

通过多维可靠性方法预测未来世界 新冠肺炎



Gaidai Oleg, 王科霖*

上海海洋大学工程学院, 上海 200000

摘要: 【目的】新冠肺炎是一种传染性疾病, 具有高传染性, 可在全球范围内传播, 发病率和死亡率都很高, 给全球公共卫生带来了巨大负担。由于流行病波的非平稳性和复杂性, 对这种现象进行建模具有挑战性。很少有数学模型可以使用, 因为流行病数据通常不是正态分布的。【方法】本文描述了一种新的生物系统可靠性方法, 特别适用于多区域环境和卫生系统, 在足够长的时间内观察, 从而对高致病性病毒爆发概率进行可靠的长期预测。传统的多区域过程时间观测统计方法不能有效地处理大区域维数和不同区域观测之间的相互相关性。在这项研究中, 选择了世界所有国家每天记录的新冠肺炎患者人数。【结果】数学算例中典型的流感流行阈值约为当地人口的 20%, 则未来 100 年记录和预测的新冠肺炎感染率均低于 1.4%。【结论】这项工作旨在对最先进的方法进行基准测试, 这使得从动态观察到的患者人数中提取必要的信息成为可能, 同时考虑到相关的地域映射。本文提出的方法开辟了准确预测多区域生物系统流行病爆发概率的可能性。

关键词: 新冠肺炎; 疫情暴发; 可靠性; 概率预测; 动态系统; 公共卫生; 数理生物学

DOI: [10.57237/j.wjms.2022.01.001](https://doi.org/10.57237/j.wjms.2022.01.001)

A Multi-dimensional Reliability Method for Predicting the Future of COVID-19 in the World

Gaidai Oleg, Wang Kelin*

College of Engineering, Shanghai Ocean University, Shanghai 200000, China

Abstract: [Objectives] Novel coronavirus disease 2019 (COVID-19) is a contagious disease with high transmissibility to 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 epidemic waves, it is challenging to model such a phenomenon. Few mathematical models can be used because epidemic data are generally not normally distributed. [Methods] 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 the highly pathogenic virus outbreak probability. 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. For this study, new COVID-19 daily numbers of recorded patients in all 195 world countries were chosen. [Results] In the mathematical example, typical influenza epidemic thresholds are about 20% of the local population, COVID-19 infection rates predicted for any

*通信作者: 王科霖, 2763634512@qq.com

world country in given day for the next 100 years were found less than 1.41%. [Conclusions] This work aims to benchmark state of the art method, which makes it possible to extract the necessary information from dynamically observed new daily patient numbers, while taking into account relevant territorial mapping. The method proposed in this paper opens up the possibility of accurately predicting epidemic outbreak probability for multi-regional biological systems.

Keywords: COVID-19; Epidemic Outbreak; Reliability; Probability Forecast; Dynamic System; Public Health; Mathematical Biology

1 引言

新冠肺炎和其他近期类似流行病的统计方面在现代研究界受到广泛关注[1-8]。一般来说，使用传统的理论统计方法计算现实的生物系统可靠性因子[1-12]和实际流行条件下的爆发概率是相当具有挑战性的[13-24]。后者通常是由于许多程度的系统自由度和控制动态生物系统的随机变量，分布在广阔的地形上。原则上，复杂生物系统的可靠性可以通过足够的测量或直接的蒙特卡罗模拟直接准确地估计。然而，对于新冠肺炎，唯一可用的观察数字受到 2020 年初的限制。受后一种论点的启发，作者为生物和卫生系统引入了一种新的可靠性方法，以更准确地预测和管理流行病爆发。他的研究重点是世界的新冠肺炎[25-31]，重点关注同一气候带内不同国家之间的相互关系。Gondauri 等[32]研究不同国家病例的新冠肺炎病毒传播情况。选择整个世界是因为其新冠肺炎起源以及广泛的健康观察和在线相关研究[33-45]。

生命周期数据或极值理论(EVT)的统计建模在医学或工程学中很普遍。例如，Gumbel 使用 EVT 来估计各种人群的人口状况 [46, 47]。Aarssen[48] 和 Galambos[49]的研究分别论证了关于预期寿命上限支持和反对。通常在这些领域撰写的论文假定从指数分布获得的参数双变量寿命分布以获得统计相关数据 [50-53]。Romeo[54]提出了一种新方法，该方法使用了幂方差函数 copulas（例如，Clayton、Gumbel 和 Inverse Gaussian copulas）、条件采样和用于生存分析的数值逼近。Beisel 等[55]论证了 EVT 已被用于预测进化遗传学中的突变，并进一步从 EVT 中开发一个可能性框架，用于确定突变的适应度效应。Joyce 等[56]在此 EVT 理论的基础上，使用 Beta-Burr 分布确定适应度效应。而 Kristensen 等[57]讨论了使用双变量逻辑回归模型，然后将其用于访问患有行走障碍的多发性硬化症患者以及用于视觉识别的认知实验。

最后，Chen 等[58]的研究使用了 EVT 来估计世界爆发流感的概率，展示了 EVT 预测流行病的潜力。同样，Thomas 等[59]使用 EVT 来预测和检测流感流行的异常情况。由于没有太多的统计研究来预测流感或传染病爆发或传播的可能性，本文提出的多维可靠性方法将能够更好地洞察和预示疾病的传播。

2 模型的建立

在本文中，流行病的爆发被视为可能在任何时间发生在特定国家的任何地区的意外事件，因此空间传播被考虑在内。此外，引入特定的无量纲因子 λ 来预测后者在任何时间、任何地点的流行病风险。

生物系统受到遍历性环境的影响。另一种选择是将过程视为依赖于特定的环境参数，其时间变化可以单独建模为遍历过程。

从公共网站检索 2020 年 2 月至今世界所有国家的新冠肺炎发病率数据[60]。由于这个有价值的数据集是每个世界地区的，所考虑的生物系统可以被视为具有高度相互关联的区域成分/维度的多自由度 (MDOF) 动态系统。在最近的研究中，Chu[9]使用线性对数模型来预测新冠肺炎的发展。

请注意，虽然这项研究的目的是通过预测来减少未来流行病爆发的风险，但它只关注每日登记的患者人数，而不是症状本身。Sudre 等[61]研究了长期持续的新冠肺炎症状，即所谓的“长期新冠肺炎”及其风险因素，并讨论了是否有可能在初期预测该疾病。Elizabeth 等[8]研究了新冠肺炎致死相关因素，并量化了一系列与新冠肺炎致死相关的临床因素。图 1 介绍了世界各国家的地图。

Daily new confirmed COVID-19 cases, Apr 26, 2022

Our World
in Data

7-day rolling average. Due to limited testing, the number of confirmed cases is lower than the true number of infections.

World

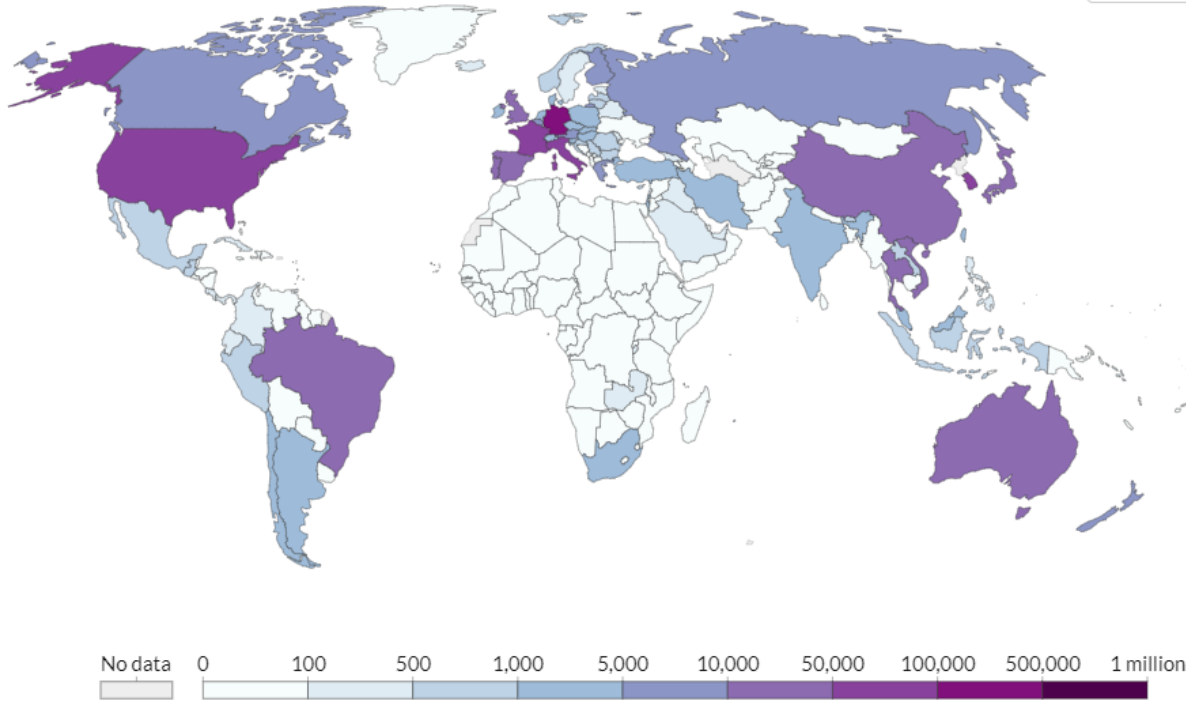


图 1 世界的地图

Figure 1 Map of the World with COVID-19 cases

在足够长的时间间隔 $(0, T)$ ，测量的 MDOF 健康响应向量过程为 $R(t) = (X(t), Y(t), Z(t), \dots)$ 。整个时间跨度 $(0, T)$ 上的一维全局最大值表示为， $X_T^{\max} = \max_{0 \leq t \leq T} X(t)$ ， $Y_T^{\max} = \max_{0 \leq t \leq T} Y(t)$ ， $Z_T^{\max} = \max_{0 \leq t \leq T} Z(t)$ ，...

足够长的时间 T 主要是指相对于动态系统自相关时间的较大值。令 X_1, \dots, X_{N_X} 为过程 $X(t)$ 离散单调递增

时刻 $t_1^X < \dots < t_{N_X}^X$ 在 $(0, T)$ 的时间局部最大值

对于其他 MDOF 响应分量，类似的定义如下。为简单起见，所有分量，因此它的最大值被假定为非负数。

$TTX_1, \dots, X_{N_X} X(t) t_1^X < \dots <$

$t_{N_X}^X (0, T) Y(t), Z(t), \dots Y_1, \dots, Y_{N_Y}; Z_1, \dots, Z_{N_Z} R(t)$

目标是准确估计系统故障概率，即超越概率：

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

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

P 是响应分量 $\eta_X, \eta_Y, \eta_Z, \dots$ 的临界值不超过的概率； \cup 表示逻辑统一运算“or”；和 $p_{X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \dots}$ 是整个时间跨度 $(0, T)$ 上全局最大值的联合概率密度。

然而，由于后者的高维度和可用数据集的限制，直接估计后者的联合概率分布是不可行的。更具体地说，当 $X(t)$ 超过 η_X ，或 $Y(t)$ 超过 η_Y 或 $Z(t)$ 超过 η_Z 等时刻，系统被视为立即失效。固定失效水平 $\eta_X, \eta_Y, \eta_Z, \dots$ 当然，当然对于 $R(t)$ 的每个一维响应分量都是独立的，

如 $X_{N_X}^{\max} = \max \{X_j; j = 1, \dots, N_X\} = X_T^{\max}$ ， $Y_{N_Y}^{\max} = \max \{Y_j; j = 1, \dots, N_Y\} = Y_T^{\max}$ ， $Z_{N_Z}^{\max} = \max \{Z_j; j = 1, \dots, N_Z\} = Z_T^{\max}$

现在，局部最大值时刻 $[t_1^X < \dots < t_{N_X}^X; t_1^Y < \dots < t_{N_Y}^Y; t_1^Z < \dots < t_{N_Z}^Z]$ 以单调非递减的顺序排列成一个单一的合并时间向量 $t_1 \leq \dots \leq t_N$ 。注意 $t_N = \max \{t_{N_X}^X, t_{N_Y}^Y, t_{N_Z}^Z, \dots\}$ ， $N = N_X + N_Y + N_Z + \dots$ 。在这种情况下， t_j 表示 MDOF 结构响应分量的局部最大值

$X(t)$ 或 $Y(t)$,或 $Z(t)$ 等。这意味着具有 $R(t)$ 的时间记录,只需连续并同时筛选一维响应分量局部最大值,并记录其超过MDOF极限向量 $(\eta_X, \eta_Y, \eta_Z, \dots)$ 的任何分量 X, Y, Z, \dots 。在合并的时间向量 $t_1 \leq \dots \leq t_N$ 之后,将局部一维响应分量最大值合并为一个时间非递减向量 $\vec{R} = (R_1, R_2, \dots, R_N)$ 。即,每个局部最大值 R_j ,是实际遇到的局部最大值对应于 $X(t)$ 或 $Y(t)$,或 $Z(t)$ 等。最后,引入统一的极限向量 (η_1, \dots, η_N) ,每个分量 η_j 都是 η_X, η_Y 或 η_Z 等,取决于

$X(t)$ 或 $Y(t)$,或 $Z(t)$ 等中的哪一个对应于当前运行指数 j 的局部最大值

现在,引入缩放参数 $0 < \lambda \leq 1$,人为地同时降低所有响应分量的极限值,即新的MDOF极限向量 $(\eta_X^\lambda, \eta_Y^\lambda, \eta_Z^\lambda, \dots)$,其中 $\eta_X^\lambda \equiv \lambda \cdot \eta_X, \eta_Y^\lambda \equiv \lambda \cdot \eta_Y, \eta_Z^\lambda \equiv \lambda \cdot \eta_Z, \dots$ [62]。引入统一的极限向量 $(\eta_1^\lambda, \dots, \eta_N^\lambda)$,每个分量为 $\eta_j^\lambda, \eta_X^\lambda, \eta_Y^\lambda$ 或 η_Z^λ ,以此类推。后者自动将概率 $P(\lambda)$ 定义为 λ 的函数,从公式(1)中可以看出 $P \equiv P(1)$ 。不超过概率 $P(\lambda)$ 可以估计如下

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

在下文中,概述了基于条件的串联近似背后的原理,第一个近似是一步记忆近似,因此与马尔可夫链近似有一些相似之处。然而,需要强调的是,这种第

一近似并不等同于这样的近似[63]。在实践中,相邻的 R_j 之间的依赖性是不可忽视的;因此,引入了下一步(将被称为条件水平 $k=1$)记忆近似法

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

对于 $2 \leq j \leq N$ (条件等级 $k=2$)。公式(4)引入的近似值可以进一步表示为

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

其中 $3 \leq j \leq N$ (将称为条件水平 $k=3$),以此类推不同条件水平 k 的取值[64]。目的是及时监测本地首先发生的每个独立故障,从而避免串联的本地相互关联的超出。

公式(5)提出了统计独立性假设的后续改进。后一种近似以更高的精度捕获相邻最大值之间的统计依赖性效应。由于初始MDOF过程 $R(t)$ 被假定为遍历的,因此是平稳的,概率 $p_k(\lambda) := \text{Prob}\{R_j > \eta_j^\lambda \mid R_{j-1} \leq \eta_{j-1}^\lambda, R_{j-k+1} \leq \eta_{j-k+1}^\lambda\}$,对于 $j \geq k$ 将独立于 j 但仅取决于条件 k 。因此,不超过概率可以近似为平均条件超过率方法[65]

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

注意,公式(6)来自于公式(1),通过忽略 $\text{Prob}(R_1 \leq \eta_1^\lambda) \approx 1$ 得到的,由于设计失效概率必须很小,也假设 $N \gg k$ 。公式(6)类似于众所周知的平均上交率方程超过概率[62, 65]。条件参数 k 有明显的收敛性

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

注意,对于 $k=1$ 的公式(6)变成一个众所周知的非超限概率关系,其平均上行速率函数为

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

其中 $v^+(\lambda)$ 表示上述由缩放的MDOF系统响应 $(\frac{x}{\eta_X}, \frac{y}{\eta_Y}, \frac{z}{\eta_Z}, \dots)$ 组合而成的非维向量 $R(t)$ 的响应水平 λ 的平均上交率。公式(8)中平均上交率由Rice公式[66]给出, p_{RR} 是 (R, \dot{R}) 的联合概率密度, \dot{R} 是时间导数 $R'(t)$ 。公式(8)依赖于泊松假设,即高 λ 水平(在本文中是 $\lambda \geq 1$)的上交事件可以被假定为是独立的。后者可能不适用于窄带响应和更高级别的动力系统,这些系统在不同的维度上表现出串联故障,随后是由极端事件之间的内在相互依赖性引起的,表现为在集合向量 $\vec{R} = (R_1, R_2, \dots, R_N)$ 内出现高度相关的局部最大值集群。

上文中,已经使用了平稳性假设。然而,所提出的方法也可以处理非平稳情况。对于非平稳情况,在 $m=1, \dots, M$ 个季节性流行情况的散点图,每个短期季节性状态的概率为 q_m ,所以 $\sum_{m=1}^M q_m = 1$ 。接下来,让我们介绍一下长期方程

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

$p_k(\lambda, m)$ 与公式(7)中的函数相同, 但对应于一个特定的短期季节性流行状态, 数字为 m 。

请注意, 作者在前几年成功验证了针对各种一维动态系统的建议方法的准确性[62, 64]。

下面简单介绍下外推法[62, 65], 作为失效概率分布尾部外推的基础, 渐近为 Gumbel 分布类型。后一种方法假设在一般情况下外推所需的参数函数类可以建模为类似于 Gumbel 分布和一般极值(GEV)分布之间的关系。

从采样时间序列中观察到的可用极值并不一定构成渐近分布, 或者至少证明它们是真正渐近的是一项不平凡的任务。这意味着将我们感兴趣的领域扩展到亚渐近水平的相关性。上面介绍的函数通常在尾部是规则的, 特别是对于 $p_k(\lambda)$ λ 接近和超越 1.更准确地说, 对于 $\lambda \geq \lambda_0$, 分布尾部的行为类似于 $\exp\{-(a\lambda + b)^c + d\}$ 其中 a, b, c, d 是适当拟合的常数, 用于适当的尾部截断 λ_0 值。因此, 可以得到

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

接下来, 通过绘制 $\ln\{\ln(p_k(\lambda)) - d_k\}$ 与 $\ln(a_k\lambda + b_k)$, 通常可以观察到几乎完美的线性尾部行为。

通过最小化以下误差函数 F , 在对数水平上进行优化是很有帮助的, 相对于四个参数 a_k, b_k, c_k, p_k, q_k 而言

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

其中 λ_1 为合适的分布尾部截断值, 即最显著的波高值, 置信区间宽度仍可接受。参数 a_k, b_k, c_k, p_k, q_k 的最佳值也可以使用 NAG 数值库中包含的顺序二次规划(SQP)方法来确定[65]。

权重函数 ω 可以定义为 $\omega(\lambda) = \{\ln CI^+(\lambda) - \ln CI^-(\lambda)\}^{-2}$, 其中 $(CI^-(\lambda), CI^+(\lambda))$ 是一个置信区间(CI), 由模拟或测量数据集的经验估计[62, 64]。当参数 $c = \lim_{k \rightarrow \infty} c_k$ 等于 1 或接近于 1 时, 分布就接近于 Gumbel 分布。

对于任何一般的遍历波高或风速过程, 可以假设超过阈值 λ 的条件超标序列构成泊松过程; 但是, 通常是非同质的。因此, 对于接近 1 的 λ 水平, $p_k(\lambda)$ 的 p -%置信区间(CI)的近似极限可以如下给出

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

其中 $f(p)$ 是根据逆正态分布估计的, 例如 $f(90\%) = 1.65, f(95\%) = 1.96$ 。其中 N 是在分析向量 \vec{R} 中组装的局部最大值的总数。

3 世界的新冠 COVID-19

流感样流行病的预测长期以来一直是流行病学和数学生物学关注的焦点。众所周知, 公共卫生动态是一个高度非线性的多维和空间互相关动态系统, 分析起来总是具有挑战性。以前的研究使用了多种方法来模拟流感样病例。本节说明了上述方法的效率, 该方法使用应用于现实生活中的新冠肺炎数据集的新方法, 呈现为新的每日记录的感染患者时间序列, 分布在大范围内。

新冠肺炎和流感是具有高度传播性的传染病, 可在全世界范围内传播, 具有相当高的发病率和死亡率。它们最常在秋末、冬季和初春季节性发生, 主要在冬季达到其流行的高峰。由甲型和乙型流感病毒引起的季节性流感流行通常每年发生在温带地区的冬季, 给全世界的公共卫生带来了巨大的负担, 据世界卫生组织(WHO)称[2], 全世界每年约有 300-500 万例严重疾病和 25-50 万例死亡。为了统一所有三个测量的时间序列 X, Y, Z, \dots , 进行了以下的缩放:

$$X \rightarrow \frac{X}{\eta_X}, Y \rightarrow \frac{Y}{\eta_Y}, Z \rightarrow \frac{Z}{\eta_Z}, \dots \quad (13)$$

使所有三个响应都是无量纲的, 并且具有相同的失效限制等于 1。失效限制 $\eta_X, \eta_Y, \eta_Z, \dots$, 或者换句话说, 流行病阈值, 不是明显的选择。最简单的选择是, 不同的国家设定的失效限制等于相应国家人口占当地人口的百分比, 基本上使 X, Y, Z, \dots 等于每个国家每天感染的百分比。

接下来, 将三个测量时间序列中的所有局部最大值合并为一个时间序列, 方法是保持时间非递减顺序: $\vec{R} = (\{X_1, Y_1, Z_1, \dots\}, \dots, \{X_N, Y_N, Z_N, \dots\})$, 其中每个集合 $\{X_j, Y_j, Z_j, \dots\}$ 是根据这些局部最大值出现的非递减次数进行排序。

图 2 以 195D 向量 \vec{R} 表示每天新记录的病人数量, 由相应区域人口中每百万人口中的每日新增患者人数组成, 按公式(13)进行缩放。请注意, 向量 \vec{R} 是由具有不同流行病背景的不同区域成分组成的。索引 j 只是在非递减的时间序列中遇到的局部最大值的运行索引。

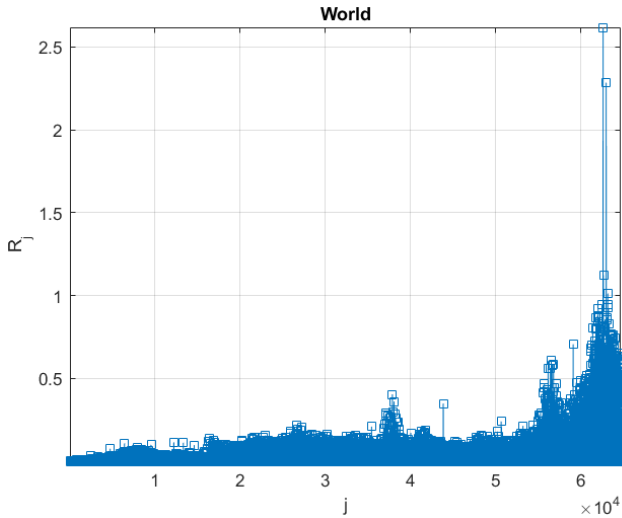


图2 以 195D 向量 \vec{R} 表示的每日新记录患者数量占相应区域人口的百分比

Figure 2 Number of new daily recorded patients as 195D vector \vec{R} . Scaled by Eq. (12) in percent of corresponding country population

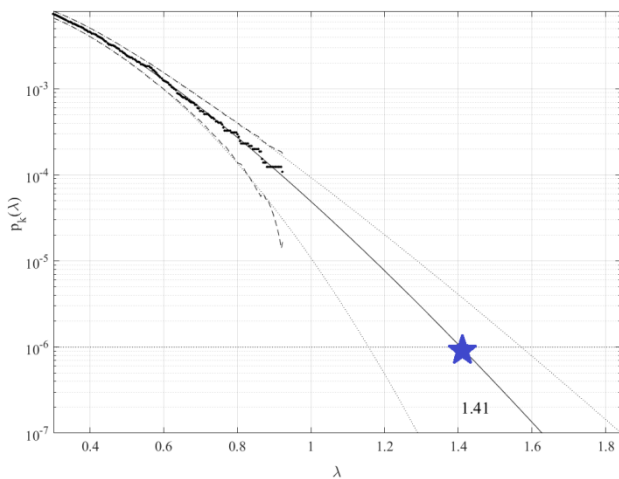


图3 $p_k(\lambda)$ 向临界水平(用星号表示)的 100 年回报水平外推占当地人口的百分比, 虚线表示外推的 95%CI, 横轴为当地人口的百分比

Figure 3 100 years return level extrapolation of $p_k(\lambda)$ towards critical level (indicated by star) in percent to local population. Extrapolated 95% CI indicated by dotted lines. Percentage of local population on horizontal axis

图3显示了根据公式(10)对流行病爆发的100年回归期的推断, 由水平虚线表示, 并在一定程度上超越了, 使用 $\lambda = 3 \cdot 10^{-3}\%$ 临界值, 横轴为当地人口百分比。虚线表示根据公式(11)推断的95%置信区间。根据公式(6)中的 $p(\lambda)$ 与公式(1)中的目标失效概率 $1 - P$ 直接相关。因此, 根据公式(6)可以估算出系统失效概率 $1 - P \approx 1 - P_k(1)$ 。注意, 在公式(6)中,

N 对应于统一响应向量 \vec{R} 中局部极大值的总数。由于出现了与 k 有关的收敛, 故条件参数 $k = 5$ 已经足够, 见公式(7)。图3显示了合理狭窄的95%CI, 后者是所提方法的一个优点。

典型的流感流行阈值约为当地人口的20%[51], 未来100年记录和预测的新冠肺炎感染率均低于1.4%。请注意, 在这项研究中, 仅研究了每日新增的患者人数, 而不是累计人数。

请注意, 上述方法新颖, 但由于其能够对卫生系统多维度进行处理, 并根据非常有限的数据集进行准确的外推, 故可以有效地利用可用的测量数据集。需要注意的是, 图3中的预测无量纲 λ 水平(星号表示), 代表未来几年世界任何地区发生疫情的概率。

4 结论

传统的时间序列卫生系统可靠性分析方法并不具备有效处理拥有高维度和不同系统响应之间交叉相关的系统的优势。该方法的主要优点是能够研究高维非线性动力系统的可靠性。

尽管算例简单, 但本研究成功地提供了一种新颖的多维建模策略和理论方法来实施流行病的预测。已经讨论了如何正确设置每个国家的流行病学警报限值(故障限值)。

本文以世界各国记录的新冠肺炎病例(2020.1.22-2022.4.6)数据集, 构建了一个全维度(195D)模型。将该方法作为一个多维实时系统应用于每日新增患者人数。文中给出了该方法背后的理论推理。注意, 直接使用测量或蒙特卡罗模拟进行动态生物系统可靠性分析是有吸引力的, 然而, 动态系统的复杂性及其高维性要求开发具有强鲁棒性和准确的技术, 以处理有限的数据集, 尽可能有效地利用可用数据。

主要结论为: 如果世界当地环境和流行病学条件下的公共卫生系统管理得当。预测100年重现期风险水平为1.4%, 因为每日新增患者占当地流行病爆发人群的百分比比较低。因此, 在当前的国家卫生管理条件和动态清零政策下, 新冠肺炎不会对中国卫生系统构成任何未来威胁。

本研究旨在进一步开发一种通用、稳定且直观的多维可靠性方法。本文介绍的方法之前已通过应用到广泛的仿真模型进行验证, 适用于一维系统响应, 通常可以获得非常准确的预测, 测量和数值模拟的时间序列响应都可以进行分析。结果表明, 所提出的方法

产生了合理的置信区间。因此，所提出的方法可能成为各种非线性动态生物系统可靠性研究的合适工具。最后，此方法可应用许多公共卫生系统，所呈现的新冠肺炎示例不会以任何方式限制新方法的适用范围。

参考文献

- [1] Chen J., Lei X., Zhang L., Peng B., 2015, Using Extreme Value Theory Approaches to Forecast the Probability of Outbreak of Highly Pathogenic Influenza in Zhejiang, China, PLoS ONE 10 (2): e0118521. doi: 10.1371/journal.pone.0118521
- [2] 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]. Available: <http://www.who.int/mediacentre/factsheets/fs211/en/index.html>.
- [3] 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. doi: 10.1371/journal.pmed.1001051 PMID: 21750666.
- [4] Soebiyanto RP, Adimi F, Kiang RK. Modeling and predicting seasonal influenza transmission in warm countries using climatological parameters. PLoS One. 2010; 5: e9450. doi: 10.1371/journal.pone.0009450 PMID: 20209164.
- [5] Mugglin AS, Cressie N, Gemmell I. Hierarchical statistical modelling of influenza epidemic dynamics in space and time. Stat Med. 2002; 21 (18): 2703–2721. PMID: 12228886.
- [6] 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. doi: 10.1371/journal.pone.0069305 PMID: 23894447.
- [7] Lee HC, Wackernagel H. Extreme value analyses of US P&I mortality data under consideration of demographic effects. 2007 [cited 10 June 2014]. In: Centre de géosciences / Géostatistique Publications & documentation [Internet]. Fontainebleau: Mine ParisTech. Available: http://cg.ensmp.fr/bibliotheque/public/LEE_Rapport_00600.pdf. doi: 10.1021/jp204223t PMID: 25565476
- [8] Elizabeth J. et al. 2020, “Factors associated with COVID-19-related death using OpenSAFELY”, Nature, Vol 584(20), <https://doi.org/10.1038/s41586-020-2521-4>
- [9] 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>
- [10] Gnedenko EM. Sur la distribution limite du terme maximum d’une série aléatoire. Ann Math. 1943; 44 (3):423–453.
- [11] Falzarano, J., Su, Z., Jamnongpipatkul, A. (2012), “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.
- [12] Su, Z. 2012, “Non-linear response and stability analysis of vessel rolling motion in random waves using stochastic dynamical systems”, Texas University, Texas.
- [13] Madsen HO., Krenk S., Lind NC., 1986. “Methods of structural safety”. Englewood Cliffs, Prentice-Hall Inc.
- [14] Ditlevsen O., Madsen HO., 1996. “Structural reliability methods”. Chichester (World): John Wiley & Sons, Inc.
- [15] Melchers RE. 1999, “Structural reliability analysis and prediction”. New York, John Wiley & Sons, Inc.
- [16] Choi S-K, Grandhi RV, Canfield RA., 2007. “Reliability-based structural design”, London, Springer-Verlag.
- [17] Gumbel EM. Statistics of extremes. Columbia: Columbia University Press; 1958.
- [18] Pickands J. Statistical inference using extreme order statistics. Ann Stat. 1975; 3 (1): 119–131.
- [19] 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.
- [20] Zheng L, Ismail K, Meng XH. Freeway safety estimation using extreme value theory approaches: A comparative study. Accident Anal Prev. 2014; 62: 32–41. doi: 10.1016/j.aap.2013.09.006 PMID: 24129319
- [21] McNeil AJ, Frey R, Embrechts P. Quantitative risk management: Concepts, techniques and tools. America: Princeton University Press; 2005.
- [22] 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]. Eidgenössische Technische Hochschule Zürich 1855 -. [about 1 screen]. Available: <http://www.math.ethz.ch/~patie/VaREvT.pdf>, 2000.
- [23] Sumi A, Kamo KI. MEM spectral analysis for predicting influenza epidemics in Japan. Envir Health Prev Med. 2012; 17: 98–108. doi: 10.1007/s12199-011-0223-0 PMID: 21647571.

- [24] Songchitruksa P, Tarko Andrew P. The extreme value theory approach to safety estimation. *Accident Anal Prev.* 2006; 38: 811–822. PMID: 16546103.
- [25] Singanayagam, Anika, et al. "Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, World, January to May 2020." *Eurosurveillance* 25.32 (2020): 2001483. DOI: 10.2807/1560-7917.ES.2020.25.32.2001483.
- [26] World, J. T., et al. "Weathering the COVID-19 storm: Lessons from hematologic cytokine syndromes." *Blood Reviews* (2020): 100707. DOI: 10.1016/j.envpol.2021.116576.
- [27] 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. DOI: 10.48550/arXiv.2103.01742
- [28] 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.
- [29] 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 World, UK[J]. *The British Journal of Social Work*, 2021. DOI: 10.1093/bjsw/bcab166.
- [30] Mahase E. Covid-19: Is the UK heading for another omicron wave?. 2022. *BMJ* 2022; 376 doi: <https://doi.org/10.1136/bmj.o738>
- [31] Rutter M, Lanyon P C, Grainge M J, et al. COVID-19 infection, admission and death among people with rare autoimmune rheumatic disease in World: results from the RECORDER project [J]. *Rheumatology*, 2021. DOI: 10.1093/rheumatology/keab794.
- [32] 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 - Volume 17 Issue 4, Article No: em209, <https://doi.org/10.29333/ejgm/7869>
- [33] Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med.*2020. <https://doi.org/10.1056/nejmoa2001017> PMID: 31978945.
- [34] 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](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).
- [35] He F, Deng Y, Li W. Coronavirus Disease 2019 (COVID-19): What we know? *J Med Virol.* 2020; 2019:0-2. <https://doi.org/10.1002/jmv.25766> PMID:32170865
- [36] 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>
- [37] 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): 565-74. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)
- [38] Zhou P, Yang XL, Wang XG, 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> PMID:32015507 PMCID: PMC7095418
- [39] Organization WH. Coronavirus disease 2019 (COVID-19) Situation Report - 70. 30 March 2020.
- [40] Wood PHN. The Mathematical Theory of Infectious Diseases and its applications. *Immunology.* 1978; 34 (5): 955-6.
- [41] Bailey NTJ. The total size of a general stochastic epidemic. *Biometrika,* 1953a; 40: 177. <https://doi.org/10.1093/biomet/40.1-2.177>
- [42] 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>
- [43] Lan L, Xu D, Ye G, et al. Positive RT-PCR Test Results in Patients Recovered from COVID-19. *JAMA.* Published online February 27, 2020. <https://doi.org/10.1001/jama.2020.2783> PMID:32105304
- [44] 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): 700-21. <https://doi.org/10.1098/rspa.1927.0118>
- [45] 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/s0022172400027595> PMID:13212043 PMCID: PMC2217790
- [46] Gumbel, E. (1958). *Statistics of extremes*. New York: Columbia University Press.
- [47] Gumbel, E. J. (1937). *La duree extreme de la vie humaine* (Vol. 520). Hermann et cie.

- [48] Aarssen, K. & De Haan, L. (1994). On the maximal life span of humans. *Mathematical Population Studies* 4 (4): 259–281. doi: 10.1080/08898489409525379.
- [49] Galambos, J. & Macri, N. (2000). The life length of humans does not have a limit. *Journal of Applied Statistical Science* 9 (4): 253–264.
- [50] Block, H. W., & Basu, A. P. (1974). A continuous, bivariate exponential extension. *Journal of the American Statistical Association*, 69(348), 1031-1037.
- [51] Rakocevic B, Grgurevic A, Trajkovic G, et al. Influenza surveillance: determining the epidemic threshold for influenza by using the Moving Epidemic Method (MEM), Montenegro, 2010/11 to 2017/18 influenza seasons [J]. *Eurosurveillance*, 2019, 24 (12).
- [52] Sarkar, S. K. (1987). A continuous bivariate exponential distribution. *Journal of the American Statistical Association*, 82 (398), 667-675.
- [53] Gupta, R. D., & Kundu, D. (1999). Theory & methods: Generalised exponential distributions. *Australian & New Zealand Journal of Statistics*, 41 (2), 173-188.
- [54] Romeo, J.S., Meyer, R. & Gallardo, D.I. Bayesian bivariate survival analysis using the power variance function copula. *Lifetime Data Anal* 24, 355–383 (2018). <https://doi.org/10.1007/s10985-017-9396-1>
- [55] Beisel, C. J., Rokyta, D. R., Wichman, H. A., & Joyce, P. (2007). Testing the extreme value domain of attraction for distributions of beneficial fitness effects. *Genetics*, 176(4), 2441-2449.
- [56] Joyce, P., & Abdo, Z. (2018). Determining the distribution of fitness effects using a generalised Beta-Burr distribution. *Theoretical Population Biology*, 122, 88-96.
- [57] Kristensen, S. B., & Bibby, B. M. (2020). A bivariate logistic regression model based on latent variables. *Statistics in Medicine*, 39 (22), 2962-2979.
- [58] Chen, J., Lei, X., Zhang, L., & Peng, B. (2015). Using extreme value theory approaches to forecast the probability of outbreak of highly pathogenic influenza in Zhejiang, China. *PloS one*, 10 (2), e0118521.
- [59] Thomas, M., & Rootzen, H. (2019). Real-time prediction of severe influenza epidemics using Extreme Value Statistics. *arXiv preprint arXiv:1910.10788*.
- [60] Hannah Ritchie, Edouard Mathieu, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Esteban Ortiz-Ospina, Joe Hasell, Bobbie Macdonald, Diana Beltekian and Max Roser (2020) - "Coronavirus Pandemic (COVID-19)". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/coronavirus' [Online Resource]<https://ourworldindata.org/covid-cases#daily-confirmed-cases-per-million-people>
- [61] Sudre C., et al. Attributes and predictors of long COVID, *Nature Medicine*, 2021, VOL 27, pp. 626–631, <https://doi.org/10.1038/s41591-021-01292-y>
- [62] Naess A., Moan T. 2013, “Stochastic dynamics of marine structures”, Cambridge University Press.
- [63] Karpa O., 2015 “Development of bivariate extreme value distributions for applications in marine technology”, PhD thesis, Norwegian University of Science and Technology.
- [64] Naess, A., Gaidai, O., 2009, “Estimation of extreme values from sampled time series” *Structural Safety*. Vol 31, No 4, pp 325-334.
- [65] Numerical Algorithms Group, 2010. *NAG Toolbox for Matlab*. Oxford, World: NAG Ltd.
- [66] Rice, S. O. 1944. “Mathematical analysis of random noise”. *Bell System Tech. J.* 23: 282–332.