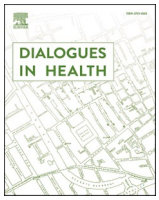
Dialogues in Health 3 (2023) 100157

Contents lists available at ScienceDirect

Dialogues in Health

journal homepage: www.elsevier.com/locate/dialog



Influenza-type epidemic risks by spatio-temporal Gaidai-Yakimov method

Oleg Gaidai [a], Vladimir Yakimov [b,*], Eric-Jan van Loon [c]

[a] Shanghai Ocean University, Shanghai, China
[b] Central Marine Research and Design Institute, Saint Petersburg, Russia
[c] Ghent University, Ghent, Belgium

ARTICLE INFO

Keywords:
COVID-19
Public health
Risk forecast
AI (artificial intelligence)
SARS-CoV-2
Epidemic outbreak

ABSTRACT

Background: Global public health was recently hampered by reported widespread spread of new coronavirus illness, although morbidity and fatality rates were low. Future coronavirus infection rates may be accurately predicted over a long-time horizon, using novel bio-reliability approach, being especially well suitable for environmental multi-regional health and biological systems. The high regional dimensionality along with cross-correlations between various regional datasets being challenging for conventional statistical tools to manage.
Methods: To assess future risks of epidemiological outbreak in any province of interest, novel spatio-temporal technique has been proposed. In a multicenter, population-based environment, assess raw clinical data using state-of-the-art, cutting-edge statistical methodologies.
Results: Authors have developed novel reliable long-term risk assessment methodology for future coronavirus infection outbreaks.
Conclusions: Based on national clinical patient monitoring raw dataset, it is concluded that although underlying data set data quality is questionable, the proposed method may be still applied.

1. Introduction

COVID-19 statistics and public health implications (SARS-CoV-2) in a view of recent outbreaks have been the focus of contemporary research, [1–10]. In general, traditional theoretical statistical techniques may find it challenging to forecast breakout probability and realistic bio-system reliability factors in actual epidemic settings, [11–28]. The latter is typically brought on by variety of bio-system and environmental degrees of freedom along with random factors that control dynamic biological and health systems, dispersed across a wide geographical area. Direct Monte Carlo (MC) simulations or observations may be used to assess risks within a complex biological system. But COVID-19 only offers data from observations for the first day of 2020. In order to predict and potentially prevent epidemic outbreaks, authors have developed specific reliability approach suitable for engineering, bio and health systems, [29–37]. Various researchers have investigated recent COVID-19 outbreaks in China, [38–45], focusing on connections between several sites within same climatic zones; further studies on statistical national variations may be found in [46]. China was chosen based on its COVID-19 origin, as well as its extensive public health observations available online, [47–50]. This study utilized statistics, obtained from the official website and database, maintained by PRC (People Republic of China) National Health Commission, [2]. PRC embraces twenty-three provinces, five autonomous regions, four municipalities, two special administrative regions, included Taiwan and excluding Tibet (as a region almost not affected by COVID-19), there is therefore thirty-four administrative units to study.

EVT (Extreme Value Theory) methodology being frequently used within both bio-engineering and medical research, [1]. Ref. [4] EVT was used in study by the authors to estimate future influenza outbreaks risks in China. While in [25] EVT has also been used to identify anomalies both before and after the flu epidemic. Numerous statistical research studies have been done to assess risks of an influenza epidemic or other infectious disease breakouts, hence recently proposed novel methodology intended to provide yet better epidemiological understanding, along with improved potential illnesses spread indicators, [51–56]. Since an epidemic breakout is treated as an unforeseen occurrence that might occur at any time in any area within a particular country, geographic dispersion is taken into account in this study. The latter pandemic hazard may be predicted at any time and everywhere because to a special non-dimensional component that is also present. Biosystems can occasionally be impacted by environmental factors.

* Corresponding author.
E-mail address: YakimovVV@cniimf.ru (V. Yakimov).

https://doi.org/10.1016/j.dialog.2023.100157
Received 20 August 2023; Received in revised form 24 September 2023; Accepted 24 October 2023
Available online 27 October 2023

Statistical methods have been applied in a number of recent researches, notably for the linear log model, to forecast the progression of COVID-19, [1,34]. Despite our stated research goals of predicting epidemic breakouts and lowering their risk (by early diagnostics and prognostics), the research only considers the daily number of patients recorded and ignores the symptoms themselves. Earlier studies have employed variety of techniques, to model influenza-like diseases. The success of the methodologies outlined above is demonstrated in this part using a novel approach to the actual unfiltered COVID-19 dataset, given as daily-recorded infected number of patients time-series, geographically/spatially scattered. Both COVID-19 (SARS-CoV-2) and influenza being contagious types of diseases, with a tendency to spread internationally, having low rates of morbidity and mortality. They are most common throughout the year's late fall, winter, and early spring months, with the winter being their peak season. According to estimates from WHO (World Health Organization), epidemics of seasonal influenzas, caused by influenza A, B viruses, resulted in between 3 and 5 million critical illness cases, and between 250,000 and 500,000 fatalities globally each year, placing a strain on global public health, [28,63–65].

Many research studies, such as [36–39], have compiled substantial current information on traditional reliability methods. Applying these conventional methods to an actual bio-engineering or public health system, however, may be challenging since bio-dynamic components usually depend on numerous degrees of freedom, and environmental random variables. Theoretically, it is well possible to precisely assess reliability of a real and complex bio-engineering systems, if underlying dataset being large enough. Such biological systems can be simulated using MC methods, or be represented by actual extensive measurements. It is often challenging to assess bio and public health system reliability, utilizing conventional public health and bio-reliability methods. The latter is frequently caused by chaotic forces that control bio-systems and very flexible systems. Complex bio-system reliability may be well assessed directly, either by doing direct extensive MC simulations, or if having substantial measured clinical dataset, [11–24]. Public and digital health becoming more and more attractive subject for spatio-temporal data analysis in the globalizing world. Numerous researches have been done recently to investigate spatio-temporal analytic techniques, and apply them to a variety of sectors. The spatio-temporal analysis landscape may be examined and projected from 3 angles: the spatio-temporal analysis technique, the development of the spatio-temporal data model and platform, and the application scenarios for the spatio-temporal data analysis. For contemporary literature review on spatio-temporal data analysis methods, see [66].

Fig. 1. Map of China with provinces, including Taiwan.

2. Spatio-temporal Gaidai-Yakimov method

This study has utilized COVID-19 incidence data for all of China's administrative provinces from February 2020 till the end of year 2022, available at public websites, [2,3]. Biological or public health system under consideration may be seen as MDOF (Multi-Degree-Of-Freedom) spatio-temporal nonlinear dynamic system, having inter-correlated geographical/administrative dimensions/components. Fig. 1 presents map of China provinces, national territorial mapping is an important feature of the suggested methodology.

Let's take a look at a MDOF biological or health system, represented by biosystem's key/critical components (in this study those are daily patient numbers per administrative units) $(X(t), Y(t), Z(t), \ldots)$ that have been either simulated, or measured over representative observational duration $(0, T)$. One-dimensional (1D) bio-system key component's global maxima being denoted as $X_T^{\max} = \max_{0\le t\le T} X(t)$, $Y_T^{\max} = \max_{0\le t\le T} Y(t)$, $Z_T^{\max} = \max_{0\le t\le T} Z(t), \ldots$. Let $X_1, \ldots, X_{N_X}$ be temporally non-decreasing bio-system component's process $X = X(t)$ local-maxima, recorded at temporally increasing time-moments $t_1^X < \ldots < t_{N_X}^X$ within $(0, T)$. Similar definitions will follow for other MDOF bio-system's components $Y(t), Z(t), \ldots$ i.e., $Y_1, \ldots, Y_{N_Y}$; $Z_1, \ldots, Z_{N_Z}$ etc. Biological or public health system component's local-maxima were assumed here being positive just for simplicity, then

$$P = \iiint_{(0,0,0,\ldots)}^{(\eta_X,\eta_Y,\eta_Z,\ldots)} p_{X_T^{\max},Y_T^{\max},Z_T^{\max},\ldots}\left(x_T^{\max}, y_T^{\max}, z_T^{\max}, \ldots\right) dx_T^{\max} dy_T^{\max} dz_T^{\max} \ldots \quad (1)$$

being target dynamic biological or public health survival probability, with limit/critical/hazard values of bio-system's critical components, being denoted here as η_X, η_Y, η_Z, …; and $\cup$ being logical unity operator «or»; $p_{X_T^{\max},Y_T^{\max},Z_T^{\max},\ldots}$ being joint PDF (Probability Density Function) of bio-system component's global maxima. If bio-system's NDOF (Number-of-Degrees-Of-Freedom) being large, it is not always practically feasible to directly estimate target joint PDF $p_{X_T^{\max},Y_T^{\max},Z_T^{\max},\ldots}$ and hence corresponding bio-system survival probability P. The latter biological or public health system survival probability P needs to be estimated, along with related system's expected lifetime, following Eq. (1). Biological or public health system's 1D components $X, Y, Z, \ldots$being re-scaled and then non-dimensionalized

$$X \to \frac{X}{\lambda \eta_X}, Y \to \frac{Y}{\lambda \eta_Y}, Z \to \frac{X}{\lambda \eta_X}, \ldots \quad (2)$$

resulting in all biological or public health system's components becoming non-dimensional, having identical target failure/hazard limits $\lambda = 1$, with bio-system's target hazard/ failure probability $1 - P = 1 - P(1)$. Eq. (2) may be employed now to determine $P(\lambda)$ as a function of non-dimensional bio-system scaling parameter λ. 1D bio-system component's local-maxima now have been merged to obtain single synthetic temporally non-decreasing system vector $\vec{R} = (R_1, R_2, \ldots, R_N)$ in agreement with merged temporal system vector $t_1 \le \ldots \le t_N$, $N \le N_X + N_Y + N_Z + \ldots$. Each element R_j being system component's local-maxima, once encountered, corresponding to either $X(t)$ or $Y(t)$, or $Z(t)$ or other bio-system's key components. Constructed synthetic bio-system $\vec{R}$-vector yields 0 data-loss, as illustrated by Fig. 2.

Hence, temporally increasing bio-system synthetic 1D vector $\vec{R}$, along with corresponding temporally increasing occurrence moments $t_1 < \ldots < t_N$, have been fully introduced. Scaling system parameter $0 < \lambda \le 1$ has been introduced, in order to artificially, simultaneously decrease limit/hazard values for all relevant bio-system components, namely MDOF system hazard/limit vector $(\eta_X^\lambda, \eta_Y^\lambda, \eta_z^\lambda, \ldots)$ having $\eta_X^\lambda \equiv \lambda \bullet \eta_X$, $\equiv \lambda \bullet \eta_Y$, $\eta_z^\lambda \equiv \lambda \bullet \eta_Z$, … being introduced. Unified system limit vector $(\eta_1^\lambda, \ldots, \eta_N^\lambda)$ being introduced with each biosystem's component η_j^λ being either η_X^λ, η_Y^λ or η_z^λ etc. The latter naturally defines survival probability $P(\lambda)$ as a smooth function of λ, with $P \equiv P(1)$, see Eq.

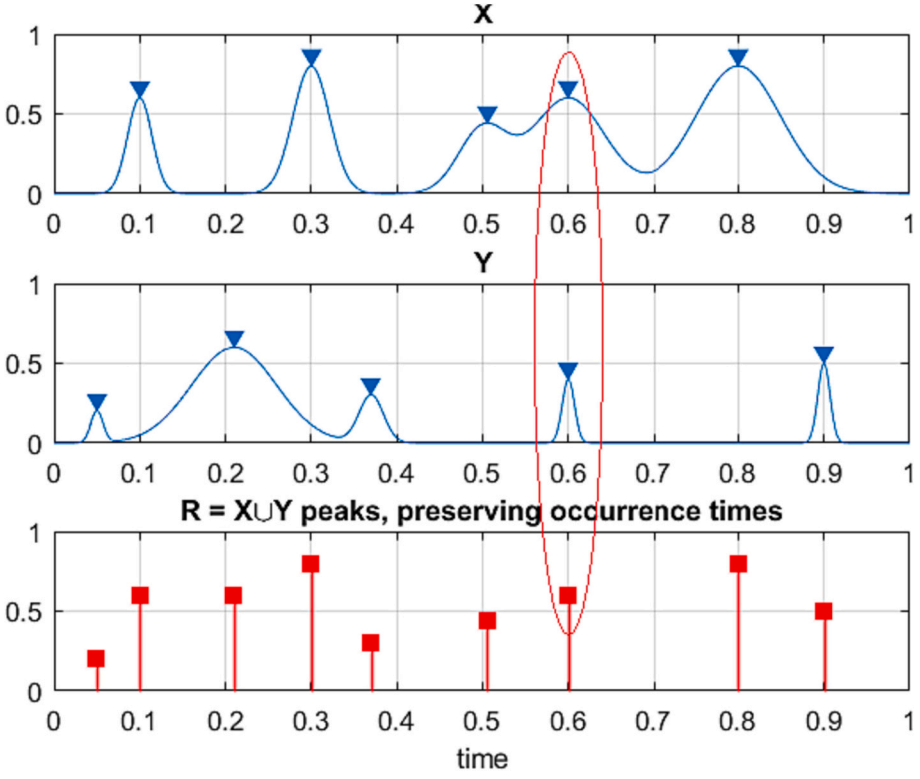


Fig. 2. Example of 2 bio-system components X, Y, merged into 1D synthetic vector $\vec{R}$, ellipse marks simultaneous occurrence of local-maxima in 2 different system components.

(1). Non-exceedance (or survival) probability $P(\lambda)$ may be now assessed as

$$P(\lambda) = \text{Prob}\{R_N \le \eta_N^\lambda, \ldots, R_1 \le \eta_1^\lambda\}$$

$$= \text{Prob}\{R_N \le \eta_N^\lambda \mid R_{N-1} \le \eta_{N-1}^\lambda, \ldots, R_1 \le \eta_1^\lambda\} \cdot \text{Prob}\{R_{N-1} \le \eta_{N-1}^\lambda, \ldots, R_1 \le \eta_1^\lambda\}$$

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

Dependency between the neighboring local-maxima R_j is often non-negligible; hence following 1-step (conditioning-number $k = 1$) memory approximation being introduced

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

for $2 \le j \le N$ (conditioning-number $k = 2$). Approximation, set by Eq. (3) may be further developed as

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

where $3 \le j \le N$ (with conditioning-number $k = 3$), etc. Idea now is to monitor every independent bio-system's hazard/failure, in temporally increasing order, hence avoiding cascading component's inter-correlated local exceedances. Eq. (4) presents statistical independence assumption's subsequent refinements. The latter approximation captures statistical dependency effects between neighboring bio-system component's local-maxima, with steadily increasing accuracy. Since the original MDOF system process $\boldsymbol{R}(t)$ has been assumed to be ergodic, and hence stationary, the probability $p_k(\lambda) := \text{Prob}\left\{R_j > \eta_j^\lambda \;\middle|\; R_{j-1} \le \eta_{j-1}^\lambda, R_{j-k+1} \le \eta_{j-k+1}^\lambda\right\}$ for $j \ge k$ being independent of j and only being dependent on conditioning-number k. Hence, non-exceedance (survival) probability may be now approximated, as in modified (4-parameter) Weibull method, [58]

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

In Eq. (6) exponent $N \bullet p_k(\lambda) \ll 1$, as design failure/hazard probability is of a small order of magnitude by design; it is also has been assumed that $N \gg k$. Eq. (5) being similar to a well-known mean up-crossing-rate formula, expressing exceedance probability. There is typical convergence, with respect to conditioning-number k

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

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

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

with $\nu^+(\lambda)$ being mean up-crossing-rate function of the non-dimensional system level λ for the above constructed non-dimensional bio-system vector $R(t)$ constructed from the scaled MDOF bio-system vector $\left(\frac{X}{\eta_X}, \frac{Y}{\eta_Y}, \frac{Z}{\eta_Z}, \ldots\right)$. The mean up-crossing-rate function being given by the Rice formula, see Eq. (8), with $p_{R\dot{R}}$ being system's joint PDF for $\left(R, \dot{R}\right)$ with $\dot{R}$ being temporal derivative $\dot{R}(t)$, see [35]. Eq. (8) relied on well-known Poisson's assumption, stating that high λ levels up-crossing events (in the current study, it is $\lambda \ge 1$) may be assumed nearly independent. The latter may not always be the case e.g., for narrow-band systems, exhibiting cascading/clustering failures/hazards in different system components, temporally subsequent, caused by intrinsic inter-dependencies between critical/extreme/hazard events, manifesting themself through appearance of a highly-correlated biosystem key component's local-maxima groups/clusters, present within constructed bio-system vector $\vec{R} = (R_1, R_2, \ldots, R_N)$. In this section system stationarity/ergodicity assumption was used, but advocated methodology may also treat reasonably nonstationary cases. For non-stationary bio-systems, in-situ scatter diagram of $m = 1, .., M$ bio-epidemiological/environmental seasonal conditions, with each short-term bio-environmental state having individual probability q_m, so that $\sum_{m=1}^{M} q_m = 1$. Let one introduce the long-term statistical equation

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

with $p_k(\lambda, m)$ being identical function, following Eq. (6), but corresponding to specific in-situ short-term bio-environmental epidemic state with the number m. The above-introduced functions $p_k(\lambda)$ being typically regular in PDF tail, specifically for values of λ when approaching, and exceeding 1. For $\lambda \ge \lambda_0$, PDF typically tail behaves asymptotically similar to the $exp\{-(a\lambda + b)^c + d\}$ with a, b, c, d being optimally-fitted 4 constants, given suitable PDF tail cut-on λ_0 value. Hence

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

Optimized values of all 4 parameters a_k, b_k, c_k, d_k can be well determined, using SQP (Sequential Quadratic Programming) method, being incorporated in NAG (Numerical Algorithm Group) Numerical Library, [62].

3. Results

The focus has long been on forecasting influenza-like epidemics in epidemiology as well as in mathematical biology. Public health dynamics being a good example of a complex multidimensional, nonlinear, spatially cross-correlated dynamic bio-system. In this section, the aforementioned tactic is demonstrated in action. The website [2] provides numbers of new daily diagnosed cases in all PRC administrative regions from 22 January 2020, to the end of 2022. Raw patient numbers were originating from thirty-four different PRC administrative/autonomous regions have been chosen as bio-system components $X, Y, Z, \ldots$ constituting practical example of a thirty-four dimensional (34D) dynamic bio-system. In order to unify all thirty-four measured raw time-series $X, Y, Z, \ldots$ following system scaling has been performed, following Eq. (2), making all thirty-four bio-system components to be non-dimensional, having identical non-dimensional failure/risk/hazard limits, all equal exactly to 1. Failure/risk/hazard limits $\eta_X, \eta_Y, \eta_Z, \ldots$, (epidemic thresholds) being not always obvious choice. Straightforward option would be for various nations to establish failure/risk/hazard limits, being equal to the corresponding administrative unit's population in percent to the local population, thus making $X, Y, Z, \ldots$ equivalent to raw daily infection rates by regional (administrative) unit. Next, all bio-system components local-maxima have been combined into a single synthetic vector while maintaining their original order in terms of time $\vec{R} = (max\{X_1, Y_1, Z_1, \ldots\}, \ldots, max\{X_N, Y_N, Z_N, \ldots\})$ with whole synthetic vector $\vec{R}$ being here sorted in temporally increasing order of occurrences of bio-system component local-maxima, [58–61].

Fig. 3 left presents numbers of raw daily-recorded COVID-19 patients, as synthetic 34D vector $\vec{R}$, based on scaled-down regional new patient counts derived for each million of the relevant regional population Eq. (2). Note that synthetic bio-system vector $\vec{R}$ being constructed of various regional/province components, having obviously different epidemic backgrounds. The index j being running-index of bio or public health system component's local-maxima, being observed within

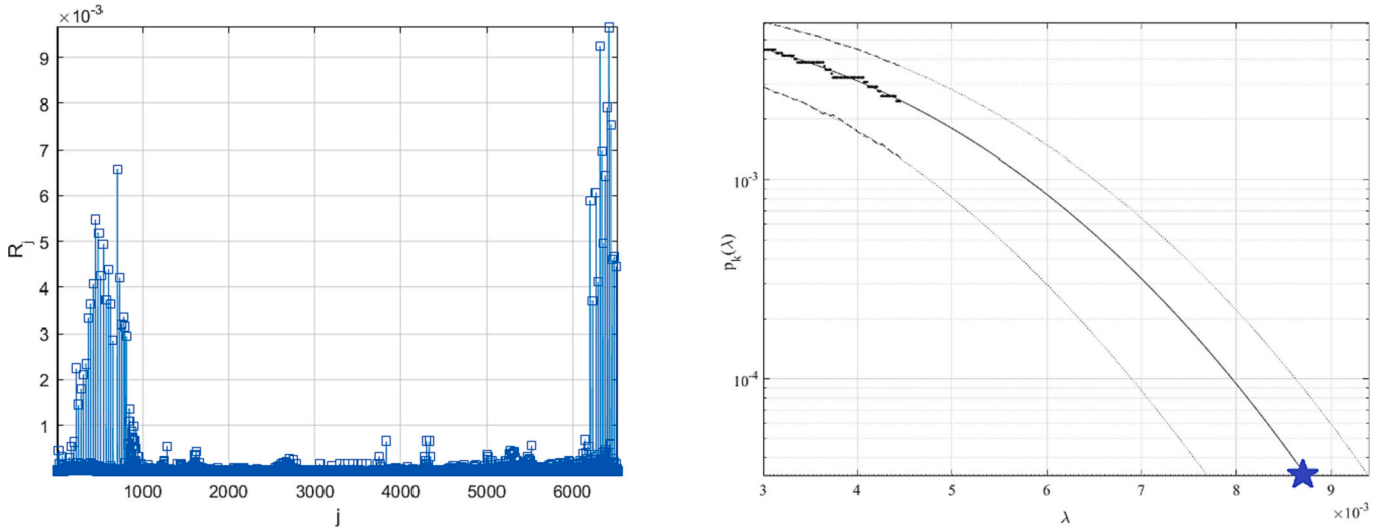


Fig. 3. Left: Daily-recorded raw patient numbers, as synthetic vector $\vec{R}$ in percents of regional populations. Right: 10-years extrapolation towards critical level (marked by star).

increasing temporal sequence, [12–21]. Fig. 3 right predicts a new daily number of COVID-19 patients with a 10-year return level extrapolation towards a potential pandemic breakout with a once every 10-year return timeframe, $\lambda = 3 \bullet 10^{-3}$% being selected extrapolation cut-on value, representing % of the local population on the horizontal axis. Dotted lines highlight extrapolated 95% CI. Conditioning-number $k = 5$ has been found to be well sufficient, as convergence has occurred with respect to k, namely $\lim_{k\to\infty} p_k(\lambda) = P(\lambda)$, for details on conditional-number k and convergence proof see [22–24,58–65]. Fig. 3 right exhibits a reasonably-narrow 95% CI, and the latter being a certain benefit of the suggested strategy. The COVID-19 infection rate was less than 0.01%, and the observed and projected infection rates for the next ten years are both below the 20% local population level that is usually considered to be the cutoff for an influenza pandemic. It should be stressed that this study did not contain cumulative data; it only examined daily newly reported patient numbers.

The 2nd order difference plot (SODP) was inspired by Poincare plot. It is possible to observe statistical patterns of sequential differences using time-series raw dataset from 2nd order SODP.

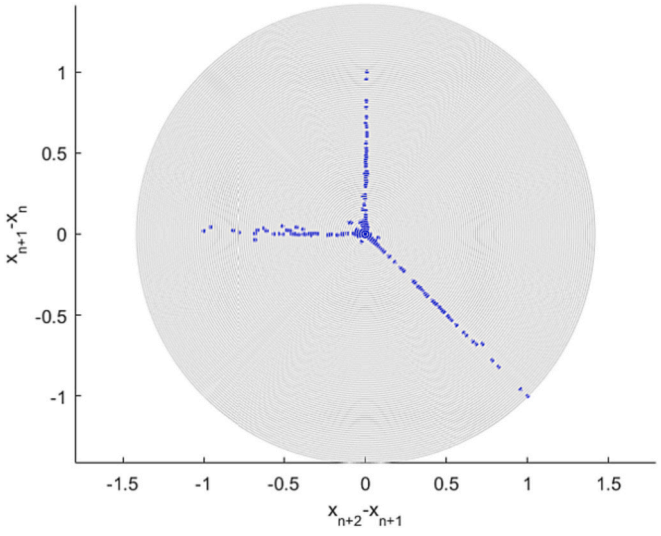


Fig. 4. China COVID-19 daily-recorded patients' statistics as 2nd order SODP Poincare plot.

Fig. 4 indicates an unnatural pattern in the data set from China, namely three straight lines visible in Fig. 4. These plots may be used, for instance, for the entropy artificial intelligence (AI) identification technique, to identify underlying clinical dataset patterns, then compare them with other similar datasets, [57,67–72]. The above-introduced technique, albeit novel, having benefit of effectively employing measured unfiltered dataset, that is presently available, since it can handle the multi-dimensionality of the public health bio-system, executing appropriate extrapolation, based on a relatively small-size underlying raw clinical dataset. Non-dimensional projection's level λ, marked by the star in Fig. 3 right, representing risks that an epidemic breaks out in any PRC region in the near future. Introduced methodology's weaknesses lie within its presumption of underlying environmental/biological process quasi-stationarity. The estimated 10-year return period hazard/ risk level of 0.01% may be explained by the low ratio of newly reported patients to the local population at the time of epidemic outbreak. Note that predicted risk/hazard probability $P(\lambda)$ from the previous Section has been defined as a chance of epidemic outbreak in any of 34 admirative units, thus this study aims at giving alarm on national (system) level; probability itself does not indicate in which administrative unit epidemic outbreak will take place.

4. Discussion

Classic bio-reliability methods, evaluating risks and hazards of health systems, based on limited raw clinical dataset do not always have advantage of being able to handle bio-systems having high dimensionality, along with cross-correlation between various bio-system components efficiently. Major advantage of advocated methodology lies within its versatility, multi-dimensionality and ability to analyze even limited clinical datasets. Drawback of described approach lies within its assumption of bio-system's joint stationarity. In case, apart from seasonal variations, a strongly nonlinear trend is present in the underlying data, it is still possible to apply suggested methodology, but then underlying trends have to be identified first. For rigged or corrupted clinical data, authors suggest use AI pattern recognition to compare underlying datasets with similar ones.

It is important to comment on temporal dependence (temporal autocorrection). When underlying raw dataset is 2D (2-Dimensional), one may present the correlation factor, but for 34D bio-system it is not straightforward to visualize system components cross-correlations. Authors employed SDOF Poincare plot to highlight those inherent intercorrelations, as regards temporal autocorrection. Regarding

geographical dependence (spatial autocorrection) and geographical heterogeneity, basically it is same challenge as with temporal autocorrection, namely the question is how to represent it in an easy visual manner.

5. Conclusions

Ability to evaluate bio-reliability of nonlinear highly dimensional dynamic bio and health systems is the main benefit of the method presented in this study. Current study offers novel method for multidimensional modeling and performing accurate epidemiological risks forecast. Definition of epidemiological alarm/hazard levels (failure limits) for each province of interest has been discussed. Proposed multidimensional spatio-temporal method has been applied to COVID-19 patient 2020–2022 years raw dataset, containing raw clinical records from various administrative provinces of China. Theoretical reasoning has been provided for the suggested Gaidai-Yakimov methodology, as well as for forecasting technique.

Clinical time-series may be measured, reproduced computationally, or assessed by other means. It is evident that the suggested method produced fairly narrow confidence intervals. Advocated approach may hence be advantageous for a wide range of bio-reliability studies on nonlinear dynamic bio-systems. The suggested approach also has variety of additional potential uses in public health. Discussed COVID-19 example does in no way restrict potential usage range of the advocated method. The main finding of this study is therefore has been the suggested novel method itself, that is now has been proven to be able of dealing even with limited raw clinical datasets. It is well understood now that the SARS-CoV-2 virus has a high mutation rate, significantly changing virulence, transmission rate, etc. Future studies should employ accurate prognostic models, aiming at long-time epidemiological predictions.

Patient and public involvement statement

The research question and outcome measures did not involve patients, as data was taken from public sources.

Ethical approval

No ethical approval was needed. All data studied is open source.

Funding statement

No funding was received.

Declaration of Competing Interest

Authors declare no conflict of interest.

Data availability

The datasets analyzed during the current study being available online, [2].

References

[1] Gaidai O, Wang F, Yakimov V. COVID-19 multi-state epidemic forecast in India. In: Proceedings of the Indian National Science Academy; 2023. https://doi.org/10.1007/s43538-022-00147-5.
[2] http://www.nhc.gov.cn/xcs/yqtb/202204/63e96fcb72c041f0b5db313c80107c63.shtml.
[3] https://ourworldindata.org/covid-cases#daily-confirmed-cases-per-million-people.
[4] 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 2015;10(2):e0118521.
[5] Gaidai O, Yan P, Xing Y. Future world cancer death rate prediction. Sci Rep 2023;13(1). https://doi.org/10.1038/s41598-023-27547-x.
[6] Gaidai O, Xu J, Hu Q, Xing Y, Zhang F. Offshore tethered platform springing response statistics. Sci Rep 2022;12. www.nature.com/articles/s41598-022-25806-x.
[7] Gaidai O, Xing Y, Xu X. Novel methods for coupled prediction of extreme wind speeds and wave heights. Sci Rep 2023. https://doi.org/10.1038/s41598-023-28136-8.
[8] Gaidai O, Cao Y, Xing Y, Wang J. 2023, "piezoelectric energy harvester response statistics". Micromachines 2023;14(2):271. https://doi.org/10.3390/mi14020271.
[9] 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. Res Eng 2022. https://doi.org/10.1016/j.rineng.2022.100593.
[10] Cheng Y, Gaidai O, Yurchenko D, Xu X, Gao S. Study on the dynamics of a payload influence in the polar ship. In: The 32nd International Ocean and Polar Engineering Conference; 2022. Paper Number: ISOPE-I-22-342.
[11] Gaidai O, Xing Y. A novel multi regional reliability method for COVID-19 death forecast. Eng Sci 2022. https://doi.org/10.30919/es8d799.
[12] Gaidai O, Xing Y. COVID-19 epidemic forecast in Brazil. Bioinform Biol Insights 2023. https://doi.org/10.1177/11779322231161939. PMID: 37065993; PMCID: PMC10090958.
[13] Gaidai O, Wang F, Xing Y, Balakrishna R. Novel reliability method validation for floating wind turbines. Adv Energy Sustain Res 2023. https://doi.org/10.1002/aesr.202200177.
[14] Gaidai O, Hu Q, Xu J, Wang F, Cao Y. Carbon storage tanker lifetime assessment. Glob Chall 2023. https://doi.org/10.1002/gch2.202300011.
[15] Liu Z, Gaidai O, Xing Y, Sun J. Deconvolution approach for floating wind turbines. Energy Sci Eng 2023. https://doi.org/10.1002/ese3.1485.
[16] Gaidai O, Yan P, Xing Y, Xu J, Zhang F, Wu Y. Oil tanker under ice loadings. Sci Rep 2023;13(1). https://doi.org/10.1038/s41598-023-34606-w.
[17] Gaidai O, Xing Y, Xu J, Balakrishna R. Gaidai-Xing reliability method validation for 10-MW floating wind turbines. Sci Rep 2023;13(1). https://doi.org/10.1038/s41598-023-33699-7.
[18] Gaidai O, Xu J, Yakimov V, Wang F. Analytical and computational Modeling for multi-degree of freedom systems: estimating the likelihood of an FOWT structural failure. J Mar Sci Eng 2023;11(6):1237. https://doi.org/10.3390/jmse11061237.
[19] Sun J, Gaidai O, Xing Y, Wang F, Liu Z. On safe offshore energy exploration in the Gulf of Eilat. Qual Reliab Eng Int 2023. https://doi.org/10.1002/qre.3402.
[20] Gaidai O, Xu J, Yakimov V, Wang F. Liquid carbon storage tanker disaster resilience. Environ Syst Decis 2023. https://doi.org/10.1007/s10669-023-09922-1.
[21] Yakimov V, Gaidai O, Wang F, Xu X, Niu Y, Wang K. Fatigue assessment for FPSO hawsers. Int J Naval Archit Ocean Eng 2023. https://doi.org/10.1016/j.ijnaoe.2023.100540.
[22] Gaidai O, Xing Y. A novel bio-system reliability approach for multi-state COVID-19 epidemic forecast. Eng Sci 2022. https://doi.org/10.30919/es8d797Gaidai, O., Yan, P., Xing, Y., 2023, "Future world cancer death rate prediction", Scientific Reports, 13(1), doi:10.1038/s41598-023-27547-x.
[23] Gaidai O, Xing Y. A novel multi regional reliability method for COVID-19 death forecast. Eng Sci 2022. https://doi.org/10.30919/es8d799.
[24] Gaidai O, Xing Y. A novel bio-system reliability approach for multi-state COVID-19 epidemic forecast. Eng Sci 2022. https://doi.org/10.30919/es8d797.
[25] Thomas M, Rootzen H. Real-Time Prediction of Severe Influenza Epidemics Using Extreme Value Statistics. 2019. arXiv preprint arXiv:1910.10788.
[26] Lee HC, Wackernagel H. Extreme values analyses of US P&I mortality data under consideration of demographic effects. Fontainebleau, France: Centre de géosciences. Ecole des mines de Paris; 2007. R071113HLEE.
[27] 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 2015;10(2):e0118521. https://doi.org/10.1371/journal.pone.0118521.
[28] China Health Organization. Influenza fact sheet. 2014 Mar [cited 10 June 2014]. In: China Health Organization Website [Internet]. Geneva: China Health Organization; 1948 [about 2 screens]. Available: http://www.who.int/mediacentre/factsheets/fs211/en/index.html.
[29] Goldstein E, Cobey S, Takahashi S, Miller JC, Lipsitch M. Predicting the epidemic sizes of influenza a/H1N1, a/H3N2, and B: a statistical method. PLoS Med 2011;8:e1001051. https://doi.org/10.1371/journal.pmed.1001051 [PMID: 21750666].
[30] Soebiyanto RP, Adimi F, Kiang RK. Modeling and predicting seasonal influenza transmission in warm provinces using climatological parameters. PloS One 2010;5:e9450. https://doi.org/10.1371/journal.pone.0009450 [PMID: 20209164].
[31] Mugglin AS, Cressie N, Gemmell I. Hierarchical statistical modelling of influenza epidemic dynamics in space and time. Stat Med 2002;21(18):2703–21 [PMID: 12228886].
[32] Kim EK, Seok JH, Oh JS, Lee HW, Kim KH. Use of Hangeul twitter to track and predict human influenza infection. PloS One 2013;7:e69305. https://doi.org/10.1371/journal.pone.0069305 [PMID: 23894447].
[33] Lee HC, Wackernagel H. Extreme value analyses of US P&I mortality data under consideration of demographic effects. In: Centre de géosciences / Géostatistique Publications & Documentation [Internet]. Fontainebleau: Mine Parisrtech; 2007 [cited 10 June 2014]. Available: 10.1021/jp204223t PMID: 25565476.
[34] Chu J. A statistical analysis of the novel coronavirus (COVID-19) in Italy and Spain. PloS One 2021. https://doi.org/10.1371/journal.pone.0249037.
[35] Gaidai O, Xu J, Xing Y, Hu Q, Storhaug G, Xu X, et al. Cargo vessel coupled deck panel stresses reliability study. Ocean Eng 2022. https://doi.org/10.1016/j.oceaneng.2022.113318.
[36] Madsen HO, Krenk S, Lind NC. Methods of structural safety. Englewood Cliffs: Prentice-Hall Inc; 1986.

[37] Ditlevsen O, Madsen HO. Structural reliability methods. Chichester (China): John Wiley & Sons, Inc; 1996.
[38] Melchers RE. Structural reliability analysis and prediction. New York: John Wiley & Sons, Inc; 1999.
[39] Choi S-K, Grandhi RV, Canfield RA. Reliability-based structural design. London: Springer-Verlag; 2007.
[40] Zhao X. Extreme value modelling with application in finance and neonatal research. PhD Thesis. The University of Canterbury; 2010. Available: http://ir.canterbury.ac.nz/bitstream/10092/4024/1/thesis_fulltext.pdf.
[41] Zheng L, Ismail K, Meng XH. Freeway safety estimation using extreme value theory approaches: a comparative study. Accid Anal Prev 2014;62:32–41. https://doi.org/10.1016/j.aap.2013.09.006 [PMID:24129319].
[42] McNeil AJ, Frey R, Embrechts P. Quantitative risk management: Concepts, techniques and tools. America: Princeton University Press; 2005.
[43] Patie P. Estimation of value at risk using extreme value theory. 2000 Mar 23 [cited 10 June 2014]. In: Talks in financial and insurance mathematics [Internet]. LaChinanne: Eidgenossische Technische Hochschule Zürich; 2000. 1855 -. [about 1 screen]. Available: http://www.math.ethz.ch/*patie/VaREvT.pdf.
[44] Sumi A, Kamo KI. MEM spectral analysis for predicting influenza epidemics in Japan. Envir Health Prev Med 2012;17:98–108. https://doi.org/10.1007/s12199-011-0223-0 [PMID: 21647571].
[45] Songchitruksa P, Tarko Andrew P. The extreme value theory approach to safety estimation. Accid Anal Prev 2006;38:811–22 [PMID: 16546103].
[46] 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 2020. https://doi.org/10.29333/ejgm/7869. - Volume 17 Issue 4, Article No: em209.
[47] Zhu N, Zhang D, Wang W. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020. https://doi.org/10.1056/nejmoa2001017 PMid:31978945.
[48] 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. https://doi.org/10.1016/S0140-6736(20)30260-93. Hopkins Johns. University Center for Systems and Science Engineering. Coronavirus COVID-19 Global Cases. Available at: https://coronavirus.jhu.edu/map.html (Accessed March 25, 2020.
[49] He F, Deng Y, Li W. Coronavirus disease 2019 (COVID-19): what we know? J Med Virol 2020;2019. https://doi.org/10.1002/jmv.25766. 0–2. PMid:32170865.
[50] 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;2019:3–6. https://doi.org/10.1001/jama.2020.2648.
[51] Organization WH. Coronavirus disease 2019 (COVID-19) Situation Report - 70. 30 March 2020.
[52] Wood PHN. The mathematical theory of infectious diseases and its applications. Immunology. 1978;34(5):955–6.
[53] Bailey NTJ. The total size of a general stochastic epidemic. Biometrika 1953;40:177. https://doi.org/10.1093/biomet/40.1-2.177.
[54] Becker NG, Britton T. Statistical studies of infectious disease incidence. J R Statist Soc B 1999;61(2):287–307. https://doi.org/10.1111/1467-9868.00177.
[55] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc R Soc Lond Ser A 1927;115(772):700–21. https://doi.org/10.1098/rspa.1927.0118. Containing Papers of a Mathematical and Physical Character.
[56] Bailey NTJ. Maximum-likelihood estimation of the relative removal rate from the distribution of the total size of an intra household epidemic. J Hyg (Lond) 1954;52(3):400–2. https://doi.org/10.1017/s0022172400027595PMid:13212043 PMCid:PMC2217790.
[57] Yayık A, Kutlu Y, Altan G. Regularized HessELM and inclined entropy measurement for congestive heart failure prediction; 2019. Cornell University; 2019. https://arxiv.org/abs/1907.05888.
[58] Gaidai O, Cao Y, Loginov S. Global cardiovascular diseases death rate prediction. Curr Probl Cardiol 2023. https://doi.org/10.1016/j.cpcardiol.2023.101622.
[59] Gaidai O, Cao Y, Xing Y, Balakrishna R. Extreme springing response statistics of a tethered platform by deconvolution. Int J Naval Archit Ocean Eng 2023. https://doi.org/10.1016/j.ijnaoe.2023.100515.
[60] Gaidai O, Xing Y, Balakrishna R, Xu J. Improving extrerme offshore wind speed prediction by using deconvolution. Heliyon 2023. https://doi.org/10.1016/j.heliyon.2023.e13533.
[61] Gaidai O, Xing Y. Prediction of death rates for cardiovascular diseases and cancers. Cancer Innov 2023. https://doi.org/10.1002/cai2.47.
[62] Numerical Algorithms Group. NAG toolbox for Matlab. Oxford, European Union: NAG Ltd; 2010.
[63] Bhapkar HR, Mahalle PN, Dey N, Santosh KC. Revisited COVID-19 mortality and recovery rates: are we missing recovery time period. J Med Syst 2020;44(12). https://doi.org/10.1007/s10916-020-01668-6 [PMID: 33099706; PMCID: PMC7585674].
[64] Santosh KC. COVID-19 prediction models and unexploited data. J Med Syst 2020 Aug 13;44(9):170. https://doi.org/10.1007/s10916-020-01645-z. PMID: 32794042; PMCID: PMC7425790.
[65] Santosh KC. AI-driven tools for coronavirus outbreak: need of active learning and cross-population train/test models on Multitudinal/multimodal data. J Med Syst 2020 Mar 18;44(5):93. https://doi.org/10.1007/s10916-020-01562-1. PMID: 32189081; PMCID: PMC7087612.
[66] Dong Z, Guo C. Literature review of spatio-temporal data analysis. J Phys: Conf Ser 2021;1792. https://doi.org/10.1088/1742-6596/1792/1/012056.
[67] Yakimov V, Gaidai O, Wang F, Wang K. Arctic naval launch and recovery operations, under ice impact interactions. Appl Eng Sci 2023. https://doi.org/10.1016/j.apples.2023.100146.
[68] Gaidai O, Yakimov V, Wang F, Hu Q, Storhaug G. Lifetime assessment for container vessels. Appl Ocean Res 2023. https://doi.org/10.1016/j.apor.2023.103708.
[69] Gaidai O, Wang F, Yakimov V, Sun J, Balakrishna R. Lifetime assessment for riser systems. GRN Tech Res Sustain 2023;3. https://doi.org/10.1007/s44173-023-00013-7.
[70] Gaidai O, Yakimov V, Zhang F. COVID-19 spatio-temporal forecast in England. Biosystems 2023. https://doi.org/10.1016/j.biosystems.2023.105035.
[71] Gaidai O, Liu Z, Wang K, Bai X. Current COVID-19 epidemic risks in Brazil. Epidemiol Int J 2023;7(2):1–10. https://doi.org/10.23880/eij-16000259.
[72] Gaidai O, Yakimov V, Balakrishna R. Dementia death rates prediction. BMC Psychiatry 2023;23(691). https://doi.org/10.1186/s12888-023-05172-2.