Models of transmission of COVID-19 with time under the influence of meteorological determinants

Atin Adhikari*

Department of Biostatistics, Epidemiology,
and Environmental Health Sciences,
Jiann-Ping Hsu College of Public Health,
Georgia Southern University, Statesboro, Georgia 30460, U.S.A.

Shilpi Ghosh†

Department of Biotechnology, University of North Bengal,
Siliguri, West Bengal, PIN: 734013, India

Moon M. Sen‡

Department of Science and Mathematics,
Biology Division, East Georgia State College,
Swainsboro, Georgia 30401, U.S.A.

Rathin Adhikari§

Centre for Theoretical Physics, Jamia Millia Islamia
(Central University), New Delhi-110025, India

Abstract

Based on the statistical analyses of the data on the number of confirmed COVID-19 cases and meteorological determinants in some of the severely affected cities in Spain, Italy and the USA, some models are constructed showing the relationship of I' (the number of infected individuals divided by the total population of a city) with temperature, relative humidity, wind velocity and time. Three models are based on the data before lockdown/travel restrictions in these cities, and the other three models are based on data both before and after lockdown/travel restrictions. These models, in general, indicate that the transmission of COVID-19 could be relatively high either for elevated temperatures with lower relative humidity or for lower temperatures with higher relative humidity conditions. Although one model indicates exponential increase in number of infection cases with time, the more statistically significant models show that the number of cases varies quadratically with time. We have discussed in short, how all these features could be linked with the alterations of structural characteristics of the SARS-CoV-2 virus. Finally, the possibility of natutal disappearance of COVID-19 pandemic, at the global level, has been discussed in the context of the most statistically significant model.

* aadhikari@georgiasouthern.edu

† shilpighosh@nbu.ac.in

‡ mmsen@ega.edu

§ rathin@ctp-jamia.res.in

Corresponding Author: Rathin Adhikari

1. Introduction

The highly pathogenic coronavirus disease 2019 (COVID-19) has become a pandemic after its initial outbreak in Wuhan, Hubei province of China, during December, 2019. According to the recent report of the World Health Organization, the disease has spread to six continents and 210 countries and has affected 4.42 million people with 0.30 million deaths as on 16th May, 2020 (WHO, 2020). The causative organism for COVID-19 is severe acute respiratory syndrome coronavirus 2 (SARS-CoV2), a genus belonging to family Coronaviridae. Clinically, patients with COVID-19 develop respiratory symptoms, which is very similar to other respiratory virus infections. Multiple symptoms may be involved with COVID-19, including respiratory (cough, shortness of breath, sore throat, rhinorrhea, hemoptysis, and chest pain), gastrointestinal (diarrhea, nausea, and vomiting), musculoskeletal (muscle ache), and neurologic (headache or confusion) types. More common signs and symptoms are fever (83% - 98%), cough (76% - 82%), and shortness of breath (31% - 55%) (Wu et al. 2020). Historically, coronaviruses gained prominence during the outbreak of severe acute respiratory syndrome (SARS) during 2002-2003 with severe acute respiratory syndrome coronavirus (SARS-CoV) as the causative agent. The virus infected 8098 individuals with a mortality rate of 9% across twenty-six countries worldwide. In contrast, the incidence of COVID-19 infection has crossed more than 4.42 million to date worldwide, indicating increased transmission ability of SARS-CoV-2 (Cheng et al.,2007; Walls et al., 2020).

According to recent evidence, SARS-CoV-2 virus is primarily transmitted between people through respiratory droplet, direct contact with infected people and indirect contact with surfaces in the immediate environment or with objects used on the infected person (Chan et al., 2020; Li et al., 2020; Liu et al., 2020). Some scientific studies have provided the initial evidence for the viability of SARS-CoV2 virus in aerosols for hours, suggesting their plausible airborne transmission (van Doremalen et al., 2020). Besides, their human transmissibility can be influenced by the environment in which pathogen and host meet (Pica and Bauvier 2012). Like other influenza virus infections, relative humidity and temperature are expected to affect the incidence of COVID-19 particularly through airborne respiratory droplets (Casanova et al., 2010; Pica and Bauvier, 2012; Lowen and Steel, 2014). The temperature and relative humidity could affect the physicochemical characteristics of infectious droplet including, pH and salt concentration of droplets as well as viral membrane lipid and surface proteins and thus influencing their transmission (Lowen and Steel, 2014; Yang and

Marr, 2012). Apart from temperature and relative humidity, the wind velocity also could play a role in the transmission of the virus.

In most of the earlier works related to the spreading of the epidemic, the rate of increase of the number of infected individuals has been studied in relation to the various meteorological determinants (Wang et al., 2020; Prata et al., 2020; Sajadi et al., 2020). However, the rate would vary between cities because of their differences of total population, area and population density. Hence, in the present communication a different approach has been followed to study the spreading of COVID-19 by considering homogeneous mixing in the populations of various major cities (Spain, Italy and the USA) along with the meteorological determinants and population density. Under this consideration, the relevant quantity is the proportionate mixing of the number of infected individuals I, within the total population N of a city, denoted as I' which is equal to $\frac{I}{N}$. In this work, in the context of infection due to SARS-CoV-2 virus, we have found the relationship between I' with temperature, relative humidity, wind velocity and time.

The section 2 of the manuscript mentions the data sources and their collection time period. In section 3, regression models were built up from the statistical analysis of the data on the number of infected persons in various cities, the related meteorological data and the time elapsed after the initial reporting of infected cases. We have found out that how the proportionate mixing I' could depend on temperature, relative humidity, wind velocity, and time in a statistically significant way. In section 4, the results of the statistical analysis corresponding to various models have been presented with the help of different contour plots, which show the combination of temperature, relative humidity and wind velocity for favourable or unfavourable COVID-19 transmission. Also it has been shown that how the number of infected cases for a particular population, could evolve with time under various meteorological conditions. In section 5 some possible biochemical and biophysical explanation for the stability of the virus under different meteorological conditions (as indicated by the models) have been discussed. In section 6, it has been discussed through illustrative example, how these models indicate the possibility of the world to be free from COVID-19 in a natural way due to the change in the meteorological factors in the environment. Finally, in section 7, based on models, some precautionary measures have been discussed.

2. Data sources and time period of data collection

All data used for analysis were available in public databases. The data on the number of COVID-19 infection cases of several major cities in Spain, Italy, and the USA (Countries with relatively severe COVID-19 outbreak) from the following sources - (1) Data of Spain were collected from the database of the Centro de Coordinací on de Alertas y Emergencias Sanitarias (CCAES), Spain. The center is responsible for coordinating information management and supporting the response to national or international health alert or emergencies, (2) Data of Italy were collected from the database of The Ministry of Health (Italian: Ministero della Salute), which is a governmental agency of Italy and is led by the Italian Minister of Health. (3) Data on different cities of USA were collected from USAFacts, which is a not-for-profit, nonpartisan civic initiative providing the most comprehensive and understandable source of government data available in the US. Data on meteorological factors were collected from the World Weather Online database.

We have considered data of the cities Madrid, Catalonia and Pais Vasco-Basque in Spain during the period 4th March, 2020 to 14th March, 2020 before lockdown/travel restrictions and 15th March to 29th March after lockdown/travel restrictions. We have considered data of the cities Milan, Bologna and Venice in Italy during the period 26th February, 2020 to 9th March, 2020 before lockdown and 10th March to 31st March after lockdown. We have considered data of the cities of New York, San Francisco, Atlanta, Seattle, Chicago and Los Angeles in the USA during the period 2nd March, 2020 to 14th March, 2020 before travel restrictions/stay at home order orders. Data of the cities of New York, Chicago, Los Angeles in USA were also considered during the period 16th March to 31st March after travel restrictions/stay at home orders.

3. Regression Models for I'

Although we have considered data of different cities both before and after lockdown/travel restrictions, the global features of the spreading of COVID-19 in relation to climatic conditions should be examined before the implementation of restrictions when there were no human interventions on the transmission of the virus. Therefore in our study, the statistical analyses of data before restrictions have been done separately in which such features are expected to be present. However, assuming that the early parts of restrictions were not so stringent in the selected large cities, the combination of both before and after restrictions

data have also been considered separately for statistical analysis. But while interpreting any results from the analyses of the combined data, one needs to be careful about this assumption.

The major cities of three countries have been chosen because they are expected to have almost homogeneous mixing of the population. On the other hand, small cities have different heterogeneous issues of the population, like change in population density, social behavior and movement of people could be more manifested, which are complicated to be taken into account for the statistical analysis. As our primary concern is to find the response of the SARS-CoV2 to different meteorological factors as well as the global properties of the spreading of the virus over space and time, the major cities with a higher population are more appropriate for the statistical analysis of the data.

In building up regression models, our approach is different from the conventional models like SIR models (Anderson and May,1979; Jong et al., 1995; Dietz 1985) in which S, I and R correspond to the number of susceptible, infected and recovered individuals, respectively. In some modified models, the number of deaths due to viral infections have also been considered. As the aim of our study is to determine the influence of the meteorological factors on the spreading of virus infection, we are mainly concerned with the number of infected individuals I. The number of persons recovered or dead also depends on various other factors, like immunity of the person, hospital facilities nearby, etc. which are not related to the virus explicitly. Although these numbers do influence S in the early period of the pandemic (which we are considering), they can be easily ignored because of being relatively much smaller than I. In finding the connection of the spreading of the virus with the meteorological factors and time, the proportionate mixing $I' = I/N$ will be considered as a parameter for the spreading of virus and, instead of I, the evolution of the parameter I' will be studied in this work. This has the advantage as the upper limit of the quantity I' is normalized to unity for a total population N in any city. It is legitimate to assume that the evolution of I' depends only on the properties of the virus namely, its' response to different meteorological factors and response to the human host. So to find the evolution of I', one may consider together all the data on the number of infected cases corresponding to different major cities.

We have not considered a priori any specific form for evolution of I' with time Ti, like conventional model that takes into account an exponential increase in the number of infected cases with respect to time. The reason behind this is that in the SIR models, the rate

of increase in the number of infected persons I with time Ti will be proportional to the proportion of infectious contacts (I/N). The model assumes that further infections occur due to the direct contact with the infected people and one follows the principle of mass action in chemistry where the proportionate mixing I/N stands for the concentration of some substance. But in the case of viral infections, it should be considered as the contacts of a large number of viruses with the population N to some extent like the scattering of elementary particles. This is because infections will be happening due to virus which could be in the air, which could be on the surface of some material, which could be due to the large number of viruses coming from the sneezing or cough of infected people. Furthermore, viruses coming from one infected person, may infect a large or a few number of people depending on the contacts with the numbers of individuals. That tells that the rate of increase in the number of infected individuals should not be considered to be proportional to the number of already infected persons I. So considering the viral infection problem as the one similar to the law of mass action in chemistry may not be correct and the rate of increase may not be proportional to I/N. In that case, the exponential time dependence which is a solution to I coming from the first order differential equation in I and time Ti in which $\frac{dI}{dTi}$ is proportional to I, may not be correct. For these reasons, in building up regression models, in the time evolution part of I', we have explored other possible functional forms with time Ti as variable. However, we have also explored the possibility of exponential time dependence, although that is not necessarily equivalent to considering that $\frac{dI}{dTi}$ is proportional to I and will be explained further in the later part of the paper.

In the absence of our understanding of the relationship of I' with time Ti at this point of discussion, it is kept as $I' \propto F_1(Ti)$ where $F_1(Ti)$ is some function of time. For building up models, the relationship of I' with various meteorological factors like temperature (T in ^0C), relative humidity (H in percentage) and wind velocity (W in Km/hr) are written in terms of another function $F_2(T, H, W)$ which has no explicit Ti dependence and are separate from F_1 and we consider $I' \propto F_2(T, H, W)$. However, T, H and W may or may not vary with time and may have no explicit time dependence. So we may write I' as

$$I' = F_2(T, H, W)\ F_1(Ti) \tag{1}$$

where the constants of proportionality are absorbed in the functions F_1 and F_2. Corresponding to each additive term on the right hand side, one proportionality constant is expected.

These constants could be related to the interactions of virus with individuals and with environment, its' self replication processes and so on. In the product of the two functions F_1 and F_2, it is expected that each additive terms should have Ti or its higher power as a factor, but T, H and W may or may not be present. This is because at $Ti = 0$, I' is assumed to be zero. The rate R of increase of I' with time can be written, as the total derivative for which $R = R_{tot}$ is written as

$$R_{tot} = \frac{dI'}{dTi} = \frac{dF_1(Ti)}{dTi} F_2(T, H, W) + \frac{dF_2(T, H, W)}{dTi} F_1(Ti). \tag{2}$$

where T, H and W are assumed to vary smoothly with time Ti. However, under the assumption that T, H and W are not changing with time Ti (which is in general not true, but may be considered as an approximation for a few days when the values of T, H and W are not changing significantly) we can write R as partial derivative for which $R = R_{part}$ is written as

$$R_{part} = \frac{\partial I'}{\partial Ti} = \frac{\partial F_1(Ti)}{\partial Ti} F_2(T, H, W). \tag{3}$$

One may note that in our discussion, $\frac{dF_1(Ti)}{dTi} = \frac{\partial F_1(Ti)}{\partial Ti}$. We will refer R as R_{part} in general. On the right-hand side of the above equations all terms are related to the properties of the spreading of virus. To identify these features from our statistical analysis, data on the number of infection cases of different cities were taken by considering $I' = I/N$ corresponding to different T, H, W, Ti values. From the proper fitting with data points, we will determine the two functions F_1 and F_2.

To smoothen the local fluctuations in the data (like say the number of cases recorded after a day or the variations of temperature, relative humidity or wind velocity) over a short period, which should not appear as a global effect, all the data were averaged over a period of three days. These averaged data points were tried to fit with various possible functional forms of F_1 and F_2. We have considered various forms for F_1 and F_2 including linear, bi-linear, tri-linear, non-linear and other simpler forms. For F_1 exponential form has also been tried. For statistical analysis two types of data are considered - one before lockdown/travel restrictions and another combining both before and after lockdown/travel restriction data, as stated earlier. In this work, six possible models are presented - three with before lockdown/travel restrictions and three with both before and after restrictions data. For two models, F_2 is function of only T and H as shown in Table -I. For the other four models (in which for two before lockdown/travel restrictions data and for the other

two both before and after restrictions data has been used), the function F_2 also depends on wind velocity (Table -II).

TABLE I. **Regression models without wind velocity**

Model	Fitted Relationship	Adjusted $\bar{R}^2$
(A) (before & after lockdown/travel restrictions)	$I' = (6.10328 \times 10^{-7}\ T + 6.95678 \times 10^{-8}\ H - 1.04797 \times 10^{-8}\ T\ H)\ Ti^2$ p values: $0.0103923, 0.000104728, 0.0262125$	0.667117
(B) (before lock-down /travel restrictions)	$I' = (-0.0000450058 + 4.65018 \times 10^{-6}\ T + 6.85532 \times 10^{-7}\ H - 6.78549 \times 10^{-8}\ T\ H)\ Ti^2$ p values: $2.68574 \times 10^{-8}, 1.93585 \times 10^{-10}, 1.5796 \times 10^{-8}, 2.6393 \times 10^{-10}$	0.742233

Different regression models with the relationship of I' with T, H and Ti. p values of the numerical coefficients are in the same order of the coefficients in the fitted equations.

In deciding these models we have taken care of mainly four criteria: 1) adjusted $\bar{R}^2$ value which indicates the goodness of fit, 2) the p value for the parameters appearing as coefficients with the various variables in F_1 and F_2 which indicates what is the probability of the values of those parameters not to be so, 3) the randomness in the residue plots (not shown in the paper) indicating somewhat appropriateness of the choice of functional forms, 4) predictive power of the model beyond the data points. We have imposed the condition $p < 0.05$ which determines the model to be statistically significant. For statistical analysis, *Mathematica* (URL: https://www.wolfram.com/mathematica) has been used. Based on above four criteria, we shall discuss at the end, which one out of six models presented here appears more appropriate.

In Table-I, we have presented two models where F_2 only depends on temperature T and relative humidity H but not on wind velocity W. In model (A), we have considered

both before and after lockdown/travel restrictions data for statistical analysis. F_2 depends linearly in T and H with positive coefficients. Then there is bilinear term TH with negative coefficients. This is expected to be related to the condition showing the lack of stability of the virus under certain temperature and relative humidity subject to the strength of the linear terms in T and H. The p values of the parameters are given in the same order as they

TABLE II. **Regression models with wind velocity**

Model	Fitted Relationship	Adjusted $\bar{R}^2$
(C) (before & after lockdown/ travel restrictions)	$I' = (7.78951 \times 10^{-7}\ T + 1.30087 \times 10^{-7}\ H - 1.56195 \times 10^{-8}\ T\ H - 2.94368 \times 10^{-9}\ H\ W)\ Ti^2$ p values: 0.00169275, 0.000116177, 0.00290938, 0.0301774	0.682489
(D) (before lock-down /travel restrictions)	$I' = (1.21246 \times 10^{-6}\ T + 6.96867 \times 10^{-8}\ H - 1.84773 \times 10^{-8}\ T\ H - 2.29219 \times 10^{-7}\ W)\ Ti^2$ p values: 6.99381×10^{-7}, 0.00020347, 8.14599×10^{-7}, 0.0003567	0.60485
(E) (before & after lockdown /travel restrictions)	$I' = (0.0000266886\ T + 4.99888 \times 10^{-6}\ H - 5.93897 \times 10^{-7}\ T\ H - 1.26361 \times 10^{-7}\ H\ W)\ e^{0.118656\ Ti}$ p values: 0.0223086, 0.00746086, 0.0148184, 0.0181516, 7.94451×10^{-12}	0.677819
(F) (before lock-down /travel restrictions)	$I' = (-0.0000385506 + 4.2459 \times 10^{-6}\ T + 6.19105 \times 10^{-7}\ H - 6.26551 \times 10^{-8}\ T\ H - 1.30748 \times 10^{-7}\ W)\ Ti^2$ p values: $7.03166 \times 10^{-7}, 1.03671 \times 10^{-9}, 7.43134 \times 10^{-8}, 8.35951 \times 10^{-10}, 0.00880254,$	0.776309

Different regression models with the relationship of I' with T, H, W and Ti. p values of the numerical coefficients are in the same order of the coefficients in the fitted equations.

appear in the fitted equation in Table-I. As I' varies with Ti^2, the number of infected cases also varies like that. From eqn (3), it follows that $R = R_{part} \propto Ti$, so the rate of increase in the number of infections, also varies linearly in time. Based on the four criteria mentioned earlier, the model (B) seems to be better than the model (A) for which particularly various p values are much smaller than the model (A) indicating much better statistical significance. Also the adjusted $\bar{R}^2$ value is higher in comparison to model (A). Like model (A), in the case of the model (B) also, F_2 has no dependence with wind velocity (W), and the rate $R = R_{part}$ varies linearly in time Ti.

The Table II presents another set of four models in which F_2 depends on wind velocity W also, apart from its' dependence on T and H. In model (C), both before and after lockdown/travel restrictions data have been considered. The F_2 function for this model, has similar linear dependence with T and H like models (A) and (B) in Table -I but besides, it contains two bilinear terms $T\ H$ and $H\ W$. Here, the coefficient of the bilinear term with W has p value just slightly less than 0.05 although p values for other numerical coefficients corresponding to other variables in F_2 are somewhat better. F_1 function is proportional to Ti^2 and $R_{part} \propto Ti$ like models (A) and (B). For the model (D), only the data before lockdown/travel restrictions have been considered. In this case, p values for different numerical coefficients are somewhat better than the model (C) although the goodness of fit reflected through adjusted $\bar{R}^2$ is slightly worse in comparison to model (C). But residuals plot (not shown here) seems better than the model (C). Unlike model (C), for model (D) F_2 is linear in W. F_1 varies as Ti^2 like all other previous models. For model (E), both before and after lockdown/travel restrictions data have been considered. In this case, p values for different numerical coefficients are not that good like model (D). Although goodness of fit is slightly better than the model (D). The functional form of F_2 is very similar to the model (C). However, F_1 has exponential Ti dependence unlike all other previous models. Model (F) is very similar to Model (B). However, unlike model (B), in the case of the model (F) there is W dependence in F_2. It has the best goodness of fit as seen through the highest adjusted $\bar{R}^2$ value among all models. Except for the coefficient of W dependent term in F_2, the p values for other coefficients corresponding to different terms are quite low. F_1 varies as Ti^2 like the other four previous models.

Based on the predictive power of the model beyond data points considered, we have eliminated several models with other possible forms of F_1 and F_2. One the basis of the

other three criteria mentioned earlier, one may draw the following conclusions: considering all six models, it seems that the models having F_2 without wind velocity W, are somewhat better than those with W. Also the models based on before lockdown/travel restrictions data seems to be better. The model (B) appears to be the best among all these models as the goodness of fit with data reflected through $\bar{R}^2$ value is better than all others except model (F), p values of various numerical coefficients are much lower than those for other models, residual plots also shows that the model differs lesser from the observed data points in comparison to other models. However, among models with wind velocity W , model (F), seems to be better. If we consider both before and after lockdown/travel restrictions data, then the model (A) seems to be slightly better because of the lesser number of parameters involved in the model.

Five of these models (A), (B), (C), (D), (F) show that I' varies as Ti^2 whereas model (E) shows I' varies exponentially with time Ti. This means that for the first four of these models, the rate $R = R_{part}$ is

$$\frac{\partial I'}{\partial Ti} = F_2(T, H, W)Ti. \tag{4}$$

This means there is always acceleration in the increase of I' and hence also an aceleration in the increase in the number of infectious individuals and $\frac{\partial^2 I'}{\partial Ti^2}$ is always non-zero. This will be further explained at the end of the result section 4. For the model (E), R is written as

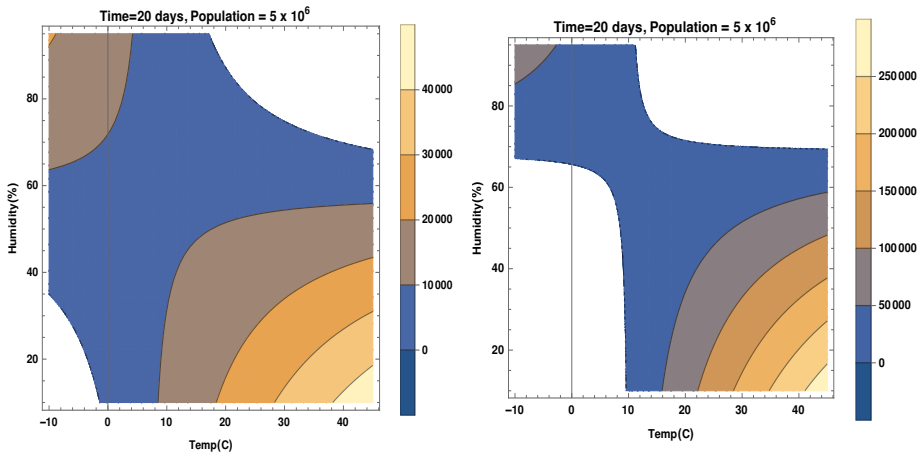
$$\frac{\partial I'}{\partial Ti} = C_2 F_2(T, H, W)e^{C_2 Ti} \tag{5}$$

Here, even the higher derivatives of I' still depends on time. Had F_2 been a constant or 1, in other words, had the number of infectious individuals not depended on T, H, W, in that case only, one could write $\frac{dI'}{dTi} \propto C_3 I'$ which is somewhat like SIR models considered extensively. But from the statistical analysis it is found that F_2 has functional dependence with T, H and also in some models with W and by writing $\frac{dI'}{dTi} \propto C_3 Log[F_2]I'$ one does not get back I' appropriately. So it seems difficult to get SIR kind of models even corresponding to the model (E) which has exponential time dependence.

4. Results for different models

The number of infected cases for a population of 5×10^6 is shown in contour plots in Figure 1 at different temperatures and relative humidities corresponding to model (A) and (B). One may compare different zones of temperature and relative humidity where the virus

is expected to cause more or less infection or no infection. To get the number of cases, I' has been multiplied by the total population N. In Figure 1, we have considered I' after 20 days and have assumed that the same temperature and relative humidity is there for entire 20 days. We have discussed the variations of temperature and humidity just before conclusion. The number of cases is not so important in the figure rather the understanding of how virus is more viable for certain meteorological determinants is important . For example, for lower temperature around 0^0C or below, the number of infected cases is expected to be more with higher relative humidity above about 65 % than that with relative humidity around 30% to 60 %. When the temperature rises above 10^0C then the number of cases may gradually rise with lower relative humidity values with the maximum number of infection cases at a relative humidity below 20% and the temperature above 35^0C. Conversely, with higher relative humidity above 70 % the number of cases is expected to drop significantly. From Figure 1, it follows that according to model (B) around 5^0C to 10^0C unless relative



(Model A) (Model B)

FIG. 1. Contour plots of number of virus-infected persons with temperature and relative humidity in the first 20 days after the initial cases of infection. Different colors in the legend shows different ranges of the number of infected cases. Models are mentioned below each figure.

humidity is above 60%, the probability of viral infection is expected to be very low. On the other hand, for higher temperatures above 15^0C and relative humidity above 70%, the probability of infection is very low. However, for temperatures higher than 20^0C and the relative humidity lower than 40%, the infection probability is very high, even higher than that expected in model (A). The unshaded regions in Figure 1 correspond to absence of viral

infections for the corresponding temperatures and relative humidities.

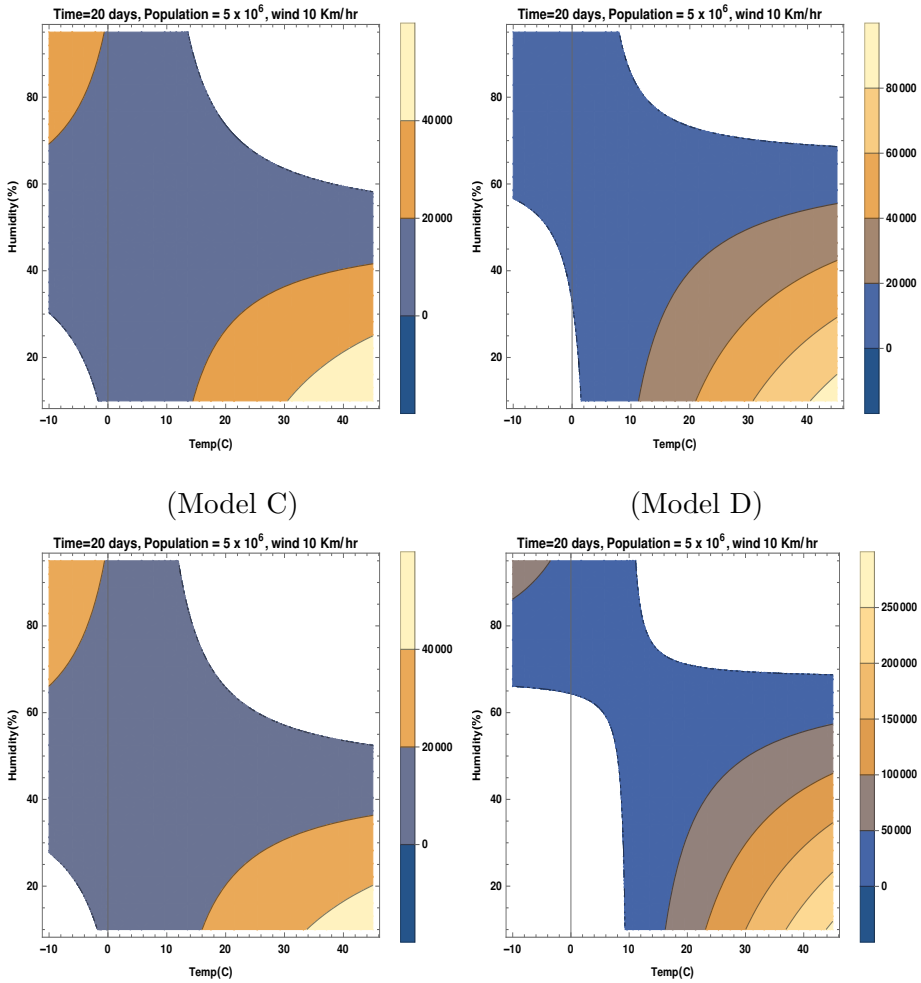


FIG. 2. Contour plots of number of virus infected persons with temperature, relative humidity and wind velocity in first 20 days after the initial cases of infection. Different colors in the legend shows different ranges of the number of infected cases. Models are mentioned below each figure.

From Figure 2 Model (C), it is evident that when wind velocity is considered at 10 Km/hr, the probability of infection with variation in temperature and relative humidity is very much similar to that of model (A) in Figure 1, except that the probability of infection is almost nil for temperature above 30^0C and relative humidity above 60%, whereas for model (A) it is above 70% relative humidity. Also, in Model (C), the probability of infection is almost nil for temperature around -10^0C and relative humidity below 30%, whereas for model (A) it is below 35% relative humidity.

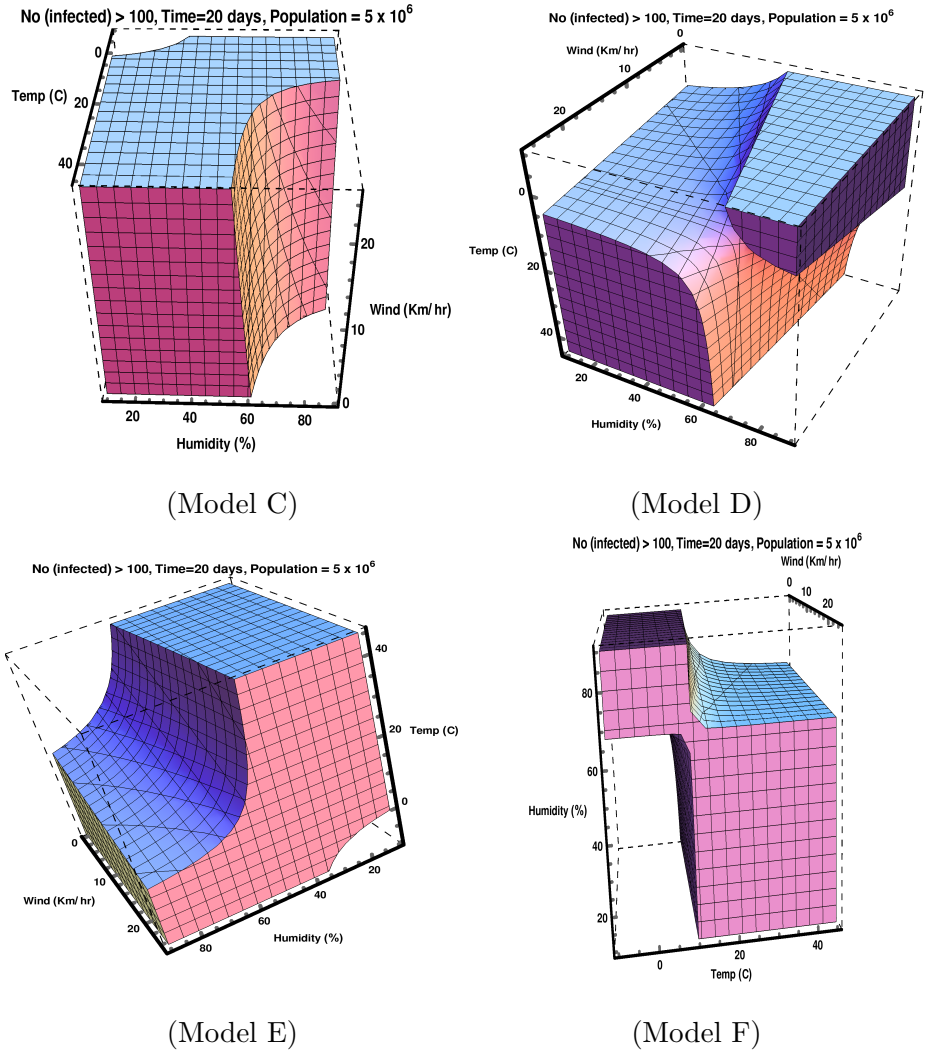


FIG. 3. Allowed region in 3d space of Temperature, relative humidity and wind velocity for which no. of virus infected persons are greater than 100 in first 20 days after the initial spreading. Models are mentioned below each figure.

In Figure 2 Model (D), for a fixed wind velocity of 10 Km/hr, the regions of no viral infections for both low and high temperatures are more in comparison to Model (A) and less in comparison to Model (B). In Figure 2 Model (E), for a fixed wind velocity of 10 Km/hr, the region for no viability of virus infections is more in the high temperature region than all other earlier mentioned models and is less in the low temperature region than all other earlier mentioned models. In Figure 2 Model (F), for a fixed wind velocity of 10 Km/hr, the regions for the viability and no viability of virus infections are almost the same as Model (B). One may note that the functions F_1 and F_2 of Model (F) are like Model (B) except

that F_2 in Model (F) also depends on W linearly.

The effect due to the variation of wind velocity in the range 0 to 25 Km/hr can be seen in Figure 3 for all the models presented in Table II. For these models, the regions for no viability of virus infections, related to lower temperature and lower relative humidity and also related to higher temperature and higher relative humidity; becomes larger for higher wind velocity. Probably it happens because virus are blown away more by the stronger winds even to the areas where hosts do not come into contact with the virus easily. However, statistically the dependance of I' on wind velocity W, seems not that significant like temperature and relative humidity as reflected in the relatively greater values of p parameters corresponding to W terms in comparison to those for non-W terms.

To get the actual number of infected cases I' is to be multiplied by the total population having almost homogeneous mixing and also it is required to take care of the variations of different meteorological factors ($T,\ H,\ W$) with time Ti. We have illustrated how to take care of such variations, in the earlier section. In various Figures except the last one, we have not considered the variations of $T,\ H,\ W$ with time Ti. Although wind velocity appears to be not so significant statistically, but according to Figure 3 Model (F), the higher wind velocity possibly could reduce the transmission of the virus by a small amount as observed in Figure 2 and 3 for Model (F).

We have not analysed the data of infected cases where lock-down/travel restrictions have been imposed quite early like the case of India say. But even in such cases, the functional form of F_2 and F_1 in I', should not differ except that some scaling on I' should be considered. This scaling will differ from place to place depending on the stringent nature of the restrictions. One may find such scaling by taking the ratio of the expected number of infected cases from the models with total population N after taking into account the time Ti and also the variation of various meteorological parameters at different time; to the actual number of infected cases. This could be useful only to estimate approximately the actual number of infected cases in the future. But the general features regarding spreading of the virus as reflected in different figures shown in this work, remain the same. Such scaling should be considered for smaller cities also to understand and estimate approximately the number of SARS-CoV-2-infected cases, where the movement of people are different with respect to large cities.

In Figure 4, the rate of increase of I' written as $\frac{\partial I'}{\partial Ti}$ in equation (4) is shown for three

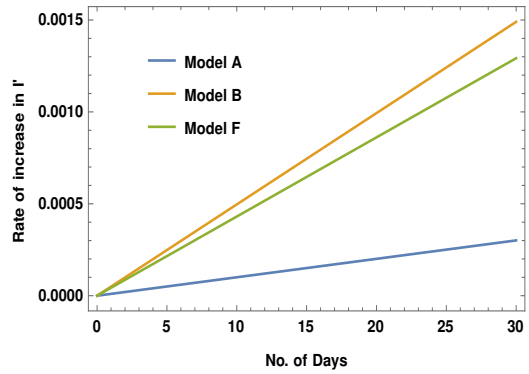


FIG. 4. Plot of $\frac{\partial I'}{\partial Ti}$ versus Ti for temperature 35^0C, relative humidity 55% in models A, B and F. For model F, wind velocity is 15 Km/hr.

statistically significant models (A), (B) and (F) for temperature 35^0C and relative humidity 55%. One can see that $\frac{\partial I'}{\partial Ti}$ is increasing linearly with time. However, the gradients of the straight lines corresponding to different models are different. For the chosen temperature and relative humidity, $\frac{\partial I'}{\partial Ti}$ for model (B), is always higher than those for the model (A) and (F). As stated earlier in section 3, there is non-zero rate of increase of $\frac{\partial I'}{\partial Ti}$ which implies that the increase in I', is associated with some acceleration with time Ti. Let us consider the motion of an object with some acceleration, then total displacement of the object is $\frac{1}{2}f\ Ti^2$ with initial zero velocity. If we compare this displacement with I', then acceleration f associated with I' could be written as $f = 2F_2$. Acceleration is understood from F_1, however, its magnitude is controlled by the environmental factor F_2. So the magnitude of the acceleration depends on the temperature, relative humidity and in some models also on wind velocity. If we consider further, the variation of temperature or relative humidity with time Ti, then $\frac{\partial I'}{\partial Ti}$ will get the further contribution from the second term on the right-hand side of the equation (2). In this case, $\frac{\partial I'}{\partial Ti}$ will vary to some extent with Ti^2 also and the plots of $\frac{\partial I'}{\partial Ti}$ versus Ti will show deviations from the straight lines in the above figure.

5. Exploring Link of F_1 and F_2 with biophysical and biochemical features of SARS-CoV-2

In statistically significant models (A), (B) and (F), F_1- the time evolution part of I', varies as Ti^2. From the relationship of I' with time Ti shown in Table I and II, there is increase of I' with the increase of time Ti. However, as discussed in the context of Figure 4, there is also increase in the rate of increase of I' with time Ti. The increase in I' may be linked with various kind of transmission of the virus. Violent respiratory events like coughs and sneezes help transmitting virus. There could be long-range air-borne transmission of viruses as well as short-range large droplet transmission. Also the virus could be transmitted through direct contact. All these together, will result in an increase in the number of infected persons as well as an increase in I'. Comparing with the motion of an object, the transmission of virus could be thought of as the velocity which is initially zero. The reason behind the increase in the rate of increase of I' with time is linked with F_2 as discussed earlier.

Let us consider the statistically most significant model (B) in which F_2 depends on temperature T and relative humidity H. Our statistical analysis through function F_2 indicates a complex interaction between COVID-19 infection and meteorological parameters. The infection is expected to enhance under two different combinations of environmental conditions i.e. lower temperature-higher humidity and higher temperature-low to ambient humidity as found in Figure 1 Model (B). Such nature of interaction can be explained by the physico-chemical characteristics of respiratory droplets or aerosol and the structural features of the virus itself. The enveloped virus SARS-CoV-2 has a positive sense, single-stranded RNA genome which encode four important structural proteins, namely spike glycoprotein (S), envelope protein (E), matrix glycoprotein (M), and nucleocapsid protein (N). The S glycoprotein is surface exposed and mediates the viral entry into host cells. The S-protein comprises two regions S1 and S2, where S1 is responsible for binding to the receptor of the host cell and S2 is for the fusion of the viral and cellular membrane. Like SARS-CoV, SARS-CoV-2 attaches the host cells through the binding of receptor-binding domain (RBD) of S1 region to the angiotensin converting enzyme 2 (ACE2) (Schoeman and Fielding, 2019; Walls et al. 2020). Although, during infection SARS-CoV and SARS-CoV-2 recognize the same target protein, the pandemic impact of the later happens to be more severe with continuous worldwide increase in the number of infection cases and mortality.

Several factors including, the structural features of virus for persistence in the environ-

ment and human contact might be associated with the severe infection of SARS-CoV-2. Recently, an experimental study based on Cryo-EM structure of the SARS-CoV-2 spike protein in prefusion conformation has suggested higher binding affinity of SARS-COV-2 with human ACE2 than SARS-COV (Wrapp et al., 2020). Another study on the protein-protein interaction and molecular dynamics simulations of RBD-ACE2 complex for SARS-COV-2 and SARS-CoV showed significantly lower binding free energy of the SARS-COV-2 RBD-ACE2 interaction (-50.43 kcal/mol) compared with SARS-CoVRBD-ACE2 interaction (-36.75 kcal/mol) and thus suggesting higher binding affinity of SARS-COV-2RBD-ACE2 interaction (He et al., 2020). Ou et al., (2020) analyzed the SARS-COV-2RBD mutations worldwide and found the equilibrium dissociation constant of three RBD mutants to be two orders magnitude lower than the prototype Wuhan-Hu-1 strain indicating remarkable increase in the infectivity of the mutated viruses.

Apart from the increased binding affinity of SARS-COV-2 to ACE2, the severity of COVID-19 can also be determined by the stability of the virus in the environment. Although most of the studies on the effect of meteorological factors on the COVID-19 have speculated the decline in infection with an increase in environmental temperature and humidity (He et al., 2020; Wang et al., 2020), the number of infection cases in reality is increasing sharply with the rise in temperature. One may note that this increase at higher temperature is the general feature of all the models in Table I and II particularly for lower humidity. The persistence of the virus at higher temperatures can be explained in the context of the stability of the spike protein because it is a critical component determining the infectivity. A lengthy molecular dynamics simulation of trimeric spike proteins of SARS-COV-2 and SARS-CoV has shown that the spike protein of SARS-COV-2 has significantly lower total free energy (-67,303.28 kcal/mol) than the spike protein of SAR-CoV (-63,139.96 kcal/mol). Similarly, the free energy of the RBD of SARS-COV-2 spike protein is relatively lower than that of SARS-CoV. The results thus explained increased stability of SARS-COV-2 spike protein at higher temperature (He et al. 2020). Several studies have been carried out on the effect of relative humidity on the survival and infectivity of enveloped and non-enveloped viruses (Benbough 1971; Shaman and Kohn; Pica and Bouvier, 2012; Yang and Marr, 2012; Marr et al. 2019). When the virus is released into the environment as part of a respiratory fluid droplet, relative humidity of environment controls the amount of water evaporated from the droplet until equilibrium with the surrounding air is maintained/reached (Prussin et al.,

2018). The respiratory viral droplets or aerosol on exposure to lower to ambient relative humidity environment are subject to evaporation due to the vapour pressure gradient between its surface and air. As evaporation proceeds, the water vapour pressure at droplet surface decreases. The presence of lipid membrane in the enveloped viruses has been evidenced to protect their capsids from damage due to change in humidity leading to survival at lower humidity conditions (Benbough, 1971; Yang and Marr, 2012).

Hence, a considerable higher infection level of SARS-COV-2 at low to ambient relative humidity and higher temperature as predicted in the present study could be related to the presence of structural features of the viral envelope and relatively stable spike protein with significantly lower free energy. Moreover, at lower relative humidity and higher temperature the respiratory droplet will undergo evaporation at a higher rate and the desiccated state of droplet will remain unaffected by the changes in temperature, causing higher infection rates. Infection is further enhanced due to more mobility of the smaller droplet size under high temperature and low humidity as such droplets can easily enter the human host. As predicted in this study, the experiments conducted by Prussin et al. (2018) showed that at 37^0C temperature Phi6, a surrogate of influenza and coronaviruses, had the highest infectivity at 20-40% relative humidity. However, in the same experiment the virus showed a significantly higher level of infection throughout the entire range of relative humidity at lower temperature of 14 and 19^0C. In our models for such temperature, infection occurs over a wide range of humidity but the level of infection is not very high.

On the other hand, our statistical analysis also showed an increase in viral infection at low temperature and high relative humidity as observed earlier for seasonality of influenza virus infection, which occurs at a significantly higher level during winters when average outdoors daily temperatures remains lower and relative humidity is higher (Lowen and Steel, 2014). The finding is also supported by the results of Prussin et al. (2018) reporting higher infectivity of Phi6 at lower temperature $(10 - 20^0C)$ when relative humidity was kept constant at 75%. Probably at higher humidity with appropriate droplet size SARS-CoV-2 could remain wet by keeping it at slightly higher temperature with respect to the lower surface temperature of the droplet and thus remains viable for infection.

However, our study indicates a complete reduction in infectivity of the virus under lower temperature and lower humidity condition. Under such condition, probably SARS-CoV-2 virus could not remain wet inside the droplets which are of smaller size and the lipid and

the protein structure inside the virus gets deformed resulting in a reduction in viability.

Furthermore, the predicted decrease in viral infectivity at a higher temperature and higher humidity in our models could be due to the reduced rate of evaporation of respiratory droplets at higher humidity, which consequently makes them highly susceptible to higher temperature resulting in non-viability of SARS-CoV-2 for infection. Furthermore, corresponding aerosols with bigger size falls down to the surface and loses the mobility. This also could be a reason for the reduction in the viability of infection of SARS-CoV-2 at both high temperature and high humidity.

6. Statistical Model-inspired Rise and fall in the spreading of COVID-19

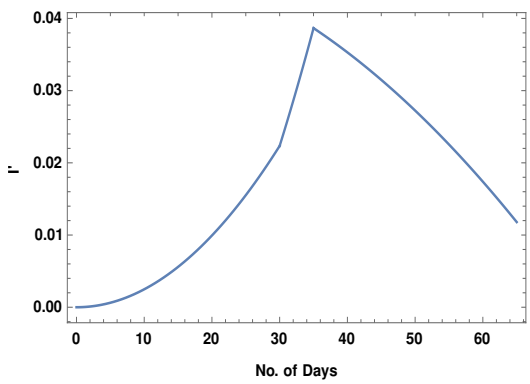


FIG. 5. Plot showing rise and fall of I' implying rise and fall of the spreading of virus SARS-CoV-2 according to Model B due to variation of relative humidity from 55 % in first 30 days to 40 % in the next 5 days and then to 75% after 35 days with temperature always at 35^0C.

In Figure 5, considering most statistically significant model (B), an illustrative example of rise and fall of I' which also means the rise and fall of the spreading of the virus, has been presented. Here, the variation of relative humidity with time Ti for a fixed temperature at 35^0C has been considered. The variation of both the temperature and relative humidity, also could have been considered but for simplicity, we have refrained from showing that. Also for simplicity, we have assumed the change in relative humidity is almost instantaneous. In this example, in the first 30 days, the relative humidity is $H_1 = 55\%$ and in the next 5 days it is $H_2 = 40\%$ and after 35 days it is $H_3 = 75\%$ and the temperature is always at $T_1 = 35^0C$. In general for model (B), $I' = F_2(T, H)Ti$ where $F_2(T, H)$ for Model (B) is given in Table I.

To plot I' in these three time periods the following functions for I' has been used:

$$
\begin{aligned}
I'_1 &= F_2(T_1, H_1)Ti^2 \\
I'_2 &= F_2(T_1, H_2)Ti^2 + \{F_2(T_1, H_1) - F_2(T_1, H_2)\}\, 30^2 \\
I'_3 &= F_2(T_1, H_3)Ti^2 + \{F_2(T_1, H_2) - F_2(T_1, H_3)\}\, 35^2
\end{aligned} \tag{6}
$$

in which I'_1 is to be considered with Ti varying from 0 to 30, I'_2 is to be considered with Ti varying from 30 to 35 and I'_3 is to be considered with Ti varying from 35 to 65. From Figure 1 Model (B), one can see that at temperature 35^0C, the number of infected cases is more for relative humidity 40% than that for relative humidity 55 %. So in Figure 5, there is steeper rise in I' during 30 to 35 days than that during 0 to 30 days. The temperature 35^0C and relative humidity 75% is in the region where viral infections are not viable in Figure 1 Model (B). F_2 is actually negative in this region which should be interpreted as the fall in the spreading of the virus with which I' is related. Once the weather is in such no viable region (which can be seen in Figure 1 and Figure 2 for various models) for the SARS-CoV-2 viral infections for some 10-14 days, one might expect that COVID-19 could go away. In statistical fitting the increase in I' has been considered with time Ti. In Figure 5, the period from 35 to 65 days, corresponds to a fall in the spreading of the virus. We have considered that the function F_1 (which indicates the time evolution) and F_2 (which indicates the interactions of virus with the environment), will remain the same even during this period. Figure 5 is an illustrative example of how large cities could be free from COVID-19 in a natural way due to the change in the meteorological factors in the environment.

7. Conclusion

Based on statistical significance, the model (B) can be considered as most appropriate to understand the future outcome of COVID-19 in big cities. From Figure 1 model (B), it appears that higher precautionary measures would be required in the cities during summer season with higher temperature and lower relative humidity, and during winter with low temperature and high relative humidity. On the other hand, the viability of SARS-CoV2 seems to be reduced at higher temperature with higher humidity and at lower temperature with lower humidity conditions in the environment, as depicted by the unshaded regions of Figure 1 and Figure 2. The model (B) suggests that keeping the relative humidity below 60% in low temperature areas in hospitals and labs, such as in a cold room where temperature

is below 10^0C, could prevent the viral infection among staff working in those conditions. In general, during summer days the relative humidity is higher and the temperature is lower in the evening and therefore, based on the model result, going out of home in the evening would be safer than that around noon.

From data analysis on the spreading of COVID-19 virus, it is found that the increase in number of infected cases depends quadratically in time in a statistically significant way. This indicates the involvement of a certain kind of acceleration in the spreading of the virus which is related to the function F2. These features are expected for the spreading of any type of viruses which infect human host. However, further studies on exploring such acceleration and the relationship of F2 with the structural features of SARS-CoV2 would be worthwhile.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We are grateful to USAFacts, U.S.A., the Centro de Coordinać ı on de Alertas y Emergencias Sanitarias (CCAES), Spain, Ministero della Salute, Italy, and World Weather Online for providing access to their collected data. No funding was received for this study.

References

WHO. 2020. Coronavirus disease (COVID-2019) Situation Report, 117

Wu, Y.C., Chen, C.S., Yu, J. 2020. The outbreak of COVID-19: An overview. J. Chinese Medical Assoc. 83, 217-220.

Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T., Veesler, D. 2020. Structure, function, and antigenicity of the SARSCoV-2 spike glycoprotein. Cell, 180, 1-12.

Cheng, C. C.V., Lau, K.P.S., Woo, C. Y. P., Yuen, K.Y. 2007. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. Clin. Microbiol. Rev. 20, 660–694.

Chan, J., Yuan, S., Kok, K.H., Wang, K.K., Chu, H., Yang, J., et al. 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 395, 514-523. doi: 10.1016/S0140-

6736(20)30154-9

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., et al. 2020. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. New Engl. J. Med. 382, 1199-1207.

Liu, J., Liao, X., Qian, S., Yuan, J., Wang, F., Leu, Y., et al. 2020. Community transmission of severe acute respiratory syndrome coronavirus 2, Shenzhen, China. Emerg. Infect. Dis.26 (6), doi.org/10.3201/eid2606.2002 39

van Doremalen, N., Bushmaker, T., Morris, D., Holbrook, M.G., Gamble, A., Williamson B.N., et al. 2020. Aerosol and surface stability of SARS-CoV- as compared with SARS-CoV-1. New Engl. J. Med. 382, 1564-1567. doi: 10.1056/NEJMc2004973

Pica, N., Bouvier, N.M. 2012. Environmental factors affecting the transmission of respiratory viruses. Curr. Opinion Virol. 2, 90–95.

Casanova, L.M., Jeon, S., Rutala, W.A., Weber, D.J., Sobsey, M.D. 2010. Effects of air temperature and relative humidity on coronavirus survival on surfaces. App. Environ. Microbiol. 76, 2712–2717.

Lowen, A.C., Steel J. 2014. Roles of humidity and temperature in shaping influenza seasonality. J. Virol. 88, 7692–7695.

Yang, W., Marr, L.C. 2012. Mechanisms by which ambient humidity may affect viruses in aerosols. App. Environ. Microbiol. 78, 6781–6788.

Wang, M., Jiang, A., Gong, L., Lu, L., Guo, W., Li, C., et al. 2020. Temperature significantly change COVID-19 transmission in 429 cities. medRxiv preprint doi: https://doi.org/10.1101/2020.02.22.20025791

Prata, D. N., Rodrigues , W., Bermejo P.H.2020. Temperature significantly changes COVID-19 transmission in (sub) tropical cities of Brazil. Sci. Tot. Environ. 729, 138862

Sajadi, M.M., Habibzadeh, P., Vintzileos, A., Shokouhi, S., Miralles-Wilhelm, F., Amoroso,A. 2020. Temperature, humidity, and latitude analysis to predict potential spread and seasonality for COVID-19. http://dx.doi.org/10.2139/ssrn.3550308.

Anderson, R. M. and May, R. M., 1979. Population biologies of infectious diseases: Part I, Nature, 280, 361-367.

Jong M. de , Diekmann O. , Hkeesterbee H. , How does transmission of infection depend on population size, Epidemic Models: Their Structure and Relation to Data (1995) 84.

Dietz K. , Schenzle D. , Proportionate mixing models for age-dependent infection trans-

mission, Journal of Mathematical Biology 22 (1985) 117.

Schoeman, D., Fielding, B.C. 2019. Coronavirus envelope protein: current knowledge. Virology J. 16, 69

Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C. L., Abiona, O., Graham, B.S., McLellan, J.S. 2020. Cryo-EM Structure of the 2019-nCoV spike in the prefusion conformation. Science 367, 1260–1263.

Ou, J., Zhou, Z., Zhang, J., Lan, W., Zhao, S., Wu, J., et al. 2020. BD mutations from circulating SARS-CoV-2 strains enhance the structure stability and infectivity of the spike protein. bioRxiv, Preprint server for Biology doi: https://doi.org/10.1101/2020.03.15.991844

He, J., Tao, H., Yan, Y., Huang, S-Y., Xiao, Y. 2020. Molecular mechanism of evolution and human infection with SARS-CoV-2. Viruses 2020, 12, 428; doi:10.3390/v12040428

Benbough, J.E. 1971. Some factors affecting the survival of airborne viruses. J. Gen. Virol. 10, 209 –220.

Yang, W., Marr, L.C. 2012. Mechanisms by which ambient humidity may affect viruses in aerosols. Appl. Environ. Microbiol. 78, 6781- 6788.

Prussin, A.J., Schwake, D. O., Lin, K., Gallagher, D.L., Buttling, L., Marr, L.C. 2018. Survival of the enveloped virus Phi6 in droplets as a function of relative humidity, absolute humidity, and temperature. App. Environ. Microbiol. 84, e00551-18

Lowen, A.C., Steel, J. 2014. Roles of humidity and temperature in shaping influenza seasonality. J. Virol. 88, 7692-7695.

Shaman, J., Kohn, M. 2009. Absolute humidity modulates influenza survival, transmission, and seasonality. Proc. Nat. Acad. Sci. 106, 3243–248.

Marr LC, Tang JW, Van Mullekom J, Lakdawala SS. 2019. Mechanistic insights into the effect of humidity on airborne influenza virus survival, transmission and incidence. J. R. Soc. Interface 16:20180298.

MATHEMATICA REFERENCE: Mathematica 12.1 Title: Mathematica Edition: Version 12.1 Publisher: Wolfram Research, Inc. URL: https://www.wolfram.com/mathematica