Correcting pandemic data analysis through environmental fluid dynamics

Cite as: Phys. Fluids 33, 067116 (2021); doi: 10.1063/5.0055299
Submitted: 27 April 2021 · Accepted: 17 May 2021 ·
Published Online: 22 June 2021

Talib Dbouka and Dimitris Drikakis

AFFILIATIONS
University of Nicosia, Nicosia CY-2417, Cyprus

Note: This paper is part of the special topic, Flow and the Virus.
a) Electronic mail: dbouk.t@unic.ac.cy
b) Author to whom correspondence should be addressed: drikakis.d@unic.ac.cy

ABSTRACT
It is well established that the data reported for the daily number of infected cases during the first wave of the COVID-19 pandemic were inaccurate, primarily due to insufficient tracing across the populations. Due to the uncertainty of the first wave data mixed with the second wave data, the general conclusions drawn could be misleading. We present an uncertainty quantification model for the infected cases of the pandemic's first wave based on fluid dynamics simulations of the weather effects. The model is physics-based and can rectify a first wave data's inadequacy from a second wave data's adequacy in a pandemic curve. The proposed approach combines environmental seasonality-driven virus transmission rate with pandemic multiwave phenomena to improve statistical predictions' data accuracy. For illustration purposes, we apply the new physics-based model to New York City data.

I. INTRODUCTION

Modeling and analysis of global epidemiology are very challenging.1 The COVID-19 pandemic has been widely spreading due to airborne virus transmission.2–5 The coronavirus pandemic has impacted the economics and the environment at a global worldwide scale6,7 since March 2020.

The daily number of infected cases, reported during the first wave of the pandemic, was inaccurate due to several factors:

• Insufficient tracing across the population.
• Inaccuracy of testing equipment.8
• Lack of reproducible confirmation tests.
• Inaccuracy of management by public and private health institutions.

Many online digital libraries, platforms, and institutions continue to utilize inaccurate data from the first wave. However, statistical comparisons between the first and second wave could lead to misleading conclusions for the reasons mentioned above.

The number of deaths is a more reliable source of information than the number of infected cases. This has been a topic of recent debate.9,10 Ioannidis et al.9 investigated the second vs the first wave of COVID-19 deaths. From available data on the age vs the number of deaths, they correlated the second vs the first wave of COVID-19 deaths to the shifts in age distribution and nursing home fatalities.

Graichen10 investigated the difference between the first and the second (and third wave) of COVID-19 to form a German and European perspective. Other research works emerge along similar lines. Yue et al.11 tried to estimate the actual size of a pandemic from surveillance systems. Stamatakis et al.12 and Huang13 investigated pandemic data reported in the United Kingdom to examine associations between lifestyle risk factors and mortality outcomes. The above is due to an unknown lack of representativeness that can affect the magnitude and direction of effect estimates.

We recently showed that two pandemic outbreaks would be inevitable due to environmental weather seasonality.14 Our findings were based on high-fidelity, multiphase fluid dynamics, and heat and mass transfer simulations of airborne virus transmission.15 We combined the simulation results with epidemiological modeling enhanced by a new airborne infection rate index (AIR) and meteorological data.14 It is generally believed that the high-temperature and high-humidity environment is conducive to reducing the transmission rate of the new coronavirus. In our previous research,14,15 we showed a complex mechanism associated with the effects of weather conditions on virus transmission. It incorporates combinations of three major parameters: relative humidity, temperature, and wind speed.

The daily infected cases’ published data during the first wave is inaccurate and thus can be misleading. We develop a new uncertainty
quantification model using environmental-climate data that corrects the daily number of infected cases during the first wave, thus improving the second and first waves’ comparative analysis. For illustration purposes, we apply the new model to correct the first wave data reported for NYC between March 2020 and March 2021 for the daily number of infections.

II. MATHEMATICAL MODEL DEVELOPMENT

The pandemic's second wave data constitute a more reliable source of information than the pandemic's first wave's data recorded starting from early March 2020.

In March 2021, thus after one year’s period of the COVID-19 pandemic, we know the following:

- The order of magnitude of the total number of deaths from COVID-19 did not change significantly between the first and the second wave periods.16
- The order of magnitude of the number of patients in intensive care units (ICUs) did not change importantly between the first and the second wave periods.16

Given the above, one can define an accurate mortality rate, $\Psi$, in an infected population as the following:

$$\Psi(n) = \beta(n) \times \frac{D_c(n)}{I_c(n)}.$$  

Moreover, $\Psi'(n)$ could be considered as an inaccurate representation of $\Psi(n)$ which is per unit time and depends on the wavenumber ($n$) of the pandemic such that

$$\Psi'(n) = \beta(n) \times \frac{D_c(n)}{I_c(n)}.$$  

Note that the inaccuracy of $\Psi'$ is placed only in $I_c$. $n = 1$ and $n = 2$ denote the first and second wave of the pandemic; $I_c$ and $D_c$ are the stationary cumulative number of infected cases and deaths due to the infection; $\beta$ is the transmission rate of infection per unit time inside a general population $N_p$. In other words, $\beta$ represents the probability per unit time that a susceptible individual becomes infected.

Liu et al.17 shed light on the role of seasonality in the spread of the COVID-19 pandemic. Dbouk and Drikakis14 quantified the relationship between the weather seasonality and the transmission rate $\beta$ through extensive fluid mechanics simulations for crucial weather parameters such as temperature, relative humidity and wind. They quantitatively showed how a weather-dependent $\beta$ could produce two pandemic waves during one year. Our analysis and modeling approaches are based on the following premises:

1. The weather seasonality is a major driving force behind two pandemic waves occurred annually.
2. The virus fatal strength level did not change importantly between two similar seasons between the first and second wave of the pandemic, i.e., winter 2020 vs winter 2021.
3. The social behaviors and global restriction strategies did not change significantly between the two waves’ periods (i.e., masks, social distance, lockdowns, etc.).
4. The age-pyramid of a country did not change significantly between the two pandemic waves.
Using the above, the mortality rate among the stationary cumulative number of infected individuals should be approximately the same in two consecutive waves of the pandemic, i.e.,
\[
\Psi(n, t) = \Psi(n + 1, t + \Delta t) \quad \text{if} \quad \beta(n, t) = \beta(n + 1, t + \Delta t),
\]
where $\Delta t = t_n - t_{n-1}$ denotes the $n^{th}$-wave period of a pandemic. The final time $t_n$ is determined by
\[
\left| \frac{\partial \beta(n)}{\partial t} \right| \leq \epsilon,
\]
where $\epsilon$ is a positive infinitesimal value. It is worth noting that $\left| \frac{\partial \Psi(n)}{\partial t} \right| = \left| \frac{\partial \Psi(n)}{\partial t} \right|$. It is well established that the non-cumulative number of infected cases $I(t)$ reported during the first pandemic wave is inaccurate, primarily due to the insufficient tracing across the population. Therefore, we aim to correct the inaccurate (old) data, $I(t)$, to obtain the more accurate (new) data, $I_n(t)$.

\[
\Psi'(t) = \beta(1) \times \frac{\Psi(1)}{I(1)}.
\]

Substituting $n = 1$ in Eq. (3) and using Eqs. (1) and (5) yield

\[
I(1) = \frac{\Psi'(1)}{\Psi(2)} \times I(1).
\]

III. NYC CASE

Figure 1 shows the weather-dependent transmission rate ($\beta$) in NYC between March 2020 and March 2021. A maximum transmission rate of 0.5 per day$^{-1}$, related to the coronavirus airborne concentration rate, means that the probability is $P = 1$ (100%) for a susceptible individual infected in two days due to the weather conditions (wind speed, temperature, and relative humidity). Figure 1(a) shows the NYC weather history data with the hat symbol denoting daily weather data averaged per month. Figure 1(b) illustrates the weather-dependent transmission rate (airborne infection rate index ($AIR = \beta$)), showing three trends labeled as high, medium, and low separated by the respective threshold values of 0.4 and 0.3; see Dbouk and Drikakis. The reported pandemic curves data for NYC are shown in Fig. 2 as cumulative number of infections [Fig. 2(a)] and cumulative number of deaths [Fig. 2(b)]. The maximum stationary values for the cumulative infected cases are $I_{c(1)}$ and $I_{c(2)}$ for the first and second waves. Similarly, the cumulative number of deaths is denoted by $D_{c(1)}$ and $D_{c(2)}$ (corresponding to null slopes approximately). Figure 2(c) shows the weather-dependent transmission rate ($\beta$) as obtained by Dbouk and Drikakis$^{14}$ with $\beta(1)$ and $\beta(2)$ representing the transmission rate

![NYC](image)

**FIG. 2.** The first ($n = 1$) and second ($n = 2$) waves of the COVID-19 pandemic in NYC between March 2020 and March 2021. (a) The cumulative number of infections ($I_{c}$); (b) the cumulative number of deaths ($D_{c}$); (c) the weather-dependent transmission rate ($\beta$) as obtained by Dbouk and Drikakis. $\beta(1)$ and $\beta(2)$ are the transmission rate during the first and the second wave of the pandemic, respectively. The pandemic data can be downloaded free of charge from.

![NYC](image)

**FIG. 3.** The mortality rate $\Psi$, per unit time, in the infected population of NYC between March 2020 and March 2021 showing the two waves of the pandemic. An inaccurate $\Psi'(1)$ is observed different from an accurate $\Psi'(2)$. This asymmetry between $\Psi'(1)$ and $\Psi'(2)$ is a measure of uncertainty quantification of the first wave data’s inadequacy concerning the second wave data’s adequacy. The daily number of infected cases $I'(1)$ reported during the first pandemic wave also needs to be corrected according to $\Psi'(1) \approx \Psi'(2)$. 
during the first and the second wave of the pandemic, respectively. This pandemic data can be downloaded free of charge from. \[ \text{\cite{18}} \] It can be observed from Fig. 2(c) that the symmetry line is well aligned with the maximum stationary values of the cumulative number of deaths \( I_c (1) \) and with the maximum cumulative number of infections \( I_f (1) \). This sheds light on the physics-based behavior of the computed transmission rate \( \beta \) of Fig. 2(c) driven by the force of weather seasonality in NYC. \[ \text{\cite{14}} \]

Figure 3 shows the mortality rate in the infected population of NYC between 03 March 2020 and 03 March 2021, computed using Eq. (1) with \( n = 1 \), and Eq. (2) with \( n = 2 \). If both waves data were accurate, we should have \( \Psi (1) \approx \Psi (2) \). \[ \Psi (1) \text{ is, however, different from } \Psi (2). \] Thus, any asymmetry between \( \Psi (1) \) and \( \Psi (2) \) is a measure of uncertainty quantification associated with the first wave. The above prompt to correct the inaccurate daily number of infected cases \( I_f (1) \) reported during the first pandemic wave.

The corrected first wave data for the daily number of infected cases in NYC are shown in Fig. 4. One can see that the correction results in increasing the infected cases fourfold.

IV. CONCLUSIONS AND PERSPECTIVES

The coronavirus pandemic data for the daily number of infections vs the number of deaths during the first wave were incomplete. They lead to misleading conclusions if considered as a data reference. The data inaccuracy for the first wave was primarily due to insufficient tracing across the population. Unfortunately, various online digital libraries and platforms continue to adopt, host, and diffuse these inaccurate data from the first wave, followed by more accurate data of the daily number of infections from the second and subsequent waves.

We proposed a new fluid dynamics, physics-based uncertainty quantification, and correction model that rectifies the first wave data’s inadequacy. As an illustration example, we applied the new model to correct the pandemic’s first wave data for the daily number of infected cases reported in NYC, USA. The proposed model is limited to regions that witnessed more than one pandemic wave. It can be used to correct their first wave data reported for the daily number of infected cases.

Environmental temperature and humidity affect the ability of the virus to infect, but they are not themselves the decisive factor in preventing the spread of the virus. We cannot rely on seasonal temperature rises to suppress the epidemic. Instead, we should focus more on the formulation and implementation of active epidemic prevention control policies. Social protective measures such as social distancing and face masks will remain important during a pandemic.

ACKNOWLEDGMENTS

The authors would like to thank the Editor-in-Chief and Physics of Fluids staff for their assistance during the peer-review and publication of the manuscript.

APPENDIX A: MATHEMATICAL DERIVATION

The reader can find the computational models and the associated assumptions in previous studies published by the authors. \[ \text{\cite{3,14}} \] For the carrier bulk multiphase fluid mixture, we have employed the compressible multiphase mixture Reynolds-averaged Navier–Stokes equations in conjunction with the \( k – \omega \) turbulence model in the shear-stress-transport formulation. The governing equations are detailed in many textbooks. \[ \text{\cite{19,20}} \]

The models accounted for the concentration variation in saliva and predicted airborne virus concentration in expelled saliva droplets under different environmental conditions. From hundreds of computational fluid dynamics simulations, we developed a reduced-order model (ROM) as a new virus airborne infection rate (AIR) index that is directly proportional to the virus concentration rate (CR). AIR is employed to quantify the potential of airborne coronavirus survival under different climate conditions (average temperature, relative humidity, and wind speed) in several worldwide cities.

1. Airborne virus particles in saliva: Concentration Rate

For an initial uniform distribution of virus particles, the concentration, \( C \), decreases in each airborne saliva droplet as a function of time at different proportions and different rates such that

\[ C / C_0 = e^{-7.5 \sqrt{D_0 t}} / d_0 \] \[ (A1) \]

where \( d_0(t) \) is the saliva droplet diameter that varies with time depending on the evaporation rate and the saliva droplets size distribution. \( D_0(t) \) is the time-dependent diffusion coefficient of a virus particle in a saliva droplet given by \( D_0(t) = k_B T(t) / 3 \eta u(t) d_0 \). \( d_0 \) is the virus capsid external mean diameter (\( d_0 = 100 \text{nm} \)), \( T(t) \) the time-dependent effective temperature of the saliva droplet, \( k_B \)
the Lattice Boltzmann constant, and $\mu(t)$ the time-dependent liquid saliva viscosity.

A positive concentration rate (CR) is defined as follows:

$$CR = \partial(C/C_0)/\partial t + 0.5.$$  \hfill (A2)

2. Airborne Infection Rate

The CR is directly proportional to the virus survivability. It provides an appropriate indicator for the airborne transmission, which is defined as an “Airborne Infection Rate (AIR)”: \begin{equation}
AIR = CR.
\end{equation} \hfill (A3)

The CR values between 0 and 0.5 are bounded between 0 and 1 using the operator (\langle \rangle), which transforms a dimensional physical variable $\xi$ into a dimensionless one denoted by $\xi'$ such that

$$\xi' = \frac{\xi - \min(\xi)}{\max(\xi) - \min(\xi)}.$$ \hfill (A4)

where min and max are the minimum and maximum values of $\xi$, respectively.

Many results are obtained for $CR'$ at different weather conditions from several advanced computational fluid dynamics multiphase simulations. Then, all the data points are well fitted as a function of the relative humidity (RH), the wind speed (U), and the temperature (T),

$$CR' = F \cdot (RH' + U') \cdot \sin(T') + \cos(T'),$$ \hfill (A5)

where $F$ is given by

$$F = 0.125(1 - (2T^* - 1)^2).$$ \hfill (A6)

Note that $CR'$ can be transformed back into $CR$ using Eq. (A4) with $\min(CR) = 0, \max(CR) = 0.5$.

In the Subsection, we will show how CR can be incorporated into epidemiology models, e.g., through a weather-dependent transmission rate $\beta$.

3. Weather-dependent epidemiological model

The extensive high-fidelity simulations led to $CR = AIR$ as a function of $T, RH, U$. We consider AIR as a good indicator for airborne virus transmission and proposed it as a flow physics relevant parameter in epidemiological models.

As a physics-based simulation model, we considered the standard a standard SIR model \footnote{D. M. Rao, A. Chernyakhovsky, and V. Rao, “Modeling and analysis of global epidemiology of avian influenza,” Environ. Modell. Software 24, 124–134 (2009).} given by

$$\frac{dS}{dt} = -\beta S I/N,$$ \hfill (A7)

$$\frac{dI}{dt} = \beta S I/N - \gamma I,$$ \hfill (A8)

$$\frac{dR}{dt} = \gamma I.$$ \hfill (A9)

$\beta = AIR$ is a physics-based weather-dependent parameter transmission rate, and $\gamma$ is the recovery rate coefficient that depends on the individual’s health and immunity system. The $\beta$ and $\gamma$ parameters represent the probability per unit time that a susceptible individual becomes infected and the probability per unit time that an infected person becomes recovered and immunized. $t$ is time, and $N$ is the population number. $S, I,$ and $R$ are the number of susceptible, infected, and recovered individuals, respectively. $\beta$ represents the probability per unit time that a susceptible individual becomes infected. $\gamma$ represents the probability per unit time that an infected person becomes recovered and immune.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the authors.

REFERENCES

1. D. M. Rao, A. Chernyakhovsky, and V. Rao, “Modeling and analysis of global epidemiology of avian influenza,” Environ. Modell. Software 24, 124–134 (2009).

2. J. L. Domingo, M. Marques, and J. Rovira, “Influence of airborne transmission of SARS-CoV-2 on COVID-19 pandemic. A review,” Environ. Res. 188, 109861 (2020).

3. T. Dbouk and D. Drikakis, “On coughing and airborne droplet transmission to humans,” Phys. Fluids 32, 053310 (2020).

4. T. Dbouk and D. Drikakis, “On respiratory droplets and face masks,” Phys. Fluids 32, 063303 (2020).

5. T. Dbouk and D. Drikakis, “On airborne virus transmission in elevators and confined spaces,” Phys. Fluids 33, 011905 (2021).

6. Q. Wang and X. Han, “Spillover effects of the united states economic slowdown induced by COVID-19 pandemic on energy, economy, and environment in other countries,” Environ. Res. 196, 110363 (2021).

7. M. F. Bashir, B. J. M. Bilal, B. Komal, M. A. Bashir, T. H. Farooq, N. Iqbal, and M. Bashir, “Correlation between environmental pollution indicators and COVID-19 pandemic: A brief study in californian context,” Environ. Res. 187, 109652 (2020).

8. C. F. Manski, “Bounding the accuracy of diagnostic tests, with application to COVID-19 antibody tests,” Epidemiology 32, 162–167 (2021).

9. P. Ioannidis, C. Axors, and D. G. Contopoulos-Ioannidis, “Second versus first wave of COVID-19 deaths: Shifts in age distribution and in nursing home fatalities,” Environ. Res. 195, 110856 (2021).

10. H. Graichen, “What is the difference between the first and the second/third wave of COVID-19?—German perspective,” J. Orthop. Trauma 32, A1 (2021).

11. B. Yue, H. E. Clapham, and A. R. Cook, “Estimating the size of a COVID-19 epidemic from surveillance systems,” Epidemiology 31, 567–569 (2020).

12. E. Stamatakis, K. Owen, L. Shepherd, B. Drayton, M. Hamer, and A. Bauman, “Is cohort representativeness passe? Matching the UK biobank sample to source population characteristics and recalculating the associations between lifestyle risk factors and mortality,” Epidemiology 32, 179–188 (2021).

13. Y. Huang, “Representativeness is not representative,” Epidemiology 32, 189–193 (2021).

14. T. Dbouk and D. Drikakis, “Fluid dynamics and epidemiology: Seasonality and transmission dynamics,” Phys. Fluids 33, 021901 (2021).

15. T. Dbouk and D. Drikakis, “Weather impact on airborne coronavirus survival,” Phys. Fluids 32, 093312 (2020).

16. WHO, “Who coronavirus disease (COVID-19) dashboard data,” https://covid19.who.int/ (2021).

17. X. Liu, J. Huang, C. Li, Y. Zhao, D. Wang, Z. Huang, and K. Yang, “The role of seasonality in the spread of COVID-19 pandemic,” Environ. Res. 195, 110874 (2021).

18. See https://covidtracking.com/data/state/new-york for “The COVID Tracking Project at The Atlantic.”

19. C. Yeoh and J. Tu, Computational Techniques for Multi-Phase Flows (Elsevier, 2010).

20. D. Drikakis and W. Rider, High-Resolution Methods for Incompressible and Low-Speed Flows (Springer, 2015).

21. W. O. Kermack, A. G. McKendrick, and G. T. Walker, “A contribution to the mathematical theory of epidemics,” Proc. R. Soc. London. Ser. A 115, 700–721 (1927).