



METHOD ARTICLE

A novel statistical method for long-term coronavirus modelling [version 1; peer review: 1 approved]

Oleg Gaidai ¹, Ping Yan¹, Yihan Xing ², JingXiang Xu¹, Yu Wu¹

¹Engineering Research Center of Marine Renewable Energy, Shanghai Ocean University, Shanghai, China

²Department of Mechanical and Structural Engineering and Materials Science, University of Stavanger, Stavanger, Norway

V1 First published: 10 Nov 2022, 11:1282
<https://doi.org/10.12688/f1000research.125924.1>

Latest published: 10 Nov 2022, 11:1282
<https://doi.org/10.12688/f1000research.125924.1>

Abstract

Background: Novel coronavirus disease has been recently a concern for worldwide public health. To determine epidemic rate probability at any time in any region of interest, one needs efficient bio-system reliability approach, particularly suitable for multi-regional environmental and health systems, observed over a sufficient period of time, resulting in a reliable long-term forecast of novel coronavirus infection rate. Traditional statistical methods dealing with temporal observations of multi-regional processes do not have the multi-dimensionality advantage, that suggested methodology offers, namely dealing efficiently with multiple regions at the same time and accounting for cross-correlations between different regional observations.

Methods: Modern multi-dimensional novel statistical method was directly applied to raw clinical data, able to deal with territorial mapping. Novel reliability method based on statistical extreme value theory has been suggested to deal with challenging epidemic forecast. Authors used MATLAB optimization software.

Results: This paper described a novel bio-system reliability approach, particularly suitable for multi-country environmental and health systems, observed over a sufficient period of time, resulting in a reliable long-term forecast of extreme novel coronavirus death rate probability. Namely, accurate maximum recorded patient numbers are predicted for the years to come for the analyzed provinces.

Conclusions: The suggested method performed well by supplying not only an estimate but 95% confidence interval as well. Note that suggested methodology is not limited to any specific epidemics or any specific terrain, namely its truly general. The only assumption and limitation is bio-system stationarity, alternatively trend analysis should be performed first. The suggested methodology can be used in various public health applications, based on their clinical survey data.

Keywords

COVID-19, Epidemic outbreak, Probability forecast, Public health, Mathematical biology

Open Peer Review

Approval Status

1

version 1

10 Nov 2022



[view](#)

1. **Oluwakemi E Abiodun** , Landmark University, Omu-Aran, Nigeria

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the **Emerging Diseases** and **Outbreaks** gateway.

Corresponding author: Yihan Xing (yihan.xing@uis.no)

Author roles: **Gaidai O:** Conceptualization, Software; **Yan P:** Resources; **Xing Y:** Validation; **Xu J:** Supervision; **Wu Y:** Formal Analysis

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2022 Gaidai O *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Gaidai O, Yan P, Xing Y *et al.* **A novel statistical method for long-term coronavirus modelling [version 1; peer review: 1 approved]** F1000Research 2022, **11**:1282 <https://doi.org/10.12688/f1000research.125924.1>

First published: 10 Nov 2022, **11**:1282 <https://doi.org/10.12688/f1000research.125924.1>

Introduction

Statistical aspects of coronavirus disease 2019 (COVID-19) and other similar recent epidemics were receiving much attention in the modern research community.¹ 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.^{1–14} 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 estimated straightforwardly by having enough measurements or by direct Monte Carlo simulations.¹⁰ For COVID-19, however, the only available observation numbers are limited as the observations are only available from the beginning of the year 2020 up to now. 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, this study was focused on COVID-19 epidemics in northern China, with focus on cross-correlations between different provinces within same climatic zone. For other studies related to statistical variations per country see,¹⁵ where spatial lags were addressed. China was chosen to test the methodology proposed in this study, because of its COVID-19 origin and extensive health observations and related research available online.^{15–29}

Statistical modelling of lifetime data or extreme value theory (EVT) is widespread in medicine and engineering. For example, Gumbel used EVT to estimate the demographic of various populations in.³⁰ Similarly, in Ref. 31, the author builds on this EVT theory to determine the fitness effect using a Beta-Burr distribution. While in Ref. 32, the author discusses using a bivariate logistic regression model, which was then used to access multiple sclerosis patients with walking disabilities and in a cognitive experiment for visual recognition. Finally, Ref. 33 is a paper of relevance, which used EVT to estimate the probability of an influenza outbreak in China. The author demonstrated a forecasting prediction potential amid the epidemic in this paper. While in Ref. 34 similarly used EVT to predict and detect anomalies of influenza epidemics.

In this paper epidemic outbreak is viewed as unexpected incident that may occur at any province of a given country at any time, therefore 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 being 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 provinces of the People's Republic of China (PRC) from February 2020 until today were retrieved from the official public PRC health website, for simplicity only northern provinces were selected for this study. As this dataset is organized by province (more than 30 provinces in China), the biological system under consideration can be regarded as a multi-degree of freedom (MDOF) dynamic system with highly inter-correlated regional components/dimensions. Some recent studies have already used statistical tools to predict COVID-19 development, for linear log model see Ref. 9, these studies however did not address fully dynamic space-time dynamic bio-system as this study does.

Note that while this study aims at reducing risk of future epidemic outbreaks by predicting them, it is solely focused on daily registered patient numbers and not on symptoms themselves. 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, see Ref. 35, for mortality research see Ref. 36.

Methods

This section presents novel bio-system reliability method, presently invented by authors and previously not existing. Advocated methodology used generalized extreme value (GEV) theory as a basis for its buildup. For numerical part authors used commercial software **MATLAB**, (Mathworks, V 8.6), namely its optimization routines, otherwise authors have used extrapolation code available from **ACER**. Only the code available from **ACER** was used to complete all sections of the methods – both numerical part as well as final extrapolation.

The novel method introduces MDOF (multi-degree of freedom) health response vector process $\mathbf{R}(t) = (X(t), Y(t), Z(t), \dots)$ that was measured over a sufficiently long period of time $(0, T)$. Unidimensional global maxima over the entire time span $(0, T)$ denoted as $X_T^{\max} = \max_{0 \leq t \leq T} X(t)$, $Y_T^{\max} = \max_{0 \leq t \leq T} Y(t)$, $Z_T^{\max} = \max_{0 \leq t \leq T} Z(t), \dots$

Let X_1, \dots, X_{N_x} be time local maxima of the process $X(t)$ consequent in time, recorded at discrete time instants $t_1^X < \dots < t_{N_x}^X$ that are monotonously increasing in $(0, T)$. A similar definition follows¹ for other MDOF response components $Y(t), Z(t), \dots$ with Y_1, \dots, Y_{N_y} ; Z_1, \dots, Z_{N_z} and so on. For simplicity, all $\mathbf{R}(t)$ components, and therefore its maxima are assumed to be non-negative.

The target is to estimate system failure probability, in other words the probability of exceedance

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

where $P = \iiint_{(0,0,0,\dots)}^{(\eta_X, \eta_Y, \eta_Z, \dots)} p_{X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \dots}(X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \dots) dX_T^{\max} dY_{N_Y}^{\max} dZ_{N_Z}^{\max} \dots$ being the probability of non-exceedance for critical values of response components $\eta_X, \eta_Y, \eta_Z, \dots$; \cup denotes logical unity operation «or»; and $p_{X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \dots}$ being joint 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 dataset 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 $\eta_X, \eta_Y, \eta_Z, \dots$ are, of course, individual for each unidimensional response component of $\mathbf{R}(t)$. $X_{N_X}^{\max} = \max \{X_j; j = 1, \dots, N_X\} = X_T^{\max}$, $Y_{N_Y}^{\max} = \max \{Y_j; j = 1, \dots, N_Y\} = Y_T^{\max}$, $Z_{N_Z}^{\max} = \max \{Z_j; j = 1, \dots, N_Z\} = Z_T^{\max}$, and so on.

Now, the local maxima time instants $[t_1^X < \dots < t_{N_X}^X; t_1^Y < \dots < t_{N_Y}^Y; t_1^Z < \dots < t_{N_Z}^Z]$ are sorted in monotonously non-decreasing order into one single merged time vector $t_1 \leq \dots \leq t_N$. 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 $\mathbf{R}(t)$ time record, one just needs continuously and simultaneously screen for unidimensional response component local maxima and record its exceedance of 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 $\bar{\mathbf{R}} = (R_1, R_2, \dots, R_N)$ following the merged time vector $t_1 \leq \dots \leq t_N$, see **Figure 1**. That is to say, each local maxima R_j is, in fact, actual encountered local maxima corresponding to either $X(t)$ or $Y(t)$, or $Z(t)$ and so on. Finally, the unified limit vector (η_1, \dots, η_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, the scaling parameter $0 < \lambda \leq 1$ is introduced to, artificially, simultaneously decrease limit values for all response components, namely the new MDOF limit vector $(\eta_1^\lambda, \eta_2^\lambda, \eta_3^\lambda, \dots)$ with $\eta_X^\lambda \equiv \lambda \cdot \eta_X, \eta_Y^\lambda \equiv \lambda \cdot \eta_Y, \eta_Z^\lambda \equiv \lambda \cdot \eta_Z, \dots$ is introduced, see Ref. 37. The unified limit vector $(\eta_1^\lambda, \dots, \eta_N^\lambda)$ is introduced with each component η_j^λ is either $\eta_X^\lambda, \eta_Y^\lambda$ 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 **Equation (1)**. Non-exceedance probability $P(\lambda)$ can be introduced as follows

$$\begin{aligned} P(\lambda) &= \text{Prob}\{R_N \leq \eta_N^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \\ &= \text{Prob}\{R_N \leq \eta_N^\lambda \mid R_{N-1} \leq \eta_{N-1}^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \cdot \text{Prob}\{R_{N-1} \leq \eta_{N-1}^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \\ &= \prod_{j=2}^N \text{Prob}\{R_j \leq \eta_j^\lambda \mid R_{j-1} \leq \eta_{j-1}^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \cdot \text{Prob}\{R_1 \leq \eta_1^\lambda\} \end{aligned} \tag{2}$$

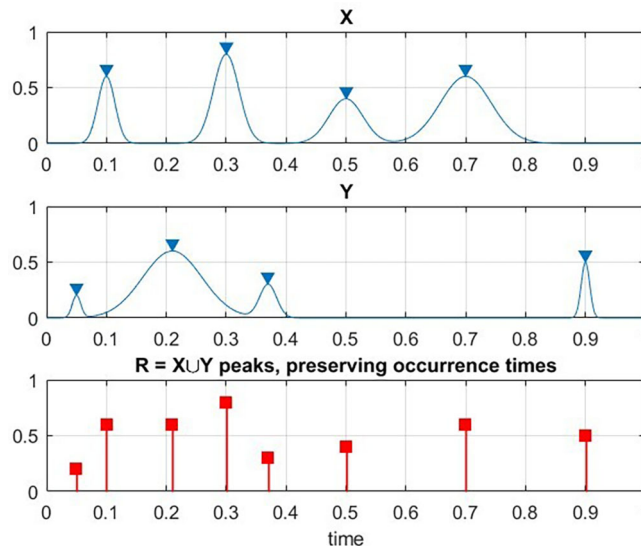


Figure 1. Illustration on how two exemplary processes X and Y are combined into new synthetic vector R.

Next, cascade of approximations which is based on conditioning is briefly outlined. In practice,^{11–14,38} the dependence between the neighboring R_j is not negligible; thus, the following one-step (will be called conditioning level $k = 1$) memory approximation is introduced

$$\text{Prob}\left\{R_j \leq \eta_j^{\lambda} \mid R_{j-1} \leq \eta_{j-1}^{\lambda}, \dots, R_1 \leq \eta_1^{\lambda}\right\} \approx \text{Prob}\left\{R_j \leq \eta_j^{\lambda} \mid R_{j-1} \leq \eta_{j-1}^{\lambda}\right\} \quad (3)$$

for $2 \leq j \leq N$ (conditioning level $k = 2$). The approximation introduced by Equation (3) can be further expressed as

$$\text{Prob}\left\{R_j \leq \eta_j^{\lambda} \mid R_{j-1} \leq \eta_{j-1}^{\lambda}, \dots, R_1 \leq \eta_1^{\lambda}\right\} \approx \text{Prob}\left\{R_j \leq \eta_j^{\lambda} \mid R_{j-1} \leq \eta_{j-1}^{\lambda}, R_{j-2} \leq \eta_{j-2}^{\lambda}\right\} \quad (4)$$

where $3 \leq j \leq N$ (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 (4) exhibits subsequent refinements with respect to the statistical independence assumption. These approximations increasingly accurately approximate statistical dependence between neighboring maxima. Since the original MDOF process $R(t)$ has been assumed ergodic and thus stationary, the probability $p_k(\lambda) := \text{Prob}\left\{R_j > \eta_j^{\lambda} \mid R_{j-1} \leq \eta_{j-1}^{\lambda}, R_{j-k+1} \leq \eta_{j-k+1}^{\lambda}\right\}$ for $j \geq k$ is independent of j , however dependent on the conditioning level k . Thus, non-exceedance probability may be approximated as in the average conditional exceedance rate method, see Refs. 30, 37 for more details on exceedance probability

$$P_k(\lambda) \approx \exp(-N \cdot p_k(\lambda)), k \geq 1. \quad (5)$$

Note that Equation (5) follows from Equation (1) if neglecting $\text{Prob}(R_1 \leq \eta_1^{\lambda}) \approx 1$, as design failure probability must be epsilon order $o(1)$, with $N \gg k$. Equation (5) is analogous to the well-known mean up-crossing rate equation for the stochastic process probability of exceedance.^{10,37} There is convergence with respect to k , called here conditioning level

$$P = \lim_{k \rightarrow \infty} P_k(1); p(\lambda) = \lim_{k \rightarrow \infty} p_k(\lambda) \quad (6)$$

Note that Equation (5) for $k = 1$ is equivalent to a well-known non-exceedance probability relationship with the mean up-crossing rate function

$$P(\lambda) \approx \exp(-v^+(\lambda)T); v^+(\lambda) = \int_0^{\infty} \zeta p_{R\dot{R}}(\lambda, \zeta) d\zeta \quad (7)$$

where $v^+(\lambda)$ 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 $\left(\frac{X}{\eta_X}, \frac{Y}{\eta_Y}, \frac{Z}{\eta_Z}, \dots\right)$. The mean up-crossing rate is given by the Rice's formula given in Equation (7) with $p_{R\dot{R}}$ being joint probability density for (R, \dot{R}) with \dot{R} being time derivative $R'(t)$. Equation (7) relies on the Poisson assumption that is up-crossing events of high λ levels (in this paper, it is $\lambda \geq 1$) can be assumed to be independent. The latter may not be the case for narrowband responses and higher-level dynamical 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 $\vec{R} = (R_1, R_2, \dots, R_N)$.

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 $m = 1, \dots, M$ 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

$$p_k(\lambda) \equiv \sum_{m=1}^M p_k(\lambda, m) q_m \quad (8)$$

with $p_k(\lambda, m)$ being the same function as in Equation (6) 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.^{10,37}

Implementing extrapolation method

Introduced by Equation (5) functions $p_k(\lambda)$ are 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

$$p_k(\lambda) \approx \exp\{-(a_k\lambda + b_k)^{c_k} + d_k\}, \lambda \geq \lambda_0 \tag{9}$$

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. Optimal values of the parameters a_k, b_k, c_k, p_k, q_k may also be determined applying sequential quadratic programming (SQP) methods, incorporated in NAG Numerical Library.³⁹ Methods described above have been applied as described in methods section. Authors used MATLAB (Mathworks, V 8.6) (RRID:SCR_001622) commercial tool as a basis for their numerical purposes. For more specific author developed code routines, related to the extrapolation method by Equation (9), see ACER. Note that ACER is a repository, containing not only the code, but user manual, examples and references. In this study only extrapolation part of ACER was used. In other words current study presents novel theoretical methodology, but using ACER software previously developed by some of the authors.

Ethical consideration

Authors confirm that all methods were performed in accordance with the relevant guidelines and regulations according to the Declarations of Helsinki.

Use case

Methods described in this paper are novel and state of art. Prediction of influenza type epidemics has long been the focus of attention in mathematical biology and epidemiology. It is known that public health dynamics is a seasonally and spatially varying dynamic system that is always challenging to analyse. 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 different regions.

COVID-19 and influenza are contagious diseases with high transmissibility and ability to spread. Seasonal influenza epidemics caused by influenza A and B viruses typically occur annually during winter, presenting a 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).³

This section presents a real-life application of the above-described method. The statistical data in the present section are taken from the official website of the National Health Commission of the people's Republic of China. The website provides the number of newly diagnosed cases every day in 13 administrative regions in northern China from 22 January 2020 to 6 April 2022. Patient numbers from thirteen different Chinese administrative regions were chosen as components X, Y, Z, \dots thus constituting an example of a thirteen dimensional (13D) dynamic biological system.

In order to unify all three measured time series X, Y, Z , the following scaling was performed as follows

$$X \rightarrow \frac{X}{\eta_X}, Y \rightarrow \frac{Y}{\eta_Y}, Z \rightarrow \frac{Z}{\eta_Z}, \dots \tag{10}$$

making all three responses non-dimensional and having the same failure limit equal to 1. Failure limits $\eta_X, \eta_Y, \eta_Z, \dots$, or in other words, epidemic thresholds, are not set values and must be decided. The simplest choice would be for different countries to set failure limits equal to the percentage per corresponding country population, making X, Y, Z, \dots equal to percentage of daily infected per country. In this study, however, twice maxima of daily infected per country have been chosen as failure limits. Note that failure limits may be chosen differently for different dynamic bio-systems. Although the latter choice obviously introduces bias (accumulation point) at $\lambda = 0.5$ if the number of countrys is large, in this study the number of regions is not that large (below 20 national regions, see Ref. 30 for similar USA study) and proper extrapolation technique may easily circumvent the above mentioned accumulation point bias.

Next, all local maxima from three measured time series were merged into one single time series by keeping them in time non-decreasing order: $\vec{R} = (\max\{X_1, Y_1, Z_1, \dots\}, \dots, \max\{X_N, Y_N, Z_N, \dots\})$ with the whole \vec{R} vector being sorted according to non-decreasing times of occurrence of these local maxima.

Figure 2 presents new daily recorded patients number plotted as a time-space 2D surface using MATLAB. Figure 3 presents the number of new daily recorded patients as a 13D vector \vec{R} , consisting of assembled regional new daily patient

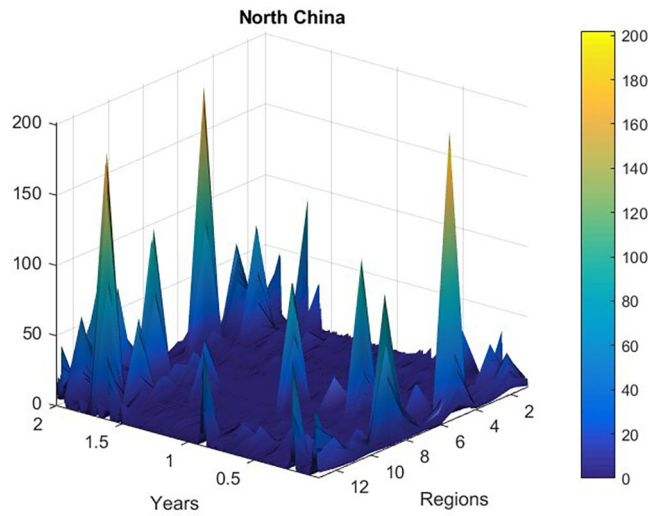


Figure 2. New daily recorded patients number plotted as a 2D time-space surface, data was according to <http://www.nhc.gov.cn>.

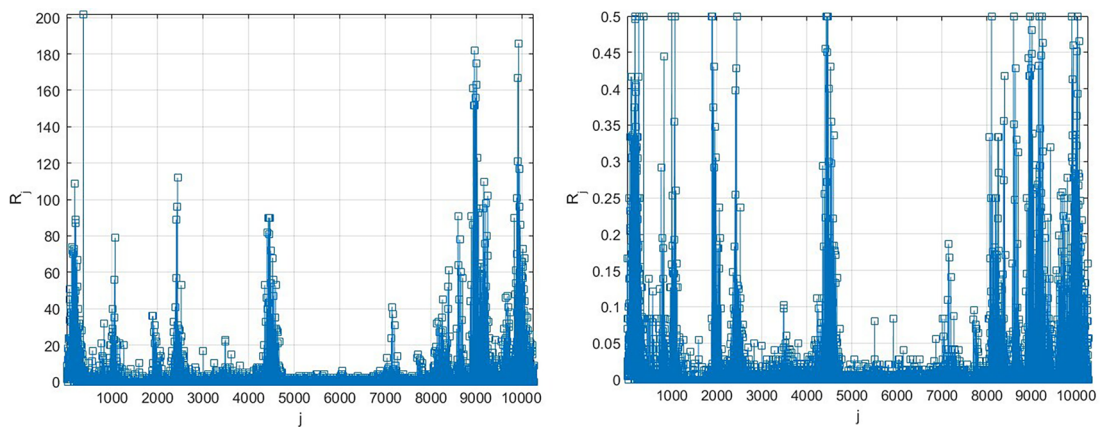


Figure 3. Number of new daily recorded patients as 13D vector \vec{R} . Left: as it is, Right: scaled by Equation (10).

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 Equation (9) towards epidemic outbreak with 100 year return period, indicated by the horizontal dotted line, and somewhat beyond, $\lambda = 0.1$ cut-on value was used. Dotted lines indicate extrapolated 95% confidence interval according to Equation (10). According to Equation (5) $p(\lambda)$ is directly related to the target failure probability $1 - P$ from Equation (1). Therefore, in agreement with Equation (5), system failure probability $1 - P \approx 1 - P_k(1)$ can be estimated. Note that in Equation (5), 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 occurrence of convergence with respect to k , see Equation (6). Figure 4 exhibits reasonably narrow 95% CI. The latter is an advantage of the proposed method.

Note that while being novel, the above-described methodology has a clear advantage of utilizing available measured data set quite efficiently due to its ability to treat health system multi-dimensionality and perform accurate extrapolation based on quite limited data set. Note that, predicted non-dimensional λ level, indicated by star in Figure 4, represents probability of epidemic outbreak at any northern province in China in the years to come.

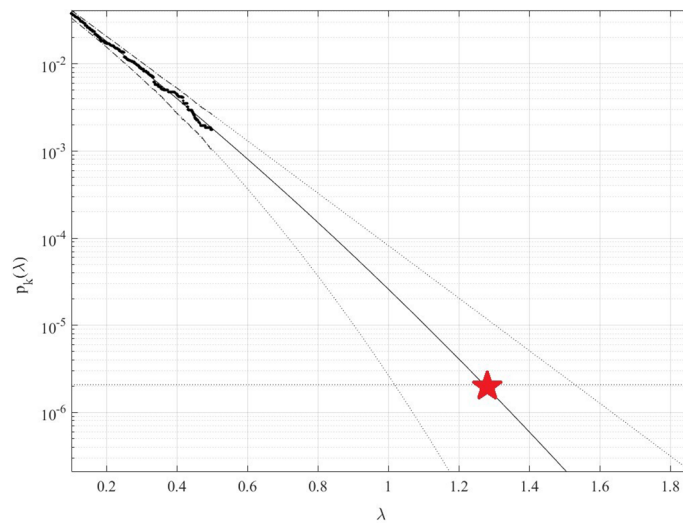


Figure 4. 100 years return level (horizontal dotted line) extrapolation of $p_k(\lambda)$ towards critical (system failure) level (indicated by star) and beyond. Extrapolated 95% CI indicated by dotted lines, using <https://github.com/cran/acer>.

Discussion

Traditional health bio-systems reliability methods dealing with observed/measured time series do not have the ability to efficiently deal with high dimensionality and cross-correlation between different system responses. The main advantage of the introduced methodology is its ability to study high dimensional non-linear dynamic systems reliability.

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, if it is assumed to be stationary in time. Proper setting of epidemiological alarm limits (failure limits) per province has been discussed, see Section Use case.

This paper studied recorded COVID-19 patient numbers from thirteen different Chinese northern provinces, constituting an example of a thirteen dimensional (13D) and ten-dimensional (10D) dynamic biological system respectively observed in 2020-2022. The novel reliability method was applied to new daily patient numbers as a multidimensional system in real-time.³⁰

The main conclusion is that if the public health system under local environmental and epidemiologic conditions in northern China is well managed. Predicted 100 year return period risk level λ of epidemic outbreak is reasonably low. However, there is an ultra-low risk of a future epidemic outbreak in both countries, at least in 100 years horizon.

This study outlines a general-purpose, robust and straightforward multidimensional bio-system reliability method. The method introduced in this study has been previously successfully validated by application to a wide range of engineering models,^{30,40-49} but only for one and two-dimensional system responses and, in general, very accurate predictions were obtained. Both measured and simulated time series responses can be analysed using the proposed method. It is shown that the method produced an acceptable 95% confidence interval, see Figure 4. Thus, the suggested methodology may be used as a tool in various non-linear dynamic biological systems reliability studies. The presented COVID-19 example does not limit potential areas of new method applicability by any means.

Data availability

The datasets analyzed during the current study are publicly available from the daily recorded patients dataset in China during 2020-2022 years, are available at <http://www.nhc.gov.cn/>. Daily recorded patients data was structured per province and per calendar day, namely it was straightforward to extract and systematize joint statistical distribution as a function of both space and time.

Software availability

Source code along with demo and user manual and examples used for extrapolation available from: ACER this third-party software is under license GPL-3.

Acknowledgements

The authors declare no conflict of interest. No funding was received.

References

- Elizabeth J, *et al.*: **Factors associated with COVID-19-related death using OpenSAFELY.** *Nature.* 2020; **584**(20): 430–436.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Chen J, Lei X, Zhang L, *et al.*: **Using Extreme Value Theory Approaches to Forecast the Probability of Outbreak of Highly Pathogenic Influenza in Zhejiang, China.** *PLoS One.* 2015; **10**(2): e0118521.
[PubMed Abstract](#) | [Publisher Full Text](#)
- World Health Organization: **Influenza fact sheet.** *World Health Organization website.* Geneva: World Health Organization; 2014 Mar [cited 10 June 2014]. 1948 -. [about 2 screens].
[Reference Source](#)
- Goldstein E, Cobey S, Takahashi S, *et al.*: **Predicting the epidemic sizes of influenza A/H1N1, A/H3N2, and B: a statistical method.** *PLoS Med.* 2011; **8**: e1001051.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Soebiyanto RP, Adimi F, Kiang RK: **Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters.** *PLoS One.* 2010; **5**: e9450.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mugglin AS, Cressie N, Gemmell I: **Hierarchical statistical modelling of influenza epidemic dynamics in space and time.** *Stat. Med.* 2002; **21**(18): 2703–2721.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kim EK, Seok JH, Oh JS, *et al.*: **Use of Hangeul Twitter to track and predict human influenza infection.** *PLoS One.* 2013; **8**: e69305.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lee HC, Wackernagel H: **Extreme value analyses of US P&I mortality data under consideration of demographic effects.** *Centre de géosciences/Géostatistique Publications & documentation.* Fontainebleau: Mine ParisTech; 2007 [cited 10 June 2014].
[PubMed Abstract](#) | [Publisher Full Text](#) | [Reference Source](#)
- Chu J: **A statistical analysis of the novel coronavirus (COVID-19) in Italy and Spain.** *PLoS One.* 2021; **16**: e0249037.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Naess A, Gaidai O: **Estimation of extreme values from sampled time series.** *Struct. Saf.* 2009; **31**(4): 325–334.
[Publisher Full Text](#)
- Madsen HO, Krenk S, Lind NC: *Methods of structural safety.* Englewood Cliffs: Prentice-Hall Inc; 1986.
- Ditlevsen O, Madsen HO: *Structural reliability methods.* Chichester (UK): John Wiley & Sons, Inc; 1996.
- Melchers RE: *Structural reliability analysis and prediction.* New York: John Wiley & Sons, Inc; 1999.
- Choi S-K, Grandhi RV, Canfield RA: *Reliability-based structural design.* London: Springer-Verlag; 2007.
- Gondaouri D, Mikautadze E, Batiashvili M: **Research on COVID-19 Virus Spreading Statistics based on the Examples of the Cases from Different Countries.** *Electron J. Gen. Med.* 2020; **17**(4). Article No: em209.
[Publisher Full Text](#)
- Zhu N, Zhang D, Wang W, *et al.*: **A Novel Coronavirus from Patients with Pneumonia in China, 2019.** *N. Engl. J. Med.* 2020; **382**: 727–733.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Wu JT, Leung K, Leung GM: **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. Hopkins Johns. University Center for Systems and Science Engineering. Coronavirus COVID-19 Global Cases. (Accessed March 25, 2020).
[Publisher Full Text](#) | [Reference Source](#)
- He F, Deng Y, Li W: **Coronavirus Disease 2019 (COVID-19): What we know?** *J. Med. Virol.* 2020; **92**: 719–725.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Wu Z, McGoogan JM: **Characteristics of and Important Lessons from the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases from the Chinese Center for Disease Control and Prevention.** *JAMA.* 2020; **323**: 1236–1239.
[Publisher Full Text](#)
- Lu R, Zhao X, Li J, *et al.*: **Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding.** *Lancet.* 2020; **395**(10224): 565–574.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Zhou P, Yang XL, Wang XG, *et al.*: **A pneumonia outbreak associated with a new coronavirus of probable bat origin.** *Nature.* 2020; **579**: 270–273.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zhu N, Zhang D, Wang W, *et al.*: **A Novel Coronavirus from Patients with Pneumonia in China, 2019.** *N. Engl. J. Med.* 2020; **382**(8): 727–733.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Organization WH: **Coronavirus disease 2019 (COVID-19) Situation Report - 70.** 30 March 2020.
- Wood PHN: **The Mathematical Theory of Infectious Diseases and its applications.** *Immunology.* 1978; **34**(5): 955–956.
- Bailey NTJ: **The total size of a general stochastic epidemic.** *Biometrika.* 1953; **40**: 177–185.
[Publisher Full Text](#)
- Becker NG, Britton T: **Statistical studies of infectious disease incidence.** *J. R. Statist. Soc. B.* 1999; **61**(2): 287–307.
[Publisher Full Text](#)
- Lan L, Xu D, Ye G, *et al.*: **Positive RT-PCR Test Results in Patients Recovered from COVID-19.** *JAMA.* Published online February 27, 2020; **323**: 1502–1503.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kermack WO, McKendrick AG: **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**(772): 700–721.
[Publisher Full Text](#)
- Bailey NTJ: **Maximum-likelihood estimation of the relative removal rate from the distribution of the total size of an intra household epidemic.** *J. Hyg. (Lond).* 1954; **52**(3): 400–402.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gaidai O, Xing Y, Xu X: **COVID-19 epidemic forecast in USA East coast by novel reliability approach.** *Res. Sq.* 2022.
[Publisher Full Text](#)
- Joyce P, Abdo Z: **Determining the distribution of fitness effects using a generalised Beta-Burr distribution.** *Theor. Popul. Biol.* 2018; **122**: 88–96.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kristensen SB, Bibby BM: **A bivariate logistic regression model based on latent variables.** *Stat. Med.* 2020; **39**(22): 2962–2979.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Chen J, Lei X, Zhang L, *et al.*: **Using extreme value theory approaches to forecast the probability of outbreak of highly pathogenic influenza in Zhejiang, China.** *PLoS One.* 2015; **10**(2): e0118521.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Thomas M, Rootzen H: **Real-time prediction of severe influenza epidemics using Extreme Value Statistics.** arXiv preprint arXiv: 1910.10788. 2019.
- Sudre C, *et al.*: **Attributes and predictors of long COVID.** *Nat. Med.* 2021; **27**: 626–631.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Williamson EJ, Walker AJ, Bhaskaran K, *et al.*: **Factors associated with COVID-19-related death using OpenSAFELY.** *Nature.* 2020; **584**(20): 430–436.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Naess A, Moan T: *Stochastic dynamics of marine structures.* Cambridge University Press; 2013.
- Gumbel EM: *Statistics of extremes.* Columbia: Columbia University Press; 1958.
- Numerical Algorithms Group: *NAG Toolbox for Matlab.* Oxford, UK: NAG Ltd; 2010.

40. Xing Y, Gaidai O, Ma Y, *et al.*: **A novel design approach for estimation of extreme responses of a subsea shuttle tanker hovering in ocean current considering aft thruster failure.** *Appl. Ocean Res.* 2022; **123**: 103179.
[Publisher Full Text](#)
41. Gaidai O, Wang F, Wu Y, *et al.*: **Offshore renewable energy site correlated wind-wave statistics.** *Probabilistic Eng. Mech.* 2022; **68**: 103207.
[Publisher Full Text](#)
42. Sun J, Gaidai O, Wang F, *et al.*: **Extreme riser experimental loads caused by sea currents in the Gulf of Eilat.** *Probabilistic Eng. Mech.* 2022; **68**: 103243.
[Publisher Full Text](#)
43. Xu X, Wang F, Gaidai O, *et al.*: **Bivariate statistics of floating offshore wind turbine dynamic response under operational conditions.** *Ocean Eng.* 2022; **257**: 111657.
[Publisher Full Text](#)
44. Gaidai O, Xing Y, Wang F, *et al.*: **Improving extreme anchor tension prediction of a 10-MW floating semi-submersible type wind turbine, using highly correlated surge motion record.** *Front. Mech. Eng.* 2022; **8**.
[Publisher Full Text](#)
45. Xu X, Xing Y, Gaidai O, *et al.*: **A novel multi-dimensional reliability approach for floating wind turbines under power production conditions.** *Front. Mar. Sci.* 2022; **9**.
[Publisher Full Text](#)
46. Gaidai O, Xing Y, Balakrishna R: **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.
[Publisher Full Text](#)
47. Cheng Y, Gaidai O, Yurchenko D, *et al.*: **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.
48. Gaidai O, Storhaug G, Wang F, *et al.*: **On-Board Trend Analysis for Cargo Vessel Hull Monitoring Systems.** *The 32nd International Ocean and Polar Engineering Conference, Paper Number: ISOPE-I-22-541.* 2022.
49. Gaidai O, Xu X, Naess A, *et al.*: **Bivariate statistics of wind farm support vessel motions while docking.** *Ships and offshore structures.* 2020; **16**(2): 135–143.

Open Peer Review

Current Peer Review Status: 

Version 1

Reviewer Report 29 November 2022

<https://doi.org/10.5256/f1000research.138279.r155545>

© 2022 Abiodun O. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Oluwakemi E Abiodun 

Department of Physical Sciences, Landmark University, Omu-Aran, Nigeria

This study came up with a new way to look at the dependability of biosystems. This method was tested for a long enough time to make a credible long-term prediction of the probability of the extreme new coronavirus mortality rate. This methodology was applicable to the environmental and health systems in numerous nations. Provinces that were examined estimated their future maximum patient populations. This study introduces a method for assessing the dependability of multidimensional biosystems. The suggested methodology can be utilized as a tool in a variety of non-linear, dynamic biological systems reliability studies.

Oleg and his colleagues come up with the MDOF (multiple degrees of freedom) health response vector process. They do this in a new way. In addition, a novel new statistical approach—a generalized extreme value (GEV)—that was directly applied to unprocessed clinical data and is capable of dealing with territorial mapping was used as the foundation for model construction. This part of the study gives future researchers a new idea about how they might try to predict how an outbreak will change over time.

This paper is, in my opinion, an excellent contribution to the literature.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to

ensure full reproducibility?

Partly

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.**Reviewer Expertise:** Mathematical and statistical modeling of infectious disease**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research