



METHOD ARTICLE

**REVISED** Gaidai reliability method for long-term coronavirus modelling [version 3; peer review: 1 approved, 3 approved with reservations]

Previously titled: A novel statistical method for long-term coronavirus modelling

Oleg Gaidai <sup>1</sup>, Ping Yan<sup>1</sup>, Yihan Xing <sup>2</sup>, JingXiang Xu<sup>1</sup>, Yu Wu<sup>1</sup>

<sup>1</sup>Engineering Research Center of Marine Renewable Energy, Shanghai Ocean University, Shanghai, China

<sup>2</sup>Department of Mechanical and Structural Engineering and Materials Science, University of Stavanger, Stavanger, Norway

**V3** First published: 10 Nov 2022, 11:1282  
<https://doi.org/10.12688/f1000research.125924.1>  
 Second version: 22 May 2023, 11:1282  
<https://doi.org/10.12688/f1000research.125924.2>  
 Latest published: 21 Nov 2023, 11:1282  
<https://doi.org/10.12688/f1000research.125924.3>

**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**

**Open Peer Review**

Approval Status    

	1	2	3	4
<b>version 3</b> (revision) 21 Nov 2023			 view	 view
<b>version 2</b> (revision) 22 May 2023		 view		
<b>version 1</b> 10 Nov 2022	 view	 view		

- Oluwakemi E Abiodun** , Landmark University, Omu-Aran, Nigeria
- Adnan Khan** , Lahore University of Management Sciences, Lahore, Pakistan
- Md. Ashraful Babu**, International University of Business Agriculture and Technology, Dhaka, Bangladesh  
Independent University, Bangladesh, Dhaka, Bangladesh
- Eduard Campillo-Funollet** , Lancaster University, Lancaster, UK

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.

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

## 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



This article is included in the **Bioinformatics** gateway.



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

**Corresponding authors:** Oleg Gaidai ([o\\_gaidai@just.edu.cn](mailto:o_gaidai@just.edu.cn)), Yihan Xing ([yihan.xing@uis.no](mailto: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:** © 2023 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.* **Gaidai reliability method for long-term coronavirus modelling [version 3; peer review: 1 approved, 3 approved with reservations]** F1000Research 2023, 11:1282 <https://doi.org/10.12688/f1000research.125924.3>

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

**REVISED Amendments from Version 2**

The authors have addressed language and formatting concerns. Equations have been improved, references extended.

**Any further responses from the reviewers can be found at the end of the article**

**Introduction**

Statistical aspects of coronavirus disease 2019 (COVID-19) and other similar recent epidemics were receiving much attention in the modern research community.<sup>1–11</sup> 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.<sup>12–19</sup> The latter is usually due to the fact that dynamic biological and environmental systems possess high number of degrees of freedom (in other words dimensional components), as well as system dependency on location – making bio-system of interest spatio-temporal. In principle, the reliability of a complex biological system may be accurately estimated straightforwardly by having enough measurements or by direct Monte Carlo simulations.<sup>20–23</sup> 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.<sup>20</sup> 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 Ref. 24, 25, where spatial lags were addressed.<sup>26–29</sup>

Statistical modelling of lifetime data or extreme value theory (EVT) is widespread in medicine and engineering.<sup>32–40</sup> For example, in Ref. 13, authors employed EVT theory to determine the fitness effect using a Beta-Burr distribution. While in Ref. 14, 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, 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 authors similarly used EVT to predict and detect anomalies of influenza epidemics.<sup>41–44</sup> 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.<sup>45–50</sup>

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 bio-system failure/hazard factor  $\lambda$  is introduced to unify various bio-system components having different failure/hazard limits.<sup>51–58</sup>

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.<sup>59–65</sup> The incidence data of COVID-19 in all provinces of the People's Republic of China (PRC) from February 2020 until end of 2022 was 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. 25, these studies however did not address fully dynamic space-time dynamic bio-system as this study does.<sup>66–72</sup>

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. 18, for mortality research see Ref. 1.

**Gaidai method**

This section presents theoretical details of novel Gaidai bio-reliability method. For numerical part authors used commercial software MATLAB, (Mathworks, V 8.6), namely its optimization routines, otherwise authors have used extrapolation code available from modified Weibull. Only the code available from modified Weibull was used to complete all sections of the methods – both numerical part as well as final extrapolation.

Novel Gaidai bio-reliability method introduces MDOF (multi-degree of freedom) health bio-system vector process  $(X(t), Y(t), Z(t), \dots)$  that was measured over a sufficiently long (representative) period of time  $(0, T)$ . Unidimensional

bio-system components 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 follows for other MDOF bio-system 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 bio-system 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

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

where  $P = \iiint_{(0,0,\dots)}^{(\eta_X, \eta_Y, \eta_Z, \dots)} p_{X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \dots}(X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \dots) dX_T^{\max} dY_T^{\max} dZ_T^{\max} \dots$  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 (PDF) 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  $\eta_X$ , or  $Y(t)$  exceeds  $\eta_Y$ , or  $Z(t)$  exceeds  $\eta_Z$ , 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 system one-dimensional component.  $X_{N_X}^{\max} = \max\{X_j; j = 1, \dots, N_X\} = X_T^{\max}$ ,  $Y_{N_Y}^{\max} = \max\{Y_j; j = 1, \dots, N_Y\} = Y_T^{\max}$ ,  $Z_{N_Z}^{\max} = \max\{Z_j; j = 1, \dots, N_Z\} = Z_T^{\max}$ , and so on.

Now, bio-system components 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  $t_N = \max\{t_{N_X}^X, t_{N_Y}^Y, t_{N_Z}^Z, \dots\}$ ,  $N \leq N_X + N_Y + N_Z + \dots$ . 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 bio-system 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  $X, Y, Z, \dots$ . In order to unify all three measured time series  $X, Y, Z$ , the following scaling was performed as follows

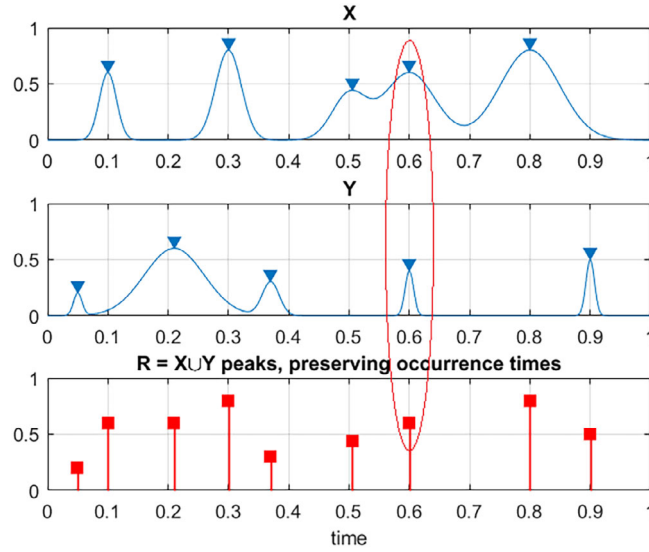
$$X \rightarrow \frac{X}{\eta_X}, \quad Y \rightarrow \frac{Y}{\eta_Y}, \quad Z \rightarrow \frac{Z}{\eta_Z}, \dots \quad (2)$$

making all bio-system components non-dimensional and having the same failure limit equal to  $\lambda = 1$ . Unidimensional bio-system component's local maxima are merged into a single temporal non-decreasing system vector  $\mathbf{R}(t) \equiv \vec{R} = (R_1, R_2, \dots, R_N)$  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 bio-system components local maxima corresponding to either  $X(t)$  or  $Y(t)$ , or  $Z(t)$  and so on. Finally, the unified limit/hazard bio-system vector  $(\eta_1, \dots, \eta_N)$  is introduced with each component  $\eta_j$  is either  $\eta_X, \eta_Y$  or  $\eta_Z$  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  $j$ . In case of simultaneous occurrence of various bio-system components  $X(t), Y(t), Z(t), \dots$  local maxima at the same time instant  $t_j$ , the maximum of those limits will be taken, for example -  $\eta_j = \max\{\eta_X, \eta_Y, \eta_Z, \dots\}$ .

Now, the scaling parameter  $0 < \lambda \leq 1$  is introduced to, artificially, simultaneously decrease limit values for all bio-system components, namely the new MDOF limit vector  $(\eta_1^\lambda, \eta_2^\lambda, \eta_3^\lambda, \dots)$  with  $\eta_1^\lambda \equiv \lambda \cdot \eta_1, \eta_2^\lambda \equiv \lambda \cdot \eta_2, \eta_3^\lambda \equiv \lambda \cdot \eta_3, \dots$  is introduced, see [Ref. 28](#). The unified limit vector  $(\eta_1^\lambda, \dots, \eta_N^\lambda)$  is introduced with each component  $\eta_j^\lambda$  is either  $\eta_X^\lambda, \eta_Y^\lambda$  or  $\eta_Z^\lambda$  and so on. The latter automatically defines probability  $P(\lambda)$  as a function of  $\lambda$ , note that  $P \equiv P(1)$  from [Equation \(1\)](#). Non-exceedance probability  $P(\lambda)$  can be introduced as follows

$$\begin{aligned} P(\lambda) &= \text{Prob}\{R_N \leq \eta_N^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \\ &= \text{Prob}\{R_N \leq \eta_N^\lambda | R_{N-1} \leq \eta_{N-1}^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \cdot \text{Prob}\{R_{N-1} \leq \eta_{N-1}^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \\ &= \text{Prob}(R_1 \leq \eta_1^\lambda) \cdot \prod_{j=2}^N \text{Prob}\{R_j \leq \eta_j^\lambda | R_{j-1} \leq \eta_{j-1}^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \end{aligned} \quad (3)$$

Next, cascade of approximations which is based on conditioning is briefly outlined. In practice,<sup>32–36</sup> 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



**Figure 1.** Illustration on how two exemplary processes  $X$  and  $Y$  are combined into new synthetic vector  $R$ . Red ellipse highlights case of simultaneous maxima for different components.

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

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

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

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 (5) 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  $p_k(\lambda) := \text{Prob}\{R_j > \eta_j^\lambda | R_{j-1} \leq \eta_{j-1}^\lambda, R_{j-k+1} \leq \eta_{j-k+1}^\lambda\}$  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 Ref. 28 for more details on exceedance probability

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

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.<sup>28</sup> There is convergence with respect to  $k$ , called here conditioning level

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

Note that Equation (6) for  $k=1$  is equivalent to a well-known non-exceedance probability relationship with the mean up-crossing rate function

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

where  $v^+(\lambda)$  denotes the mean up-crossing rate of the bio-system non-dimensional level  $\lambda$  for the above assembled non-dimensional vector  $R(t)$  assembled from scaled MDOF bio-system scaled components  $(\frac{X}{\eta_X}, \frac{Y}{\eta_Y}, \frac{Z}{\eta_Z}, \dots)$ . The mean up-crossing rate is given by the Rice's formula, given in Equation (7) with  $p_{R\dot{R}}$  being joint PDF 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  $\lambda$  levels (in this

paper, it is  $\lambda \geq 1$ ) can be assumed to be independent. The latter may not be the case for narrowband bio-system components 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

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

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.<sup>28,29</sup>

### Implementing modified Weibull extrapolation method

Introduced by Equation (5) functions  $p_k(\lambda)$  are regular in the tail, specifically for values of  $\lambda$  approaching and exceeding 1.<sup>17</sup> 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  $\lambda_0$  value. Therefore, one can write

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

Next, by plotting  $\ln\{\ln(p_k(\lambda)) - d_k\}$  versus  $\ln(a_k\lambda + b_k)$ , often nearly perfectly linear tail behavior 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.<sup>30</sup> 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 modified Weibull. Note that modified Weibull is a repository, containing not only the code, but user manual, examples and references. In this study only extrapolation part of modified Weibull was used. In other words, current study presents novel theoretical methodology, but using modified Weibull software previously developed by some of the authors.<sup>31</sup>

### 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).<sup>20</sup>

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 the end of 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, according to Equation (2).<sup>39–41</sup>



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 countries is large, in this study the number of regions is not that large (below 20 national regions) 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 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.

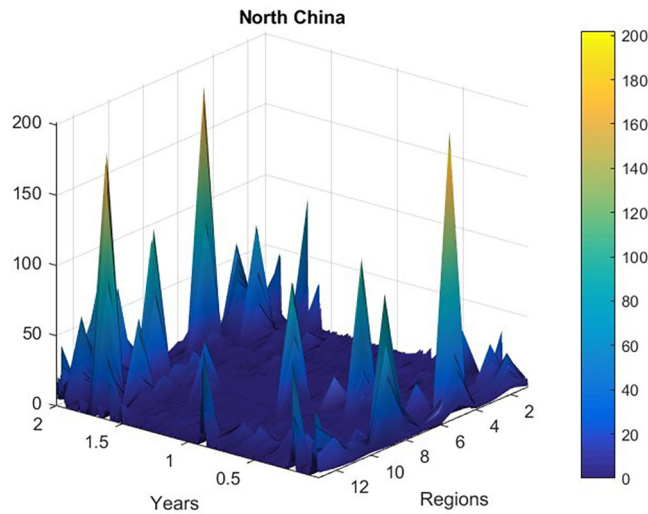


Figure 2. New daily recorded patients number plotted as a 2D time-space surface, data was according to <http://www.nhc.gov.cn>.

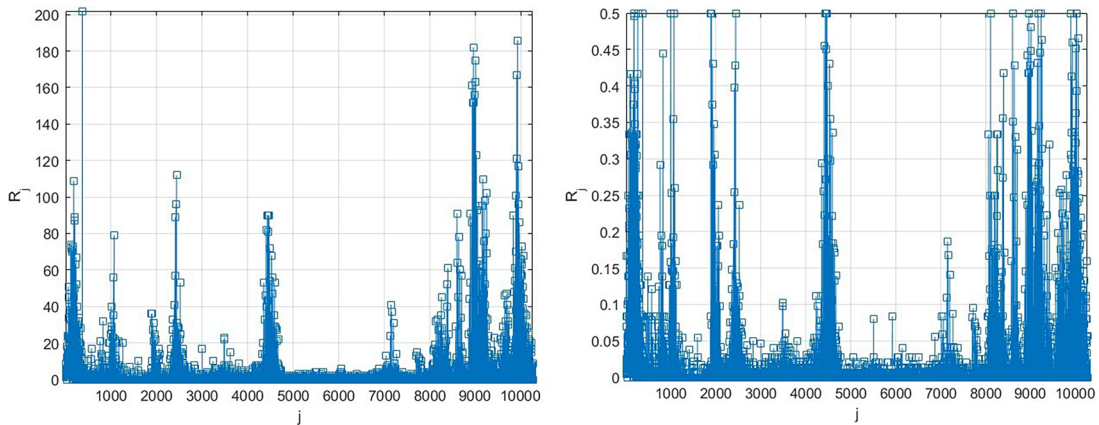
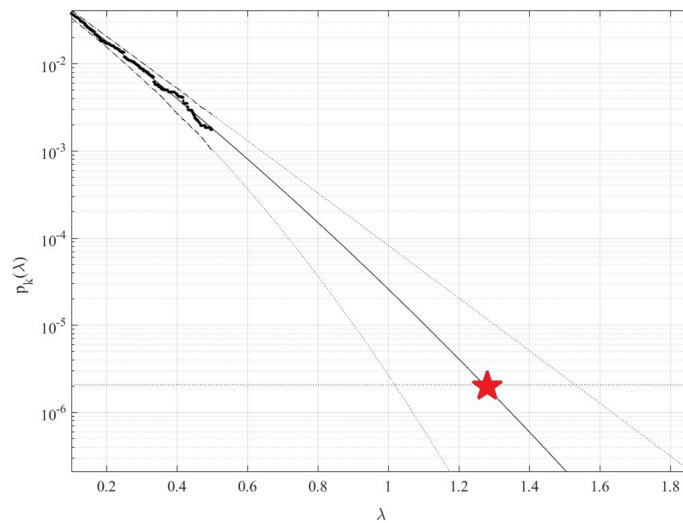


Figure 3. Number of new daily recorded patients as 13D vector  $\vec{R}$ . Left: as is, Right: scaled by Equation (10).



**Figure 4. 100 years return level (horizontal dotted line) extrapolation of  $p_k(\lambda)$  towards critical (biosystem failure) level (indicated by star) and beyond.** Extrapolated 95% CI indicated by dotted lines, using modified Weibull extrapolation technique.

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 bio-system 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  $\lambda$  level, indicated by star in Figure 4, represents probability of epidemic outbreak at any northern province in China in the years to come.

## 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 bio-system components. 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.

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  $\lambda$  of epidemic outbreak is reasonably low. However, there is an ultra-low risk of a future epidemic outbreak in chosen country of interest, 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,<sup>11,12</sup> but only for one and two-dimensional bio-system components and, in general, very accurate predictions



were obtained. Both measured and simulated bio-system components time series 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 patient's 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

- Gaidai O, Xing Y: **A novel bio-system reliability approach for multi-state COVID-19 epidemic forecast**. *Engineered Science*. 2022. [Publisher Full Text](#)
- Gaidai O, Xing Y: **A Novel Multi Regional Reliability Method for COVID-19 Death Forecast**. *Engineered Science*. 2022. [Publisher Full Text](#)
- Gaidai O, Wang F, Yakimov V: **COVID-19 multi-state epidemic forecast in India**. *Proceedings of the Indian National Science Academy*. 2023. [Publisher Full Text](#)
- Gaidai O, Wang F, Wu Y, et al.: **Offshore renewable energy site correlated wind-wave statistics**. *Probabilistic Engineering Mechanics*. 2022; **68**. [Publisher Full Text](#)
- Gaidai O, Cao Y, Xing Y, et al.: **Extreme springing response statistics of a tethered platform by deconvolution**. *International Journal of Naval Architecture and Ocean Engineering*. 2023. [Publisher Full Text](#)
- Gaidai O, Xing Y, Balakrishna R, et al.: **Improving extreme offshore wind speed prediction by using deconvolution**. *Heliyon*. 2023. [Publisher Full Text](#)
- Gaidai O, Xing Y: **Prediction of death rates for cardiovascular diseases and cancers**. *Cancer Innovation*. 2023. [Publisher Full Text](#)
- Xu X, Xing Y, Gaidai O, et al.: **A novel multi-dimensional reliability approach for floating wind turbines under power production conditions**. *Frontiers in Marine Science*. 2022; **9**. [Publisher Full Text](#)
- 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](#)
- 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.
- Gaidai O, Cao Y, Xing Y, et al.: **Piezoelectric Energy Harvester Response Statistics**. *Micromachines*. 2023; **14**(2): 271. [Publisher Full Text](#)
- Gaidai O, Cao Y, Loginov S: **Global cardiovascular diseases death rate prediction**. *Current Problems in Cardiology*. 2023. [Publisher Full Text](#)
- Joyce P, Abdo Z: **Determining the distribution of fitness effects using a generalised Beta-Burr distribution**. *Theoretical Population Biology*. 2018; **122**: 88–96. [PubMed Abstract](#) | [Publisher Full Text](#)
- Kristensen SB, Bibby BM: **A bivariate logistic regression model based on latent variables**. *Statistics in Medicine*. 2020; **39**(22): 2962–2979. [PubMed Abstract](#) | [Publisher Full Text](#)
- Zhang J, Gaidai O, Ji H, et al.: **Operational reliability study of ice loads acting on oil tanker bow**. *Heliyon*. 2023. [Publisher Full Text](#)
- Thomas M, Rootzen H: **Real-time prediction of severe influenza epidemics using Extreme Value Statistics**. arXiv preprint arXiv: 1910.10788. 2019.
- Xing Y, Gaidai O: **Multi-regional COVID-19 epidemic forecast in Sweden**. *Digital Health*. 2023; **9**. [Publisher Full Text](#)
- Sudre C, et al.: **Attributes and predictors of long COVID**. *Nature Medicine*. 2021; **27**: 626–631. [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**. *Statistics in Medicine*. 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](#)
- Chu J: **A statistical analysis of the novel coronavirus (COVID-19) in Italy and Spain**. *PLoS One*. 2021. [Publisher Full Text](#)
- Gaidai O, Xing Y: **COVID-19 Epidemic Forecast in Brazil**. *Bioinformatics Biology Insights*. 2023. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gaidai O, Xu X, Xing Y: **Novel deconvolution method for extreme FPSO vessel hawser tensions during offloading operations**. *Results in Engineering*. 2023; **17**. [Publisher Full Text](#)

28. Gaidai O, Xu J, Hu Q, *et al.*: **Offshore tethered platform springing response statistics**. *Scientific Reports*. 2022; **12**.  
[Reference Source](#)
29. Gaidai O, Xing Y, Xu X: **Novel methods for coupled prediction of extreme wind speeds and wave heights**. *Scientific Reports*. 2023.  
[Publisher Full Text](#)
30. Numerical Algorithms Group: *NAG Toolbox for Matlab*. Oxford, UK: NAG Ltd; 2010.
31. Rice SO: **Mathematical analysis of random noise**. *The Bell System Technical Journal*. 1944; **23**: 282–332.
32. Madsen HO, Krenk S, Lind NC: *Methods of structural safety*. Englewood Cliffs: Prentice-Hall Inc; 1986.
33. Ditlevsen O, Madsen HO: *Structural reliability methods*. Chichester (UK): John Wiley & Sons, Inc; 1996.
34. Melchers RE: *Structural reliability analysis and prediction*. New York: John Wiley & Sons, Inc; 1999.
35. Choi S-K, Grandhi RV, Canfield RA: *Reliability-based structural design*. London: Springer-Verlag; 2007.
36. Gaidai O, Xu J, Yan P, *et al.*: **Novel methods for reliability study of multi-dimensional non-linear dynamic systems**. *Scientific Reports*. 2023; **13**, 3817.  
[Publisher Full Text](#)
37. Pickands J: **Statistical inference using extreme order statistics**. *Annals of Statistics*. 1975; **3**(1): 119–131.
38. Zhao X: *Extreme value modelling with application in finance and neonatal research*. PhD Thesis. The University of Canterbury; 2010.  
[Reference Source](#)
39. Zheng L, Ismail K, Meng XH: **Freeway safety estimation using extreme value theory approaches: A comparative study**. *Accident Analysis and Prevention*. 2014; **62**: 32–41.  
[PubMed Abstract](#) | [Publisher Full Text](#)
40. McNeil AJ, Frey R, Embrechts P: *Quantitative risk management: Concepts, techniques and tools*. America: Princeton University Press; 2005.
41. Gaidai O, Wang F, Xing Y, *et al.*: **Novel Reliability Method Validation for Floating Wind Turbines**. *Advanced Energy and Sustainability Research*. 2023.  
[Publisher Full Text](#)
42. Sumi A, Kamo KI: **MEM spectral analysis for predicting influenza epidemics in Japan**. *Environmental Health and Preventive Medicine*. 2012; **17**: 98–108.  
[PubMed Abstract](#) | [Publisher Full Text](#)
43. Songchitruksa P, Tarko AP: **The extreme value theory approach to safety estimation**. *Accident Analysis and Prevention*. 2006; **38**: 811–822.  
[PubMed Abstract](#)
44. Gaidai O, Cao Y, Xu X, *et al.*: **Offloading operation bivariate extreme response statistics for FPSO vessel**. *Scientific reports*. 2023.  
[Publisher Full Text](#)
45. Gondauri D, Mikautadze E, Batiashvili M: **Research on COVID-19 Virus Spreading Statistics based on the Examples of the Cases from Different Countries**. *Electronic Journal of General Medicine*. 2020; **17**(4). Article No: em209.  
[Publisher Full Text](#)
46. Zhu N, Zhang D, Wang W, *et al.*: **A Novel Coronavirus from Patients with Pneumonia in China, 2019**. *The New England Journal of Medicine*. 2020; **382**: 727–733.  
[PubMed Abstract](#) | [Publisher Full Text](#)
47. 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](#)
48. He F, Deng Y, Li W: **Coronavirus Disease 2019 (COVID-19): What we know?** *Journal of Medical Virology*. 2020; **92**: 719–725.  
[PubMed Abstract](#) | [Publisher Full Text](#)
49. 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](#)
50. 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](#)
51. 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](#)
52. Organization WH: **Coronavirus disease 2019 (COVID-19) Situation Report - 70**. 30 March 2020.  
[Reference Source](#)
53. Wood PHN: **The Mathematical Theory of Infectious Diseases and its applications**. *Immunology*. 1978; **34**(5): 955–956.
54. Bailey NTJ: **The total size of a general stochastic epidemic**. *Biometrika*. 1953; **40**: 177–185.  
[Publisher Full Text](#)
55. Becker NG, Britton T: **Statistical studies of infectious disease incidence**. *Journal of the Royal Statistical Society: Series B*. 1999; **61**(2): 287–307.  
[Publisher Full Text](#)
56. Lan L, Xu D, Ye G, *et al.*: **Positive RT-PCR Test Results in Patients Recovered from COVID-19**. *JAMA*. 2020 Apr 21; **323**: 1502–1503.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
57. 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](#)
58. Bailey NTJ: **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**(3): 400–402.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
59. Gaidai O, Xu J, Yakimov V, *et al.*: **Analytical and Computational Modeling for Multi-Degree of Freedom Systems: Estimating the Likelihood of an FOWT Structural Failure**. *Journal of Marine Science and Engineering*. 2023; **11**(6): 1237.  
[Publisher Full Text](#)
60. Sun J, Gaidai O, Xing Y, *et al.*: **On safe offshore energy exploration in the Gulf of Eilat**. *Quality and Reliability Engineering International*. 2023.  
[Publisher Full Text](#)
61. Gaidai O, Xu J, Yakimov V, *et al.*: **Liquid carbon storage tanker disaster resilience**. *Environment Systems and Decisions*. 2023.  
[Publisher Full Text](#)
62. Yakimov V, Gaidai O, Wang F, *et al.*: **Fatigue assessment for FPSO hawasers**. *International Journal of Naval Architecture and Ocean Engineering*. 2023.  
[Publisher Full Text](#)
63. Yakimov V, Gaidai O, Wang F, *et al.*: **Arctic naval launch and recovery operations, under ice impact interactions**. *Applications in Engineering Science*. 2023.  
[Publisher Full Text](#)
64. Gaidai O, Yakimov V, Wang F, *et al.*: **Lifetime assessment for container vessels**. *Applied Ocean Research*. 2023.  
[Publisher Full Text](#)
65. Gaidai O, Wang F, Yakimov V, *et al.*: **Lifetime assessment for riser systems**. *GRN TECH RES SUSTAIN*. 2023; **3**.  
[Publisher Full Text](#)
66. Gaidai O, Yakimov V, Zhang F: **COVID-19 spatio-temporal forecast in England**. *Biosystems*. 2023.  
[Publisher Full Text](#)
67. Gaidai O, Liu Z, Wang K, *et al.*: **Current COVID-19 Epidemic Risks in Brazil**. *Epidemiology International Journal*. 2023; **7**(2): 1–10.  
[Publisher Full Text](#)
68. Gaidai O, Yakimov V, Balakrishna R: **Dementia death rates prediction**. *BMC Psychiatry*. 2023; **23**(691).  
[Publisher Full Text](#)
69. Gaidai O, Yakimov V, Wang F, *et al.*: **Floating wind turbines structural details fatigue life assessment**. *Scientific Reports*. 2023; **13**(1).  
[Publisher Full Text](#)
70. Gaidai O, Yakimov V, Wang F, *et al.*: **Safety design study for energy harvesters**. *Sustainable Energy Research*. 2023; **10**(1).  
[Publisher Full Text](#)
71. Gaidai O, Yakimov V, van Loon E: **Influenza-type epidemic risks by spatio-temporal Gaidai-Yakimov method**. *Dialogues Health*. 2023; **3**(2).  
[Publisher Full Text](#)
72. Gaidai O, Yakimov V, Niu Y, *et al.*: **Gaidai-Yakimov reliability method for high-dimensional spatio-temporal biosystems**. *Biosystems*. 2023.  
[Publisher Full Text](#)

# Open Peer Review

Current Peer Review Status:    

Version 3

Reviewer Report 01 April 2024

<https://doi.org/10.5256/f1000research.158999.r242811>

© 2024 Campillo-Funollet E. 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.

 **Eduard Campillo-Funollet** 

Mathematics and Statistics, Lancaster University, Lancaster, England, UK

The manuscript presents the application of an extreme-value methodology to the analysis of epidemiological data. The methodology have been applied in the past to other types of data. The manuscript illustrates the methodology using data from the COVID-19 pandemic in China.

I have several reservations about the manuscript, in particular regarding the technical background and the conclusions.

Major concerns:

- The claim that this study "fully space-time dynamic biosystem" (Introduction, paragraph beginning with "Biological systems...") is not justified. If anything, the proposed method does not account for the dynamics of the disease, and in fact it relies on the assumption of stationary data.
- Similarly, the assumption of the process being stationary is a strong assumption in the context of an evolving epidemic. It is mentioned briefly but its implications should be discussed.
- The description of the methods mixes using three components, with using many. See for example the paragraph before equation 2. The presentation would be clearer if the notation was adapted to use a generic number of components  $n$ , instead of using three and saying "etcetera", for exapmle in "X,Y,Z,...". This also affects the rest of the equations in the presentaiton, for example  $\eta_j$  is not clearly defined in (3).
- The assumption of an exponential tail ("exceedance probability", eq. 6) should be clearly justified in paper, rather than citing an application paper.
- The 100-year return level in Fig 4, which summarises very well the applicability of the method, should be discussed in detail. The data is from a two year period during the COVID-19 pandemic, and the method essentially captures the peaks of new cases during that time. However, the changes in the population (e.g. immunity or vaccinations) make the past data not representative of the current or future situations, and from an epidemiological perspective, make the 100-year return level difficult to apply in practice. Could the methodology be applied in shorter time frames, maybe even allowing for some validation of the method?

Other minor comments:

- Several references are not relevant to the paper For example, 4-11 are cited as related to epidemics, and in fact they are about renewable energy, wind turbines and other topics. Similarly with 14, 15.
- Several citation styles are used in the paper (see for example the first paragraph of the introduction)

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

Partly

**Is the description of the method technically sound?**

Partly

**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?**

Yes

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

No

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Epidemiology, statistics

**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, however I have significant reservations, as outlined above.**

Reviewer Report 18 December 2023

<https://doi.org/10.5256/f1000research.158999.r226132>

© 2023 Babu M. 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.



**Md. Ashraful Babu**

<sup>1</sup> International University of Business Agriculture and Technology, Dhaka, Dhaka Division, Bangladesh

<sup>2</sup> Department of Physical Sciences, Independent University, Bangladesh, Dhaka, Dhaka Division, Bangladesh

The article presents a new method for long-term coronavirus forecasting using statistical extreme value theory and MATLAB optimization. The rationale for this method is clearly explained, addressing the limitations of existing models in varying epidemic scenarios. While the description of the method is technically sound, its reproducibility is partly limited due to insufficient detailed procedural and algorithmic information. Additionally, the full reproducibility of results is constrained as the underlying source data is not entirely available. The conclusions drawn about the method's performance are well-supported by the findings.

To enhance the scientific soundness and utility of the article, the authors should consider:

1. Providing more detailed procedural and algorithmic descriptions, enabling others to replicate the method more accurately.
2. Including or making accessible the complete source data used in their study, allowing for full validation and replication of results.

Addressing these points will significantly improve the paper's comprehensiveness and its applicability in different research contexts.

**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?**

Partly

**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:** Operations Research, Machine Learning, Deep Learning, Optimization, Time Series Forecasting

**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, however I have significant reservations, as outlined above.**

---

## Version 2

Reviewer Report 14 November 2023

<https://doi.org/10.5256/f1000research.144806.r174507>

© 2023 Khan A. 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.

**Adnan Khan**

Department of Mathematics, Lahore University of Management Sciences, Lahore, Pakistan

The authors have addressed many of the language and formatting concerns. There remain a few language issues in that the construct is rather cumbersome. For example

*"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."*

However, while not ideal in my view, the paper can be understood in its current form.

My concern regarding the method still remains, it seems that all work using this technique has involved one of the authors of the current work. This fact is of some concern to me.

**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?**

Yes

**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 Biology



**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, however I have significant reservations, as outlined above.**

Author Response 15 Nov 2023

**Oleg Gaidai**

*"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."*

**Answer:** thanks, passages corrected.

However, while not ideal in my view, the paper can be understood in its current form.

My concern regarding the method still remains, it seems that all work using this technique has involved one of the authors of the current work. This fact is of some concern to me.

**Answer:** thanks, now manuscript and method title modified to stress authorship of Oleg Gaidai (the leading author).

**Competing Interests:** No competing interests were disclosed.

---

### Version 1

Reviewer Report 21 February 2023

<https://doi.org/10.5256/f1000research.138279.r157163>

© 2023 Khan A. 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.



**Adnan Khan** 

Department of Mathematics, Lahore University of Management Sciences, Lahore, Pakistan

The authors want to propose a new methodology to make reliable long term forecast of COVID-19 infection rates, in particular they want to consider multidimensional data that may have cross correlations.

Comments:

The paper is not well written and very hard to read. There are grammatical errors as well as sentences which are very unclear to me. A few examples from the first section are given below:

- "The latter is usually due to many degrees of system freedom and random variables

governing dynamic biological systems, spread over extensive terrain." (Hard to understand)

- "For example, Gumbel used EVT to estimate the demographic of various populations in." (Incomplete sentence)
- "Moreover, a specific non-dimensional factor  $\lambda$  is introduced to predict the latter epidemic risk at any time and any place." (What exactly is the biological meaning of this parameter, this seems to be a critical component of the study and really needs a lot of detail when it is introduced.)

Many of the equations are not well explained e.g. Eq 2 third equality has a typo  $\eta_{1j}$ , similarly equation 5 needs to be explained further.

In the discussion section there is a mention of 'both countries', not clear which country besides China is being referred to.

These are just some examples of unclear text, hence my recommendation below.

Recommendation:

The aims and results are not well explained. My recommendation is that the authors may want to extensively edit the paper, making their goals, method and results clearer and the paper more readable.

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

No

**Is the description of the method technically sound?**

Partly

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

No

**If any results are presented, are all the source data underlying the results available to ensure full reproducibility?**

Yes

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

No

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Mathematical Biology

**I confirm that I have read this submission and believe that I have an appropriate level of**

**expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.**

Author Response 23 Feb 2023

**Oleg Gaidai**

### **Response to reviewer comments**

Note: since publication time of this methodology in 2022, a number of other applications of this method has been published, to mention few:

- 1) Gaidai, O., Cao, Y., Loginov, S., 2023, "Global cardiovascular diseases death rate prediction", Current Problems in Cardiology, <https://doi.org/10.1016/j.cpcardiol.2023.101622>
- 2) 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, <https://doi.org/10.1016/j.ijnaoe.2023.100515>
- 3) Gaidai, O., Xing, Y., Balakrishna, R., Xu, J., 2023, "Improving extreme offshore wind speed prediction by using deconvolution", Heliyon, <https://doi.org/10.1016/j.heliyon.2023.e13533>
- 4) Gaidai, O., Xing, Y., 2023, "Prediction of death rates for cardiovascular diseases and cancers", Cancer Innovation, <http://doi.org/10.1002/cai2.47>
- 5) Gaidai, O., Wang F., Yakimov, V., 2023, "COVID-19 multi-state epidemic forecast in India", Proceedings of the Indian National Science Academy, DOI: 10.1007/s43538-022-00147-5

***Making proposed method well validated***, reference list is updated now. Equations now restructured, so that all equations are collected only in one Section.

The paper is not well written and very hard to read. There are grammatical errors as well as sentences which are very unclear to me. A few examples from the first section are given below:

"The latter is usually due to many degrees of system freedom and random variables governing dynamic biological systems, spread over extensive terrain." (Hard to understand)

Answer: Changed to: "The latter is usually due to the fact that dynamic biological and environmental systems possess high number of degrees of freedom (in other words dimensional components), as well as system dependency on location – making bio-system of interest spatio-temporal."

"For example, Gumbel used EVT to estimate the demographic of various populations in." (Incomplete sentence)

Answer: Sentence removed.

"Moreover, a specific non-dimensional factor  $\lambda$  is introduced to predict the latter epidemic risk at any time and any place." (What exactly is the biological meaning of this parameter,

this seems to be a critical component of the study and really needs a lot of detail when it is introduced.)

Answer: Changed to: "introduced to **unify** various bio-system components having different failure limits". As seen later in the text  $\lambda$  being just a non-dimensional parameter, used to indicate system failure, namely when  $\lambda$  reaches 1.

Many of the equations are not well explained e.g. Eq 2 third equality has a typo ", similarly equation 5 needs to be explained further.

Answer: Equation corrected, now its Eq. (3).

In the discussion section there is a mention of 'both countries', not clear which country besides China is being referred to.

Answer: Thanks a lot for your valuable comments, changed now to a "chosen country of interest"

Recommendation:

The aims and results are not well explained. My recommendation is that the authors may want to extensively edit the paper, making their goals, method and results clearer and the paper more readable.

Answer: Text and equations modified. In a view of recent publication on this method authors now refer to already published papers: 1)-5) (see above).

**Competing Interests:** no Competing Interests

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](mailto:research@f1000.com)

**F1000Research**