Engineered Science
DOI: https://dx.doi.org/10.30919/es8d797

A Novel Bio-System Reliability Approach for Multi-State COVID-19 Epidemic Forecast

Oleg Gaidai[1] and Yihan Xing[2,*]

Abstract

COVID-19 was reported to spread worldwide with specific mortality and burdened worldwide public health. Due to the non-stationarity and complicated nature of COVID-19, it is challenging to model such a phenomenon, delivering reliable long-term forecasts of the extreme death rate. The present study describes a novel bio-system reliability approach, particularly suitable for multi-regional environmental health systems, observed over a sufficient period, resulting in a reliable long-term forecast of the COVID-19 registration rate. This study analysed COVID-19 patient numbers from different US states, constituting an example of a multi-state model observed during the years 2020-2022. Traditional statistical methods dealing with temporal observations of multi-regional processes do not have the advantage of dealing efficiently with extensive regional dimensionality and cross-correlation between different regional observations. The present study presents a novel statistical method to analyse raw clinical data using a multicenter, population-based and medical survey data-based bio-statistical approach. Non-dimensional parameter λ is introduced to measure multi-state risk level. The suggested method predicts a 100-year return period risk level along with its 95% confidence interval band. This methodology can be used in various public health applications based on clinical survey data.

Keywords: COVID-19; Epidemic outbreak; Probability forecast; Public health; Mathematical biology.
Received: 30 July 2022; Revised: 11 November 2022; Accepted: 29 November 2022.
Article type: Research article.

1. Introduction

Statistical aspects of COVID-19 and other similar recent epidemics were receiving much attention in the modern research community.[1-7] Generally, it is quite challenging to calculate realistic biological system reliability factors and outbreak probabilities under actual epidemic conditions by using conventional theoretical statistical methods.[8-12] The latter is usually due to many degrees of system freedom and random variables governing dynamic biological systems spread over extensive terrain. In principle, the reliability of a complex biological system may be accurately estimate straightforwardly by having enough measurements or by direct Monte Carlo simulations. For COVID-19, however, the only available observation numbers are limited by the beginning of the year 2020. Motivated by the latter argument, the authors have introduced a novel reliability method for biological and health systems to predict and manage epidemic outbreaks more accurately. His study focused on COVID-19 epidemics in the USA[13-45] focusing on cross-correlations between different provinces within the same climatic zone. For other studies related to statistical variations per country, see[46] USA East coast was chosen because of its COVID-19 origin and extensive health observations and related research available online.[46-86]

Statistical modelling of lifetime data or extreme value theory (EVT) is widespread in medicine or engineering. For example, EVT was to estimate the demographic of various populations.[1] In [48,49], authors used EVT to estimate the probability of an influenza outbreak in the USA East coast. The authors demonstrated a forecasting prediction potential amid the epidemic in the present study. While similarly[50,51] used EVT to predict and detect anomalies of influenza epidemics. As there is not much statistical research done to predict the probability of influenza or contagious diseases outbreak or their spread, the proposed new method will

[1] *Shanghai Engineering Research Center of Hadal Science and Technology, College of Engineering Science and Technology, Shanghai Ocean University, Shanghai 201306, China.*
[2] *Department of Mechanical and Structural Engineering and Materials Science, University of Stavanger, Norway.*
**E-mail: yihan.xing@uis.no* (Y. Xing)

provide better insight and an indication of the possible spread of diseases.

In the present study, an epidemic outbreak is viewed as an unexpected incident that may occur in any province of a given country at any time; therefore, the spatial spread is accounted for. Moreover, a specific non-dimensional factor λ is introduced to predict the latter epidemic risk at any time and any place.

Biological systems are subjected to ergodic environmental influences. The other alternative is to view the process as dependent on specific environmental parameters whose variation in time may be modelled as an ergodic process on its own.

The incidence data of COVID-19 in all US states from February 2020 until today were retrieved from the public website.[47] As this valuable data set is per the US state, the biological system under consideration can be regarded as a multi-degree of freedom (MDOF) dynamic system with highly inter-correlated regional components/dimensions. For the linear log model, some recent studies have already used statistical tools to predict COVID-19 development.[52]

Note that while this study aims to reduce the risk of future epidemic outbreaks by predicting them, it is solely focused on daily registered patient numbers and not symptoms. For long-lasting COVID-19 symptoms, the so-called "long COVID" and its risk factors and whether it is possible to predict a protracted course early in the disease for mortality research.[72] Fig. 1 presents the map of US East coast states.

2. Methodology

The MDOF health response vector process $R(t) = (X(t), Y(t), Z(t), \dots)$ that has been measured over a sufficiently long-time interval $(0,T)$. Unidimensional global maxima over the entire time span $(0,T)$ denoted as $X_T^{\max} = \max_{0\le t\le T} X(t)$, $Y_T^{\max} = \max_{0\le t\le T} Y(t)$, $Z_T^{\max} = \max_{0\le t\le T} Z(t), \dots$.

By sufficiently long time T one primarily means a large value of T with respect to the dynamic system auto-correlation time. Let $X_1, \dots, X_{N_X}$ be consequent in time local maxima of the process $X(t)$ at discrete monotonously increasing time instants $t_1^X < \cdots < t_{N_X}^X$ in $(0,T)$. The analogous definition follows for other MDOF response components $Y(t)$, $Z(t)$, ... with $Y_1, \dots, Y_{N_Y}$; $Z_1, \dots, Z_{N_Z}$ and so on. For simplicity, all $R(t)$ components, and therefore its maxima are assumed to be non-negative.

The target is to accurately estimate system failure probability, namely the probability of exceedance

$$1 - P = \mathrm{Prob}(X_T^{\max} > \eta_X \cup Y_T^{\max} > \eta_Y \cup Z_T^{\max} > \eta_Z \cup \dots) \quad (1)$$

with

$$P = \iiint_{(0,\ 0,\ 0,\ \dots)}^{(\eta_X,\ \eta_Y,\ \eta_Z,\ \dots)} p_{X_T^{\max},\ Y_T^{\max},\ Z_T^{\max},\dots}(X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \dots)\, dX_T^{\max} dY_{N_Y}^{\max} dZ_{N_z}^{\max} \dots \quad (2)$$

the probability of non-exceedance for critical values of response components η_X, η_Y, η_Z,...; $\cup$ being logical unity operation «or»; and $p_{X_T^{\max},\ Y_T^{\max},\ Z_T^{\max},\dots}$ being joint

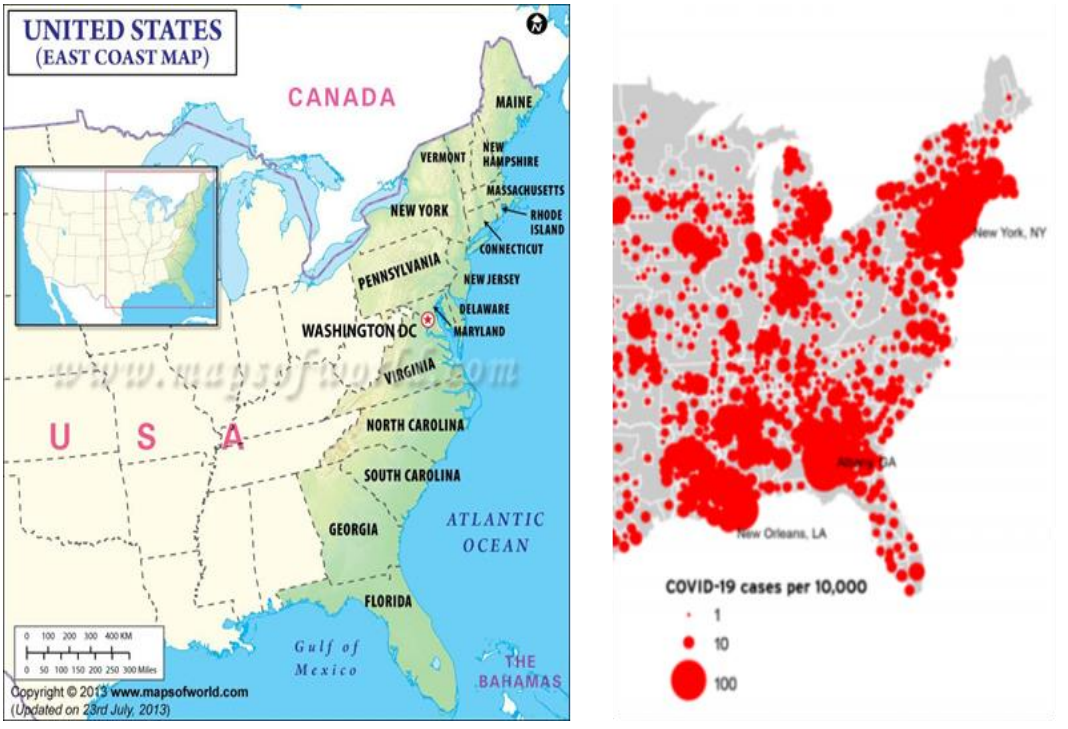


Fig. 1 Left: Map of USA East coast with states. Right: COVID cases map.[47]

probability density of the global maxima over the entire period $(0,T)$.

However, it is not feasible to estimate the latter joint probability distribution directly due to its high dimensionality and available data set limitations. More specifically, the moment when either $X(t)$ exceeds η_X, or $Y(t)$ exceeds η_Y, or $Z(t)$ exceeds η_Z, and so on, the system is regarded as immediately failed. Fixed failure levels η_X, η_Y, η_Z,...are, of course, individual for each unidimensional response component of $R(t)$. $X_{N_X}^{\max} = \max\{X_j; j = 1, \dots, N_X\} = X_T^{\max}$, $Y_{N_Y}^{\max} = \max\{Y_j; j = 1, \dots, N_Y\} = Y_T^{\max}$, $Z_{N_z}^{\max} = \max\{Z_j; j = 1, \dots, N_Z\} = Z_T^{\max}$, and so on.

Now, the local maxima time instants $\left[t_1^X < \cdots < t_{N_X}^X;\ t_1^Y < \cdots < t_{N_Y}^Y;\ t_1^Z < \cdots < t_{N_Z}^Z\right]$ are sorted in monotonously non-decreasing order into one single merged time vector $t_1 \le \cdots \le t_N$ .Note that $t_N = \max\{t_{N_X}^X,\ t_{N_Y}^Y,\ t_{N_Z}^Z,\ \dots\}$, $N = N_X + N_Y + N_Z + \dots$. In this case t_j represents local maxima of one of MDOF structural response components either $X(t)$ or $Y(t)$, or $Z(t)$ and so on. That means that having $R(t)$ time record, one just needs continuously and simultaneously screen for unidimensional response component local maxima and record its exceedance of the MDOF limit vector $(\eta_X,\ \eta_Y,\ \eta_Z, \dots)$ in any of its components $X, Y, Z, \dots$ Local unidimensional response component maxima are merged into one temporal non-decreasing vector $\vec{R} = (R_1, R_2, \dots, R_N)$ following the merged time vector $t_1 \le \cdots \le t_N$. That is to say, each local maxima R_j is actual encountered local maxima corresponding to either $X(t)$ or $Y(t)$, or $Z(t)$ and so on. Finally, the unified limit vector $(\eta_1, \dots, \eta_N)$ is introduced with each component η_j is either η_X, η_Y or η_Z and so on, depending on which of $X(t)$ or $Y(t)$, or $Z(t)$ *etc*., corresponds to the current local maxima with the running index j.

Now, scaling parameter $0 < \lambda \le 1$ is introduced to artificially simultaneously decrease limit values for all response components, namely the new MDOF limit vector $(\eta_X^\lambda,\ \eta_Y^\lambda,\ \eta_z^\lambda, \dots)$ with $\eta_X^\lambda \equiv \lambda \cdot \eta_X, \equiv \lambda \cdot \eta_Y, \eta_z^\lambda \equiv \lambda \cdot \eta_Z, \dots$ is introduced.[48-62] The unified limit vector $(\eta_1^\lambda, \dots, \eta_N^\lambda)$ is introduced with each component η_j^λ is either η_X^λ, η_Y^λ or η_z^λ and so on. The latter automatically defines probability $P(\lambda)$ as a function of λ, note that $P \equiv P(1)$ from Eqs. (1) and (2). Non-exceedance probability $P(\lambda)$ can be estimated as follows

$$P(\lambda) = \text{Prob}\{R_N \le \eta_N^\lambda, \dots, R_1 \le \eta_1^\lambda\}$$
$$= \text{Prob}\{R_N \le \eta_N^\lambda \mid R_{N-1} \le \eta_{N-1}^\lambda, \dots, R_1 \le \eta_1^\lambda\} \cdot \text{Prob}\{R_{N-1} \le \eta_{N-1}^\lambda, \dots, R_1 \le \eta_1^\lambda\}$$
$$= \prod_{j=2}^{N} \text{Prob}\{R_j \le \eta_j^\lambda \mid R_{j-1} \le \eta_{1j-}^\lambda, \dots, R_1 \le \eta_1^\lambda\} \cdot \text{Prob}(R_1 \le \eta_1^\lambda) \quad (3)$$

In the following, the principle behind a cascade of approximations based on conditioning is outlined.[63] In practice, the dependence between the neighbouring R_j is not negligible; thus, the following one-step (will be called conditioning level $k = 1$) memory approximation is introduced

$$\text{Prob}\{R_j \le \eta_j^\lambda \mid R_{j-1} \le \eta_{j-1}^\lambda, \dots, R_1 \le \eta_1^\lambda\} \approx$$
$$\text{Prob}\{R_j \le \eta_j^\lambda \mid R_{j-1} \le \eta_{j-1}^\lambda\} \quad (4)$$

for $2 \le j \le N$ (conditioning level $k = 2$). The approximation introduced by Eqs. (3) and (4) can be further expressed as

$$\text{Prob}\{R_j \le \eta_j^\lambda \mid R_{j-1} \le \eta_{j-1}^\lambda, \dots, R_1 \le \eta_1^\lambda\} \approx$$
$$\text{Prob}\{R_j \le \eta_j^\lambda \mid R_{j-1} \le \eta_{j-1}^\lambda, R_{j-2} \le \eta_{j-2}^\lambda\} \quad (5)$$

where $3 \le j \le N$ (will be called conditioning level $k = 3$), and so on. The idea is to monitor each independent failure that happened locally first in time, thus avoiding cascading local inter-correlated exceedances.

Equation (5) presents subsequent refinements of the statistical independence assumption. The latter approximations capture the statistical dependence effect between the neighbouring maxima with increased accuracy. Since the original MDOF process $R(t)$ was assumed ergodic and therefore stationary, the probability $p_k(\lambda) \coloneqq \text{Prob}\{R_j > \eta_j^\lambda \mid R_{j-1} \le \eta_{j-1}^\lambda,\ R_{j-k+1} \le \eta_{j-k+1}^\lambda\}$ for $j \ge k$ will be independent of j but only dependent on conditioning level k. Thus, non-exceedance probability can be approximated as in the average conditional exceedance rate method, see[64]

$$P_k(\lambda) \approx \exp(-N \cdot p_k(\lambda)),\ k \ge 1. \quad (6)$$

Note that Eq. (6) follows from Eq. (1) by neglecting $\text{Prob}(R_1 \le \eta_1^\lambda) \approx 1$, as design failure probability must be minuscule, also assumed $N \gg k$. Eq. (5) is similar to the well-known mean up-crossing rate equation for the probability of exceedance.[53,63] There is evident convergence with respect to the conditioning parameter k

$$P = \lim_{k\to\infty} P_k(1); \quad p(\lambda) = \lim_{k\to\infty} p_k(\lambda) \quad (7)$$

Note that Eq. (6) for $k = 1$ turns into a well-known non-exceedance probability relationship with the mean up-crossing rate function

$$P(\lambda) \approx \exp(-\nu^+(\lambda)\,T); \quad \nu^+(\lambda) = \int_0^\infty \zeta p_{R\dot{R}}(\lambda, \zeta) d\zeta \quad (8)$$

where $\nu^+(\lambda)$ denotes the mean up-crossing rate of the response level λ for the above assembled non-dimensional vector $R(t)$ assembled from scaled MDOF system response $\left(\frac{X}{\eta_X}, \frac{Y}{\eta_Y}, \frac{Z}{\eta_Z}, \ldots\right)$. The mean up-crossing rate is given by Rice's formula given in Eqs. (7) and (8) with $p_{R\dot{R}}$ being joint probability density for $(R, \dot{R})$ with $\dot{R}$ being time derivative $R'(t)$, see[65]. Eq. (8) relies on the Poisson assumption that up-crossing events of high λ levels (in the present study, it is $\lambda \geq 1$) can be assumed to be independent. The latter may not be the case for narrowband responses and higher-level dynamic systems that exhibit cascading failures in different dimensions, subsequent in time, caused by intrinsic inter-dependency between extreme events, manifesting itself in the appearance of highly correlated local maxima clusters within the assembled vector $\vec{R} = (R_1, R_2, \ldots, R_N)$.

In the above, the stationarity assumption has been used. However, the proposed methodology can also treat the nonstationary case. For nonstationary case, the scattered diagram of $m = 1, .., M$ seasonal epidemic conditions, each short-term seasonal state has the probability q_m, so that $\sum_{m=1}^{M} q_m = 1$. Next, let one introduce the long-term equation

$$p_k(\lambda) \equiv \sum_{m=1}^{M} p_k(\lambda, m)\, q_m \qquad (9)$$

with $p_k(\lambda, m)$ being the same function as in Eq. (9) but corresponding to a specific short-term seasonal epidemic state with the number m.

Note that the accuracy of the suggested approach for a large variety of one-dimensional dynamic systems was successfully verified by authors in previous years.[53,63]

Next, the following extrapolation method [67-71] is briefly introduced, as it will be used as a basis for the failure probability distribution tail extrapolation, asymptotically being Gumbel distribution type. The latter approach assumes that the class of parametric functions needed for extrapolation in a general case can be modelled similarly to the Gumbel distribution and the general extreme value (GEV) distribution. The above introduced $p_k(\lambda)$ as functions are often regular in the tail, specifically for values of λ approaching and exceeding 1. More precisely, for $\lambda \geq \lambda_0$, the distribution tail behaves similar to $\exp\{-(a\lambda + b)^c + d\}$ with a, b, c, d being suitably fitted constants for suitable tail cut-on λ_0 value. Therefore, one can write

$$p_k(\lambda) \approx \exp\{-(a_k\lambda + b_k)^{c_k} + d_k\}, \quad \lambda \geq \lambda_0 \qquad (10)$$

Next, by plotting $\ln\{\ln(p_k(\lambda)) - d_k\}$ versus $\ln(a_k\lambda + b_k)$, often nearly perfectly linear tail behaviour is observed.

It is helpful to do the optimisation on the logarithmic level by minimising the following error function F with respect to the four parameters a_k, b_k, c_k, p_k, q_k

$$F(a_k, b_k, c_k, p_k, q_k) = \int_{\lambda_0}^{\lambda_1} \omega(\lambda)\,\{\ln(p_k(\lambda)) - d_k + (a_k\lambda + b_k)^{c_k}\}^2 d\lambda, \quad \lambda \geq \lambda_0 \qquad (11)$$

with λ_1 being a suitable distribution tail cut-off value, namely the most significant wave height value, where the confidence interval width is still acceptable. Optimal values of the parameters a_k, b_k, c_k, p_k, q_k may also be determined using a sequential quadratic programming (SQP) method incorporated in the NAG Numerical Library.[64]

Weight function ω can be defined as $\omega(\lambda) = \{\ln \mathrm{CI}^+(\lambda) - \ln \mathrm{CI}^-(\lambda)\}^{-2}$ with $(\mathrm{CI}^-(\lambda), \mathrm{CI}^+(\lambda))$ being a confidence interval (CI) empirically estimated from the simulated or measured dataset, see [81-86]. When the parameter $c = \lim_{k\to\infty} c_k$ is equal to 1 or close to it, the distribution is close to the Gumbel distribution.

For any general ergodic wave height or wind speed process, the sequence of conditional exceedances over a threshold λ can be assumed to constitute a Poisson process; however, in general, non-homogeneous one. Thus, for levels of λ approaching 1, the approximate limits of a p-% confidence interval (CI) of $p_k(\lambda)$ can be given as follows

$$\mathrm{CI}^{\pm}(\lambda) = p_k(\lambda)\left(1 \pm \frac{f(p)}{\sqrt{(N-k+1)p_k(\lambda)}}\right). \qquad (12)$$

with $f(p)$ being estimated from the inverse normal distribution, for example, $f(90\%) = 1.65$, $f(95\%) = 1.96$. with N being the total number of local maxima assembled in the analysed vector $\vec{R}$, see Eqs. (10-12).

3. Results and discussion

Prediction of influenza-like epidemics has long been the focus of attention in epidemiology and mathematical biology. It is well known that public health dynamics is a highly non-linear multidimensional and spatially cross-correlated dynamic system that is always challenging to analyse. Previous studies have used a variety of approaches to model influenza-like cases. This section illustrates the efficiency of the above-described methodology using the new method applied to the real-life COVID-19 data sets, presented as a new daily recorded infected patient time series spread over large terrains. COVID-19 and influenza are contagious diseases with high transmissibility to spread worldwide with considerable morbidity and mortality. They occur most frequently seasonally in late autumn, winter and early spring, reaching their peak prevalence mostly in winter. Seasonal influenza epidemics caused by influenza A and B viruses typically occur annually during winter in temperate regions and present an enormous burden on worldwide public health, resulting in

around 3–5 million cases of severe illness and 250,000–500,000 deaths worldwide each year, according to the World Health Organization (WHO).[2]

This section presents a real-life application of the above-described method. The statistical data in the present section are taken from the official US website.[47] The website provides the number of newly diagnosed cases every day in US states from 22 January 2020 to 6 April 2022. Patient numbers from fourteen different most affected US states were chosen as components $X, Y, Z, ..$ thus constituting an example of a fourteen dimensional (14D) dynamic biological system.

To unify all 14 measured time series X, Y, Z, the following scaling was performed

$$X \to \frac{X}{\eta_X},\ Y \to \frac{Y}{\eta_Y},\ Z \to \frac{Z}{\eta_Z}, \ldots \tag{13}$$

making all 14 responses non-dimensional and having the same failure limit equal to 1. Failure limits, or in other words, epidemic thresholds, were chosen differently for different provinces in the present study $\eta_X, \eta_Y, \eta_Z, \ldots$ were set equal to the observed two years maxima, twice increased. Next, all local maxima from 14 measured time series were merged into one single time series by keeping them in time non-decreasing order: $\vec{R} = (\max\{X_1, Y_1, Z_1, \ldots\}, \ldots, \max\{X_N, Y_N, Z_N, \ldots\})$ with each the whole vector $\vec{R}$ being sorted according to non-decreasing times of occurrence of these local maxima.

Figure 2 presents the number of new daily recorded patients as a surface per state and per day. It is seen that the 2D surface presented in Fig. 2 is highly irregular, which might be an indication of inaccuracies inherent in the data set.

Figure 3 presents the number of new daily recorded patients as a 14D vector $\vec{R}$, consisting of assembled regional new daily patient numbers. Note that vector $\vec{R}$ does not have physical meaning on its own, as it is assembled of different regional components with different epidemic backgrounds. Index j is just a running index of local maxima encountered in a non-decreasing time sequence.

Figure 4 presents 100 years return level extrapolation according to Eq. (9) towards epidemic outbreak with a 100-year return period, indicated by the horizontal dotted line, and somewhat beyond, $\lambda = 0.1$ cut-on value was used. The dotted lines indicate extrapolated 95% confidence interval according to Eq. (10). According to Eq. (6) $p(\lambda)$ is directly related to the target failure probability $1 - P$ from Eq. (1). Therefore, in agreement with Eq. (6), system failure probability $1 - P \approx 1 - P_k(1)$ can be estimated. Note that in Eq. (6), N corresponds to the total number of local maxima in the unified response vector $\vec{R}$. Conditioning parameter $k = 3$ was found to be sufficient due to the occurrence of convergence with respect to k, see Eq. (7). Fig. 4 exhibits a reasonably narrow 95% CI. The latter is an advantage of the proposed method.

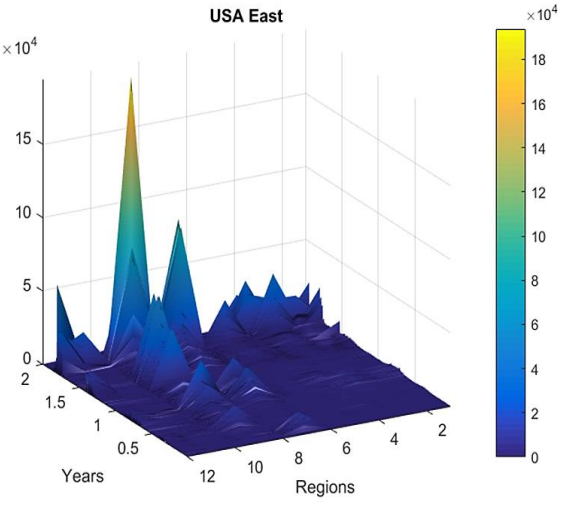


Fig. 2 Number of new daily recorded patients as a surface per state and per day.

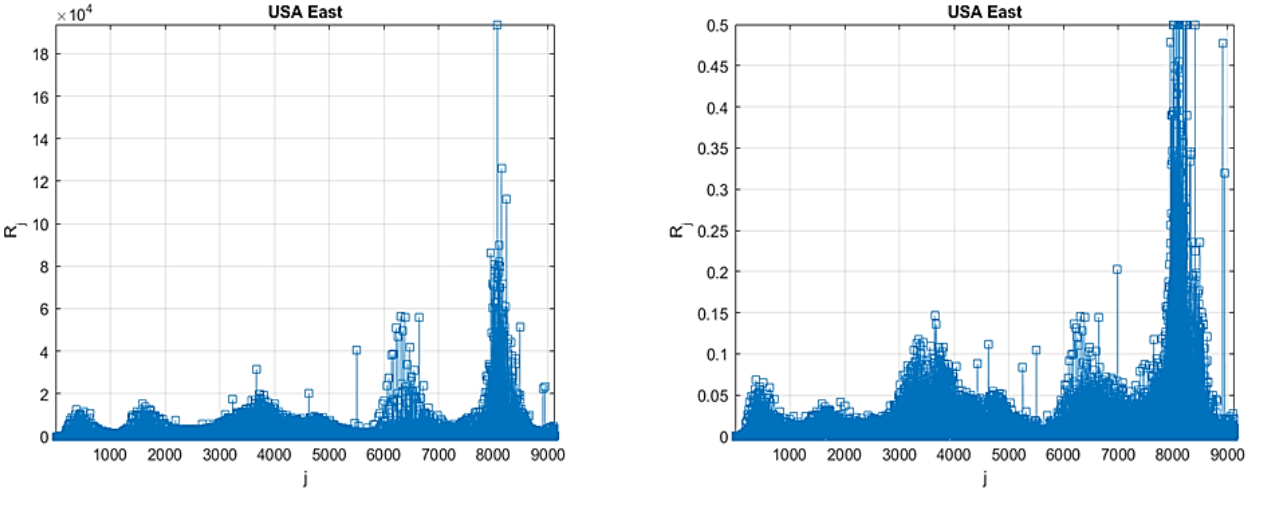


Fig. 3 Number of new daily recorded patients as 14D vector $\vec{R}$. Left: as it is, Right: scaled by Eq. (13).

4. Method limitations

Note that while being novel, the above-described methodology has a clear advantage of efficiently utilising available data sets due to its ability to treat health system multi-dimensionality and perform accurate extrapolation based on a relatively limited data set. Note that the predicted non-dimensional λ level, indicated by the star in Fig. 4, represents the probability of an epidemic outbreak in any US state in the years to come.

As with any statistical analysis, there are several inherent limitations. As the suggested method is quite general and does not introduce any additional assumptions, there are still limitations within the analysed data. To mention major:

- Fake and manipulated data
- Inaccuracies and irregularities in the data itself, including outliers and missing records
- Non-stationarity and presence of an unknown trend
- Insufficient data volume.

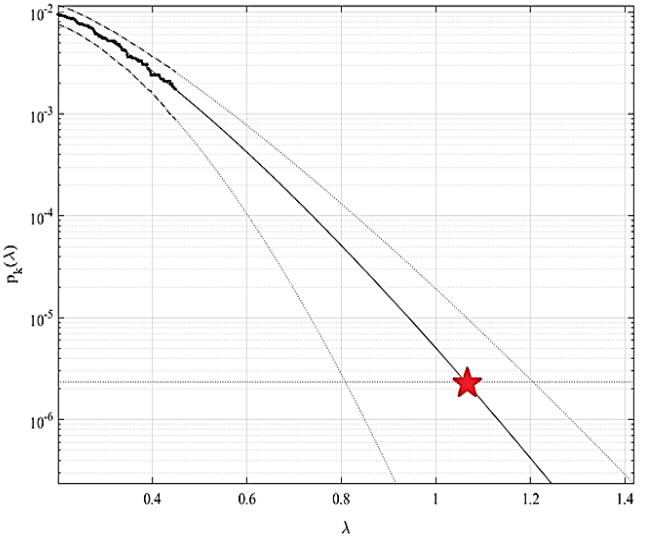


Fig. 4 100 years return level (horizontal dotted line) extrapolation of p_k (λ) towards critical level (indicated by a star) and beyond. Extrapolated 95% CI indicated by dotted lines.

5. Conclusions

Despite the simplicity, the present study successfully offers a novel multidimensional modelling strategy and a methodological avenue to implement the forecasting of an epidemic during its course. The authors studied recorded COVID-19 patient numbers from fourteen different US states, constituting an example of a fourteen-dimensional (14D) observed in 2020-2022. The novel reliability method was applied to new daily patient numbers as a real-time multidimensional system. The theoretical reasoning behind the proposed method is given in detail. Note that the use of direct either measurement or Monte Carlo simulation for dynamic biological system reliability analysis is attractive; however, dynamic system complexity and its high dimensionality require the development of novel robust and accurate techniques that can deal with a limited data set at hand, utilising available data as efficient as possible.

The main conclusion is that the public health system under local environmental and epidemiologic conditions in the USA East coast is well managed. The main conclusion is that the public health system under local environmental and epidemiologic conditions in the USA East coast is well managed. The predicted 100-year return period risk level λ of the epidemic outbreak is reasonably low compared to the reference value of one. However, there is a medium risk of a future epidemic outbreak, at least within 100 years.

This study aimed to develop a general-purpose, robust, and straightforward multidimensional reliability method. The method introduced in the present study has been previously validated by application to a wide range of simulation models, but for only one-dimensional system responses and, in general, very accurate predictions were obtained. Both measured and numerically simulated time series responses can be analysed. It is shown that the proposed method produced a reasonable confidence interval. Thus, the suggested methodology may become appropriate for various non-linear dynamic biological systems reliability studies. Finally, the suggested method can be used in many public health applications. The presented COVID-19 example does not limit areas of new method applicability.

Acknowledgements

The datasets analysed during the current study are available online.[47] https://github.com/nytimes/covid-19-data. The authors confirm that all methods were performed in accordance with the relevant guidelines and regulations according to the Declarations of Helsinki.

Conflict of Interest

There is no conflict of interest.

Supporting Information

Not applicable.

References

[1] J. Chen, X. Lei, L. Zhang, B. Peng, Using extreme value theory approaches to forecast the probability of outbreak of highly pathogenic influenza in Zhejiang, China, *PLoS One*, 2015, **10**, e0118521, doi: 10.1371/journal.pone.0118521.

[2] World Health Organization. Influenza fact sheet March 2014 (cited 10 June 2014). In: World Health Organization website

[Internet]. Geneva: World Health Organization 1948 - [about 2 screens]. Available: http://www.who.int/mediacentre/factsheets/fs211/en/index.html.
[3] E. Goldstein, S. Cobey, S. Takahashi, J. C. Miller, M. Lipsitch, Predicting the epidemic sizes of influenza A/H1N1, A/H3N2, and B: A statistical method, *PLoS Medicine*, 2011, **8**, e1001051, doi: 10.1371/journal.pmed.1001051.
[4] R. P. Soebiyanto, F. Adimi, R. K. Kiang, Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters, *PLoS One*, 2010, **5**, e9450. doi: 10.1371/journal.pone.0009450.
[5] A. S. Mugglin, N. Cressie, I. Gemmell, Hierarchical statistical modelling of influenza epidemic dynamics in space and time, *Statistics in Medicine*, 2002, **21**, 2703-2721, doi: 10.1002/sim.1217.
[6] E.-K. Kim, J. Seok, J. S. Oh, H. W. Lee, K. Kim, Use of hangeul twitter to track and predict human influenza infection, *PLoS One*, 2013, **7**, e69305. doi: 10.1371/journal.pone.0069305.
[7] H. C. Lee, H. Wackernagel, Extreme value analyses of US P&I mortality data under consideration of demographic effects. 2007 (cited 10 June 2014). In: Centre de géosciences / Géostatistique Publications & documentation, available: http://cg.ensmp.fr/bibliotheque/public/LEE_Rapport_00600.pdf, doi: 10.1021/jp204223t.
[8] O. Gaidai, F. Wang, Y. Wu, Y. Xing, A. Medina, J. Wang, Offshore renewable energy site correlated wind-wave statistics, *Probabilistic Engineering Mechanics*, 2022, **68**, 103207, doi: 10.1016/j.probengmech.2022.103207.
[9] H. O. Madsen, S. Krenk, N. C. Lind, Methods of structural safety, Englewood Cliffs, Prentice-Hall Inc., 1986.
[10] O. Ditlevsen, H. O. Madsen, Structural reliability methods, Chichester (UK): John Wiley & Sons, Inc, 1996
[11] R. E. Melchers, Structural reliability analysis and prediction, New York, John Wiley & Sons, Inc, 1999.
[12] S.-K. Choi, R. V. Grandhi, R. A.Canfield, Reliability-based structural design, London, Springer-Verlag, 2007.
[13] A. H. Dehkordi, M. Alizadeh, P. Derakhshan, P. Babazadeh, A. Jahandideh, Understanding epidemic data and statistics: A case study of COVID-19, *Journal of Medical Virology*, 2020, **92**, 868-882, doi: 10.1002/jmv.25885.
[14] M. Desjardins, A. Hohl, E. Delmelle, Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: detecting and evaluating emerging clusters, *Applied Geography (Sevenoaks, England)*, 2020, **118**, 102202-102202, doi: 10.1016/j.apgeog.2020.102202.
[15] A. Hohl, E. Delmelle, M. Desjardins, Y. Lan, Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States, *Spatial and Spatio-Temporal Epidemiology*, 2020, **34**, 100354-100354, doi: 10.1016/j.sste.2020.100354.
[16] A. Hohl, E. Delmelle, M. Desjardins, Rapid detection of COVID-19 clusters in the United States using a prospective space-time scan statistic: An update, *Sigspatial Special*, 2020, **12**, 27-33.
[17] R. Paul, A. A. Arif, O. Adeyemi, S. Ghosh, D. Han, Progression of COVID-19 from urban to rural areas in the United States: a spatiotemporal analysis of prevalence rates, *The Journal of Rural Health*, 2020, **36**, 591-601, doi: 10.1111/jrh.12486.
[18] N. Pearce, J. P. Vandenbroucke, T. J. VanderWeele, S. Greenland, Accurate statistics on COVID-19 are essential for policy guidance and decisions, *American Journal of Public Health*, 2020, **110**, 949-951, doi: 10.2105/AJPH.2020.305708.
[19] L. M. Rossen, A. M. Branum, F. B. Ahmad, P. Sutton, R. N. Anderson, Excess deaths associated with COVID-19, by age and race and ethnicity—United States, 26 January–3 October, 2020, *Morbidity and Mortality Weekly Report*, 2020, **69**, 1522.
[20] M. Shah, M. Sachdeva, R. P. Dodiuk-Gad, COVID-19 and racial disparities, *Journal of the American Academy of Dermatology*, 2020, **83**, e35, doi: 10.1016/j.jaad.2020.04.046.
[21] S. H. Jacobson, J. A. Jokela, Non–COVID-19 excess deaths by age and gender in the United States during the first three months of the COVID-19 pandemic, *Public Health*, 2020, **189**, 101-103, doi: 10.1016/j.puhe.2020.10.004.
[22] D. B. G. Tai, A. Shah, C. A. Doubeni, I. G. Sia, M. L. Wieland, The disproportionate impact of COVID-19 on racial and ethnic minorities in the United States, *Clinical Infectious Diseases*, 2021, **72**, 703-706, doi: 10.1093/cid/ciaa815.
[23] D. Weinberger, J. Chen, T. Cohen, F. Crawford, F. Mostashari, D. Olson, V. Pitzer, N. Reich, M. Russi, L. Simonsen, A. Watkins, C. Viboud, Estimation of excess deaths associated with the COVID-19 pandemic in the United States, March to May *JAMA Internal Medicine*, 2020, **180**, 1336-1344, doi: 10.1001/jamainternmed.2020.339.
[24] B. Doğan, M. Ben Jebli, K. Shahzad, T. H. Farooq, U. Shahzad, Investigating the effects of meteorological parameters on COVID-19: case study of new jersey, United States, *Environmental Research*, 2020, **191**, 110148 - 110148, doi: 10.1016/j.envres.2020.110148.
[25] E. Kuhl, Data-driven modeling of COVID-19—Lessons learned, *Extreme Mechanics Letters*, 2020, **40**, 100921, doi: 10.1016/j.eml.2020.100921.
[26] J. A. Gold, L. M. Rossen, F. B. Ahmad, P. Sutton, Z. Li, P. P. Salvatore, B. R. Jackson, Race, ethnicity, and age trends in persons who died from COVID-19—United States, May–August 2020, *Morbidity and Mortality Weekly Report*, 2020, **69**, 1517.
[27] H. Xu, R. Basu, How the United States flunked the COVID-19 test: some observations and several lessons, The American *Review of Public Administration*, 2020, **50**, 568-576, doi:

10.1177/0275074020941701.
[28] C. T. Rentsch, F. Kidwai-Khan, J. P. Tate, L. S. Park, J. T. King Jr, M. Skanderson, R. G. Hauser, A. Schultze, C. I. Jarvis, M. Holodniy, V. L. Re 3rd, K. M. Akgün, K. Crothers, T. H. Taddei, M. S. Freiberg, A. C. Justice, Covid-19 testing, hospital admission, and intensive care among 2, 026, 227, United States veterans aged 54–75 years, *MedRxiv*, 2020.
[29] C. F. Baum, M. Henry, Socioeconomic factors influencing the spatial spread of COVID-19 in the United States, *Miguel, Socioeconomic Factors influencing the Spatial Spread of COVID-19 in the United States*, 29 May, 2020.
[30] I. Arpaci, K. Karatas, M. Baloglu, A. Haktanir, COVID-19 phobia in the United States: Validation of the COVID-19 Phobia Scale (C19P-SE), *Death Studies*, 2022, **46**, 553-559.
[31] W. H. Finch, M. E. Hernández Finch, Poverty and Covid-19: rates of incidence and deaths in the United States during the first 10 weeks of the pandemic, *Frontiers in Sociology*, 2020, **47**.
[32] A. Güemes, S. Ray, K. Aboumerhi, M. R. Desjardins, A. Kvit, A. E. Corrigan, B. Fries, T. Shields, R. D. Stevens, F. C. Curriero, R. Etienne-Cummings, A syndromic surveillance tool to detect anomalous clusters of COVID-19 symptoms in the United States, *Scientific Reports*, 2021, **11**, 1-11.
[33] D. Liang, L. Shi, J. Zhao, P. Liu, J. Sarnat, S. Gao, J. Schwartz, Y. Liu, S. Ebelt, N. Scovronick, H. H. Chang, Urban air pollution may enhance COVID-19 case-fatality and mortality rates in the United States, *The Innovation*, 2020, **1**, 100047, doi: 10.1016/j.xinn.2020.100047.
[34] D. Weinberger, T. Cohen, F. Crawford, F. Mostashari, D. Olson, V. Pitzer, N. Reich, M. Russi, L. Simonsen, A. Watkins, C. Viboud, Estimating the early death toll of COVID-19 in the United States, Medrxiv, 2020, doi: 10.1101/2020.04.15.20066431
[35] M. Roser, H. Ritchie, E. Ortiz-Ospina, J. Hasell, Coronavirus disease (COVID-19)–Statistics and research, Our World in data, 2020, 4.
[36] M. G. Baker, T. K. Peckham, N. S. Seixas, Estimating the burden of United States workers exposed to infection or disease: a key factor in containing risk of COVID-19 infection, *PLoS one*, 2020, **15**, e0232452, doi: 10.1371/journal.pone.0232452.
[37] A. Mollalo, B. Vahedi, K. M. Rivera, GIS-based spatial modeling of COVID-19 incidence rate in the continental United States, *Science of the Total Environment*, 2020, **728**, 138884, doi: 10.1016/j.scitotenv.2020.138884.
[38] J.-M. Chen, Novel statistics predict the COVID-19 pandemic could terminate in 2022, *Journal of Medical Virology*, 2022, **94**, 2845-2848, doi: 10.1002/jmv.27661.
[39] C. E. Rodriguez-Diaz, V. Guilamo-Ramos, L. Mena, E. Hall, B. Honermann, J. S. Crowley, S. Baral, G. J. Prado, M. Marzan-Rodriguez, C. Beyrer, P. S. Sullivan, G. A. Millett, Risk for COVID-19 infection and death among Latinos in the United States: examining heterogeneity in transmission dynamics, *Annals of Epidemiology*, 2020, **52**, 46-53, doi: 10.1016/j.annepidem.2020.07.007.
[40] F. Lu, A. T. Nguyen, N. Link, M. Santillana, Estimating the cumulative incidence of COVID-19 in the United States using four complementary approaches, *MedRxiv*, 2020.
[41] N. A. Halpern, K. S. Tan, United States resource availability for COVID-19, *Society of Critical Care Medicine*, 2020.
[42] H. Singh, S. Bawa, Predicting COVID-19 statistics using machine learning regression model: Li-MuLi-Poly, *Multimedia Systems*, 2021, **28**, 113-120, doi: 10.1007/s00530-021-00798-2.
[43] Q. Deng, Dynamics and development of the COVID-19 epidemic in the United States: a compartmental model enhanced with deep learning techniques, *Journal of Medical Internet Research*, 2020, **22**, e21173, doi: 10.2196/21173.
[44] A. Mollalo, K. M. Rivera, B. Vahedi, Artificial neural network modeling of COVID-19 (COVID-19) incidence rates across the continental United States, *International Journal of Environmental Research and Public Health*, 2020, **17**, 4204, doi: 10.3390/ijerph17124204.
[45] M. Sharma, S. Sharma, The rising number of COVID-19 cases reflecting growing search trend and concern of people: a google trend analysis of eight major countries, *Journal of Medical Systems*, 2020, **44**, 117, doi: 10.1007/s10916-020-01588-5.
[46] D. Gondauri, E. Mikautadze, M. Batiashvili, Research on COVID-19 virus spreading statistics based on the examples of the cases from different countries, *Electronic Journal of General Medicine*, 2020, **17**, em209, doi: 10.29333/ejgm/7869.
[47] https://github.com/nytimes/covid-19-data.
[48] O. Gaidai, Y. Cao, S. Loginov, 2023, "Global cardiovascular diseases death rate prediction", Current Problems in Cardiology, doi: 10.1016/j.cpcardiol.2023.101622.
[49] J. Chen, X. Lei, L. Zhang, B. Peng, Using extreme value theory approaches to forecast the probability of outbreak of highly pathogenic influenza in Zhejiang, China, *PLoS One*, 2015, **10**, e0118521, doi: 10.1371/journal.pone.0118521.
[50] M. Thomas, H. Rootzén, Real-time prediction of severe influenza epidemics using Extreme Value Statistics, 2019: arXiv: 1910.10788, https://arxiv.org/abs/1910.10788.
[51] H. C. Lee, H. Wackernagel, Extreme values analyses of US P&I mortality data under consideration of demographic effects. Centre de géosciences, Ecole des mines de Paris, Fontainebleau, France, R071113HLEE.
[52] J. Chu, A statistical analysis of the COVID-19 (COVID-19) in Italy and Spain, *PLoS One*, 2021, **16**, e0249037, doi: 10.1371/journal.pone.0249037.
[53] O. Gaidai, Y. Xing, "Prediction of death rates for cardiovascular diseases and cancers", Cancer Innovation, 2023,

doi: 10.1002/cai2.47
[54] O. Gaidai, F. Wang, V. Yakimov, "COVID-19 multi-state epidemic forecast in India", Proceedings of the Indian National Science Academy, 2023, doi: 10.1007/s43538-022-00147-5.
[55] O. Gaidai, Y. Xing, R. Balakrishna, J. Xu, "Improving extreme offshore wind speed prediction by using deconvolution", Heliyon, 2023, doi: 10.1016/j.heliyon.2023.e13533.
[56] O. Gaidai, Y. Xing, X. Xu, COVID-19 epidemic forecast in USA East coast by novel reliability approach, *Research square*, 2022, doi: 10.21203/rs.3.rs-1573862/v1.
[57] X. Xu, Y. Xing, O. Gaidai, K. Wang, K. S. Patel, P. Dou, Z. Zhang, A novel multi-dimensional reliability approach for floating wind turbines under power production conditions, *Frontiers in Marine Science*, 2022, **9**, 970081, doi: 10.3389/fmars.2022.970081.
[58] O. Gaidai, Y. Xing, R. Balakrishna, Improving extreme response prediction of a subsea shuttle tanker hovering in ocean current using an alternative highly correlated response signal, *Results in Engineering*, 2022, **15**, 100593, doi: 10.1016/j.rineng.2022.100593.
[59] Y. Cheng, O. Gaidai, D. Yurchenko, X. Xu, S. Gao, Study on the Dynamics of a Payload Influence in the Polar Ship, The 32nd International Ocean and Polar Engineering Conference, Paper Number: ISOPE-I-22-342, 2022.
[60] Gaidai, O., Cao, Y., Xing, Y., Balakrishna, R., 2023, "Extreme springing response statistics of a tethered platform by deconvolution", International Journal of Naval Architecture and Ocean Engineering, doi: 10.1016/j.ijnaoe.2023.100515.
[61] O. Gaidai, X. Xu, A. Naess, Y. Cheng, R. Ye, J. Wang, Bivariate statistics of wind farm support vessel motions while docking, *Ships and Offshore Structures*, 2021, **16**, 135-143, doi: 10.1080/17445302.2019.1710936.
[62] A. Naess, O. Gaidai, Estimation of extreme values from sampled time series, *Structural Safety*, 2009, **31**, 325-334, doi: 10.1016/j.strusafe.2008.06.021.
[63] A. Naess, T. Moan, Stochastic Dynamics of Marine Structures. Cambridge: Cambridge University Press, 2012, doi: 10.1017/cbo9781139021364
[64] Numerical Algorithms Group, 2010. NAG Toolbox for Matlab. Oxford, UK: NAG Ltd.
[65] S. O. Rice, Mathematical analysis of random noise, *Bell System Technical Journal*, 1944, **23**, 282-332, doi: 10.1002/j.1538-7305.1944.tb00874.x.
[66] X. Zhao, Extreme value modelling with application in finance and neonatal research, PhD Thesis, The University of Canterbury, 2010.
[67] L. Zheng, K. Ismail, X. Meng, Freeway safety estimation using extreme value theory approaches: a comparative study, *Accident Analysis & Prevention*, 2014, **62**, 32-41, doi: 10.1016/j.aap.2013.09.006.
[68] A. J. McNeil, R. Frey, P. Embrechts, Quantitative risk management: Concepts, techniques and tools, America: Princeton University Press, 2005.
[69] P. Patie, Estimation of value at risk using extreme value theory, 2000 23 March [cited 10 June 2014]. In: Talks in financial and insurance mathematics [Internet]. Lausanne: Eidgenossische Technische Hochschule Zürich 1855 -. [about 1 screen]. Available: http://www.math.ethz.ch/*patie/VaREvT.pdf, 2000.
[70] A. Sumi, K.-I. Kamo, MEM spectral analysis for predicting influenza epidemics in Japan, *Environmental Health and Preventive Medicine*, 2012, **17**, 98-108, doi: 10.1007/s12199-011-0223-0.
[71] P. Songchitruksa, A. P. Tarko, The extreme value theory approach to safety estimation, *Accident Analysis & Prevention*, 2006, **38**, 811-822, doi: 10.1016/j.aap.2006.02.003.
[72] E. J. Williamson, A. J. Walker, K. Bhaskaran, S. Bacon, C. Bates, C. E. Morton, H. J. Curtis, A. Mehrkar, D. Evans, P. Inglesby, J. Cockburn, H. I. McDonald, B. MacKenna, L. Tomlinson, I. J. Douglas, C. T. Rentsch, R. Mathur, A. Y. S. Wong, R. Grieve, D. Harrison, H. Forbes, A. Schultze, R. Croker, J. Parry, F. Hester, S. Harper, R. Perera, S. J. W. Evans, L. Smeeth, B. Goldacre, Factors associated with COVID-19-related death using OpenSAFELY, *Nature*, 2020, **584**, 430-436, doi: 10.1038/s41586-020-2521-4.
[73] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G. F. Gao, W. Tan, A COVID-19 from patients with pneumonia in China, 2019, *New England Journal of Medicine*, 2020, **382**, 727-733, doi: 10.1056/nejmoa2001017.
[74] J. T. Wu, L. Leung, G. M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, Lancet. 2020, 1-3, doi: 10.1016/S0140-6736(20)30260-93.
[75] F. He, Y. Deng, W.-N. Li, Coronavirus disease 2019 (COVID-19): what we know? *Journal of Medical Virology*, 2020, **2019**, 0-2, doi: 10.1002/jmv.25766.
[76] Z. Wu, J. M. McGoogan, Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China, *JAMA*, 2020, **323**, 1239, doi: 10.1001/jama.2020.2648.
[77] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W. J. Liu, D. Wang, W. Xu, E. C. Holmes, G. F. Gao, G. Wu, W. Chen, W. Shi, W. Tan, Genomic characterisation and epidemiology of 2019 COVID-19 : implications for virus origins and receptor binding, *The Lancet*, 2020, **395**, 565-574, doi: 10.1016/S0140-6736(20)30251-8.
[78] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W.

Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, Z.-L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, *Nature*, 2020, **579**, 270-273, doi: 10.1038/s41586-020-2012-7.

[79] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G. F. Gao, W. Tan, A COVID-19 from patients with pneumonia in China, 2019, *New England Journal of Medicine*, 2020, **382**, 727-733, doi: 10.1056/nejmoa2001017.

[80] W. H. Organisation, Coronavirus disease 2019 (COVID-19) Situation Report - 70. 30 March 2020.

[81] R. J. Kryscio, The mathematical theory of infectious diseases and its applications, *Mathematical Biosciences*, 1977, **34**, 172, doi: 10.1016/0025-5564(77)90044-x.

[82] N. T. J. Bailey, The total size of a general stochastic epidemic, *Biometrika*, 1953, **40**, 177, doi: 10.2307/2333107.

[83] N. G. Becker, T. Britton, Statistical studies of infectious disease incidence, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 1999, **61**, 287-307, doi: 10.1111/1467-9868.00177.

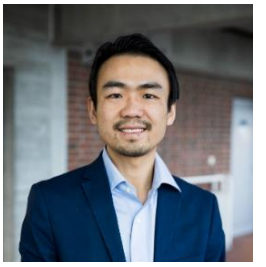
[84] L. Lan, D. Xu, G. Ye, C. Xia, S. Wang, Y. Li, H. Xu, Positive RT-PCR test results in patients recovered from COVID-19, *JAMA*, 2020, **323**, 1502, doi: 10.1001/jama.2020.2783.

[85] W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics, Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character, 1927, **115**, 700-721, doi: 10.1098/rspa.1927.0118.

[86] N. T. J. Bailey, Maximum-likelihood estimation of the relative removal rate from the distribution of the total size of an intra-household epidemic, *Journal of Hygiene*, 1954, **52**, 400-402, doi: 10.1017/s0022172400027595.

Author Information

Prof. Oleg Gaidai *is specialised in statistics and reliability. He has research interests within bio-statistics, risk and reliability study, non-linear dynamic systems.*



Prof. Yihan Xing *is Professor, University of Stavanger, Norway. He has major experience in dynamic systems modelling. He has research interests within digital twin modelling, extreme value analysis and stochastic design methods.*

