th</sup> groups are almost the same as the 3<sup>rd</sup> and 4<sup>th</sup> groups in the tropics where the country is densely populated but the number infected by the virus is low or a country with a small population and corona virus infection is rarely found.



Interestingly, the 4<sup>th</sup> group consisting of Andorra, Faroe Islands, Iceland, Luxembourg, San Marino, Vatican City are countries with a very small population and narrow area. Countries such as China, Spain, France, Iran, Italy and United States are very susceptible to this virus attack, with a massive number of infected people coupled with a very large number of deaths. In **Figure 3** on the left, in addition to the high number of infected, also seen is the very large number of deaths in these 4 countries.



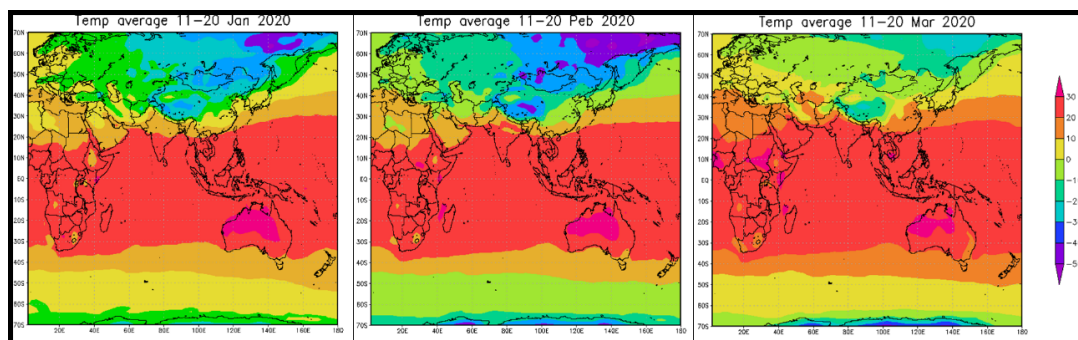
**Figure 3** The clustering in the subtropical region. Picture in the left, the color on the circle shows the cluster and the size of the circle is the number of deaths from COVID-19. Whereas, the colors in the right picture shows a group, where red, dark red, yellow and light green indicate a group of clusters of countries 1 to 4.

#### Air temperature at the time of the corona outbreaks

The tropics are characterized by warm to hot temperatures throughout the year. Importantly, the changing of temperature is small compared with regions outside the tropics. Rainfall is largely abundant, but it becomes increasingly seasonal with its distance from the equator causing this area to be very humid.

Further analysis using 2-meter (2 m) temperatures in early 2020 shows that in the beginning of this year, from January to March, the temperature in the tropics is higher in contrasts to the other regions as depicted in **Figure 4**. In tropical maritime areas where there is an abundant heat source, it allows the area to produce latent heat that keeps it warm compared to other places. This condition has the potential to restrain the rate of spread of COVID-19 during the early part of the year. There was no significant difference in average air temperatures between countries that have positive cases of corona in the tropics, namely Brazil, Ecuador, Indonesia, Malaysia and the Philippines, in contrast to the places where corona virus caused many casualties namely China, Italy, America and Spain with cold temperatures. In some of these regions the temperature is 0 - 10 °C and in other parts 0 - 10 °C. Wuhan which is the origin of COVID-19 in early January 2020 has a temperature of 0 - 10 °C. This temperature is lower in February, where COVID is increasingly spreading outside China, and then decreases in intensity in March. Conversely, in European and American subtropics countries during the month of January the number of infected are still low, however, in March it increased tremendously, even though the temperature in that place increased in as well.





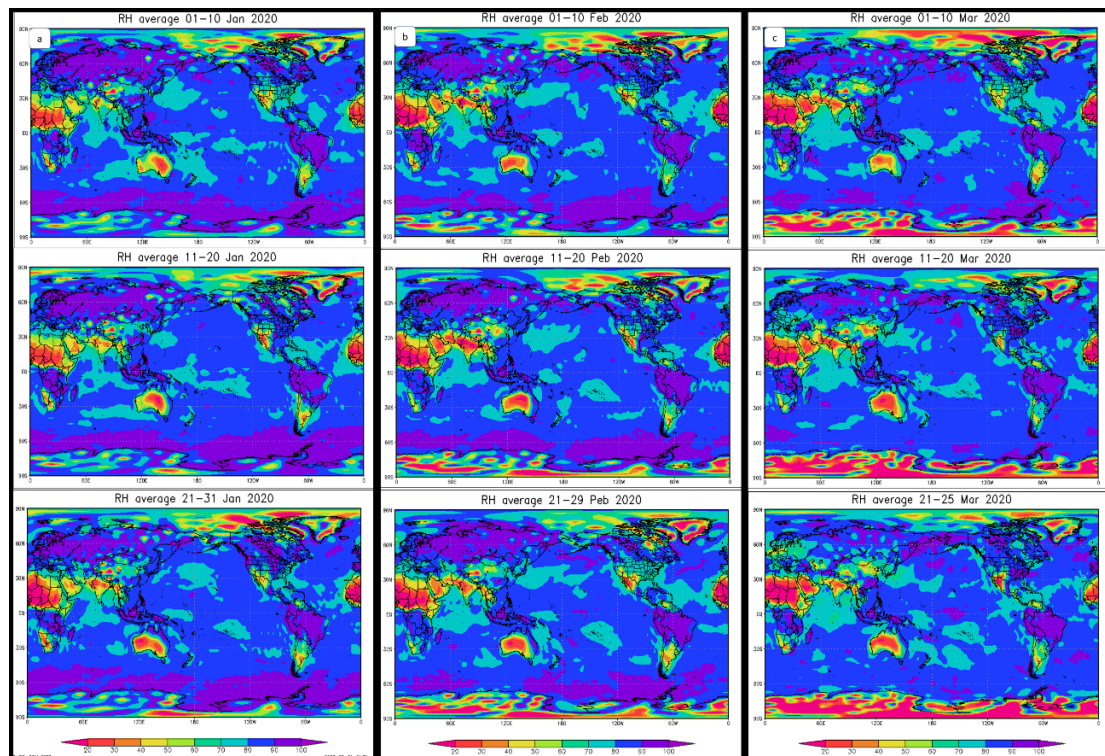
**Figure 4** The average air temperature in early 2020.

Temperature variability in the tropics is very low, so if the analysis only uses temperature, it will not provide a clarity of the relationship between the number of positive cases of coronavirus and the number of deaths that actually varies in this area. The next analysis uses relative humidity or RH which turns out to have variations in humidity in the tropics as illustrated in **Figure 5**.

The results of moisture reanalysis data show that the relative humidity is more variable than the temperature, especially in the tropics region. Moreover, on the land of the equator region, the humidity reached more than 90 % in the early 3 months of the year. While in oceans or small land rarely inhabited by humans in equatorial oceans, the humidity is between 70 - 90 % as shown in **Figure 5**. In the position further north, 10 °N or 10 °S to at the 30 °N/S, humidity varies between 20 - 90 %. Countries in the northern and central Africa, the northern Indian subcontinent and the northern part of the Malay Peninsula are places with little humidity value compared to other places in the tropics.

The subtropics region which is located between 30 - 50 °N has low humidity and has very dynamic changes. The worst places due to COVID-19 infection and a high number of deaths are in this area. The eastern region of China where the spread of COVID-19 originated, the humidity increased in March and at the same time the plague at that place began to subside. While in this same month, in European countries the humidity dropped along with increased infection.

The subtropics in the south also have high humidity fluctuations. Although in this area the population is not as dense as in the north at the same latitude, but in this region it has densely populated cities like some cities in Australia and South Africa. Meanwhile, in the northern equator between 5 - 30 °N in Africa, the Middle East and West Asia have low humidity, where in this area the air temperature ranges from 0 - 20 °C. Some countries located in this region have high infections, but there are also those with low infection rates. In places where the infection and poverty are low, such as in Saudi Arabia, etc. These prosperous countries are very efficient in preventing COVID-19 infections. In Saudi Arabia for example, the prohibition of direct visits was imposed from the beginning of the infection, even to countries that have not reported any coronavirus infections, for to the Saudi authorities they considered it to be dangerous. Even the hajj pilgrimage that Muslims usually do every year is forbidden. But in this region, there are countries having increased in infection rates such as Turkey and Israel, although the positive case-fatality rate is rather low.



**Figure 5** The average relative humidity in early 2020.

### Discussion

The current COVID-19 outbreak has reached in almost all countries in the world, 3 months since the it was first noted in Wuhan [1,2], the capital of the Hubei province of Central China. Tens of 1,000 of people died, millions of infected people have to be helped and the potential healthy people must be protected immediately. The factors that support the spread of the virus must be considered, and one of those is the weather conditions.

The association between temperature in the cities affected with COVID-19 deserves special attention (Sajadi *et al.*, 2020). The spread of corona virus variants such SARS-CoV, MERS-CoV and influenza is affected by temperature and humidity [10]. Human coronaviruses (HCoV-229E, HCoV-HKU1, HCoV-NL63, and HCoV-OC43) detected in strong winter seasonality between December and April, and are undetectable in summer months in temperate regions [20-22]. Sajadi *et al.* (2020) found that the distribution of COVID-19 infection restricted by latitude, temperature, and humidity are consistent with the behavior of a seasonal respiratory virus. The low temperature and humidity hampers the stability of coronavirus.

Subtropical countries such as China, Spain, Italy, Iran and the United States greatly suffered by severe attacks of COVID-19. In these countries the number of people infected and deaths are very large and the number continues to increase in the latter 3 countries. Previous research shows that low temperatures and humidity in this area coincide with coronavirus attacks, which were confirmed in this study by adding the dynamic variable of humidity that affects the infection rate as shown in 2<sup>nd</sup> cluster.

The 2<sup>nd</sup> cluster are countries with potential like infection as the 1<sup>st</sup> group. These countries with potential threats of COVID-19 infections includes Austria, Australia, Azerbaijan, Belgium, Bahrain, Brazil, Belarus, Canada, Switzerland, Czech Germany, Denmark, Dominican Republic, Algeria, Ecuador, Estonia, Egypt, Finland, France, Georgia, Croatia, Indonesia, Ireland, Israel, India, Iraq, Iran, Japan,

Cambodia, South Korea, Kuwait, Lebanon, Sri Lanka, Lithuania, Monaco, Macedonia, Mexico, Malaysia, Nigeria, Netherlands, Norway, Nepal, New Zealand, Oman, Philippines, Pakistan, Qatar, Romania, Russia, Sweden, Singapore and Thailand. These countries are located in tropical and sub-tropical regions. The warm and humid tropical countries such as Brazil, Ecuador, Indonesia, Malaysia and the Philippines have a lot of infections and high mortality. Large populations in these countries make the conditions potentially vulnerable to a COVID-19 pandemic, despite they are warm and humid place.

Sajadi *et al.* (2020) suggested that by using weather predictions can tentatively establish the potential community spread of COVID-19. They predicted that in 2 months onwards (March and April) the temperature will rise dramatically across many areas in the Northern Hemisphere, which proved true for Spain and the USA. However, it is necessary to add moisture dynamics to the prediction of COVID-19 spreading, especially in tropical countries that are relatively warm.

## Conclusions

Application of the (dis)similarity and the choice of distance measures based on geographical location at the onset of the outbreak of COVID-19 that was calculated by K-means algorithm shows the characteristics of the spreading of the virus in the early stages of the pandemic. In the countries that have large populations, although in a relatively warm place and the temperature is quite high, still there is a need to be vigilant. Moreover, weather changes at any time where the distribution of warm temperatures moved from the south to the north during the upcoming summer (July), in these equatorial areas, it is necessary to consider the dynamic humidity factor which will decrease when entering the dry season, especially in the dry season where there is generally low temperatures at night. The countries such as Austria, Australia, Azerbaijan, Belgium, Bahrain, Brazil, Belarus, Canada, Switzerland, Czech Germany, Denmark, Dominican Republic, Algeria, Ecuador, Estonia, Egypt, Finland, France, Georgia, Croatia, Indonesia, Ireland, Israel, India, Iraq, Iran, Japan, Cambodia, South Korea, Kuwait, Lebanon, Srilanka, Lithuania, Monaco, Macedonia, Mexico, Malaysia, Nigeria, Netherlands, Norway, Nepal, New Zealand, Oman, Philippines, Pakistan, Qatar, Romania, Russia, Sweden, Singapore, and Thailand could be potentially vulnerable of attack by COVID-19, if they will not act seriously to make an effort to prevent it.

## References

- [1] N Zhu, Z Zhang, W Wang, X Li, B Yang, J Song, X Zhao, B Huang, W Shi, R Lu, P Niu and F Zhan. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 2020; **382**, 727-33.
- [2] C Huang, Y Wang, X Li, L Ren, J Zhao, Y Hu, L Zhang, G Fan, J Xu, X Gu, Z Cheng, T Yu, J Xia, Y Wei, W Wu, X Xie, W Yin, H Li, M Liu, Y Xiao, H Gao, L Guo, J Xie, G Wang, R Jiang, Z Gao, Q Jin, J Wang and B Cao. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; **395**, 497-506.
- [3] Coronavirus disease (COVID-19) weekly epidemiological update and weekly operational update, Available at: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>, accessed January 2020.
- [4] EL Gralinski and VD Menachery. Return of the coronavirus: 2019-ncov. *Viruses* 2020; **12**, 135.
- [5] P Zhou, XL Yang, XG Wang, B Hu, L Zhang, W Zhang, HR Si, Y Zhu, B Li, CL Huang, HD Chen, J Chen, Y Luo, H Guo, RD Jiang, MQ Liu, Y Chen, XR Shen, X Wang, XS Zheng, K Zhao, QJ Chen, F Deng, LL Liu, B Yan, FX Zhan, YY Wang, GF Xiao and ZL Shi. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020; **579**, 270-3.
- [6] D Xu, Z Zhang, F Chu, Y Li, L Jin, L Zhang, GF Gao and FS Wang. Genetic Variation of SARS Coronavirus in Beijing Hospital. *Emerg. Infect. Dis.* 2004; **10**, 789-94.
- [7] Y Fan, K Zhao, ZL Shi and P Zhou. Bat coronaviruses in China. *Viruses* 2019; **11**, 1-14.
- [8] EC Teeling, MS Springer, O Madsen, P Bates, SJ O'Brien and WJ Murphy, A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* 2005; **307**, 580-4.

- [9] Worldometers: Covid-19 coronavirus pandemic, Available at: <https://www.worldometers.info/coronavirus>, accessed January 2020.
- [10] KH Chan, JS MalikPeiris, SY Lam, LLM Poon, KY Yuen and WH Seto. The effects of temperature and relative humidity on the viability of the SARS coronavirus. *Adv Virol.* 2011; 734690, 1-7.
- [11] MER Darnell, K Subbarao, SM Feinstone and DR Taylor. Inactivation of the coronavirus that induces severe acute respiratory syndrome, SARS-CoV. *J. Virol. Meth.* 2004; **121**, 85-91.
- [12] MK Ijaz, AH Brunner and SA Sattar. Survival characteristics of airborne human coronavirus 229E. *J. Gen. Virol.* 1985; **66**, 2743-8.
- [13] J Sizun, MWN Yu and PJ Talbot. Survival of human coronaviruses 229E and OC43 in suspension and after drying on surfaces: a possible source of hospital-acquired infections. *J. Hosp. Infect.* 2000; **46**, 55-60.
- [14] HF Rabenau, J Cinatl, B Morgenstern, G Bauer, W Preiser and HW Doerr. Stability and inactivation of SARS coronavirus. *Med. Microbiol. Immunol.* 2005; **194**, 1-6.
- [15] J Tan, L Mu, J Huang, S Yu, B Chen and J Yin. An initial investigation of the association between the SARS outbreak and weather: With the view of the environmental temperature and its variation. *J. Epidemiol. Commun. Health.* 2005; **59**, 186-92.
- [16] J Yuan, H Yun, W Lan, W Wang, SG Sullivan, S Jia and AH Bittles. A climatologic investigation of the SARS-CoV outbreak in Beijing, China. *Am. J. Infect. Control.* 2006; **34**, 234-6.
- [17] QC Cai, J Lu, QF Xu, Q Guo, DZ Xu, QW Sun, H Yang, GM Zhao and QW Jiang. Influence of meteorological factors and air pollution on the outbreak of severe acute respiratory syndrome. *Public Health* 2007; **121**, 258-65.
- [18] LM Casanova, S Jeon, WA Rutala, DJ Weber and MD Sobsey. Effects of air temperature and relative humidity on coronavirus survival on surfaces. *Appl. Environ. Microbiol.* 2010; **76**, 2712-7.
- [19] GBDI Collaborators. Mortality, morbidity, and hospitalisations due to influenza lower respiratory tract infections, 2017: An analysis for the global burden of disease study 2017. *Lancet. Respir. Med.* 2019; **7**, 69-89.
- [20] C Viboud, WJ Alonso and L Simonsen. Influenza in tropical regions. *PLoS Med.* 2006; **3**, e89.
- [21] KB Feshbach, WJ Alonso, V Charu, J Tamerius, L Simonsen, MA Miller and C Viboud. Latitudinal variations in seasonal activity of influenza and respiratory syncytial virus (RSV): A global comparative review. *PLoS One* 2013; **8**, e54445.
- [22] Y Li, RM Reeves, X Wang, Q Bassat, WA Brooks, C Cohen, DP Moore, M Nunes, B Rath, H Campbell and H Nair. Global patterns in monthly activity of influenza virus, respiratory syncytial virus, parainfluenza virus, and metapneumovirus: A systematic analysis. *Lancet. Glob. Health* 2019; **7**, e1031-e45.
- [23] MM Sajadi, P Habibzadeh, A Vintzileos, S Shokouhi, FM Wilhelm and A Amoroso. Temperature and latitude analysis to predict potential spread and seasonality for covid-19, Available at: <http://dx.doi.org/10.2139/ssrn.3550308>, accessed January 2020.
- [24] European Centre for Disease Prevention and Control, Available at: <https://www.ecdc.europa.eu>, accessed January 2020.
- [25] Simplemaps, Available at: <https://simplemaps.com>, accessed January 2020.
- [26] Physical Sciences Laboratory, Available at: <https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.pressure.html>, accessed January 2020.
- [27] C Marzban and S Sandgathe. Cluster analysis for verification of precipitation fields. *Weather Forecast.* 2006; **21**, 824-38.
- [28] S Gokila, KA Kumar and A Bharath. Clustering and classification in support of climatology to mine weather data: A review. *Int. J. Comput. Algorithm.* 2015; **4**, 45-8.
- [29] B Murugesakumar, K Anandakumar and A Bharathi. Improved fuzzy K-means cluster algorithm to analyse weather data in coimbatore region. *Int. J. Adv. Eng. Res. Dev.* 2017; **4**, 840-6.
- [30] W Dong, K Yang, Q Xu, L Liu and J Chen. Spatio-temporal pattern analysis forevaluation of the spread of humaninfections with avian influenza A(H7N9) virus in China, 2013-2014. *MC Infect. Dis.* 2017; **17**, 1-17.

- [31] CR Vicente, KH Herbinger, CC Junior, CM Romano, ASA Cabidelle and G Froschl. Determination of clusters and factors associated with dengue dispersion during the first epidemic related to Dengue virus serotype4 in Vitória, Brazil. *PLoS One* 2016; **12**, e0175432.
- [32] RM McCloskey and AY Poon. A model-based clustering method to detect infectious disease transmission out breaks from sequence variation. *PLoS Comput. Biol.* 2017; **13**, e1005868.
- [33] C Yuan and H Yang. Research on K-value selection method of K-MeansClustering algorithm. *J. Multidiscip. Sci.* 2019; **2**, 227-35.
- [34] JP Ortega, NNA Ortega, and D Romero. Balancing effort and benefit of K-means clustering algorithms in big data realms. *PLoS One* 2018; **13**, 1-19.
- [35] MZ Sáska and S Dombay. Seasons's shifts in some depressions of the Eastern Carpathians, based on daily temperature analysis. *In: Proceedings of the Air and Water–Components of the Environment*, Cluj-Napoca, Romania. 2020, p. 213-22.