Proceedings of the Indian National Science Academy (2023) 89:154–161
https://doi.org/10.1007/s43538-022-00147-5

RESEARCH PAPER

COVID-19 multi-state epidemic forecast in India

Oleg Gaidai[1] · Fang Wang[1] · Vladimir Yakimov[2]

Received: 25 April 2022 / Accepted: 17 December 2022 / Published online: 9 February 2023

Abstract

Clinical importance Novel coronavirus disease is spread worldwide with considerable morbidity and mortality and presents an enormous burden on worldwide public health. Due to the non-stationarity and complicated nature of novel coronavirus waves, it is challenging to model such a phenomenon. Few mathematical models can be used because novel coronavirus data are generally not normally distributed. This paper describes a novel 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 advantage of dealing efficiently with extensive regional dimensionality and cross-correlation between infection rate and mortality.

Objective To determine extreme novel coronavirus death rate probability at any time in any region of interest. Traditional statistical methods dealing with temporal observations of multi-regional processes do not have the advantage of dealing efficiently with extensive regional dimensionality and cross-correlation between different regional observations.

Design Apply modern novel statistical methods directly to raw clinical data.

Setting Multicenter, population-based, medical survey data based bio statistical approach.

Main outcome and measure Due to the non-stationarity and complicated nature of novel coronavirus, it is challenging to model such a phenomenon. Few mathematical models can be used because novel coronavirus data are generally not normally distributed. This paper describes 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.

Conclusions and relevance 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

Introduction

Statistical aspects of COVID-19 and other similar recent epidemics were receiving much attention in the modern research community, (https: prsindia.org, covid-19, cases; Acharya and Porwal 2020; Chen et al. 2015a; Thomas and Rootzen 2019; Lee and Wackernagel 2007a; Sudre et al. 2021; Chen et al. 2015b). 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, (World Health Organization 2014; Goldstein et al. 2011; Soebiyanto et al. 2010; Mugglin et al. 2002; Kim et al. 2013; Lee and Wackernagel 2007b; Falzarano et al. 2012; Su 2012; Xing et al. 2022; Madsen et al. 1986). This paper shares the same methodology as previously was introduced by authors in Gaidai et al. (2022c); Gaidai et al. (2022d).

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. For COVID-19, however, the only available observation numbers are limited

✉ Vladimir Yakimov
YakimovVV@cniimf.ru

[1] Shanghai Ocean University, Shanghai, China

[2] Central Marine Research and Design Institute, Saint Petersburg, Russia

by the beginning of the year 2020. Motivated by the latter argument, the authors have introduced a novel reliability method for biological and health systems to predict and manage epidemic outbreaks more accurately. His study was focused on COVID-19 epidemics in India (Singanayagam, et al. 2020; Maishman et al. 2021; Gareth, et al. 2021; Tom et al. 2021; Mahase 2022; Rutter et al. 2021), with focus on cross-correlations between different states within same climatic zone. For other studies related to statistical variations per country see e.g. (Gondauri et al. 2020). India was chosen of course because of its extensive COVID-19 health observations and related research available online (Gondauri et al. 2020; Zhu et al. 2020a; Wu et al. 2020; He et al. 2019; Wu and McGoogan 2020; Lu et al. 2020; Zhou et al. 2020; Zhu et al. 2020b; Organization 2020; Wood 1978; Bailey 1953; Becker and Britton 1999; Kermack and McKendrick 1927; Bailey 1954).

Statistical modelling of lifetime data or extreme value theory (EVT) is widespread in medicine or engineering. For example, Gumbel used EVT to estimate the demographic of various populations in https: prsindia.org, covid-19, cases. In (Chen et al. 2015a) authors used EVT to estimate the probability of an influenza outbreak in India. The author demonstrated a forecasting prediction potential amid the epidemic in this paper. While in Thomas and Rootzen (2019) similarly used EVT to predict and detect anomalies of influenza epidemics. As there is not much statistical research done to predict the probability of influenza or contagious diseases outbreak or its spread, the newly proposed novel method will be able better insight and an indication of the possible spread of diseases.

In this paper epidemic outbreak is viewed as unexpected incident that may occur at any state of a given country at any time, therefore spatial spread is accounted for. Moreover, 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 twenty-five India states from February 2020 until today were retrieved from the public website (https:prsindia.org, covid-19, cases.). As this valuable data set is per India state, the biological system under consideration can be regarded as a multi-degree of freedom (MDOF) dynamic system with highly inter-correlated regional components/dimensions. Some recent studies have already used statistical tools to predict COVID-19 development. 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 e.g. (Sudre et al. 2021), for mortality research see e.g. (J et al. 2020). Figure 1 presents map of India states.

Method

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

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

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

where $P = \iiint_{(0,0,0,\dots)}^{(\eta_X,\eta_Y,\eta_Z,\dots)} p_{X_T^{\max},Y_T^{\max},Z_T^{\max},\dots}(X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \dots) dX_T^{\max} dY_{N_Y}^{\max} dZ_{N_Z}^{\max} \dots$ is 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 data set limitations.

More specifically, the moment when either $X(t)$ exceeds η_X, or $Y(t)$ exceeds η_Y, or $Z(t)$ exceeds η_Z, and so on, the system is regarded as immediately failed. Fixed failure levels $\eta_X, \eta_Y, \eta_Z, \dots$ are, of course, individual for each unidimensional response component of $R(t)$. $X_{N_X}^{\max} = \max\{X_j; j = 1, \dots, N_X\} = X_T^{\max}$, $Y_{N_Y}^{\max} = \max\{Y_j; j = 1, \dots, N_Y\} = Y_T^{\max}$, $Z_{N_Z}^{\max} = \max\{Z_j; j = 1, \dots, N_Z\} = Z_T^{\max}$, and so on, (Xing et al. 2022; Gaidai et al. 2022a; Sun et al. 2022; Xu et al. 2022a; Gaidai et al. 2022b; Gaidai et al. 2022c; Gaidai et al. 2022d; Xu et al. 2022b; Gaidai et al. 2022e; Cheng et al. 2022; Gaidai et al. 2022f; Gaidai et al. 2020; Gaidai et al. 2022g; Gaidai et al. 2022h; Gaidai et al. 2022i; Balakrishna et al. 2022).

Now, the local maxima time instants $\left[t_1^X < \cdots < t_{N_X}^X; t_1^Y < \cdots < t_{N_Y}^Y; t_1^Z < \cdots < t_{N_Z}^Z\right]$ are sorted in monotonously non-decreasing order into one single merged time vector $t_1 \le \cdots \le t_N$. Note that $t_N = \max\{t_{N_X}^X, t_{N_Y}^Y, t_{N_Z}^Z, \dots\}$

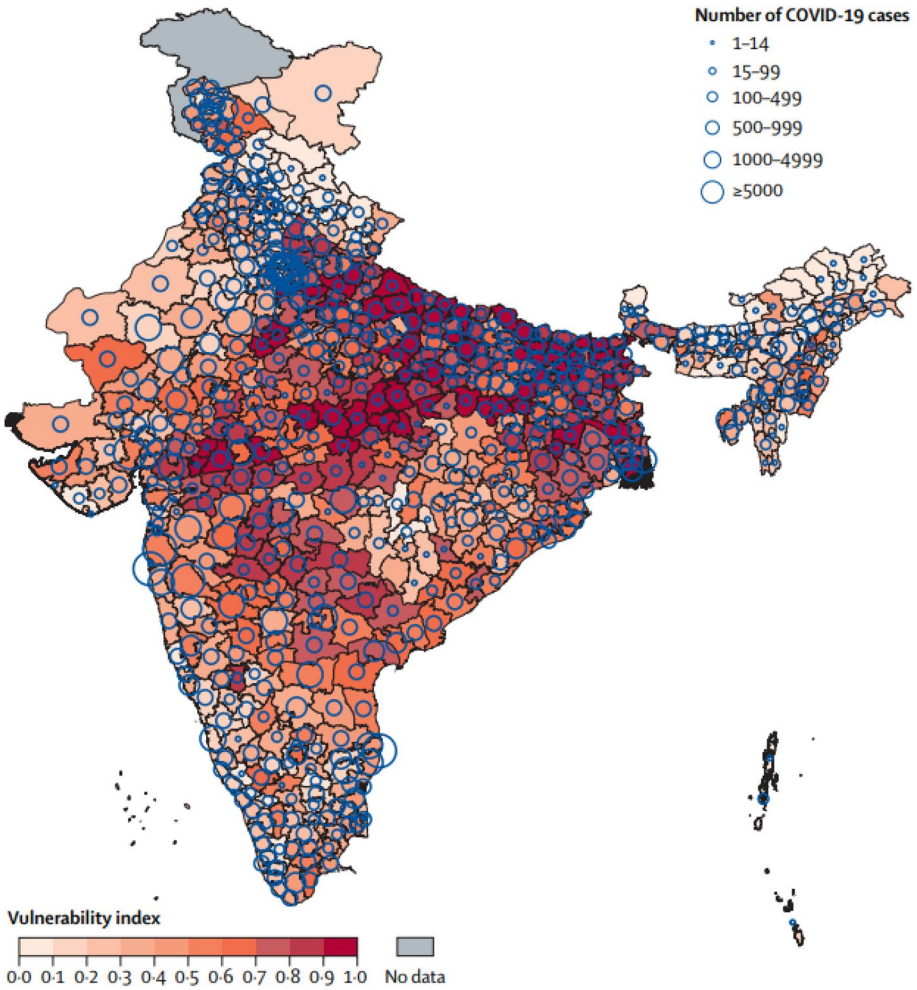

Fig. 1 Map of India with COVID-19 cases in affected states, (https: prsindia.org, covid-19, cases)

$,N = N_X + N_Y + N_Z + \ldots$. In this case t_j represents local maxima of one of MDOF structural response components either $X(t)$ or $Y(t)$, or $Z(t)$ and so on. That means that having $R(t)$ time record, one just needs continuously and simultaneously screen for unidimensional response component local maxima and record its exceedance of MDOF limit vector $(\eta_X, \eta_Y, \eta_Z, \ldots)$ in any of its components $X, Y, Z, \ldots$ Local unidimensional response component maxima are merged into one temporal non-decreasing vector $\vec{R} = (R_1, R_2, \ldots, R_N)$ following the merged time vector $t_1 \leq \cdots \leq t_N$. 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 $(\eta_1, \ldots, \eta_N)$ is introduced with each component η_j is either η_X, η_Y or η_Z and so on, depending on which of $X(t)$ or $Y(t)$, or $Z(t)$ etc., corresponds to the current local maxima with the running index j.

Now, scaling parameter $0 < \lambda \leq 1$ is introduced to artificially simultaneously decrease limit values for all response components, namely the new MDOF limit vector $(\eta_X^\lambda, \eta_Y^\lambda, \eta_z^\lambda, \ldots)$ with$\eta_X^\lambda \equiv \lambda \bullet \eta_X$,$\equiv \lambda \bullet \eta_Y$,$\eta_z^\lambda \equiv \lambda \bullet \eta_Z$, … is introduced, see (Naess and Moan 2013). The unified limit vector $(\eta_1^\lambda, \ldots, \eta_N^\lambda)$ is introduced with each component η_j^λ is eitherη_X^λ, η_Y^λ or η_z^λ and so on. The latter automatically defines probability $P(\lambda)$ as a function of λ, note that $P \equiv P(1)$ from Eq. (1). Non-exceedance probability $P(\lambda)$ can be estimated as follows

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

In practice, 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.

$$\begin{aligned} &\text{Prob}\{R_j \leq \eta_j^\lambda | R_{j-1} \leq \eta_{j-1}^\lambda, \ldots, R_1 \leq \eta_1^\lambda\} \\ &\quad \approx \text{Prob}\{R_j \leq \eta_j^\lambda | R_{j-1} \leq \eta_{j-1}^\lambda\} \end{aligned} \tag{3}$$

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

$$\begin{aligned}&\text{Prob}\{R_j \le \eta_j^\lambda | R_{j-1} \le \eta_{j-1}^\lambda, \ldots, R_1 \le \eta_1^\lambda\}\\ &\quad \approx \text{Prob}\{R_j \le \eta_j^\lambda | R_{j-1} \le \eta_{j-1}^\lambda, R_{j-2} \le \eta_{j-2}^\lambda\},\end{aligned} \tag{4}$$

where $3 \le j \le N$ (will be called conditioning level $k = 3$), and so on. The idea is to monitor each independent failure that happened locally first in time, thus avoiding cascading local inter-correlated exceedances. Equation (4) presents subsequent refinements of the statistical independence assumption. The latter approximations capture the statistical dependence effect between the neighboring maxima with increased accuracy. Since the original MDOF process $R(t)$ was assumed ergodic and therefore stationary, probability $p_k(\lambda) := \text{Prob}\{R_j > \eta_j^\lambda | R_{j-1} \le \eta_{j-1}^\lambda, R_{j-k+1} \le \eta_{j-k+1}^\lambda\}$ for $j \ge k$ will be independent of j but only dependent on conditioning level k. Thus non-exceedance probability can be approximated as in the average conditional exceedance rate method, see (Naess and Moan 2013)

$$P_k(\lambda) \approx \exp(-N \cdot p_k(\lambda)), k \ge 1. \tag{5}$$

Note that Eq. (5) follows from Eq. (1) by neglecting $\text{Prob}(R_1 \le \eta_1^\lambda) \approx 1$, as design failure probability must be minuscule, also assumed $N\epsilon k$. Equation (5) is similar to the well-known mean up-crossing rate equation for the probability of exceedance (Xing et al. 2022; Naess and Moan 2013). There is evident convergence with respect to the conditioning parameter k

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

Note that Eq. (5) for $k = 1$ turns into a well-known non-exceedance probability relationship with the mean up-crossing rate function.

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

where $\nu^+(\lambda)$ denotes the mean up-crossing rate of the response level λ for the above assembled non-dimensional vector $R(t)$ assembled from scaled MDOF system response $\left(\frac{X}{\eta_X}, \frac{Y}{\eta_Y}, \frac{Z}{\eta_Z}, \ldots\right)$. The mean up-crossing rate is given by the Rice's formula given in Eq. (7) with $p_{R\dot{R}}$ being joint probability density for $\left(R, \dot{R}\right)$ with $\dot{R}$ being time derivative $R'(t)$, see (Rice 1944). Equation (7) relies on the Poisson assumption that is up-crossing events of high λ levels (in this paper, it is $\lambda \ge 1$) 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 interdependency between extreme events, manifesting itself in the appearance of highly correlated local maxima clusters within the assembled vector $\vec{R} = \left(R_1, R_2, \ldots, R_N\right)$.

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

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

with $p_k(\lambda, m)$ being the same function as in Eq. (6) but corresponding to a specific short-term seasonal epidemic state with the number m. The above introduced $p_k(\lambda)$ as functions are often regular in the tail, specifically for values of λ approaching and exceeding 1. More precisely, for $\lambda \ge \lambda_0$, 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. One can then write

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

Next, by plotting $\ln\left\{\ln\left(p_k(\lambda)\right) - d_k\right\}$ versus $\ln\left(a_k\lambda + b_k\right)$, often nearly perfectly linear tail behaviour is typically observed. Optimal values of the parameters a_k, b_k, c_k, p_k, q_k may also be determined using a sequential quadratic programming (SQP) method incorporated in the NAG Numerical Library (Numerical Algorithms Group 2010).

Results

Prediction of influenza-like epidemics has long been the focus of attention in epidemiology and mathematical biology. It is well known that public health dynamics is a highly non-linear multidimensional and spatially cross-correlated dynamic system that is always challenging to analyse. Previous studies have used a variety of approaches to model influenza-like cases. This section illustrates the efficiency of the above-described methodology using the new method applied to the real-life COVID-19 data sets, presented as a new daily recorded infected patient time series, spread over large terrains.

Influenza and COVID-19 are contagious diseases having high transmissibility with certain mortality. They typically occur seasonally in late autumn, winter, early spring, reaching its peak in winter. Seasonal influenza epidemics are typically caused by influenza A, B viruses typically occur annually during winter time presenting certain burden on national public health, resulting in about 3–5 million cases of severe illness, along with 250,000–500,000 deaths worldwide annually, according to World Health Organization (WHO) (World Health Organization 2014).

This section presents a real-life application of the above-described method. The statistical data in the present section are taken from the official India website (https: prsindia.

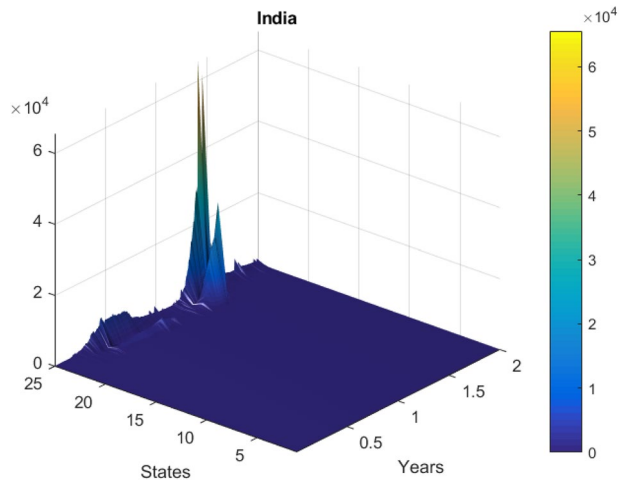

Fig. 2 Daily recorded patients numbers per country and per day

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

Figure 3 presents the number of new daily recorded patients as a 25D vector $\vec{R}$, consisting of assembled regional new daily patient numbers. Note that vector $\vec{R}$ does not have physical meaning on its own, as it is assembled of different regional components with different epidemic backgrounds. Index j is just a running index of local maxima encountered in a non-decreasing time sequence.

Figure 4 presents 100 years return level extrapolation according to Eq. (9) towards epidemic outbreak with 100 year return period, indicated by the horizontal dotted line, and somewhat beyond, $\lambda = 0.1$ cut-on value was used. Critical 100 years return level is indicated by star in Fig. 4.

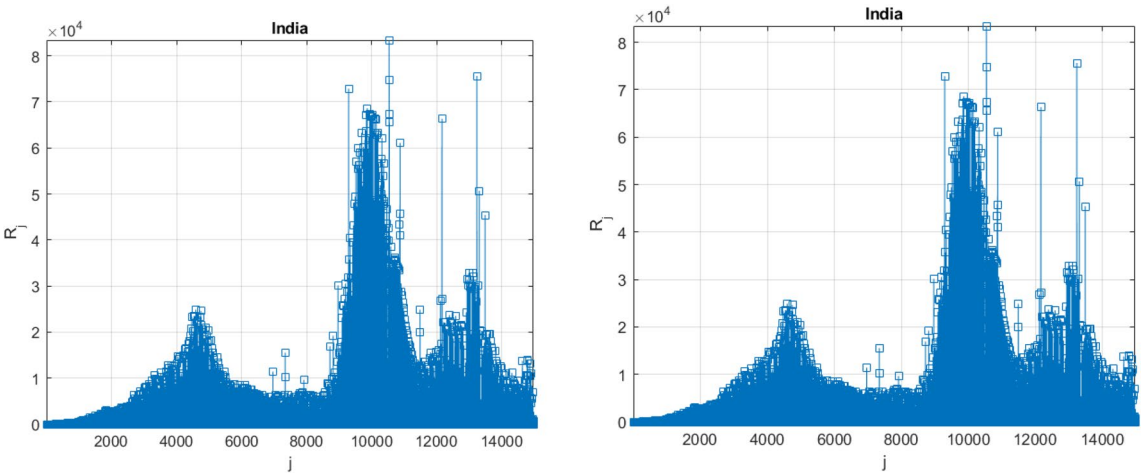

Fig. 3 Number of new daily recorded patients as 25D vector $\vec{R}$. Left: as it is, right: scaled by Eq. (10)

org, covid-19, cases). The website provides the number of newly diagnosed cases every day in India from 22 January 2020 to 6 April 2022. Patient numbers from twenty-five different India states were chosen as components $X, Y, Z, \dots$. thus constituting an example of a twenty-five dimensional (25D) dynamic biological system, see Fig. 2. In order to unify all 25 measured time series $X, Y, Z, \dots$. the following scaling was performed

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

making all three responses non-dimensional and having the same failure limit. Failure limits, or in other words, epidemic thresholds, were chosen differently for different states in this paper $\eta_X, \eta_Y, \eta_Z, \dots$ were set equal to observed two years maxima, twice increased. Next, all local maxima

Note that predictions for shorter (and more realistic) return periods (e.g. few years) can also be easily extracted from Fig. 4, with the only limitation of underlying assumption of bio-system quazi-stationarity.

Dotted lines indicate extrapolated 95% confidence interval according to Eq. (10). According to Eq. (5) $p(\lambda)$ is directly related to the target failure probability $1 - P$ from Eq. (1). Therefore, in agreement with Eq. (5), system failure probability $1 - P \approx 1 - P_k(1)$ can be estimated. Note that in Eq. (5), N corresponds to the total number of local maxima in the unified response vector $\vec{R}$. Conditioning parameter $k = 5$ was found to be sufficient due to occurrence of convergence with respect to k, see Eq. (6). Figure 4 exhibits reasonably narrow 95% CI. The latter is an advantage of the proposed method.

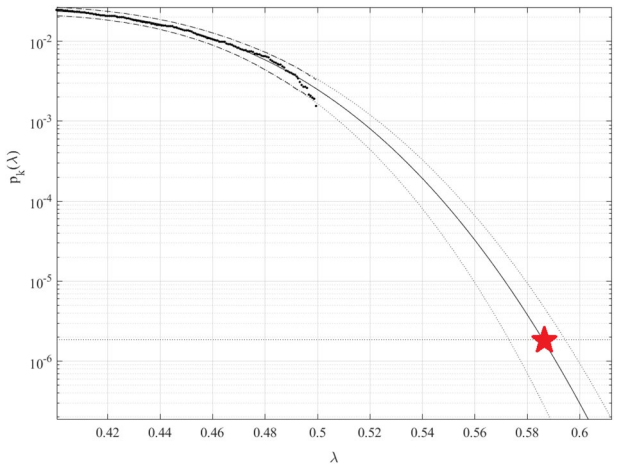

Fig. 4 100 years return level (horizontal dotted line) extrapolation of $p_k(\lambda)$ towards critical 100 years return level (indicated by star) and beyond. Extrapolated 95% CI indicated by dotted lines

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 Fig. 4, represents probability of epidemic outbreak at any India state in the years to come.

Conclusions

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

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.

This paper studied recorded COVID-19 patient numbers from twenty-five different most COVID affected India states, constituting an example of a twenty-five dimensional (25D) observed in 2020–2022. The novel reliability method was applied to new daily patient numbers as a multidimensional system in real-time. The theoretical reasoning behind the proposed method is given in detail. Note that the use of direct either measurement or Monte Carlo simulation for dynamic biological system reliability analysis is attractive; however, dynamic system complexity and its high dimensionality require the development of novel robust and accurate techniques that can deal with a limited data set at hand, utilizing available data as efficient as possible.

The main conclusion is that if the public health system under local environmental and epidemiologic conditions in India is well managed. Predicted 100 year return period risk level λ of epidemic outbreak is very low.

Various authors with different approaches have shown the usage of statistics through EVT and other models in medicine. One such method used the block maxima (BM) approach, while another used the Peak Over Threshold (POT) approach to estimate the distribution of extremes. Even though both these studies showed their suitability for estimating the extreme values, each of them had its limitations, with one of them requiring a large amount of data.

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

Acknowledgements This paper shares the same methodology as previously was introduced by authors in preprint [21], however the case studied was different.

Author contributions All authors contributed equally.

Funding No funding received.

Data availability See https://prsindia.org/covid-19/cases, (https: prsindia.org, covid-19, cases) for the data used.

Declarations

Conflict of interest Authors declare no conflict of interest.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Acharya, R., Porwal, A.: A vulnerability index for the management of and response to the COVID-19 epidemic in India: an ecological study. Lancet Glob. Health (2020). https://doi.org/10.1016/S2214-109X(20)30300-4

Bailey, N.T.J.: The total size of a general stochastic epidemic. Biometrika **40**, 177 (1953). https://doi.org/10.1093/biomet/40.1-2.177

Bailey, N.T.J.: Maximum-likelihood estimation of the relative removal rate from the distribution of the total size of an intra household epidemic. J. Hyg (lond). **52**(3), 400–402 (1954). https://doi.org/10.1017/s0022172400027595

Balakrishna, R., Gaidai, O., Wang, F., Xing, Y., Wang, S.: A novel design approach for estimation of extreme load responses of a 10-MW floating semi-submersible type wind turbine. Ocean Eng. (2022). https://doi.org/10.1016/j.oceaneng.2022.112007

Becker, N.G., Britton, T.: Statistical studies of infectious disease incidence. J. R. Statist. Soc. B. **61**(2), 287–307 (1999). https://doi.org/10.1111/1467-9868.00177

Chen, J., Lei, X., Zhang, L., Peng, B.: Using extreme value theory approaches to forecast the probability of outbreak of highly pathogenic influenza in Zhejiang China. Plos One **10**(2), e0118521 (2015)

Chen, J., Lei, X., Zhang, L., Peng, B.: Using extreme value theory approaches to forecast the probability of outbreak of highly pathogenic influenza in Zhejiang China. PLoS ONE **10**(2), e0118521 (2015). https://doi.org/10.1371/journal.pone.0118521

Cheng, Y., Gaidai, O., Yurchenko, D., Xu, X., Gao, S.: "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)

Falzarano, J., Su, Z., Jamnongpipatkul, A.: "Application of stochastic dynamical system to non-linear ship rolling problems", Proceedings of the 11th International Conference on the Stability of Ships and Ocean Vehicles, Athens, Greece. (2012)

Gaidai, O., Xu, X., Naess, A., Cheng, Y., Ye, R., Wang, J.: Bivariate statistics of wind farm support vessel motions while docking. Sh. Offshore Struct. **16**(2), 135–143 (2020)

Gaidai, O., Wang, F., Wu, Y., Xing, Y., Medina, A., Wang, J.: Offshore renewable energy site correlated wind-wave statistics. Probabilistic Eng. Mech. (2022). https://doi.org/10.1016/j.probengmech.2022a.103207

Gaidai, O., Xing, Y., Wang, F., Wang, S., Yan, P., Naess, A.: 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). https://doi.org/10.3389/fmech.2022b.888497

Gaidai, O., Xing, Y., Xu, X.: COVID-19 epidemic forecast in USA East coast by novel reliability approach. Res. square (2022). https://doi.org/10.21203/rs.3.rs-1573862/v1

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 Eng. (2022e). https://doi.org/10.1016/j.rineng.2022.100593

Gaidai, O., Fu, S., Xing, Y.: Novel reliability method for multidimensional nonlinear dynamic systems. Mar. Struct. (2022). https://doi.org/10.1016/j.marstruc.2022g.103278

Gaidai, O., Yan, P., Xing, Y.: A novel method for prediction of extreme wind speeds across parts of Southern Norway. Front. Environ. Sci. (2022). https://doi.org/10.3389/fenvs.2022.997216

Gaidai, O., Yan, P., Xing, Y.: Prediction of extreme cargo ship panel stresses by using deconvolution. Front. Mech. Eng. **86**, 34–45 (2022). https://doi.org/10.3389/fmech.2022.992177

Gaidai, O., Yan, P., Xing, Y., Xu, J., Wu, Y.: A novel statistical method for long-term coronavirus modelling. F1000 Res. **11**, 1282 (2022)

Gaidai, O., Storhaug, G., Wang, F., Yan, P., Naess, A., Wu, Y., Xing, Y., Sun, J.: "On-Board Trend Analysis for Cargo Vessel Hull Monitoring Systems", The 32nd International Ocean and Polar Engineering Conference, Paper number: ISOPE-I-22–541. (2022f)

Gareth, Maze Ss D, Benskin L.: Serious Statistical Flaws in Hastie, et al. Vitamin D concentrations and COVID-19 infection in UK Biobank Analysis. (2021)

Goldstein, E., Cobey, S., Takahashi, S., Miller, J.C., Lipsitch, M.: Predicting the epidemic sizes of influenza A/H1N1, A/H3N2, and B: a statistical method. PLoS Med. **8**, e1001051 (2011). https://doi.org/10.1371/journal.pmed.1001051

Gondauri, 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. **17**(4), em209 (2020). https://doi.org/10.29333/ejgm/7869

He, F., Deng, Y., Li, W.: Coronavirus disease 2019 (COVID-19): what we know? J. Med. Virol. (2020). https://doi.org/10.1002/jmv.25766

https://prsindia.org/covid-19/cases

J, et al.: Factors associated with COVID-19-related death using OpenSAFELY. Nature (2020). https://doi.org/10.1038/s41586-020-2521-4

Kermack, W.O., McKendrick, A.G.: A contribution to the mathematical theory of epidemics. Proc. Royal Soc. London Ser. A Contain. Pap. Math. Phys. Charact. **115**(772), 700–21 (1927). https://doi.org/10.1098/rspa.1927.0118

Kim, E.K., Seok, J.H., Oh, J.S., Lee, H.W., Kim, K.H.: Use of Hangeul twitter to track and predict human influenza infection. PLoS One **7**, e69305 (2013). https://doi.org/10.1371/journal.pone.0069305

Lee H. C, Wackernagel H.: Extreme values analyses of US P&I mortality data under consideration of demographic effects. Centre de géosciences. Ecole des mines de Paris, Fontainebleau, France, R071113HLEE. (2007a)

Lee HC, Wackernagel H.: Extreme value analyses of US P&I mortality data under consideration of demographic effects. [cited 10 June 2014]. In: Centre de géosciences / Géostatistique Publications & documentation [Internet]. Fontainebleau: Mine Parisrtech. http://cg.ensmp.fr/bibliotheque/public/LEE_Rapport_00600.pdf. doi: https://doi.org/10.1021/jp204223t(2007b)

Lu, R., Zhao, X., Li, J., et al.: Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet **395**(10224), 565–574 (2020). https://doi.org/10.1016/S0140-6736(20)30251-8

Madsen, H.O., Krenk, S., Lind, N.C.: Methods of structural safety. Prentice-Hall Inc, Englewood Cliffs (1986)

Mahase, E.: Covid-19: is the UK heading for another omicron wave? BMJ (2022). https://doi.org/10.1136/bmj.o738

Maishman T, Schaap S, Silk D S, et al.: Statistical methods used to combine the effective reproduction number, R(t), and other related measures of COVID-19 in the UK[J]. (2021) https://doi.org/10.48550/arXiv.2103.01742

Mugglin, A.S., Cressie, N., Gemmell, I.: Hierarchical statistical modelling of influenza epidemic dynamics in space and time. Stat. Med. **21**(18), 2703–2721 (2002)

Naess, A., Moan, T.: Stochastic dynamics of marine structures. Cambridge University Press (2013)

Numerical Algorithms Group: NAG Toolbox for Matlab. India: NAG Ltd, Oxford (2010)

Organization WH.: Coronavirus disease 2019 (COVID-19) Situation Report - 70. (2020)

Rice, S.O.: Mathematical analysis of random noise. Bell Syst. Tech. J. **23**, 282–332 (1944)

Rutter, M., Lanyon, P.C., Grainge, M.J., et al.: COVID-19 infection, admission and death among people with rare autoimmune rheumatic disease in India: results from the RECORDER project [J]. Rheumatology (2021). https://doi.org/10.1093/rheumatology/keab794

Singanayagam, Anika, et al.: Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19,

India. Eurosurveillance. **25**(32), 2001483 (2020). https://doi.org/10.2807/1560-7917.ES.2020.25.32.2001483

Soebiyanto, R.P., Adimi, F., Kiang, R.K.: Modeling and predicting seasonal influenza transmission in warm states using climatological parameters. PLoS One. **5**, e9450 (2010). https://doi.org/10.1371/journal.pone.0009450

Su, Z.: Non-linear response and stability analysis of vessel rolling motion in random waves using stochastic dynamical systems. Texas University, Texas (2012)

Sudre, C., et al.: Attributes and predictors of long COVID. Nat. Med. **27**, 626–631 (2021). https://doi.org/10.1038/s41591-021-01292-y

Sun, J., Gaidai, O., Wang, F., Naess, A., Wu, Y., Xing, Y., van Loon, E., Medina, A., Wang, J.: Extreme riser experimental loads caused by sea currents in the Gulf of Eilat. Probabilistic Eng. Mech. (2022). https://doi.org/10.1016/j.probengmech.2022.103243

Thomas, M., & Rootzen, H.: Real-time prediction of severe influenza epidemics using Extreme Value Statistics. arXiv preprint arXiv:1910.10788. (2019)

Tom, K., Paul, C., Alina, A., et al.: Exploring the impact of the first wave of covid-19 on social work practice: a qualitative study in India, UK [J]. Br. J. Soc. Work (2021). https://doi.org/10.1093/bjsw/bcab166

Wood, P.H.N.: The mathematical theory of infectious diseases and its applications. Immunology **34**(5), 955–956 (1978)

World Health Organization. Influenza fact sheet. 2014 Mar [cited 10 June 2014]. In: World Health Organization website [Internet]. Geneva: World Health Organization 1948–[about 2 screens]. http://www.who.int/mediacentre/factsheets/fs211/en/index.html.

Wu, J.T., Leung, K., Leung, G.M.: Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet (2020). https://doi.org/10.1016/S0140-6736(20)30260-93

Wu, Z., McGoogan, J.M.: 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 **2019**, 3–6 (2020). https://doi.org/10.1001/jama.2020.2648

Xing, Y., Gaidai, O., Ma, Y., Naess, A., Wang, F.: 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). https://doi.org/10.1016/j.apor.2022.10317

Xu, X., Wang, F., Gaidai, O., Naess, A., Xing, Y., Wang, J.: Bivariate statistics of floating offshore wind turbine dynamic response under operational conditions. Ocean Eng. (2022). https://doi.org/10.1016/j.oceaneng.2022a.111657

Xu, X., Xing, Y., Gaidai, O., Wang, K., Patel, K., Dou, P., Zhang, Z.: A novel multi-dimensional reliability approach for floating wind turbines under power production conditions. Front. Mar. Sci. (2022). https://doi.org/10.3389/fmars.2022.970081

Zhou, P., Yang, X.L., Wang, X.G., et al.: A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature (2020). https://doi.org/10.1038/s41586-020-2012-7

Zhu, N., Zhang, D., Wang, W., et al.: A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. (2020a). https://doi.org/10.1056/nejmoa2001017

Zhu, N., Zhang, D., Wang, W., et al.: A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. **382**(8), 727–733 (2020b). https://doi.org/10.1056/NEJMoa2001017